Continuous attractor working memory and provenance of channel models

THÈSE N^{O} 7845 (2017)

PRÉSENTÉE LE 22 SEPTEMBRE 2017
À LA FACULTÉ DES SCIENCES DE LA VIE
LABORATOIRE DE CALCUL NEUROMIMÉTIQUE (IC/SV)
PROGRAMME DOCTORAL EN NEUROSCIENCES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Alexander Kevin SEEHOLZER

acceptée sur proposition du jury:

Prof. C. Petersen, président du jury
Prof. W. Gerstner, directeur de thèse
Dr G. Mongillo, rapporteur
Prof. A. Compte, rapporteur
Prof. F. Schürmann, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2017

“The entire universe has been
neatly divided into things to
(a) mate with,
(b) eat,
(c) run away from,
and (d) rocks.”
— Terry Pratchett, Equal Rites

To Polona and Liam.

Acknowledgements

First I would like to thank my supervisor, Prof. Wulfram Gerstner, for giving me the chance of coming to the LCN, where the ideas float through the corridors. Throughout the short internship and my subsequent Ph.D., he always had an open door for discussions and placed trust in my work and abilities. Equally, I want to thank Moritz Deger, who collaborated tightly with me on several projects. The ICGenealogy project could not have happened without the conviction and drive that Tim Vogels put forward. Essential was also the non-questioning dedication and Zen of William Podlaski, who kept us all from taking things too seriously at times. I want to thank both of them, as well as Rajnish Ranjan, for the tight collaboration ongoing since several years. I again have to thank Wulfram for letting me partake in this sideproject-turned-fulltime.

What was special to me about the LCN is the degree of collegiality: we might take this for granted, but I hear it seems to be rare for labs to provide an environment in which you do not feel you have to outcompete each other. For making this a tight-knit bunch of colleagues and friends, I want to thank (roughly in lab-chronology): Richard Naud, Eleni Vasilaki, Nicolas Fremaux, Guillaume Hennequin, Mohammadjavad Faraji, Danilo Rezende, Tilo Schwalger, David Kastner, Olivia Gozel, Vasiliki Liakoni, Marco Lehmann, Aditya Gilra. In particular, I would like to thank my musical office for collaborations and discussions: Samuel(e) Muscinelli, Johanni Brea, and Florian Colombo. Special thanks go out to some lab mates, in addition to the above, for an unforgettable period of spike synchronization: Christian Pozzorini, Friedemann Zenke, Felipe Gerhard, Kerstin Preuschoff, Lorric Ziegler, Carlos Stein NDB, Moritz Deger, Tim Vogels, Skander Mensi, Dane Corneil, Everton Agnes. Dopamine also lead to a strong assembly formation with Daniel Miner and Cliff(ord) Kerr.

Let's thank some family and friends! My parents, Gabi and Hans, who always supported me although they never understood what I'm doing for a living. My sisters, especially Niki, who I had to shield from suitors whenever she came to visit. I want to thank Simone Seeger, for years of support. My friends, Michael Naab, Marcus Jost, and Arne Geisel, for staying close although I'm not around often. Big thanks go out to all my friends at the Catalyst, who gave me a life next to family and the Ph.D., and taught me how to improvise, organize, and overcome stage-fright.

And, finally, I want to thank Polona. For being a loving partner. For readily approximating the single-mom state during my writing of this thesis. For grounding me when I need it. And, of course, for having a family with me. ❤

Lausanne, 15 May 2017 A. S.

Abstract

The brain is a complex biological system composed of a multitude of microscopic processes, which together give rise to computational abilities observed in everyday behavior. Neuronal modeling, consisting of models of single neurons and neuronal networks at varying levels of biological detail, can synthesize the gaps currently hard to constrain in experiments and provide mechanistic explanations of how these computations might arise. In this thesis, I present two parallel lines of research on neuronal modeling, situated at varying levels of biological detail.

First, I assess the provenance of voltage-gated ion channel models in an integrative meta-analysis that investigates a backlog of nearly 50 years of published research. To cope with the ever-increasing volume of research produced in the field of neuroscience, we need to develop methods for the systematic assessment and comparison of published work. As we demonstrate, neuronal models offer the intriguing possibility of performing automated quantitative analyses across studies, by standardized simulated experiments. We developed protocols for the quantitative comparison of voltage-gated ion channels, and applied them to a large body of published models, allowing us to assess the variety and temporal development of different models for the same ion channels over the time scale of years of research. Beyond a systematic classification of the existing body of research made available in an online platform, we show that our approach extends to large-scale comparisons of ion channel models to experimental data, thereby facilitating field-wide standardization of experimentally-constrained modeling.

Second, I investigate neuronal models of working memory (WM). How can cortical networks bridge the short time scales of their microscopic components, which operate on the order of milliseconds, to the behaviorally relevant time scales of seconds observed in WM experiments? I consider here a candidate model: continuous attractor networks. These can implement WM for a continuum of possible spatial locations over several seconds and have been proposed for the organization of prefrontal cortical networks. I first present a novel method for the efficient prediction of the network-wide steady states from the underlying microscopic network properties. The method can be applied to predict and tune the “bump” shapes of continuous attractors implemented in networks of spiking neuron models connected by nonlinear synapses, which we demonstrate for saturating synapses involving NMDA receptors. In a second part, I investigate the computational role of short-term synaptic plasticity as a synaptic nonlinearity. Continuous attractor models are sensitive to the inevitable variability of biological neurons: variable neuronal firing and heterogeneous networks decrease the time that memories are accurately retained, eventually leading to a loss of memory functionality on behaviorally relevant time scales. In theory and simulations, I show that short-term plasticity can control the time scale of memory retention, with facilitation and depression playing antagonistic roles in controlling the drift and diffusion of locations in memory. Finally, we place quantitative constraints on the combination of synaptic and network parameters under which continuous attractors networks can implement reliable WM in cortical settings.

Key words: voltage-gated ion channels, spiking neuron models, neuronal networks, working memory, short-term plasticity, continuous attractors

Zusammenfassung

Das Gehirn ist ein komplexes System vieler mikroskopischer Prozesse, die den rechnerischen Fähigkeiten des Verhaltens unterliegen. Neuronale Modellierung, bestehend aus Modellen einzelner Neuronen und neuronaler Netze unterschiedlicher Detailtreue, kann fehlende Daten derzeit noch undurchführbarer Experimente künstlich überbrücken, um mechanistische Erklärungen für das Entstehen dieser rechnerischen Fähigkeiten zu finden. In dieser Dissertation stelle ich zwei parallele Forschungsschwerpunkte der neuronalen Modellierung, situiert in unterschiedlicher Detailtreue, vor.

Als Erstes untersuche ich die Herkunft von Ionenkanalmodellen in einer integrativen Metaanalyse von in den letzten 50 Jahren publizierten Arbeiten. Um dem stetig zunehmenden Volumen von Gehirnforschung gerecht zu werden, sollten Verfahren zur systematischen Auswertung publizierter Arbeiten entwickelt werden. Wir zeigen, dass neuronale Modelle den interessanten Ansatz bieten, studienübergreifende automatisierte und quantitative Analysen, mittels simulierter Experimente, durchzuführen. Wir entwickelten Protokolle für den quantitativen Vergleich spannungsabhängiger Ionenkanäle und wendeten diese auf einen Fundus publizierter Modelle an, wodurch wir die Vielfalt und zeitliche Entwicklung verschiedener Modelle der gleichen Ionenkanäle über Jahre der Forschung hinweg untersuchen konnten. Unsere Klassifizierung der existierenden Forschung ist in einer Online-Plattform zugänglich und eignet sich zum weitläufigen Vergleich experimenteller Daten mit Ionenkanalmodellen – dies erleichtert die forschungsfeldweite Standardisierung von experimentell beschränkter Modellierung.

Als Nächstes untersuche ich neuronale Modelle des Arbeitsgedächtnisses (AG). Auf welche Weise überbrücken kortikale Netzwerke die millisekundenschnellen Zeitskalen ihrer mikroskopischen Bestandteile, um schließlich verhaltensrelevante Zeitskalen im Sekundenbereich von AG-Experimenten zu erlangen? Wir betrachten ein mögliches Modell: Netzwerke mit kontinuierlichen Attraktoren (KA). Diese können ein AG implementieren, das ein Kontinuum räumlicher Lagen über mehrere Sekunden hinweg speichert und wurden als Organisationsprinzip des Präfrontalen Cortex vorgeschlagen. Zunächst stelle ich eine neuartige Methode zur effizienten Vorhersage der Netzwerkweiten stationären Zustände, als Funktion der mikroskopischen Eigenschaften des Netzwerks, vor. Dies kann angewendet werden, um die "BumpProfile von KA in Netzwerken von pulsbasierten Neuronen mit nichtlinearen Synapsen (z.B. saturierende Synapsen mit NMDA-Rezeptoren) vorherzusagen und um Netzwerkparameter einzustellen. In einem zweiten Teil untersuche ich die algorithmische Rolle einer weiteren synaptischen Nichtlinearität, der Kurzzeitplastizität. KA-Modelle sind anfällig gegenüber der unvermeidlichen Variabilität kortikaler Neuronen: Variable Feuerraten und heterogene Netzwerke verringern die Zeitspanne, über die hinweg Erinnerungen im Modell akkurat behalten werden. In Theorie und Simulationen zeige ich, dass Kurzzeitplastizität die Zeitskala des Gedächtniserhalts im Modell steuern kann, wobei Kurzzeit-Fazilitierung und -Depression entgegengesetzten Einfluss auf die Drift und Diffusion der Erinnerungen im AG ausüben. Schließlich setzen wir Grenzwerte für kombinierte Netzwerk- und Synapsenparameter, unter denen diese Modelle zuverlässiges AG im kortikalen Umfeld realisieren könnten.

Stichwörter: Spannungsabhängige Ionenkanäle, Pulsbasierte Neuronenmodelle, Neuronale Netze, Arbeitsgedächtnis, Kurzzeitplastizität, Kontinuierliche Attraktoren

Contents

Acknowledgements **i**

Abstract (English/Deutsch) **iii**

List of figures **xi**

List of tables **xiii**

1 Introduction **1**

1.1 Basic concepts 2
1.2 Detailed and simplified neuronal models 5
1.2.1 Detailed models of single neurons 5
1.2.2 Voltage-gated ion channels 7
1.2.3 Simplified models of neuronal networks 10
1.3 Models of working memory in simplified neuronal networks 13
1.3.1 Working memory in prefrontal cortex 14
1.3.2 Attractor models of working memory 16
1.3.3 Are continuous attractors used in biological neuronal circuits? 21
1.4 Thesis contribution 23

2 ICGenealogy: Mapping the function of neuronal ion channels in model and experiment **29**

2.1 Author contributions 29
2.2 Abstract 30
2.3 Introduction 30
2.4 Results 31
2.4.1 Categorizing ion channel models by metadata and ancestor-descendant relationships 31
2.4.2 Defining functional groups of models through voltage clamp protocols and clustering 33
2.4.3 Ion channel model groups defined by common metadata show variability in behavior 37
2.4.4 Automated comparison of new ion channel models and experimental data 40
2.5 Discussion 40

2.6 Methods and Materials . . . 44
  2.6.1 Pyramidal cell model (Figure 1) . . . 44
  2.6.2 The *ModelDB* database, *NEURON* language and nomenclature . . . 45
  2.6.3 Collection of metadata . . . 45
  2.6.4 Ancestor-descendant relationships . . . 47
  2.6.5 Voltage-clamp protocol . . . 48
  2.6.6 Data extraction & processing . . . 50
  2.6.7 Similarity measure . . . 50
  2.6.8 Clustering . . . 51
  2.6.9 Assessment of protocols . . . 52
  2.6.10 Generation of Circos plots . . . 53
  2.6.11 Generation of genealogy figures . . . 53
  2.6.12 Relational database, API and web interface . . . 54
  2.6.13 Code availability . . . 54
  2.6.14 Electrophysiology . . . 54
2.7 ICGenealogy web application . . . 54
  2.7.1 Software stack & web application . . . 55
  2.7.2 Database . . . 56
  2.7.3 Conclusion . . . 60

3 Continuous attractor networks: Efficient low-dimensional approximation 63
3.1 Author contributions . . . 63
3.2 Abstract . . . 63
3.3 Introduction . . . 63
3.4 Results . . . 65
  3.4.1 General equations for the approximation of stable states in ring-attractors 66
  3.4.2 Approximation of ring-attractor profiles in rate models . . . 69
  3.4.3 Approximation of ring-attractor profiles in spiking networks . . . 70
3.5 Discussion . . . 76
3.6 Methods . . . 79
  3.6.1 Rate network model . . . 79
  3.6.2 Spiking network model . . . 79
  3.6.3 Self-consistent equations . . . 82
  3.6.4 Spiking simulations . . . 87

4 Continuous attractor networks: Effects of short-term plasticity on drift and diffusion 89
4.1 Author contributions . . . 89
4.2 Abstract . . . 89
4.3 Introduction . . . 90
4.4 Results . . . 92
  4.4.1 Drift and diffusion in continuous attractor models . . . 92
  4.4.2 Ring-attractor in spiking networks with short-term plasticity . . . 94

4.4.3 Theoretical description of drift and diffusion in the presence of short-term plasticity . . . 97
4.4.4 Effects of short-term plasticity on diffusion . . . 103
4.4.5 Effects of short-term plasticity on drift . . . 106
4.4.6 Application: Effects of short-term plasticity on working memory function 117
4.4.7 Application: Construction of attractors . . . 125
4.4.8 Application: Control of uncertainty by short-term plasticity . . . 129
4.5 Discussion . . . 133
4.6 Methods . . . 139
4.6.1 Projection of dynamics onto the attractor manifold . . . 139
4.6.2 Drift and diffusion in presence of short-term plasticity . . . 148
4.6.3 Spiking network model . . . 154
4.6.4 Firing rate approximation of the spiking network . . . 156
4.6.5 Frozen noise and expected squared field magnitude . . . 161
4.6.6 Steady-state approximation and parameter optimization . . . 166
4.6.7 Numerical methods . . . 169

A Secondary publications 173
A.1 Synaptic patterning and the timescales of cortical dynamics . . . 173
A.2 Multi-contact cooperativity in spike-timing dependent structural plasticity stabilizes networks . . . 174
A.3 Algorithmic Composition of Melodies with Deep Recurrent Neural Networks . 175
A.4 Deep Artificial Composer: A Creative Neural Network Model for Automated Melody Generation . . . 176

B Supplementary Material: ICGenealogy 179

C Supplementary Material: Efficient low-dimensional approximation 193

D Supplementary Material: Effects of short-term plasticity on drift and diffusion 197
D.1 Extension: Synaptic connections with saturating NMDA conductance . . . 197
D.2 Extension: Simplified theory by projection onto Fourier modes . . . 200
D.3 Analysis of CVs in the attractor state . . . 202

Bibliography 207

Curriculum Vitae 231

List of Figures

1.1 Ion channel models collected from ModelDB 8
1.2 Attractor models for visuospatial working memory 17

2.1 The choice of ion channel model influences the behavior of a simulated neuron 32
2.2 Ion channel models can be categorized by metadata and ancestor-descendant relationships 34
2.3 Voltage-clamp protocols for the quantitative analysis of ion channel models . . 36
2.4 Quantitative analysis of Kv ion channel models: functional map and clusters of common behavior 38
2.5 Ion channel model groups defined by common subtype, neuron type and brain area show variability in behavior 39
2.6 Automated analysis of new models and experimental data 41
2.7 The ICGenealogy website allows for the interactive visualization of all data and analysis on the resource http://icg.neurotheory.ox.ac.uk 42
2.8 The IonChannelGenealogy web application 57
2.9 SQL database diagram and relational structure 58

3.1 Stable firing rate profiles in continuous attractors 66
3.2 Approximation of bump shapes in attractor networks 68
3.3 Application: Optimization of network connectivity 70
3.4 Application: Prediction of firing rate profiles in spiking continuous attractor networks 71
3.5 Efficient optimization of self-consistency errors 74
3.6 Applications: Shape prediction over parameter ranges / Optimization of network parameters 75

4.1 Continuous attractor models are susceptible to diffusion and drift (illustration) 93
4.2 Drift and diffusion in a spiking ring-attractor model with short-term plasticity . 96
4.3 Calculation of diffusion and drift on the ring-attractor (illustration) 98
4.4 Stability of stable firing rate profiles across parameters 102
4.5 Effects of short-term plasticity on diffusion 104
4.6 Diffusion magnitude as function of facilitation and depression parameters . . . 105
4.7 Prediction of drift-field in spiking network without facilitation 107
4.8 Prediction of drift-field in spiking network with facilitation 108

4.9 Prediction of drift in spiking network with varying STP parameters 109
4.10 Prediction of expected magnitude of drift fields from noise parameters 111
4.11 Self-consistency of squared magnitude predictions 112
4.12 Comparison of theoretically predicted fields to simulations 114
4.13 Comparison of theoretical predictions and simulations for distributions of final positions 115
4.14 Effects of short-term facilitation on mutual information 118
4.15 Stability of memory function depends on network parameters 122
4.16 Design of an attractor network with unimodal steady-state distribution 128
4.17 Short-term plasticity modulates uncertainty in bistable attractor networks . . . 130
4.18 Dependence of diffusion strength B on shape parameters 135
4.19 Numerical evaluation of diffusion strength 150
4.20 Pre-factor of normalizer for vanishing facilitation 151
4.21 Numerical evaluation of drift field 153
4.22 Steady-state approximation of short-term plasticity 157
4.23 Prediction of bump shapes for spiking networks with short-term plasticity . . . 167

B.1 Graphical description of the five voltage-clamp protocols used for ion channel model analysis. 181
B.2 Cluster indexes for Kv and Nav classes. 182
B.3 Cluster indexes for Cav, KCa and Ih classes. 183
B.4 Comparison of intra- and inter-subtype variability with intra- and inter-cluster variability. 184
B.5 Nav and Cav class genealogy and clustering. 187
B.6 KCa and Ih class genealogy and clustering. 189
B.7 K+ current recordings from Drosophila Kenyon cells. 191

D.1 Steady state of synaptic activation with logistic NMDA saturation 199
D.2 Comparison of full theory to approximation by projection onto Fourier coefficients203
D.3 Short-term plasticity does not affect spiking statistics 204
D.4 Diffusion estimated by theory only 205

List of Tables

2.1 Parameters for reversal potential and inside and outside concentrations used in simulation protocols for five ion type classes 48
2.2 Voltage-clamp protocol parameters for the five ion type classes 49

4.1 Upper bounds on system-sizes for stable continuous attractor memory in prefrontal cortex 123

B.1 Table of subtypes for each ion type class. 180
B.2 Table of omitted files. 186

C.1 Parameters for spiking simulations 194
C.2 Connectivity parameters of rate models 194
C.3 Parameters optimized in spiking networks 195

D.1 Parameters for spiking simulations 204
D.2 Network parameters for spiking simulations 205

1 Introduction

The work presented here is split into research on two parallel lines of research on neuronal modeling at different levels of biological detail. In a study presented in the first part of this thesis (Chapter 2), we show the results of a meta-analysis of a large body of published voltage-gated ion channel models. This study organizes existing knowledge and shows how models of the same biological ion channels change as they are reused and adapted. I then present two studies that investigate the long time scales of working memory, emerging through the collective network dynamics of continuous attractor models. The first study, presented in Chapter 3, demonstrates a method to predict the steady-state firing rate profiles of continuous attractors implemented in networks of simplified spiking neuron models. In a second study, in Chapter 4, we investigate the effects of short-term facilitation and depression on the slow dynamics of drift and diffusion in continuous attractor models.

This introduction is divided into three sections that reflect the bipartite structure of the results presented here. In Section 1.1 we begin by introducing fundamental concepts. I then present an introduction to detailed models of single neurons in Section 1.2, including a review of ion channel models to provide some background for the work presented in Chapter 2. After, I introduce simplified models of neurons and neuronal networks, with a focus on the models used in the remainder of the thesis. In Section 1.3 we turn to working memory models. To provide some biological and experimental background, I start by reviewing literature on working memory circuits in the prefrontal cortex. Then, in Section 1.3.2, I introduce attractor models and discuss the feasibility of continuous attractor models in a cortical setting, which motivates the work presented in Chapters 3 and 4.

Throughout this introduction, I will additionally place into context four additional studies to which I contributed during my thesis. These are summarized in Section A of the Appendix.

1.1 Basic concepts

In the following, I give a short introduction to the basic concepts of neuronal function, focusing on a level of biological detail that is used by models introduced later in the introduction. References in the following are mostly omitted, as this introduction is based on the textbook of Kandel et al. (2000), as well as the textbooks by Wulfram Gerstner et al. (Gerstner and Kistler, 2002; Gerstner et al., 2014).

Neurons Nerve cells, or neurons, are one major class of cells found in the nervous system, the second being glial cells, which classically have been thought to primarily provide structure and support to neurons. In this work, I will consider exclusively neurons, and networks formed by them, as the substrate of signal propagation and computation in the nervous system, although recent research hints at more complex roles of glial cells (Perea et al., 2014).

Neurons are electrically excitable cells that receive and propagate signals in nervous tissue by both electrical and chemical processes. As such, neurons typically possess a *membrane potential* V_m, a voltage difference across the cellular membrane, which arises from concentration differences of ions, most prominently Na^+ (sodium), K^+ (potassium), and Cl^- (chloride), between the intra- and extracellular medium. For neurons at rest (not firing action potentials, see below), a *resting potential* arises through the dynamical balance of ionic currents through ion channels as well as active ion pumps in the cellular membrane. Typical values for the resting potential of neurons are in the range of $V_m = -60mV$ to $V_m = -70mV$. In the reductionist view, the spatial structure of neurons can roughly be partitioned into three functionally different areas: *dendrites*, *soma,* and *axon*. Dendrites receive signals from other neurons, while the soma integrates these signals and initiates signal propagation by generating an output signal (action potential, see below), which travels through the axon to reach other neurons. Simplified neuron models often neglect the spatial structure of neurons, considering neurons to be *point neurons* (but see below).

Action potentials and synaptic connections The neurons I consider in this work (regardless of the many possible sub-classifications of cell types; see Markram et al. (2015) for a recent approach on cortical cells), can propagate signals by eliciting *action potentials*, also referred to as *spikes*. Action potentials are characterized by a transient depolarization of the membrane potential, actively generated by voltage-gated ion channels on the cell membrane of the axon initial segment. These voltage-gated conductances will be introduced in more detail below, as part of the *Hodgkin-Huxley model*. The action potential as an electrical signal can be transmitted between neurons by means of chemical signaling through *synaptic connections*[1] which are found along the axon. At synapses, the presynaptic axon and the postsynaptic neuron are in close spatial proximity, allowing *neurotransmitter* (chemicals released at the synapse in response to presynaptic action potentials) to bridge the distance to the postsynaptic membrane via the *synaptic cleft*. On the postsynaptic side, these neurotransmitters bind to receptors on ligand-gated ion channels[2] that permit the flow of ions through the membrane and thus convert the chemical signal back into deflections of the membrane potential (*postsynaptic potentials*, or *PSP*).

---

[1] Electrical synapses, which are bidirectional electrical couplings between neurons, are also found in the nervous system.

Action potentials serve as the basic unit of signal transmission and computation in the nervous system. The flow of the signal transmitted by the action potential is typically directed: the depolarization travels from the soma through the axon, where it, in turn, influences postsynaptic neurons through synaptic connections. The basic computation of a neuron lies in its integration of presynaptic input, and the all-or-nothing response of the action potential: 1) the joint presynaptic activity leads to the generation of an action potential only if the neuronal membrane potential is depolarized sufficiently; 2) if a neuron generates an action potential, it is a stereotypical response that will be transmitted to neurons that are synaptically connected to this neuron.

Neurotransmitters and receptors The post-synaptic potentials elicited by a single action potential are determined by several factors. Depending on the type of neurotransmitter that is released, one can distinguish between two classes of synapses for the vertebrate central nervous system: *excitatory synapses*, which cause depolarization of the postsynaptic membrane potential (excitatory PSP, or *EPSP*) and *inhibitory* synapses that cause a hyperpolarization (inhibitory PSP, or *IPSP*). In the setting of the cerebral cortex that I will consider in the context of working memory networks, the majority of neurons with excitatory synaptic transmission are pyramidal neurons relying on *glutamate* as a neurotransmitter, while inhibitory synaptic transmission relies on mainly *GABA* (γ-aminobutyric acid) which is released from inhibitory interneurons (see Gupta (2000) and Markram et al. (2015) for classifications of types of interneurons). Cortical presynaptic neurons can be classified as distinctly excitatory or inhibitory, according to the type of neurotransmitter that their synapses onto postsynaptic neurons use. It should be noted that this clear assignment of neuron types to single neurotransmitters does not hold generally throughout brain areas: there is evidence for co-release of several types of neurotransmitters from the same synapses (Strata and Harvey, 1999), e.g. GABA and dopamine in striatum (Tritsch et al., 2012).

The effect of any type of neurotransmitter is, in turn, determined by the postsynaptic composition of receptors to which it binds. In the central nervous system, glutamate elicits EPSPs mainly by binding to receptors on two different types of ligand-gated ion channels, that are both conductive to Na^+ and K^+ ions. First, *AMPA* (α-amino-3-hydroxy-5-methyl-4-isoxalone propionic acid) glutamate receptors, which elicit ionic currents that have relatively fast rise and decay times. Second, the class of *NMDA* (N-methyl-D-aspartate) glutamate receptors, which elicit currents that decay on considerably slower time scales. The ionic current passing through channels gated by NMDA receptors shows an additional voltage dependence, due to an additional binding of extracellular Mg^{2+} (magnesium), which is cleared from the channel as the cellular membrane is depolarized. Finally, NMDA channels are additionally permeable to Ca^{2+} (calcium) ions, which has important implications in synaptic plasticity (see below). Inhibitory IPSPs, on the other hand, are mediated in the central nervous system mainly by the binding of *GABA* to ligand-gated Cl^{-} permeable channels. Since the concentration of chloride is higher in the extracellular medium, the opening of these channels leads to an influx of chloride into the cell, which hyperpolarizes the membrane potential. It should be noted that, for other types of neurotransmitter besides glutamate and GABA, the postsynaptic receptor composition can reverse the response to a given neurotransmitter: e.g. acetylcholine can produce both inhibitory and excitatory postsynaptic responses (Kandel et al., 2000, p. 184).

---

[2] Next to these ionotropic receptors, there also exist metabotropic receptors that do not directly lead to ion fluxes through the membrane, but instead rely on second messengers.

Finally, the *synaptic efficacy* (often measured by the amplitude or slope of single PSPs) can be subject to changes over time, both on fast and slow time scales. Depending on mostly the activity of presynaptic neurons, the synaptic efficacy can undergo *short-term plasticity* on short time scales up to several seconds. Since it will be referred to below, I will provide here a brief overview of the molecular basis of short-term plasticity (see Zucker and Regehr 2002 for a review) – its computational relevance will be discussed in Section 1.3.3. When the action potential of a neuron reaches its presynaptic terminals, the opening of voltage-gated calcium channels (see below, Section 1.2.2), which are abundant at the *active zones* (sites of neurotransmitter release), will lead to a localized increase of intracellular calcium concentration (Kandel et al., 2000). The increased calcium concentration, in turn, causes the fusion of synaptic vesicles (the presynaptic “containers” of neurotransmitter) to the active zones, and subsequent release of neurotransmitter into the synaptic cleft. Regulating this release process are three factors: the probability of release p, the number of active zones n, and the available pool of releasable synaptic vesicles, where p increases with increasing calcium concentration. For each action potential, the amount of neurotransmitter released is random and proportional to $n \cdot p$. As several spikes occur in short succession*, short-term facilitation* will lead to a transient increase in the release probability p. The origin of this effect can be traced (amongst others, see Zucker and Regehr (2002)) to residual calcium in the presynaptic compartment, accumulated over several action potentials. *Short-term depression*, on the other hand, is a decrease in synaptic strength under successive spikes, which is probably linked to depletion of the pool of readily releasable synaptic vesicles (Zucker and Regehr, 2002). Synapses can also undergo longer lasting changes, which is thought to be the substrate of learning and memory. Here, the calcium permeability of NMDA receptors plays an important role, which, through the activation of cellular processes, can cause long-lasting potentiation or depression of synapses. For reviews on these long-term forms of synaptic plasticity see (Abbott and Nelson, 2000; Sjöström and Nelson, 2002; Shouval et al., 2010; Zenke and Gerstner, 2017).

1.2 Detailed and simplified neuronal models

The levels of description of neuronal models are as varied as the experimental data that is collected in the greater field of neuroscience (Dayan, 2006; Sejnowski et al., 2014). Here, I set a lower bound on granularity, in considering models that describe the dynamics of membrane potentials of single neurons and ionic currents.

In the first part below, I introduce the basic differential equation for neuronal membranes, and active generation of action potentials based on the work first put forward by Hodgkin and Huxley (1952b). With this foundation, the second part focuses on voltage-gated ion channels and their models in more detail, to provide the reader with some background for the work presented in Chapter 2. In a final part of this section, I then review simpler models of neuronal dynamics, which will be extensively used in Chapters 3 and 4.

1.2.1 Detailed models of single neurons

The resistor-capacitor (RC) circuit is a model of the dynamics of the neuronal membrane potential (Koch, 2004; Gerstner and Kistler, 2002). The membrane isolates a charge Q between the intra- and extracellular medium, thus acting as a capacitor with a membrane potential $V_m = Q/C$, where C is the membrane capacitance (in units of farads). A change in voltage corresponds to a capacitive current $I_C = C\frac{dV_m}{dt}$. We assume the membrane to be a perfect insulator, thus currents can only flow through ion channels that are situated in the membrane and allow ions to pass through them. The "resistor" in the RC circuit is introduced by a constant leak current through ion channels in the membrane, which reverses at the resting potential E_L (modeled through Ohm's law):

$$I_L = g_L(V_m - E_L),$$

where g_L is the leak conductance, usually given in units of siemens. Additionally, we include dynamical currents through other ion channels, which will be introduced below. Kirchhoff's law of current conservation then states that (under the absence of external currents) the total current must vanish, which yields a basic differential equation for the neuronal membrane:

$$C\frac{dV_m}{dt} = -I_L - I_{\text{Other}}. \tag{1.1}$$

Detailed models introduce other currents mediated by voltage-gated ion channels, that can actively generate action potentials. While this complicates the analysis of the resulting dynamics considerably, it yields a biophysically accurate description of the membrane potential dynamics. The basic formulation of this framework goes back to seminal work by Hodgkin and Huxley (1952b). They introduced two currents mediated by voltage-gated ion channels, permeable to either sodium (Na^+) or potassium (K^+) ions:

$$I_{\text{Other}} = g_{\text{Na}} m^3 h (V_m - E_{\text{Na}}) + g_{\text{K}} n^4 (V_m - E_{\text{K}}). \tag{1.2}$$

Here, for each of the two currents, g_{Na} and g_{K} are the maximal conductances, and E_{Na} and E_{K} are reversal potentials.

The effective conductance of each current is gated by the gating variables m, h and n. The variables m and h model the activation of ion channels, while n models the inactivation of sodium channels at high voltages. All three gating variables obey differential equations of the form (for $x \in \{m, h, n\}$):

$$\tau_x(V_m) \frac{dx}{dt} = -\left[x - x_0(V_m)\right]. \tag{1.3}$$

The functional dependence of the time constants τ_x as well as the steady-state values x_0 (constrained between 0 and 1) on the potential V_m determines the dynamics of the currents in response to the membrane potential and can be estimated from experiments Hodgkin and Huxley (1952b). Without giving the detailed functions here, we will nevertheless shortly describe how the interplay of gated sodium and potassium currents can generate action potentials. Increasing the potential V_m sufficiently above the resting potential leads to a fast increase of the variable m, thus allowing a transient strong current of positively charged potassium ions to enter the neuron, which depolarizes the membrane. On slower timescales, the high membrane potential leads to inactivation of the sodium current by a decrease of the variable h, as well as activation of the potassium current by an increase of the variable n. The outward directed potassium current quickly repolarizes the membrane and ends the action potential – due to the delayed activation of the potassium current, this class of potassium channel is commonly referred to as a *delayed rectifier.*

While the model proposed by Hodgkin and Huxley (1952b) was constrained by measurements of the giant squid axon, the same mathematical framework of ion channel dynamics is still in wide use today, mainly for simulation studies. Detailed models now often implement several different types of voltage-gated ion channel models (e.g. Marder and Goaillard (2006); Markram et al. (2015)). This has also produced a rich backlog of published voltage-gated ion channel models, which we consider in the next section and in Chapter 2.

Although we do not consider such models here, it should be noted that the Hodgkin-Huxley formalism of Eqs. (1.1)-(1.3) can be readily extended to incorporate the spatial geometry of neurons. This can be achieved by discretizing the spatial extent of neurons into iso-potential compartments which allow current to flow between them (see e.g. Koch (2004)). Such compartmental models can be simulated efficiently (Brette et al., 2007) (using for example the popular NEURON simulation environment (Carnevale and Hines, 2006)), and have found wide adoption in the field (Hines et al., 2004; Parekh and Ascoli, 2013; Markram et al., 2015).

1.2.2 Voltage-gated ion channels

In this subsection we introduce some more detailed background on ion channels and their models, focusing on the voltage-gated ion channels analyzed in Chapter 2. As we argue in the introduction (Section 2.3), the presence of different ion channel models can crucially affect single neuron dynamics, signal processing of neurons, and synaptic plasticity. Thus, the particular types of ion channel models used in detailed neuronal models will impact the interpretability of simulation results in light of biological fidelity. An increasingly large number of ion channel models for the NEURON simulator (Carnevale and Hines, 2006) are being made available through the public code repository ModelDB (Hines et al., 2004) (Fig. 1.1A). However, the large numbers of different model implementations for similar ion channel types and redundant nomenclature (see below) make this body of digital knowledge hard to navigate. In the work presented in Chapter 2 we developed methods for the automated comparison and classification of ion channel models, and applied them to 2378 published ion channel models (since the time of publication, we have increased the number of channel models to 2952), that fall into 5 major classes, depending on ions and gating mechanisms (Fig. 1.1A, inset). Here, we provide the reader with some background on voltage-gated ion channel models and the 5 major classes of ion channel models that we analyzed.

The dynamics of single biological ion channels are inherently stochastic, with currents that switch stochastically (in a voltage-dependent manner) between discrete conductive states (Koch, 2004). Kinetic schemes can be used to model the voltage-dependent transition rates between several conformational changes of subunits responsible for opening channels, usually implemented as Markov models (Destexhe et al., 1994; Koch, 2004). In the macroscopic limit (averaging over more than a few hundreds of channels (Koch, 2004)) the resulting compound current I_k for a channel k of most types can be well approximated by generalized Hodgkin-Huxley equations of the form:

$$I_k = g_k m^{p_k} h^{q_k} (V_m - V_k),$$

where g_k is the maximal conductance and V_k the reversal potential. The channel activation and inactivation are described by dynamical activation (m) and inactivation (h) variables, which obey equations similar to Eq. (1.3). Different integer exponents p_k and q_k can be fit to experimental data of the conductance in response to voltage steps. The decision of which modeling approach to use depends on the desired level of accuracy, the speed of implementations, and the experimental data available to fit models (Destexhe and Huguenard, 2000). Accordingly, the landscape of published models is also diverse: implementations that we encountered in the study of Chapter 2 varied from stochastic Markov models with multiple states of activation and inactivation to models implemented in the simpler Hodgkin-Huxley formalism. To be able to accommodate all possible kinetics, the analysis performed in Chapter 2 remains agnostic to the model implementations, analyzing only generated ionic currents from models.

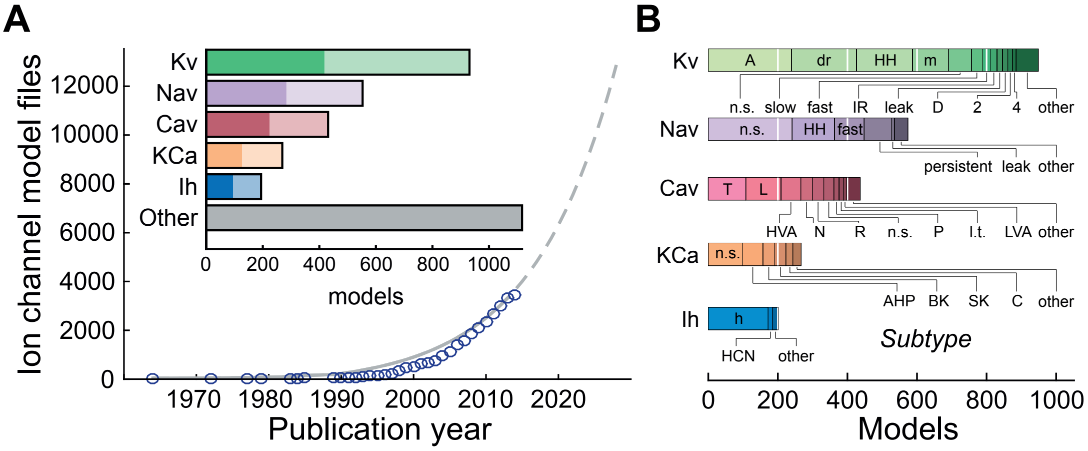


Figure 1.1 – **Ion channel models collected from ModelDB.** **A** Cumulative number of single channel models collected from ModelDB (Hines et al., 2004) (blue circles), together with illustrative exponential fit (dashed gray line). Inset displays the number of total and unique (light and solid, respectively) channel models in five main classes, as well as model files catalogued that do not fall into the five classes (*other*). **B** Partition of the five classes of panel A by the nomenclature (*Subtype*) provided by the channel model authors, ordered from left to right according to size. Uncommon subtypes are grouped together (*other*).

Voltage-gated ion channels can in principle be classified by their genetic composition: e.g. the voltage-gated sodium channels can be grouped into the 12 classes $K_V1 - K_V12$ (and their respective subclasses) (Gutman et al., 2005) (for other ion types references are given below). In the biophysical literature, on the other hand, more descriptive naming schemes are common (Hille, 2001), where currents are often denominated by their kinetic properties (e.g. "delayed-rectifier" potassium currents, or "T type" (transient) calcium channels). These descriptive denominations are quite loose and can correspond to several genetic subclasses, indicating further diversity in each functional class. Nevertheless, in published ion channel models they do occur frequently: in the assay of published models presented in Chapter 2 we found a large diversity of legacy labeling, with both genetic and descriptive naming schemes (these were more frequent) occurring ("Subtype" in Fig. 1.1B). Even for ion channel models labeled by similar names, we were able to show a large diversity of model kinetics. Additionally, many models of ion channels are labeled with ambiguous names, or did not have any names indicating their function: e.g. the subtype "HH" in Fig. 1.1B is composed of channel models with Hodgkin-Huxley type dynamics, that did not indicate any further denomination. One of the results of the study in Chapter 2 is an automated sorting of the ambiguous "backlog" of nomenclature of published ion channel models.

In the following, we introduce the 5 major classes of voltage-gated ion channels that occur in Chapter 2, along with their general relevance and some of the established nomenclature in the descriptive naming scheme (since channel models denominated by this scheme frequently occur in the study).

Voltage-gated sodium (Na$_V$) The opening of sodium channels contributes to the rapid rise of the action potential in nerve cells (reversal potential around $E_{\text{Na}} = 50mV$), allowing for a fast influx of sodium which depolarizes the membrane. Typically, sodium channels have fast voltage-dependent inactivation ($1-2ms$) (Yu and Catterall, 2004). Next to these, persistent sodium channels (NaP) are frequent, that do not show any (or very slow) inactivation (Koch, 2004). Genetically, 10 different voltage-gated sodium channels are reported, showing mostly similar functional properties (Catterall et al., 2005a).

Voltage-gated potassium (K$_V$) The reversal potential of potassium currents is usually around $E_K = -80mV$, thus hyperpolarizing the membrane by an outward current of potassium ions, for example in the falling phase of action potentials (Yu and Catterall, 2004). Next to the delayed rectifier (dr) currents introduced in the Hodgkin-Huxley model above, which activate slowly and do not (or very slowly) inactivate, there are several other common voltage-gated potassium currents. A type (transient inactivating) currents activate fast and inactivate on fast to intermediate timescales ($\sim 10-100ms$), which allows cells to generate action potentials at low frequencies (Koch, 2004; Gutman et al., 2005). Slowly activating and inactivating K2 channels inactivate at longer time constants ($0.2-2s$) (Gerstner and Kistler, 2002). M (muscarinic) type potassium currents are slowly activating and not inactivating channels influencing subthreshold excitability (Gutman et al., 2005). Genetically, 40 different voltage-gated potassium channels across 12 subfamilies can be differentiated (Gutman et al., 2005).

In our study, we also included models of inward rectifier potassium channels in this class. Also referred to as anomalous rectifiers, these channels permit potassium currents to flow only into the cell. Since they are often activated by hyperpolarization of the membrane beyond the resting potential, these currents can stabilize the resting potential without initiating action potentials (Nichols and Lopatin, 1997). Genetically, there currently are 15 different inward rectifying potassium currents across 7 subfamilies on record (Kubo et al., 2005).

Voltage-gated calcium (Ca$_V$) The calcium flux through these channels is always directed inward, due to the low intracellular concentration of Ca^{2+} ions, with a range of sensitivities to voltage, and a range of inactivation time constants (Koch, 2004). Ca$_V$ channels occur in almost all types of nerve cells and are found mostly in dendrites, the soma, and presynaptic compartments (Koch, 2004). Functionally, the influx of calcium in response to depolarization is needed for the release of neurotransmitter from presynaptic terminals (Kandel et al., 2000) and underlies several intracellular regulatory processes (Yu and Catterall, 2004), importantly activating second messengers that underlie synaptic plasticity. Common subtypes include (Koch, 2004): L type (long lasting) channels, activated at high voltages and inactivated by small increases in the intracellular calcium levels; T type (transient) channels, that activate at lower voltages than L-types and deactivate strongly; and N type (neither L nor T) that have properties intermediate between L and T types. Due to almost vanishing intracellular concentration of calcium ions, modeling these ionic currents involves the Hodgkin-Goldman-Katz equations, instead of Ohm's law (see for example Gerstner and Kistler (2002); Koch (2004)). Genetically, 10 different voltage-gated calcium channels across 3 subfamilies have been identified (Catterall et al., 2005b).

Calcium gated potassium (K_{Ca}) Potassium currents from these ion channels are, in addition to being voltage-dependent, modulated by the intracellular calcium concentration. Simple modeling schemes involve making the activation variable dependent on the intracellular calcium concentration $m(V_m, [\mathrm{Ca}^{2+}]_{\mathrm{intra}})$, which implies that neuronal models must additionally model the dynamics of the calcium concentration (see e.g. Koch (2004)). A common type of KCa channel is the fast activating C type, which as an additional potassium current contributes to repolarization of the membrane during action potentials. The calcium-dependent afterhyperpolarization channel (AHP) is also of functional importance: this slowly inactivating current depends only on the calcium concentration, and its buildup through the calcium influx during successive spiking can lead to adaptation of neuronal firing. Commonly occurring descriptive separations for human KCa channels are big conductance (BK); intermediate conductance (IK); and small conductance (SK) (Vergara et al., 1998). Currently, one can separate 8 different KCa channels across 5 subfamilies by their genetic composition (Wei et al., 2005).

Hyperpolarization activated cation (I_h) I_h currents are mixed cation currents composed of both K^+ and Na^+, with reversal potentials around $-20mV$ (Biel et al., 2009). Similar to inward rectifier potassium channels (above), the associated channels are activated by hyperpolarization of the membrane beyond values close to the resting potential, leading to inward directed currents. Thus, the current can lead to subthreshold adaptation (Gerstner et al., 2014) and generally a stabilization of the membrane potential against both depolarizing and hyperpolarizing currents, amongst a variety of other roles (Biel et al., 2009). The underlying channel family is that of hyperpolarization-activated cyclic nucleotide-gated (HCN) channels, which separates into 4 different members (Robinson and Siegelbaum, 2003). Importantly, HCN channels are actively regulated by the binding of cyclic nucleotides, most prominently cAMP – this makes I_h currents subject to neuromodulation by neurotransmitters that up- or down-regulate cAMP levels (Robinson and Siegelbaum, 2003).

1.2.3 Simplified models of neuronal networks

The Hodgkin-Huxley type models that I introduced in the previous section are, as a coupled system of 4 (or more) nonlinear differential equations, hard to analyze. One can approximate the 4-dimensional Hodgkin-Huxley model with sodium and potassium channels by a two-dimensional approximation (see e.g. Gerstner and Kistler (2002)) which actively generates action potentials. Depending on the phenomena to be investigated, however, even this level of description is often considered too detailed. Since action potentials are usually stereotypical (for a given cell) events of very short duration, they can be approximated by discrete events in time and the dynamics underlying their generation and time course skipped altogether.

This class of neuron models, called *integrate-and-fire neurons*, can be seen as a consensus between abstraction and biological detail (Dayan, 2006): they are both analyzable and efficient to simulate in large numbers (Brette et al., 2007), while still retaining the event-based character of neuronal transmission by action potentials. In Chapter 3 and Chapter 4 we simulate and analyze networks that consist of connected models of *leaky integrate-and-fire* (LIF) neurons. Here, I introduce the model and how networks of connected neurons can be modeled by introducing synaptic interactions.

Leaky integrate-and-fire neurons (see e.g. Burkitt 2006a for a review) model the temporal dynamics of a neuronal membrane $V(t)$ by a membrane equation similar to Eq. (1.1):

$$C\frac{dV}{dt} = -g_L(V - E_L) - I_{\mathrm{Syn}}, \tag{1.4}$$

where I have included the leak current and a synaptic current I_{Syn} that will be introduced below. The generation of spikes from the model is abstracted by introducing a threshold potential V_{thr}: if the membrane potential crosses the threshold V_{thr} at time t_f from below, the neuron is considered to elicit an action potential, and its membrane potential is immediately reset to a subthreshold value V_{reset}. To capture the duration of spikes and refractoriness of neurons[3], LIF neurons often have a fixed refractory period τ_{refr}, during which the voltage remains clamped at V_{reset}.

While I only use simple LIF models in this thesis, it is noteworthy that several extensions to the LIF model have been proposed, mainly to capture the more complex dynamics of the Hodgkin-Huxley model with several voltage-gated ion channels. Voltage-dependent subthreshold nonlinearities (Fourcaud-Trocmé et al., 2003; Brette and Gerstner, 2005; Badel et al., 2008; Hansel and Mato, 2001) can account for spike initiation dynamics, and simplified spike-triggered and voltage-dependent currents can produce complex spiking dynamics (Rauch et al., 2003; Brette and Gerstner, 2005; Mensi et al., 2016), for example spike-frequency adaptation or bursting. Importantly, parameters for such extended LIF models can be efficiently estimated from experimental data, and can accurately predict the subthreshold membrane potential (Pozzorini et al., 2015) as well as spike-times (Jolivet et al., 2008; Rossant et al., 2010) of biological neurons. This compression of the detailed neuronal behavior into models with few parameters has been used for classification of biological neuron types (Mensi et al., 2012), as well as the approximation of detailed compartmental neuron models (Rossant et al., 2011; Rössert et al., 2016).

In Chapter 3 and Chapter 4 we use LIF neurons to model networks of synaptically connected excitatory and inhibitory neurons. Synaptic transmission is implemented by specifying the dynamics of postsynaptic currents in response to presynaptic action potentials. Since these dynamics can be different for varying types of ligand-gated ion channels (and associated neurotransmitter), the total synaptic current I_{Syn} above will be composed of individual currents I_k, each arising from a different class of ion channels k. A common conductance-based description of synaptic transmission models these individual currents by Ohm's law (Gerstner and Kistler, 2002):

---

[3] After eliciting a spike, neurons can often not immediately fire an action potential due to sodium inactivation and residual potassium activation – this is called the (absolute) refractory period.

$$I_{\text{Syn}} = \sum_k I_k = \sum_k \bar{g}_k (V - E_k) \sum_j s_{k,j}(t), \tag{1.5}$$

where $\bar{g}_k$ is the maximal conductance for each synaptic connection and ion channel, and E_k is the reversal potential which depends on the ions the channel k is permeable to (Koch, 2004). The synaptic gating variables $s_{k,j}$ model the fraction of open channels for the synaptic connection from neuron j, which is increased by neurotransmitter released in response to the presynaptic spike train. For currents through AMPA and NMDA receptor channels, E_k is usually chosen to be $0mV$, thus leading to a depolarization of the membrane potential. For inhibitory currents mediated through GABA-gated channels E_k is chosen well below the resting potential, around $-75mV$.

Next to the stochastic opening of ion channels (see the last section), the presynaptic release of neurotransmitter from active zones is also subject to stochasticity (Zador, 1998; Koch, 2004). In an additional study (summarized in the Appendix, in Section A.2) we investigated the interplay between stochastic synaptic transmission and long-term synaptic plasticity in a network model with multiple synaptic contacts. For stochastic models of synaptic transmission, see Destexhe et al. (1998). In Chapter 3 and Chapter 4, we will instead rely on widely used (Burkitt, 2006a) simplified deterministic dynamics of the synaptic gating variables, which makes them analytically tractable:

$$\frac{ds_{k,j}(t)}{dt} = -\frac{s_{k,j}(t)}{\tau_k} + \sum_{t_f \in T_j} \delta(t - t_f). \tag{1.6}$$

Here, τ_k is the decay time constant of the conductance through the channel k, and T_j are the times ($\leq t$) that the presynaptic neuron fired action potentials. Realistic and simulated synapses will always experience a short transmission delay, which I neglect in this notation. The parameters of Eq. (1.6) depend on the ion channel considered, in particular the time constants τ_k are usually fast for AMPA channels ($\sim 2ms$), intermediate for GABA channels ($\sim 10ms$) and long for NMDA ($\sim 100ms$) (Gerstner and Kistler, 2002). Additionally, as mentioned in the introduction, NMDA receptor channels are voltage-dependent, which can be included in Eq. (1.5) by adding a voltage-dependent term (Jahr and Stevens, 1990b). For NMDA the linear differential equation Eq. (1.6) is often replaced by a non-linear relationship that saturates for high presynaptic firing rates (Destexhe et al., 1994; Wang, 1999) – in Chapter 3 we analyze networks with such non-linear synaptic transmission.

The simplest description of neuronal networks that still includes the activity of single neurons it that of *rate-based models*. In these, the successive spikes generated by single neurons are replaced by their firing rates ν. For single neurons, a common definition of the firing rate of neurons is the spike-count rate over a given interval T (see e.g. Dayan and Abbott (2001); Gerstner and Kistler (2002)): $\nu = n_{\mathrm{spk}}/T$. Here, n_{spk} is the number of action potentials the neuron emits during the interval T. For networks in a *steady state*, where the statistics of the synaptic current I to neurons stay constant over time, we can replace the neuronal dynamics by the steady state *input-output relation (Dayan and Abbott, 2001):*

$$\nu = F(I).$$

Importantly, the input-output relation can be measured from cortical neurons and allows relating the steady-states of simplified neuronal networks to biology (La Camera et al., 2008).

The advantage of using LIF models is that one can calculate their input-output relation (Burkitt, 2006a). Matching the irregular firing of cortical neurons (Softky and Koch, 1993; Compte et al., 2003), networks of recurrently connected excitatory and inhibitory neurons can display asynchronous and irregular firing (Amit and Brunel, 1997; Brunel, 2000a; Renart et al., 2007), with spiking statistics close to that of Poisson processes[4]. For neurons that receive input from many presynaptic neurons that fire uncorrelated and with these Poisson statistics, the mean and fluctuations of synaptic inputs in Eq. (1.4) can be approximated, which, in turn, allows derivation of the neuronal input-output relation (Brunel and Sergi, 1998; Brunel and Wang, 2001b; Fourcaud and Brunel, 2002). This central analytical result will be used in Chapter 3 and Chapter 4 to derive rate-based approximations for the steady-state spiking activity of networks of excitatory and inhibitory LIF neurons which implement spatial working memory (see below). Extended analytical methods can be used to calculate the response of populations of neurons to time-dependent inputs (Fourcaud and Brunel, 2002; Burkitt, 2006b).

1.3 Models of working memory in simplified neuronal networks

Having introduced simplified models of single neurons and connected networks of neurons in the last section, I now turn to neuronal models of working memory. Conceptual models of working memory include separation of attentional mechanisms, episodic buffers, and mechanisms tasked with integration of different modalities from visuospatial and verbal/auditory origins (reviewed in Baddeley (2012)). Mechanistic computational models of working memory have to provide the means to bridge the short time scales of their microscopic components to the time scales of working memory on the order of seconds. I will focus here on one possible solution: attractor networks, in particular networks with *continuous attractors*, which we will investigate in detail in Chapters 3 and 4.

Why use simplified network models? Such models can help understand the circuit properties that can generate working memory function by exploring and predicting the aspects of neuronal circuitry currently inaccessible in experiments (Amit, 1998; Sejnowski et al., 2014). Experimentally, the properties of single neurons (Pozzorini et al., 2015) and synaptic connections (Wang et al., 2006) can be measured. In addition, the detailed activity (Constantinidis et al., 2001; Compte et al., 2003) and connectivity (Goldman-Rakic, 1995; Wang et al., 2006) of small subsets of neurons can be assessed, and macroscopic correlates of the activity of populations of neurons can be recorded (Curtis and D'Esposito, 2003). However, linking microscopic neuronal properties and the emergent circuit dynamics requires knowledge of the full circuit structure. In simulations, connectivity can be synthesized, the emergent circuit function predicted, and eventually compared to experiments (Dayan, 2006). The reductionist approach of simplified models is by definition an approximation of the true cortical substrate, as are more detailed simulation approaches (Koch and Buice, 2015). In contrast to detailed simulations (Markram et al., 2015), however, the reductionist approach tries to identify, in analysis and simulation, the *necessary* conditions for the emergence of computational properties, as for example, working memory.

---

[4] The (homogeneous) Poisson process generates, independently and randomly, temporal events at a given rate r, with a coefficient of variation of the distribution of waiting times $CV = 1$; see for example Dayan and Abbott (2001).

In the first section below, I introduce the notion of working memory and provide some biological and experimental background to the correlates of working memory, focusing on visuospatial working memory in the prefrontal cortex. I then give an overview of classes of working memory models in simplified neuronal networks and introduce the continuous attractor models used in Chapters 3 and 4. In a third section, I will highlight the intrinsic instability of continuous attractor models – this is the topic of Chapter 4, where we investigate the effects of short-term synaptic plasticity on the stability and time scales of working memory function in continuous attractor models.

1.3.1 Working memory in prefrontal cortex

Working memory (WM) is the computational ability to represent, transiently store, and recall behaviorally relevant information, on time scales up to minutes (Goldman-Rakic, 1995). Underlying the emergent computational ability of working memory is the concerted activity of highly interconnected neuronal networks (Chaudhuri and Fiete, 2016). In contrast to long-term (and short-term) memory, in which information is stored in the neural circuity through structural modifications due to learning (Goldman-Rakic, 1991; Arnsten et al., 2012), working memory processes have been compared to a “mental sketchpad” (Baddeley, 2012). The involved neural circuitry is, through evolution (Arnsten et al., 2012; Duarte et al., 2017) and prior structuring (Amit and Mongillo, 2003; Mongillo et al., 2003), specialized to readily provide its function.

In primates, the *prefrontal cortex* (PFC) is an area crucial for working memory, as assessed in experimental paradigms which require the storage of information in working memory circuits over *delay periods* of several seconds. Across multiple different behavioral paradigms, the delay-period activity of neurons in PFC seems to be the neural correlate of keeping information in working memory (for reviews, see Goldman-Rakic (1995); Funahashi (2006); Compte (2006)).

In particular, many PFC neurons show elevated and *persistent activity* during these delay periods – this is the basis for attractor models of working memory introduced in the next section.

In Chapters 3 and 4 we investigate a class of models (*continuous attractor models*, see below) that have been proposed as a model for working memory of locations in visual space. This type of working memory can be assessed in the *visuospatial oculomotor delayed response* (ODR) task (Fig. 1.2A) (Funahashi et al., 1989; Constantinidis et al., 2001). The ODR task allowed experimenters to show that many PFC neurons with elevated activity during delay periods have "memory fields" (Goldman-Rakic, 1995): the activity of neurons is selectively tuned to increase with memories of specific locations in the visual field. In the following, I introduce the ODR task in a little more detail, since its paradigm of cueing, maintenance and readout of memories over delay periods is reflected in models used here. In ODR experiments (Funahashi et al., 1989; Constantinidis et al., 2001; Compte et al., 2003), subjects are presented with several possible target locations in the visual field (Fig. 1.2A, squares), arranged in a (usually circular) fashion around a fixation point (Fig. 1.2A, circle). A cue signal is given by illuminating a single, randomly chosen target (Fig. 1.2A, red square at 0°) for a short duration ($0.5s$ in Funahashi et al., 1989). During a delay period of several seconds following the cue signal, the subject has to maintain fixation and prepare a saccadic response to the previously illuminated location upon a "go" signal (the disappearance of the fixation point). The task is designed such that subjects need to internally maintain a working memory of the cue's location during the delay period, which is then used to effectuate the saccade. Electrophysiological recordings of neurons in primate PFC during ODR tasks showed that many neurons show *persistent activity* during the delay period: in the absence of external stimuli, these neurons fire at elevated rates, with firing rates depending on the location of the previously shown cue. Importantly, neurons represent in their activity several locations: firing rates of direction-selective neurons during the delay period were found to be distributed around a preferred direction in a bell-shaped manner (Fig. 1.2A, inset). This *continuous mapping* of neuronal mnemonic responses to the memorized spatial locations (see also next section) was termed "memory fields" (Goldman-Rakic, 1995).

How does the activity in PFC correlate with memory function? First, small lesions of PFC can lead to degradation of behavioral performance in ODR tasks for restricted areas of the visual field, implying an essential contribution of local circuitry to the mnemonic representation of visual space (Funahashi, 2006). Second, a majority of persistent activity in PFC seems to represent visual memory (retrospective memory), while internal representations of preparatory motor signals (prospective memory) are less frequent (Takeda and Funahashi, 2002). More generally, mixed selectivity of neuronal firing with respect to different tasks (Warden and Miller, 2010) and stimulus dimensions (Rao, 1997; Rainer et al., 1998) is often observed (Fusi et al., 2016), which one would expect if the neural substrate is to be used efficiently (Rigotti et al., 2010). For example, PFC neurons also show sustained firing during delay periods of auditory memory tasks (Constantinidis and Wang, 2004). Third, PFC is highly inter-connected with other cortical areas, which also show delay activity during ODR tasks, e.g. the parietal cortex earlier in the dorsal visual stream (Chafee and Goldman-Rakic, 1998). Thus, PFC might not be solely responsible for working memory function, with working memory representations possibly distributed across several areas (Curtis and D'Esposito, 2003; Constantinidis and Wang, 2004; Guo et al., 2017). More recent evidence collected from fMRI studies shows that even early sensory areas maintain persistent activity during delay periods (Sreenivasan et al., 2014). In this view, it has been hypothesized that PFC might play the role of a top-down controller that maintains optimized sensory representations in earlier sensory regions (Curtis and D'Esposito, 2003; Sreenivasan et al., 2014). Finally, next to persistent firing, heterogeneous dynamical responses of neurons in PFC with temporal dynamics have been reported during working memory activity: neurons have been shown to encode task-related activity (Stokes, 2015), for example the evolution of time during delay periods (Brody et al., 2003b; Machens et al., 2010). A recent study by Murray et al. (2016) (see also Balaguer-Ballester et al. (2011), for a similar approach across several frontal cortical areas) showed that, on a population level, the joint activity of PFC neurons converges to subspaces of neuronal activity on which activity remains stable during delay periods. This indicates that, despite possible dynamical encoding of additional task dimensions, populations of PFC neurons act together as a neural substrate that encodes working memory representations in a stable fashion.

In summary, the activity of neurons in PFC plays an integral role in working memory function, which can be evaluated in delayed response tasks involving visuospatial working memory. The bell shaped location-specific "mnemonic tuning" of neurons that show persistent firing during the delay period naturally leads to a continuous representation of visual space in PFC. This is the basis of continuous attractor models of visual working memory, which are introduced in the next section.

1.3.2 Attractor models of working memory

Most microscopic constituents of cortical networks do not retain their state for more than hundreds of milliseconds (Chaudhuri and Fiete, 2016). The ability of cortical circuits to retain information in working memory over time scales of several seconds is therefore probably an emergent property of networks of neurons (although cellular origins also exist; see Major and Tank (2004)) . A widespread working hypothesis is that recurrently connected cortical networks realize working memory function through *attractor dynamics* (Durstewitz et al., 2000b; Major and Tank, 2004; Compte, 2006; Chaudhuri and Fiete, 2016). In this view, cortical working memory networks are *multi-stable*: several states (neuronal activity across the network) are stable under the recurrent network dynamics (*stable fixed points*), and thus can be used as working memory representations that are self-sustained. To load sensory information into working memory, feed-forward excitatory input from cortical areas that process sensory signals leads to the elevated firing of a subgroup of neurons and sets the network state close to one of the stable states. In the absence of sensory input, the network dynamics converge to the closest stable state and remain there over delay periods. The elevated activity of some neurons in the stable attractor state will be visible as persistent activity, which can be propagated

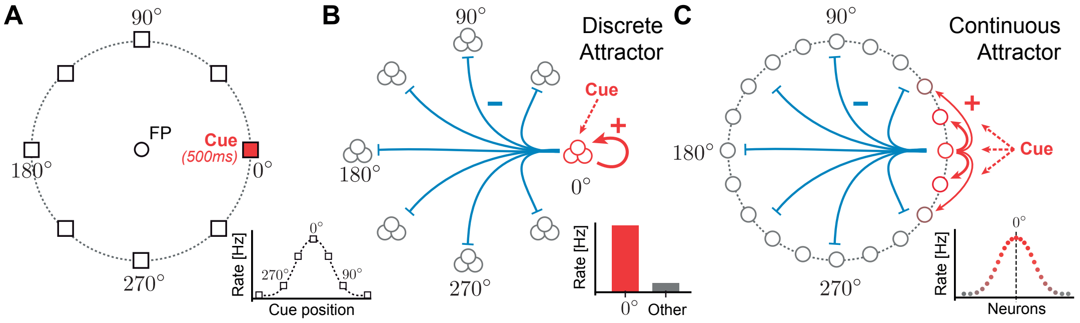


Figure 1.2 – **Attractor models for visuospatial working memory.** **A** Experimental setup for oculomotor delayed response tasks (ODR; adapted from Funahashi et al. (1989)). Subjects are presented with a fixation point (FP) and are tasked to perform, after a delay period of several seconds, a saccade to one randomly chosen location out of eight possible target locations (white squares). At the start of the delay period, the target location is visually cued by illuminating it for a short duration (500ms), here illustrated by the red square at 0° (Cue). Inset: Bell-shaped firing rate tuning of a single PFC neuron during the delay period (illustration), as a function of the cue position. **B** Discrete attractor model. Each possible location is represented by a population of several neurons (3 circles represent one population). Due to excitatory feedforward input from sensory areas encoding the cue signal (arrow to population at 0°), the associated population of neurons (red circles) starts firing at elevated rates (inset). The elevated firing rates persist over delay periods: they are self-sustained by strong excitatory feedback (+, red arrow), as well as inhibition of all populations (blue lines). **C** Continuous (ring) attractor model. Each neuron in the network represents a specific preferred angular position in visual space, approximating a continuum of possible cue locations. Sensory input representing the cue signal (red arrows) arrives at all neurons close to the cue location. Self-sustained activity is generated by strong excitatory feedback between neurons representing similar angles (+, red arrows), together with global inhibition (blue lines). Inset: Self-sustained activity takes the shape of a “bump” profile of elevated firing rates across all neurons (in contrast to the inset in panel A, which shows the firing rate response of *a single neuron* for varying cue positions), here centered at the neuron representing 0°. For clarity, only outgoing excitatory synaptic connections from neurons at 0° are shown in panels B and C.

through feedforward connections to other cortical areas, in order to "read" from this working memory implementation.

In network models of excitatory and inhibitory LIF neurons, attractors dynamics can be realized through strong excitatory recurrent connections between subgroups of excitatory neurons, combined with inhibitory feedback ((Amit and Brunel, 1997; Brunel and Wang, 2001b; Curti and Amit, 2004; Romani et al., 2006); see Brunel (2003); Compte (2006) for reviews). In these models, the recurrent excitatory connections will self-sustain the firing of excitatory neurons in an attractor state. Inhibitory feedback, on the other hand, counterbalances the excitatory input to neurons and limits the rate of excitatory firing and ensures the existence of a stable, nonselective *spontaneous state* where all neurons fire equally at low rates (Amit and Brunel, 1997). Low firing rates in attractor states can also be achieved by the nonlinearity of saturating NMDA currents, or short-term synaptic depression Wang (1999). Many attractor models rely on the asynchronous firing of neurons (Compte, 2006) (but see below), which can stabilize attractor states by reducing the fluctuations of synaptic currents generated by many such neurons. Long synaptic time constants further support this stabilizing effect. In particular, NMDA receptor currents stabilize attractor states in models (Wang, 1999; Compte et al., 2000; Brunel and Wang, 2001b), and have been shown to be crucial for working memory performance in experiments (Arnsten et al., 2012; Wang et al., 2013).

Discrete attractor models How can external sensory information be represented in attractor networks? Some visual working memory paradigms present discrete items, e.g. faces (Scalaidhe et al., 1997), or objects (Miller et al., 1996). For each of these stimuli, neurons in PFC are found that are selectively responsive. Possibly, the discrete set of spatial locations of the ODR task (Fig. 1.2A; see last section) could also be interpreted as such a set of discrete items. *A discrete attractor* gives a possible model implementation for working memory of such discrete items (illustrated in Fig. 1.2B; e.g. Brunel and Wang (2001b)). First, excitatory neurons are grouped by the items they are selective to (with neurons responsive to several items possibly assigned to several groups) – in Fig. 1.2B each stimulus location is represented by a separate population of neurons. Strong excitatory feedback connections are introduced between neurons of the same groups (red arrow in Fig. 1.2B), while inhibitory feedback is realized through an intermediate population of inhibitory neurons (blue lines in Fig. 1.2B illustrate the effectively resulting inhibition). After a cue signal activates a group of neurons (location 0° in Fig. 1.2B), the network settles into an attractor state where only this population fires at high rates and all other groups are inhibited (inset in Fig. 1.2B). Such a discrete attractor structure can result from Hebbian learning, which strengthens neuronal connections for neurons that are active together (Mongillo et al., 2003; Amit and Mongillo, 2003; Mongillo et al., 2005; Zenke et al., 2015; Carrillo-Reid et al., 2016).

Continuous attractor models The discrete attractor paradigm does not reflect possible "similarities" between stimuli. In the ODR task target locations are sampled from a continuum of possible locations. Beyond the experimental setup, cortical working memory networks presumably have developed to store any continuously valued spatial location in working memory, which could prove beneficial, for example, during everyday visual navigation. Indeed, as noted above, in ODR experiments it has been found that many PFC neurons fire at elevated rates for a range of angular positions of the visual space (Funahashi et al., 1989). Each neuron responds for angular positions distributed around a preferred angle (highest response) in a bell-shaped fashion (cf. Fig. 1.2A, inset). An attractor network structure that produces such tuning curves and stores a continuum of items is the *continuous attractor* model (Fig. 1.2C). Using this model for the ODR task (Compte et al., 2000), excitatory neurons are assigned a preferred angle from a continuum of possible angles along a circle in the visual field[5]. Since this circular layout of locations in the ODR task leads to a ring of angular positions, the resulting models are often referred to as *ring-attractor* models. Recurrent excitatory feedback connections are chosen such that neurons representing similar angles are more strongly connected, leading to *distance dependent connectivity* (distances are in angular space) (Fig. 1.2C, red arrows from neuron at 0°). Again, inhibitory feedback is mediated by an intermediate population of inhibitory neurons (blue lines in Fig. 1.2B illustrate the effectively resulting inhibition). After the signal activates neurons around the presented cue location (location 0° in Fig. 1.2C), the excitatory neurons settle into a self-sustained continuous "bump" shaped profile of firing rates (Fig. 1.2C, inset) – this profile jointly represents the angular position currently in working memory by its center[6]. Importantly, measuring the tuning of neuronal responses during delay periods (memory fields, see the last section) for excitatory neurons will yield the bell-shaped tuning also found in PFC neurons during the ODR task (Fig. 1.2A, inset). Although they will not be used here, I want to point out that the principle of distance dependent connectivity can be extended to models of other topologies, to yield continuous attractors on lines, toroids, and spheres (reviewed in Knierim and Zhang (2012)), mixtures of several ring attractors (Romani and Tsodyks, 2010), and mixtures of discrete and ring attractors (Roudi and Treves, 2008).

Mean-field analysis of attractor models It is noteworthy that attractor models mostly require that neuronal populations fire asynchronously (Compte, 2006), in agreement with experimental recordings from PFC (Compte et al., 2003) (but see next below). The asynchronous regime of firing is important theoretically since it makes network models amenable to analysis by *mean-field* approaches (Brunel, 2003). In this regime, one can approximate the statistics of neuronal firing by Poisson statistics. Neurons that fire at similar rates can be grouped into sub-populations, which permits calculating the statistics of synaptic input to sub-population as a function of the firing rates of all others. Using the input-output relation of neurons (see the end of Section 1.2.3), one can then derive self-consistent equations for the firing rates of each population, which can be (numerically) solved to predict the firing rates in attractor states (Amit and Brunel, 1997; Brunel and Wang, 2001b). In discrete attractors (also with overlapping populations, see Curti and Amit (2004); Mongillo et al. (2005)), the grouping of neurons into homogeneous sub-populations is given by the structure of the network – in the above example, a possible split would consist of all inhibitory neurons, one group of "active" excitatory neurons, and all other neurons. In continuous attractor models, no such intrinsic separation is given since the network structure does not designate subgroups and attractor states show heterogeneous profiles of firing rates in the excitatory population (Fig. 1.2C, inset). In Chapter 3, we present a method that circumvents this problem by parametrizing the firing rate profile: this allows deriving a set of self-consistent equations that constrain the firing rate profile of continuous attractors. Importantly, our method relies only on the knowledge of the input-output relation, and can be applied to networks of LIF neurons in the presence of non-linear NMDA saturation (Chapter 3) and short-term plasticity (Chapter 4).

---

[5] These angles can be randomly assigned: e.g. sampling from the uniform distribution over all circular angles, see e.g. Romani and Tsodyks (2010). A simpler modeling choice is to cover the whole angular range by evenly spacing all neurons, as illustrated in Fig. 1.2C.

[6] Similar to the population vector coding of motor cortex, see Georgopoulos et al. (1989).

Alternatives to attractor models I would like to point out some alternative approaches to equip recurrent neuronal networks with the long time scales useful for working memory function (see Chaudhuri and Fiete (2016) for a recent overview). In a recently published review (Duarte et al. (2017); summarized in the Appendix, Section A.1), we review the impact of local and regional variations in synaptic composition and circuit structure on the emergence of long time scales. The review includes a study that used derivative feedback control (Lim and Goldman, 2013), relying on balanced excitation and inhibition (see the end of next section) with fast inhibitory feedback, to achieve long retention times in recurrent networks. This principle can also be used to stabilize continuous attractor structures (Lim and Goldman, 2014). A second recently emerging principle relies on the transient retention of network states in slowly decaying activity traces left in synapses by short-term plasticity (Maass et al., 2002; Mongillo et al., 2008; Stokes, 2015): the spontaneous reactivation of sub-populations with facilitated recurrent synapses allows the efficient (since firing is not persistent) storage of memories over long delay times (Mongillo et al., 2008). Different to asynchronously firing attractor networks (see above), these networks show synchronous population activity during the delay period (Lundqvist et al., 2011), which have been hypothesized to underlie bursts of oscillatory activity observed in PFC during working memory (Lundqvist et al., 2016). Finally, *long short-term memory* (LSTM) neurons used in machine learning approaches have working memory by design (Hochreiter and Schmidhuber, 1997): changes to their internal states are protected by "read" and "write" mechanisms, and can thus retain their state indefinitely. While these models contain small recurrent circuits explicitly designed for this function, the useful principle of write-protecting internal states could be approximated by neuronal circuitry (O'Reilly and Frank, 2006; Marblestone et al., 2016). In the appendix, I present two studies (Colombo et al. (2016, 2017), summarized in the Appendix, Section A.3 and Section A.4) in which we used variants of these models to extract and reproduce the long time scale structure of music.

To summarize, I have introduced discrete and continuous attractor models of working memory, on the example of the ODR task. The continuous attractor model, which I focus on in this work, can represent a continuum of possible spatial locations through a firing rate profile that is self-sustained and can be centered at any possible location. In the next section, I will review evidence for whether the continuous attractor framework can be, and is, used in cortical networks.

1.3.3 Are continuous attractors used in biological neuronal circuits?

While the continuous attractor model was developed from rather theoretical considerations, it is a promising candidate computational principle for the how recurrently connected (Constantinidis and Wang, 2004) cortical working memory circuits could implement the storage of continuous-valued items. Models for the ODR task based on ring-attractors (Compte et al., 2000) have been used to explain behavioral performance (Macoveanu et al., 2006, 2007; Roggeman et al., 2014; Almeida et al., 2015), predict the effects of neuromodulation (Cano-Colino et al., 2013, 2014), or the implications of cognitive impairment (Murray et al., 2012; Cano-Colino and Compte, 2012). Next to modeling working memory, continuous attractors have also been applied to model head-direction cells, the oculomotor system, and hippocampal spatial representations (see Knierim and Zhang (2012); Moser et al. (2014); Burak (2014); Wu et al. (2016) for recent reviews, as well as the introduction of Chapter 4). Specific evidence for continuous attractor-like dynamics of working memory in PFC has recently been put forward by Wimmer et al. (2014). There, the authors show that the trial-to-trial variability in behavioral reports during an ODR task is compatible with *continuous* shifts of the encoded working memory – this would be expected from continuous attractors, but not from discrete attractors representing the possible target locations. Experimental evidence for continuous attractor dynamics also come from the limbic system (Knierim and Zhang, 2012; Yoon et al., 2013) and, recently, circuits of Drosophila responsible for path integration and self-orientation (Seelig and Jayaraman, 2015).

Stabilization of continuous attractors: short-term plasticity The capacity of continuous attractor models to store a continuum of possible states relies crucially on a degree of perfect symmetry only obtainable in a model system (Brody et al., 2003a; Barak and Tsodyks, 2014; Chaudhuri and Fiete, 2016). As we review in Chapter 4 (Section 4.3 and Section 4.4.1), any transient or permanent deviation from symmetry in models will lead to a degradation of the stored information, which makes continuous attractors highly susceptible to noise. For example, the spiking variability of neurons will introduce transient deviations from the mean firing rates for each neuron and lead to an accumulation of random changes in the angles maintained in working memory. More crucially, biological variability (e.g. heterogeneous neuronal parameters, or random sparse network connectivity) will introduce permanent deviations, that lead to systematic biases in the positions storable in such a network for longer times. In cortical settings both sources of variability can be expected to be present and continuous attractors will always be subject to degradation over time.

Does this exclude continuous attractors as candidate computational structures for cortical working memory? On the timescale of hours to days, it has been shown that the degrading effects of systematic biases in continuous attractors implemented in networks of spiking neurons could be reduced by homeostatic synaptic plasticity (Renart et al., 2003). A stabilization mechanism that could be useful to counteract variability on shorter time scales is short-term synaptic plasticity of recurrent synaptic connections. In simplified rate networks, short-term facilitation has been shown to decrease the effect of systematic biases (Itskov et al., 2011b). In spiking networks, studies have reported similar effects on both systematic (Hansel and Mato, 2013) and random degradation of memories (Pereira and Wang, 2015). Neither of these studies, however, investigated the functional dependence of these effects on the strength of facilitation. Next to short-term facilitation, biological synapses commonly also show short-term depression, where possible implications on biologically plausible implementations of continuous attractor working memory remain unclear (see Section 4.3).

In Chapter 4 we quantify in detail the effects of short-term synaptic plasticity on both random and systematic sources of variability, both by theory and simulation. In particular, we use our theory to predict the performance of working memory systems with short-term plasticity, depending on the underlying sources of variability. Thereby, we can place constraints on the properties of networks that can implement working memory in continuous attractor structures. Finally, I would like to point to Section 4.5, where we discuss other means of increasing the robustness of continuous attractors to noise, which could be employed in concert with short-term plasticity to stabilize biological implementations of continuous attractor memory.

Computational roles of short-term plasticity In Section 1.1 we have already introduced the molecular basis of short-term plasticity (STP). To describe the dynamics of the release of neurotransmitter under facilitation and depression in Chapter 4, we use a simplified deterministic model (Markram et al., 1998; Tsodyks et al., 1998, 2000) that neglects the stochasticity introduced by the probabilistic nature of neurotransmitter release (Pan and Zucker, 2009). Such simplified models of STP (notably also Abbott (1997); Dittman et al. (2000); Lindner et al. (2009)) have shown that STP plays a key role for dynamic gain control of neuronal information transmission (Lindner et al., 2009; Pfister et al., 2010) (see Abbott and Regehr (2004); Anwar et al. (2017) for reviews). In discrete attractor networks, STP has been shown to exert control over the stability of attractor states (Mongillo et al., 2005; Sussillo et al., 2007; Rolls et al., 2013) and noise-induced transitions between them (Bibitchkov et al., 2002; Mejias and Torres, 2009; Miller, 2013). In Chapter 4 we show that this principle extends to continuous attractors: short-term depression and facilitation control the time scales of dynamics on the attractor manifold in an antagonistic fashion, which includes noise induced movement.

Finally, a recently proposed role of short-term plasticity in the context of attractor models of working memory is the variability of neuronal firing. Attractor models that rely on strong recurrent excitation show decreased variability in attractor states (Renart et al., 2007; Barbieri and Brunel, 2007), contrary to what was observed in cortical networks (Compte et al., 2003; Renart et al., 2010). This decrease in variability is also apparent in the models of continuous attractors investigated in Chapter 3 and Chapter 4 (but see Barbieri and Brunel (2007)). A different mechanism of achieving attractor working memory in recurrently connected networks is by mediating self-sustained firing not through increasing the mean of currents, but by increased fluctuations (Renart et al., 2007), which is achieved by scaling the recurrent connectivity such that excitation and inhibition are both strong and balance each other. Recent studies have shown that the synaptic nonlinearities introduced by short-term plasticity can equip such balanced networks with robust bi-stability, both in models implementing discrete (Mongillo et al., 2012; Romani et al., 2006) and continuous (Hansel and Mato, 2013) attractors (see also the studies of Rosenbaum and Doiron (2014); Lim and Goldman (2014) for balanced continuous attractors without short-term plasticity). Since this leads beyond the scope considered here, extensions of the results obtained in this thesis to balanced architectures of continuous attractors will be left to future work.

1.4 Thesis contribution

This thesis compiles the results that I have obtained during my Ph.D. from 2012 to 2017 at the EPFL under the supervision of Prof. Wulfram Gerstner. The primary goal of the thesis was to investigate biologically plausible mechanisms of eliciting long time scales of activity in cortical networks which could be applied for the implementation of working memory circuits in cortical networks. To approach this question, I developed novel theoretical methods that allow the prediction of the steady states and slow timescale dynamics of continuous attractor models implemented in spiking neuronal networks with nonlinear synapses. A second significant contribution developed in parallel, in a collaboration with the laboratory of Prof. Tim Vogels at the University of Oxford, with the goal of developing methods for the automated evaluation and comparison of published neuronal models.

The thesis is divided into three main chapters. My personal contribution to each of these is provided at the beginning of each chapter in a section called *Author contributions*. Next to these chapters, in **Section A** of the Appendix, I present four additional studies to which I contributed during my Ph.D., together with a description of my contributions. As I have already placed these studies into context in the introduction above, they will not be referred to in the following.

In **Chapter 2,** I present the (published) results of the collaboration with the University of Oxford, which document an online resource for published ion channel models (ICGenealogy, 2016b). A large and ever increasing number of published neuronal models is archived in an online repository (Hines et al., 2004), including several thousand models of ion channels that are currently not systematically labeled nor quantitatively compared. This is a problematic state for the field of neuronal modeling since it is common practice for modelers to adapt previously published ion channel models for use in their work. Without ongoing experimental verification, this may introduce systematic changes or even errors into later generations of models and may affect the biological interpretability of results. **To improve this state of affairs, we developed methods for the automated evaluation and classification of ion**

channels, allowing us to quantitatively assess and this large body of published ion channel models. For the quantitative assessment of ion channel kinetics, we devised a standardized set of voltage-clamp protocols that are commonly applied for the experimental characterization of ion channels and the fitting of Hodgkin-Huxley models (Ranjan et al., 2011). These protocols allow us to evaluate, in a model-free manner, the kinetics of ion channel models, regardless of their underlying implementation. In a novel approach, the same protocols can be applied in experiments to collect similar data from biological ion channels, which we use to analyze both experimental and synthetic data by the same method. All published channels model implementations were curated to be able to simulate them under the protocols – this dataset has been made publicly available and presents a potential resource for further studies seeking to perform large-scale simulation experiments across published channel models for any given type of ion channel. A further significant contribution arises from a manual extraction of previously not isolated metadata from all associated publications, thereby compiling a second layer of categorical data that can help researchers assess models. Importantly, we record whether a model implementation is based on an earlier published model, thereby establishing a "genealogy" of ion channel models that can help trace systematic changes over generations of model reuse. Using the dataset collected in this fashion, we show a high degree of variability in the previously existing labeling of channel models and provide a new classification of models based on a clustering of models into classes with similar kinetic responses. For most channel types of the dataset, we were able to isolate several clusters of distinct kinetics, thereby sorting several thousand published models into a few hundred different classes. All our results are made available to the scientific community in an online resource (ICGenealogy, 2016b), which constitutes one of the major contributions of this project. Extending the previously existing archive of models (Hines et al., 2004), our project enables modelers looking for existing channel models to browse all models on record sorted by quantitative or categorical data, and compare response kinetics of any number of models. Further, our resource allows the quantitative evaluation of user-uploaded models and experimental recordings against all other models in the database. In our study, we demonstrate this by identifying a model that closely matches recordings from Kenyon cells in Drosophila. **In summary, this project contributed to the field by creating a novel interactive resource for the categorization of new and existing ion channel models as well as experimental recordings.** In ongoing work, we are keeping the resource updated to integrate newly published models and extend our data sources to integrate additional repositories of published models. The methods and infrastructure developed in this project provide, for the first time, the means to evaluate experimental recordings of ionic currents against a large body of published ion channel models. As more experimental and model data is added, we hope our resource can contribute to the field-wide standardization of experimentally-constrained modeling by providing a platform for the automated and continuous integration of ion channels in model and experiment.

In Chapters 3 and 4 I turn towards the dynamics of continuous attractor models. As explained in the introduction above, continuous attractors can provide a neuronal substrate that retains information about a given input stimulus over extended periods of time, in the context of cortical working memory. My aim was to analyze continuous attractor dynamics for varying neuronal properties, in particular, nonlinear synapses. In **Chapter 3** I present the first part of the results of this investigation, in which we propose a novel and generalizable method for the efficient approximation of the steady-state "bumps" of continuous attractor models. To our knowledge, there currently exists no broadly applicable method to efficiently relate the microscopic properties of the neuronal network to the emerging bump steady states of continuous attractor models, which are their central feature. While simplified models admit analytical solutions of the system dynamics, spiking neuronal dynamics and nonlinear synaptic transmission both complicate theoretical approaches. When these more complex models are reduced to a firing rate description, the resulting high-dimensional coupled nonlinear equations require iterative numerical solutions, which are possibly as costly as simulating the full neuronal dynamics. **In a semi-numerical approach to this problem, we present a novel method for the prediction of self-consistent bump solutions applicable to a wide range of neuronal networks.** Our method relies on the optimization of a low-dimensional parametrization of the continuously varying firing rates of the bump state. For neurons that admit an input-output relation in the steady state, the method allows us to use numerically efficient optimization approaches to constrain the parametrization and predict the bump shape of the full network dynamics. Importantly, since the dimensionality of the optimization target is small (with respect to the network size), we can use this approach to optimize network parameters for the emergence of particular bump shapes. We apply our method to a widely used class of continuous attractor networks implemented in spiking networks with nonlinear saturating NMDA synapses and show that we can programmatically tune these networks to display a variety of steady states. For spiking networks, the method incorporates the mean-field theory of Brunel and Wang (2001b), and extends its applicability to the setting of continuous attractors. The method can be of utility for modelers working on spiking continuous attractors since it allows the efficient tuning of network parameters to match physiological quantities. For example, we apply the method in the next part (Chapter 4) to networks with short-term plasticity. There, we re-tune network parameters for a large set of plasticity parameters, in order to make the bump states of networks with different short-term plasticity parameters comparable. Additionally, the method can be used to efficiently predict the effects of varying microscopic network properties on the emerging bump state, across ranges of parameters. In the study we demonstrate this by predicting a widening of the bump shape under cortical disinhibition in a spiking network, confirming previous simulation results that related these effects to cognitive impairment in schizophrenia patients (Murray et al., 2012). **In summary, our method contributes a novel and generalizable approach for the theoretical analysis and tuning of the steady states of neuronal systems with spatially varying firing rates.**

In **Chapter 4** we then investigate continuous attractor networks with short-term facilitation and depression as synaptic nonlinearities. Continuous attractor models are known to be susceptible to the biological variability of cortical networks (see Section 1.3.3 of the Introduction), which will lead to drift and diffusion along the attractor manifold. As earlier research in simplified models showed (these were discussed in Section 1.3.3), short-term synaptic facilitation can slow drift. However, whether, and to which degree, this effect persists in spiking neuronal networks that show diffusion remain unclear. Additionally, synapses rarely are only facilitating, but also show short-term depression. The effects of short-term synaptic depression on the time scales of dynamics along continuous attractor manifolds are scarcely investigated, as are the combined effects of both facilitation and depression. The central question of this research was whether, and to which degree, short-term facilitation and depression can act as a biologically plausible mechanism to control the time scales of drift and diffusion in spiking continuous attractor networks, and possibly lead to slow dynamics on long time scales. In particular, we were interested whether short-term plasticity would allow such networks to implement reliable working memory function under the presence of biological variability. **To answer this question, we developed a novel theory that predicts, quantitatively, the drift and diffusion of the memories encoded in continuous attractors with short-term plasticity.** The theory presents a major extension to a previously published approach (Burak and Fiete, 2012): our novel contribution accounts for short-term synaptic plasticity in the prediction of diffusion, and demonstrates the applicability of the approach to more complex neuron models. In the limit of vanishing short-term plasticity, our theory recovers the results of Burak and Fiete (2012) for the prediction of diffusion in simpler spiking neuron models. A further major extension consists in the analytical prediction of the magnitude of directed drift on the attractor, which generalizes the rate-based approach of Itskov et al. (2011a) to continuous attractor networks with neurons described by input-output relations and dynamic depressing synapses (the authors assumed a fixed level of depression). We additionally extend the approach of Renart et al. (2003), where the authors estimated drift fields numerically for spiking networks without short-term plasticity. Together with the approach presented in Chapter 3, our theory provides a theoretical toolset for the analysis of the steady states and quasi-stationary slow time scale dynamics of continuous attractor networks implemented in spiking neuronal networks with short-term synaptic plasticity. **To validate our theory against simulations, we implemented and compared a large set of spiking continuous attractor networks with varying parameters of facilitation and depression.** This study is, to our knowledge, the first to do so systematically for biologically plausible spiking networks (with the partial exception of Pereira and Wang, 2015, who investigated the effects of the facilitation time scale). We required particular care in ensuring that networks retained nearly identical bump states as the parameters of facilitation and depression were varied, to exclude possible secondary effects on the slow network dynamics. Network parameters fulfilling this demand were obtained using the theory presented in Chapter 4, for which we extended the mean-field theory of Brunel and Wang (2001b) to include currents with short-term synaptic plasticity.

Assuming an underlying organization of cortical networks as continuous attractors, the theory developed here can make general predictions about the dynamical properties of such putative networks. As an original result, the theory we developed allows us to constrain the combinations of synaptic and network (mainly the network size) properties that sufficiently slow drift and diffusion to allow continuous attractor models to perform working memory function even under substantial cortical variability. The resulting network sizes for experimentally constrained parameters of short-term plasticity are in anatomically reasonable ranges (Collins et al., 2016) between $10^4 - 10^5$ neurons. Our result is thus in support of the feasibility of cortical networks implementing the continuous attractor paradigm, especially since additional mechanisms can further increase the robustness of working memory implemented in such networks (see Discussion in Chapter 4). We have performed these experiments for integrate-and-fire neurons and one given shape of the bump state. The free parameters of the theory, however, can be extracted from experimentally observable quantities of cortical networks that implement working memory: these include the input-output relations of neurons, and the putative shape of the bump state, which can be related to experimentally measured memory fields of PFC neurons (see Section 1.3.2). **A further original finding of our study is that both drift and diffusion on continuous attractor networks will be affected in an antagonistic fashion by short-term facilitation and depression, with facilitation being able to strongly reduce both.** This yields a novel implication for facilitation as a cortical stabilization mechanism for continuous attractor working memory (see Discussion in Chapter 4): strong facilitation will, especially in larger networks, remove most diffusive displacements. This might be difficult to unify with behavioral reports during working memory experiments, which show variable (diffusive) errors to be present. We propose one possible explanation with interesting implications: facilitation and depression might be under neuromodulatory control, which would render the working memory system more flexible and could allow diffusion to reappear. In our study, we demonstrate that this principle of neuromodulatory control of short-term plasticity could be used to represent uncertainty in working memory circuits.

2 ICGenealogy: Mapping the function of neuronal ion channels in model and experiment

William F Podlaski[†], **Alexander Seeholzer**[†], **Lukas N Groschner**, **Rajnish Ranjan**, **Gero Miesenböck** and **Tim P Vogels**; [†]equal contribution

Published in: eLife 2017;6:e22152 (DOI: 10.7554/eLife.22152.001; Podlaski et al. (2017))

2.1 Author contributions

The idea for the study was conceived by TPV, WP, RR, and me.

WFP, RR, and TPV developed the voltage-clamp protocols and NEURON scripts to run simulated experiments of ion channel models, collect the resulting current traces, and preprocess them. Metadata was primarily collected by WFP, with support by RR and me. WFP and I developed and performed the dimensional reduction (PCA) procedure, as well as the cluster analysis. WFP analyzed the Drosophila data, which was collected by LG and GM.

I developed all web-application technology, except the graphical data browser (see below). The graphical data browser was created by phyramid.com, in tight coordination me. My contribution: devising and implementing the database structure; data import and curation; creation and maintenance of the web-application; creation and maintenance of the web-application stack and server. Additionally, I modified the NEURON scripts to be executable in on the web-server environment. I developed and curated Python code to generate the Github repository containing all modified channel models. This contribution is additionally detailed in Section 2.7, where I included an extended summary of the web-application technology. This summary was not part of the publication.

All authors wrote the original draft and revised version. WFP and I produced all figures, except Fig. 1, which was produced by WFP. RR provided additional visualizations, which were not included in the publication.

2.2 Abstract

Ion channel models are the building blocks of computational neuron models. Their biological fidelity is therefore crucial for the interpretation of simulations. However, the number of published models, and the lack of standardization, make the comparison of ion channel models with one another and with experimental data difficult. Here, we present a framework for the automated large-scale classification of ion channel models. Using annotated metadata and responses to a set of voltage-clamp protocols, we assigned 2378 models of voltage- and calcium-gated ion channels coded in *NEURON* to 211 clusters. The *IonChannelGenealogy* (ICGenealogy) web interface provides an interactive resource for the categorization of new and existing models and experimental recordings. It enables quantitative comparisons of simulated and/or measured ion channel kinetics, and facilitates field-wide standardization of experimentally-constrained modeling.

2.3 Introduction

Ion channels play crucial roles in neuronal signal processing (Koch and Segev, 2000; Cai et al., 2004; Goldberg et al., 2008) and plasticity (Sjöström and Nelson, 2002; Shah et al., 2010; Debanne et al., 2003). Interactions among the many different ion channels expressed by a single cell can lead to extraordinarily complex dynamics, whose dissection necessitates computational modeling, as first demonstrated by Hodgkin and Huxley in 1952 for action potential generation (Hodgkin and Huxley, 1952b). Simulation environments like *NEURON* (Hines and Carnevale, 2001; Carnevale and Hines, 2006) can be used to create biophysical neuron models with realistic morphologies, ionic currents, and channel densities (Figure 2.1A), facilitating the integration of experimental data into models (Mainen and Sejnowski, 1995; Stuart and Spruston, 1998; Migliore et al., 1999; Poirazi et al., 2003; Destexhe and Paré, 1999; Traub et al., 2003). More than a thousand neuronal models, and several thousand individual ion channel models, are archived in the online database *ModelDB* (Hines and Carnevale, 2004), which enables other researchers to verify original claims, and to reuse and extend existing neuron models in the light of new results.

Matching model and experiment is essential for biophysical neuron models, in which many components have a direct biological counterpart (Brette et al., 2007). For example, pyramidal neuron models have been shown to reproduce the recorded spiking activity of these cells accurately with a particular set of ion channels (Traub et al. (2003); Figure 2.1B, gray traces; see Methods). However, the dynamics can change, sometimes dramatically, when one of the modeled ion channel currents is exchanged for an identically-labeled model from a different publication on *ModelDB* (Figure 2.1B, colored traces, Figure 2.1C). This example underscores the importance of selecting ion channel models, yet there is currently no standardized experimental dataset against which to validate them.

Furthermore, the increasingly large number of models on ModelDB (e.g., over 300 new ion channel models in 2014 alone; Shepherd Lab (2015)), with non-standardized labeling and a high degree of redundancy, makes it difficult to understand how ion channel models relate to each other and to biology. For example, a researcher looking to use an existing *A-type* potassium channel model will find over 250 *A-type* models, spanning a range of behaviors (Figure 2.1C, blue). Instead of a thorough and time-consuming fitting of appropriate ion channel dynamics, it is common for modelers to adapt previously published ion channel models for their own purposes. However, this may introduce experimentally unverified systematic changes or even errors into later generations of models and may have dramatic effects on the biological interpretation of the results.

To facilitate informed choices among this bewildering variety of ion channel models, we categorized 2378 published voltage- and calcium-dependent ion channel models in *NEURON* that are available on *ModelDB*. We cataloged all relevant information about each ion channel model from the associated literature, including its *pedigree* relations: whether a given ion channel model is based on previous models, and, if so, which ones. Additionally, we compared the kinetics of each ion channel model in standardized voltage-clamp protocols. The resulting maps of ion channel behavior show model variability and diversity, and point to the computational and experimental sources that were used to fit each model. Our efforts have grouped 2378 ion channel models into 211 clusters, dramatically simplifying the search for an appropriate ion channel model.

We present our findings in an annotated, interactive web-interface with a short video manual (ICGenealogy (2016b); http://icg.neurotheory.ox.ac.uk), that allows filtered search of individual ion channel models by metadata and relational information, and the comparison of channel model kinetics. The underlying database is freely and programmatically accessible via a web application programming interface (API). In an effort to make our resource compatible with experimental data and new ion channel models, we offer the possibility to upload and assess the similarity of experimentally recorded current traces (as well as new models and model traces) in the same topology. We show an example of the use of this comparison through the analysis of an unclassified ion channel model, as well as an experimentally recorded voltage-dependent potassium current from *Drosophila melanogaster*. In summary, we provide a framework for the direct and automated comparison of models and experiments to facilitate experimentally constrained modeling and quantitative characterization of ion channel behavior.

2.4 Results

2.4.1 Categorizing ion channel models by metadata and ancestor-descendant relationships

To build a map of ion channel model function, we categorized and analyzed a widely-used subset of 2378 voltage- and calcium-dependent ion channel models (*".mod"* files) in the

NEURON simulation environment (Hines and Carnevale, 2001; Carnevale and Hines, 2006). A set of “metadata” was extracted manually from the associated journal articles for each ion channel model file (Figure 2.2A, top): reference information (Ref. Info, including author(s) of the model code), ion channel information (I.C. Info: ion selectivity, gating mode, subtype), system information (Sys. Info: brain area, neuron type, neuron region, animal model), as well as additional comments (Other: e.g. temperature constraints, see Methods).

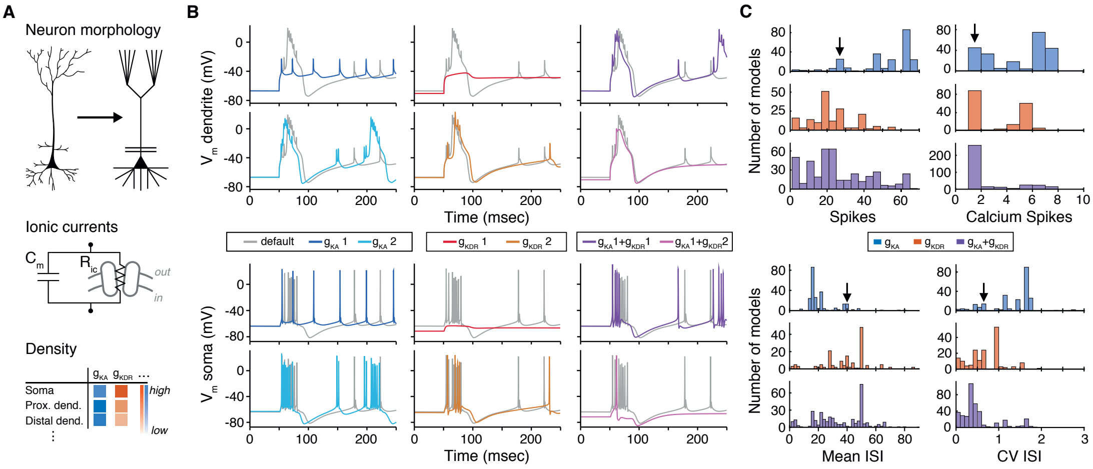


Figure 2.1 – The choice of ion channel model influences the behavior of a simulated neuron. **A**: Biophysical neuron models are composed of a detailed multicompartmental morphology, several active ion channel conductances, and a density of each conductance that depends on the specific compartment. **B**: Simulation of a detailed layer 2/3 pyramidal neuron model, adapted from Traub et al. (2003) (see Methods for details). The neuron model was stimulated with a 1.5nA current step beginning at 50msec, while recording the membrane potential in the apical dendrite (top) and soma (bottom). Simulations were first run using the original conductances from Traub et al. (“default”, gray). Left: the default A-type potassium model (g_{KA}) was replaced with two other A-type models ($g_{KA}1$: dark blue, Hay et al. (2011), ModelDB ID no. 139653; $g_{KA}2$: light blue, Traub et al. (2004), ModelDB ID no. 45539). Middle: the default delayed rectifier potassium model (g_{KDR}) was replaced with two other delayed rectifier models ($g_{KDR}1$: red, Zhou and Hablitz (1996), ModelDB ID no. 3660; $g_{KDR}2$: orange, Durstewitz et al. (2000a), ModelDB ID no. 82849). Right: both A-type and delayed rectifier models were replaced with other models ($g_{KA}1 + g_{KDR}1$, purple; $g_{KA}1 + g_{KDR}2$, magenta). **C**: Model from **B** was simulated for 1000msec with a 1.5nA current step. Firstly, the default A-type current model was replaced with each of the 243 A-type-labeled model on ModelDB (blue). Secondly, the default delayed rectified current was replaced with each of the 188 delayed rectifier models on ModelDB (red). Finally, the default A-type and delayed rectifier currents were replaced with a random sample of approximately 1% of all possible combinations of A-type and delayed rectifier models on ModelDB (purple). Summary measures are shown for total number of spikes, total number of calcium spikes, mean inter-spike interval (ISI) and coefficient of variation (CV) of ISI during the 1000msec period. Black arrows represent the simulation results for the default model.

Based on ion selectivity and gating mode, the majority of all ion channel models (~75%) fell into five classes (Figure 2.2B): voltage-dependent potassium (Kv), voltage-dependent sodium (Nav), voltage- dependent calcium (Cav), calcium-dependent potassium (KCa) and hyperpolarization-activated cation (Ih) channel models. We recorded 74 different subtype labels across all classes (Kv: 32, Nav: 19, Cav: 20, KCa: 11, Ih: 5; Figure 2.2B, cf. Figure 2.2–source data 1). Prominent modeled neuron types were *pyramidal*, *interneuron*, *granule cell*, and *basket cell* (Figure 2.2C), and prominent brain areas included *hippocampus*, and *cortex* (Figure 2.2D). Other metadata also showed a broad variety across ion channel models (ICGenealogy, 2016b).

To denote family relations (Figure 2.2A, bottom), ion channel model *A* was labeled as a “descendant” of an older “ancestor” ion channel model *B* if the publication reporting *A* cited the publication for *B* as the source or starting-point of its channel dynamics, or if the code of models *A* and *B* were sufficiently similar (Methods). By establishing a citation relationship between different models, we effectively create a *genealogy* of neuronal ion channel models, which describes their lineage (not to be confused with the actual genetic ancestry of different types of ion channels). Visualizing family relationships makes it apparent that many ion channel models form large families, often with a highly-cited hub model that has many descendants (Figure 2.2E). On the other hand, there are a large number of small families and model singletons that imply *de novo* ion channel model creation, lack of appropriate citations, or translation from other simulators (noted in the metadatum *comments*). Subtype labeling mapped well onto families, but family identity did not guarantee homogeneity of subtype or vice versa – all individual subtypes were found across several families (see Figure 2.2E for Kv, and Figure 2.4–Figure supplements 1 & 2A,F for other ion type classes).

Family relationships and metadata thus help to distinguish ion channel models, but the lack of standardized annotations in a common nomenclature, as well as the sheer abundance of models make it difficult to infer the degree of their functional diversity. Based on metadata alone, it is thus difficult to choose an ion channel model for appropriation into one’s own work.

2.4.2 Defining functional groups of models through voltage clamp protocols and clustering

To quantify the functional relationships between ion channel models, we used a set of voltage-clamp simulation protocols, in kind with those developed for the experimental characterization of ion channels and model fitting (Hodgkin and Huxley, 1952a; Willms et al., 1999; Ranjan et al., 2011). We chose this procedure to assess the spectrum of possible dynamics in a model-free manner, i.e. without explicitly taking into account the underlying equations. This allows for the comparison of ion channel models strictly based on their behavior, and, as we discuss later, the direct comparison with experimental data.

Using the *NEURON* simulation environment (Hines and Carnevale, 2001; Carnevale and

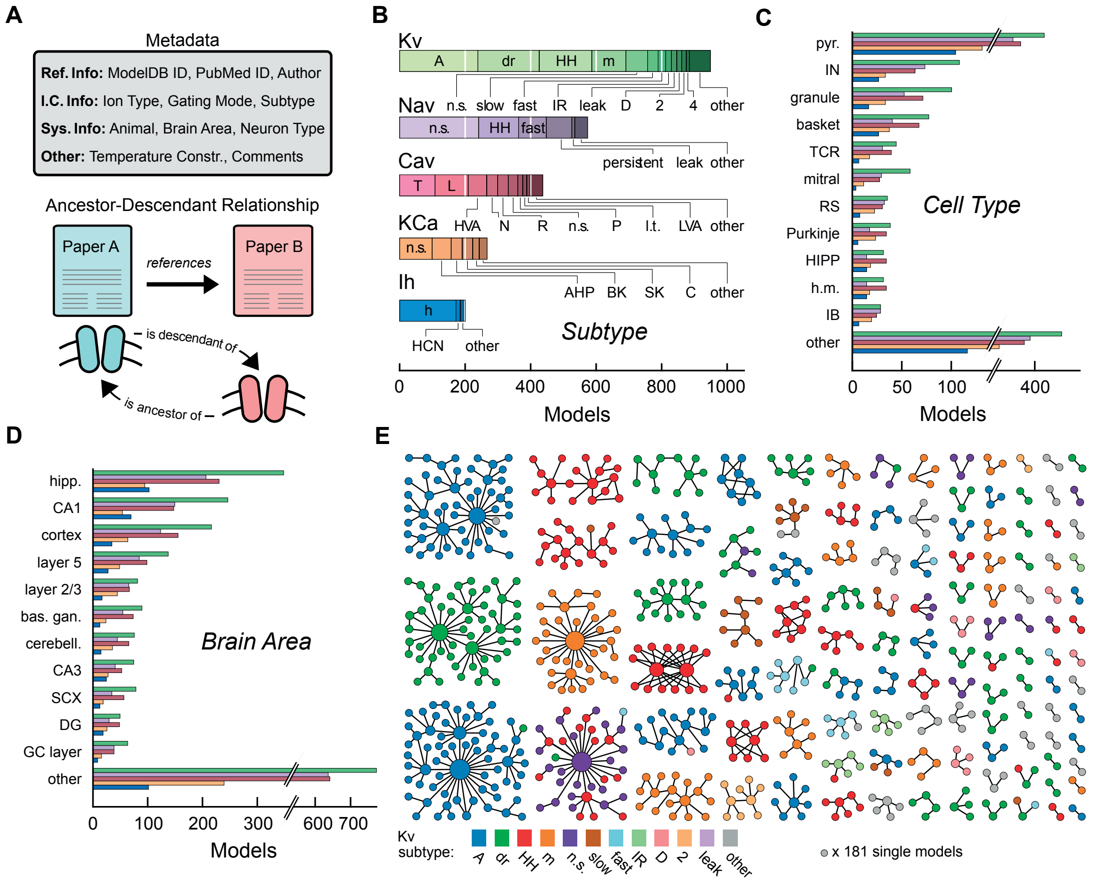


Figure 2.2 – Ion channel models can be categorized by metadata and ancestor-descendant relationships. **A**: Metadata were manually extracted from ModelDB and associated journal articles (top). Ancestor-descendant relationship (bottom) was established between different models (see main text for description). **B**: Models were divided into five classes based on ion type: voltage-dependent potassium (Kv), voltage-dependent sodium (Nav), voltage-dependent calcium (Cav), calcium-dependent potassium (KCa), and hyperpolarization-activated cation (Ih). Each class is divided into subtypes, ordered from left to right according to group size. Uncommon subtypes are grouped together (other). **C, D**: Histogram of cell types and brain areas for each ion type, ordered from top to bottom by the number of models. Colors as in **B**. **E**: Pedigree graph displaying families of the Kv class, sorted by family size. Each node represents one model, colored by subtype, and edges represent ancestral relations between models (panel **A**, bottom). Note that unconnected models (181 total) are not shown. A: A-type, dr: delayed rectifier, HH: Hodgkin-Huxley, m: m-type, n.s.: not specified, IR: inward rectifier, HVA: high-voltage activating, N: N-type, R: R-type, P: P-type, l.t.: low threshold, LVA: low-voltage activating, AHP: after-hyperpolarization, BK: big conductance, SK: small conductance, HCN: Hyperpolarization-activated cyclic nucleotide-gated, pyr.: pyramidal, IN: interneuron, TCR: thalamocortical relay, RS: regular spiking, HIPP: hilar perforant-path associated, h.m.: hilar mossy, IB: intrinsic bursting, hipp: hippocampus, bas. gan.: basal ganglia, cerebell.: cerebellum, SCX: somatosensory cortex, DG: dentate gyrus, GC: granule cell.

Figure 2.2–Source data 1. Table of subtypes for each ion type class.

Hines, 2006), each ion channel model was placed individually into a model soma and its current responses to five voltage-clamp protocols were recorded (Figure 2.3A, left; see Methods, Figure 2.3–Figure supplement 1 and Table 2.2 for full description). The protocols were designed to probe the gating characteristics of ion channels, i.e. activation, deactivation and inactivation, as well as temporal dynamics during voltage ramping and repeated action potentials. Protocol parameters were adjusted for each of the five ion classes separately. Current responses were normalized to remove the dependence on the maximum conductance, and subsampled at particular regions of interest (Figure 2.3A, dashed areas) to obtain a trace of characteristic data points for each protocol (Figure 2.3B, Figure 2.3–Figure supplement 1F). Using principal component analysis (PCA) across all traces of a particular ion channel type, we obtained a final D-dimensional score for each ion channel model, accounting for at least 99% of the variance across all channel models in each class. The dimensionality D varied between 16 and 29 dimensions for the five classes (Kv: 16, Nav: 21, Cav: 29, KCa: 16, Ih: 16). The Euclidean distance between any two given model scores was termed their “similarity” (Figure 2.3C, top). Finally, we used Ward’s clustering method (Ward, 1963) on the model scores to establish an agglomerative hierarchy of ion channel model clusters (Figure 2.3C, bottom).

A suitable number of clusters was obtained through a variety of published cluster indexes (see Methods, Figure 2.3–Figure supplements 4,5). For the Kv class, this resulted in 60 clusters with distinct responses (Figure 2.4) and small intra-cluster variability (Figure 2.3–Figure supplement 6). The other classes divided similarly into 38, 43, 44, and 26 clusters for Nav, Cav, KCa, and Ih, respectively (see Figure 2.4–Figure supplements 1,2). We named clusters according to the most common label of their members and we denoted the ion channel model closest to the mean score coordinate of each cluster as its reference model. While many clusters are relatively homogeneous in terms of subtype label, there are several that feature a mix of different subtypes (see the section on variability below). Therefore, the subtype label of clusters should be used as a guide for data exploration rather than as a strict classifier.

We found that most ancestor-descendant families fell within one cluster, indicating consistency between the family relations collected from the papers and ion channel model behavior (Figure 2.4A-B, Figure 2.4–Figure supplements 1,2B,C & G,H). However, a common subtype label did not guarantee a common cluster identity (Figure 2.4B, Figure 2.4–Figure supplements 1,2 C,H). Many models with the same subtype fell into different clusters. For example, the ~250 A-type-labeled Kv ion channel models fell into 14 clusters (although only five clusters comprised over 90% of them, Figure 2.4B). These clusters contained few other subtype labels, suggesting that *A-type* is generally a consistent label for at least five similar, yet distinct kinetic behaviors. Moreover, the similarity between these clusters was generally high (and thus they were plotted within the same vicinity on the wheel of the ‘Circos’ plot, Figure 2.4D; see also Methods). Other subtype labels across all ion channel types showed similar results (Figure 2.4–Figure supplements 1,2 C,H). Interestingly, for four of the five ion type classes (KCa being the exception), most isolated single-model clusters corresponded to genealogical singletons, supporting the idea that these ion channel models are indeed unique,

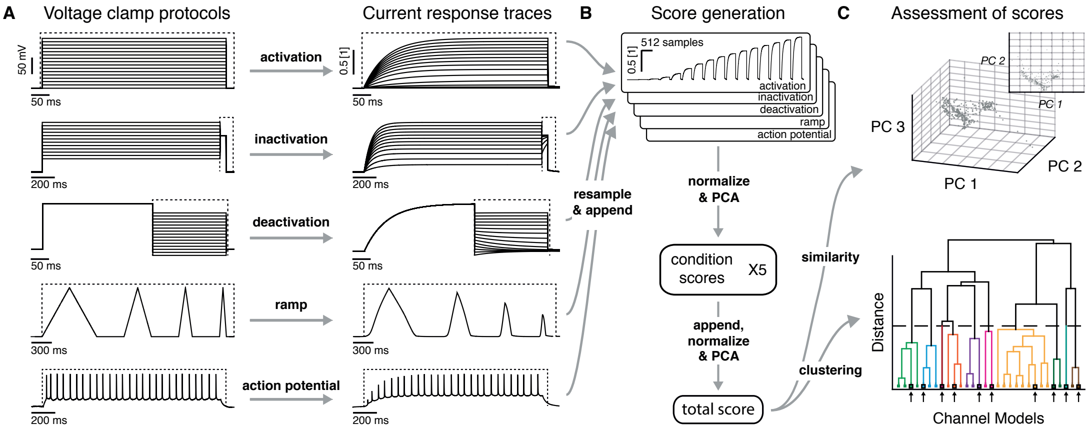

A
Voltage clamp protocols
50 mV
50 ms
200 ms
50 ms
300 ms
200 ms
activation
inactivation
deactivation
ramp
action potential
Current response traces
0.5 [1]
50 ms
200 ms
50 ms
300 ms
200 ms
B
Score generation
512 samples
0.5 [1]
activation
inactivation
deactivation
ramp
action potential
resample & append
normalize & PCA
condition scores
X5
append, normalize & PCA
total score
C
Assessment of scores
similarity
clustering
PC 3
PC 1
PC 2
PC 2
PC 1
Distance
Channel Models

Figure 2.3 – Voltage-clamp protocols for the quantitative analysis of ion channel models. **A**: Left: five voltage clamp protocols were used to characterize ion channel responses recorded in single compartment somata simulated in *NEURON* (see Figure 2.3–Figure supplement 1 and Tables 2.1, 2.2 for full description). Multiple lines indicate a series of increasing voltage steps with the same time sequence. Right: current response traces are shown for an example model. Dashed regions indicate response times used for data analysis. **B**: Current responses were subsampled and appended, then dimensionality-reduced by principal component analysis (PCA) to form a *condition score* vector for each protocol. These score vectors were further normalized and dimensionality-reduced to form a *total score* vector. **C**: The first three principal components of the score vector are shown for Kv ion channel models (top). Scores were clustered using an agglomerative hierarchical clustering technique (bottom). Distinct clusters (noted by colors) form when a cutoff (dashed line) is introduced in the distance between hierarchical groupings, chosen based on several cluster indexes (see Figure 2.3–Figure supplements 2,3). *Cluster representative models* (bold squares with arrows) are selected as reference models for each cluster (see Methods).

Figure 2.3–Figure supplement 1. Graphical description of the five voltage-clamp protocols used for ion channel model analysis.

Figure 2.3–Figure supplement 2. Cluster indexes for Kv and Nav classes.

Figure 2.3–Figure supplement 3. Cluster indexes for Cav, KCa and Ih classes.

Figure 2.3–Figure supplement 4. Comparison of intra- and inter-subtype variability with intra- and inter-cluster variability.

Figure 2.3–Source data 1. Table of omitted files.

and do not appear isolated simply due to missing ancestor-descendant links. However, this was not true for all genealogical singletons, many of which were kinetically aligned to larger families and consequently fell into the same clusters. In conclusion, clustering allowed us to identify 211 distinct groups of ion channel models that share similar behavior, regardless of publication context or subtype labeling.

2.4.3 Ion channel model groups defined by common metadata show variability in behavior

The variability in behavior of identically-labeled ion channel models in different clusters may stem from various sources. There is substantial evidence that individual neurons of any given type display heterogeneity in ion channel expression and regulation (Marder and Goaillard, 2006; Schulz et al., 2008). Furthermore, characterizing an ionic current using the average response across a population may not be sufficient to capture the appropriate behavior at the neuronal level (Golowasch et al., 2002), as there may be several distinct "solutions" (Prinz et al., 2004). Diversity and variation in ionic currents even within a single cell type may arise from such mechanisms as splice variants, differential subunit combination, and post-translational modification like phosphorylation (Schulz et al., 2008; Li et al., 2007; Campiglio and Flucher, 2015; Levitan, 1994; Misonou et al., 2004).

This biologically variability may also contribute to the diversity in ion channel models that we observe in our database. Consistent with this notion, the behavior of groups of ion channel models defined by common subtype, neuron type and brain area (Figure 2.5A, plotted data points) is often more diverse than that of models defined by a common cluster (Figure 2.5A, dashed line). More specifically, we find no clear correspondence of any given cluster with categories such as brain area and neuron type (Figure 2.5B). Nearly every cluster contains ion channel models that have been used in pyramidal cells (Figure 5B, left, blue) of both cortex and hippocampus (Figure 5B, right, blue & green). In the same vein, e.g., *A-type*-labeled models that have been used in pyramidal cells of the hippocampus (117 models) are found in 9 clusters (cf. ICGenealogy (2016b)).

A portion of the variability may also stem from non-biological sources, such as differences in experimental setup, as well as model fitting, and idiosyncratic changes to individual ion channel model implementations. Consistently, we find that models defined by common families (connected directly or indirectly through ancestor-descendant relationships) can occasionally fall into different clusters (Figure 2.4A).

It is not possible to disentangle how much of the variability in ion channel kinetics is due to each of these components. While our resource provides, for the first time, a catalogue of all models created for each system, and how they relate to one another, we remain agnostic about the sources of variability seen in the models that we analyze (cf. Discussion).

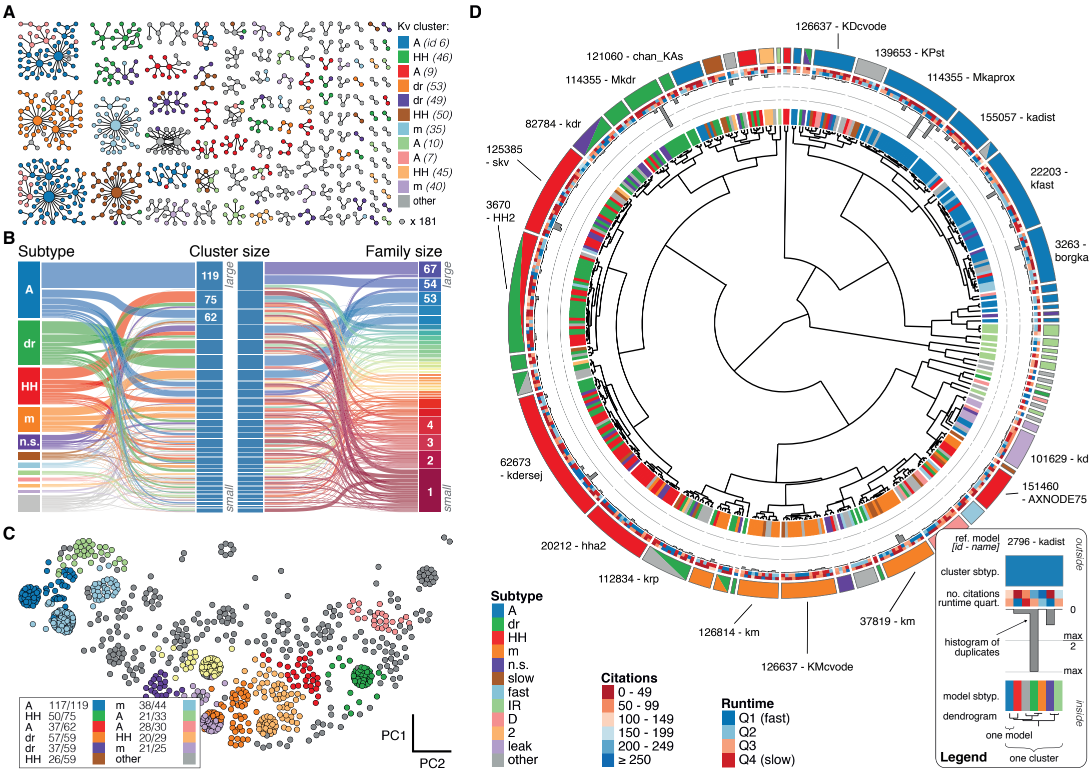


Figure 2.4 – Quantitative analysis of Kv ion channel models: functional map and clusters of common behavior. **A** Pedigree graph of the Kv class (cf. Figure 2.2E), colored by membership in the 11 largest clusters in the class (named by most prevalent subtype, bottom). Membership to other clusters is indicated by grey color. Cluster ID is given in parentheses for easy comparison with website. **B**: 'Sankey' diagram for the Kv ion type class, showing the relation between subtype, cluster identification and family identification, each ordered from top to bottom by increasing group size. The 11 most common subtypes are shown in color, with all others grouped together in grey. Small families (size 1 to 6 members) are grouped together. **C**: Plot of Kv models in the first two principal components of score space. Colors indicate membership in one of the 11 largest clusters in the class, with membership to other clusters colored in grey. Clusters are named by their most common subtype, with the proportion of that subtype specified in the legend. Points lying very close to each other have been distributed around the original coordinate for visualization. **D**: 'Circos' diagram of the Kv ion type class. All unique ion channel models are displayed on a ring, organized by cluster identification. From outside to inside, each segment specifies: cluster reference model (only displayed for large clusters), cluster subtype(s) (all subtype labels that contribute at least 30%), number of citations, runtime, number of duplicates, model subtype, as well as a dendrogram of cluster connections (black) and family relations (grey). A: A-type, dr: delayed rectifier, HH: Hodgkin-Huxley, m: m-type, n.s.: not specified, IR: inward rectifier. See Figures 2.4–Figure supplements 1 & 2 for other ion type classes.

Figure 2.4–Figure supplement 1. Nav and Cav class genealogy and clustering.

Figure 2.4–Figure supplement 2. KCa and Ih class genealogy and clustering.

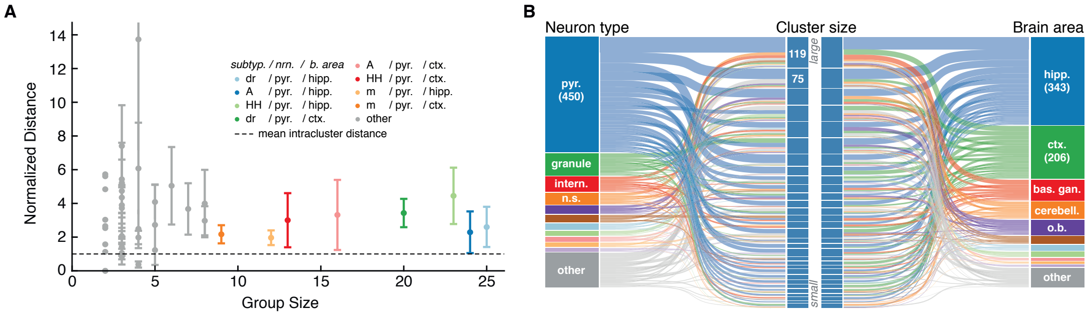


Figure 2.5 – Ion channel model groups defined by common subtype, neuron type and brain area show variability in behavior. **A**: Kv models are grouped by common subtype, neuron type, and brain area. The mean pairwise distance in score space between all models within a group is plotted against group size, with errorbars corresponding to the first and third quartiles (middle 50% of distances). The largest eight groups are shown in color. Distances are normalized relative to the mean pairwise distance between models in the same *cluster* (dashed line; averaged across all clusters with at least one pairwise distance greater than zero, 43 of 60 clusters). Most groups have a larger mean distance than the average cluster (above the dashed line). **B**: 'Sankey' relational diagram for metadata, neuron type, cluster identification and brain area, for the Kv ion type class. Bars are stacked histograms, i.e. the height of bars indicate the relative number of models. Left column: partition of channels by 11 most prevalent neuron types. Middle column: partition of channels into assigned clusters. Right column: partition of channels by 11 most prevalent brain areas. Links between columns are colored according to neuron type (left) and brain area (right). Numbers in left/right column are number of channels per label. pyr: pyramidal, intern.: interneuron, n.s.: not specified, hipp.: hippocampus, ctx: cortex, bas. gan.: basal ganglia, cerebell: cerebellum, o.b.: olfactory bulb.

2.4.4 Automated comparison of new ion channel models and experimental data

Our analysis framework, accessible through the web interface, enables the automated analysis of new ion channel models as well as experimental data (Figure 2.6A). To illustrate this process, we uploaded and tested a previously uncatalogued Kv model from a hippocampus CA1 pyramidal cell model (*kad.mod* from Hsu et al. (2015); ModelDB ID no. 184054). We compared its scores and response traces to the presently available 931 Kv models (Figure 2.6B) and determined its relation to previous ion channel implementations. We found that the ion channel model fits well within a cluster of mostly pyramidal A-type-labeled ion channel models used in simulations of rodent hippocampus, thereby verifying the assumed characteristics of the model.

The framework can also be used for the comparison of experimental data and models. To illustrate, we uploaded and tested an experimental dataset of recordings from Kenyon cells in *Drosophila melanogaster*. Voltage-gated cationic currents across the membranes of these neurons are thought to be dominated by A-type K+ channels, in particular Shal/Kv4 (Gasque et al., 2005). This renders Kenyon cells a suitable neuronal cell type to test the biological relevance of our voltage-clamp protocols in an *in vivo* setting. Current responses were recorded in targeted whole-cell patch clamp experiments *in vivo* (see Methods). The recordings were performed using our five standardized voltage-clamp protocols, allowing us to transform and compare the experiments directly to ion channel models in the same 'score' space (Figure 2.6C). The comparison revealed a close match to an existing model from our resource, and thus characterizes the behavior of the ion channel as similar to a mammalian Kv4 ion channel (Figure 2.6C; Fineberg et al. (2012); *Kv4_csi*, ModelDB ID no. 145672).

2.5 Discussion

Neuroinformatics has become an increasingly important part of neuroscience research, as new technology and large-scale research projects push the field into the realm of *big data* (Akil et al., 2011; Ferguson et al., 2014; Grillner, 2014; Tripathy et al., 2014). Importantly, the need for assessment and aggregation of published knowledge extends beyond experimental data, and has recently started encompassing computational models of neural function (Hines and Carnevale, 2004; Gleeson et al., 2013). Here, we have performed a meta-analysis of voltage- and calcium-dependent ion channel models coded in the *NEURON* programming language available in the database *ModelDB* (Hines and Carnevale, 2004). Our approach of combining metadata extracted from publications with a kinetics-based analysis allowed us to provide detailed information regarding the identity of each ion channel model in the resource, filling in missing or ambiguous data, and validating the functional properties of channel models against their sometimes ambiguous nomenclature. Furthermore, we provide a framework for the large-scale comparison of models, both with each other and with experiments using the same standardized protocols, thus paving the way towards a unified characterization of ion channel function.

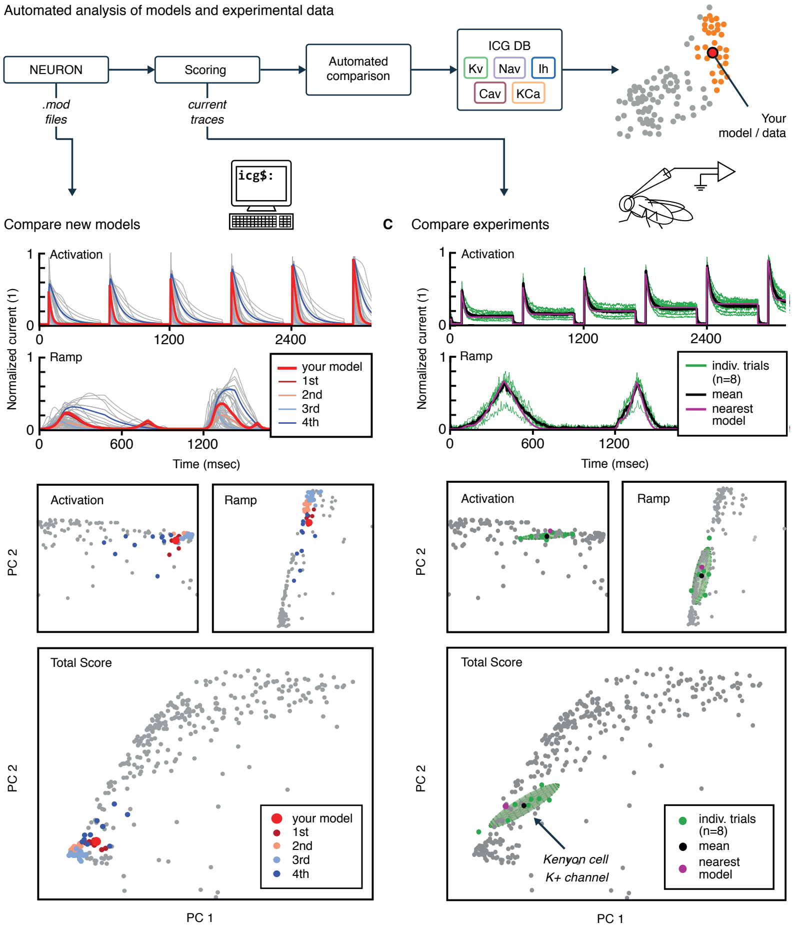
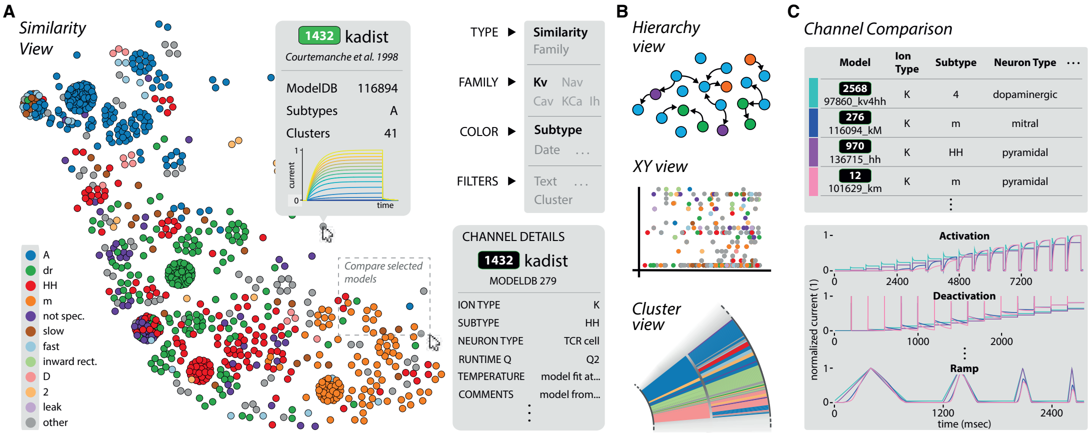


Figure 2.6 – Automated analysis of new models and experimental data. **A**: Flowchart of data processing steps involved in automated comparison. Source code for model files written in *NEURON* can be uploaded to the website, and current responses are automatically generated. Current traces are processed to compute scores, which are compared to all models in the resource (illustrated in **B**). Additionally, raw current traces obtained experimentally (or from models in other languages) can be uploaded and analyzed directly (illustrated in **C**). **B**: Example analysis and comparison of a new ion channel model (*kad.mod* from Hsu et al. (2015); ModelDB ID no. 184054). Top: Segments of the current response traces (red) for activation (voltage steps 10 - 60mV) and ramp protocols (first half), along with the closest four clusters (other colors: mean currents, grey lines: individual currents). Bottom: first two principal components of score space for activation and ramp protocols, as well as total score. **C**: Example analysis of in vivo recordings of a K+ current from Drosophila Kenyon cells (see Methods for details and Figure 2.6–Figure supplement 1 for full traces) and comparison to ICG resource. Top: mean (n=8 recordings, black) and individual recordings (green). Bottom: mean (black dot) and individual experimental recordings (green dots) plotted in the first two principal components of score space (ellipsoid illustrates the variance across individual recordings). Comparison is made to the nearest (in score space) ion channel model in the resource (magenta; *Kv4_csi*, ModelDB ID no. 145672). Grey dots in **B** and **C** are scores of Kv channel models in the resource.

Figure 2.6–Figure supplement 1. K+ current recordings from Drosophila Kenyon cells.


A
Similarity View
A
dr
HH
m
not spec.
slow
fast
inward rect.
D
2
leak
other
current
time
1
0
TYPE ▶ Similarity Family
FAMILY ▶ Kv Nav Cav KCa Ih
COLOR ▶ Subtype Date ...
FILTERS ▶ Text ... Cluster
Compare selected models
B
Hierarchy view
XY view
Cluster view
C
Channel Comparison
normalized current (1)
Activation
Deactivation
Ramp
1-
0-
0
2400
4800
7200
1000
2000
1200
time (msec)

1432 kadist
Courtemanche et al. 1998
ModelDB 116894
Subtypes A
Clusters 41

CHANNEL DETAILS
1432 kadist
MODELDB 279
ION TYPE K
SUBTYPE HH
NEURON TYPE TCR cell
RUNTIME Q Q2
TEMPERATURE model fit at...
COMMENTS model from...

Model	Ion Type	Subtype	Neuron Type	···
2568 97860_kv4hh	K	4	dopaminergic	
276 116094_kM	K	m	mitral	
970 136715_hh	K	HH	pyramidal	
12 101629_km	K	m	pyramidal	

Figure 2.7 – The ICGenealogy website allows for the interactive visualization of all data and analysis on the resource (ICGenealogy, 2016b) (http://icg.neurotheory.ox.ac.uk). **A**: Schematic of similarity view on website. Channel models of the Kv family are displayed in the first two principal components of score space, colored by subtype (legend on left). Hovering over models brings up information tooltip (center), and clicking on a model displays Channel Details (bottom right). Selected models can be compared by click-and-drag (see instruction manual in Supplementary file 1 for more details). **B**: Schematic of other three views available on the website. Hierarchy view: models are displayed in a graph with edges that represent family relations. XY view: any two selected metadata are plotted against each other. Cluster view: models are organized in a ring partitioned by clusters. **C**: The channel comparison displays selected channels side-by-side with metadata (top) and current responses (bottom).

The voltage-clamp protocols used in this study were designed to efficiently probe the kinetics of all ion channel types considered here (Hodgkin and Huxley, 1952a; Willms et al., 1999; Ranjan et al., 2011). Measuring the kinetic responses of each channel model allowed us to compare models regardless of their specific implementation. Notably, our method is amenable to the addition of other protocols that may be better suited to separate certain models. However, there is evidence that simple step and ramp current pulses are sufficient to probe the underlying kinetics of neurons (Druckmann et al., 2011), similar to the voltage-clamp protocols that we use here. Additionally, the simplicity of our protocols makes experimental comparison easier.

Additionally, our study can be extended beyond the selection of ion channel models considered here. We limited our analysis to voltage-dependent and calcium-dependent ion channel models coded in the *NEURON* language, but, given the appropriate protocols, other types of ion-channels can be included. The same protocols can also be used to integrate models written for other simulators, or even simulator-independent formats, e.g., NeuroML (Cannon et al., 2014). We have taken steps to integrate our resource and visualizations tightly with existing online resources, notably *ModelDB* (Hines and Carnevale, 2004).

The end result of our work is a dramatically reduced group of candidate ion channel models to test when looking for particular ion channel dynamics. Of the 2378 models in our resource, we could identify 1132 models as unique, and further reduced this to 211 groups with substantially different kinetics. However, this does not eliminate the task of finding the most appropriate ion channel model current, and we stress that the partition of channel models into clusters of similar response properties does not imply that models in the same cluster are necessarily redundant. Clustering is not trivial, and while we have used several measures to determine an appropriate partition of models, we cannot escape a certain level of ambiguity. Intra-cluster differences may still be important depending on the particular simulation at hand. Since the responses of different channels vary slowly and continuously rather than in discrete steps along the dimensions of the manifold of scores (see e.g. Figure 2.4B), the data may also be amenable to more sophisticated clustering and machine-learning approaches. To this end, the raw response data and scores have been made publicly available (ICGenealogy, 2016b).

Furthermore, the variability in behavior of ionic currents seen here and elsewhere (Marder and Goaillard, 2006) suggests that there is no clear answer to the question of which ion channel model (and parameters) to use for a given neuronal simulation at hand. However, for modellers who would like to be more diligent about the sources of ion channel models and the comparability across models of the same underlying biological phenomenon, our resource takes a step in this direction.

On a larger scope, it has been suggested previously that neuron model parameters should be viewed as regions, rather than as individual representative points in parameter space (Goldman et al., 2001). It may be possible that the variability seen at the level of individual ion channel models covers such regions, and organizes into a handful of distinct “solutions” to particular model behaviors on the neuronal scale (Leary et al., 2014). In this sense, our database would lend itself to systematic analyses of the co-variability of sets of published ion channel models that are able to elicit desired behaviors in conductance-based neuron models (see also Figure 2.1), in line with previous work (Prinz et al., 2003).

We provide an interactive browser (ICGenealogy, 2016b), which acts as a complement to existing resources such as ModelDB. It allows the comparison of channel models in five views: a similarity view focusing on the channel’s response kinetic scores (Figure 2.7A), a hierarchical tree view focusing on genealogical data (Figure 2.7B, top), an XY view to sort data by a given set of metadata dimensions (Figure 2.7B, middle), and a circular cluster view (Figure 2.7B, bottom). All these views feed a central comparison tool (Figure 2.7C), in which the metadata and traces for user-selected channel models can be viewed side-by-side. For specific examples of how to utilize this browser and to search for specific ion channel models, please refer to the instruction video (ICGenealogy, 2016b) and manual (Supplementary file 1).

Because our voltage-clamp protocols are inspired by experimental procedures, ion channel models can be compared directly to experiments in an automated fashion. We have taken the first steps in this direction by showing a comparison of both a new model and an experimental dataset to the resource here (Figure 2.6). While it is beyond the scope of the current study to integrate ion channel information from IUPAR/BPS Guide to Pharmacology (Pawson et al., 2014), Channelpedia (Ranjan et al., 2011) or other sources, these are important future steps which would help standardize nomenclature. Beyond its usefulness for cataloguing ion channel model behavior and pedigree, our resource will enable better experimentally-constrained modeling, and presents a first step towards a unified functional map of ion channel dynamics in model and experiment.

2.6 Methods and Materials

2.6.1 Pyramidal cell model (Figure 1)

A model of a layer 2/3 pyramidal cell was adapted from a previous study (Traub et al., 2003), ModelDB ID no. 20756. It contained 68 soma-dendritic compartments and 6 axonal compartments, with the following active conductances: leak (g_L), transient (inactivating) Na+ (g_{NaF}), persistent (noninactivating) Na+ (g_{NaP}), delayed rectifier K+ (g_{KDR}), transient inactivating K+ (g_{KA}), slowly activating and inactivating K+ (g_{K2}), muscarinic receptor-suppressed K+ (g_{KM}), fast voltage- and calcium-dependent K+ (g_{KC}), a slow calcium-dependent K+ (g_{KAHP}), low-threshold inactivating Ca2+ (g_{CaT}), high-threshold non-inactivating Ca2+ (g_{CaH}), hyperpolarization-activating cation conductance (g_{Ih}). We refer the reader to Traub et al. (2003) for channel kinetics, distribution and other details of the model.

The neuron model was simulated in the *NEURON* simulation environment (Hines and Carnevale, 2001; Carnevale and Hines, 2006), with a current step input injected into the apical dendrite, following Figure 2 of Traub et al. (2003). The protocol was as follows: 400 msec at -0.15 nA, followed by 1000 msec at 1.5 nA. A subset of this trace is shown in Figure 2.1B, comprising 50 msec at -0.15 nA and the first 200 msec at 1.5 nA. The gray traces in Figure 2.1B show the default behavior of the neuron model in response to injected input. The following four measures were computed for each spike train: total number of spikes, total number of calcium spikes, mean inter-spike interval (ISI) and coefficient of variation of ISI.

The stimulation paradigm was repeated in the presence of alternate ion channel models for g_{KA}, taken from ModelDB (243 models total). All other ionic conductances, parameters and distributions remained the same. This was further done with alternate ion channel models for g_{KDR} in a separate simulation (188 models total), with g_{KA} set back to the original model. Finally, this was done in the case of replacing both g_{KA} and g_{KDR} currents. A random subset of approximately 1% (441 of the total 45684) of pairs of alternate ion channel models were run together.

2.6.2 The *ModelDB* database, *NEURON* language and nomenclature

The *ModelDB* database archives published neuron and network models (Hines and Carnevale, 2004). It contains over 1000 entries, with thousands of ion channel models. At the time of analysis, 496 of the entries were implemented in the *NEURON* language (Carnevale and Hines, 2006; Hines and Carnevale, 2001), making it the most used simulation environment on the database. Customizable ion channel models are coded in *NEURON* in so-called *.mod* files (with suffix ".mod") (Hines and Carnevale, 2000). Mod files for all *NEURON* entries were downloaded from the *ModelDB* website. Each mod file was renamed by adding the *ModelDB* ID as a prefix to the file, in the following way: *ID_name.mod*, where *ID* is the *ModelDB* ID, and *name* is the original name of the *.mod* file. 46 *.mod* files contained more than one current of the same ion type, and were separated into distinct files for each one. The suffix "icgXY" was appended to the name, where X was the ion type, and Y was the number of the current, beginning with 1 (e.g., 1234_kv_icgK2.mod for the second Kv current in file 1234_kv.mod). Furthermore, some ModelDB entries contained more than one file with the same name. These files were added separately and given unique names by appending a version number to the name – e.g. "_v2" for the second file. The total number of files collected from the database was 3495.

2.6.3 Collection of metadata

Metadata information was collected using all information in journal articles and files associated with each *ModelDB* entry of interest (SOM). Each field is listed below and defined. Note that some channel models may have missing entries for information that was not stated explicitly in the journal articles or *ModelDB*. Further, we stress that metadata items corresponding to the intended neuron type, brain area and animal are strictly associated to the modeling context, and are not necessarily representative of the experimental ion channels found in that particular neuron, brain area or animal.

- **ModelDB ID**. Identification number associated with each entry on *ModelDB*. All channels from the same entry have the same number.
- **PubMed ID**. PubMed citation ID of journal articles associated with this channel's *ModelDB* entry; may contain multiple elements, and can be empty for a select few *ModelDB* entries for which no articles were found on PubMed.
- **ion type**. The ion type, or permeability, of the channel model, as listed in the journal article and *.mod* file. The following ion types were analyzed: potassium (K), sodium (Na), calcium (Ca), nonspecific (NS). If models contained more than one current, all ion types were recorded separately. Other ion types were registered but not included in the analysis.
- **gating mode/mechanism**. The dynamic simulation variable that modulates the kinetics of the model, such as voltage (v), calcium (ca), voltage and calcium (v/ca),

sodium (na), chloride (cl), light (o), and g-protein coupled (g). Only v channel models were included in the analysis, with the exception of ca and v/ca models exclusively for the K ion type.

- **subtype**. The listed ion channel type, as detailed in the journal article or the *.mod* file. Subtypes were listed as mentioned without conformation to any naming convention, e.g., (Ashburner et al., 2000; Yu and Catterall, 2004). If no subtype was given, then the subtype was recorded as *not specified*. A full list of all recorded subtypes is found in Figure 2.2–source data 1.
- **author**. Listed author(s) of the *.mod* file (programmers). If authors were not specified in the *.mod* file or on *ModelDB*, we recorded the field as *not specified*.
- **animal**. The animal model (and age, if specified) emulated in simulations, either stated explicitly, or inferred from the journal article.
- **brain area/layer**. The emulated brain area and layer of the simulation, as stated explicitly, or inferred from the journal article.
- **neuron type**. The emulated neuron type of the simulation. May be several types, or listed as *general* if no neuron type was specified.
- **neuron region**. The neuron region that the ion channel is found in, divided into dendrites, soma, axon, axon hillock, or specific areas of dendrites or axon.
- **comments**. Comments from the *.mod* file itself and any other information about the channel and model from the journal article, such as previous models or experimental data that were used to constrain the model.
- **runtime [ms]**. Elapsed CPU time for running 10 repetitions of a single voltage-clamp protocol (action-potential). In plots and on the web interface, we simplify model runtimes by assignment to one of four quartiles of the distribution of runtimes of all models in each class.
- **temperature**. Details about the model's temperature dependence, and also the temperature at which simulations and/or experiments were performed as described in the journal article.
- **citations**. Estimated number of citations as available through Google Scholar, scraped monthly to update the entries.

A total of 3495 *.mod* files were collected from *ModelDB*. 366 of these files were tools, full neuron models, or other items that do not function as ion channel models. Out of the remaining 3150 files, *.mod* files were placed into one of five groups: voltage-dependent potassium (Kv), voltage-dependent sodium (Nav), voltage-dependent calcium (Cav), calcium-dependent potassium (KCa), and hyperpolarization-activated cation (Ih). The calcium-dependent potassium group contained both ca and v/ca channel models without any distinction. These five groups accounted for 2378 files, references for which are available in Supplementary file 2. Mod files that did not fit this description were omitted from the analysis. This included pumps and active dynamics (290), receptor models (370), and models with other gating dependencies or ion types (68).

We note that the voltage-dependence of ion channels may stem from different underlying mechanisms. This includes traditional voltage-gated ion channels that contain a voltage-sensitive domain (Catterall, 1995), as well as dependence that occurs indirectly through interaction with other molecules such as polyamines and magnesium (Nichols and Lopatin, 1997), or intracellular signalling cascades (Kase and Imoto, 2012). The ion channel models considered here are often agnostic to the biophysical mechanism, and different types of models can be used to model an arbitrary voltage-conductance relationship (Destexhe and Huguenard, 2000). Therefore, voltage-dependence as discussed in this work refers to the functional relationship between voltage and channel conductance, and does not generally depend on any particular biological mechanism.

2.6.4 Ancestor-descendant relationships

The genealogy of ion channel models was defined by an ancestor-descendant relationship. Each channel model was linked to previous models if a relationship was listed in the journal article. We denoted the groups of models connected by these relations as “families”. This relation could be specific, along the lines of “the A channel model’s kinetics were adapted from the B channel model in a previous journal article”. Other times, the description was vague, e.g., stating that the neuron model was adapted from a previous one, with no explicit reference to ion channel kinetics. When no information was listed about previous kinetics, both in the journal article and model files themselves, the channels were assumed to have no ancestors and to have been created de novo. However, in many situations obvious similarity in *.mod* file code as verified by a `diff` command was sufficient to link models to previous ones. In these cases, relations were established even when they were not stated in the journal articles. This task was done by hand, and as such is prone to mistakes. We repeated the collection of metadata, including ancestry relations, a second time in order to correct for potential errors – we hope to correct any remaining missing or superfluous ancestral relations with the help of user submissions (ICGenealogy, 2016b).

We note that the use of the word *genealogy* in this work is used exclusively to describe the ancestry of ion channel *models*. It does not refer to the genetic lineage of ion channels as found in biology. Furthermore, this genealogy does not necessarily conform to a pedigree as defined by journal article reference information, as sometimes ion channel models are combined from several papers, or references may be missing.

Table 2.1 – Parameters for reversal potential and inside and outside concentrations used in simulation protocols for five ion type classes. Ionic concentrations were not used for Ih currents.

	E_{rev} (mV)	$[ion]_{in}$ (mM)	$[ion]_{out}$ (mM)
Kv	-86.7	85.0	3.3152396
Nav	50.0	21.0	136.3753955
Cav	135.0	8.1929e-5	2.0
KCa	-86.7	85.0	3.3152396
Ih	-45.0	-	-

2.6.5 Voltage-clamp protocol

Mod files were run individually in a *NEURON* simulation by generating a single soma compartment of length and diameter equal to $20\,\mu m$ and cytoplasmic resistivity of $150\,\Omega cm$. A passive conductance was set to $3.334 \cdot 10^{-5} S/cm^2$. The simulation temperature was set to 37 °C. Reversal potentials were specified separately for each ion type. Some models (172 files) featured explicit calculation of the reversal potential, so internal and external concentration values were added as extra variables to make these equivalent. Parameter values for reversal potential and ion concentrations can be found in Table 2.1.

A particular model was placed in the soma and a series of five voltage clamp experiments were run (Figure 2.3–Figure supplements 1 & 2), with the current output being recorded. Based on the desired effect of each protocol, only particular sections of the protocols were used in comparing the kinetics (Figure 2.3–Figure supplement 1, dashed lines; also noted in Table 2.2). The activation protocol featured a single voltage step level, meant to capture the activation kinetics of the model. The inactivation protocol featured a varying voltage step for a long duration, followed by a second fixed voltage step, measuring the inactivation due to the first step. The deactivation protocol featured a single voltage step at a high voltage, followed by a second voltage step of varying amplitude, meant to measure the deactivation kinetics that occur as the voltage is changed from one level to the other. The ramp protocol featured a series of four up and down ramping voltages, at different slopes. Finally, the action potential protocol features voltage deflections as recorded from the soma of a neuron exhibiting a regular spiking pattern. This was recorded from a L2/3 pyramidal neuron of P14 rat somatosensory cortex (R. Ranjan, unpublished). Each of the five ion type groups featured different voltage values and durations based on differences in time constants, reversal potentials, and voltage ranges at which each ion channel class is known to be active – no quantitative comparisons were made between classes. Additionally, calcium gated channel models were simulated at seven different calcium concentrations based on known concentrations (Neher and Sakaba, 2008). Values were expressed in concentration as $10^{-x} mM$, with x taking the following values: $2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0$. Voltage-clamp protocols are available for download from the ICG website (ICGenealogy, 2016b).

Table 2.2 – Voltage-clamp protocol parameters for the five ion type classes. Times are stated in units of ms, voltages in units of mV. See Figure 2.3–Figure supplement 1 for graphical description. Items T_A^* and T_B^* represent the starting and ending times, respectively, of the regions used for analysis (dashed areas in Figure 2.3 of the main text, as well as Figure 2.3–Figure supplement 1).

	Ion Type	V_0	V_1	V_2	ΔV	T_1	T_2	T_3	T_A^*	T_B^*
Act	Kv	-80	-80	70	10	100	500	100	100	700
	Nav	-80	-80	70	10	20	50	30	18	100
	Cav	-80	-80	70	10	100	500	100	98	700
	KCa	-80	-80	70	10	100	500	100	95	605
	Ih	-40	-150	0	10	100	2000	100	95	2105

	Ion Type	V_0	V_1	V_2	V_3	ΔV	T_1	T_2	T_3	T_4	T_A^*	T_B^*
Inact	Kv	-80	-40	70	30	10	100	1500	50	100	1600	1700
	Nav	-80	-40	70	30	10	100	1500	50	100	1580	1750
	Cav	-80	-40	70	30	10	100	1500	50	100	1580	1750
	KCa	-80	-40	70	30	10	100	1500	50	100	1595	1700
	Ih	-40	-150	-40	-120	10	100	1000	300	100	1095	1405

	Ion Type	V_0	V_1	V_2	V_3	ΔV	T_1	T_2	T_3	T_4	T_A^*	T_B^*
Deact	Kv	-80	70	-100	40	10	100	300	200	100	400	600
	Nav	-80	70	-100	40	10	20	10	30	20	29	80
	Cav	-80	70	-100	40	10	100	300	200	100	380	700
	KCa	-80	70	-100	40	10	100	300	200	100	395	605
	Ih	-40	-140	-110	0	10	100	1500	500	400	1595	2105

	Ion Type	V_0	V_1	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_A^*	T_B^*
Ramp	Kv	-80	70	100	800	400	400	400	200	400	100	100	100	2800
	Nav	-80	70	100	800	400	400	400	200	400	100	100	98	2800
	Cav	-80	70	100	800	400	400	400	200	400	100	100	98	2800
	KCa	-80	70	100	800	400	400	400	200	400	100	100	100	2800
	Ih	-80	70	100	800	400	400	400	200	400	100	100	100	2800

	Ion Type	T_1	T_A^*	T_B^*
AP	Kv	1800	100	1800
	Nav	1800	98	1800
	Cav	1800	98	1800
	KCa	1800	95	1655
	Ih	1800	95	1655

A substantial number of *.mod* files (952 files) had to be slightly modified to work with the procedure, in one of the following ways: (1) reversal potential was renamed and made a global variable to be accessed from *.hoc* file, (2) NONSPECIFIC CURRENT was changed to a USEION statement with the correct ion type (3) extra functions and/or data were included through *.inc* files, data tables or extra *.mod* files (4) max conductance was made nonzero (arbitrarily set to 1.0) (5) file was split into multiple files for each current present (6) POINTER variables were removed (7) internal temperature initialization was removed, and temperature dependence was set to use the global variable “celsius”. The changed files will be made available upon publication.

All ion channel models were taken from the *ModelDB* repository, as published, with small changes as noted above. This included the assumption that model parameters as chosen by the authors were set to reproduce a given, desired, dynamical behavior which matches experimental data or other constraints. We did not consider changing the internal parameters of given models – this would prevent any feasible comparison of the models, since most given models would be able to generate a large variety of behaviors under changing parameter settings (data not shown).

A small number of *.mod* files were omitted from the analysis due to problems with running the simulation protocols (44 files, Figure 2.3–Supplementary data 1). These included files that did not compile for unknown reasons, and files that produced abnormal oscillations, or extreme values. The count of these files was 16 for Kv, 20 for Nav, 5 for Cav, 2 for KCa, and 1 for Ih.

2.6.6 Data extraction & processing

The most recent version at the time of writing, version 7.3, of the *NEURON* language (Carnevale and Hines, 2014) was used to run the simulation protocols. All simulations were run in Ubuntu 12.04.5 on a single core of a Intel Core i7 @ 2.67GHz with 24 Gigabytes of RAM. *NEURON* models were injected with the five different voltage clamp protocols described above and integrated at a timestep of $dt = 5e^{-2}ms$. The resulting current traces were processed in the following manner. First, inward currents (represented as negative deflections from baseline) were flipped by multiplying the entire trace by -1. Next, all current traces were normalized by dividing by the maximum trace value in order to make the result invariant to the max conductance parameter (e.g., $\bar{g}$ in Hodgkin-Huxley sodium current $I_{Na} = \bar{g}m^3h(V_m - E_{Na})$). We reasoned that the maximum current amplitude depends on the number of channel models in a particular area and was thus not related to the kinetic behavior of the channels. The traces were then subsampled at a resolution of 512 data points within the regions of interest stated above. Finally, for protocols containing graded steps (activation, inactivation, deactivation) the subsampled responses across all c graded voltage steps were appended into one representative vector of length $L = 512 * c$. For calcium gated channels, we performed a similar procedure for each of the k calcium concentrations separately, and then appended them into one representative vector of length $L = 512 * k * c$. See Figure 2.3–Figure supplement 1F for a schematic of this process.

2.6.7 Similarity measure

To remove the time dependence of current response waveforms, we performed discrete principal component analysis (PCA)(Ramsay and Silverman, 2005) across the temporal dimension, similar to approaches in spike sorting (Lewicki, 1998). To this end, the subsampled and appended current responses for each protocol across all N channels in a family yielded a NxL dimensional data matrix, in which we normalized each column by Z-scoring: we subtracted its mean and then divided by its standard deviation. This matrix was then dimensionality reduced by PCA across the L temporal entries, where we chose the reduced dimensionality to capture 99% of the variability. To normalize the range of scores across conditions while keeping the covariance structure, we divided the score vector of each protocol by the standard deviation of all score entries of this protocol. These normalized scores (denoted by *condition scores*) were then combined into a final score vector. Further correlations across protocols were removed by again dimensionality-reducing by PCA (99% variance criterion) to yield a final score vector for each model. Since response traces were relatively noise-free, a high PCA dimensionality can be chosen to capture current response dimensions that are rare across the population of models. The precise value of the variance criterion in both PCA steps, although slightly changing the resulting scores, did not affect the clustering results reported above. Similarity between two channel models was then defined as the Euclidean distance of their dimensionality reduced scores.

In summary, the principal components calculated in the first step represented response curves along which the current traces were projected to yield intermediate scores. The second transformation was a linear mixing matrix, which combined these intermediate scores. Final scores had between 16 and 29 dimensions depending on the family analyzed, which additionally allowed the efficient storage of the characteristics of the thus compressed response properties in our resource. The linear PCA transformations, once calculated, can be applied to additional channel models and their current responses and allow us to efficiently score new channels and easily evaluate them against all other channels in the resource (ICGenealogy, 2016b).

2.6.8 Clustering

For clustering of channel scores, we used Ward's minimal variance linkage (Ward, 1963) for hierarchical clustering, as implemented in the MATLAB Statistics Toolbox (R2015A, The MathWorks Inc., Natick, MA). This method can be used to produce a division of the set of all channel models into an arbitrary number of "similar" clusters, the number of which has to be constrained by internal criteria (we assumed no a-priori existence of classes in this dataset) (Halkidi et al., 2001). To this end we employed a range of internal clustering evaluation measures, which indicate the emergence of an appropriate number of clusters. Although the evaluation of these measures requires some heuristics, they have been well established and can guide the decision as to which number of clusters to choose. Concretely, these are: the Silhouette criterion (Rousseeuw, 1987), the Dunn index (Dunn, 1973), the Davies-Bouldin index (Davies and Bouldin, 1979), and the Calinski-Harabasz measure (Caliński and Harabasz, 1974), also implemented in the MATLAB Statistics Toolbox (R2015A, The MathWorks Inc., Natick, MA). For the Dunn index, the Silhouette index and the Calinksi-Harabasz measure, high values indicate mostly compact and well-separated clusters. The Davies-Bouldin index also indicates compactness and separation, however for low values. For details and reviews on these clustering indexes see e.g., Milligan and Cooper (1985) and Halkidi et al. (2001).

Values for the indexes and heuristics applied to arrive at the cluster numbers of the main text are given in Figures 2.3–Figure supplement 4 & 5. Due to the natural partition of our dataset into five conditions used to calculate the final score, we also included a measure of heterogeneity computed directly on the traces of each condition, which we termed the “Inner distance”. For a given condition, let $\mathbf{s}_j \in \mathbb{R}^{512*c}$ be the appended (over all possible voltage steps c) subsampled current responses, where $j \in \{1,\ldots,n_{\text{channels}}\}$ runs over all channels. Let $\{C_k | k \in \{1,\ldots,n_{\text{clusters}}\}\}$ be a clustering of all channels – a collection of sets, such that each channel index j is contained in a single set. Let $\mathbf{c}_k = \frac{1}{|C_k|}\sum_{j\in C_k}\mathbf{s}_j$ be the mean response trace of each cluster. The inner distance is then calculated as the scatter around the mean, averaged over all clusters:

$$d_{\text{inner}} = \frac{1}{n_{\text{clusters}}}\sum_k \frac{1}{|C_k|}\sum_{j\in C_k}\|\mathbf{s}_j - \mathbf{c}_k\|. \tag{2.1}$$

To make the measure comparable across different conditions, which might have different values of c (the number of voltage steps), we define the norm as $\|\mathbf{s}_j\| = \frac{1}{c*512}\sum_{i=1}^{c*512}|s_j(i)|$.

A number of additional linkage methods (complete, single, average) and metrics (cityblock, squared euclidean) were also evaluated. While giving comparable performance on a synthetic test set, they yield mostly inferior subsections of the full set of channels with very high numbers of single elements being isolated as separate clusters.

2.6.9 Assessment of protocols

To qualitatively assess the necessity of the voltage-clamp protocols for separation of labeled subtypes, the condition scores of all channel models of a particular subtype were compared with those of other subtypes (Figure 2.3–Figure supplement 6). We show that certain protocols are more important for differentiating particular subtypes: for example, Kv models of the *m-type* show a large distinction from *A-type*, *dr* and *HH* subtypes in the condition scores of the action potential protocol, whereas *A-type* channel models show distinct condition scores in the activation, inactivation and deactivation protocols. The protocols chosen here thus exploit a necessary range of response kinetics; the general method of deriving a final score from each of the conditions, however, is amenable to straight-forward extension by further protocols or second-order features extracted from the response traces, as for example peak response values and time-scales (Lewicki, 1998; Druckmann and Chklovskii, 2012). Each of these could be incorporated in the analysis as additional condition scores.

An alternative for the characterization of ion channel dynamics would be given by a model-based approach: by fitting the parameters of a single *super-model* to closely approximate the dynamics of all other channel models on hand, one could characterize channel models by the resulting parameter values. However, we could not assume that such a single super-model would capture the full dynamical diversity we were presented with in our dataset. Similarly, the diverse kinetics of biological ion channels measured in experiments might not be captured adequately by a single super-model. We thus chose a standardized "model-free" approach, which can be extended (see above) should the need for more detailed discrimination of channel kinetics (from model and experiment alike) arise.

2.6.10 Generation of Circos plots

'Circos' plots displaying the clustering results together with genealogical links were generated using the Circos visualization tool (Krzywinski et al., 2009), combined with TreeDyn (Chevenet et al., 2006) to create circular dendrograms (Figure 2.4D).

The Circos plot is a visualization technique that enables the comparison of functional similarity to metadata information for each cluster separately. All *unique* models of a given class are arranged in a circle (numbers of duplicate models are shown in the gray histogram along the circle). *Plot rings*: For each model, the following information is shown (from inside to outside; refer to legend): clustering dendrogram, subtype label, histogram count of duplicate models, model runtime information, number of citations of the accompanying paper, and most common subtype label(s) of each cluster (all subtypes that contribute 30% or more). Location along the circle corresponds to functional (clustering) information, whereas color corresponds to metadata information, such as subtype label. *Location*: position along the circle was established by the circular dendrogram at the center. This dendrogram was created with an agglomerative hierarchical clustering algorithm as defined above, and arranges models in such a way that similar models are in adjacency, and all models in the same cluster appear in one continuous group. The outer ring of the plot denotes the extent of each individual cluster. Groups of models defined by cluster were visually displaced from others by adding a small white space between clusters. *Color*: Three color legends accompanying the graph define the color relationships plotted. The two large rings on the inside and outside are colored by subtype label (of individual models and of clusters, respectively), of which the 11 most common are displayed, with all others in gray. Two smaller rings just inside of the outer ring, denoting number of citations and runtime, are colored on a red-blue scale.

2.6.11 Generation of genealogy figures

Pedigree plots were generated using Gephi (Bastian et al., 2009), and then manipulated and ordered manually for visualization (Figure 2.2E, Figure 2.4A). Coloring was chosen according to subtype label as well as cluster identity. 'Sankey' diagrams (Sankey, 1898; Schmidt, 2008) were generated in Javascript and D3.js (Bostock et al., 2011) (Figure 2.4B, Figure 2.5B). Subtype coloring was chosen as for the pedigree plots. Subtype labels, clusters and families were arranged from top to bottom by size.

All other figures were generated using MATLAB (R2015A, The MathWorks Inc., Natick, MA) and Python 2.7 with matplotlib 1.4.2.

2.6.12 Relational database, API and web interface

All collected metadata, as well as final scores and clustering results were organized in a relational MYSQL database, which is openly queryable through a web API. Details about database structure and the implementations of the web-application and API will be made available on the website (ICGenealogy, 2016b). The graphical channel browser frontend was developed in Javascript and D3.js (Bostock et al., 2011) by Phyramid Ltd, Bucharest, Romania.

2.6.13 Code availability

Code for the generation of current response traces in *NEURON* as well as for the analysis of current traces will be made available on the website (ICGenealogy, 2016b).

2.6.14 Electrophysiology

K+ currents were recorded from Drosophila Kenyon cells in targeted in vivo whole-cell voltage clamp experiments as previously described (Murthy and Turner, 2013). Male NP7175-GAL4;UAS-mCD8-GFP flies were immobilized and fixed to a perfusion chamber using wax. Cuticle, adipose tissue, trachea and perineural sheath were removed in a window large enough to expose the posterior brain. The preparation was continuously superfused with extracellular solution containing (in mM) 103 NaCl, 3 KCl, 26 $NaHCO_3$, 1 NaH_2PO_4, 1.5 $CaCl_2$, 4 $MgCl_2$, 5 TES, 8 trehalose, 10 glucose and 7 sucrose (pH 7.3 when equilibrated with 5% CO_2 and 95% O_2). Tetrodotoxin was added at a final concentration of $1\,\mu M$. Borosilicate glass electrodes (14-16 MΩ) were filled with pipette solution containing (in mM) 140 potassium aspartate, 1 KCl, 10 HEPES, 4 MgATP, 0.5 Na_3GTP and 1 EGTA (pH 7.3). All experiments were performed at room temperature ($21-23\,^\circ C$). Signals were recorded with a MultiClamp 700B Microelectrode Amplifier, lowpass-filtered at 10 and digitized at 50 kHz using a Digidata 1440A digitizer controlled via the pCLAMP 10 software (all Molecular Devices). Capacitive transients and linear leak currents were subtracted using a P/4 protocol and all traces were corrected for the liquid junction potential (Neher, 1992). Voltage pulse protocols were applied as indicated for Kv (Figure 2.3–Figure supplements 1 & 2) and data were analyzed in MATLAB. Resulting current traces were processed analogously to model current traces, as specified in section *Data extraction & processing*.

2.7 ICGenealogy web application

In this section, I describe the structure and implementation of the components that form the IonChannelGenealogy (ICG) web application (ICGenealogy, 2016b). This summary was not part of the publication.

The ICG web application can be separated into three parts (see Fig. 2.8A). First, a *SQL database* stores all ion channel models currently available on ICG, including metadata and quantitative data (traces) for each model. Second, a *web application* accesses this database, constructs views from the data and delivers these to the website user through a web server – the application also provides a web application programming interface (API). Finally, a *graphical browser* accesses the ICG database through an and locally creates visualizations on the device of the user. The graphical browser has been developed in collaboration with a third party, and will only be discussed shortly below (in Section 2.7.1).

2.7.1 Software stack & web application

Here, I describe the software stack used to provide the ICG application to the public. The later parts of this section are split by the specific functionality provided through the stack. See Fig. 2.8B for an illustration of how the different parts interlink. Version numbers for the software used are omitted from this text to prevent any possible security exploits, but are available from the author upon request.

The ICGenealogy web application is built around the core of Django (Django, 2017), which is a free and open-source Python based web framework. Django is extensively maintained, widely used, and offers hands-on solutions for several tasks which would otherwise require extensive additional work, e.g. the handling of user accounts and user access to the website. The externally accessible functionality of the web application (Fig. 2.8A, right hand side) is generated by application code written for Django, but has to be sent to the user securely and efficiently. For this, Django lies embedded in a stack of additional free open-source software (Fig. 2.8B).

All software is compiled and run on the Ubuntu Linux operating system (Ubuntu, 2017) using the Supervisor (Supervisor, 2017) process controller for convenient monitoring and starting of the application. External user requests are passed through the NGINX web-server, which serves static files (e.g. cascading style-sheets, javascript files, images) directly and relays all application requests further down the stack. Requests to the main site (not the graphical browser) are received by uWSGI (uWSGI, 2017), an open-source application server implementing the wsgi (Eby, 2003) specification, which allows it to communicate with the Python interpreter. Finally, through Python the actual web application is served by Django, which accesses and modifies the relational database running on the open-source MySQL (MySQL, 2017) database server.

Most requests relayed to Python/Django depend on the current state of the database, which in the current application changes on slow time scales compared to the user traffic. We thus cache responses to many requests (especially to API calls, see below). For this, Django stores request-response pairs in system memory (RAM) using the Memcached (Memcached, 2017) caching backend.

Web API An application programming interface (API) is used make the ICGenealogy database accessible by JSON encoded textual representations. The Django framework treats responses to user requests in conjunction with the Django-REST-framework (Django-REST, 2017). This allows using the Python-internal objects that represent the database contents to construct the responses. All request-response pairs to the API are stateless, i.e. they depend only on the contents of the request, and not on earlier requests of the same user. This enables us to cut down on the time required to compute responses by using the Django caching backend (see above).

User submission The functionality of user submission of models and experimental traces was written in Django. After files have been uploaded and locally stored, NEURON (Hines and Carnevale, 2001) simulations, the analysis of data produced by NEURON, and the generation of figures using the analysis results, have to be performed. To prevent the Django application from blocking until these tasks are finished it is desirable to provide this functionality running “in the background”. This is implemented using the distributed task queue Celery (Celery, 2017) together with the message broker RabbitMQ (RabbitMQ, 2017). In this setup, the Django web application adds tasks to the Celery queue, which are then asynchronously executed when resources become available. Messages about finished jobs and possible errors are passed back to the web application front-end with asynchronous notifications, implemented with Django-notifications (Django-Notifications, 2017). This setup allows easily scaling the computational resources designated to running simulations and analysis tasks in the background, since additional “worker” machines can be dynamically added to the same task queue as the demand for computational resources grows.

Graphical browser The channel browser front-end was developed by Phyramid (www.phyramid.com) in close collaboration with the author. The browser is implemented using Node.js to provide an asynchronous web server that communicates internally with the API and is connected to user requests through the NGINX web-server as a proxy (see Fig. 2.8B).

2.7.2 Database

The ICG database is a relational MySQL (MySQL, 2017) database that holds both metadata extracted from the literature (see Section 2.6.3 for details), as well as quantitative data extracted from simulations and the results of the analysis of these data quantitative. The database structure is automatically generated from Python (Python, 2017) based definitions of the data structure in the Django web framework (Django, 2017) (see Section 2.7.1 above). Using Django to maintain the database structure has the advantage that migrations can be generated programmatically, tested locally, and applied to the web application server during only short downtimes. Further, designing the database structure in this way makes it independent of the database backend used, facilitating later migrations of the database to backends possibly different from MySQL.

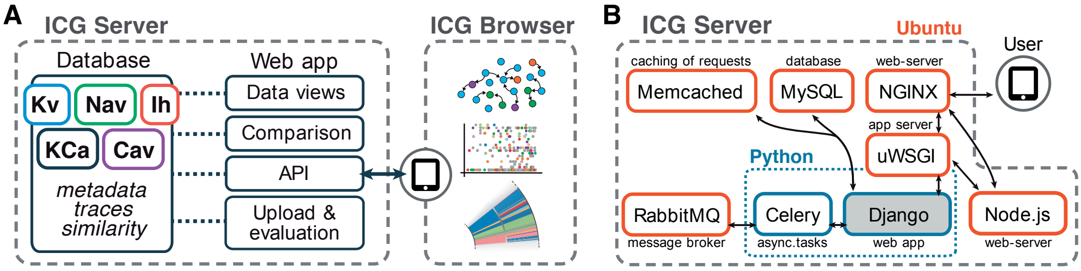

A
ICG Server
Database
Kv
Nav
Ih
KCa
Cav
metadata
traces
similarity
Web app
Data views
Comparison
API
Upload &
evaluation
ICG Browser
B
ICG Server
Ubuntu
User
caching of requests
Memcached
database
MySQL
web-server
NGINX
app server
uWSGI
Python
RabbitMQ
message broker
Celery
async.tasks
Django
web app
Node.js
web-server

Figure 2.8 – **The IonChannelGenealogy web application. A** All data for ion channel models in the five main classes (Kv=voltage gated potassium, Nav=voltage gated sodium, Ih=inward rectifier potassium, KCa=calcium gated Potassium, Cav=voltage gated calcium) are stored in a SQL database. A web application accesses this database, implements data views (including the comparison of channel models), and allows access to the database by a web application interface (API). The graphical browser (ICG browser) application was developed to use the API endpoints to create graphical data visualizations on the device of the website's user (right dashed area). **B** Software stack used to run the ICG web application in the production environment. See main text for details.

See Fig. 2.9 for a diagram of tables in the database, and the relations between them. Edges in the diagram ending in a single black dot indicate directed "one to many" relationships – e.g. each *Channel* can have several *Meta*s related to it to it (while each *Meta* only has a single channel assigned to it). Edges ending in double black dots represent bidirectional associations between two tables – e.g. each *Channel* can have several *Reference*s related to it, while each *Reference* is associable to several *Channel*s. Bidirectional relations are realized by additional associative tables (omitted from the diagram), which contain two columns representing the two *id*s that are associated by an entry in the table. The datatypes listed in each column refer to the Django field type (which is mapped to MySQL built-in datatypes, see https://docs.djangoproject.com/en/1.9/ref/models/fields).

In the following, I shortly elaborate on the role of each table in the database related to the ICG web application. The listing below and Fig. 2.9 (except for "User") contain only tables that are exclusively used for ICG. Several Django components and add-ons used in the application also require database storage, which brings the total count of tables currently used in the production environment to 69. Some of the ICG related tables have been omitted from Fig. 2.9 for brevity and are marked by an asterisk (*) in the listing below.

Channel Contains one entry per channel model in the database. Channels are assigned a single *id_modeldb*, which links our database to the corresponding channel models entries on ModelDB (Hines and Carnevale, 2004). Other relations to several tables exist, see below.

ChannelRelation Introduces annotated directed self-associations for Channel entries. *Parent* Channels are linked to *child* Channels, while specifying a label and a type for these relations (see below). The ancestor-descendant relationship (see Section 2.6.4) of

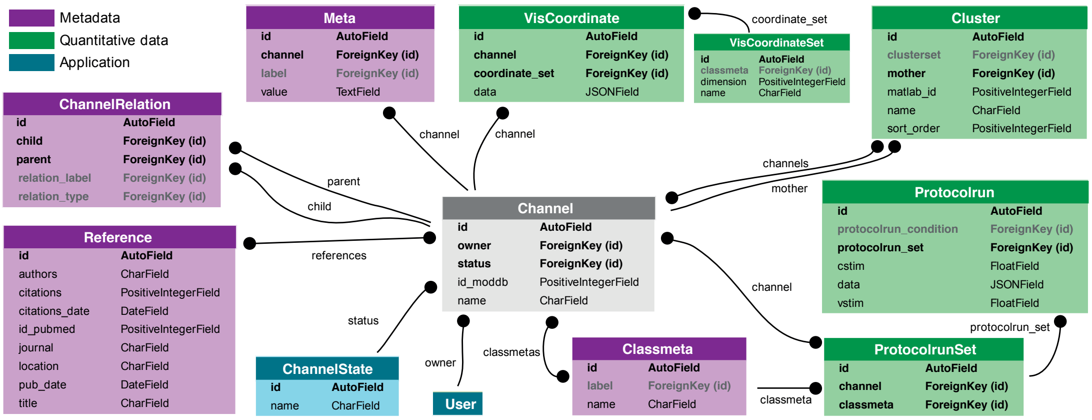

Metadata
Quantitative data
Application
Meta
id AutoField
channel ForeignKey (id)
label ForeignKey (id)
value TextField
VisCoordinate
id AutoField
channel ForeignKey (id)
coordinate_set ForeignKey (id)
data JSONField
coordinate_set
VisCoordinateSet
id AutoField
classmeta ForeignKey (id)
dimension PositiveIntegerField
name CharField
Cluster
id AutoField
clusterset ForeignKey (id)
mother ForeignKey (id)
matlab_id PositiveIntegerField
name CharField
sort_order PositiveIntegerField
ChannelRelation
id AutoField
child ForeignKey (id)
parent ForeignKey (id)
relation_label ForeignKey (id)
relation_type ForeignKey (id)
channel
channel
parent
child
channels
mother
Channel
id AutoField
owner ForeignKey (id)
status ForeignKey (id)
id_moddb PositiveIntegerField
name CharField
Protocolrun
id AutoField
protocolrun_condition ForeignKey (id)
protocolrun_set ForeignKey (id)
cstim FloatField
data JSONField
vstim FloatField
Reference
id AutoField
authors CharField
citations PositiveIntegerField
citations_date DateField
id_pubmed PositiveIntegerField
journal CharField
location CharField
pub_date DateField
title CharField
references
status
channel
protocolrun_set
owner
classmetas
ChannelState
id AutoField
name CharField
User
Classmeta
id AutoField
label ForeignKey (id)
name CharField
classmeta
ProtocolrunSet
id AutoField
channel ForeignKey (id)
classmeta ForeignKey (id)

Figure 2.9 – **SQL database diagram and relational structure.** Each colored square represents a table in the database, with the table name displayed in bold large letters. Tables are colored by the general class of data they store, partitioned in Metadata, Quantitative data, and Application (data). Edges between table represent relations between tables, see main text for a detailed description. The columns of each table are listed as rows below the header, with their datatype displayed next to them. Columns in gray text link to tables that have been omitted from the diagram for brevity.

Channels (which we extracted from publications) is implemented by this table. This table contains many-to-one relations to ChannelRelationLabel and ChannelRelationType, which have been omitted from the diagram.

ChannelRelationLabel* Allows specifying additional unique types for a ChannelRelation. Examples are “Identical copy” or “Slightly modified” for the ancestor-descendant relationship.

ChannelRelationType* Allows specifying several disjoint sets of channel relations. While currently only the “Ancestry” type is used (for the “ancestor-descendant” relationship), I chose to normalize the ChannelRelation functionality to a large degree, so as to be able to easily add additional sets of relations later on.

ChannelState Defines several possible states that each Channel can have in the web application (including “Public”, “Draft”, “Deleted”). This table is used as a coarse filter for display of channels on the front-end and has been introduced in preparation for user submission. It further allows the ICG editors to prepare new entries to the database without displaying them on the front-end.

ClassMeta ClassMetas are metadata entries by which Channels can be sorted into classes. These are stored in a normalized fashion to allow the storage of several categories of String based metadata: each ClassMeta is assigned to a category designated by the ClassMetaLabel. For example, metadata for channel subtypes are stored as separate ClassMeta entries (with names e.g. “A”, “dr”, “HH”) that all have the same ClassMetaLabel “Subtype”.

ClassMetaLabel* Groups ClassMetas together into categories. Examples are “Class” (which contains ClassMetas “Kv”, “Nav”, “Cav”, etc.), “Subtype” or “Animal model”.

Cluster Store the clustering result presented in the main text (a partition of all channels into disjoint groups). Each separate cluster is one entry in this table, with an association from it to all Channels that are contained in this cluster. Additionally, a sort order is stored, that is used to store the sorting of clusters according to similarity, as obtained from the hierarchical clustering. A second unidirectional association to Channels is included, that designates the “mother” Channel of each cluster (named “representative model” in Chapter 2). Clusters are grouped into ClusterSets by a relation which was omitted from the diagram.

ClusterSet* Groups together several Clusters into clusterings of a full set of Channels. E.g. all Clusters obtained from clustering Kv channels are contained linked to the same ClusterSet (“Kv clustering”). This normalization has been chosen to allow the storage of several different clusterings of the same set of Channels (e.g. as “Kv clustering 1”, “Kv clustering 2”).

ICGSubmission Represents user submissions made through the web application. This table mainly stores links to uploaded *.mod* files, time series, and generated figures. See also below on how this is used by the web application.

Meta Metadata entries that are different for each Channel. These are stored in a normalized fashion to allow the storage of several String based metadata: each Meta is assigned to a category designated by the MetaLabel. For example, each comment added to a Channels is stored in a separate Meta entry, that is linked to the MetaLabel “Comments”. The difference to the ClassMeta table is in the relation of Channels to Metadata – each Meta has only a single Channel assigned to it (e.g. each “Comment” is entered for a single Channel), while ClassMetas can be assigned to several Channels (several Channels can be assigned to the “A” subtype, which is a ClassMeta).

MetaLabel* Groups Metas together into categories. Examples are “Comments” or “Runtime”.

Protocolrun Stores the current time series obtained from voltage-clamp simulations (see Section 2.6.5) as JSON encoded arrays in the *data* column. For each time series, we store the index of the applied voltage protocol (e.g. the Kv “Activation” protocol consists of 16 voltage steps) as *vstim*. Similarly, we store the index of the applied calcium protocol – only KCa channels are simulated under 7 calcium concentrations. Protocolruns are associated to Channels indirectly, by the relation to ProtocolrunSet.

ProtocolrunCondition* Groups Protocolruns into similar categories, which are referred to as “conditions” – these represent the 5 different protocols currently used to measure channel model kinetics (see Section 2.6.5), e.g. “Activation”.

ProtocolrunSet Groups together ProtocolrunSets and associates them to a Channel as well as a ClassMeta. The link to ClassMeta is introduced to be able to separate

ProtocolrunsSets for the channels that are contained in several ClassMetas: for example, a channel can appear both in the "Kv" and "Nav" ClassMetas with different internal conductances activated, which will result in different Protocolrun data depending on active ClassMeta.

Reference Stores data for publications that are associated to ModelDB entries.

User This is the table storing user data from the *django.contrib.auth.User* model. This is used to associate channels to the user accounts of logged in editors.

VisCoordinate Stores data for visualization on the graphical browser, which are usually the scores calculated for channels (Section 2.6.7). Coordinates are stored as JSON encoded arrays in the *data* column.

VisCoordinateSet Groups together VisCoordinates and associates them to a ClassMeta. Similar to the ProtocolrunSets, the link to ClassMeta here is introduced to be able to separate VisCoordinateSets for channels that are contained in several ClassMetas.

2.7.3 Conclusion

In summary, above we have sketched current architecture and structure of the ICG web application and the database at the time of writing. It should be noted that, as a software project, the ICG web application and database are subject to ongoing improvements and changes.

The web application is formed by a set of interlinked open-source software components built around a Django-based core. The architecture was devised in this way to ensure scalability with the number of requests. By using the NGINX web-server as the public entry point to the web application, it can act as a load-balancer that relays requests to more than one web server serving the ICG application. Similarly, the Memcached caching backend can be shared between several servers. As mentioned above, for asynchronous simulation and analysis, additional Celery instances can be added to the pool of available computational resources, should the need arise.

Scalability with respect to the storage and efficient retrieval of quantitative data (time series) however, should be improved in the future. Currently, quantitative data are stored as JSON encoded strings in the SQL database at a fixed subsampling rate of the original time series. Already now, comparing and displaying 100s of channels (in the "Compare channels" section of the web application) leads to combined latencies for database queries and transmission of time series to the user on the order of seconds. For ICG to be able to act as a database not only for simulated data but also experimental data, this minimum sampling rate should be lowered, or better, arbitrary sampling rates should be allowed that are as low as the data provided by the contributor. Databases specialized in storage time-series are abundant (see e.g. (Merdanović, 2017)), and externalizing the storage of quantitative data to a carefully chosen candidate should be performed as the ICG project grows. For the display of data on the front-end, several solutions already exist that combine time series databases with efficient graphing on the front-end (e.g. (Grafana, 2017; OpenTSDB, 2017)).

The SQL database architecture was normalized to a large degree, by storing most metadata in the *Meta* and *ClassMeta* structures. This prepares the database for the inclusion of additional metadata categories as the need arises. In particular, changes already on the horizon include preparing the database for the inclusion of models (and experimental data) linked to sources (e.g. (Ranjan et al., 2011; Gleeson et al., 2013; NeuroML, 2017)) other than ModelDB (Hines and Carnevale, 2004). Further, direct links from models to the ICG Github repository containing the original and modified channel files (ICGenealogy, 2016a), as well as direct links to ModelDB *.mod* files (instead of the general entry) will also be included. For a general roadmap of upcoming changes, we refer the reader to the "Development" section on the ICG website (ICGenealogy, 2016b).

3 Continuous attractor networks: Efficient low-dimensional approximation

Alexander Seeholzer, **Moritz Deger** and **Wulfram Gerstner** (in preparation)

3.1 Author contributions

I conceived the study, with support by MD. I wrote all simulation code and carried out all experiments and analyses. The optimization approach was conceived by me and MD and subsequently realized by me. I wrote the manuscript, with help from MD and WG.

3.2 Abstract

Continuous "bump" attractors are an established model of cortical working memory for continuous variables and can be implemented using various neuron and network models. Here, we develop a generalizable approach for the approximation of bump states of continuous attractor networks implemented in networks of both rate-based and spiking neurons. The method relies on a low-dimensional parametrization of the spatial shape of firing rates, allowing to apply efficient numerical optimization methods. Using our theory, we can establish a mapping between network structure and attractor properties that allows the prediction of the effects of network parameters on the steady state firing rate profile and the existence of bumps, and vice-versa, to fine-tune a network to produce bumps of a given shape.

3.3 Introduction

Behaving animals commonly need to transiently memorize information about the environment. For example, as an animal overlooks the visual scenery, locations of certain salient stimuli need to be recorded and stored. Such information does not need to be stored in long-term memories. Rather, working memory must provide a quickly accessible computational substrate for storing information over short durations. While long-term memory is thought to be stored in the efficacy of synaptic connections in the brain (Hebb, 1949; Martin et al., 2000; Takeuchi et al., 2014), a possible substrate for working memory may be transiently stable states of neuronal activity across cortical networks (Goldman-Rakic, 1995; Curtis and D'Esposito, 2003; Chaudhuri and Fiete, 2016).

As model implementations of this concept, localized spatial profiles of neural activity have been proposed for the internal representation of sensory stimuli (Wilson and Cowan, 1973; Amari, 1977; Ben-Yishai et al., 1995). First, neurons are associated to the presence of physical quantities through elevated responses during and after the presentation of stimuli, akin to receptive fields. For example, the presentation of stimuli at varying angular positions in the visual field evokes persistent and elevated firing rates in selective groups of neurons of the prefrontal cortex during delay periods (Funahashi et al., 1989). Choosing recurrent connection weights (or connection probabilities) which are stronger between neurons that are responsive to similar stimuli, together with feedback inhibition limiting the total firing rates in a network, allows this class of models to display *bumps* of self-sustained activity: neuronal activity that is localized in the space of possible stimuli. Since these states are stable attractive states, and all possible such states form a continuum, these models are often referred to as *continuous attractors*. The elevated firing of neurons responsive to similar stimuli is then seen as the working memory representation of physical quanta stored in the network, e.g. spatial orientations (Zhang, 1996), or angular positions in the visual field (Compte et al., 2000). Similar computational circuits might also serve as the basis of persistent internal representations in hippocampal areas (Itskov et al., 2011a; Yoon et al., 2013).

Continuous attractor models with simplified shapes of connectivity or neuronal input-output relations can be analyzed and often exactly solved (Wilson and Cowan, 1973; Amari, 1977; Ben-Yishai et al., 1995; Bressloff and Cowan, 2003; Fung et al., 2010; Itskov et al., 2011b; Laing, 2014), or may generally be approximated in the linear input-output regime of balanced networks (Rosenbaum and Doiron, 2014). However, the inclusion of biologically plausible nonlinearities, like nonlinear neuronal input-output relations (Compte et al., 2000; Renart et al., 2003), neuronal adaptation (Brette and Gerstner, 2005; Roach et al., 2015), or synaptic nonlinearities like short-term plasticity (Zucker and Regehr, 2002; Itskov et al., 2011b) and saturating NMDA kinetics (Destexhe et al., 1994; Compte et al., 2000), complicate the mathematical solution of these systems considerably and make a derivation of the stable firing rate profile unfeasible. Therefore, such systems are usually studied by explicit simulations of the underlying dynamics or by numerical optimization of approximated equations for all neurons (Rosenbaum and Doiron, 2014; Spiridon and Gerstner, 2001). While these procedures in principle allow the prediction of the network activity as a function of the parameters, they involve computationally demanding numerical optimization of high-dimensional systems of equations, possibly as costly as simulating the full neuronal dynamics. Thus, currently, relating the microscopic network parameters to the resulting emergent bump states involves repeated and possibly time consuming simulations. For example, this makes the matching of the network steady states to physiologically constrained features tedious.

Here, we present a generalizable approach for the approximation of the network-wide steady states of continuous attractors. Our approach allows the prediction of the shape of steady-state firing rate profiles, under nonlinear neuronal dynamics and varying configurations of the underlying microscopic system, without having to solve the dynamics of the full, high-dimensional system. Our novel method relies on a low-dimensional parametrization of the network's firing rate profile, which allows us to derive computationally tractable systems of self-consistent equations describing the attractor steady-state, akin to mean-field approaches for networks with discrete attractors (Brunel and Wang, 2001b). These equations can be used to efficiently predict the dependence of the firing rate profile on microscopic network parameters. Importantly, because the dimension of the parameterization of the spatial activity profile is low, our approach makes optimization of the microscopic network parameters for the appearance of desired bump profiles feasible. We apply our method to both networks of simplified rate neurons, and networks of complex, conductance-based integrate-and-fire neurons with saturating and voltage-dependent nonlinear NMDA transmission.

3.4 Results

Mean-field approaches (see e.g. Amit and Brunel (1997); Brunel and Wang (2001b)) that predict the steady states of recurrently connected neuronal networks usually rely on dimensionality reduction. The number of equations describing the dynamics is reduced by partitioning neurons into groups of "similar" neurons, and deriving expressions which describe the average statistics for these coupled groups in the steady states. For example, the simplest such partition consists in considering the mean firing rates of excitatory and inhibitory neurons separately, e.g. all excitatory neurons fire with similar mean rates given the same input:

$$\nu_E = F_E\left(\mathrm{input}_E\right).$$

If the groups of neurons are now homogeneously coupled, i.e. the connections between neurons depend only on the groups of the neurons involved, one can derive the input to neurons of each group in dependence of the firing rates of the groups only. This leads to a closed system of self-consistency equations describing the coupled steady-state firing rates:

$$\begin{aligned}\nu_E &= F_E\left(\mathrm{input}_E\left[\nu_E, \nu_I\right]\right),\\ \nu_I &= F_I\left(\mathrm{input}_I\left[\nu_E, \nu_I\right]\right).\end{aligned}$$

In the steady states of continuous attractor models (see Fig. 3.1), neurons fire at different rates, making a clear partition into discrete groups of similarly firing neurons difficult. Therefore, the solution of such systems usually relies on the explicit simulation of the neuronal dynamics of all neurons along the spatial dimension, or a numerical solution of the coupled

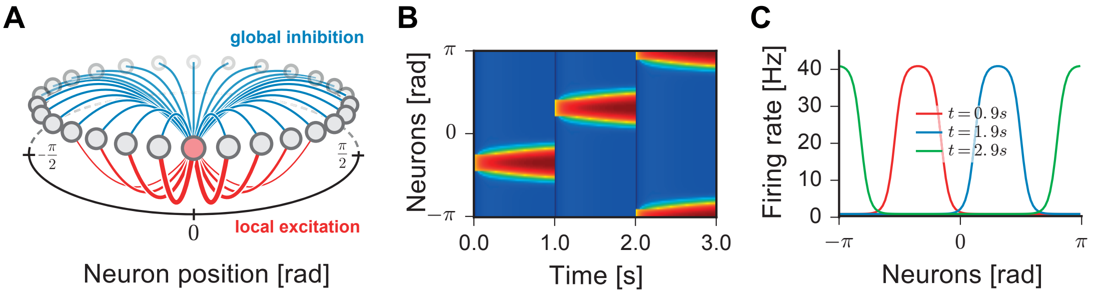


Figure 3.1 – **Stable firing rate profiles in continuous attractors.** **A:** Neurons (circles) are assigned a position on a ring and connected with distance-dependent connection weights to all other neurons: the firing of a neuron (red circle) inhibits all other neurons (global inhibition, blue lines) and strongly excites neurons close it (local excitation, red lines). **B:** Example simulations of a continuous attractor. At times $t = 0, 1, 2s$ the system is reset to localized activity centered at different positions. With time the activity bump broadens towards a stable state. Colors indicate firing rates as in panel C. **C:** Firing rates of the network shown in panel B, measured close to the stable states at $t = 0.9, 1.9, 2.9s$. Plots are generated using the rate model introduced in Section 3.6.1.

self-consistency equations for all neurons.

Here, by using the continuity of the shape of the attractor states, we demonstrate that continuous attractors are amenable to dimensional reduction, by parametrizing the attractor state by a low-dimensional family of functions. In Section 3.4.2, we check our method on networks of simple rate neurons, for which the method might not yield much improvement over simulations or numerical solutions of the steady states. For the spiking networks considered in Section 3.4.3, we show that our approach speeds up predictions of the steady states considerably and further makes these networks amenable to the optimization of network parameters.

3.4.1 General equations for the approximation of stable states in ring-attractors

As a concrete class of continuous attractors, we consider the ring-attractor model, in which stable bumps of neuronal activity are freely translatable along all positions on a circle. Ring-attractor models can be constructed by placing N neurons (rate-based or spiking) at equally spaced angular positions θ along the ring (Fig. 3.1A) (Wilson and Cowan, 1973; Amari, 1977; Ben-Yishai et al., 1995). We choose the angular space to consist of positions $\theta \in [-\pi, \pi)$, where we identify the ends of the interval: a neuron at position $\theta = \pi - \epsilon$ is the neighbor of a neuron at position $\theta = -\pi$. At short angular distances, recurrent connections are chosen to be strong and excitatory, while neurons further apart in angular space effectively inhibit each other's firing (Fig. 3.1A). Due to the symmetry in connectivity with respect to distance, these networks can form a continuous manifold of stable states for sufficiently strong connections: the network activity in response to external inputs converges to firing rate profiles centered around some angular position. The position can be, for example, controlled by providing an external input to the network centered around any desired position (Fig. 3.1B). The stereotypical shape of the resulting firing rate profiles (Fig. 3.1C) is invariant with respect to translations in angular space.

The continuity of the firing rate profile allows us to parametrize the firing rates in the population by a small number of parameters, and to derive equations from the underlying model that constrain these parameters. Here, inspired by shapes observed in simulations, we choose to parameterize the firing rate profile by a generalized Gaussian function, where we assume without loss of generality that the distribution is centered at $\theta = 0$ (cf. Fig. 3.2A):

$$g(\theta) = g_0 + g_1 \exp\left(- \left[\frac{|\theta|}{g_\sigma} \right]^{g_r} \right). \tag{3.1}$$

Here, g_0 controls the baseline firing rate and $g_0 + g_1$ will be the maximum firing rate of the profile. The parameters g_σ and g_r control the width and steepness of the profile, respectively. If we know the distance-dependent connectivity w between neurons and their input-output relation F, we can predict the expected neuronal firing at any position θ in the population (cf. Fig. 3.2B). Crucially, we want the firing rate profile $g(\theta)$ to be generated by the neuronal dynamics – we thus identify $g(\theta)$ at the point θ with the firing rate $\nu(\theta)$ of a neuron at position θ. Finally, we replace the synaptic input to the neuron at position θ with the contributions from all neurons firing at rates g along the ring. For any given position θ_i along the ring, this yields a self-consistent equation in the function g:

$$\begin{aligned} g(\theta_i) &= F(\text{input}_{\theta_i}[g]) \\ &= F\left(\int_{-\pi}^{\pi} d\varphi\, w(\varphi - \theta_i) g(\varphi) \right), \end{aligned} \tag{3.2}$$

with the corresponding *self-consistency error*

$$\text{Err}_i \equiv g(\theta_i) - F(\text{input}_{\theta_i}[g]). \tag{3.3}$$

In principle, this procedure can yield up to N coupled error functions, one for each of the N neurons. One could then minimize the quadratic error $\sum_i \text{Err}_i^2$ with respect to the parameters $\{g_0, g_1, g_\sigma, g_r\}$ to find an approximate solution of the system. However, since the evaluation of each error function can be costly (e.g. in spiking networks, see Section 3.4.3.2), we propose a low-dimensional approximation to constrain the set of 4 free parameters: we pick only 4 points $\theta_1, \ldots, \theta_4$, at which we evaluate the errors. This assumes that the shapes of firing rate profiles maintained by the network are well approximated by the function $g(\theta)$, which we found to be the case for all networks considered here. This leaves the choice of points θ_i to evaluate. To ensure that errors are evaluated across different firing rates, we set the position of these points to cover a range of function values $h_i = g(\theta_i)$: we choose the top of the distribution $\theta_1 = 0$ with $h_1 = g_1 + g_0$, as well as the lowest point $\theta_4 = \pi$ with $h_4 = g_0$ (circles in Fig. 3.2C). The remaining intermediate points (triangles in Fig. 3.2C) are dynamically

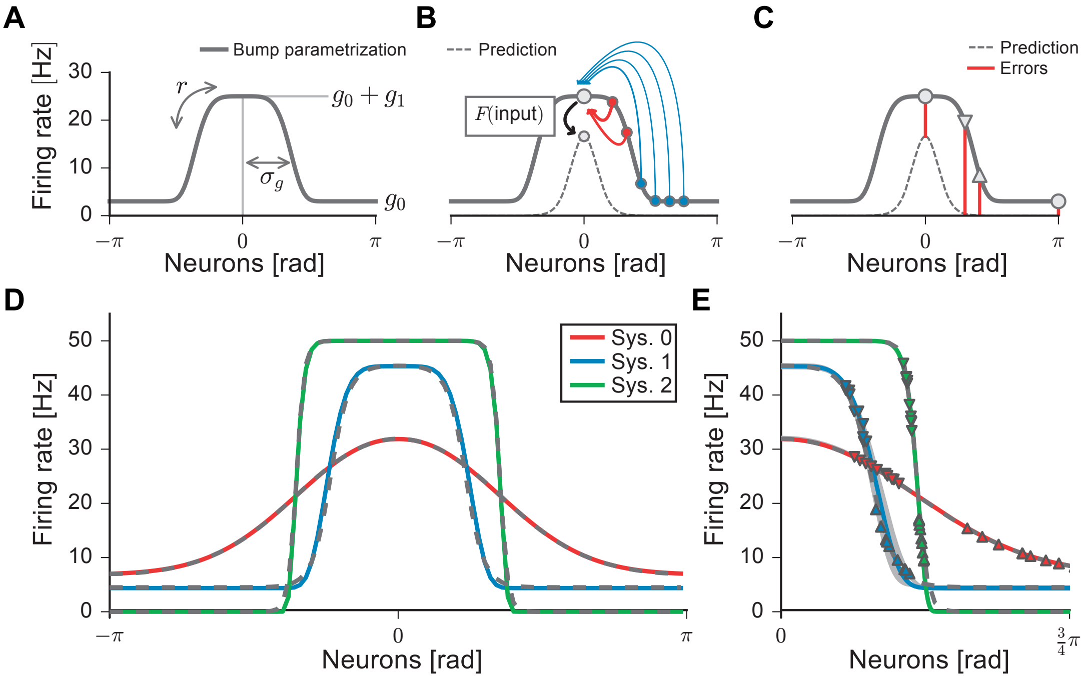


Figure 3.2 – **Approximation of bump shapes in attractor networks. A:** Parametrization of firing rate profiles by 4 shape parameters: g_0 and g_1 control the baseline and maximal firing rates, σ_r controls the width, and r controls the steepness. **B:** Assuming a spatial profile given by the parametrization g, the input to any neuron can be calculated by summing all synaptic inputs: in the example, the neuron at position 0 receives some strong local excitatory input (red thick arrows) and weaker inhibitory input from more distant neurons (blue thin arrows). The firing rate prediction (dashed line) for any neuron can then be calculated as a function of its input. This is illustrated for a neuron at 0 (circle). **C:** Self-consistency errors between the current parametrization (gray line) and the resulting firing rate prediction (dashed line) are calculated at a small number of points along the firing rate profile. Intermediate points (triangles) are positioned dynamically during optimization. **D:** Optimized bump parametrization (dashed lines) for systems with three different connectivities (solutions of the full system are plotted in color). **E:** Dependence of solution on intermediate point placement. Left points are given by downward triangles, rightward points by upward triangles (compare to panel **C**). Single optimization runs (light gray lines) together with median parameters (dashed lines, same as in D) and full solutions (colored lines).

positioned (see Section 3.6.3.1 in Methods): their position depends on the function g such that they always sample intermediate function values h_i.

3.4.2 Approximation of ring-attractor profiles in rate models

The proposed method is, in principle, applicable to any neuron model with a defined input-output relation F. The shape of the stable attractor profiles will, however, depend on the concrete choice of neuron model and the microscopic parameters, in particular the parameters governing the connectivity between neurons. To test the ability of the low-dimensional approximation proposed in the last section to correctly predict the shapes of firing rate profiles, we implemented the ring-attractor model introduced above (see Fig. 3.1) in a network of rate-based neurons with tanh input-output function and a generalized Gaussian recurrent connectivity (see Section 3.6.1 in Methods). For the rate neuron models chosen here, the self-consistency errors Eq. (3.3) are given by (see Section 3.6.3.2 in Methods):

$$\begin{aligned}\mathrm{Err}_i &= g(\theta_i) - \frac{\nu_{\max}}{2}\left[1+\tanh\left(\frac{\tau_s}{s_0}\frac{N}{2\pi}\mathrm{input}_{\theta_i}\left[g\right]\right)\right],\\ &= g(\theta_i) - \frac{\nu_{\max}}{2}\left[1+\tanh\left(\frac{\tau_s}{s_0}\frac{N}{2\pi}\int_{-\pi}^{\pi} d\varphi\, w(\varphi-\theta_i)g(\varphi)\right)\right],\end{aligned} \tag{3.4}$$

where τ_s is the time-constant of synaptic inputs, $\nu_{\max}$ is the maximal firing rate and s_0 is an input scale.

3.4.2.1 Prediction of stable firing rate profiles

To approximate the firing rate profiles that these networks admit as self-consistent solutions, we minimize the error functions Eq. (3.4) with respect to the parameters $\{g_0, g_1, g_\sigma, g_r\}$ as free variables. We find that, for a range of connectivity parameters (see Table C.2), the predicted shapes converge to unique solutions. This solution matches the steady state of the microscopic network simulations accurately (Fig. 3.2D). This is the case both for attractor states that lie in the linear regime of the neuronal input-output relations (Fig. 3.2D, red line) as well as for highly nonlinear attractor dynamics in which neuronal firing reaches saturation values, leading to plateau-shaped firing rate profiles (Fig. 3.2D, blue and green lines).

As discussed above, the placement of intermediate sampling points (Fig. 3.2C, triangles) is not constrained by theory, but remains a free parameter of our approach. We chose these points at positions θ_i such that they sample given function values $h_i = g(\theta_i)$. To investigate the dependence of the prediction on the placement of intermediate sampling points, we calculated several predictions while randomly varying the choices of h_i (Fig. 3.2E, triangles). We find that this hardly affects the converged solutions.

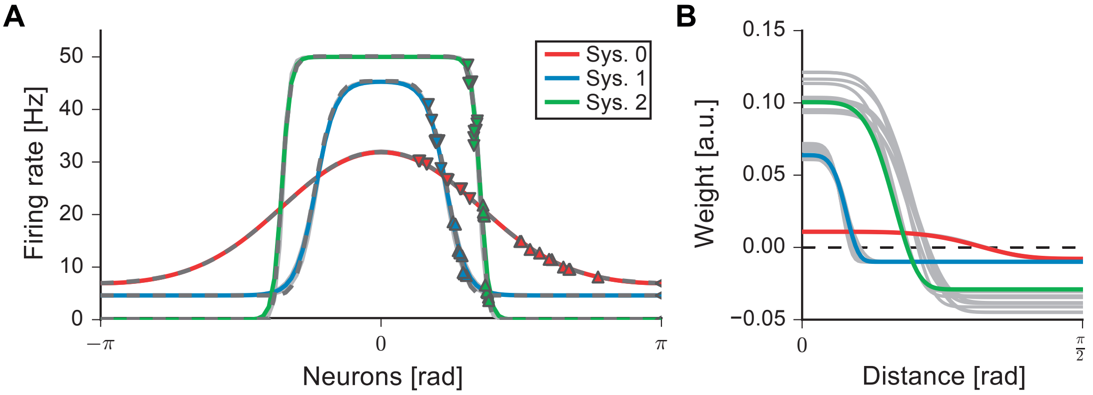


Figure 3.3 – **Application: Optimization of network connectivity. A:** Stable network firing rate profiles resulting from 10 optimization runs of the network local connectivity profiles. The optimization target profiles are plotted in colors, together with single optimization runs (light gray lines, almost overlapping with colored lines) and one example highlighted optimization result (dashed lines) each. Left optimization points are given by downward triangles, rightward points by upward triangles (compare to Fig. 3.2C,E). **B:** Results of the 10 optimized connectivity profiles (light gray lines) together with one example profile (colored lines). Solutions for System 2 are degenerate, while others are fairly unique.

3.4.2.2 Optimization of network parameters

While our low-dimensional system of self-consistent equations can be used for the prediction of the steady-state firing rates, they can also be used in the inverse way, to optimize any of the network parameters. We demonstrate this here, using the shape of the recurrent connectivity as an example. However, such optimizations can include further parameters of the network model (see Section 3.4.3.3).

To optimize the network connectivity parameters, we keep the parametrization parameters $\{g_0, g_1, g_\sigma, g_r\}$ fixed to the desired values of the shape of the firing rate profile. We then optimize the self-consistent equations Eq. (3.2) for values of the recurrent connectivity parameters $\{w_0, w_1, w_\sigma, w_r\}$ which lead to solutions of the equations and produce the desired bump profile. In Fig. 3.3 we show the results of this procedure for the three systems also investigated in Fig. 3.2. The procedure yields network connectivities that fulfill the desired properties (Fig. 3.3A), largely independently of the points θ_i chosen for the evaluation of the errors. Importantly, for some shapes the solutions show a degeneracy (Fig. 3.3B, gray lines), in the sense that several connectivity parameter sets are found that produce the same stable firing rate profile.

3.4.3 Approximation of ring-attractor profiles in spiking networks

In complex spiking neuron models, the steady-state input-output relations often involve integral functions (Brunel and Hakim, 1999; Brunel and Wang, 2001b) that are not amenable

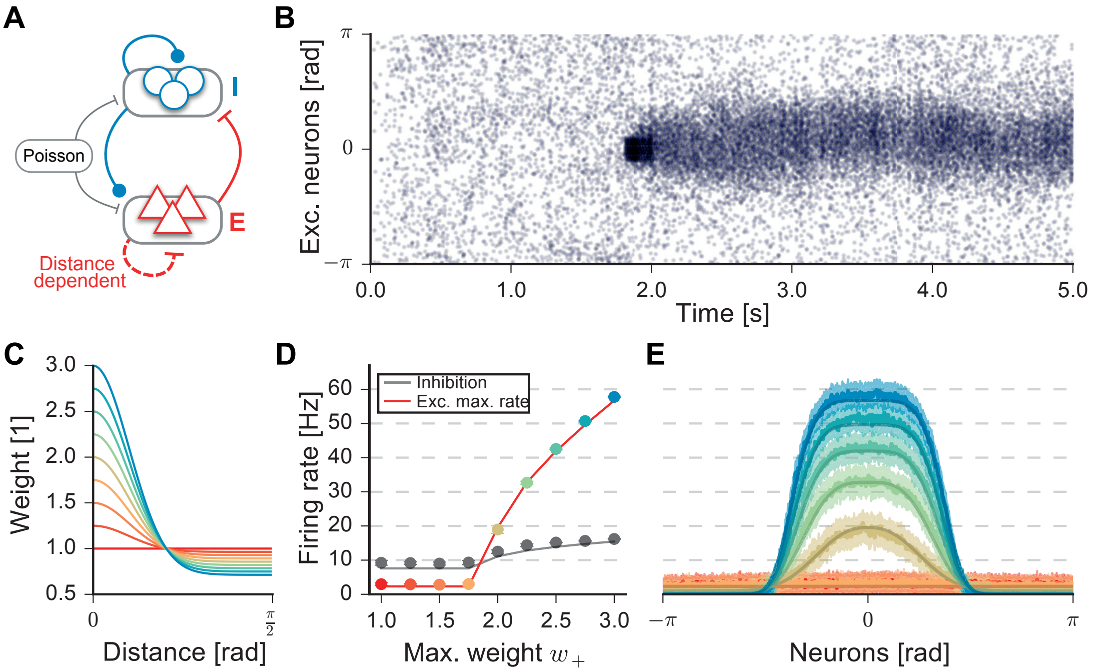


Figure 3.4 – **Application: Prediction of firing rate profiles in spiking continuous attractor networks**. **A:** Recurrently connected spiking network of a population of excitatory (E, red triangles) and inhibitory (I, blue circles) neurons. Networks are fully connected with uniform weights, except for E-E connections (dashed red line), which are distance-dependent. All neurons receive excitatory input with spikes generated by homogeneous Poisson processes. **B:** Example simulation: E neurons fire asynchronously and irregularly until an external cue is given (centered at 0 for 0.2s starting at $t = 1.8s$). After this stimulus, a bump of elevated activity sustains itself around the point of stimulation ($w_+ = 2.0$). **C:** Distance-dependent E-E connectivity as a function of the parameter w_+ (maximal strength or recurrent connections). Values range between $w_+ = 1$ (red) and $w_+ = 3$ (blue). **D:** Appearance of the bump firing rate profile as a function of the connectivity parameter w_+. Theoretical predictions (lines) and simulation results (maximum of fit of g to firing rates recorded over 1.5s of delay activity, mean over 5 repetitions, errors show 95% CI) for maximal firing rates of E (colored) and I (gray) neurons. Colors similar to those in panel C. **E:** Shape of the optimized firing rates profiles (solid colored lines) compared to mean firing rates measured from spiking simulations (thin colored lines) (mean ±0.5STD of firing rates recorded from 1.5s of delay activity in 5 repetitions) for all values of w_+ in panel C and D (similar colors).

to further theoretical analysis. Further, the introduction of nonlinear NMDA transmission at excitatory synapses (Wang, 1999; Compte et al., 2000) complicates the analysis of such models considerably: voltage-dependent gating of the maximal NMDA conductance (by the voltage-dependence of the Mg^{2+} block) and saturation of the NMDA at high conductances necessitates the numerical solution of a 4-dimensional system of self-consistent equations for a relatively simple 2 population mean-field model (Brunel and Wang, 2001b) (see below).

Here, we demonstrate that firing rate profiles in a continuous attractor network of spiking neurons with such nonlinear NMDA transmission are amenable to the same approach as described above, which still involves evaluating comparatively few equations. The spiking network we implement (similar to Compte et al. (2000) and used with variations in e.g. Renart et al. (2003); Murray et al. (2012); Wei et al. (2012); Pereira and Wang (2015); Almeida et al. (2015), see Section 3.6.2 in Methods) consists of two fully connected populations of conductance-based integrate-and-fire neurons: a population of inhibitory neurons with unstructured all-to-all connectivity, and a population of excitatory neurons, with distance-dependent recurrent excitatory connections (Fig. 3.4A). In addition, all neurons receive excitatory background input mediated by spikes generated by Poisson processes. These network can be tuned such that they possess a bi-stability (Fig. 3.4B): a uniform state with spontaneous spiking activity in the excitatory population (the inhibitory population is always uniformly spiking) coexists with an “evoked” spatially inhomogeneous bump-state that appears after an external cue input is given to a subgroup of excitatory neurons (stimulus is present at $t = 1.8 - 2.0s$ in Fig. 3.4B).

3.4.3.1 Self-consistent equations for networks of spiking neurons

For the excitatory population, we again parametrize the spatial profile of firing rates by Eq. (3.1), which allows us to derive self-consistent equations for any neuron in the excitatory population. We construct self-consistent equations for the excitatory firing rates at positions θ_i as in Eq. (3.2). However, these now will depend additionally on the inhibitory firing rate ν_I. Also, the voltage-dependence of the differential equations leads to an additional self-consistent equation for the mean-voltage $\bar{V}$. For any position θ, the excitatory self-consistent equations are of the form (see Section 3.6.3.3 in Methods for detailed expressions):

$$\begin{aligned} g(\theta) &= F(\mathrm{input}_\theta\left[g\right], \nu_I, \bar{V}(\theta)) \\ &\equiv F\left(\int_{-\pi}^{\pi} d\varphi\, w(\varphi-\theta)\psi\left(g(\varphi)\right), \nu_I, \bar{V}(\theta)\right), \end{aligned} \tag{3.5}$$

$$\bar{V}(\theta) = G\left(\mathrm{input}_\theta\left[g\right], \nu_I, \bar{V}(\theta)\right). \tag{3.6}$$

The function $\psi(g)$ expresses the mean synaptic activation under presynaptic Poisson spiking at rate g. For accuracy, we chose to measure ψ numerically for the model of nonlinear NMDA conductance of the recurrent excitatory synapses given in the network (see 3.6.3.3 in

Methods).

To constrain the free parameters of $g(\theta)$, we again pick 4 points $\theta_i \in \{\theta_1, \dots, \theta_4\}$, each now yielding 2-dimensional error functions

$$Err_i = \begin{pmatrix} g(\theta_i) - F\left(\text{input}_{\theta_i}\left[g\right], \nu_I, \bar{V}(\theta_i)\right) \\ \bar{V}(\theta_i) - G\left(\text{input}_{\theta_i}\left[g\right], \nu_I, \bar{V}(\theta_i)\right) \end{pmatrix}. \tag{3.7}$$

The resulting 8 equations are optimized for the 4 parameters of the parametrization g, as well as the additional 4 variables $\bar{V}(\theta_i)$. The inhibitory population, on the other hand, is assumed to be homogeneous. Its activity can be described by a single mean firing rate ν_I and the average voltage in the inhibitory population $\bar{V}_I$, resulting in a pair of additional self-consistency errors that constrain these two variables:

$$Err_I = \begin{pmatrix} \nu_I - F(\text{input}_I, \nu_I, \bar{V}_I) \\ \bar{V}_I - G(\text{input}_I, \nu_I, \bar{V}_I) \end{pmatrix}, \tag{3.8}$$

where $\text{input}_I = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\varphi \psi\left(g(\varphi)\right)$ is the mean recurrent excitatory input to inhibitory neurons.

As mentioned above, in a certain range of parameters the spiking system possesses two dynamically stable states (Fig. 3.4B): the uniform state and the “evoked” bump state. In this bistable regime, the associated self-consistency Equations (3.5)-(3.6) must have an an additional unstable solution (Strogatz, 2000). Even for parameters in which the bump-state is the only stable state of the system, the uniform state will still be a (unstable) solution of the self-consistent equations. Accordingly, numerical solutions of the errors Eqs. (3.7)-(3.8) sometimes converge to the uniform state or an unstable intermediate solution, even if a stable bump state at higher firing rates exists. In the following we consider only the solutions with the highest spatial modulation found under repeated solutions (see Discussion).

3.4.3.2 Prediction of firing rate profiles from network properties

Above, we derived error functions constraining the parametrization of firing rate profiles for spiking networks Eq. (3.7)-(3.8). Here, we use these to predict the dependence of the spatial shape of the firing rate profile on of a bifurcation parameter w_+, which is the maximal strength of recurrent excitatory connections. At $w_+ = 1$ the connection profile is homogeneous, while at larger values local connections are stronger (Fig. 3.4C). The strength of long range connections is calculated by a normalization condition (see Section 3.6.2.3 in Methods).

As w_+ is increased (Fig. 3.4C) above a critical value, a spatially inhomogeneous bump state appears in simulations of the spiking network (Fig. 3.4E). Our theory predicts this dependence of the network state on the bifurcation parameter, while approximating to a large degree the

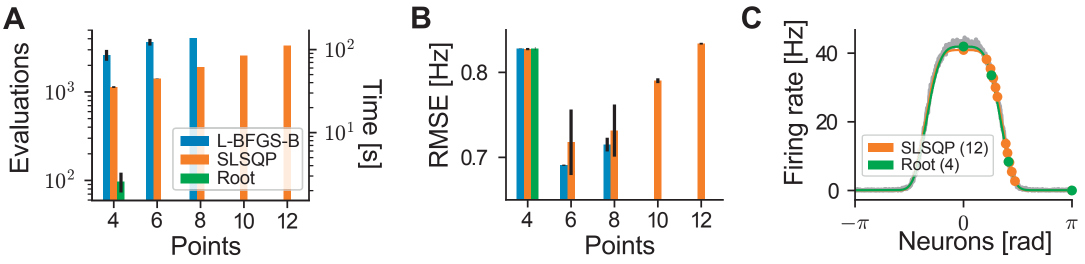


Figure 3.5 – **Efficient optimization of self-consistency errors**. **A:** Evaluations of self-consistency errors until convergence for three different optimization methods ($w_+ = 2.5$) (10 repetitions, errors show 95% CI). *L-BFGS-B* and *SLSQP* both minimize the sum of squared errors, *Root* minimizes the error vector directly. Right-hand axis shows wall-clock times of optimization procedures. **B:** Averaged root mean square error (RMSE) between optimized firing rate profiles and mean firing rates measured from simulations ($w_+ = 2.5$, data from Fig. 3.4C,D,E). RMSE was calculated for all (800) neurons ($n = 10$, as in panel A). **C:** Optimized firing rate profiles, together with placement of sampling points for two combinations of optimizers and points (Root on 4 points, SLSQP on 12 points). Gray line shows mean firing rate measured from simulations ($w_+ = 2.5$, as in Fig. 3.4C,D,E).

changing shape of the rate profile as the parameter is increased (Fig. 3.4D,E). The firing rates of the inhibitory population and their increase with the parameter w_+ are also well described (Fig. 3.4D, black dots and lines).

As mentioned above, the error functions Eq. (3.7) could, in principle, be evaluated at an arbitrary number of points. To constrain the 4 parameters of the parametrization $g(\theta)$, we chose only the minimal number of points. This reduces the necessary number of evaluations of the errors Err_i (Fig. 3.5A). Further, since in this case the dimensions of the optimization variables $\{g_0, g_1, g_\sigma, g_r\}$ and the error vector coincide, application of a more efficient numerical optimization method (*Root*, see Section 3.6.3.5) allows for faster optimization (Fig. 3.5A, green bar), which reduces the needed time from $\sim 30s$ to close to $2s$ (Fig. 3.5A, right hand axis). We observe, however, that adding additional points does slightly influence the resulting prediction (Fig. 3.5B), where more sampling points placed in the flanks of the bump tend to reduce slightly the predicted maximal firing rates (Fig. 3.5C, orange points).

In a second experiment, we show that the theory can be used to efficiently predict the effect that changing network parameters have on the shape of the resulting firing rate profile. Similar to the simulated experiment in Murray et al. (2012, Fig. 3), we systematically reduced the strength of recurrent excitatory-to-excitatory (g_{EE}) and inhibitory-to-excitatory (g_{EI}) conductances of the network from the baseline of the network presented in Fig. 3.4. Such changes of the ratio of excitation to inhibition have been hypothesized to occur under cortical disinhibition observed in schizophrenia (Marin, 2012; Murray et al., 2012). Recovering the result presented in the study, we see that the width of the bump profile[1] depends mostly on

---

[1] Note, that the network presented here has a wider profile at baseline than the one investigated in Murray et al. (2012).

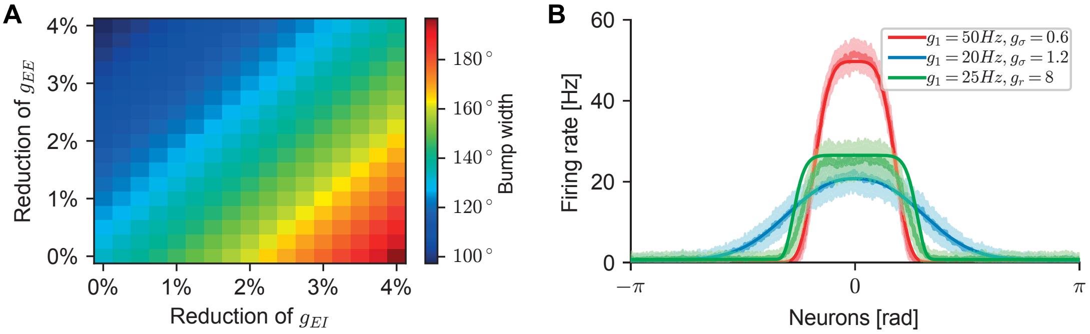


Figure 3.6 – **Applications: Shape prediction over parameter ranges / Optimization of network parameters**. All panels use the spiking network model of Fig. 3.4. **A** Theoretically predicted width of the firing rate profiles (full width, as predicted by the parameter $2 \cdot g_\sigma$) under varying strengths of recurrent excitatory-to-excitatory (g_{EE}) and inhibitory-to-excitatory (g_{EI}) synaptic connections. Compare to the similar experiment performed in spiking network simulations in Murray et al. (2012, Fig. 3). The bifurcation parameter was kept fixed at $w_+ = 2.5$. **B** Network parameters optimized for 3 different shapes of the stable firing rate profile (see Table C.3 for details). Solid dark lines are theoretical predictions using the same method as in Section 3.4.3.2. Lighter lines and shaded areas are mean ±0.5STD of firing rates recorded over 3.5s of delay activity in 5 repetitions.

the ratio of recurrent conductances, and thus undergoes significant widening under disinhibition. As we have shown in Fig. 3.5A, the optimization procedure for each datapoint is fast and thus enables these type of parameter scans for wide ranges of values under many parameters.

These results show that our approach can be used to accurately describe the firing rate profiles of bump-attractor networks of recurrently connected spiking excitatory and inhibitory neurons, across a range of parameters. While evaluating the error function at more points can lead to slightly increased accuracy of the prediction, the impact on optimization performance is significant, increasing the needed time by an order of magnitude.

3.4.3.3 Optimization of network parameters

As demonstrated above for a rate-based network, our low-dimensional approximation of continuous attractors allows the optimization of network parameters. Here, we demonstrate that this approach extends well to continuous attractors implemented in recurrently connected spiking neural networks with nonlinear synaptic transmission. As in the case of the rate network, this is achieved by fixing some desired properties of the firing rate profile, while minimizing the error functions Eqs. (3.7)-(3.8) with respect to several network parameters. Here, we included the shape parameters w_σ and w_+ of the distance-dependent connectivity, as well as the strengths of all recurrent synaptic connections: g_{EE}, g_{IE}, g_{EI}, g_{II} (cf. Table C.3 for details).

To find networks that admit a given shape of the firing rate profile, we first fixed the firing rate modulation g_1 and width g_σ, while optimizing the connectivity parameters w_+, w_σ as well as the strength of all recurrent excitatory and inhibitory transmission. In total, there were 17 variables (see Table C.3 for listings and optimization results) which were optimized (see Section 3.6.3.5 for details). Varying these free parameters allows us to optimize the remaining parameters of the spiking networks through a range of shapes of the stable firing rate profile, from rather thin bumps of high activity (Fig. 3.6, red) to wide bumps of low activity (Fig. 3.6, blue). To check whether we could optimize the spiking network to show saturated flat-top shapes at low firing rates, we fixed $g_1 = 25Hz$ and the sharpness parameter to a high value $g_r = 8$, while optimizing the parameter g_σ (Fig. 3.6, green).

Similar optimization results for all three bump shapes were achieved by imposing additionally a low basal excitatory firing rate ν_E^{basal}, which constrains firing rates in the uniform (spatially homogeneous) state for $w_+ = 1$. For the optimizations above, the basal rates were unconstrained and varied between $1Hz$ and $5Hz$. Thereby, the number of optimization variables is increased to 20, since additionally a basal inhibitory firing rate and basal inhibitory and excitatory mean voltages need to be introduced to calculate the self-consistency error of the basal firing rate ν_E^{basal}. It should be noted, that the value of this constraint affects the possible bump shapes: for example, setting $\nu_E^{\text{basal}} = 1Hz$ did not yield a converging optimization for the blue and green curves of Fig. 3.6 – this could be alleviated by relaxing to the higher value $\nu_E^{\text{basal}} = 3Hz$.

3.5 Discussion

We have presented a framework for the approximation of self-sustained firing-rate profiles in continuous attractor neural networks. Analytical computation of the steady states of continuous attractors is often not possible, since it involves the solution of high-dimensional systems of nonlinear equations, making numerical solutions necessary. Moreover, the spatially inhomogeneous firing rate profiles of these networks prohibit dimensional reduction of these equations by separation of neurons into homogeneous populations, as is usually done in mean-field approaches (Amit and Brunel, 1997; Brunel and Wang, 2001b). Here, we propose a simple approach, consisting in approximating the continuous firing rate profiles by a family of functions g with only 4 parameters. These parameters are constrained by equations expressing the microscopic dynamics of the neurons and synapses involved in the model, and can be optimized to find admissible solutions. As we have shown, this can be used for the efficient mapping of the effects that different network parameters have on the bump shape. Next to predicting the emergent steady states of attractors, the utility of the low-dimensional approximation is that the derived self-consistent equations are efficiently optimizable: we were able to use standard numerical methods to constrain the parameters of spiking networks to show desired firing-rate profiles.

In the main text we have formulated our approach as generally as possible, to emphasize that the method does not rely on a specific neuronal (rate or spiking) model, as long as a prediction of firing rates given the synaptic input can be derived. Therefore, the theory could be extended easily to other neuron models, or connectivities where connection probabilities are distance-dependent (e.g. Hansel and Mato (2013)). The approach presented here could also be used to predict the activity of two-dimensional attractor models implemented in "sheets" of neurons, that are often used in the context of hippocampal networks (Tsodyks, 1999; Samsonovich and McNaughton, 1997; Itskov et al., 2011a). Assuming isotropy of the connectivity (if connection strengths depend only on the Euclidean distance between neurons), a two-dimensional generalized Gaussian function with g_0, g_1, g_r and a single parameter g_σ as before could be used to approximate activity states. For non-isotropic networks, further width parameters could be introduced and constrained by sampling at additional points.

We have shown that our approach is amenable to the inclusion of synaptic nonlinearities like saturating NMDA transmission (which is captured by the synaptic activation function ψ, cf. Eq. (3.5)). Other sources of nonlinear synaptic transmission, for example the activity-dependent short-term plasticity of synapses (Zucker and Regehr, 2002) that is often investigated in the context of working memory models (Mongillo et al., 2008; Itskov et al., 2011b; Hansel and Mato, 2013), can be similarly incorporated: calculation (or numerical estimation) of a compound function ψ that describes the steady-state values of synaptic activation under all nonlinear processes affecting synaptic transmission would suffice to adapt the theory (see Section 4.6.6, where we apply the same method to networks with short-term synaptic plasticity). Similar to the estimation of the mean-voltage in spiking networks demonstrated here, the theory could also be extended to incorporate adaptation effects on the steady-state firing rates of neurons (Brette and Gerstner, 2005).

For the prediction of firing rates in spiking networks we have adapted a theory for the description of mean-field firing rates of conductance based integrate-and-fire neurons (Brunel and Wang, 2001b) to predict spatial firing rate profiles in recurrently connected networks (Compte et al., 2000). Mean firing rates and mean voltages in this theory are generally expectation values over ensembles of neurons that can be assumed to have homogeneous activity. Strictly speaking, here we violate this assumption by taking the rate prediction of theoretical ensembles as the prediction of the firing rates (and mean voltages) of single neurons at given positions along the ring attractor. However, since we are investigating the stationary state of networks, the mean firing rates calculated from this theory can also be interpreted as the time averaged firing rate of single neurons (Brunel and Sergi, 1998). Further, the approximation of the recurrent synaptic inputs in the steady state usually relies on the averaging over presynaptic ensembles of homogeneously firing neurons. Nonetheless, the theory still works quite well, which might be due to the fact that we are calculating the synaptic drive as an integral over a continuum of presynaptic neurons, thereby effectively averaging out deviations of the synaptic drive that are to be expected in single neuron samples from such ensembles.

Although bump attractor states are interesting from a functional point of view, they are not the only solutions to ring-like networks. As we have seen, spiking networks can show bi-stability, in which both a stable uniform state and a stable evoked bump-state co-exist. Here, we have neglected solutions of our theory that converged to the uniform state (or intermediate unstable solutions) if a bump-state at larger firing rates was also found as a solution. While it goes beyond the scope of the current work, our theory could be extended to also predict the dynamical stability of the states that are found: similar to approaches in networks that admit a splitting into discrete populations (Amit and Brunel, 1997) one could measure the magnitude of perturbations to the firing rates at the points θ_i resulting from perturbations to the parametrization: the general perturbation $g(g_0, g_1, g_\sigma, g_r) \rightarrow g(g_0 + \delta g_0, g_1 + \delta g_1, g_\sigma + \delta g_\sigma, g_r + \delta g_r)$ would translate into both perturbations of the rates $g(\theta_i) = \nu_0(\theta_i) + \delta\nu(\theta_i)$ and the prediction $F(\text{input}_{\theta_i}\left[g_0\right], \bar{V}(\theta_i)) + \delta F(\theta_i)$. Stability can then be determined by comparing the scale of the input perturbations $\delta\nu(\theta_i)$ to the predicted output perturbations $\delta F(\theta_i)$ (Amit and Brunel, 1997). It is worth noting, that calculating such linear perturbations will involve the derivative of the synaptic activations ψ (cf. Eq. (3.5)), which will also have to be estimated numerically in cases where no analytical formula for ψ can be found.

Our choice of parametrization of the firing rate profile was heuristic, guided by the shapes observed in numerical simulations. The framework presented here, though, could be used with any other family of functions parametrized by a small number of parameters, which should be adapted to the shapes to be approximated. For example, multi-modal ring-attractor profiles resulting from narrower connectivities (see e.g. Laing et al. (2002), or Wei et al. (2012) for a spiking network similar to the one investigated here) can not be approximated by the unimodal family chosen here. Since the topology of ring-attractors is periodic, a natural candidate for such a generalization would be the family of finite Fourier series. However, the nonzero frequency components necessary to faithfully approximate shapes that deviate far from simple (e.g. cosine-shaped) unimodal distributions might require a large number of Fourier coefficients for parametrization.

Throughout this report, we chose as many positions θ_i along the attractor manifold as there are free parameters of the profile g to be constrained. In principle, the number of free parameters can be chosen independently of the number of positions, by performing numerical optimization on a dimension-agnostic sum of squared errors. We have shown that matching the error dimension to the number of parameters permits using efficient optimization methods that significantly speed up the optimization. However, we have also investigated under-determined (see Fig. 3.6B) and over-determined (see Fig. 3.5A-C) systems, which also converged to similar solutions. Finally, when optimizing the network parameters for desired spatial profiles (Fig. 3.3 and Fig. 3.6), choosing optimization goals outside the space of possible solutions of the network dynamics did not allow the procedure to converge. Thus, our approach could be used to estimate the boundaries of the solution space for a given neural network, by starting the optimization at a known solution and varying the shape parameters until convergence fails.

3.6 Methods

3.6.1 Rate network model

We study a network of $N = 100$ recurrently connected rate neurons indexed by $i \in \{0, ..., N-1\}$, where each neuron i is described by a single variable $\nu_i(t)$ which denotes the firing rate of the neuron (Miller and Fumarola, 2012).

The neuronal firing rate is given by a nonlinear input-output function F of a synaptic variable s_i (assuming that the membrane time constant is considerably faster than the synaptic variable):

$$\nu_i(t) = F(s_i(t)) = \frac{\nu_{\max}}{2}\left[1 + \tanh\left(\frac{s_i(t)}{s_0}\right)\right]. \tag{3.9}$$

Here, $\nu_{\max} = 50\text{Hz}$ sets the maximal firing frequency, and $s_0 = 1$ is the (dimensionless) scale of the synaptic input.

The input to neuron i is mediated through the synaptic variable s_i:

$$\dot{s}_i(t) = -\frac{s_i(t)}{\tau_s} + \sum_{j=0}^{N-1} w_{ij}\nu_j(t), \tag{3.10}$$

where $\dot{s}(t) = \frac{d}{dt}s(t)$ denotes the temporal derivative, $\tau_s = 100ms$ is the synaptic time constant, and w_{ij} are the recurrent connection weights (see below), and $0 \le i \le N-1$.

Neurons are organized at circular positions $\theta_i = i \cdot \frac{2\pi}{N} - \pi \in [-\pi, \pi)$ with identified boundaries, such that neuron 0 is the direct neighbor of neuron $N-1$. The recurrent connections depend only on the distance between neurons in the resulting angular space: the connection w_{ij} from neuron j to neuron i is given by a generalized Gaussian function, with 4 free parameters controlling its shape:

$$w_{ij} = w(\theta_i - \theta_j) = \frac{1}{N}\left(w_0 + w_1 \exp\left[-\left(\frac{\left|\min\left(\left|\theta_i - \theta_j\right|, 2\pi - \left|\theta_i - \theta_j\right|\right)\right|}{w_\sigma}\right)^{w_r}\right]\right). \tag{3.11}$$

The parameters $\{w_0, w_1, w_\sigma, w_r\}$ used for the networks of Fig. 3.2 and Fig. 3.3 are given in Table C.2.

3.6.2 Spiking network model

Spiking simulations are based on a reimplementation of a popular ring-attractor model of visuospatial working memory (Compte et al., 2000) (and used with variations in e.g. Renart et al. (2003); Murray et al. (2012); Wei et al. (2012); Pereira and Wang (2015); Almeida et al. (2015)) in the NEST simulator (Diesmann and Gewaltig, 2007). Parameters were modified from the original publication to produce the results shown in Fig. 3.4, Fig. 3.5, and Fig. 3.6 (see Table C.1 for parameter values). For completeness we restate the definition of the model here.

3.6.2.1 Neuron model

Neurons are modeled by leaky integrate-and-fire dynamics with conductance based synaptic transmission (Compte et al., 2000; Brunel and Wang, 2001b). The network consists of recurrently connected populations of N_E excitatory and N_I inhibitory neurons, both additionally receiving external spiking input with spike times generated by N_{ext} independent, homogeneous Poisson processes, with mean rates ν_{ext}. Following Compte et al. (2000), we assume that external excitatory inputs are mediated by fast AMPA receptors, while recurrent excitatory currents are mediated only by slower NMDA channels.

The neuronal dynamics for neurons in both excitatory and inhibitory populations are governed by the following system of differential equations indexed by $i \in \{0, ..., N_{E/I} - 1\}$ (with different sets of parameters for each population):

$$\begin{aligned} C_m \dot{V}_i(t) &= -I_i^L(t) - I_i^{\text{Ext}}(t) - I_i^{\text{I}}(t) - I_i^{\text{E}}(t), \\ I_i^P &= g_P \, s_i^P(V_i(t), t) \, (V_i(t) - V_P), \end{aligned} \tag{3.12}$$

where $P \in$ {L,Ext,I,E}. Here, C_{m} is the membrane capacitance and V_{L}, V_{E}, V_{I} are the reversal potentials for leak, excitatory currents, and inhibitory currents, respectively. The parameters g_P for $P \in$ {L,Ext,I,E} are fixed scales for leak (L), external input (Ext) and recurrent excitatory (E) and inhibitory (I) synaptic conductances, which are dynamically gated by the (possibly voltage dependent) gating variables $s_i^P(V, t)$. In the main text we refer to the conductance scales of excitatory neurons by the “strength of synaptic connections” $g_{\text{EE}} = g_{\text{E}}$ and $g_{\text{EI}} = g_{\text{I}}$. Similarly, for inhibitory neurons we refer to the conductance scales by the “strengths” $g_{\text{IE}} = g_{\text{E}}$ and $g_{\text{II}} = g_{\text{I}}$. The gating variables s_i^P are described in detail below, however we set the leak conductance gating variable to $s_i^L = 1$.

The model neuron dynamics (Eq. (3.12)) are integrated until their voltage reaches a threshold V_{thr}. At any such time, the respective neuron emits a spike and its membrane potential is reset to the value V_{res}. After each spike, voltages are clamped to V_{res} for a refractory period of τ_{ref} (see Table C.1 for parameter values).

3.6.2.2 Synaptic gating variables

The synaptic gating variables $s_i^P(t)$ for $P \in$ {Ext,I} for external and inhibitory currents are exponential traces of the firing times t_j of all presynaptic neurons j:

$$\dot{s}_i^P(t) = -\frac{s_i^P(t)}{\tau_P} + \sum_{j \in \text{pre}(P)} \sum_{t_j} \delta\left(t - t_j\right), \tag{3.13}$$

where the sum runs over all neurons presynaptic to the neuron i regarding the connection P.

For the recurrent excitatory gating variables s_i^{E} a nonlinear NMDA model is used (Wang, 2001). This model has second order kinetics for NMDA channel activation (Destexhe et al., 1994), which result in a saturation of channels. Together with a voltage dependence $\mathrm{Mg}(V_i)$ of the conductance (due to the release of the Mg^{2+} block, see Jahr and Stevens (1990a)) this yields the following dynamics:

$$\begin{aligned}
s_i^{\mathrm{E}}(V,t) &= \mathrm{Mg}(V_i)\sum_{j=1}^{N_E} w_{ij}^{\mathrm{E}} y_j(t), && (3.14)\\
\dot{y}_j &= -\frac{y_j}{\tau_{\mathrm{E}}} + \alpha x_j(t)\left(1-y_j\right), && (3.15)\\
\dot{x}_j &= -\frac{x_j}{\tau_{\mathrm{E,rise}}} + \sum_{t_j}\delta\left(t-t_j\right), && (3.16)\\
\mathrm{Mg}(V) &= \frac{1}{1+\gamma\exp\left(-\beta V\right)}. && (3.17)
\end{aligned}$$

See Table C.1 for parameter values used in simulations.

3.6.2.3 Network connectivity

All connections except for the recurrent excitatory connections are all-to-all and uniform. The recurrent excitatory connections are chosen to be distance-dependent. As in the rate model, each neuron of the excitatory population with index $i \in \{0,\dots,N_E-1\}$ is assigned an angular position $\theta_i = i\cdot\frac{2\pi}{N_E} - \pi \in [-\pi,\pi)$. Recurrent excitatory NMDA connections w_{ij}^E from neuron j to neuron i are then given by the Gaussian function $w^E(\theta)$:

$$w_{ij}^E = w^E(\theta_i-\theta_j) = w_0 + (w_+ - w_0)\exp\left(-\left[\min\left(\left|\theta_i-\theta_j\right|, 2\pi - \left|\theta_i-\theta_j\right|\right)\right]^2 \frac{1}{2\sigma_w^2}\right).$$

Additionally, for each neuron we keep the integral over all recurrent connection weights normalized, resulting in the normalization condition $\frac{1}{2\pi}\int_{-\pi}^{\pi} d\varphi\, w^E(\varphi) = 1$. This normalization ensures that varying the maximum weight w_+ will not change the total recurrent excitatory input if all excitatory neurons fire at the same rate. Here, we choose w_+ as a free parameter and constrain the baseline connection weight to

$$w_0 = \frac{w_+\sigma_w \mathrm{erf}\left(\frac{\pi}{\sqrt{2}\sigma_w}\right) - \sqrt{2\pi}}{\sigma_w \mathrm{erf}\left(\frac{\pi}{\sqrt{2}\sigma_w}\right) - \sqrt{2\pi}}.$$

3.6.3 Self-consistent equations

3.6.3.1 Placement of sampling points

Self consistent equations (Eq. (3.2)) are constructed for both rate-based and spiking neurons by using the low-dimensional parametrization Eq. (3.1) described in the main text. As mentioned there, we choose the top of the firing rate profile $\theta_1 = 0$, as well as the lowest point $\theta_4 = \pi$. For the intermediate points $0 < \theta_i < \pi$ for $i \in \{2,3\}$, we sample the firing rate profile by inverting the function g to give a sample at a desired height $h_i = g_0 + a_i \left(g_1 - g_0\right)$ with $0 < a_i < 1$. This yields a relation for the position which depends on the shape parameters g_r and g_σ:

$$\theta_i = -g_\sigma \log(a_i)^{\frac{1}{g_r}} .$$

For all figures except for Fig. 3.5 and Fig. 3.6, the intermediate points were chosen by setting $a_2 = 0.2$ and $a_3 = 0.8$, although we show (Fig. 3.2 and Fig. 3.3) that the exact choice of points affects the solutions only slightly.

In Fig. 3.5 we iterate through even numbers $p \geq 4$ of sampling points. As before, we choose $\theta_1 = 0$, as well as the lowest point $\theta_p = \pi$. Generalizing the placement of 4 points described above, the remaining $p - 2$ points were chosen as $\theta_k = {}^{0.4}/_{k}$ for $1 < k \leq {}^{p}/_{2}$, and $\theta_k = 1 - {}^{0.4}/_{k}$ for ${}^{p}/_{2} < k \leq p$. For the optimization of spiking network parameters shown in Fig. 3.6, we chose 7 sampling points: we first chose $p = 6$ points by the scheme just described, then added the point $\theta_7 = 0.5$.

3.6.3.2 Derivation of input-output functions for the rate network

For the rate network, we set Eq. (3.10) to zero and solve for the steady-state input s_i, which yields

$$\begin{aligned} s_i &= \tau_s \sum_{j=0}^{N-1} w_{ij} \nu_j = \tau_s \sum_{j=0}^{N-1} w_{ij} g(\theta_j) \\ &\approx \frac{\tau_s N}{2\pi} \int_{-\pi}^{\pi} d\varphi\, w(\theta_i - \varphi) g(\varphi). \end{aligned}$$

Here, we have replaced the activity of neurons in the network by our parametrization g. In the second line we approximated the summation $\frac{1}{N}\sum_{j=0}^{N-1}$ by the integral $\frac{1}{2\pi}\int_{-\pi}^{\pi} d\varphi$ and exploited that the connectivity w is only dependent on the angular distance between the neuron i (at position θ_i) and neurons j (at varying positions $\theta_j = \varphi$), to replace $w_{ij} \to w(\theta_i - \varphi)$. We use this steady state input in Eq. (3.9) to arrive at Eq. (3.4).

3.6.3.3 Derivation of input-output functions for the spiking network

In the rate model presented in Section 3.6.1 the firing rates are given by Eq. (3.9). In the spiking network, we have to approximate the expected firing rates of neurons. To this end, we first replace the synaptic activation variables $s^P(V,t)$ for $P \in \{I, E, \text{ext}\}$ by their expectation values under Poisson input. For the linear synapses this yields

$$\begin{aligned}\left\langle s^{\text{ext}}(t)\right\rangle_{\text{Poisson}} =& \tau_{\text{ext}}\nu_{\text{pre}},\\ \left\langle s^{\text{I}}(t)\right\rangle_{\text{Poisson}} =& \tau_{\text{I}}\nu_I.\end{aligned}$$

The nonlinear synaptic activation of NMDA synapses under stimulation with Poisson processes at rates ν_j was estimated by simulating Eqs. (3.15)-(3.16) under varying presynaptic firing rates and fitting an interpolating function to the temporal means of the synaptic activation $\psi(\nu_j) \equiv \langle y_j \rangle_t$. An analytical approximation of the function $\psi(\nu_j)$ was stated in Brunel and Wang (2001b, p. 80) (see Fig. D.1 in Section D.1 for a comparison to simulations). We instead chose to numerically fit this function to simulated data, since for higher firing rates the analytical approximation tended to over-estimate the synaptic activations.

We then define the expected recurrent excitatory input, assuming presynaptic Poisson firing, by

$$J_i \equiv \frac{1}{N_E}\sum_{i=0}^{N_E-1} w_{ij}^E \psi(\nu_j). \tag{3.18}$$

Following Brunel and Wang (2001b), we linearize the voltage dependence Eq. (3.17) at the mean voltage $\langle V \rangle$ and reduce the differential equations of Eq. (3.12) to dimensionless form. The resulting expressions depend only on the mean firing rates and mean voltages of excitatory and inhibitory neurons (see Section 3.6.3.4 for the detailed expressions and derivations):

$$\begin{aligned}\tau_i \dot{V}_i &= -(V_i - V_L) + \mu_i + \sigma_i\sqrt{\tau_i}\eta_i(t) \qquad (3.19)\\ \mu_i &= \mu_i\left(J_i, \nu_I, \nu_{\text{ext}}, \langle V_i\rangle\right)\\ \sigma_i &= \frac{g_{\text{ext}}}{C_m}\left(\langle V\rangle - V_E\right)\tau_{\text{ext}}\sqrt{\tau_i N_{\text{ext}}\nu_{\text{ext}}}.\\ \tau_i &= \tau_i\left(J_i, \nu_I, \nu_{\text{ext}}, \langle V_i\rangle\right)\\ \left\langle \eta_i(t)\right\rangle &= 0\\ \left\langle \eta_i(t)\eta_i(t')\right\rangle &= \frac{1}{\tau_{\text{ext}}}\exp\!\left(-\frac{\left|t-t'\right|}{\tau_{\text{ext}}}\right)\end{aligned}$$

Here, μ_i is the bias of the membrane potential due to synaptic inputs, and σ_i measures the scale of fluctuations in the membrane potential due to random spike arrival approximated by the Gaussian process η_i. Due to active synaptic conductances, the effective membrane time constant τ_i is decreased from the intrinsic membrane time-constant C_m/g_L – its value thus depends on all presynaptic firing rates (and the mean voltage, see Section 3.6.3.4).

The prediction F of the mean firing rates and $\langle V_i \rangle$ of mean voltages of populations of neurons governed by this type of differential equation can be well approximated by (Brunel and Wang, 2001b) (see also the published corrections in Brunel and Wang (2001a)):

$$\begin{aligned}
\phi\left[\mu_i, \sigma_i, \tau_i\right] &= \left(\tau_{\text{ref}} + \sqrt{\pi}\tau_i \int_{\beta(\mu_i,\sigma_i)}^{\alpha(\mu_i,\sigma_i)} du \exp(u^2)\,[1+\operatorname{erf}(u)]\right)^{-1}, && (3.20)\\
\alpha(\mu_i,\sigma_i) &= \frac{V_{\text{reset}} - V_L - \mu_i}{\sigma_i}\left(1 + \frac{\tau_{\text{ext}}}{2\tau_i}\right) + 1.03\sqrt{\frac{\tau_{\text{ext}}}{\tau_i}} - \frac{\tau_{\text{ext}}}{\tau_i}, && (3.21)\\
\beta(\mu_i,\sigma_i) &= \frac{V_{\text{reset}} - V_L - \mu_i}{\sigma_i}, && (3.22)\\
\langle V_i \rangle &= \mu_i + V_L - (V_{\text{thr}} - V_{\text{reset}})\,\phi\left[\mu_i,\sigma_i,\tau_i\right]\tau_i. && (3.23)
\end{aligned}$$

As in the rate model, we first replace the network activity ν_j on the right hand side of Eq. (3.18) by our parametrization $g(\theta_j)$. We then approximate the summation $\frac{1}{N_E}\sum_{j=0}^{N_E-1}$ with an integral $\frac{1}{2\pi}\int_{-\pi}^{\pi} d\varphi$, and replace the connectivity by its continuous equivalent $w_{ij} \to w(\theta_i - \varphi)$ to arrive at:

$$\begin{aligned}
J_i &\approx \frac{1}{2\pi}\int_{-\pi}^{\pi} d\varphi\, w^E\left(\theta_i - \varphi\right)\psi(g\left(\varphi\right))\\
&\equiv \frac{1}{2\pi}\text{input}_{\theta_i}\left[g\right].
\end{aligned}$$

We then substitute this relation in Eqs. (3.20) and (3.23) to arrive at

$$\begin{aligned}
\mu_i &= \mu_i\left(\text{input}_{\theta_i}\left[g\right], \nu_I, \nu_{\text{ext}}, \langle V_i\rangle\right),\\
\tau_i &= \tau_i\left(\text{input}_{\theta_i}\left[g\right], \nu_I, \nu_{\text{ext}}, \langle V_i\rangle\right),\\
F(\text{input}_{\theta_i}\left[g\right], \nu_I, \langle V_i\rangle) &\equiv \phi\left[\mu_i,\sigma_i,\tau_i\right],\\
G(\text{input}_{\theta_i}\left[g\right], \nu_I, \langle V_i\rangle) &\equiv \mu_i + V_L - (V_{\text{thr}} - V_{\text{reset}})\, g_i(\theta_i)\tau_i, \qquad (3.24)
\end{aligned}$$

which defines Eqs. (3.5) and (3.6) of the main text.

3.6.3.4 Detailed derivation of dimensionless voltage equations

In this section we give details on the derivation of Eq. (3.19) as well as the resulting full expressions. This closely follows Brunel and Wang (2001b, pp. 79–81), while keeping a slightly simplified notation.

We first replace all synaptic activations in Eq. (3.12) by their expected values under Poisson input, which also introduces the expected recurrent excitatory input J_i (cf. Eq. (3.18)):

$$\begin{aligned} s_i^{\text{ext}}(t) &\to N_{\text{ext}}\tau_{\text{ext}}\nu_{\text{pre}} + \Delta_{S,ext}, \\ s_i^{\text{I}}(t) &\to N_I\tau_{\text{I}}\nu_I, \\ s_i^E(t) &\to \text{Mg}(V_i)\sum_{i=0}^{N_E-1} w_{ij}^E\psi(\nu_j) = \text{Mg}(V_i)N_E J_i. \end{aligned}$$

Here, $\Delta_{S,ext}$ represents fluctuations of around the mean of s_i^{ext} due to random spike arrival at fast AMPA synapses. Since the synaptic timescales (GABA, NMDA) of the other synaptic activations are much longer, these fluctuations can be neglected. We then rearrange Eq. (3.12) to dimensionless form, which yields:

$$\begin{aligned} \frac{C_m}{g_L}\dot{V}_i &= -(V_i - V_L)\left[1 + T_I\nu_I + T_{\text{ext}}\nu_{\text{ext}} + \frac{g_E}{g_L}\text{Mg}(V_i)N_E J_i\right] \\ &\quad + (V_I - V_L)\,T_I\nu_I + (V_E - V_L)\left[T_{\text{ext}}\nu_{\text{ext}} + \frac{g_E}{g_L}\text{Mg}(V_i)N_E J_i\right] \\ &\quad + \frac{g_{\text{ext}}}{g_L}(V_i - V_E)\,\Delta_{S,\text{ext}}, \end{aligned}$$

where $T_{\text{ext}} = N_{\text{ext}}\tau_{\text{ext}}\frac{g_{\text{ext}}}{g_L}$, $T_I = N_I\tau_I\frac{g_I}{g_L}$ are effective timescales of external and inhibitory input.

To get rid of the nonlinear voltage dependence of the right hand side through $\text{Mg}(V_i)$, we linearize this function (cf. Eq. (3.17)) around the mean voltage $\langle V_i\rangle$:

$$\frac{V_i - V_E}{1 + \gamma\exp(-\beta V_i)} = \frac{\langle V_i\rangle - V_E}{\rho} + (V - \langle V\rangle)\,\frac{\rho + \beta(\langle V_i\rangle - V_E)\big(\rho - 1\big)}{\rho^2},$$

where $\rho = 1 + \gamma\exp\big(-\beta\langle V_i\rangle\big)$.

After replacing the voltage dependence in the fluctuation term by the mean voltage, we arrive at

$$\begin{aligned} \frac{C_m}{g_L}\dot{V}_i &= -(V_i - V_L)\left[1 + T_I\nu_I + T_{\text{ext}}\nu_{\text{ext}} + \big(\rho_1 + \rho_2\big)J_i\right] \\ &\quad + (V_I - V_L)\,T_I\nu_I + (V_E - V_L)\left[T_{\text{ext}}\nu_{\text{ext}} + \rho_1 J_i\right] \\ &\quad + \rho_2(\langle V_i\rangle - V_L)\,J_i + \frac{g_{\text{ext}}}{g_L}(\langle V_i\rangle - V_E)\,\Delta_{S,\text{ext}}. \\ \rho_1 &= \frac{g_{\text{E}}N_E}{g_L\rho} \\ \rho_2 &= \beta\frac{g_{\text{E}}N_E(\langle V_i\rangle - V_E)\big(\rho - 1\big)}{g_L\rho^2} \\ \rho &= 1 + \gamma\exp\big(-\beta\langle V_i\rangle\big). \end{aligned} \tag{3.25}$$

Finally, we replace the fluctuations $\Delta_{S,ext}$ by independent Gaussian noise processes with zero mean $\langle \eta_i(t) \rangle = 0$ and simpler autocorrelation $\langle \eta_i(t)\eta_j(t') \rangle = \frac{1}{\tau_{\text{ext}}} \exp(-\frac{|t-t'|}{\tau_{\text{ext}}})\delta_{ij}$, to arrive at the full form of Eq. (3.19) in the main text:

$$\begin{aligned}
\tau_i \dot{V}_i &= -(V_i - V_L) + \mu_i + \sigma_i \sqrt{\tau_i}\eta_i(t) \\
S_i &= 1 + T_I \nu_I + T_{\text{ext}} \nu_{\text{ext}} + \left(\rho_1 + \rho_2\right) J_i \\
\mu_i S_i &= (V_I - V_L)\, T_I \nu_I + (V_E - V_L)\, T_{\text{ext}} \nu_{\text{ext}} + \\
&\quad \left[\rho_1 (V_E - V_L) + \rho_2 (\langle V \rangle - V_L)\right] J_i \\
\sigma_i &= \frac{g_{\text{ext}}}{C_m} (\langle V \rangle - V_E)\, \tau_{\text{ext}} \sqrt{\tau_i N_{\text{ext}} \nu_{\text{ext}}}. \\
\tau_i &= \frac{C_m}{g_L S_i}
\end{aligned} \tag{3.26}$$

Reducing the conductance based differential equation Eq. (3.12) of the main text to the simplified form Eq. (3.26), now allows us to compute the mean firing rate as a functions of the (input-like) bias μ_i and fluctuation term σ_i, according to Eq. (3.20) of the last section.

3.6.3.5 Optimization of self-consistent equations

For each point θ_i that we choose to sample from the excitatory population, the theory of Section 3.6.3.3 yields 2 constraining Equations (3.20) and (3.23). The inhibitory population, being homogeneous and unstructured, yields 2 equations, for the 2 free variables ν_I and $\langle V_i \rangle$. Since we choose a low-dimensional parametrization for the excitatory population, the number of free variables increases only by 1 (the mean voltage $\langle V_i \rangle$) for each point θ_i that we choose to evaluate, while yielding the same 2 constraining equations. This allows us to choose at minimum 4 evaluation points to constrain the 4 free parameters of the parametrization (see Table C.3 for a listing).

The errors Err_i (and Err_I, for spiking networks) between firing rate predictions and the firing rate parametrization are numerically minimized using methods provided in the *Scipy* package (Oliphant, 2007). In particular, if the dimension of the error function matches the number of parameters, we are able to use the efficient *optimize.root* solver (*Root* in the main text), which applies a modified version of the Powell hybrid method (Powell, 1970), but does not provide constraints on valid parameter regions. Here, we implemented artificial constraints by returning a high error for dimensions that leave the bounded region. The same optimization results were achieved by using the slower *optimize.minimize* method, which allows optimization (of the sum of squared errors $\text{SSE} = \sum_i \text{Err}_i^2$, or $\text{SSE} = \sum_i \text{Err}_i^2 + \text{Err}_I^2$ for spiking networks) in constrained parameter regions via the L-BFGS-B (Byrd et al., 1995) and SLSQP (Kraft, 1988). For spiking networks, we normalized firing rate errors by the firing rate $\nu_{\max} = 100 Hz$ and voltage differences by the voltage range $V_{\text{thres}} - V_{\text{reset}}$, to ensure comparable contributions to the SSE for variables with different dimensions.

For the optimization results of Fig. 3.6 we chose 7 sampling points (see Section 3.6.3.1 for details), which yielded 16 errors, including those of the inhibitory population. These were used to optimize 17 free parameters using the SLSQP algorithm (see Table C.3 for a listing). We also tried using 8 sampling points, which brings both the number of equations and free parameters up to 20 – this yielded similar results at increased processing time (the possibly faster *Root* solver failed to converge most of the time).

Wall clock times for error functions in Fig. 3.5A were measured on a single core of a MacBook Pro with 2,6 GHz Intel Core i5 processor, using the Python benchmark *timeit.timeit* (minimum wall clock time of 100 repetitions). We first measured average time for evaluation of a single error Err_i of Eq. (3.7), which evaluated to $t_E = 4.59ms$ (100 repetitions of 10 executions). For a single evaluation of the inhibitory error Err_I of Eq. (3.8) we found $t_I = 0.98ms$ (the numerical integration performed in the calculation of $\mathrm{input_I} = \frac{1}{2\pi}\int d\varphi\psi\left(g\left(\varphi\right)\right)$ is faster). The wall clock time T (see Fig. 3.5A, right axis) for a given number n of error vector evaluations on p points was then calculated by $T = n\left(p \cdot t_E + t_I\right)$.

3.6.4 Spiking simulations

All network simulations where performed in the NEST simulator (Diesmann and Gewaltig, 2007) using fourth-order Runge-Kutta integration as implemented in the GSL package (Galassi et al., 2009). For the simulation results shown in Fig. 3.4 and Fig. 3.6B, networks underwent a transient initial period of t_{initial}. Neurons centered at a position of $\theta = 0$ then received a short and strong excitatory input mediated by additional Poisson firing onto AMPA receptors ($500ms$, 2kHz) with connections scaled down by a factor of $g_{\mathrm{signal}} = 0.5$. The external input ceased at $t = t_{\mathrm{off}}$.

Simulations were run until $t = t_{\max}$ and spikes were recorded and converted to firing rates by spike counts in a $75ms$ window shifted at a time resolution of $1ms$. For every time step, the firing rates across the whole population were then rectified (by measuring the phase of the first spatial Fourier coefficient and setting it to $\theta = 0$ by rotation of the angular space) to center the bump of activity around the position $\varphi = 0$. The resulting centered firing rates were then sampled at an interval of $60ms$ in the interval $[t_{\mathrm{off}} + 500ms, t_{\max}]$ for 5 repetitions of the network simulation with the same microscopic parameters. In Fig. 3.4B times were: $t_{\mathrm{initial}} = 1.8s$, $t_{\mathrm{off}} = 2s$, $t_{\max} = 5s$. In Fig. 3.4D,E we chose: $t_{\mathrm{initial}} = 0.5s$, $t_{\mathrm{off}} = 1s$, $t_{\max} = 3s$. For Fig. 3.6B: $t_{\mathrm{initial}} = 0.5s$, $t_{\mathrm{off}} = 1s$, $t_{\max} = 5s$.

4 Continuous attractor networks: Effects of short-term plasticity on drift and diffusion

Alexander Seeholzer, **Moritz Deger** and **Wulfram Gerstner** (in preparation)

4.1 Author contributions

I conceived the study, wrote all simulation code and carried out all experiments and analyses. The theory was developed by me, with support by MD. I wrote the manuscript, with help from MD and WG.

4.2 Abstract

Continuous attractor neuronal networks are frequently used to model cortical short-term memory of continuous-valued sensory information in the brain. However, these models have been shown to be highly sensitive to the inevitable variability of biological neurons: variable neuronal spike timing, sparse synaptic connectivity and heterogeneous neuronal populations decrease the time that memories can be accurately retained, eventually leading to a loss of memory functionality on behaviorally relevant time scales. Short-term plasticity of recurrent synapses has previously been shown to be able to exert influence on the dynamics of continuous attractor models. As a candidate mechanism for stabilization of working memory activity, short-term facilitation can stabilize the retention of memories. Short-term synaptic depression, on the other hand, can make continuous attractors more volatile. However, it remains unclear to which degree these two short-term plasticity mechanisms influence the dynamics of continuous attractor networks together, and whether they can provide a biologically plausible stabilizing mechanism for working memory systems.

Here, we present a comprehensive analysis of the effects of short-term facilitation and depression on the dynamics of memories encoded in continuous attractor models. We derive a general theoretical description of the slow dynamics of the center position, which represents the encoded memory, by drift and diffusion depending on the parameters of short-term

plasticity. The theory predicts quantitatively the effects of short-term plasticity when applied to a ring-attractor implemented in a spiking network: we find that facilitation can decrease both random displacement due to spiking variability, as well as directed drifts due to sparse connectivity and neuronal heterogeneity, while short-term depression tends to increase both. Using the theory, we are able to place constraints on the combinations of network and synapse properties that enable a stable retention of memories under the presence of biological variability. Finally, we apply the theory to derive continuous attractor networks with facilitation and depression that maintain a given steady-state distribution of memories, and show that short-term facilitation can provide a mechanism to both represent and control uncertainty of memorized information.

4.3 Introduction

As confirmed in behavioral tasks involving working memory, information about past environmental stimuli can be stored in the primate brain and be retrieved seconds later (Goldman-Rakic, 1995; Constantinidis and Wang, 2004). Strikingly, this transient storage for several seconds is achieved using ingredients operating mostly on time scales of tens of milliseconds and shorter (Chaudhuri and Fiete, 2016). A working hypothesis of neuroscience is that this function emerges from cortical neuronal network dynamics: memories are retained by self-generating cortical activity through positive feedback (Curtis and D'Esposito, 2003; Durstewitz et al., 2000b; Barak and Tsodyks, 2014), where the echoes of past stimuli reverberate over time in the network activity and thereby bridge the time scales from milliseconds (neuronal dynamics) to seconds (behavior) (Duarte et al., 2017).

Analog sensory stimuli present in the environment often are continuously structured: for example, positions of objects in the visual field, frequencies of auditory stimuli, or the intensity and position of sensory stimuli on the body, all have continuously varying values. The organization of cortical working memory circuits could reflect this continuous nature of sensory information, thereby allowing storage of all possible values and a conservation of the metric relations between stimuli (Chaudhuri and Fiete, 2016). This idea is supported by the finding that, although behavioral experiments often sample continuously structured stimuli at discrete spacings, the firing rates of neurons involved in working memory vary continuously with the presented stimuli (Goldman-Rakic, 1995; Romo et al., 1999). For example, firing rates of neurons encoding object locations during delay periods in prefrontal cortex are found to be distributed around a preferred location in a bell-shaped manner (Funahashi et al., 1989). A much researched class of cortical working memory models able to store continuously structured information is that of *continuous attractors*. Symmetries in synaptic connections mediating the self-generated reverberations allow these models to have a continuum of meta-stable states, which can be used to retain memories over time scales (delay periods) much longer than those of the single network constituents. While continuous attractor models arose primarily from theoretical considerations (Wilson and Cowan, 1973; Amari, 1977), they were soon proposed as models for cortical working memory (Camperi and

Wang, 1998; Hansel and Sompolinsky, 1998; Compte et al., 2000), path integration (Samsonovich and McNaughton, 1997; Stringer et al., 2002; Burak and Fiete, 2009), and other cortical functions (Ben-Yishai et al., 1995; Zhang, 1996; Seung, 1998) (see e.g. Knierim and Zhang (2012); Moser et al. (2014); Burak (2014); Wu et al. (2016); Chaudhuri and Fiete (2016) for recent reviews). Recently, supporting experimental evidence for continuous attractor dynamics in cortical networks (Wimmer et al., 2014) and the limbic system (Knierim and Zhang, 2012; Yoon et al., 2013) has been reported, alongside evidence for the existence of continuously structured population activity in the fly that is responsible for path integration and self-orientation (Seelig and Jayaraman, 2015) (although the circuit mechanisms responsible for its generation have yet to be identified).

Cortical working memory with a continuum of memory states might be very useful in theory. However, the meta-stability of memory states in continuous attractors leaves the encoded memories vulnerable to noisy processes that break, transiently or permanently, the crucial symmetry underlying the attractor dynamics (Chaudhuri and Fiete, 2016). For example, the fast variability of neuronal spiking leads to transient asymmetries that randomly displace encoded memories along the continuum of states (Hansel and Sompolinsky, 1998; Compte et al., 2000; Wu et al., 2008; Kilpatrick and Ermentrout, 2012), leading to *diffusion* of encoded variables. Introducing fixed (on the time scale of working memory tasks) asymmetries has more drastic effects: they introduce “frozen” biases in neuronal activity leading to *directed drift* of memories and a collapse of the continuum of attractive states to a set of discrete states. Examples of such *frozen noise* include the heterogeneity of neuronal parameters (Renart et al., 2003), the sparsity of recurrent connections (Stringer et al., 2002; Renart et al., 2003), or adding fixed noise to the recurrent weights (Tsodyks and Sejnowski, 1995; Zhang, 1996; Itskov et al., 2011b). Since both fast and frozen *biological variabilities* are to be expected in cortical settings, the feasibility of continuous attractors being implemented as computational systems of the brain is called into question.

Recently, short-term plasticity of recurrent synaptic connections has been shown to influence the susceptibility of continuous attractor networks to both fast and frozen noise. Short-term depression (Zucker and Regehr, 2002) – the use-dependent depletion of available synaptic resources – has been claimed to not have an effect on diffusion in spiking networks (Barbieri and Brunel, 2007), but has also been shown to induce directed displacement on continuous attractors in rate models (York and van Rossum, 2009; Romani and Tsodyks, 2015). Short-term facilitation (Zucker and Regehr, 2002) – the use-dependent increase of utilization of synaptic resources – has been proposed as a stabilizing mechanisms that could increase the retention time of memories in continuous attractor networks. In networks of simplified rate neurons without fast noise (Itskov et al., 2011b), this has been shown to significantly decrease the amount of drift that can be expected. In simulations of a continuous attractor implemented with spiking neurons, facilitation was reported to underlie the observed slow drift (Hansel and Mato, 2013) and was shown to reduce the amount of diffusion for a single value of facilitation strength (Pereira and Wang, 2015) (see Discussion). However, several general questions previously remained unaddressed. What are the quantitative effects of short-term facilitation in more complex neuronal models and across facilitation parameters? Can short-term depression also influence the strength of directed drift, as well as diffusion? More generally, it is currently still unclear how facilitation and depression interplay in these dynamical roles. Finally, can theory be used to predict all of these effects?

Here we present a comprehensive description of the effects of short-term facilitation and depression on one-dimensional continuous attractor models. First, we introduce the general framework and the concepts of drift and diffusion in one-dimensional ring-attractor models (Section 4.4.1). To demonstrate the application of these concepts, we then describe a spiking network implementation of a ring-attractor model (Section 4.4.2) with short-term facilitation and depression. In Section 4.4.3, we derive theoretical predictions of the amount of diffusion and drift in continuous attractor models implemented in networks of neurons with arbitrary input-output relations. These theoretical predictions are then compared with simulations of the spiking network in Section 4.4.4 and Section 4.4.5. We find that facilitation and depression play an antagonistic role in making continuous attractors robust against noise, where facilitation (depression) can decrease (increase) both diffusion and drift. The theory allows us to make predictions on the expected amount of displacement along continuous attractors, as a function of facilitation, depression, and system size. In Section 4.4.6 we apply this to measure the performance of working memory systems and to place memory-reliability constraints on the microscopic parameters of networks supposed to implement continuous attractors, even under the influence of strong biological variability. Exploring the possibilities of the theoretical framework further, we demonstrate (Section 4.4.7) that it may be used to calculate neuronal heterogeneities necessary to implement a given steady-state distribution in the memory positions, by effectively implementing a Fokker-Planck equation on the continuum of positions. Finally, in Section 4.4.8, we show that short-term facilitation can provide a mechanism to represent and control uncertainty in the retention of memories.

4.4 Results

4.4.1 Drift and diffusion in continuous attractor models

In this section we introduce the general framework for continuous attractor models used throughout this work. We consider these to be implemented in recurrently connected neuronal networks of N neurons. The firing rate ϕ_i (in units of Hz) for each neuron i for $0 \leq i < N-1$ is given as a function of the neuronal input:

$$\phi_i(t) = F(J_i(t)) \tag{4.1}$$

The input-output relation F relates the input J_i to neuron i to its firing rate. This represents a rate-based simplification of the possibly complex underlying neuronal dynamics (Gerstner and Kistler, 2002). Neurons are recurrently connected and receive input from all other

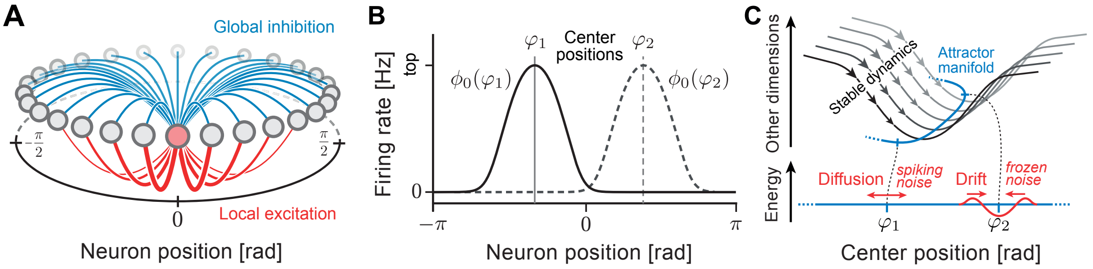

A
Global inhibition
$-\frac{\pi}{2}$
$\frac{\pi}{2}$
0
Local excitation
Neuron position [rad]
B
Firing rate [Hz]
top
0
φ_1
Center positions
φ_2
$\phi_0(\varphi_1)$
$\phi_0(\varphi_2)$
$-\pi$
0
π
Neuron position [rad]
C
Other dimensions
Energy
Stable dynamics
Attractor manifold
Diffusion
spiking noise
Drift
frozen noise
φ_1
φ_2
Center position [rad]

Figure 4.1 – **Continuous attractor models are susceptible to diffusion and drift (illustration). A:** Construction of a ring-attractor attractor network. Neurons (circles) are distributed along positions on a ring with circular boundary conditions. Excitatory connections between neurons are local, such that nearby neighbors are excited more strongly (red lines). Inhibitory connections are global (blue lines). Only outgoing connections from the neuron at position 0 (red circle) are displayed. **B:** The network dynamics converge to a unimodal firing rate profile (“bump”) $\vec{\phi}_0(\varphi)$ whose center φ can be shifted to any position on the ring. Two bumps localized at different positions φ_1 (black line) and φ_2 (dashed gray line) are shown. **C:** Illustration of the system dynamics of a 1-dimensional continuous attractor model. A one-dimensional manifold of meta-stable states exists, that correspond to the ring of possible center positions of the bump of activity (“Attractor manifold”, blue line). The remaining dimensions are stable, and perturbations from the attractor manifold along these dimensions will decay (“Stable dynamics”, black curves). Along the attractor manifold, no dynamics constrain the system, and all states have the same energy (lower plot). Small displacements are introduced through the fluctuations of neuronal firing rates (“spiking noise”) and lead to diffusion along the manifold (lower left). Fixed deviations from perfect symmetry in the system (“frozen noise”) will perturb the energy landscape (lower right, red curve) and lead to systematic drift along the attractor (lower right, red arrows).

neurons. Thus, we assume that the input $J_i(t)$ to neuron i at time t is given by the sum

$$J_i(t) = \sum_{j=0}^{N-1} w_{ij} s_j(t), \tag{4.2}$$

where w_{ij} is the connection weight for the synaptic connection from neuron j to neuron i. The variables $s_j(t)$ are the synaptic activation of connections outgoing from the presynaptic neuron j, which depend on the firing rates of the neurons j.

Without providing the details of the implementation (we give a concrete implementation in the next section), we *assume* that this network of neurons *implements a ring-attractor* (for a formal definition see Section 4.6.1.1), realized by some choice of neuronal model (Eq. (4.1)), recurrent connectivity (Eq. (4.2)) and dynamics of the synaptic activation variables s_j. A common way of implementing a ring-attractor in networks of neurons is by distance-dependent local excitation and global inhibition (Fig. 4.1A). A distance can be defined by assigning to each neuron a position $\theta_i = \frac{2\pi}{N} i - \pi \in [-\pi, \pi)$ (for $0 \le i < N$), where we identify the edges of the interval to form a ring. The recurrent connectivity between neurons is then chosen to be dependent only on the distance (along the ring of positions) between neurons Fig. 4.1, such that neurons excite their local “neighbors” and inhibit neurons that are far away (Fig. 4.1A). The resulting connectivity will have an intrinsic symmetry that gives rise to the desired dynamical properties of the system: when tuned appropriately, the network dynamics lead to the formation of a *localized and symmetric firing rate profile* (“bump”) $\phi_{0,i}(\varphi_k)$ (for $0 \leq i, k < N$), that can be centered at any given position φ_k along the ring without changing its shape (Fig. 4.1B). The network activity converges to a 1-dimensional manifold of possible center positions, which we term the “attractor manifold” (Fig. 4.1C, blue line represents a local linearization of the manifold). Linearizing the dynamical system on any point of the attractor manifold will yield $N-1$ dimensions that are constrained by the network-internal dynamics (Fig. 4.1C, black curves). These are “stable” in the sense that any deviation from the attractor manifold along these dimensions will decay (Fig. 4.1C, arrows).

The computational advantage of continuous attractors is that, along the attractor manifold, all states are dynamically equal. For the purpose of illustration, let us assume that the network dynamics are such that they minimize an energy-like function[1]. Then, the energy of all possible states along the attractor is flat (Fig. 4.1C, bottom, blue line). This metastability allows these networks to retain an approximated continuum (N different states, see Section 4.6.1.1) of stable states, realized by positioning the center of the bump of activity at any given neuron. Identifying the center with a continuous quantity (e.g. angular positions in space) allows these systems to encode and retain the quantity in principle indefinitely. However, the continuity of the stable states comes at a cost: since no dynamics constrain movement along the attractor manifold, the state encoded in the continuous attractor is vulnerable to displacement. In this work we will investigate two different types of displacement. First, fast fluctuations in the firing of single neurons transiently break the perfect symmetry of the system and introduce small random displacements along the attractor manifold, which become apparent as *diffusion* of the center position (Fig. 4.1C, lower left). Second, persistent deviations from perfect symmetry will lead to systematic *drift* along the attractor manifold, appearing as new local maxima and minima in the formerly flat energy landscape (Fig. 4.1C, lower right). As explained in the introduction, such drift can result, for example, from random changes of the recurrent weights (Zhang, 1996; Itskov et al., 2011b), sparse recurrent connectivity (Renart et al., 2003), or heterogeneity of the single neuron properties across the population (Renart et al., 2003). Since these random deviations are assumed to stay fixed (or change only very slowly) with respect to the time scale of working memory we will refer to them as “frozen” noise.

4.4.2 Ring-attractor in spiking networks with short-term plasticity

As a concrete example, we implemented a continuous ring-attractor model in a network of excitatory and inhibitory leaky integrate-and-fire models with conductance-based synaptic transmission (Fig. 4.2A, see Section 4.6.3 for details). Excitatory neurons were connected to other excitatory neurons with connection strengths depending on the distance between neurons (Fig. 4.2A, red lines). Global inhibition was implemented by inhibitory feedback from a homogeneous population of inhibitory neurons, which connect from and to all excitatory neurons uniformly (Fig. 4.2A, blue circles and lines). Additionally, all neurons receive external excitatory input with Poisson statistics. For such networks of spiking neurons, assuming that neurons fire with Poisson statistics and the network is in a stationary state, the input-output relation F can be calculated (Brunel and Wang, 2001b) (see 4.6.4), which allows us to map the network into the general framework of Eqs. (4.1)-(4.2).

---

[1] For linear synapses, e.g. with $\dot{s}_j = -\frac{s_j}{\tau_s} + \phi_j$, this Ljapunov function exists if the connectivity w_{ij} is symmetric and F is a monotonic increasing function (Hopfield, 1984).

In addition to the bump state, the spiking ring-attractor also displays a uniform state in which all excitatory neurons fire asynchronously and irregularly at low firing rates (Fig. 4.2B, left of dotted line). Activity on the attractor manifold can then be evoked by stimulating excitatory neurons localized around a given position by additional external input (Fig. 4.2B, red dots). After offset of this external cue, a self-sustained firing rate profile ("bump") emerges (Fig. 4.2B, right of dashed line, and inset) that persists until the network state is again changed by external input. For example, a short and strong uniform excitatory input to all excitatory neurons will cause a transient increase in inhibitory feedback that is strong enough to return the network to the uniform state (Compte et al., 2000).

To investigate the effects of short-term plasticity (STP) (Zucker and Regehr, 2002) on dynamics along the attractor manifold, recurrent excitatory-to-excitatory connections were subject to short-term facilitation (STF) and depression (STD) implemented by a widely used model (Tsodyks and Markram, 1997; Tsodyks et al., 1998; Mongillo et al., 2008) (see Section 4.6.3.2). Short-term plasticity modulates the effect a single spike has in excitatory-to-excitatory connections depending on the history of spikes: facilitating synapses increase their strength transiently under continued stimulation, while depressing synapses effectively weaken. Throughout this study, we varied all parameters of the short-term plasticity model (see next section). For any setting of short-term facilitation and depression parameters, networks were re-tuned (by adjusting recurrent conductance parameters and the width of distance-dependent connections, see Section 4.6.6) to display nearly identical bump shapes (similar to Fig. 4.2B, inset; see Fig. 4.4). It should be noted that synaptic depression was always present in order to saturate synaptic conductances as firing rates increase. This nonlinear behavior enables spiking networks to implement bistable attractor dynamics with relatively low firing rates (Wang, 1999; Hansel and Mato, 2013) similar to the saturating NMDA synapses used in the previous part (Section 3.6.2.2). See Section D.1 in the Appendix for a comparison of both.

In repeated simulations with initial cues provided at 20 uniformly spaced locations (Fig. 4.2C-D) we measured the trajectories of the spatial center of bump states estimated from the momentary firing rates of the excitatory population (Fig. 4.2B, turquoise line; see Section 4.6.7.1). In fully connected networks with synaptic depression and no facilitation (Fig. 4.2C, left), we find noisy displacement of the bump centers. The variance of these displacements increases linearly in time (inset in Fig. 4.2C), showing these dynamics to be

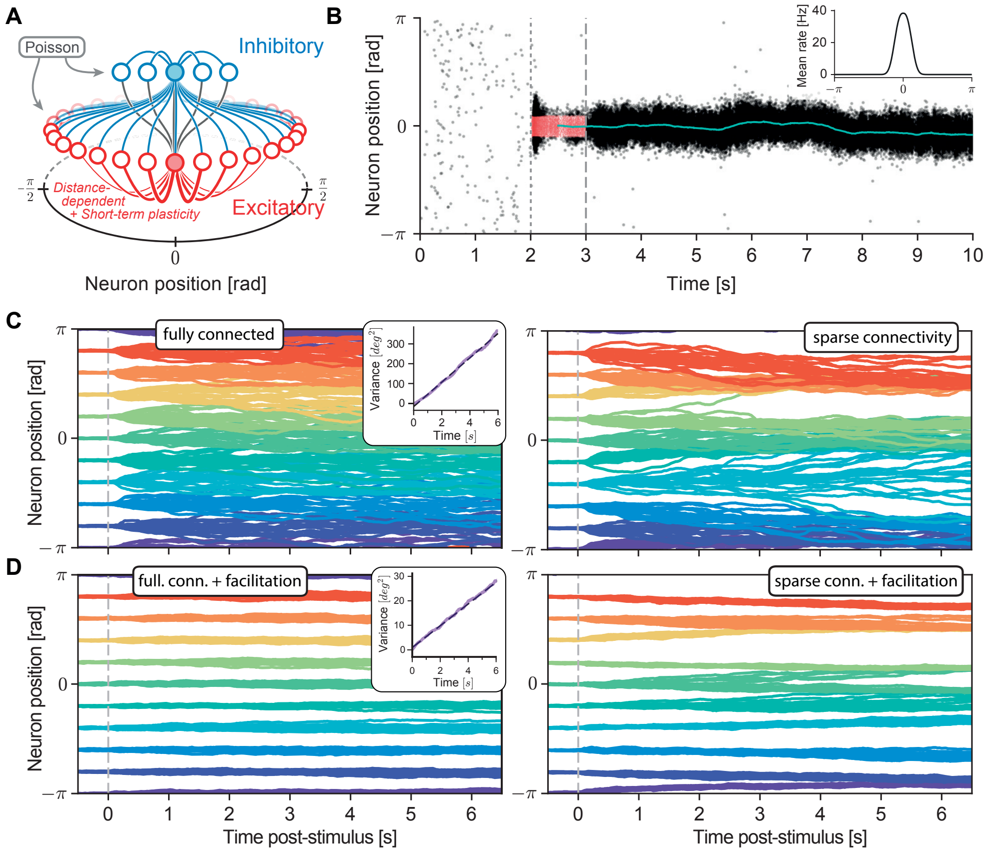


Figure 4.2 – **Drift and diffusion in a spiking ring-attractor model with short-term plasticity.** **A:** Excitatory (E) neurons (red circles) are distributed along positions on a ring with circular boundaries. Excitatory-to-excitatory (E-E) connections (red lines) are distance dependent and are subject to short-term plasticity (facilitation and depression, see Eq. (4.4)). Inhibitory (I) neurons (blue circles) project to all E and I neurons (blue lines) and receive connection from all E neurons (gray lines). Only outgoing connections from shaded circles are displayed. All neurons receive excitatory input with spikes generated by homogeneous Poisson processes. **B:** Example simulation: E neurons fire asynchronously and irregularly at low rates until a subgroup of E neurons is stimulated (external cue), causing them to spike at elevated rates (red dots, input was centered at 0, starting at $t = 2s$ lasting $1s$). During and after the stimulus (dashed line) a bump state of elevated activity forms and sustains itself after offset of the external cue. The spatial center of the population activity is estimated from the momentary firing rate (turquoise line, plotted from $t = 2.5s$ onwards). Inset: Averaged centered mean firing rates in the bump state. **C-D:** Center positions of 20 repeated simulations for 10 different initial cue positions each. Insets: Estimated variance of deviations of center positions from positions at $t = 0.5s$ (purple) as a function of time ($\langle[\varphi(t) - \varphi(0.5)]^2\rangle$), together with linear fit (dashed line). Note the changed scale of insets in C and D. **C:** A fully connected network with short-term depression ($U = 1, \tau_x = 150ms$) shows diffusive dynamics of centers (left). Sparsifying the E-E connections (with connection probability $p = 0.5$) evokes directed drift in addition to diffusion (right). **D:** Adding short-term facilitation ($U = 0.2, \tau_u = 650ms, \tau_x = 150ms$) reduces the diffusion of center positions in the fully connected network (left), and slows drift and diffusion in the sparsely connected network (right).

diffusive (Compte et al., 2000; Wu et al., 2002; Burak and Fiete, 2012; Hansel and Mato, 2013). Making excitatory-to-excitatory connections sparse (decreasing the probability of a given recurrent excitatory connection to $p = 0.5$; see Section 4.6.5) leads to fixed asymmetries that bias the center positions of bumps systematically (Fig. 4.2C, right): in addition to diffusive displacement of center positions, we observe systematic drift towards a few specific stable fixed points along the ring, designated by the realization of random connectivity. Increasing the strength of facilitation (Fig. 4.2D) leads to reduction of the magnitude of diffusion of bump centers (Fig. 4.2D, left and inset). As our theory developed in the next sections will show, facilitation also slows the directed drift of bump centers (Fig. 4.2D, right).

In summary, we introduced an implementation of a ring-attractor model in a network of leaky-integrate and fire neurons with short-term plasticity. We have seen that, in agreement with an earlier study (Pereira and Wang, 2015), short-term facilitation can reduce the diffusion along the attractor manifold. In the next section, we develop a theoretical framework that is able to quantitatively predict the effects of short-term plasticity (both depression and facilitation) on both drift and diffusion along the attractor manifold for any neuronal network given by Eqs. (4.1)-(4.2). The spiking network introduced here will be used for comparison of the theory to simulations.

4.4.3 Theoretical description of drift and diffusion in the presence of short-term plasticity

In this section we present the **central theoretical result of this chapter**: a description of the effects of short-term plasticity on drift and diffusion of the bump center on continuous attractor manifolds. The theory yields the coefficients of a one-dimensional drift-diffusion equation, that depend on short-term plasticity parameters, the shape of the metastable firing rates, as well as the neuron model chosen to implement the attractor. In the final part of this section we explain how we map the spiking network dynamics into this framework.

Under some assumptions (see below), we are able to reduce the microscopic dynamics of the continuous attractor network to a **one-dimensional Langevin equation**, describing the drift and diffusion dynamics of the center $\varphi(t)$ of the firing rate profile (see Section 4.6.1 and Section 4.6.2 in Methods):

$$\dot{\varphi} = A(\varphi) + \sqrt{B}\eta(t). \tag{4.3}$$

Here, $\eta(t)$ is white Gaussian noise with zero mean and correlation function $\langle \eta(t), \eta(t') \rangle = \delta(t - t')$. For a given neuronal implementation of the continuous attractor, we show in the Methods (Section 4.6.2) how both the ***drift field*** $A(\varphi)$ and the ***diffusion strength*** B can be explicitly calculated. Both of these depend mainly on the firing rates of the bump state maintained in the ring-attractor, the recurrent inputs to neurons in this state, the derivative of the input-output relation of neurons, and, importantly, the parameters of synaptic short-term plasticity.

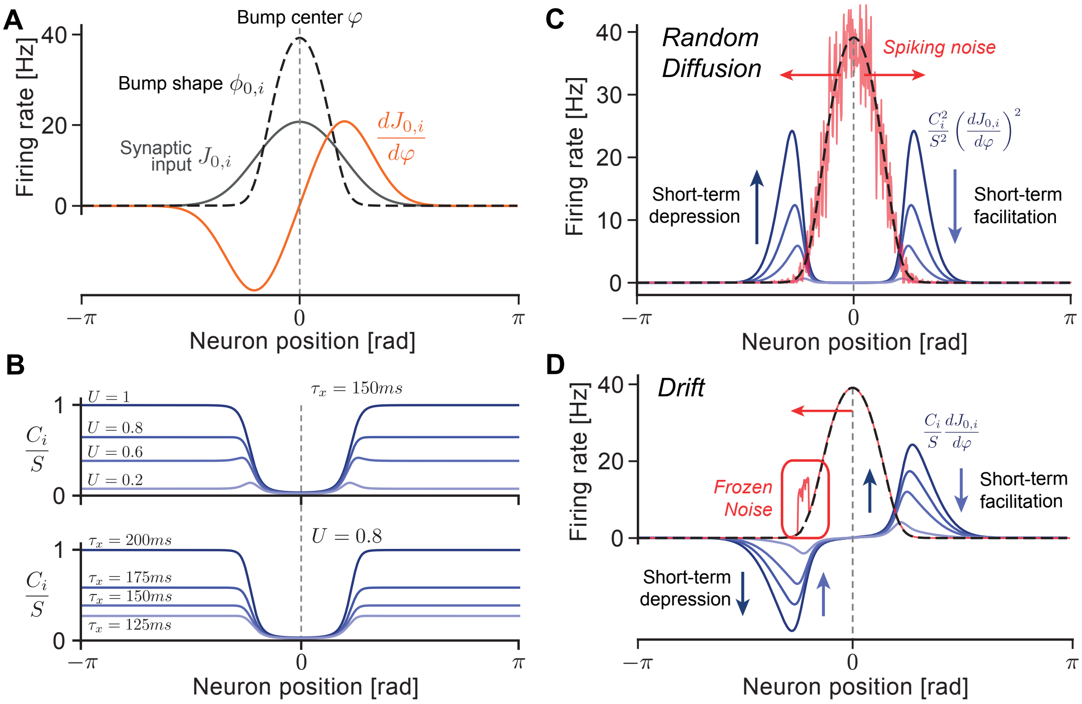


Figure 4.3 – **Calculation of diffusion and drift on the ring-attractor (illustration).** Curves are scaled and plotted together with firing rates of the bump for illustration. **A** Static quantities used in calculations. Neurons in the bump fire at rates $\phi_{0,i}$ (dashed black line) due to the steady-state synaptic input $J_{0,i}$ (gray line). Movement of the bump center causes a change of the synaptic input $\frac{dJ_{0,i}}{d\varphi}$ (orange line). **B** Short-term plasticity dependent factors $\frac{C_i}{S}$ used in calculations. Top: factors decrease with stronger facilitation, illustrated by decreasing U from $U = 1$ (dark blue) to $U = 0.1$ (light blue) for fixed $\tau_x = 150ms$. Curves are normalized to $U = 1$. Bottom: factors increase with stronger depression, illustrated by increasing τ_x from $\tau_x = 120ms$ (light blue) to $\tau_x = 200ms$ (dark blue) at fixed $U = 0.8$. Curves are normalized to $\tau_x = 200ms$. **C** The diffusion along the attractor manifold is calculated (see Eq. (4.5)) as a weighted sum of the neuronal firing rates in the bump state (dashed black line). Spiking noise (red line) is added as a random deviation from the mean rate (with standard deviation proportional to $\sqrt{rate}$) for illustration. The symmetric weighting factors (blue lines show $\frac{C_i^2}{S^2}\left(\frac{dJ_{0,i}}{d\varphi}\right)^2$ for varying U as in panel B, top) are non-zero at the flanks of the firing rate profile. Stronger (weaker) short-term depression and weaker (stronger) facilitation increases (decreases) the magnitude of weighting factors. **D** Deterministic drift is calculated as a weighted sum (see Eq. (4.7)) over systematic deviations of the neuronal firing rates from the bump state, which are introduced by frozen noise. Here, this is illustrated by a large positive firing rate deviation in the left flank (red line), which could be achieved by making these few neurons highly excitable. Since the weighting factors (blue lines show $\frac{C_i}{S}\frac{dJ_{0,i}}{d\varphi}$ for varying U as in panel B, top) are asymmetric, this deviation would cause the sum to be negative in total, resulting in movement of the center position to the left (red arrow). The weighting factors depend on short-term plasticity parameters similar to the case of diffusion (panel C).

We briefly discuss the assumptions underlying this central result, before we discuss both the drift and diffusion term in more detail below. To model synaptic transmission with short-term plasticity, we assume that the dynamics of outgoing synaptic activations s_i of neuron i are modeled by the following three dimensional system of ordinary differential equations:

$$\begin{aligned}
\dot{s}_i &= -\frac{s_i}{\tau_s} + u_i x_i \phi_i, \\
\dot{u}_i &= -\frac{u_i - U}{\tau_u} + U(1-u_i)\phi_i, \\
\dot{x}_i &= -\frac{x_i - 1}{\tau_x} - u_i x_i \phi_i.
\end{aligned} \tag{4.4}$$

Synaptic activations are given by a linear filter of the neuronal firing rates $\phi_i(t)$ multiplied by scaling factors $u_i(t)x_i(t)$. These factors modulate the synaptic transmission in an activity-dependent manner according to a rate-based version (Barak and Tsodyks, 2007; Mongillo et al., 2008; Itskov et al., 2011b) of the short-term plasticity model introduced in the spiking network (Markram et al., 1998; Tsodyks et al., 1998; Mongillo et al., 2008). It is composed of two processes. Facilitation represents the use-dependent increased utilization (u_i) of synaptic resources. As neurons fire at higher rates, more synaptic resources are utilized by the synapse which expressed as an increase of u_i from its baseline U. Depression, on the other hand, depletes the available synaptic resources (x_i), with decrease proportional to the current utilization $u_i(t)$. We further *assume* that the system of Eqs. (4.4) together with the network Eqs. (4.1)-(4.2) has a 1-dimensional manifold of meta-stable states, i.e. the network is a ring-attractor network. This is characterized by the existence of a symmetric self-sustained bump of firing rates $\phi_{0,i} = F\left(J_{0,i}\right)$ with steady-state inputs $J_{0,i}$ (for $0 \leq i < N$), that can be freely centered at any arbitrary position $\varphi \in \left\{ \frac{j}{N} 2\pi - \pi \middle| 0 \leq j < N \right\}$ under the network dynamics (see Section 4.4.1). This is illustrated in Fig. 4.3A for a bump centered at $\varphi = 0$. Our final assumption is that neuronal firing is noisy, with spike counts distributed as Poisson processes, and that we are able to replace the shot-noise of Poisson spiking by white Gaussian noise with the same mean and autocorrelation (see Section 4.6.1, and Discussion).

4.4.3.1 Diffusion strength B

We turn first to the coefficient B of Eq. (4.3), which we call the *diffusion strength*[2]. It describes the diffusive displacement of bump positions due to fluctuations in neuronal firing. In absence of $A(\varphi)$ this term causes diffusive displacement of the center φ from its initial position: the mean (over realizations) squared displacement of positions increases linearly as $\left\langle \left[\varphi(t) - \varphi(t_0)\right]^2 \right\rangle = B \cdot (t - t_0)$ (van Kampen, 1992; Gardiner, 2009; Burak and Fiete, 2009). Our theory predicts that the coefficient B of diffusion can be calculated (see Section 4.6.2.1) as a weighted sum over the neuronal firing rates (Fig. 4.3B)

---

[2] We chose this nomenclature, since Brownian motion usually defines the *diffusion constant* to be $D = B/2$.

$$B = \frac{1}{S^2} \sum_i C_i^2 \left(\frac{dJ_{0,i}}{d\varphi} \right)^2 \phi_{0,i}, \tag{4.5}$$

where $\frac{dJ_{0,i}}{d\varphi}$ is the change of the input to neuron i under shifts of the center position (Fig. 4.3A, orange line), and S is a normalizing factor that depends on short-term plasticity parameters and additionally on the derivatives $\phi'_{0,i} = \left.\frac{dF}{dJ}\right|_{J_{0,i}}$ of the neuronal input-output relation at the steady-state input $J_{0,i}$ of the bump state (cf. Eq. (4.50)). The factors C_i express the spatial dependence of the diffusion coefficient on the short-term plasticity parameters through

$$C_i = \frac{U\left(1 + 2\tau_u \phi_{0,i} + U\tau_u^2 \phi_{0,i}^2\right)}{\left(1 + U\phi_{0,i}\left(\tau_u + \tau_x\right) + U\tau_u\tau_x\phi_{0,i}^2\right)^2}. \tag{4.6}$$

Equation (4.5) shows that B is a weighted sum of the firing rates $\phi_{0,i}$ in the bump state, with weighting factors $\left(C_i/S \frac{dJ_{0,i}}{d\varphi}\right)^2$ that depend on the parameters of the synaptic dynamics (see Fig. 4.3B for an illustration). We note that the normalization factor S scales linearly with U (cf. Eq. (4.50)) and thus the linear dependences on U in C_i/S cancel. The remaining scaling is visualized in Fig. 4.3C, where we see that a) due to the squared spatial derivative $\frac{dJ_{0,i}}{d\varphi}$ of the bump shape and the squared factors C_i/S, the nonzero contributions of the sum arise primarily from the flanks of the bump; b) the weighing factors increase with stronger short-term depression (larger τ_x) and decrease with stronger short-term facilitation (smaller U, larger τ_u); cf. Fig. 4.3B. In the next sections, we analyze and compare these predictions to simulations.

Therefore, we have found an analytical expression relating the microscopic configuration of the continuous attractor network to the emerging diffusion of memory states. Our theory is based on an approach similar to Burak and Fiete (2012), where the diffusion strength B was predicted for simple synaptic dynamics in the absence of short-term plasticity. We recover the result stated there (Burak and Fiete, 2012, Eq. S18) by taking the limiting case of no facilitation (setting $U = 1$) and infinitely fast recovery from depression (setting $\tau_x = 0ms$). This simplifies the normalization factor S considerably, leaves $C_i = 1$, and yields the following simple expression for the diffusion magnitude: $B = \left(\sum_i \frac{dJ_{0,i}}{d\varphi}^2 \phi_{0,i}\right) / \left(\tau_s \sum_i \frac{dJ_{0,i}}{d\varphi}^2 \phi'_{0,i}\right)^2$.

4.4.3.2 Drift field $A(\varphi)$

The other term in Eq. (4.3) is the field $A(\varphi)$, which describes a deterministic drift depending on the current position. As discussed in the last section, these drifts appear in ring-attractor networks under the presence of systematic asymmetries (frozen noise). We derive this function to be (see Section 4.6.2.2)

$$A(\varphi) = \frac{1}{S}\sum_i C_i \frac{dJ_{0,i}}{d\varphi}\Delta\phi_i(\varphi), \tag{4.7}$$

where the terms $\Delta\phi_i(\varphi)$ are systematic firing rate deviations from the steady-state bump shape that can be calculated for a given source of frozen variability, and depend on the current position φ of the bump center. The drift field $A(\varphi)$ is calculated by equation Eq. (4.7) as a weighted sum over the firing rate deviations, with weighing factors depending on the spatial derivative of the bump shape $\frac{dJ_{0,i}}{d\varphi}$ and the parameters of the synaptic dynamics through the same factors C_i/S as for the case of diffusion. This is illustrated in Fig. 4.3D, where we see that the summands are now asymmetric with respect to the bump center, since the spatial derivative is not squared, in contrast to Eq. (4.5). **The resulting field $A(\varphi)$ predicts the directed drift of the macroscopic bump state as a function of the microscopic firing-rate perturbations introduced by frozen variability.** We find that: a) the weighting factors are again constrained to the flanks of the bump; b) the factors are scaled up and down by short-term facilitation and depression, respectively, similarly to the case of diffusion. This dependence will be analyzed and compared to simulations in Section 4.4.5.

4.4.3.3 Comparison to spiking simulations

In the previous section we introduced a spiking network implementation of a ring-attractor model, against which we will validate the theoretical predictions. This requires computing the quantities necessary to evaluate the above Eqs. (4.5) and (4.7). For the spiking network we only consider excitatory neurons that maintain the bump state in their spatial distribution of firing rates. The inhibitory population is uniformly firing and is crucial to maintain the bump state by inhibitory feedback. However, the steady-state inhibitory input to excitatory cells remains constant under shifts of the center position (since excitatory neurons are uniformly connected to all inhibitory neurons) and inhibition can thus be seen as a constant parameter. Since the resulting shape of the bump state enters our theory through the steady-state rates $\phi_{i,0}$, the inhibitory firing rate is only needed to calculate the operating point of the input-output relation F of excitatory neurons (see below).

The steady-state excitatory firing rates $\phi_{0,i}$ can be estimated in two ways. First, for a given bump-attractor network they can be measured from the spiking activity (see Section 4.6.7.1). The results presented in the next sections rely on this numerical estimation of mean firing rates, unless mentioned otherwise. On the other hand, we have presented a method for the prediction of steady-state firing rate distributions for bump-attractor networks with non-linear synaptic transmission in the previous chapter (see Section 3.4.1). The same method can be applied to networks with short-term plasticity to predict the firing rates $\phi_{0,i}$ entirely from the microscopic configuration (see Section 4.6.6). This method also yields predictions for the firing rates of inhibitory neurons. Since the theoretical predictions are very close to the numerically estimated mean-firing rates, this changes the following results only

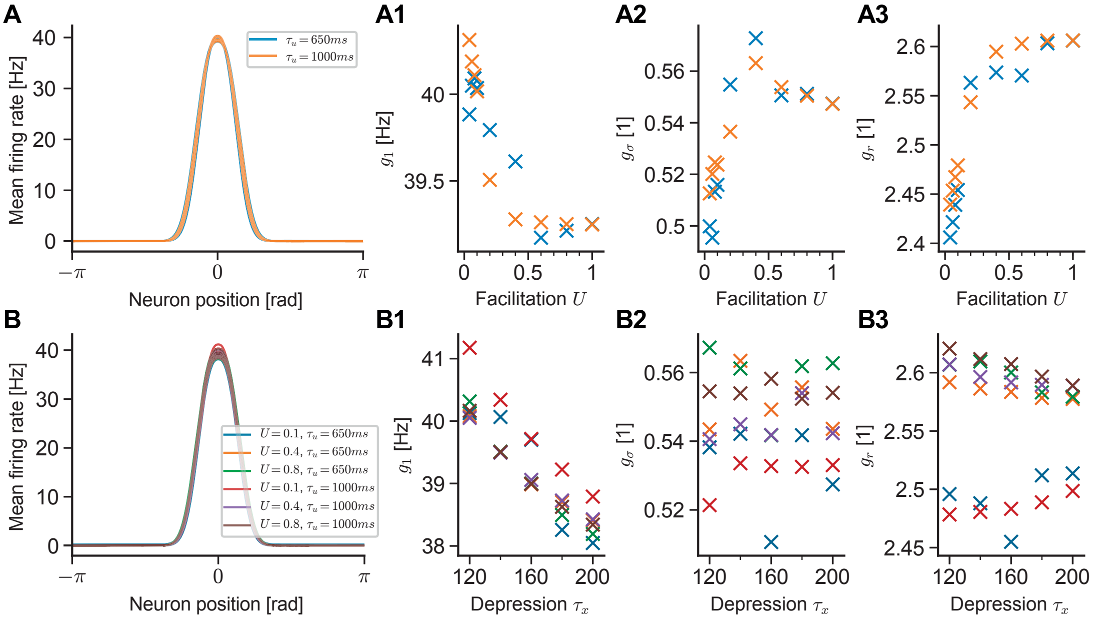


Figure 4.4 – **Stability of stable firing rate profiles across parameters.** **A** Averaged mean firing rates in bump state, measured from ~ 1000 simulations of networks tuned to display closely resembling firing rates across different values of the facilitation parameter U and the facilitation time-scale τ_u. Parameters in panels **A2-A4** are estimated by fitting the generalized Gaussian $\nu(\theta) = g_0 + g_1 \exp\left(-\left[\frac{|\theta|}{g_\sigma}\right]^{g_r}\right)$ to the measured firing rate profiles. **A1** Top firing rate g_1. **A2** Half-width parameter g_σ. **A3** Sharpness parameter g_r. **B** Same as in **A** for firing rates across different values of the depression time-scale τ_x and the facilitation time-scale τ_u.

minimally. For some figures in the following sections we will nevertheless point to Supplementary Figures that use the theoretical method of rate prediction instead of the numerical variant.

For the steady-state input $J_{0,i}$ we only consider the excitatory-to-excitatory input, which dictates the shape of the bump and changes as the center of the bump moves. The input $J_{0,i}$ is calculated by Eq. (4.2) from the recurrent connectivity w_{ij}, the steady-state firing rates $\phi_{0,i}$, and the steady state of the synaptic dynamical variables in Eq. (4.4) (see Section 4.6.4 for details). It should be noted, that $J_{0,i}$ can also be estimated from the steady-state firing rates $\phi_{0,i}$, if the input-output relation F is invertible, such that it allows an inversion of its mapping from inputs to firing rates. Thus, knowing the detailed connectivity for a neuronal network is not generally necessary to evaluate quantities predicting the amount of drift and diffusion, which could be useful for experimental applications of the theory.

The derivatives $\phi'_{0,i} = \frac{dF}{dJ}\Big|_{J_{0,i}}$ of the input-output relation at steady-state input appear in the normalization factor S in Eqs. (4.5) and (4.7). Given a steady-state recurrent excitatory input $J_{0,i}$ for each unit, as well as the firing rate of the inhibitory population, the theory of (Brunel and Wang, 2001b) can be used to derive an input-output relation F and the associated derivative – again, the details are provided in Section 4.6.4. The firing rate perturbations $\Delta\phi_{0,i}$ can be calculated for a given instantiation of frozen noise, and will be used here as a input perturbation to the input-output relation F. This is described in more detail in Section 4.4.5, as well as Section 4.6.5 in the Methods.

In summary, all quantities needed to evaluate the theoretical predictions of Eqs. (4.5) and (4.7) can be calculated, and allow us to predict the dynamics of drift and diffusion for the spiking network model. To investigate the effects of short-term plasticity on drift and diffusion on the attractor manifold, while excluding effects introduced by changing the bump shape, we tuned networks to display comparable firing rates for varying STP parameters. This was performed by an optimization procedure (see Section 4.6.6). All networks we investigated are tuned to display firing rates close to a generalized Gaussian of the shape $\nu(\theta) = g_0 + g_1 \exp\left(-\left[\frac{|\theta|}{g_\sigma}\right]^{g_r}\right)$ with $g_0 = 0$, $g_1 \approx 40Hz$, $g_\sigma \approx 0.5$, $g_r \approx 2.5$. The stability of firing rates across STP parameters was confirmed by comparing mean firing rates for all networks (Fig. 4.4). Throughout the next sections, we additionally validate this tuning by extrapolating the theory: using coefficients in our theory that are extracted from a “control network” without facilitation ($U = 1, \tau_x = 150ms$) and changing only the short-term plasticity parameters yields consistent predictions of drift and diffusion across all other networks.

4.4.4 Effects of short-term plasticity on diffusion

To investigate the effects that short-term plasticity (STP) has on diffusion, we simulated a ring-attractor implemented in a spiking network (see Section 4.4.2) with varying STP parameters. For each combination of parameters, we simulated 1000 repetitions of $13.5s$ delay activity (after cue offset) distributed across 20 uniformly spaced initial cue positions.

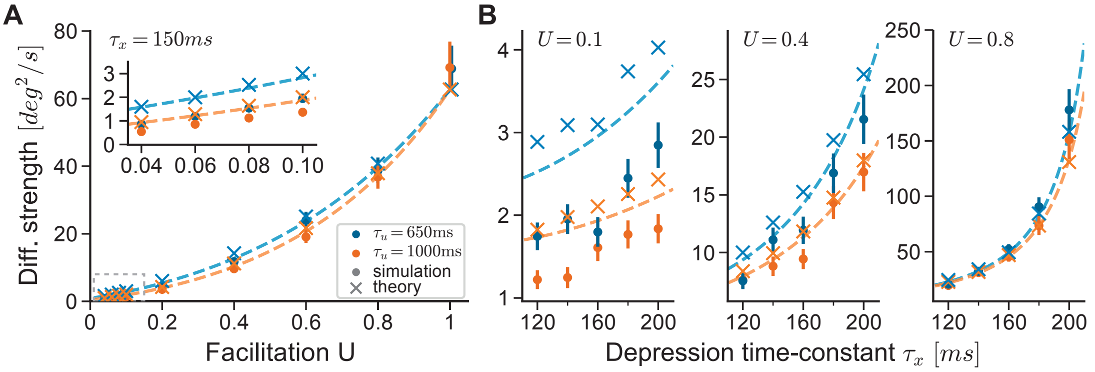


Figure 4.5 – **Effects of short-term plasticity on diffusion.** Diffusion strengths estimated from simulations (dots, error bars show 95% confidence interval, estimated by bootstrapping) compared to those predicted by theory with firing rates measured from simulations (crosses). Dashed lines show theoretical prediction using only firing rates measured from the baseline network ($U = 1, \tau_x = 150ms$). Results for two different facilitation time constants $\tau_u = 650ms$ (blue) and $\tau_u = 1s$ (orange) are displayed. **A:** Diffusion strength as a function of facilitation parameter U. Inset shows zoom of region indicated in the dashed area in the lower left. Depression time constant is $\tau_x = 150ms$. **B:** Diffusion strength as a function of depression time constant τ_x. Results for three different values of U are shown, note the change in scale.

From these simulations, the magnitude of diffusion was estimated by measuring the growth of variance (over repetitions) of the distance of the center position from its initial position as a function of time (see Section 4.6.7.1). For all parameters considered, this growth was well fit by a linear function (e.g. Fig. 4.2C,D, insets), the slope of which we compare to the theoretical prediction obtained from the diffusion strength B (Eq. (4.5)).

We find that both facilitation and depression can significantly influence the amount of diffusion along the attractor manifold (see Fig. 4.5), in an antagonistic fashion. First, decreasing the facilitation parameter U from its baseline $U = 1$ (no facilitation) towards the minimal value $U = 0$, while keeping the depression time constant $\tau_x = 150ms$ fixed, the measured diffusion strength decreases over an order of magnitude (Fig. 4.5A, dots). On the other hand, increasing the facilitation time constant τ_u from $\tau_u = 650ms$ to $\tau_u = 1000ms$ (Fig. 4.5A, orange and blue dots, respectively) only slightly reduces diffusion. Second, we find that increasing the depression time constant τ_x for fixed U, which increases the amount of depression, leads to an increase of the measured diffusion (Fig. 4.5B). Depending on the level of facilitation, increasing the depression time constant from $\tau_x = 120ms$ to $\tau_x = 200ms$ leads only to slight increases in diffusion for strong facilitation ($U = 0.1$), and up to an increase of factor 10 for weak facilitation ($U = 0.8$).

Our theory is able to closely predict this result, which we confirm in a twofold way. First, we estimate the diffusion strength by using the shape of the stable firing rate profile extracted separately for each network with different sets of parameters (Fig. 4.5, crosses). This confirms that the theory closely describes the dependence of diffusion on short-term plasticity for each

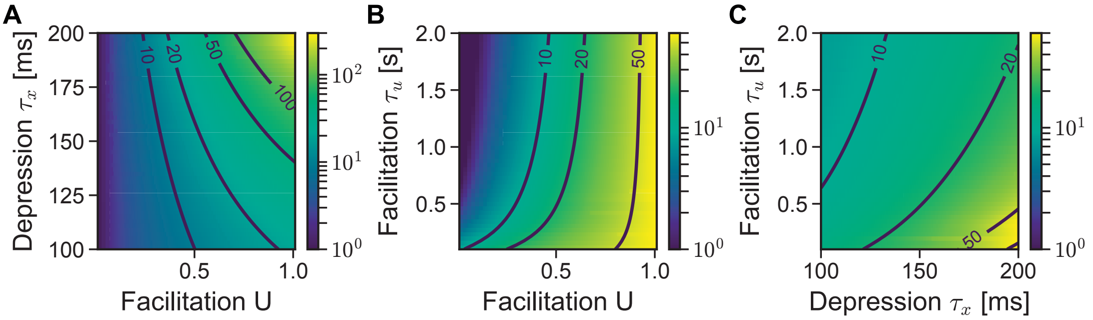


Figure 4.6 – **Diffusion magnitude as function of facilitation and depression parameters.** All color values display diffusion magnitude estimated from B in Eq. (4.3) with bump shape estimated from simulations. Units of color values are $\frac{\mathrm{idx}^2}{s}$. **A** Diffusion as function of facilitation U and depression time constant τ_x. Facilitation time constant was $\tau_u = 650ms$. **B** Diffusion as function of facilitation U and facilitation time constant τ_u. Depression time constant was $\tau_x = 150ms$. **C** Diffusion as function of depression time constant τ_x and facilitation time constant τ_u. Facilitation U was $U = 0.5$.

parameter set, while effects of possibly changed steady-state firing rates (albeit small) can not be excluded. To control for this, we confirmed this result by using the measured bump shape of the “control network” ($U = 1$, $\tau_x = 150ms$) in Eq. (4.5), then extrapolating by changing the parameters of short-term plasticity. This leads to very similar predictions (Fig. 4.5, dashed lines), and validates our results: a) the diffusion to be expected in attractor networks with similar observable quantities (mainly, the bump shape) depends only on the short-term plasticity parameters; b) the family of networks we have investigated are sufficiently similar in these quantities to be accurately predicted by the control network only.

In summary, the strength of diffusion that a ring-attractor model experiences can be controlled by facilitation and depression in an antagonistic fashion, where increased synaptic facilitation parameters U decrease the diffusion strength, and increased synaptic depression time constants τ_x increase the diffusion strength. In the region of parameters that we compared to simulation results, the facilitation time constant τ_u had only a minimal effect on the amount of diffusion expected in the system. In Fig. 4.6 we show plots exploring theoretical predictions for a larger range of parameters for τ_u – the analysis there predicts that increasing the facilitation time constants above $\tau_u = 1s$ will not lead to large reductions in the magnitude of diffusion. It should be noted that the theory tends to slightly over-estimate the diffusion (see Discussion). However, given the simplifying assumptions needed to derive the theory, the quantitative match to the spiking network is surprisingly accurate.

4.4.5 Effects of short-term plasticity on drift

Next to random diffusion of bump centers due to the variability of neuronal spiking, systematic drifts will appear in continuous attractor networks due to frozen variability. As we have argued (Section 4.4.1) and demonstrated (Section 4.4.2), frozen variability will cause fixed deviations from the perfect symmetry required to keep a manifold of possible bump center positions meta-stable. In this section, we will introduce frozen noise of varying magnitude from two different sources of variability into spiking networks, while varying short-term plasticity parameters. As in the previous sections, we will compare these to theoretical predictions obtained from Eq. (4.7) in Section 4.4.3.

In cortical networks, neurons are rarely connected to all other neurons (Braitenberg and Schütz, 1991; Wang et al., 2006). As a first source of variability we thus consider random and sparse connectivity. We realize this by randomly removing a certain percentage of excitatory-to-excitatory (EE) connections, while rescaling recurrent conductances to keep the same mean input (see Section 4.6.5). In the resulting networks, excitatory neurons are connected on average to a fraction p of all other neurons in the excitatory population, where $0 < p \leq 1$ is the connection probability. The remaining uniform connections (involving inhibitory neurons) are left unmodified, since the effects of making these sparse would have no immediate effect on the dynamics of the bump center positions.

A second source of frozen variability we consider is heterogeneity of the neuronal parameters in the excitatory population. Electrophysiological assays of principal cells in prefrontal cortex show considerable variability of the leak reversal potential V_L (Renart et al., 2003): standard deviations for regular spiking cells have been reported to 1.7mV in-vitro (Yang et al., 1996), and between 2.4mV and 5.3mV, depending on cell type, in a more recent in-vivo study (Dégenètais et al., 2002). We realize this biological variability in our model by making the leak-reversal potentials of the excitatory population randomly distributed: leak reversal potentials of excitatory neurons are given by $V_L + \Delta_L$, where Δ_L is normally distributed with zero mean and standard deviation σ_L.

The theory we have derived in Section 4.4.3 allows predicting the drift-fields $A(\varphi)$, given perturbations to firing rates $\Delta\phi_i(\varphi)$ induced by frozen noise. Briefly, we treat each realization of an arbitrary heterogeneity as small perturbations Δ_i around the perfectly symmetric system, and use an expansion to first order of the input-output relation F to calculate the resulting changes in firing rates (see Section 4.6.5 for details):

$$\Delta\phi_i(\varphi) = \frac{dF}{d\Delta_i}\Delta_i.$$

The resulting terms are used in Eq. (4.7) to predict the magnitude of the drift field $A(\varphi)$ for any center position φ.

In the following we focus on the short-term plasticity parameters U and τ_x when comparing our theoretical results to simulations. However, the facilitation time scale τ_u also affects the

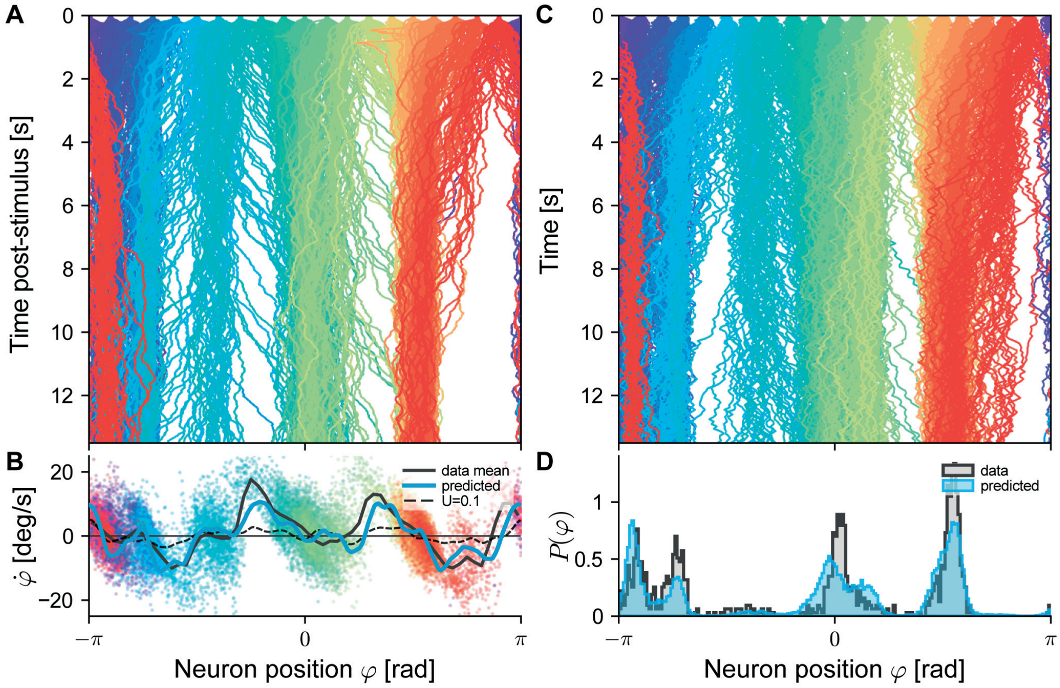


Figure 4.7 – **Prediction of drift-field in spiking network without facilitation.** The network considered has sparse connections with $p = 0.5$. **A** Centers of excitatory population activity for 50 repetitions of 13.5s delay activity, over 20 different positions of initial cues (cue offset is at time $t = 0$) colored by position of the cues. **B** Drift field as a function of the bump position. The theoretical prediction (blue line, see Eq. (4.7)) of the drift field is compared to velocity estimations along the trajectories shown in A, colored by the line they were estimated from. The thick black line shows the binned mean of data points in 60 bins. For comparison, the predicted drift field for $U = 0.1$ is plotted (thin dashed line). **C** The same trajectories, obtained by forward-integrating the one-dimensional Langevin equation Eq. (4.3). **D** Histograms of final positions at time $t = 13.5$ for data from spiking simulations (gray areas, data from A) and forward solutions of the Langevin equations (blue areas, data from B). Histograms (200 bins) are normalized to probability distributions over $[-\pi, \pi)$. STP parameters were $U = 1, \tau_u = 650ms, \tau_x = 150ms$.

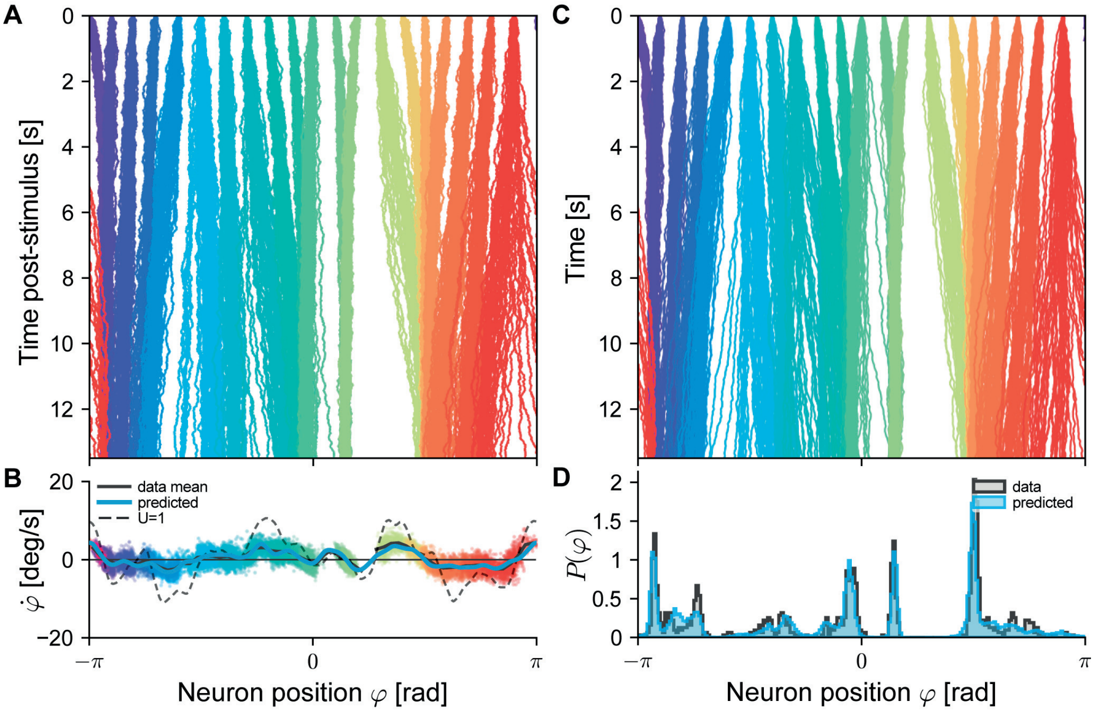


Figure 4.8 – **Prediction of drift-field in spiking network with facilitation.** Same realization of the sparse network ($p = 0.5$) as in Fig. 4.7 with strong facilitation ($U = 0.1$). Other STP parameters were $\tau_u = 650ms, \tau_x = 150ms$. See Fig. 4.7 for legend.

drift-fields. While we did not perform extensive comparisons of the effects of τ_u to simulations, we give examples and discuss these briefly at the end of the next section. Furthermore, in Section 4.4.6, we analyze theoretically the effects of all short-term plasticity parameters on the combined effects of drift and diffusion.

In the following sections, we will compare simulations of the spiking network with these two types of heterogeneities to theoretical predictions. First, we will provide an illustrative example of this comparison for a single realization of sparse connectivity (Section 4.4.5.1). We then develop the theory further, allowing us to predict the expected magnitude of the resulting drift fields directly from the noise parameters, and compare this result to simulations (Section 4.4.5.2). Finally, we evaluate the accuracy of theoretical predictions of single drift fields, under varying facilitation parameters and parameters of frozen noise (Section 4.4.5.1).

4.4.5.1 Example: Prediction of drift by theory

To illustrate the approach, we simulated spiking networks with one fixed instantiation of sparse connectivity ($p = 0.5$), while varying the facilitation parameter U. In a network without short-term facilitation (Fig. 4.7A) the missing connections introduce non-negligible systematic displacements of the center positions. Numerically extracting the magnitude of

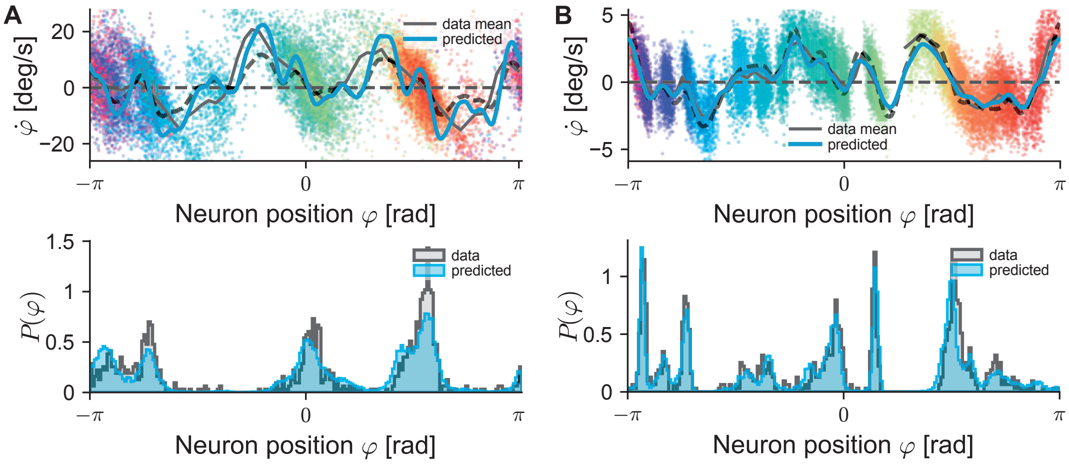


Figure 4.9 – **Prediction of drift in spiking network with varying STP parameters.** Sparse network ($p = 0.5$) with noise as in Fig. 4.7 (see panels B,D there for legends). **A** Drift field (top) and distribution of final positions (bottom) for network with increased depression time constant ($U = 0.8, \tau_x = 200ms, \tau_u = 650ms$). Dashed line (top plot) indicates drift-field prediction for $\tau_x = 150ms$. **B** Same as in panel A, for network with increased facilitation time constant ($U = 0.1, \tau_x = 150ms, \tau_u = 1s$). Dashed line (top plot) indicates drift-field prediction for $\tau_u = 650ms$.

displacements (see Section 4.6.7.1 for details) yields a drift-field in good agreement with the theoretical prediction (Fig. 4.7B). At points where the drift field prediction crosses from positive to negative values we expect stable fixed points of the center position dynamics – these are always in close match with simulation results, where we see trajectories converging to these points. Similarly, unstable fixed points (negative-to-positive crossings) can be seen to lead to a diffusive separation of trajectories converging to either side (e.g. Fig. 4.7A, around $\varphi = -\frac{\pi}{2}$). In regions where the positional drifts are predicted to lie close to zero (e.g. Fig. 4.7A, around $\varphi = 0$) the effects of diffusive dynamics are more pronounced. Since both drift and diffusion in this system are closely predicted by our theory, we proceed by simplifying the full network dynamics to only the 1-dimensional Langevin equation Eq. (4.3) with coefficients predicted by theory, and sampled trajectories by forward integration (see Section 4.6.7.3 for details). As expected, this yields quite very comparable dynamics (Fig. 4.7C). Finally, when comparing the center positions after $13.5s$ of delay activity between the full spiking simulation and the simple 1-dimensional Langevin system, we find similar distributions (Fig. 4.7D).

Including strong short-term facilitation ($U = 0.1$) reduces the measured as well as the predicted drift fields by nearly an order of magnitude (Fig. 4.8A,B). The resulting drift-field resembles a scaled-down version of the drift-field for the unfacilitated case (Fig. 4.8B, dashed line). As before, forward integrating the associated Langevin-dynamics yields trajectories similar to those of the full spiking network (Fig. 4.8C). Analogously, distributions of final positions are in close match (Fig. 4.8D).

In summary, for a single network with sparse connectivity ($p = 0.5$) we found our theoretical predictions to be in close agreement with simulation results obtained from a spiking network of neurons. Trajectories obtained from the Langevin-equations resemble closely those of the spiking network. Thus, reducing the full system to a position-dependent drift and a general diffusion coefficient seems to capture the most prominent system dynamics. Both theory and simulation show that introducing facilitation reduces the magnitude of the drift-field significantly (in agreement with Itskov et al. (2011b)).

As noted above, the other short-term plasticity parameters also have effects on the drift-field. We simulated the same network for varying time constants of short-term depression τ_x and facilitation τ_u. As was the case for diffusion, increasing (decreasing) τ_x yielded increased (decreased) drift fields, which also matched well with simulation results (see Fig. 4.9A, for $U = 0.8$). Similar to the diffusion case, this effect vanished as U was decreased. Increasing the facilitation time constant τ_u from $650ms$ to $1s$ did yield a very small decrease in predicted and measured drift field magnitude (see Fig. 4.9B, for $U = 0.1$). In the next section we investigate the effect of the time scale τ_x on drift fields more closely. For further results see Section 4.4.6, where we report the effects of all short-term plasticity parameters on drift and diffusion.

4.4.5.2 Drift as a function of noise parameters

The theory developed so far allows us to predict drift-fields from given realizations of frozen noise, which are controlled by parameters p (for sparse connectivity) and σ_L (for heterogeneous leak reversal-potentials). Here, we generalize this approach by calculating the expected magnitude of drift fields under the distributions of these parameters and compare the resulting theory to simulations.

When averaged over the distributions that govern the random variables of both frozen noise sources (connections and reversal potential), the mean drift fields for any given center position vanish (see Section 4.6.5). However, the squared drift field magnitude does not vanish when averaged over the distributions of the frozen noises, and evaluates to (see Section 4.6.5.3 for the derivation):

$$\langle A^2 \rangle_{\text{frozen}} = \frac{1}{S^2} \sum_i C_i^2 \left(\frac{\left(\phi'_{0,i}\right)^2}{N_E^2} \left(\frac{1}{p} - 1\right) \sum_j \left(s_{0,j}\right)^2 \left(w_{ij}^{\text{EE}}\right)^2 + \left(\frac{d\phi_{0,i}}{d\Delta_i^L}\right)^2 \sigma_L^2 \right), \tag{4.8}$$

where $s_{0,j}$ is the steady-state synaptic activation. Additionally, we have introduced the derivatives of the input-output relation with respect to the noise sources: $\phi'_{0,i} = \frac{dF}{dJ}(J_{0,i}(\varphi))$ is the derivative with respect to the steady state synaptic input, and $\frac{d\phi_{0,i}}{d\Delta_i^L}$ is the derivative with respect to the perturbation in the leak potential. In Section 4.6.5.3 we additionally show that Eq. (4.8) is independent of the center position φ, and thus equally describes the expected squared drift field for all center positions. Using this, we estimate this quantity from simulations as the variance of the drift field across positions, which then averaged over the

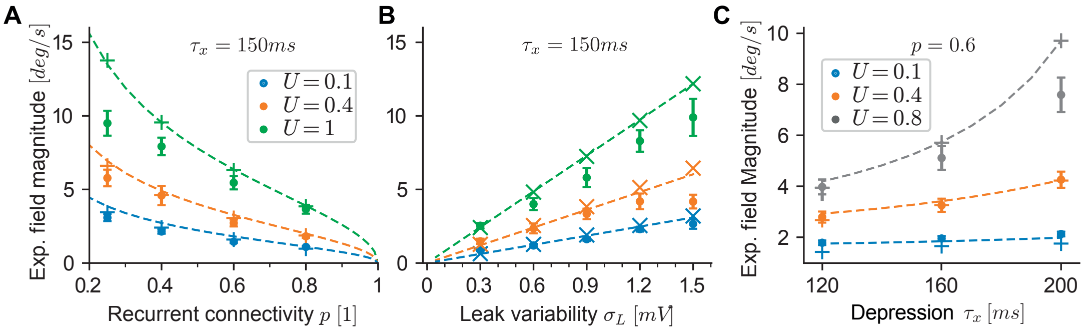


Figure 4.10 – **Prediction of expected magnitude of drift fields from noise parameters.** **A** Expected magnitude of drift fields as a function of the sparsity parameter p of recurrent excitatory-to-excitatory connections. Dots are the standard deviation of fields estimated from 400 trajectories (see main text) of each network, averaged over 18 – 20 realizations for each noise parameter and facilitation setting (error bars show 95% confidence of the mean). Theoretical predictions (dashed lines) are given by Eq. (4.9) extrapolated from the unfacilitated network ($U = 1, \tau_u = 650, \tau_x = 150$). For validation, we also estimated Eq. (4.9) with coefficients measured from each simulated network separately (plus signs). **B** Same as in panel A, with heterogeneous leak-reversal potentials as the source of frozen noise. Validation predictions are plotted as crosses. **C** Same as in panels A,B but varying the depression time constant τ_x for a fixed level of frozen noise (sparse connectivity, $p = 0.6$).

ensemble of instantiations (see Section 4.6.5.3).

In the following we will often use the root of the expected squared magnitude of Eq. (4.8):

$$\sqrt{\langle A^2 \rangle_{\text{frozen}}}, \tag{4.9}$$

which we term the *expected field magnitude*. This theoretical quantity predicts the magnitude of the deviations of drift-fields from zero that are expected from the parameters that control the frozen noise – in analogy to the standard deviation for random variables, this quantity predicts the standard deviation of the fields.

We proceeded to check the theoretical result, Eq. (4.9), against simulations. For sparse connectivity we varied the sparseness parameter p between 0.25 and 1. For heterogeneities in the leak reversal-potentials, we chose values for the standard deviation σ_L of leak-reversal potentials between $0mV$ and $1.5mV$, which lead to a similar range of drift magnitudes as the sparse networks. For simulations varying the facilitation parameter U only, we chose 3 networks with different levels of facilitation: an unfacilitated network ($U = 1$), a network with intermediate levels of facilitation ($U = 0.4$) and a strongly facilitating network ($U = 0.1$). For each combination of noise and network parameters (networks had either sparse connections or heterogeneous leaks) we realized 18 – 20 networks, and simulated for each of these 400 repetitions of $6.5s$ of delay activity, which were split across 20 uniformly spaced positions of the initial cue. We additionally investigated the effect of the depression time constant τ_x for

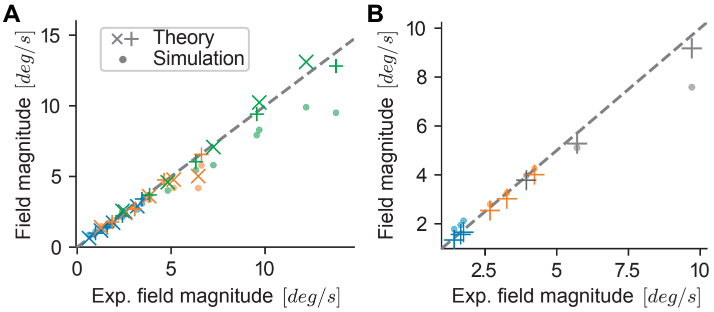


Figure 4.11 – **Self-consistency of squared magnitude predictions.** Expected magnitude of drift fields (Eq. (4.9), coefficients estimated for each simulated network separately, x-axis) plotted against averaged (18-20 repetitions, error bars are omitted to reduce visual clutter) standard deviation of fields, calculated from either: a) fields predicted by Eq. (4.7) (plus signs for sparse connectivity, crosses for variable leak-reversal potentials); b) fields estimated from simulations (dots). **A** Same data as in Fig. 4.10A and B. **B** Same data as in Fig. 4.10C.

sparse networks with $p = 0.6$, by simulating networks with combinations of short-term plasticity parameters from $U \in \{0.1, 0.4, 0.8\}$ and $\tau_x \in \{120ms, 160ms, 200ms\}$. This large set of simulated data is analyzed in the following.

For each set of repetitions, we estimated the drift-field numerically by recording displacements of bump centers along their trajectories (as in Fig. 4.7B and Fig. 4.8B, see Section 4.6.7.1 for details) and measuring the standard deviation of the resulting fields across all positions. These data are plotted as circles in Fig. 4.10A for sparse connectivity, and Fig. 4.10B for heterogeneous leak potentials. We then used Eq. (4.9) to predict the same quantity by theory. First, we used Eq. (4.9) with all parameters and coefficients estimated from each spiking simulation separately (Fig. 4.10A,B, crosses and plus signs). For validation, we then extrapolated the theoretical prediction by using coefficients in Eq. (4.8) measured from the unfacilitated network only ($U = 1, \tau_u = 650ms, \tau_x = 150ms$), then changing the facilitation and noise parameters (Fig. 4.10A,B, dashed lines).

We observe that the two theoretical predictions (crosses and plus signs, dashed lines) coincide well across all levels of frozen noise and networks. Similar to the comparison in Section 4.4.4, this shows that our network tuning across facilitation parameters is consistent. Only for $U < 1$ and sparse networks are theoretical predictions obtained from single networks slightly below the extrapolated theory, which we found resulted from the fact that bump shapes for these networks tended to be slightly reduced under sparse connectivity (e.g. the top firing rate is reduced to $\sim 35Hz$ for $U = 0.1, p = 0.25$). When compared to the simulation results (Fig. 4.10A,B, circles), the predictions are generally in good agreement. For both sources of frozen noise, increasing the strength of short-term facilitation (by decreasing U) has the effect of scaling down the magnitude of drift fields. As noise levels increase, the theory tends to over-estimate the squared magnitude of fields. Our theory relies on a linear expansion of perturbations to the firing rates (see Section 4.6.5) resulting from frozen noise, which are then scaled by the effects of short-term plasticity. Thus, we expect these over-estimations to become apparent across facilitation parameters, which is in agreement with the deviations observed here. Comparing the same data for a fixed noise level ($p = 0.6$) while varying the depression time constant τ_x (Fig. 4.10C), we observe that increasing (decreasing) the depression time constant also leads to an increase (decrease) in drift magnitude, especially for values of $U \gg 0$.

As a final self-consistency check (Fig. 4.11), we calculated the standard deviation of fields estimated by Eq. (4.7), where coefficients were obtained for each simulation separately. In Fig. 4.11 we compare these data (crosses and plus signs) to the expected field magnitude obtained from Eq. (4.9) with the same coefficients (x-axis). In addition, we plot the standard deviation of fields estimated from simulations against the same x-axis (dots). This analysis shows that for small noise parameters (point to the left in each color), the cases where the simulations deviate from the theoretical predictions of Eq. (4.9) (off-diagonal dots) are still well predicted by the theory of Eq. (4.7) (since crosses and dots coincide). Thus, deviations could stem from the small sample size ($n = 18 - 20$). On the other hand, for larger noise amplitudes (points to the right for each color) theoretical predictions again lie systematically above the simulation results.

In summary, we have derived theoretical predictions for the expected magnitude of drift fields depending only on the parameters of the underlying distributions of frozen noise Eq. (4.9)**. We have compared these predictions against simulation results and found them to be in good agreement.** Across both sources of frozen noise and the short-term depression parameters investigated here, we have seen that short-term facilitation can significantly reduce the magnitude of drift fields, leading to slower drift dynamics. Increased depression, on the other hand, leads to an increase in the magnitude of drift fields. The theoretical prediction Eq. (4.9) can, in principle, also be used to describe the effects of the remaining parameter of short-term facilitation: the facilitation time constant τ_u. While we have not extensively validated the results against simulations, an example for the small effect of τ_u was given above, in Fig. 4.9. In Section 4.4.6 we additionally report the effects of all short-term plasticity parameters on drift and diffusion.

4.4.5.3 Evaluation of single predictions

In the last section, we confirmed that the generalized approach of Eq. (4.9) can successfully predict the expected magnitude of fields as a function of parameters of the frozen noise distributions. Here, stepping down one level of detail, we compared how well single drift fields were predicted by the theory of Eq. (4.7). The results shown here are based on the simulated data set of the previous section for varying facilitation (Fig. 4.10A, B).

First, we compared each drift field extracted from simulations (see Section 4.6.7.1 for details) to the corresponding field predicted by Eq. (4.7), by measuring the root mean square error (RMSE) of the prediction to the 100 bins estimated from simulations (Fig. 4.12A). This shows

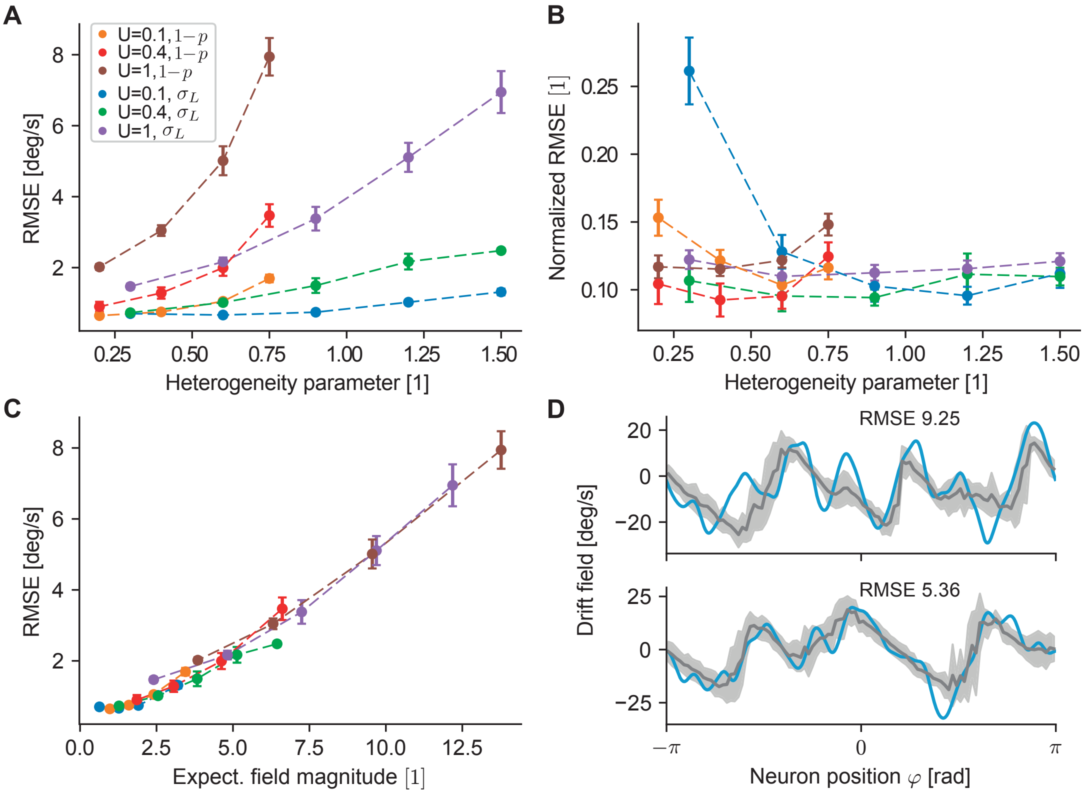


Figure 4.12 – **Comparison of theoretically predicted fields to simulations. A** Averaged root mean square error (RMSE) between predicted fields (Eq. (4.7)) and fields extracted from simulations (mean over 18-20 networks, error bars show 95% confidence of the mean). Both frozen noise parameters (σ_L and $1-p$) are plotted on the same x-axis. **B** Normalized RMSE: each RMSE is normalized by the range (max – min) of the joint data of simulated and predicted fields it is calculated on. Colors as in A. **C** Average RMSE (same data as in A) plotted as a function of the mean expected field magnitude (estimated separately for each network, then averaged). Colors as in A. **D** Worst (top) and best (bottom) match between predicted field (blue line) and field extracted from simulations (black line) of the group with the largest mean RMSE in panels A, C ($U = 1, 1-p = 0.75$). Shaded areas show 1 standard deviation of points included in the binned mean estimate (100 bins) of the extracted field.

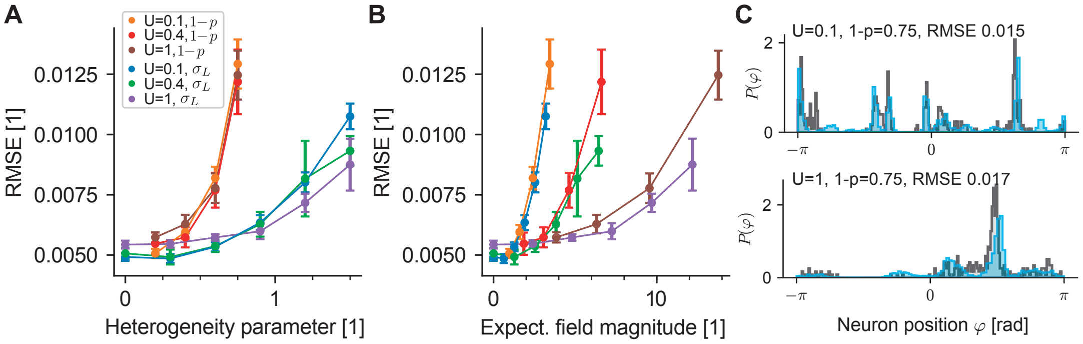


Figure 4.13 – **Comparison of theoretical predictions and simulations for distributions of final positions.** **A** Averaged root mean square error (RMSE) between histograms of final positions (100 bins) estimated from simulations (20 repetitions for 20 initial cue positions each) and those predicted from theory (200 repetitions for 20 initial cue positions each). Dots show mean over 20 networks, error bars show 95% confidence of the mean. Both frozen noise parameters (σ_L and $1-p$) are plotted on the same x-axis. **B** Average RMSE (same data as in A) plotted as a function of the mean expected field magnitude (estimated separately for each network, then averaged). Colors as in A. **C** Worst matches (largest RMSE) between predicted distribution of final positions (blue) and simulated distribution of final positions (black) in the two groups with the largest mean RMSE: top shows $U = 0.1, 1-p = 0.75$; bottom shows $U = 1, 1-p = 0.75$.

that, as heterogeneity parameters increase, predicted drift fields tend to deviate from those estimated from simulations. This is in agreement with the more general analysis of Fig. 4.10, where we showed that fields tend to be over-predicted as the expected magnitude increases. Here, we find that this error is proportional to the magnitude of drift fields: the normalized RMSE (each RMSE is normalized by the range (max − min) of the data it is calculated on) stays fairly constant across networks (Fig. 4.12B). This also becomes obvious when plotting the RMSE against the theoretically predicted expected field magnitude (Eq. (4.9)), which collapses all data onto a single, nearly linear relation (Fig. 4.12C). We stress that, even if the fields tend to be less well estimated as their magnitude increases, the theory still manages to predict fields well. Examples are given in Fig. 4.12D, where we plot the worst (top) and best (bottom) matches (as measured by the RMSE) between predictions and simulations in the group with the highest average RMSE ($U = 1, 1-p = 0.75$). In both cases, stable fixed points are well captured, although the worst match produces more zero crossings than simulated data. Again, since our theory is linear in the perturbations caused by frozen noise, we do expect such deviations to appear as the magnitude of perturbations increases. Finally, we note that these results are stable under changes of bin numbers used to estimate and compare the drift fields (data not shown).

As a second check, we investigated how well the combined prediction of drift fields and diffusion strengths was able to describe the evolution of trajectories in the spiking network. For this, we compared the distributions of final positions after 6.5s of delay period between theory and simulations. For each network considered, we recorded the final positions from simulations, calculated the coefficients of the Langevin equation Eq. (4.3), and used these to generate 10 trajectories for each simulated one, over the same delay time. We then calculated the RMSE between normalized histograms of both sets of final positions (100 bins, normalized to probability distributions on $[-\pi, \pi)$) (as in Fig. 4.7D). As both heterogeneity parameters increase, the distance between both distributions increases (Fig. 4.13A) quite similarly for all 3 networks, regardless of the values of short-term facilitation. For completeness, we also plotted the dependence of the RMSE on the expected magnitude of the drift-fields (Fig. 4.13B), which collapses both sources of heterogeneity onto single curves for each facilitation parameter U. Again, we note that even for the larger RMSE values, the predicted distributions of positions match quite well. Examples are given in Fig. 4.13C for the worst matches (as measured by the RMSE) in the two groups with the highest average RMSE ($U = 0.1, 1 - p = 0.75$ and $U = 1, 1 - p = 0.75$). Again, the choice of bin numbers did not influence this result. The number of trajectories generated from the Langevin equations for comparison does influence the absolute values of the RMSE, however, the general properties described here are robust. Finally, we would like to note estimating the distance between the two distributions by a second measure commonly used for distances between distributions (Bhattacharyya coefficient, Bhattacharyya (1943)) yielded different absolute values, but similar dependencies on facilitation and noise parameters.

4.4.6 Application: Effects of short-term plasticity on working memory function

Continuous attractor models have been proposed as models for visuospatial short-term memory in pre-frontal cortex (Compte et al., 2000; Constantinidis and Wang, 2004; Compte, 2006; Wimmer et al., 2014). In this context, the continuous attractor network is interpreted as a (noisy) cortical “storage device”: a visual cue located at some angle in the visual field evokes a properly located input in pre-frontal cortex, that encodes the position which is maintained by the network during the delay period. After the delay, a go signal is given, upon which the position is read out from the storage device and propagated to motor areas that effectuate a saccade of the eye to this position. In the previous sections, we have developed theory that allows the prediction of diffusion (Section 4.4.4) and directed drift (Section 4.4.5) of continuous attractor networks in dependence of short-term plasticity parameters and parameters of the frozen noise. So far we have mostly considered both of these separately. The ability of continuous attractor networks to reliably store continuous quantities over delay periods will, however, be influenced by both random diffusion and directed drift of center positions. Diffusion degrades the memories over time, in that center positions deviate from their initially cued positions with variance of deviations growing linear in time. Directed drift, on the other hand, leads to collapse of the continuum of fixed points to a few stable fixed points.

In this section, we will present approaches to measuring the combined effects of drift, diffusion and short-term plasticity on the working memory function of continuous attractor networks. First, we introduce mutual information as an abstract measure for the ability of such networks to store information about initial positions over time, and the effect of different sources of frozen variability and facilitation parameters. We show that mutual information can be increased and maintained by facilitation, and we demonstrate that this strongly depends on the combined effects of diffusion and drift. In the second part, we introduce system size scaling into our theory. This yields theoretical predictions on classes of working memory systems that are expected to be subjected to the same magnitude of displacements through both drift and diffusion. Thereby, we derive bounds on network sizes and facilitation parameters that will allow continuous attractors to perform accurate working memory function in the presence of frozen noise and spiking variability.

4.4.6.1 Effects of short-term facilitation on mutual information

To measure the combined effects of random diffusion and directed drift in working memory systems, we measure the mutual information (MI) of memories encoded in spiking continuous attractor networks over a delay period of $6.5s$. In this setting, we interpret the neuronal network as a noisy communication channel (Latham and Roudi, 2009) that maps a set of initial positions $\varphi(t = 0s)$ (time of the cue offset in the attractor network) to the associated final positions $\varphi(t = 6.5s)$, which are read out of the system after the delay period. Mutual information then provides a measure of the amount of information contained in the readout position about the initially encoded position (Cover and Thomas, 2012; Kilpatrick

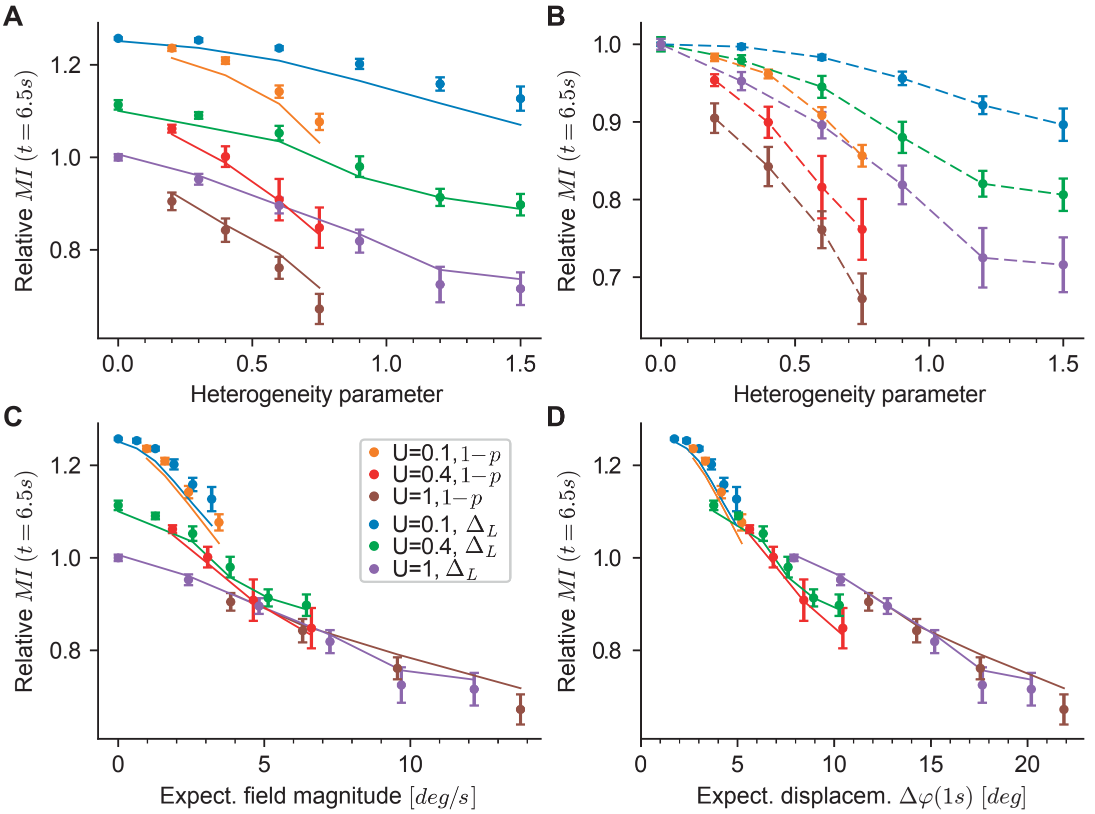


Figure 4.14 – **Effects of short-term facilitation on mutual information.** All panels show normalized mutual information (MI) of distributions of initial and final bump center positions in working memory networks. MI is calculated from 20 repetitions of 20 equally spaced initial cueing positions, with final positions after $6.5s$ of delay activity. Dots are average MI ($18-20$ realizations, error bars show 95% confidence of the mean) obtained from spiking network simulations. Lines show average MI calculated from Langevin dynamics for the same networks, repetitions and realizations (see main text). Sparse connectivity is plotted as $1-p$. **A** MI as function of network heterogeneity parameters: sparse networks ($1-p$) and leak heterogeneity (σ_L). MI values are normalized to the average MI of the spiking network without facilitation ($U=1$) and no heterogeneities (purple left-most dot). **B** Same as A, but MI normalized to the average MI of each spiking network without heterogeneities (leftmost dot for each blue, green, purple). Dashed lines connect the means, for visual guidance. **C** Same as in A, with x-axis given by the expected field magnitude, Eq. (4.9). **D** Same as in A, with x-axis given by the expected displacement $\Delta\varphi(1s)$, Eq. (4.10).

et al., 2013). Here, we investigated how the MI of initial and final positions in a continuous attractor working memory system changes as a function of short-term facilitation and the amount of frozen noise the neuronal system is subject to.

To measure MI from simulations, we simulated 30 networks with different short-term facilitation parameters and magnitudes of frozen noise. These networks were simulated for 20 repetitions of 20 initial cue positions, and 18-20 realizations of frozen noise each (same dataset as in Section 4.4.5). We recorded the center positions encoded in the network at the time of cue-offset ($t = 0$) and after $6.5s$ of delay activity, and used binned histograms (100 bins) to calculate discrete probability distributions of initial ($t = 0$) and final positions ($t = 6.5$). These probability distributions are used to calculate the MI across all repetitions and cues for each network realization (see Section 4.6.7.2). For each trajectory simulated in spiking networks, we then generated a trajectory starting at the same initial position by using the Langevin equation Eq. (4.3) that describes the drift and diffusion dynamics of center positions (see Section 4.4.3 for details). The resulting distributions of final positions (again at $t = 6.5$) for each network serve as our theoretical prediction of the mutual information for each network. Finally, we designated the network without facilitation ($U = 1, \tau_u = 650ms, \tau_x = 150ms$) and no frozen noises ($p = 1$, $\sigma_L = 0mV$) as the "control network". The average MI measured from the control network was used to normalize all other measured values of mutual information (both for spiking simulations and theoretical predictions). In the following, MI will thus always be given relative to that measured in the control network (except for Fig. 4.14B, see below).

We find that the average relative MI decreases from the level of the control network as network heterogeneities are introduced (Fig. 4.14A, purple and brown dots). This is not surprising, since the MI of the control network is set by its level of diffusion, and directed drift leads to an additional loss of information about initial positions, since trajectories tend to converge towards the stable fixed points. The effects of short-term facilitation (by decreasing the parameter U) on drift and diffusion are neatly summarized by their effect on the MI. First, diffusion is slowed, which is visible in higher offsets of the relative MI for facilitated networks (Fig. 4.14A, other dots at 0 heterogeneity). Second, the effects of frozen noise are decreased for increased facilitation. This is visualized in Fig. 4.14B, where we plot the MI relative to that of each network without frozen noise (effectively removing the offsets of facilitated networks): the slopes of the decrease of MI become less steep as facilitation is increased. The MI obtained by integration of the Langevin equations (see above) matches those of the simulations well (Fig. 4.14A, lines). From earlier results, we expected the drift-fields to be slightly over-estimated by the theory as the heterogeneity parameters increase (Fig. 4.10), which would lead to a general under-estimation of MI. We do observe this here, although for $U = 1$ the effect is slightly counter-balanced by the under-estimated level of diffusion (cf. Fig. 4.5A, right), which we expect to increase the MI. For networks with stronger facilitation ($U = 0.1$) the level of MI under-estimation is increased, since we additionally systematically over-estimate diffusion (cf. Fig. 4.5, left).

Finally, we sought to find a quantity that reduced the functional dependence between MI and the varying levels of facilitation and frozen noise. As a first candidate, we investigated the expected field magnitude (Eq. (4.9)): plotting the MI data in dependence of $\sqrt{\langle A^2 \rangle_{\text{frozen}}}$ (Fig. 4.14C) shows that the MI curves collapse onto single curves for each facilitation parameter, thereby isolating the effects that facilitation has on diffusion. To include additionally these effects, we turned to the combined dynamics of Eq. (4.3). Replacing the field $A(\varphi)$ by $\sqrt{\langle A^2 \rangle_{\text{frozen}}}$ and forward integrating the differential equation in time for an interval $\Delta t = 1s$, we arrive at the *expected displacement* in $1s$:

$$\Delta\varphi(1s) = \sqrt{\langle A^2 \rangle_{\text{frozen}}} \cdot 1s + \sqrt{B \cdot 1s}. \tag{4.10}$$

This quantity describes the total expected displacement of the center variable in $1s$ due to drift (as a function of the frozen noise distribution parameters) and diffusion. Finally, plotting the MI data against $\Delta\varphi(1s)$ we find that all data collapse on to nearly a single curve (Fig. 4.14D).

In summary, we have analyzed the MI of working memory systems with varying frozen noises and facilitation parameters. We have shown that facilitation increases the information retained in a working memory system by both decreasing diffusion and slowing drift. By using our theoretical predictions of drift and diffusion we were able to derive the expected displacement $\Delta\varphi$ which depends only on the parameters of frozen noise and the network. We have shown that this quantity describes the combined effects of drift and diffusion on working memory function as measured by the MI.

We briefly discuss these results here. First, the choice of $1s$ of forward integrated time for the expected displacement of Eq. (4.10) is arbitrary. While a choice of $\sim 2s$ lets the curves in Fig. 4.14D collapse slightly better, we chose $1s$ to avoid further heuristics. Second, we investigated only relative changes of MI in working memory systems. The absolute values of MI do depend on the amount of different initial cue positions that one imposes, since the entropy of the distribution of initial positions is an upper bound for the MI (Cover and Thomas, 2012). In our comparisons, the absolute value of MI for the control network was 3.54 bit, while the maximal measured MI was 4.45bit ($U = 0.1$, no frozen noise). The mean entropy of initial positions (this varies slightly for each realization) was 5.00bit. Thus, in our setup, the maximal achievable relative MI would be 1.41 in Fig. 4.14A,C,D, which was not achieved even under strong facilitation. Third, we expect values of MI to decrease as the length of the delay period is increased. Our choice of $6.5s$ lies at the upper bounds of the delay periods often considered in behavioral experiments (usually 3-6s) (Funahashi et al., 1989; White et al., 1994; Chafee and Goldman-Rakic, 1998). However, a more rigorous link between the MI measure and the underlying attractor dynamics would be desirable. Indeed, for noisy channel systems governed by Fokker-Planck equations, this might be feasible (Wibisono et al., 2017). While this goes beyond the scope of this work, in deriving a general Langevin description (which is equivalent to a Fokker-Planck equation; van Kampen (1992)) for continuous attractor working memory networks with short-term plasticity we have provided an initial link that could be used to derive a more comprehensive theory in the future.

4.4.6.2 Theoretical bounds on working memory function

Up to this point, we have developed a theory that can predict the magnitude of diffusion and drift in working memory systems with short-term plasticity. In addition, we have only investigated the predictions for one single size of neuronal network: that of the spiking network of $N_E = 800$ excitatory neurons. In this section we extend our theory by including the effects of the system size of the excitatory population. With this tool in hand, we demonstrate that the theory can be used to make general predictions for the expected displacements in networks, and thereby place architectural bounds on working memory systems.

To include system size dependence in our theory, we re-scaled all sums appearing in Eq. (4.3), thereby exposing the system size as an explicit parameter (see Section 4.6.5.4 for details). We find (see Eq. (4.84)) that the expected field magnitude $\sqrt{\langle A^2\rangle_{\text{frozen}}}$ (cf. Eq. (4.9)) scales with the sparse connectivity parameter p and the system size N to leading order as $1/\sqrt{pN}$, whereas the perturbations due to single unit heterogeneity scale as $1/\sqrt{N}$, both in accordance with earlier results (Zhang, 1996; Renart et al., 2003; Itskov et al., 2011b; Hansel and Mato, 2013). For the diffusion scale $\sqrt{B}$ (cf. Eq. (4.5)) we find a scaling as as $1/\sqrt{N}$, also in agreement with earlier work (for example Compte et al. (2000); Renart et al. (2003); Hansel and Mato (2013)). In the following we use these rescaled drift and diffusion strengths to estimate the interaction of system size with short-term plasticity in the effective stabilization of the ring-attractor working memory.

To measure the stability of working memories, we turn again to the measure of expected displacement during $1s$ of delay activity $\Delta\varphi(1s)$, which was derived in the last section (Eq. (4.10)). There, we have shown that the expected displacement can be related to working memory function via the measure of mutual information. Using our theory, we map out the stability of networks for given magnitudes of frozen noises, as the system size and all short-term plasticity parameters are varied. We systematically varied all possible parameters, including the system size N, while recording the expected deviation $\Delta\varphi(1s)$. We calculated this by using numerical coefficients in Eq. (4.3) that are measured from the spiking networks investigated here with $U = 1, \tau_u = 650, \tau_x = 150$ and $N_E = 800$, and extrapolating from there by changing N and short-term plasticity parameters.

It should be noted, that the expected displacement only reflects the displacement due to drift and diffusion with drift contributions averaged over all possible instantiations of frozen noise. However, since drift fields will, as we have seen in Section 4.4.5, have several local minima, the position-dependent dynamics of each realization can be quite different from the averaged case. For example, large drift fields still can contain regions that are locally dominated by diffusion (cf. Fig. 4.7). Nevertheless, this simplification allows us to make theoretical predictions for network properties, given displacement constraints that should hold across the whole attractor manifold. Further, we have seen previously that for large magnitudes of frozen noise our theory tends to over-estimate the expected magnitude of drift-fields slightly (cf. Fig. 4.10). Thus, we expect the predictions made here to be upper bounds on network

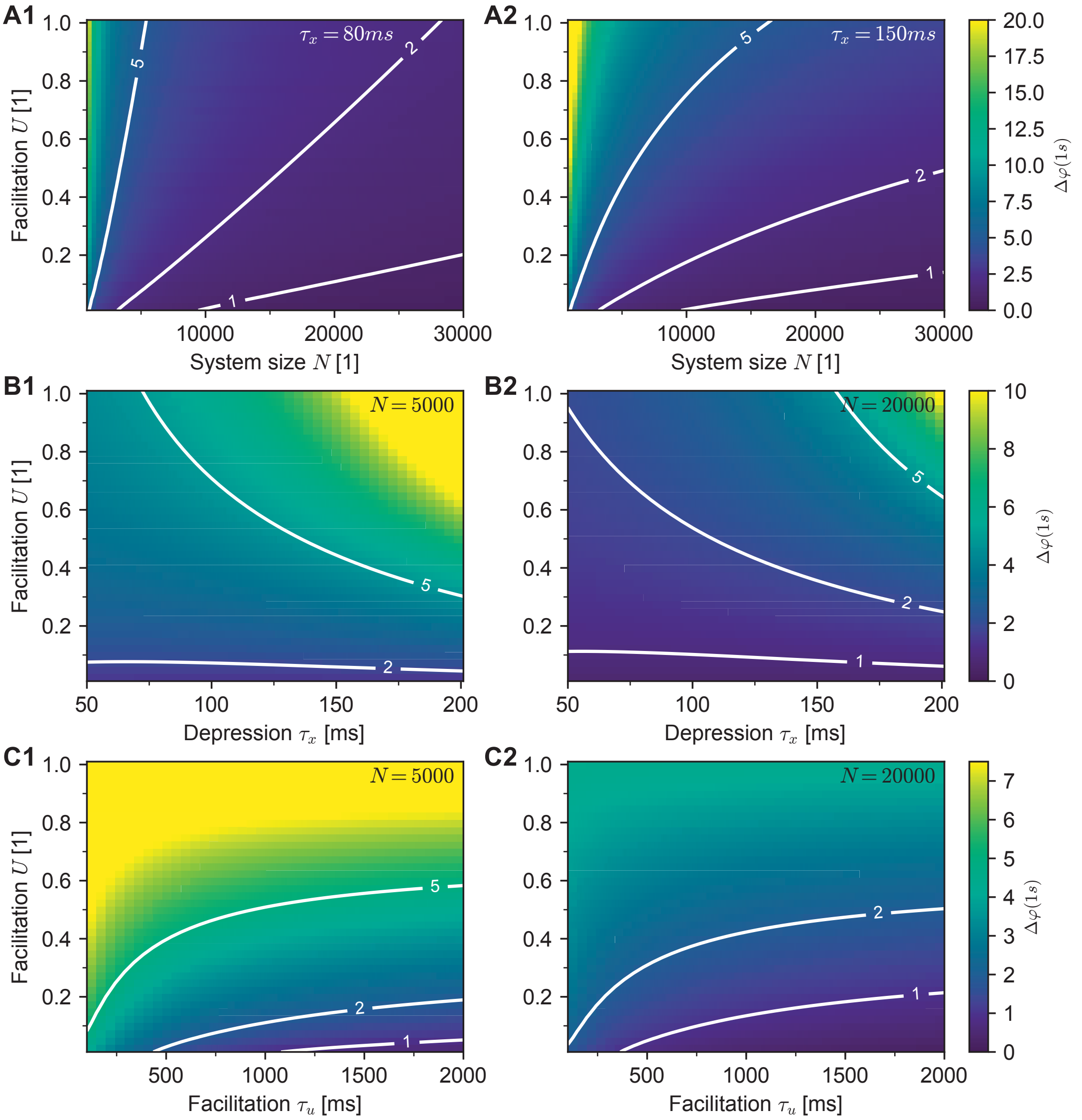


Figure 4.15 – **Stability of memory function depends on network parameters.** Colors display expected displacements $\Delta\varphi(1s)$ over $1s$ of delay activity, cf. Eq. (4.10). White lines display displacement contours for 1, 2 and 5 deg. **A1-2** Displacement as a function of system size and facilitation U, for a small depression time constant $\tau_x = 80ms$ (A1) and $\tau_x = 150ms$ (A2). In both panels $\tau_u = 650ms$. **B1-2** Displacement as a function of depression time constant τ_x and facilitation U for $N = 5000$ (B1) and $N = 20000$ (B2). In both panels $\tau_u = 650ms$. **C1-2** Displacement as a function of facilitation time constant τ_u and facilitation U for $N = 5000$ (C1) and $N = 20000$ (C2). In both panels $\tau_x = 150ms$.

STP parameters	$\Delta\varphi(1s)$	p	σ_L	**Network size** N
$U = 0.17$ $\tau_u = 563ms$ $\tau_x = 242ms$ (Wang et al., 2006, E1b)	1.5 deg (Ploner et al., 1998)	0.12 (Wang et al., 2006)	1.7mV (Yang et al., 1996, RS)	36’047
			2.4mV (Dégenètais et al., 2002, fa-RS)	57’063
		0.1728 (Wang et al., 2006)	1.7mV	35’610
			2.4mV	56’809
	1.0 deg	0.12	2.4mV	127’465
$U = 0.35$ $\tau_u = 482ms$ $\tau_x = 163ms$ (Wang et al., 2006, E1a)	1.5 deg	0.12	1.7mV	46’176
			2.4mV	73’252
		0.1728	1.7mV	45’738
			2.4mV	72’999
	1.0 deg	0.12	2.4mV	163’896

Table 4.1 – **Upper bounds on system-sizes for stable continuous attractor memory in pre-frontal cortex.** Theoretical predictions of Eq. (4.10) optimized for the number of excitatory neurons N that are needed to achieve a given level of expected displacement $\Delta\varphi(1s)$ under given parameters of short-term plasticity and frozen noises. RS: regular spiking pyramidal cells, fa-RS: fast-adapting regular spiking pyramidal cells.

parameters needed to achieve a certain expected displacement. This holds even more in light of the fact that facilitation is not the only process able to stabilize continuous attractor working memory (see Discussion).

In Fig. 4.15 we demonstrate how this procedure leads to theoretical predictions for the levels of expected deviation, in dependence of network parameters. To approximate the levels of frozen noise we expect in cortical networks, we fix the level of frozen noise by imposing sparse connectivity ($p = 0.12$; Wang et al. (2006)) and a conservative level of leak heterogeneity ($\sigma_L = 1.7mV$; Yang et al. (1996)). For example, increasing the network size to $30,000$ excitatory neurons would give expected deviations of between $1-2$ deg for $U < 0.4$ (Fig. 4.15A2), which is not an unrealistic setting for recurrent short-term facilitation at recurrent excitatory synapses found in pyramidal networks of pre-frontal cortex (Wang et al., 2006). Similarly, we find that, while increased depression time constants τ_x lead to larger displacements (Fig. 4.15A1, B1), these can also be counter-balanced by increasing the system size (Fig. 4.15A2, B2). We also find that, across system sizes, increasing the facilitation time constant τ_u beyond $1s$ affects the expected displacements only slightly (Fig. 4.15C1-2).

Finally, we optimized the theoretical result to give upper bounds on network sizes N to be expected in networks of excitatory neurons in pre-frontal cortex, that implement putative continuous attractors (see Table 4.1) (Funahashi et al., 1989; Compte et al., 2000; Constantinidis and Wang, 2004; Wimmer et al., 2014). The expected displacement is set to a value of $\Delta\varphi(1s) = 1.5$ deg (Ploner et al., 1998)[3]. Facilitation parameters were based on facilitation settings of two groups of strongly facilitating synapses found in a study of mammalian (ferret) prefrontal cortex (Wang et al., 2006). The same study reported a general probability $p = 0.12$ of pyramidal cells to be connected. However, for pairs of pyramidal cells that were connected by a facilitating synapse, the study found a high probability of reciprocal connections ($p_{rec} = 0.44$): thus if neuron A was connected to neuron B (at p), neuron B was connected to neuron A with high probability (p_{rec}), resulting in a non-random connection topology. To approximate this, we evaluated a second, slightly elevated, level of random connectivity, that included reciprocal connections: $p + p \cdot p_{rec} = 0.1728$. For the more critical leak reversal-potential standard deviations we compared values measured in two studies (Yang et al., 1996; Dégenètais et al., 2002). The maximum network sizes we find necessary (see Table 4.1, N_E) are mostly smaller than 10^5 neurons, with values depending most strongly on the value of the facilitation baseline U, and the magnitude of the leak reversal-potential heterogeneities σ_L. Since the expected field magnitude scales weakly ($1/\sqrt{p}$) with the recurrent connectivity p, increasing p lead only to comparatively small decreases in the predicted network sizes. Finally, we see that the increasing the reliability of networks further comes at a cost: decreasing the expected displacement to $\Delta\varphi(1s) = 1\,\mathrm{deg}$ more than doubles the required neurons for both facilitation settings. Nevertheless, these network sizes lie within anatomically reasonable ranges (Collins et al., 2016). **In summary, these results predict that cortical continuous attractor networks with realistic values for the strength of facilitation and depression of recurrent connections can achieve sufficient stability, even under the presence of realistic biological variability and absence of further stabilization mechanisms (see Discussion).**

---

[3] Linear fit to data approximated from Fig. 3B in Ploner et al. (1998) gives a gain variability slope of ~ 0.01, which evaluates to expected deviations between 1 deg (10 deg target eccentricity) and 2 deg (20 deg target eccentricity).

4.4.7 Application: Construction of attractors

In the earlier sections we have developed a theory that allows us to derive the expected amounts of drift and diffusion on continuous attractors. In this section, we will invert the problem: using our theory, and given the magnitude of diffusion, we derive heterogeneities in the leak reversal-potentials of excitatory neurons necessary to stabilize a given steady-state distribution of positions on the continuous attractor. More generally, since we can control both the diffusion (by short-term plasticity parameters, see Section 4.4.4) and the field implemented on the attractor, this allows us to implement systems where any trajectory of the continuous attractor system is a sample trajectory of a given Langevin equation. In the following, we first develop the necessary theory, and show that the space of drift-fields that can be created by heterogeneities of the leak-potential is limited. We then show how the drift fields can be inferred from steady-state distributions of the drift-diffusion dynamics on the attractor, and validate the theoretical considerations by calculating heterogeneities that implement a simple unimodal distribution in a network of spiking neurons.

4.4.7.1 General considerations for drift fields caused by leak heterogeneities

Here, we show that the drift-fields caused by heterogeneities in the leak reversal-potentials will always be a superposition of Fourier modes, that are determined only by the neural system the continuous attractor is implemented in. As usual, we assume all indices to be circular (taken modulo N) in the following.

In Eq. (4.7) we assumed that the bump shape $\vec{\phi}_0$ and other spatially inhomogeneous quantities stay centered at $\varphi_0 = 0$, while the firing rate perturbations $\Delta\vec{\phi}(\varphi)$ depend on the center position φ. In the following, we will use that the single leak heterogeneities Δ_i^L are independent of the bump center position, and rewrite the equation for the drift-field (Eq. (4.7)) to shift the dependence on the center position into the bump shape. Starting from Eq. (4.7), it then holds that

$$
\begin{aligned}
A(\varphi_j) &= \frac{1}{S}\sum_{i=0}^{N-1} C_i \frac{dJ_{0,i}}{d\varphi} \Delta\phi_i(\varphi) \\
&= \frac{1}{S}\sum_{i=0}^{N-1} C_i(\varphi_j) \frac{dJ_{0,i}}{d\varphi}(\varphi_j) \frac{d\phi_{0,i}}{d\Delta_i^L}(\varphi_j) \cdot \Delta_i^L, \\
&= \frac{1}{S}\sum_{i=0}^{N-1} C_{i-j} \frac{dJ_{0,i-j}}{d\varphi} \frac{d\phi_{0,i-j}}{d\Delta_{i-j}^L} \cdot \Delta_i^L.
\end{aligned}
$$

In the first equality we used Eq. (4.80). The second equality holds, since we can rewrite a shift in bump center as a shift in index of each vector quantity (see Section 4.6.1.1). This relation can be rewritten as a matrix multiplication with a circulant matrix M (each consecutive row consists of the same vector shifted to the right):

$$\begin{aligned} \mathbf{A} &= M\mathbf{q}^L, \\ M_{ij} &\equiv \frac{1}{S} C_{j-i} \frac{dJ_{0,j-i}}{d\varphi} \frac{d\phi_{0,j-i}}{d\Delta^L_{j-i}}, \\ q^L_i &= \Delta^L_i \end{aligned} \tag{4.11}$$

Here, $A_k = A(\varphi_k)$ is the vectorized drift field, q^L_i is the vectors of leak perturbations, and $0 \leq i, j < N$.

Circulant matrices have the useful property of being diagonalized by the discrete Fourier matrix $\tilde{F}_{jk} = \exp\left(-2\pi i jk/N\right)$ (Golub and Loan, 1996), where i is the imaginary unit and $0 \leq j, k < N$. The eigenvalues of the circulant matrix M are the Fourier coefficients $\tilde{F}\mathbf{m}$, where $\mathbf{m}$ is the first row vector of M with entries $m_i \equiv M_{0,i}$. We use this to transform Eq. (4.11) into Fourier-space and obtain a vector equation:

$$\tilde{F}A = \tilde{F}M\tilde{F}^{-1}\tilde{F}\mathbf{q}^L = \tilde{F}\mathbf{m} \cdot \tilde{F}\mathbf{q}^L, \tag{4.12}$$

where $\cdot$ stands for element-wise multiplication. A second way of seeing this, is by noticing that $A = M\mathbf{q}^L = \mathbf{m} \star \mathbf{q}^L$, where $\star$ is the circular convolution and again $\mathbf{m}$ is the first row of M. Thus, taking the discrete Fourier transform on both sides, by the circular convolution theorem (Golub and Loan, 1996) we get the same equation.

If all components of $F\mathbf{m}$ are nonzero, it is straightforward to solve Eq. (4.12) by calculating the element-wise solution:

$$\mathbf{q}^L = \tilde{F}^{-1} \left(\frac{(FA)_i}{(Fm)_i} \right)_{0 \leq i < N}. \tag{4.13}$$

However, in general this might not be well defined, or lead to very large Fourier coefficients of $\mathbf{q}^L$ for those modes where $\left(\tilde{F}\mathbf{m}\right)_i$ has close to zero power. To circumvent this, we calculate the Fourier coefficients for $\mathbf{q}^L$ by Eq. (4.13), setting all coefficients to zero where $\left|\left(\tilde{F}\mathbf{m}\right)_i\right| < c$ for a certain (small) cutoff value c. Effectively, we approximate A by low-pass filtering it to the nonzero frequencies of $\mathbf{m}$.

Therefore, by the considerations above (Eq. (4.12)**), we see that the space of possible drift-fields that can be implemented by heterogeneous leaks is entirely determined by the Fourier spectrum of the vector** $m_i = \frac{1}{S} C_i \frac{dJ_{0,i}}{d\varphi} \frac{d\phi_{0,i}}{d\Delta^L_i}$**: any drift field will be entirely decomposable into the nonzero Fourier modes of m.** In turn, the spatial shape of $\mathbf{m}$ is determined by the facilitation and depression parameters (primarily through C_i) as well as the other microscopic parameters of the system. Finally, it should be noted that for perturbations created by recurrently mediated heterogeneities (as for example sparse connectivity), these considerations become more complicated since the resulting firing rate perturbations depend on the position of the bump through the recurrent connectivity, and thus can not be decoupled from the position of the bump as simply as here.

4.4.7.2 Calculation of leak heterogeneities for given steady-state distributions

We now apply the above to calculate the leak heterogeneities required to realize a given steady-state probability distribution of center positions $P(\varphi)$ with $\varphi \in [-\pi, \pi)$. Assuming that $P(\varphi)$ is the steady-state solution of the Langevin-equation Eq. (4.3), one can show that (Gardiner, 2009, p. 124)[4]:

$$P(\varphi) = \mathcal{N} \exp\left(2\frac{\int_{-\pi}^{\varphi} A(\varphi')d\varphi'}{B}\right) \equiv \mathcal{N} \exp\left(-\frac{U(\varphi)}{D}\right), \tag{4.14}$$

where $\mathcal{N}$ normalizes $\int_{-\pi}^{\pi} P(\varphi)d\varphi = 1$, and we have defined the potential $U(\varphi)$ as well as the diffusion constant $D = B/2$. Here, B is given by the theory of Eq. (4.5).

By taking the logarithm of Eq. (4.14) and differentiating with respect to φ, we can solve for the field $A(\varphi)$ that will implement the desired steady-state distribution $P(\varphi)$:

$$\begin{aligned} \ln P &= \ln \mathcal{N} - \frac{U}{D}, \\ A &= -\frac{d}{d\varphi}U = D\frac{d}{d\varphi}\ln P. \end{aligned} \tag{4.15}$$

Applying Eq. (4.13) of the last section, we can then derive leak heterogeneities that will approximate the field A. As we have seen there, how well a specific continuous attractor can implement this field will depend on the Fourier spectrum of A and the microscopic details of the neural system.

4.4.7.3 Example: Gaussian distribution

To validate this approach and see how well it performs in a network of spiking neurons, we now turn to implementing a simple Gaussian distribution (Fig. 4.16). In particular, we choose

$$P(\varphi) = \mathcal{N} \exp(-\frac{\varphi^2}{2\sigma^2}) \tag{4.16}$$

with $\sigma = \pi/20$. We proceed by calculating Eq. (4.15) for this distribution, which yields

$$A(\varphi) = -D\varphi/\sigma^2.$$

To approximate this function by heterogeneities of the leak-reversal potentials, we solved Eq. (4.13) by keeping only modes with up to 12cycles spatial frequency, which resulted in a

---

[4] For the sake of simplicity, we assume here the distribution P to be contained well within the boundaries, which are assumed to be absorbing; the extension to periodic solutions is described in Gardiner (2009, p. 125).

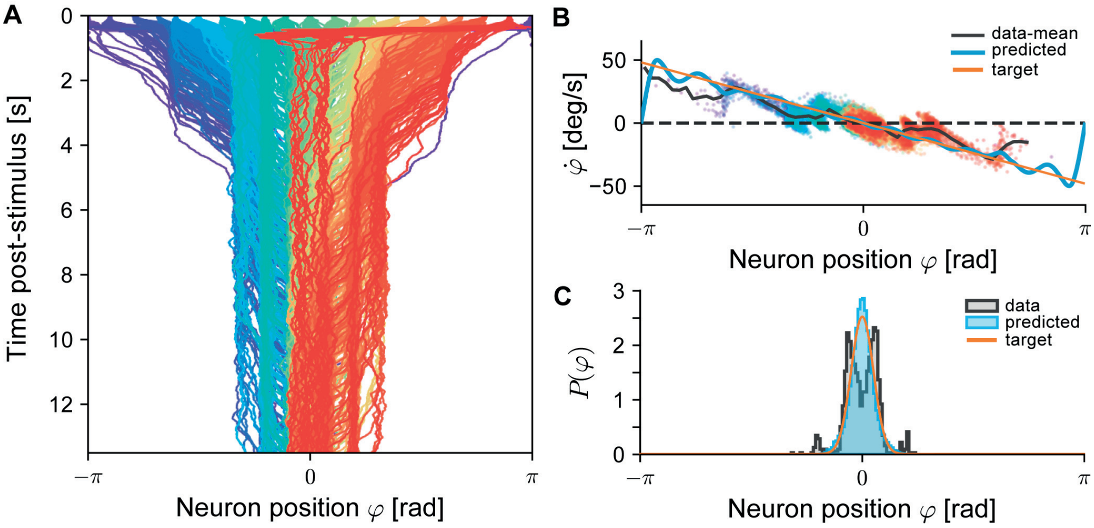


Figure 4.16 – **Design of an attractor network with unimodal steady-state distribution. A** Center positions of bumps in spiking network simulations with heterogeneities of leak reversal-potentials of excitatory neurons derived to approximate a unimodal steady-state distribution (see main text for details). Shown are 50 repetitions for 20 uniformly spaced initial cues, each simulated for $13.5s$ delay time after cue offset. Short-term plasticity parameters were $U = 0.8, \tau_u = 650ms, \tau_x = 150ms$. **B** Drift-field $\dot{\varphi}$ depending on the bump center φ. The target drift field (orange line) is approximated by the field predicted from theory (blue line). Colored dots are estimated from trajectories shown in A, alongside binned mean (black line). **C** Histograms (100 bins) of final positions of simulated trajectories shown in panel A (gray), and 500 realizations of the Langevin equation of the system predicted by theory (blue). Histograms are normalized to probability distributions on $[-\pi, \pi)$. Both approximate the target distribution (orange line).

solution for the vector of heterogeneities $\mathbf{q}^L$ with standard deviation $0.89mV$[5]. In Fig. 4.16B we have plotted the field $A(\varphi)$ (orange line, “target”) alongside the approximation (blue line, “predicted”). Note, that the non-periodic discontinuity of the target field is interpolated and results in an unstable fixed point at the boundary $\varphi = -\pi$. The probability distribution implemented by the combination of the approximated drift field and the predicted diffusion coefficient approximates well the target distribution Eq. (4.16) (Fig. 4.16C, orange line, “target”). We confirmed this by measuring a histogram of positions at $t = 15s$ of 500 realizations of the corresponding stochastic differential equation (cf. Fig. 4.16C, blue, “predicted”). To compare this to the spiking simulation, we added the vector of perturbations to the leak potentials, by setting $V_L \rightarrow -70mV + q_i^L$ for excitatory neurons $0 \leq i < N$ and simulating the resulting spiking attractor network (Fig. 4.16A).

We find that the simulations match the predicted result to a large degree: the width of the measured distribution of final positions (Fig. 4.16C, gray area) corresponds well to both the target and predicted distributions. The drift field is well reproduced by the spiking system (Fig. 4.16B), although a few additional local minima as well as points with slow dynamics appear in regions where the field is of low magnitude (see Fig. 4.16A, B, C). It should be noted that our theory is only linear in the perturbations of the leak potentials, and thus deviations from the predicted results for larger perturbations are to be expected. Due to the very strong field at the boundaries, we also noted that trajectories sometimes did not directly form stable bump profiles, and after cue-offset centered around the minimum of the field instead (Fig. 4.16A, e.g. red lines).

In summary, we designed and implemented a unimodal attractor of a given shape in the spiking network. Although the resulting drift field was not perfect, we were nevertheless able to demonstrate the applicability of the theory to derive the needed perturbations.

4.4.8 Application: Control of uncertainty by short-term plasticity

A central prediction of our theory is that short-term plasticity can control the effects of noise on the states encoded in continuous attractor models: mainly the depression time constant τ_x as well as the facilitation parameter U can influence diffusion from spiking noise, as well as drift fields that are created by frozen heterogeneities in neural networks. In this section we show that for neural networks with fixed connectivity, changing the parameters of short-term plasticity can result in transitions between different regimes of certainty of memory retention.

To demonstrate this, we designed a continuous attractor with two stable fixed points separated by an unstable fixed point at $\varphi = 0$, implemented by perturbations in the leak reversal-potentials of excitatory neurons (Fig. 4.17). Akin to model circuits for decision-making that are designed to integrate sensory evidence (Wang, 2002; Wong and Wang, 2006; Curtis and Lee, 2010), we interpret the center position as a proxy for the internal

---

[5] We note this value for comparison with the earlier sections – the heterogeneities here are, however, not normally distributed.

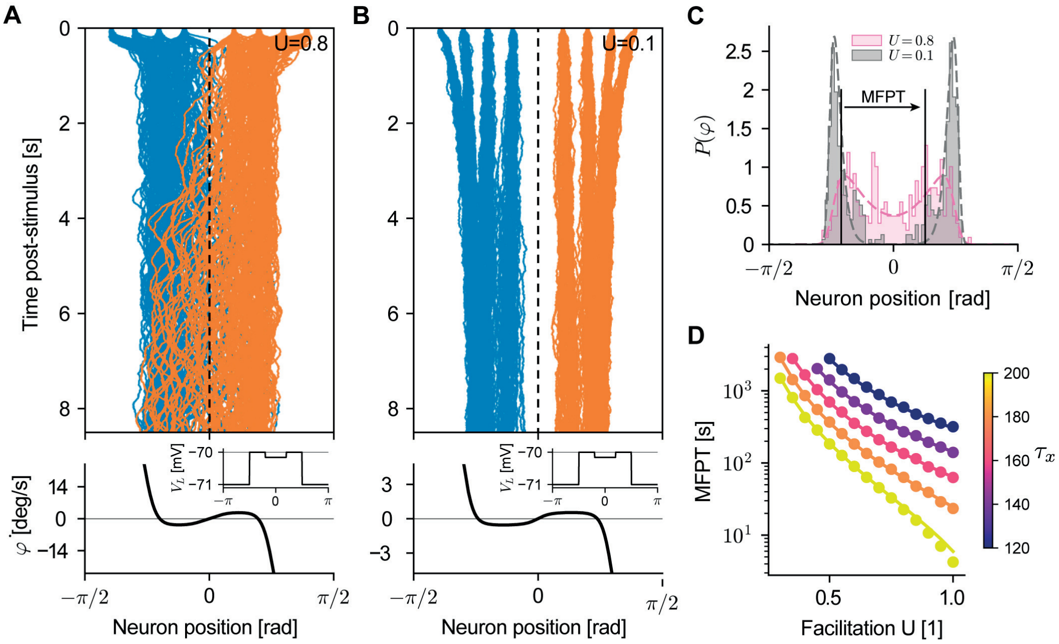


Figure 4.17 – **Short-term plasticity modulates uncertainty in bistable attractor networks. A** Top: Center positions of bumps in a spiking network with a bistable state-space (see main text for details). Shown are $8.5s$ after stimulus offset, for 50 repetitions of 20 initial positions each. The position $\varphi = 0$, which serves as a separatrix between the two states, is highlighted by a dashed line. Bottom: Positional drift field $\dot{\varphi}$ depending on the bump center φ, as predicted from theory. Inset shows leak reversal potentials that create the drift field (x-axis shows neurons, the plateau near $\varphi = 0$ is slightly lowered for visibility). Short-term plasticity parameters are $U = 0.8, \tau_u = 650ms, \tau_x = 200ms$. **B** Similar to panel A, but for a strongly facilitating network with $U = 0.1$. **C** Theoretical predictions for steady-state probability distributions of bump-centers (dashed) together with normalized histograms of the final positions after $13.5s$ of delay activity for the networks in A (pink) and B (gray). **D** Mean first-passage-times (MFPT) between the positions indicated by black horizontal lines in panel C. Lines are theoretical predictions, dots are the average of first-passage times recorded from 2000 realizations of the corresponding Langevin equations.

representation of a two-state variable, divided by $\varphi = 0$. In the absence of external input, the system dynamics lead to convergence of center positions to regions around either of the two attractive states. Switching between these states can then happen either spontaneously due to intrinsic noise in the system (diffusion) (Moreno-Bote et al., 2007), or can be externally forced by providing input to either side of the network. In the following, we show that changing the parameters of recurrent short-term plasticity on excitatory connections can significantly influence the amount of uncertainty that the network displays in its encoded positions, and in its transition between the two states. By this mechanism the network can continuously interpolate its computational mode between a "certain" mode, where the states are kept at high fidelity in a sharply constrained region of the attractor, and an "uncertain" mode, where the regions that each state occupies become less sharp, and switching between states becomes very probable.

4.4.8.1 Perturbed continuous attractor model for decision making

Starting from a continuous attractor network with strong diffusion ($U = 0.8, \tau_x = 200ms$, cf. Fig. 4.5) we created two local fixed points by systematically adding small perturbations to the leak reversal potentials of excitatory neurons (Fig. 4.17A,B insets in lower plots). While in Section 4.4.7 we introduced a framework to calculate leak perturbations, we chose to design the drift field here by hand to evade the appearance of non-predicted local minima for small fields.

For 400 excitatory neurons centered around $\varphi = -\pi$, we decreased the leak reversal potential by $1mV$. This makes these neurons less likely to fire, leaves this area strongly repelling, and thereby constrains trajectories to an area towards the middle of the attractor manifold ($\varphi = 0$). To introduce a dynamical separation of this central region into two separate basins of attraction, we decreased the leak potential of 160 neurons centered around $\varphi = 0$ by $0.04mV$. This perturbation was intentionally chosen to be small, such that displacement of trajectories elicited by strong diffusive noise could overcome it with high probability.

Confirming our design decisions, across short-term plasticity parameters, the predicted drift fields (see Fig. 4.17A,B lower plots) are strongly directed towards the center ($\varphi = 0$) for the region outside of $\sim (-\frac{\pi}{4}, \frac{\pi}{4})$. At the boundaries of this region we find two stable fixed points that are separated by an unstable fixed point at the separatrix $\varphi = 0$ (Fig. 4.17A,B, dashed vertical lines). In absence of diffusion, this system thus would act as a bistable "decision making" circuit that will have trajectories converge, after they are cued by an external input at a given position, to either the left or the right of the separatrix. Throughout the remainder of this section we will not apply any additional external inputs to the system after the initial cue (but see below).

4.4.8.2 Short-term facilitation interpolates between “certain” and “uncertain” regimes

We simulated the network introduced above for two different settings of short-term facilitation. For a weakly facilitating network $U = 0.8$ (Fig. 4.17A) the network is strongly diffusive and the encoded states readily transit from either side of the separatrix to the other. We call this the “uncertain” regime. On the contrary, for $U = 0.1$ (Fig. 4.17B) most of the diffusion is removed and trajectories very rarely cross the separatrix spontaneously (see below), making this the “certain” regime. While the drift field elicited in these networks by heterogeneities is reduced by a factor of five by facilitation (see Fig. 4.17A,B, lower plots) for $U = 0.1$, it still seems large enough to constrain trajectories to well defined regions to either side of the separatrix.

To check this, we calculated the theoretically predicted steady-state distributions for both networks, by using the predicted drift field and diffusion strength in Eq. (4.14). In both cases the theoretical predictions (Fig. 4.17C, dashed lines) and histograms of final positions of the spiking simulations (Fig. 4.17C, shaded areas) are in close agreement. For $U = 1$, diffusion is strong enough to extend the distribution of expected positions across the separatrix, with a nonzero probability density of momentary positions on the separatrix itself (Fig. 4.17C, pink). In contrast, for $U = 0.1$ (Fig. 4.17C, gray) the low diffusion and residual drift field lead to a strongly separated bimodal equilibrium distribution. It should be noted, that in the case of $U = 0.1$ the experimentally measured histogram of final positions depends more closely on the initial positions, which we limited to either side of the separatrix.

4.4.8.3 Waiting times for state transitions are modulated by STP

To study the dependence of spontaneous transitions between the states on the parameters of short-term plasticity, we turned to the theory we developed earlier. In the earlier sections (Section 4.4.4 and Section 4.4.5) we have shown that the theoretical predictions for diffusion and drift are in good agreement with simulations. Using these theoretical results, we calculated the expected average time it takes for a bump centered close to the left fixed point to move, across the separatrix, into to the right basin of attraction (Fig. 4.17C, arrow and black vertical lines). In the simplified description of this system by the Langevin equation Eq. (4.3), this is commonly referred to as the mean first-passage-time (MFPT) and can be calculated by standard methods (Gardiner, 2009, p. 136), given the diffusion B and the drift field $A(\varphi)$.

We find that the MFPT predicted from theory crosses several orders of magnitude as the parameters of facilitation and depression are varied (Fig. 4.17D). For strong diffusion ($\tau_x = 200ms$; Fig. 4.17D, yellow line) unfacilitated networks can display short waiting times on the order of several seconds (as in Fig. 4.17A). However, decreasing the facilitation parameter as low as $U = 0.5$ yields waiting times above $10^3 s$, making spontaneous transitions very unlikely. Generally, decreasing the depression time constant τ_x decreases the expected level of diffusion, and thereby increases the MFPT even for unfacilitated systems.

In summary, we have shown that short-term depression and facilitation can influence the amount of uncertainty that the network displays in its encoded positions, and can elicit control over waiting times for transitions between the two states. In this network, the waiting times for transitions achievable in facilitated networks can be quite long, quickly approaching values larger than $10^3 s$. We chose a starting point for trajectories to the left that is far from the separatrix, so the predictions above serve as lower bounds of transitions between both basins representing either state. More importantly, the tendency of trajectories to drift to either basin of attraction of the two fixed points can be drastically influenced by providing external inputs to the network, thereby representing the effect of external evidence on the internal representation of the system state. We checked this by adding a small tonic current to all neurons to the right of the separatrix (with positions $\varphi > 0$), for which we predicted a maximal increase of firing of $4.5 Hz$ across all neurons in the bump state. This decreased the mean first passage times predicted from the drift-diffusion model down to a minimum of $419 ms$ for $U = 1$ (and $9.2 s$ for $U = 0.1$).

4.5 Discussion

In this article we presented a theoretical approach for the analysis of drift and diffusion in continuous attractor models, exemplified by a one-dimensional ring attractor model. Our framework extends earlier approaches calculating the effects of noise by projection on to the attractor manifold (Wu et al., 2008; Kilpatrick and Ermentrout, 2012; Itskov et al., 2011b; Burak and Fiete, 2012) by including the dynamical effects of short-term plasticity. The theory predicts that facilitation and depression play antagonistic roles in making continuous attractors robust against the influences of both dynamic noise (introduced by spiking variability) and frozen noise (introduced by biological variability). We have confirmed the quantitative predictions of our theory in simulations of a ring-attractor implemented in a spiking network model with synaptic facilitation and depression, and found theory and simulation to be quantitatively in good agreement.

Facilitation slows down both directed drift and diffusion along the attractor manifold. Similar to an earlier theoretical approach using a simplified rate model (Itskov et al., 2011b), we find that the slowing of drift depends mainly on the facilitation parameter U, while the time constant τ_u has a less pronounced effect. While the approach of Itskov et al. (2011b) relied on the projection of frozen noise onto the derivative of the first spatial Fourier mode of the bump shape along the ring, here we reproduce and extend this result (1) for arbitrary neuronal input-output relations and (2) a more detailed spatial projection that involves the synaptic dynamics and the resulting shape of the steady state (see Section D.2 for a more detailed comparison). While we mainly focused on connectivity heterogeneities induced by sparseness, we have also investigated the case of noisy recurrent connection weights as in Itskov et al. (2011b) (see Section 4.6.5). However, the drifts generated by these heterogeneities were generally small compared to diffusion and the other sources of heterogeneity, and lead to a loss of the continuous attractor state when increased too much.

A second previous study investigated short-term facilitation and showed that it reduces drift and diffusion in a spiking network, for a fixed setting of U (although the model of short-term facilitation differs slightly from the one employed here) (Pereira and Wang, 2015). Contrary to what we find here, these authors find that an increase in τ_u leads to increased diffusion, while we find that an increase over the range they investigated ($\sim 0.5s - 4s$) would halve the diffusion. More precisely, for our shape of the bump state (which we keep fixed) we predict a reduction from ~ 26 to ~ 16 deg^2/s for a similar setting of facilitation U. The difference to our prediction may arise from an effectively increased width of the bump attractor profile for large depression time constants in Pereira and Wang (2015), which would lead to increased diffusion in our model (see below). Whether this effect persists under the two-equation model of saturating NMDA synapses used there remains to be investigated – see Section D.1 of the Appendix, where we present a possible approach using the theoretical framework presented here. Finally, increasing the time constant of recurrent NMDA conductances has been shown to also reduce diffusion (Pereira and Wang, 2015), in agreement with our theory, according to which the normalization constant S increases with τ_s (similar to the result of Burak and Fiete (2012) without short-term plasticity).

Previously, it was claimed that the presence of short-term depression does not affect the stability of the bump (Barbieri and Brunel, 2008), but in that study only a single parameter for depression ($\tau_x = 160ms$) was investigated. In contrast, we find that stronger short-term depression will increase the diffusion and directed drift along the attractor. This result seems to qualitatively agree with earlier studies in rate models which showed that synaptic depression can induce movement of bump attractors (York and van Rossum, 2009; Romani and Tsodyks, 2015; Wang et al., 2015; Mi et al., 2016), similar to neuronal adaptation (Hansel and Sompolinsky, 1998; Laing and Longtin, 2001). In particular, the study of York and van Rossum (2009) showed that for simpler rate models a regime exists where the bump state moves with constant speed along the attractor manifold. We did not find any such directed movement in our spiking networks (e.g. Fig. 4.17 shows trajectories for $\tau_x = 200ms$), and we checked that even for very large τ_x the observed dynamics of the centers are indeed diffusive as opposed to drift-dominated. We expect directed movement to be obtainable in our theory by initializing depression with asymmetry, which would result in an effective drift-term that can again be calculated as a deviation from perfect symmetry. Finally, we note that the fast spiking noise component in our networks could be stabilizing the directed movement of bumps observed in noise-free rate models, as shown in Laing and Longtin (2001): there, additive fast noise was able to cancel directed bump movement caused by single neuron adaptation.

Although we did not thoroughly investigate this here, the coefficients of Eq. (4.3) give clear predictions as to how drift and diffusion will depend on the shape of the bump state and the neural transfer function F. The relation is not trivial, since the pre-factors C_i and the normalization constant S also depend on the bump shape. For the diffusion strength Eq. (4.5), we explored this relation numerically, by artificially varying the shape of the firing rate profile (while extrapolating other quantities). Although a more thorough analysis remains to be

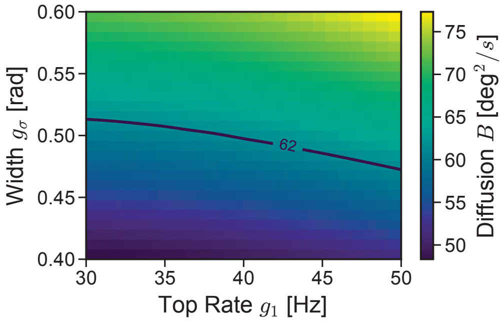


Figure 4.18 – **Dependence of diffusion strength B on shape parameters**. Diffusion was calculated from Eq. (4.5) with bump solutions $\phi_0 = g_1 \exp(-\left|\frac{x}{g_\sigma}\right|^{g_r})$. The values of $\frac{dJ_0}{d\varphi}$ and ϕ_0' were calculated by fitting and extrapolating (linearly, for $\phi_0 > 40.31$Hz) curves $\phi_0 \to \phi_0'$ and $\phi_0 \to J_0$ that were obtained from the numerical values extracted for $g_1 = 40.31$Hz, $g_\sigma = 0.51$ by theory (Section 4.6.4). Thus, any nonlinearity or saturation of the inputs and input-output relation for $\phi_0 > 40.31$Hz was not included. This approximate analysis shows that the major dependence of the diffusion expected in the system is on the bump width g_σ, although a minor dependence on g_1 is seen.

performed, our preliminary analysis shows (see Fig. 4.18) that the diffusion increases both with bump width and top firing rate, consistent with earlier findings (Compte et al., 2000; Murray et al., 2012). Additionally, our theory predicts the shape of the drift field that is generated by a local perturbation (cf. Section 4.4.7). This can be extended to the scenario of external inputs: any external input (excitatory or inhibitory) will cause a deviation $\Delta\phi_i$ from the steady-state firing rates, which, in turn, generates a drift field by Eq. (4.7). This could be used to predict the strength and location of external inputs that are needed to induce continuous shifts of the bump center at given speeds, for example when these attractor networks are designed to track external inputs (see e.g. Hansel and Sompolinsky (1998); Fung et al. (2010)).

The spiking networks we analyzed here are tuned to display balanced inhibition and excitation in the inhibition dominated uniform state (Brunel, 2000b; Wang, 2001), while the bump state relies on positive currents, mediated through strong recurrent excitatory connections (cf. Barbieri and Brunel (2007) for an analysis). Similar to other spiking network models of this class, this mean–driven bump state shows relatively low variability of neuronal inter-spike-intervals, because near the bump center the mean input is close to (or even above) the firing threshold (Renart et al., 2007; Barbieri and Brunel, 2008) (see also next paragraph). While the decreased variability appears for neurons in the center of the firing rate profile, we found that neurons at its flanks still display variable firing, with statistics close to that expected of spike trains with Poisson statistics (see Section D.3 and Fig. D.3 in the Appendix), which may be because the flank's position slightly jitters and so firing irregularity at the flank is again increased. Since the non-zero contributions to the diffusion strength are constrained to these flanks (cf. Fig. 4.3), the simple theoretical assumption of Poisson statistics of neuronal firing still matches the spiking network quite well. However, we observe that for very low diffusion (small facilitation U) our theory consistently over-estimates the expected amount of diffusion in the network (see Fig. 4.19A, B left). This may be because the slower the bump moves, the more also the firing irregularity of flank neurons decreases. Since this is inconsistent with the Poisson firing assumption of our theory, our approximations do not yield as accurate predictions in this case.

More recent bump attractor approaches allow networks to perform working memory function with a high firing variability also during the delay period (Chaudhuri and Fiete, 2016), in better agreement with experimental evidence (Compte et al., 2003). These networks show bi-stability, where both stable states show balanced excitation and inhibition (Renart et al., 2007) and the higher self-sustained activity in the delay activity is evoked by an increase in fluctuations of the input currents (noise-driven) rather than an increase in the mean input (Mongillo et al., 2012). This was also reported for a ring-attractor network (with distance dependent connections between all populations), where facilitation and depression are crucial for irregularity of neuronal activity in the self-sustained state (Hansel and Mato, 2013). To which degree our approach can be applied to these setups remains an open question. Linearizing the synaptic dynamics around a steady-state spatial profile of firing rates (or spatial profile of variances) might still be a feasible approach, although the significantly higher variability of spiking might require including the higher moments of the steady-state synaptic variables in the theory. For example, in Romani et al. (2006) the higher moments of the steady-state for a population of neurons with synaptic depression were calculated. See also the study of Mongillo et al. (2012) for a similar (albeit probabilistic) model of facilitation and depression.

We have shown that short-term plasticity offers direct and variable control over the sensitivity of attractor networks to both frozen and dynamic noise. For example, as demonstrated in Section 4.4.8, increasing facilitation of recurrent excitatory synapses (and similarly decreasing the depression time-constant, although to a lesser degree, cf. Fig. 4.15B) can reduce noise-induced state transitions. Control of short-term plasticity would represent an efficient "crank" for modulation of the timescale of computations of such networks. Indeed, evidence for neuromodulation of short-term plasticity is ubiquitous (Nadim and Bucher, 2014). By changing the properties of presynaptic calcium entry (Oh et al., 2012), inhibitory modulation mediated via $\mathrm{GABA_B}$ and adenosine $\mathrm{A_1}$ receptors can lead to increased facilitatory components in rodent cerebellar (Kreitzer and Regehr, 2000) and avian auditory synapses (Brenowitz et al., 1998). Dopamine, serotonin and noradrenaline have all been shown to differentially modulate short-term depression (and facilitation when blocking GABA receptors) at sensorimotor synapses (Barrière et al., 2008). Finally, changing the amount of background input or neuronal excitability (Nadim and Bucher, 2014) will modulate the shape of the bump state, leading to changed properties of diffusion and drift (Roach et al., 2015).

Importantly, introducing long time constants of facilitation as a stabilization mechanism for continuous attractors has been shown to make it hard to return the bump state back to uniform activity (which is usually done by providing strong inhibition to all excitatory cells) (Pereira and Wang, 2015). We have also observed that longer (or stronger) initial cues (stimuli) were needed as the strength of facilitation increased, to facilitate recurrent connections enough to stabilize the bump state. Generally, this apparent tradeoff between stability of encoded memories and the flexibility of the working memory systems could be affected by short-term depression (Mi et al., 2016) – after a strong excitatory drive to all neurons (leading to strong inhibitory feedback) residual depression could occlude the facilitation state of recurrent connections and allow for an easier reset.

In Section 4.4.6 we compared working memory systems with short-term plasticity, and derived bounds on network parameters that lead to "tolerable" levels of drift and diffusion. It should be noted, that in cortical settings we do expect only short-term facilitation to be solely responsible for the stabilization of putative continuous attractor networks. Several dynamical mechanisms other than short-term facilitation have been proposed for stabilization. The authors of Pereira and Wang (2015) investigated two further mechanisms that can make bump-attractors robust: activity dependent disinhibition of excitatory neurons (DSI, see also Carter and Wang 2007) and a cation current leading to activity-dependent increase in excitability for excitatory neurons (I_{CAN}, see also Tegnér et al. (2002)) – both are found to affect diffusion of bump positions. Fast corrective inhibitory feedback has also been shown to stabilize spatial working memory systems (in balanced networks) (Lim and Goldman, 2014). A different study has shown that dynamical control of neuronal adaptation by acetylcholine neuromodulation can be used to control the stability properties of continuous attractors (Roach et al., 2015). Next to dynamical mechanisms for stabilization, also the recurrent connectivity can affect the stability of continuous attractors (Burak and Fiete, 2009). It has been shown that, on the timescale of hours to days, homeostatic processes can be used to counteract the effects of frozen noise (Renart et al., 2003). Purposefully breaking the continuous attractor into several smaller fixed points can be achieved by introducing structured heterogeneity, and has been shown to decrease diffusion along the attractor (Kilpatrick et al., 2013). Nevertheless, our theoretical results could be used, given estimates of the "size" of bumps maintained in cortical working memory networks (Wimmer et al., 2014) and short-term plasticity parameters (Wang et al., 2006), to place upper bounds on the size of the recurrent networks involved in generating working memory function (see Section 4.4.6.2).

One would not imagine cortical networks to ever be tuned to display the perfect symmetry of continuous attractor networks, and we have confirmed (as others have before) that these circuits are intrinsically susceptible to both dynamical and frozen noise. However, continuous attractor models that are fully connected and have no heterogeneity have been successfully employed in the context of visuospatial working memory to explain behavioral performance (Macoveanu et al., 2006; Wei et al., 2012; Roggeman et al., 2014; Almeida et al., 2015), to predict the effects of neuromodulation (Cano-Colino et al., 2013, 2014), or the implications of cognitive impairment (Murray et al., 2012; Cano-Colino and Compte, 2012). Furthermore, recent evidence shows that cortical activity converges to rather low-dimensional subspaces

(Balaguer-Ballester et al., 2011; Murray et al., 2016) that might indeed implement the theoretical concept of a continuous attractor (Yoon et al., 2013; Wimmer et al., 2014). While major imperfections in biological continuous attractors could be recovered by learning processes on the time scale of hours to days (see above), we expect circuits also to be subject to ongoing but slow imperfections, requiring dynamical processes for additional stabilization from drift. We have shown that short-term plasticity is such a dynamical mechanism, able to control the reliability of attractor dynamics. Our theory can be used to make general predictions about working memory systems (Section 4.4.6), using only experimentally constrained parameters of short-term plasticity (Wang et al., 2006), the size of the cortical networks involved (Markram et al., 2015), and the shape of putative self-sustained bumps of activity in cortex. We have seen that imperfections will necessarily introduce local fixed points in the attractor (see also Kilpatrick et al. (2013)), but short-term plasticity can slow down the resulting transient dynamics, or enable spontaneous transitions between them by increasing diffusion (see Section 4.4.8).

Our results have novel implications for short-term plasticity as a stabilizing mechanism for working memory. We have shown, to our knowledge for the first time, that strong facilitation (small values of U) as a stabilizing mechanism does not only slow directed drift, but also is very efficient at removing diffusion. However, in delayed response tasks involving saccades, that presumably involve continuous attractors in the prefrontal cortex (Compte et al., 2000; Wimmer et al., 2014), one does observe an increase of variability in time (Funahashi et al., 1989): both quickly accumulating systematic errors (alike drift) (White et al., 1994) and longer increasing variable errors (with variability growing linear in time, alike diffusion) (Ploner et al., 1998) appear. Pyramidal networks of the prefrontal cortex have been reported to be highly reciprocally connected, often with strongly facilitating synapses (Wang et al., 2006). Thus, if biological circuits use mainly facilitation to stabilize continuous attractor dynamics, either a) variable errors are introduced elsewhere in the pathway between visual input and motor output (but see (Brunton et al., 2013)) or by input from of other noisy local circuits during the delay period (Selemon and Goldman-Rakic, 1988); b) intermediate fixed values of U are used, which slow diffusion and drift only slightly, and networks employ other methods of stabilization that do not affect diffusion; c) the higher variability of neuronal firing of continuous attractors in balanced networks (see above) might lead to increased and persistent diffusion, even under stronger facilitation; d) short-term plasticity is under cortical neuromodulatory control (see above), such that for less “certain” situations, as delay time progresses, diffusion can re-appear when facilitation is reduced or depression is increased. Neuromodulation could also generally provide higher flexibility (see above) to a facilitation-stabilized working memory system, by allowing for more dynamical control of the rigidity of memory representations. Interestingly, next to short-term facilitation on the timescale of seconds, other dynamic processes up-regulate recurrent excitatory synaptic connections in prefrontal cortex (Wang et al., 2006): synaptic augmentation and post-tetanic potentiation operate on longer time-scales (up to tens of seconds), and might be able to support working memory function (Hempel et al., 2000). While the long time scales of these processes might again render putative short-term memory networks inflexible, there is evidence that they might also be under tight neuromodulatory control (Sakurai and Katz, 2009).

4.6 Methods

4.6.1 Projection of dynamics onto the attractor manifold

In this section, we present the theoretical toolset needed to derive the central result presented in Section 4.4.3 of the main text – a one-dimensional Langevin equation for the dynamics of the center position of the bump profile (Eq. (4.3)). First, we formalize networks implementing ring-attractors in Section 4.6.1.1. In the following subsections, we then calculate the projection of the linearized synaptic dynamics, including facilitation and depression, onto the tangential space of the attractor manifold. We show that this yields a differential equation for the temporal dynamics of the center position. This projection will then be used in the next section (Section 4.6.2) to calculate expressions for the two different parts of the Langevin equation: diffusion (Section 4.6.2.1) and drift (Section 4.6.2.2).

To begin, we briefly restate some basic definitions, which are repeated from Section 4.4.1 of the Results. We consider recurrently connected neuronal networks of N neurons. The firing rate ϕ_i (in units of Hz) for each neuron i for $0 \leq i < N-1$ is given as a function F (input-output relation) of the neuronal input $J_i(t)$:

$$\phi_i(t) = F(J_i(t)). \tag{4.17}$$

We further assume the input $J_i(t)$ to neuron i at time t to be given by the sum

$$J_i(t) = \sum_{j=0}^{N-1} w_{ij} s_j(t), \tag{4.18}$$

where w_{ij} is the connection weight for the synaptic connection from neuron j to neuron i. The variables $s_j(t)$ are synaptic activations of connections outgoing from the presynaptic neurons j, which depend on the firing rates of the neurons j.

To each neuron we assign an angular position $\theta_i = \frac{2\pi}{N} i - \pi \in [-\pi, \pi)$ (for $0 \leq i < N$), where we identify the edges of the interval to form a ring. Throughout this section, we assume that indices are taken modulo N. As introduced in detail in the first section below, we consider networks that maintain a family of steady-state firing rate profiles $\phi_{0,i}(\varphi_k)$ ("bumps") that are centered at positions $\varphi_k = \frac{2\pi}{N} k - \pi \in [-\pi, \pi)$ (for $0 \leq k < N$). Unless stated explicitly, derivations are performed for a bump centered at the left edge of the interval $\varphi_0 = -\pi$. If needed, we adopt the following notation to introduce the dependence on the center position: for vector quantities $\square$ we will write $\square(\varphi_k)$ to indicate that the vector is shifted to be centered at φ_k.

4.6.1.1 General definition of a one-dimensional ring-attractor

Starting from the definitions above, we here formalize the notion of a 1-dimensional continuous attractor manifold implemented in a neuronal network on the discretized ring of positions $\theta_i = \frac{2\pi}{N} i - \pi \in [-\pi, \pi)$ (for $0 \leq i < N$) with periodic boundaries. The central assumption made below is that there exists a 0 eigenvalue of the system dynamics, with an eigenvector corresponding to translation of the center of the firing rate profile. Similar to the approach in Burak and Fiete (2009), this is used in the following sections for dimensional reduction of the system dynamics.

A network implements a ring-attractor, if there exists a set $\mathscr{C}$, which is a family of firing rate profiles $\phi_i(\varphi_k)$ for neuron positions i and center positions φ_k:

$$\mathscr{C} = \left\{ \phi_i(\varphi_k) = F\left(J_i(\varphi_k)\right) | 0 \leq i, k < N \right\},$$

with the following properties. First, for each center position φ_k it holds that $\phi_{k+i}(\varphi_k) = \phi_{k-i}(\varphi_k)$ for $0 \leq i < N$. Thus, firing rate profiles centered at position φ_k are symmetric around φ_k. Second, we assume that $\phi_i(\varphi_k) = \phi_{i-k}(\varphi_0) \equiv \phi_{0,i-k}$. Thus, all members of $\mathscr{C}$ are center-shifted versions of the same firing rate profile $\vec{\phi}_0$, which is centered at index 0. Further, we will denote the vector of steady-state inputs associated to this firing rate profile by $\vec{\phi}_0 = F(\vec{J}_0)$. Note that for members of $\mathscr{C}$ with $\phi_i(\varphi_k) = F(J_i(\varphi_k))$ it holds that $J_i(\varphi_k) = J_{0,i-k}$. Second, all members of $\mathscr{C}$ are fixed points under the synaptic (and neuronal) dynamics. Third, when linearizing these dynamics around any member $\vec{\phi}(\varphi_k)$ of $\mathscr{C}$, we assume the Jacobian to have one single eigenvalue with value 0, with corresponding eigenvector proportional to $\vec{\phi}(\varphi_k) - \vec{\phi}(\varphi_{k-1})$, corresponding to a shift of the bump along the manifold (Burak and Fiete, 2009). All other eigenvalues are assumed to have negative real part, thus making $\mathscr{C}$ a family of meta-stable fixed points.

A “continuous” attractor on a discretized ring is defined by the above three assumptions: the family $\mathscr{C}$ consists of fixed points that are stable under the dynamics at hand, except for a change of center position by which the network state remains in the family $\mathscr{C}$. Heuristically speaking, any change of the center position will not be subject to any network-internal dynamics. Thus, in the limit of infinite network size, $\mathscr{C}$ is a 1-dimensional continuous center-manifold of the system (see e.g. Guckenheimer and Holmes (2013)). In the following, we will refer to the family $\mathscr{C}$ as a continuous manifold, even in the case of finite system size N.

4.6.1.2 Synaptic dynamics and variability of neural firing

To model synaptic dynamics with short-term plasticity, we assume that the dynamics of outgoing synaptic activations s_i of neuron i are modeled by the following three dimensional system of ordinary differential equations (Barak and Tsodyks, 2007; Mongillo et al., 2008;

Itskov et al., 2011b):

$$\begin{aligned}
\dot{s}_i &= -\frac{s_i}{\tau_s} + u_i x_i \phi_i, \\
\dot{u}_i &= -\frac{u_i - U}{\tau_u} + U(1-u_i)\phi_i, \\
\dot{x}_i &= -\frac{x_i - 1}{\tau_x} - u_i x_i \phi_i.
\end{aligned} \tag{4.19}$$

Synaptic activations s_i are given by a linear filter of the neuronal firing rates $\phi_i(t)$ multiplied by scaling factors $u_i(t)$ (facilitation) and $x_i(t)$ (depression), which modulate the synaptic transmission according to a rate-based variant of a widely used model of synaptic short-term plasticity (Markram et al., 1998; Tsodyks et al., 1998; Mongillo et al., 2008). See Section 4.4.3 in the Results for the biological interpretation of the scaling factors.

To be able to describe diffusion on the continuous attractor, we need to extend the model by a treatment of the noise induced into the system through the variable process of neuronal spike emission. Starting from Eq. (4.19), we assume that neurons i fire according to independent Poisson processes $\xi_i(t) = \sum_k \delta(t - t_{i,k})$, where $t_{i,k}$ is a Poisson point process with time-dependent rate ϕ_i. The variability of the point process $\xi_i(t)$ introduces noise in the synaptic variables. By neglecting the shot-noise (jump-like) nature of this process, we can capture the neurally induced variability simply as white noise with variance proportional to the incoming firing rates (Gerstner and Kistler, 2002), $\xi_i(t) = \phi_i + \sqrt{\phi_i} \cdot \eta_i(t)$, where η_i are white Gaussian noise processes with mean $\langle \eta_i \rangle = 0$, and correlation function $\langle \eta_i(t)\eta_j(t') \rangle = \delta(t-t')\delta_{ij}$. This model of $\xi_i(t)$ preserves the mean and the auto-correlation function of the original Poisson processes. Here, we introduce diffusive noise for each synaptic variable separately, but later average their *linear* contributions over the large population, when projecting onto movement along the continuous manifold (see also Burak and Fiete (2012, Supplementary Material) for a discussion).

Substituting the noisy processes $\xi_i(t)$ for $\phi_i(t)$ in Eq. (4.19) results in the following system of $3 \cdot N$ coupled Ito-SDEs:

$$\begin{aligned}
\dot{s}_i &= -\frac{s_i}{\tau_s} + u_i x_i \left(\phi_i + \eta_i \sqrt{\phi_i}\right), \\
\dot{u}_i &= -\frac{u_i - U}{\tau_u} + U(1-u_i)\left(\phi_i + \eta_i \sqrt{\phi_i}\right), \\
\dot{x}_i &= -\frac{x_i - 1}{\tau_x} - u_i x_i \left(\phi_i + \eta_i \sqrt{\phi_i}\right).
\end{aligned} \tag{4.20}$$

Note that the noise inputs η_i to the synaptic variables for neuron i are all identical, since they result from the same presynaptic spike train.

4.6.1.3 Linearized System

We start analyzing the system by defining a concatenated $3 \cdot N$ dimensional column vector of state variables $\mathbf{y} = \left(\mathbf{s}^T, \mathbf{u}^T, \mathbf{x}^T\right)^T$. Given a solution of the stable firing rate profile $\vec{\phi}_0$ we can calculate the deterministic stable fixed point of this system by setting the l.h.s. of Eq. (4.20) to zero and setting $\eta_i \equiv 0$. This yields steady-state solutions for the synaptic activations, facilitation and depression variables $\mathbf{y}_0 = (\mathbf{s}_0, \mathbf{u}_0, \mathbf{x}_0)$:

$$\begin{aligned} s_{0,i} &= \tau_s u_{0,i} x_{0,i} \phi_{0,i} \\ u_{0,i} &= U \frac{1 + \tau_u \phi_{0,i}}{1 + U \tau_u \phi_{0,i}} \qquad (4.21) \\ x_{0,i} &= \frac{1 + U \tau_u \phi_{0,i}}{1 + U\left(\tau_u \phi_{0,i} + \tau_u \tau_x \phi_{0,i}^2 + \tau_x \phi_{0,i}\right)} \qquad (4.22) \end{aligned}$$

We continue by linearizing the system Eq. (4.20) at the fixed point $\mathbf{y}_0$ by a change of variables consisting of perturbations around the fixed point: $\mathbf{y} = \mathbf{y}_0 + \delta\mathbf{y} = \mathbf{y}_0 + \left(\delta\mathbf{s}^T, \delta\mathbf{u}^T, \delta\mathbf{x}^T\right)$ and $\phi_i = \phi_{0,i} + \delta\phi_i$. To reach a self-consistent linear system, we further assume a separation of time scales between the neuronal dynamics and the synaptic conductances, in that the neuronal firing rates change as an immediate function of their input. This allows replacing $\delta\phi_i = \left.\frac{d\phi_i}{dJ_i}\right|_{J_{0,i}} \sum_j \frac{dJ_i}{ds_j} \delta s_j = \phi'_{0,i} \sum_j w_{ij} \delta s_j$, where we introduce the shorthand $\phi'_{0,i} \equiv \left.\frac{d\phi_i}{dJ_i}\right|_{J_{0,i}}$. Finally, keeping only linear orders in all perturbations and considering only the unperturbed noise (we neglect multiplicative noise terms by replacing the terms $\sqrt{\phi_i} \to \sqrt{\phi_{0,i}}$), we arrive at the linearized system equivalent of Eq. (4.20):

$$\begin{aligned} \dot{\delta\mathbf{y}} =& \begin{pmatrix} -\frac{1}{\tau_s}\mathbb{1} + D\left(\mathbf{u}_0 \cdot \mathbf{x}_0 \cdot \vec{\phi}'_0\right) W & D\left(\vec{\phi}_0 \cdot \mathbf{x}_0\right) & D\left(\vec{\phi}_0 \cdot \mathbf{u}_0\right) \\ UD\left((1 - \mathbf{u}_0) \cdot \vec{\phi}'_0\right) W & -\frac{1}{\tau_u}\mathbb{1} - UD\left(\vec{\phi}_0\right) & 0 \\ -D\left(\mathbf{u}_0 \cdot \mathbf{x}_0 \cdot \vec{\phi}'_0\right) W & -D(\mathbf{x}_0 \cdot \vec{\phi}_0) & -\frac{1}{\tau_x}\mathbb{1} - D\left(\vec{\phi}_0 \cdot \mathbf{u}_0\right) \end{pmatrix} \delta\mathbf{y} \\ &+ \begin{pmatrix} \vec{\eta}\mathbf{u}_0\mathbf{x}_0\sqrt{\vec{\phi}_0} \\ \vec{\eta}U(1 - \mathbf{u}_0)\sqrt{\vec{\phi}_0} \\ -\vec{\eta}\mathbf{u}_0\mathbf{x}_0\sqrt{\vec{\phi}_0} \end{pmatrix} \\ \equiv& K\delta\mathbf{y} + L \end{aligned} \tag{4.23}$$

Here, dots between vectors indicate element-wise multiplication, the operator $D : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ creates diagonal matrices from vectors, and $W = (w_{ij})$ is the synaptic weight matrix of the network. Note that the same vector of white noises $\vec{\eta} \equiv \left(\eta_1, \ldots, \eta_n\right)^T$ appears three times.

4.6.1.4 Projection onto the center manifold

Let the current center position of the firing rate profile be $\varphi = \varphi_0 = -\pi$. We will now project the dynamical system Eq. (4.23) onto movement of the center position φ of the firing rate profile. For this, we assume that N is large enough to treat the center position φ as a continuous variable. We also use the earlier assumption (see Section 4.6.1.1) that the network has a one dimensional center-manifold of metastable fixed points – this allows unconstrained displacement only of the center position φ, while all other possible directions of change are constrained by the system dynamics. In the system at hand, this implies that the matrix K of Eq. (4.23), which captures the linearized dynamics around any of these fixed points, will have a zero eigenvalue corresponding to the eigenvector of a change of the dynamical variables under a change of position φ.

Formally, let e_r be the vector of changes in the state variables as the steady-state center φ is translated along the manifold, given by:

$$e_r = \frac{d\mathbf{y}_0}{d\varphi} = \left(\frac{d\mathbf{s}_0}{d\varphi}^T, \frac{d\mathbf{u}_0}{d\varphi}^T, \frac{d\mathbf{x}_0}{d\varphi}^T\right)^T. \tag{4.24}$$

Let the full diagonalization[6] of the matrix K be $\bar{K} = T^{-1}KT$. W.l.o.g. let e_r be the first column of the transformation matrix T, and let e_l be the first row vector of T^{-1}, with $e_l \cdot e_r = 1$. Now, the projection of the perturbation $\delta\mathbf{y}$ onto contributions along e_r isolates displacements of the center position: the perturbation $\delta\mathbf{y} = e_r\delta\varphi$ corresponds to a shift $\varphi + \delta\varphi$ in the center position (by definition, see Eq. (4.24)). We left multiply this relation with e_l to see that $e_l\delta\mathbf{y} = e_l e_r \delta\varphi = \delta\varphi$. Finally, we take the time derivative, which yields:

$$\dot{\delta\varphi} = e_l\dot{\delta\mathbf{y}}. \tag{4.25}$$

Thus, to find an equation for the temporal dynamics of changes in the center positions, we project the system Eq. (4.23) onto contributions along the e_r eigenspace by left multiplying by e_l. In the full matrix equations, this corresponds to rewriting the system Eq. (4.23) as

$$T^{-1}\dot{\delta\mathbf{y}} = \bar{K}T^{-1}\delta\mathbf{y} + T^{-1}L.$$

Restricting this onto only the first dimension (the e_r eigenspace) and using Eq. (4.25), we find that the linear dynamics vanish (since the first entry in $\bar{K}$ is zero) and we are left with:

$$\dot{\varphi} = \dot{\delta\varphi} = e_l\dot{\delta\mathbf{y}} = 0 \cdot e_l\delta\mathbf{y} + e_l L, \tag{4.26}$$

where we have assumed that $\varphi(t) = \varphi(t=0) + \delta\varphi(t)$.

Thus, left-multiplying the linearized equations Eq. (4.23) by e_l yields a differential equation for the changes $\delta\varphi$ of the center position, leaving in this case only the contributions of the noisy input L. In the following two sections, we will compute the analytical form of the left eigenvector e_l.

---

[6] Or Jordan normal form, if K is not fully diagonalizable. Since we know there exists a zero eigenvalue with one dimensional Eigenspace, the corresponding normal form will have a diagonal 0 entry with e_r being the corresponding eigenvector.

4.6.1.5 Derivation of the left eigenvector

It holds that $e_l e_r = 1$ (since $T^{-1}T = I$), and also that e_l^T is an eigenvector of K^T with eigenvalue 0: since it holds that $T^{-1}K = \bar{K}T^{-1}$, we can project this relation on to the first dimension to get $e_l K = 0$, and equivalently $K^T e_l^T = 0$. If the matrix K is symmetric, the left and right eigenvectors e_l and e_r for the same eigenvalue 0 are the transpose of each other. Unfortunately, here this is not the case (see Eq. (4.23)), and we need to compute the unknown vector e_l, which will depend on the known vector e_r.

In the following we will find a parametrized vector $\mathbf{y}'(\mathbf{y}) = \left(\mathbf{t}^T(\mathbf{y}), \mathbf{v}^T(\mathbf{y}), \mathbf{z}^T(\mathbf{y})\right)^T$ that for $\mathbf{y} = e_r$ fulfills the eigenvalue equation of the left eigenvector:

$$K^T \mathbf{y}'(e_r) = 0. \tag{4.27}$$

As we show below, we can find variables $\mathbf{y}'$ that fulfill the transposed dynamics $\dot{\mathbf{y}}' = K^T \mathbf{y}'$ and for which it holds that $\dot{\mathbf{y}}'(e_r) = 0$, thus fulfilling the condition Eq. (4.27). In this case we know that (due to uniqueness of the 1-dimensional eigenspace associated to the 0 eigenvalue) the vector $\mathbf{y}'^T$ is proportional to e_l.

The system of equations resulting from $\dot{\mathbf{y}}' = K^T \mathbf{y}'$ reads (cf. Eq. (4.23)):

$$\begin{aligned} \dot{t}_i &= -\frac{t_i}{\tau_s} + \sum_j w_{ji} u_{0,j} x_{0,j} \phi_j^{'} t_j + U \sum_j w_{ji} (1 - u_{0,j}) \phi_j^{'} v_j - \sum_j w_{ji} u_{0,j} x_{0,j} \phi_j^{'} z_j && (4.28) \\ \dot{v}_i &= -\frac{v_i}{\tau_u} + \phi_{0,i} x_{0,i} (t_i - z_i) - U \phi_{0,i} v_i && (4.29) \\ \dot{z}_i &= -\frac{z_i}{\tau_x} + u_{0,i} \phi_{0,i} (t_i - z_i) && (4.30) \end{aligned}$$

Let us first consider the system without facilitation or depression. The linearized dynamics of $\delta\mathbf{s}$ are given by the upper left $N \times N$ block K_s of Eq. (4.23), with $U \to 1$, $u_0 \to 1, x_0 \to 1$ (which recovers the case of Burak and Fiete (2012)):

$$\begin{aligned} \delta \dot{s}_i &= -\frac{1}{\tau_s} \delta s_i + \phi'_{0,i} \sum_j w_{ij} \delta s_j && (4.31) \\ &\equiv K_s \delta \mathbf{s} && (4.32) \end{aligned}$$

We assume that $w_{ji} = w_{ij}$, which in general holds for the symmetric synaptic connectivity of common models of continuous attractor networks (see below a concrete spiking model). The transposed block K_s^T then describes the linear dynamics of perturbations to the input variables $J_i = \sum_j w_{ij} s_j$ (Burak and Fiete, 2012). To see this, we differentiate $\delta J_i = \sum_j w_{ij} \delta s_j$

with respect to time, use Eq. (4.31), and use the symmetric connectivity, to arrive at:

$$\begin{aligned}\delta \dot{J}_i &= \sum_j w_{ij} \delta \dot{s}_j = \sum_j w_{ij} \left(-\frac{1}{\tau_s} \delta s_j + \phi'_{0,j} \delta J_j \right) \\ &= \left(-\frac{1}{\tau_s} + \sum_j w_{ji} \phi'_{0,j} \right) \delta J_j \\ &= \left(K_s^T \delta \mathbf{J} \right)_i . \end{aligned} \tag{4.33}$$

If $\delta \dot{s}_i = 0$ for all i, which is the case if $\delta s_i = e_r$ restricted to the first N entries, we know that $\delta \dot{J}_i = \sum_j w_{ij} \delta \dot{s}_j = 0$. Thus, in this restricted case, the left eigenvector proportional to $t_i = \delta J_i$ since it fulfills Eq. (4.27) restricted to the first $N \times N$ block.

We now consider again the full system (with facilitation and depression), where we start with the same Ansatz for parametrization of the variables t_i [7]:

$$t_i \equiv \sum_j w_{ij} \delta s_j , \tag{4.34}$$

and continue to find variables v_i and z_i that satisfy the full equations Eq. (4.27).

First, we differentiate Eq. (4.34) with respect to time and use the linear response Eq. (4.23) to obtain

$$\begin{aligned}\dot{t}_i &= \sum_j w_{ij} \delta \dot{s}_j = \sum_j w_{ij} \left(-\frac{\delta s_j}{\tau_s} + \phi'_j u_{0,j} x_{0,j} \sum_k w_{jk} \delta s_k + \phi_{0,j} x_{0,j} \delta u_j + \phi_{0,j} u_{0,j} \delta x_j \right) \\ &= -\frac{t_i}{\tau_s} + \sum_j w_{ij} u_{0,j} x_{0,j} \phi'_j t_j + \sum_j w_{ij} \phi_{0,j} x_{0,j} \delta u_j + \sum_j w_{ij} \phi_{0,j} u_{0,j} \delta x_j . \end{aligned} \tag{4.35}$$

We then equate Eqs. (4.28) and (4.35), which yields the following identity:

$$\sum_j w_{ij} \phi_{0,j} \left(x_{0,j} \delta u_j + u_{0,j} \delta x_j \right) = \sum_j w_{ji} \phi'_j \left(U(1 - u_{0,j}) v_j - u_{0,j} x_{0,j} z_j \right). \tag{4.36}$$

To relate the remaining variables δu_i and δx_i to the new variables v_i and z_i, we make a linear Ansatz:

$$v_i = \alpha_1 \delta u_i + \alpha_2 \delta x_i , \tag{4.37}$$

$$z_i = \beta_1 \delta u_i + \beta_2 \delta x_i . \tag{4.38}$$

By differentiating these equations with respect to time and equating the result to Eqs. (4.29) and (4.30), respectively, we find the following equations

---

[7] This was motivated by numerical evaluations of the left eigenvector e_l of the full system, which showed that here also $t_i = \delta J_i$.

$$\begin{aligned} -\frac{v_i}{\tau_u} + \phi_{0,i} x_{0,i} (t_i - z_i) - U\phi_{0,i} v_i =& \alpha_1 \delta \dot{u}_i + \alpha_2 \delta \dot{x}_i && (4.39) \\ -\frac{z_i}{\tau_x} + u_{0,i}\phi_{0,i} (t_i - z_i) =& \beta_1 \delta \dot{u}_i + \beta_2 \delta \dot{x}_i && (4.40) \end{aligned}$$

The linear response for $\dot{u}_i$ and $\dot{x}_i$ can be obtained from Eq. (4.23) (substituting $t_i = \sum_j w_{ij}\delta s_j$):

$$\begin{aligned} \delta \dot{u}_i &= -\delta u_i \left(\frac{1}{\tau_u} - U\phi_{0,i} \right) + U(1 - u_{0,i})\phi_i' t_i, \\ \delta \dot{x}_i &= -\delta x_i \left(\frac{1}{\tau_x} - \phi_{0,i} \right) - x_{0,i}\phi_{0,i}\delta u_i - u_0 x_0 \phi_i' t_i. \end{aligned}$$

By inserting these two equations and Eqs. (4.37) and (4.38) into Eqs. (4.39) and (4.40), we obtain two closed equations in $\delta u_i, \delta x_i, t_i$. By equating coefficients for $\delta u_i, \delta x_i, t_i$[8], we obtain solutions for the coefficients $\alpha_1, \alpha_2, \beta_1, \beta_2$ that fulfill these equations:

$$\begin{aligned} \alpha_1 &= \frac{x_{0,i}\phi_{0,i}\left[U\tau_u\tau_x\phi_{0,i} + u_{0,i}(\tau_x - \tau_u)\right]}{U\left(1 - u_{0,i}\right)\phi_i'\left[\tau_u\tau_x\phi_{0,i}\left(U - u_{0,i}^2\right) + u_{0,i}(\tau_x - \tau_u)\right]} \\ \alpha_2 &= -\frac{\tau_u\tau_x u_{0,i}\phi_{0,i}^2}{\phi_i'\left[\tau_u\tau_x\phi_{0,i}\left(u_{0,i}^2 - U\right) + u_{0,i}(\tau_u - \tau_x)\right]} \\ \beta_1 &= -\frac{\tau_u\tau_x u_{0,i}\phi_{0,i}^2}{\phi_i'\left(\tau_u\tau_x\phi_{0,i}\left(u_{0,i}^2 - U\right) + u_{0,i}(\tau_u - \tau_x)\right)} \\ \beta_2 &= \frac{u_{0,i}\phi_{0,i}\left[\tau_u\left(\tau_x\phi_{0,i}\left(U - u_{0,i}\right) - 1\right) + \tau_x\right]}{x_{0,i}\phi_i'\left[\tau_u\tau_x\phi_{0,i}\left(u_{0,i}^2 - U\right) + u_{0,i}(\tau_u - \tau_x)\right]}. \qquad (4.41) \end{aligned}$$

A little bit of further algebra shows that these coefficients together with Eqs. (4.37) and (4.38) also satisfy Eq. (4.36), as for every j it holds that $\phi_j^{'}\left(U(1 - u_{0,j})v_j - u_{0,j}x_{0,j}z_j\right) = \phi_{0,j}\left(x_{0,j}\delta u_j + u_{0,j}\delta x_j\right)$.

Thus, we have found a linear parametrization

$$\mathbf{y}'(\mathbf{y})^T = \left((W\delta\mathbf{s})^T, \left(\alpha_1\delta\mathbf{u} + \beta_1\delta\mathbf{x}\right)^T, \left(\alpha_2\delta\mathbf{u} + \beta_2\delta\mathbf{x}\right)^T\right), \qquad (4.42)$$

which fulfills $\dot{\mathbf{y}}' = K^T\mathbf{y}'$. In addition, we know that if $\mathbf{y} = e_r$, then the original system dynamics vanish since $\dot{\mathbf{y}} = K\mathbf{y} = 0$. Thus, since the parametrization is linear in the original variables, it also holds that $\dot{\mathbf{y}}' = 0$, and the parametrization satisfies Eq. (4.27). This makes $\mathbf{y}'(e_r)^T$ *proportional* to the (unique) left eigenvector e_l of the 0-eigenvalue.

---

[8] Comparing coefficients of any two of the three variables yields 4 equations, which give the same solution and satisfy the equations posed by the remaining variable.

Finally, we can evaluate the vector $\mathbf{y}'(e_r)$ by using Eq. (4.24) in Eq. (4.42):

$$\mathbf{y}'(e_r)^T = \left(\frac{d\mathbf{J}_0}{d\varphi}^T, \left(\alpha_1 \frac{d\mathbf{u}_0}{d\varphi} + \alpha_2 \frac{d\mathbf{x}_0}{d\varphi}\right)^T, \left(\beta_1 \frac{d\mathbf{u}_0}{d\varphi} + \beta_2 \frac{d\mathbf{x}_0}{d\varphi}\right)^T\right). \tag{4.43}$$

Note, that in the first component we used that $\frac{dJ_{0,i}}{d\varphi} = \sum_j w_{ij} \frac{ds_{0,j}}{d\varphi}$.

4.6.1.6 Normalization of the left eigenvector

In the last section we have found a vector $\mathbf{y}'^T$ proportional to the left eigenvector e_l. Since it holds that $e_l e_r = 1$, it remains to calculate the normalization constant S such this normalization condition is fulfilled by

$$e_l = \frac{1}{S} \mathbf{y}'(e_r)^T. \tag{4.44}$$

First, we calculate the components of the vector e_r, Eq. (4.24), using the steady-state values of Eq. (4.21) and Eq. (4.22):

$$\frac{ds_{0,i}}{d\varphi} = \tau_s \left(u_{0,i} x_{0,i} \phi'_{0,i} \frac{dJ_{0,i}}{d\varphi} + x_{0,i} \phi_{0,i} \frac{du_{0,i}}{d\varphi} + u_{0,i} \phi_{0,i} \frac{dx_{0,i}}{d\varphi}\right), \tag{4.45}$$

$$\frac{du_{0,i}}{d\varphi} = u'_{0,i} \phi'_{0,i} \frac{dJ_{0,i}}{d\varphi} = \frac{(1-U)U\tau_u}{\left(U\phi_{0,i}\tau_u + 1\right)^2} \phi'_{0,i} \frac{dJ_{0,i}}{d\varphi}, \tag{4.46}$$

$$\frac{dx_{0,i}}{d\varphi} = x'_{0,i} \phi'_{0,i} \frac{dJ_{0,i}}{d\varphi} = -\frac{U\tau_x \left(\phi_{0,i}\tau_u \left(U\phi_{0,i}\tau_u + 2\right) + 1\right)}{\left(U\phi_{0,i}\left(\tau_u + \phi_{0,i}\tau_u\tau_x + \tau_x\right) + 1\right)^2} \phi'_{0,i} \frac{dJ_{0,i}}{d\varphi}, \tag{4.47}$$

where we introduced the shorthand notations $u'_{0,i} = \frac{du_{0,i}}{d\phi_{0,i}}$, $x'_{0,i} = \frac{dx_{0,i}}{d\phi_{0,i}}$ and $\phi'_{0,i} = \left.\frac{d\phi_i}{dJ_i}\right|_{J_{0,i}}$. We have also used the steady-state values of Eq. (4.21) and Eq. (4.22) to calculate the values of $u'_{0,i}$ and $x'_{0,i}$ by differentiating with respect to $\phi_{0,i}$.

Now, using Eq. (4.44) in $e_l \cdot e_r = 1$, and plugging in Eqs. (4.24) and (4.43), we find that:

$$\begin{aligned}
S =& \mathbf{y}'(e_r)^T \cdot e_r \\
=& \tau_s \sum_i \frac{dJ_{0,i}}{d\varphi} \left(u_{0,i} x_{0,i} \phi'_{0,i} \frac{dJ_{0,i}}{d\varphi} + x_{0,i}\phi_{0,i} \frac{du_{0,i}}{d\varphi} + u_{0,i}\phi_{0,i} \frac{dx_{0,i}}{d\varphi} \right) \\
&+ \sum_i \frac{du_{0,i}}{d\varphi} \left(\alpha_1 \frac{du_{0,i}}{d\varphi} + \alpha_2 \frac{dx_{0,i}}{d\varphi} \right) + \sum_i \frac{dx_{0,i}}{d\varphi} \left(\beta_1 \frac{du_{0,i}}{d\varphi} + \beta_2 \frac{dx_{0,i}}{d\varphi} \right) \qquad (4.48) \\
=& \tau_s \sum_i \left(\frac{dJ_{0,i}}{d\varphi} \right)^2 \phi'_{0,i} \left(u_{0,i} x_{0,i} + x_{0,i} \phi_{0,i} u'_{0,i} + u_{0,i} \phi_{0,i} x'_{0,i} \right) \\
&+ \sum_i \left(\frac{dJ_{0,i}}{d\varphi} \right)^2 \phi'^2_{0,i} \left[\alpha_1 u'^2_{0,i} + \beta_2 x'^2_{0,i} + 2\alpha_2 u'_{0,i} x'_{0,i} \right] \\
=& U \sum_i \frac{\left(\frac{dJ_{0,i}}{d\varphi} \right)^2 \phi'_i}{\left[U\phi_{0,i} \left(\tau_u \left(\tau_x \phi_{0,i} + 1 \right) + \tau_x \right) + 1 \right]^3} \\
&\Bigg[\tau_s \left[\tau_u \phi_{0,i} \left(U \tau_u \phi_{0,i} + 2 \right) + 1 \right] \left[U\phi_{0,i} \left(\tau_u \left(\tau_x \phi_{0,i} + 1 \right) + \tau_x \right) + 1 \right] \\
&- \phi_{0,i} \left[(U-1)\tau_u^2 + U\tau_x^2 \left(\tau_u \phi_{0,i} + 1 \right) \left(\tau_u \phi_{0,i} \left(U\tau_u \phi_{0,i} + 2 \right) + 1 \right) \right] \qquad (4.49) \\
&- \frac{(U-1)U\tau_u^2 \tau_x \phi_{0,i} \left(\tau_u \phi_{0,i} + 1 \right)}{\left(U\tau_u \phi_{0,i} + 1 \right)} \Bigg], \qquad (4.50)
\end{aligned}$$

which defines the normalization constant S. In the last equation we used the steady-state values of Eq. (4.21) and Eq. (4.22) to calculate the values of $u'_{0,i}$ and $x'_{0,i}$ by differentiating with respect to $\phi_{0,i}$. Additionally, the coefficients of Eq. (4.41) were used.

4.6.2 Drift and diffusion in presence of short-term plasticity

In the last section, we have derived a vector (Eq. (4.44)) which projects the synaptic dynamics Eq. (4.19) onto movement along the attractor manifold, according to Eq. (4.26). **Here, will apply this projection vector to calculate the central result presented in Section 4.4.3 of the main text** – a reduction of the microscopic dynamics of continuous attractor networks to a one-dimensional Langevin equation for the center position of the bump profile (Eq. (4.3)). First, we calculate the diffusion term (characterized by a diffusion strength B), by projecting the diffusive noise resulting from variable spike emission (introduced in Section 4.6.1.2) onto the attractor manifold. Second, we project fixed deviations from the steady-state firing rate profiles onto the attractor manifold, which results in an additional position-dependent drift field $A(\varphi)$. The results of this section thereby provide a link from the full synaptic dynamics with short-term plasticity to the emerging one-dimensional dynamics of the bump center.

4.6.2.1 Diffusion strength

We use the expression for the left eigenvector Eq. (4.44) to project the individual noise contributions in Eq. (4.23) onto movement along the attractor manifold. Restating Eq. (4.26), we know that $\dot{\varphi} = e_l L = \sum_k e_{l,k} L_k$. Through the normalization by S, which sums over all neurons, the individual contributions $e_{l,k}$ become small as the number of neurons N increases (this scaling is made explicit in Section 4.6.5.4). Thus, for large networks we average the small contributions of many single noise sources, which validates the diffusion approximation of Section 4.6.1.2.

Further, we can rewrite Eq. (4.26) by introducing a single Gaussian white noise process with intensity B, that matches the correlation function of the summed noises:

$$\dot{\varphi} = \sqrt{B}\eta. \tag{4.51}$$

Here, η is a white Gaussian noise process with $\langle \eta \rangle = 0$ and $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$. To calculate the correlation function of Eq. (4.26), we first note that only terms $\langle \eta_i(t)\eta_i(t+\tau) \rangle = \delta(\tau)$ remain in expectation. Thus, we arrive at an expression for B in Eq. (4.5) of the Results section:

$$\begin{aligned}
\left\langle \dot{\varphi}(t)\dot{\varphi}(t+\tau) \right\rangle =& \langle e_l L(t) e_l L(t+\tau) \rangle \\
=& \sum_{i=1}^{n} \left(e_{l,i}^2 u_{0,i}^2 x_{0,i}^2 + e_{l,n+i}^2 U^2 (1 - u_{0,i})^2 + e_{l,2n+i}^2 u_{0,i}^2 x_{0,i}^2 \right) \phi_{0,i} \delta(\tau) \\
& + 2 \sum_{i=1}^{n} \left[\left(e_{l,i} e_{l,n+i} - e_{l,n+i} e_{l,2n+i} \right) U (1 - u_{0,i}) - e_{l,i} e_{l,2n+i} u_{0,i} x_{0,i} \right] u_{0,i} x_{0,i} \phi_{0,i} \delta(\tau) \\
=& \frac{1}{S^2} \sum_i \left(\frac{dJ_{0,i}}{d\varphi} \right)^2 \phi_{0,i} \left[u_{0,i}^2 x_{0,i}^2 + U^2 (1 - u_{0,i})^2 \phi_{0,i}'^2 \left(\alpha_1 u_{0,i}' + \alpha_2 x_{0,i}' \right)^2 \right. \\
& \qquad \left. + u_{0,i}^2 x_{0,i}^2 \phi_{0,i}'^2 \left(\beta_1 u_{0,i}' + \beta_2 x_{0,i}' \right)^2 \right] \delta(\tau) \\
& + \frac{2}{S^2} \sum_i \left(\frac{dJ_{0,i}}{d\varphi} \right)^2 \phi_{0,i} u_{0,i} x_{0,i} U (1 - u_{0,i}) \\
& \qquad \left[\phi_{0,i}' \left(\alpha_1 u_{0,i}' + \alpha_2 x_{0,i}' \right) - \phi_{0,i}'^2 \left(\alpha_1 u_{0,i}' + \alpha_2 x_{0,i}' \right) \left(\beta_1 u_{0,i}' + \beta_2 x_{0,i}' \right) \right] \delta(\tau) \\
& - \frac{2}{S^2} \sum_i \left(\frac{dJ_{0,i}}{d\varphi} \right)^2 \phi_{0,i} u_{0,i}^2 x_{0,i}^2 \phi_{0,i}' \left(\beta_1 u_{0,i}' + \beta_2 x_{0,i}' \right) \delta(\tau) \\
=& \frac{U^2}{S^2} \sum_i \frac{\left(\frac{dJ_{0,i}}{d\varphi} \right)^2 \phi_{0,i} \left(1 + 2\tau_u \phi_{0,i} + U \tau_u^2 \phi_{0,i}^2 \right)^2}{\left(U \phi_{0,i} \left(\tau_u \left(\tau_x \phi_{0,i} + 1 \right) + \tau_x \right) + 1 \right)^4} \delta(\tau) \\
\equiv& B \delta(\tau).
\end{aligned} \tag{4.52}$$

In the last equation we again have used the steady-state values of Eq. (4.21) and Eq. (4.22) to calculate the values of $u_{0,i}'$ and $x_{0,i}'$ by differentiating with respect to $\phi_{0,i}$. Additionally, the coefficients of Eq. (4.41) were used.

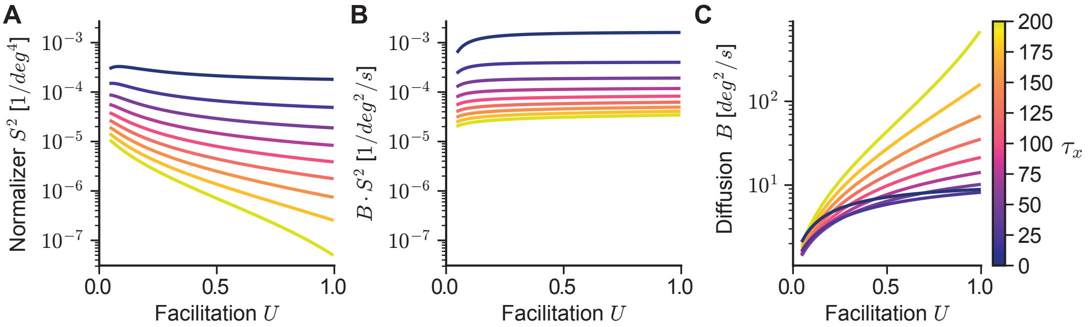


Figure 4.19 – **Numerical evaluation of diffusion strength.** All plots show quantities related to Eq. (4.52) for varying facilitation U and depression time constant τ_x (color legend on the right hand side shows values of τ_x in units of ms). Other coefficients $\phi_{0,i}$, $\frac{dJ_{0,i}}{d\varphi}$, $\phi'_{0,i}$ are estimated from a network with $U = 1, \tau_x = 150ms, \tau_u = 650ms$. **A** Normalization term S^2 ("Normalizer") of Eq. (4.52). **B** Diffusion strength B of Eq. (4.52) without the normalization term (equal to $B \cdot S^2$). **C** Full diffusion strength B of Eq. (4.52). Color legend on the right hand side shows values of τ_x in units of ms.

Here, the value of B is computed for $\varphi = \varphi_0$, but holds generally: for any other center position all vectors appearing in Eq. (4.52) are symmetrically index-shifted versions of the vectors for $\varphi = \varphi_0$, thus leaving the sum invariant.

Analysis of diffusion strength In Fig. 4.19 we plot the diffusion strength Eq. (4.52) and the value of the normalization term S^2 for varying τ_x and U. For a numerical evaluation including the facilitation time constant τ_u, we refer to Fig. 4.6. We observe that for $\tau_x \to 0$ diffusion is reduced to a large degree (Fig. 4.19C). Thus, decreasing the depression time constant is one method of reducing the diffusivity of these attractor networks, assuming that the same bump firing rates can be kept as a stable state of the network dynamics as depression is removed. Since some saturating nonlinearity of the excitatory feedback will be needed for the network to display stable persistent firing at low rates (Wang, 1999), this limit is probably hard to achieve. It should be noted, that using the same methods used to tune all spiking networks (see Section 4.6.6), we were able to find parameters that reliably lead to bistable bump states with $U = 1$, $\tau_u = 650ms$, $\tau_x = 15ms$ and bump shape similar to the other networks presented here (parameters were $g_{EE} = 9.975pS$, $g_{IE} = 6.724pS$, $g_{EI} = 2.213nS$, $g_{II} = 1.670nS$, as well as $w_\sigma = 0.425$, $w_+ = 2.95$).

We also find (see also Fig. 4.5C of the Results) that diffusion increases quickly as τ_x moves towards $\tau_x = 200ms$. In the remainder of this section we will show that this behavior is a general feature of the theory, and is due to the normalization constant S tending to become small for $U \to 1$ and large τ_x (see Fig. 4.19A).

To analyze this behavior further, we set $U = 1$ in the normalization constant S (Eq. (4.50)),

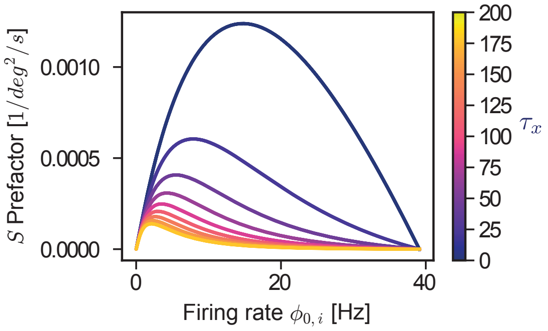


Figure 4.20 – **Pre-factor of normalizer for vanishing facilitation.** Pre-factor $\frac{dJ_{0,i}}{d\varphi}^2 \phi'_{0,i} \frac{1}{(1+\phi_{0,i}\tau_x)^3}$ for varying firing rates $\phi_{0,i}$ and depression time constant τ_x. Color legend on the right hand side shows values of τ_x in units of ms.

which yields:

$$S = \sum_i \frac{dJ_{0,i}}{d\varphi}^2 \phi'_{0,i} \frac{\left(\phi_{0,i}\tau_s\tau_x - \phi_{0,i}\tau_x^2 + \tau_s\right)}{(1+\phi_{0,i}\tau_x)^3}. \tag{4.53}$$

Inspecting Eq. (4.53), we find that S will be close to zero if $\phi_{0,i}\tau_s\tau_x - \phi_{0,i}\tau_x^2 + \tau_s \approx 0$ for all summands. Solving this equation for τ_x yields a single positive solution

$$\tau_{x,c} = \frac{1}{2}\left(\tau_s + \sqrt{\frac{\tau_s(\phi_{0,i}\tau_s + 4)}{\phi_{0,i}}}\right), \tag{4.54}$$

which approaches $\tau_{x,c} \to \tau_s$ for $\phi_{0,i} \to \infty$. Thus, summands of Eq. (4.53) with large firing rates will vanish for $\tau_x > \tau_s$. In addition, the range of firing rates that will result in nonzero summands in Eq. (4.53) is limited: the pre-factors $\frac{dJ_{0,i}}{d\varphi}^2 \phi'_{0,i}$ in Eq. (4.53) will vanish for $\phi_{0,i} \to 0$ and $\phi_{0,i} \to \max_i \phi_{0,i}$, since in both cases $\frac{dJ_{0,i}}{d\varphi} = 0$. Plotting the pre-factor $\frac{dJ_{0,i}}{d\varphi}^2 \phi'_{0,i} \frac{1}{(1+\phi_{0,i}\tau_x)^3}$ against the firing rate $\phi_{0,i}$ for varying τ_x (see Fig. 4.20), we see that as τ_x increases beyond $50ms$, the range of firing rates with positive contributions to the sum quickly decays to between 0 and 15Hz, with maxima between 2Hz and 5.5Hz.

Evaluating Eq. (4.54) for $\tau_s = 100ms$ at these firing rates, the corresponding values of $\tau_{x,c}$ are $279.1ms$ (for 2Hz) and $193.8ms$ (for 5.5Hz). This illustrates that, as τ_x approaches $200ms$, the normalization term S is closer to zero as more summands vanish, which leads to diverging diffusion strengths B by Eq. (4.52). While the values at which an increase of τ_x leads to this behavior depend on the coefficients ϕ_0 and $\frac{dJ_{0,i}}{d\varphi}$, the existence of such a divergence is a general feature of this system due to the structure of Eq. (4.53), and will be present for any stable distribution of firing rates. Thus, any bump attractor will observe escalating diffusion as τ_x is increased.

4.6.2.2 Drift field

Let the current center position of the bump be φ_k (for $0 \leq k < N$), with the index shifted firing rate profile $\vec{\phi}_0(\varphi_k)$. To calculate the effect of frozen variability, for now we assume that the effect of all sources of frozen variability can be expressed by a vector of small firing rate perturbations $\Delta\phi(\varphi_k)$ to the firing rate profile: $\vec{\phi}_0(\varphi_k) \rightarrow \vec{\phi}_0(\varphi_k) + \Delta\vec{\phi}(\varphi_k)$. These firing rate perturbations stem from any deviation of the neural system from the "baseline" case: in Section 4.6.5 we calculate the perturbations induced by sparse connectivity as well as heterogeneous leak reversal-potentials in excitatory neurons of the spiking network.

The perturbations will change with the center position φ_k of the bump, and the resulting drift field will depend on the center position around which we linearize the dynamics. For simplicity of notation, we let the steady-state quantities remain centered at $\varphi = \varphi_0 = -\pi$ (e.g. we consider the firing rate profile $\vec{\phi}_0$ centered at φ_0), and absorb the dependence on the position φ_k into the perturbations:

$$\phi_{0,i} \rightarrow \phi_{0,i} + \Delta\phi_i(\varphi_k).$$

For simplicity, we will neglect the index k for the position in the following.

The firing rate perturbations add an additional term in the linearized equations Eq. (4.23), where we again neglect the influence of the perturbation on the noise term L:

$$\delta\mathbf{y} = K\delta\mathbf{y} + \begin{pmatrix} \mathbf{x}_0\mathbf{u}_0\Delta\vec{\phi}(\varphi) \\ U(1-\mathbf{u}_0)\Delta\vec{\phi}(\varphi) \\ -\mathbf{x}_0\mathbf{u}_0\Delta\vec{\phi}(\varphi) \end{pmatrix} + L. \tag{4.55}$$

Again, we left-multiply by the left eigenvector e_l, which eliminates the linear response kernel K and yields a drift-term in the formerly purely diffusive SDE Eq. (4.51):

$$\begin{aligned}
\dot{\varphi} &= e_l \begin{pmatrix} \mathbf{x}_0\mathbf{u}_0\Delta\vec{\phi}(\varphi) \\ U(1-\mathbf{u}_0)\Delta\vec{\phi}(\varphi) \\ -\mathbf{x}_0\mathbf{u}_0\Delta\vec{\phi}(\varphi) \end{pmatrix} + \sqrt{B}\eta \\
&= \frac{1}{S}\sum_i \left[\frac{dJ_{0,i}}{d\varphi} x_{0,i}u_{0,i} + \left(\alpha_1 \frac{du_i}{d\varphi} + \alpha_2 \frac{dx_i}{d\varphi}\right) U\left(1 - u_{0,i}\right) \right. \\
&\qquad \left. - \left(\beta_1 \frac{du_i}{d\varphi} + \beta_2 \frac{dx_i}{d\varphi}\right) x_{0,i}u_{0,i} \right] \Delta\phi_i(\varphi) + \sqrt{B}\eta \\
&= \frac{U}{S}\sum_i \frac{dJ_{0,i}}{d\varphi} \frac{1 + \tau_u\phi_{0,i}\left(U\tau_u\phi_{0,i} + 2\right)}{\left(U\phi_{0,i}\left(\tau_u\tau_x\phi_{0,i} + \tau_u + \tau_x\right) + 1\right)^2} \Delta\phi_i(\varphi) + \sqrt{B}\eta \\
&\equiv A(\varphi) + \sqrt{B}\eta,
\end{aligned} \tag{4.56}$$

which defines the drift field $A(\varphi)$ of Eq. (4.7). In the last equation we have used Eqs. (4.46) and (4.47), as well as Eq. (4.41).

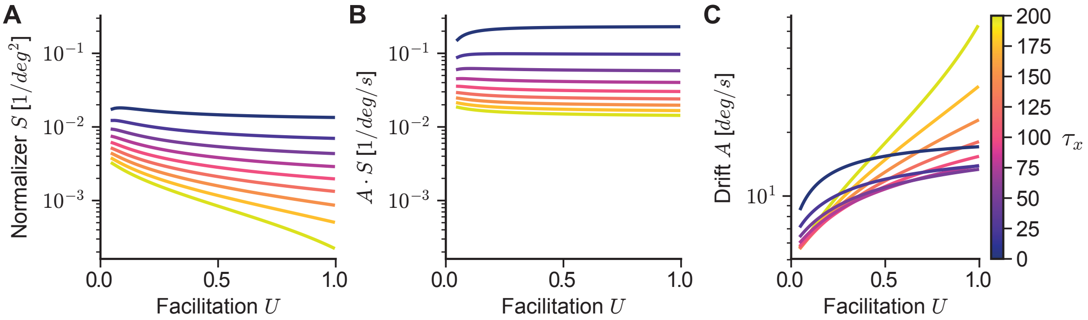


Figure 4.21 – **Numerical evaluation of drift field.** All plots show quantities related to Eq. (4.56) for varying facilitation U and depression time constant τ_x (color legend on the right hand side shows values of τ_x in units of ms). The firing rate perturbation was fixed to $\Delta\phi_i = 0.1\phi_{0,i}$ for neurons in the right half of the bump (leading to a positive drift) and 0 otherwise. Other coefficients $\phi_{0,i}, \frac{dJ_{0,i}}{d\varphi}, \phi'_{0,i}$ are estimated from a network with $U = 1, \tau_x = 150ms, \tau_u = 650ms$. **A** Normalization constant ("Normalizer") S of Eq. (4.56). **B** Drift field A of Eq. (4.56) without the normalization constant (equal to $A \cdot S$). **C** Full drift term A of Eq. (4.56). Color legend on the right hand side shows values of τ_x in units of ms.

Analysis of drift field In the limit of vanishing short-term plasticity ($U \to 1, \tau_x \to 0$) the drift term of Eq. (4.56) reduces to $A = \left(\sum_i \frac{dJ_{0,i}}{d\varphi} \Delta\phi_i(\varphi)\right) / \left(\tau_s \sum_i \frac{dJ_{0,i}}{d\varphi}^2 \phi'_{0,i}\right)$, which extends the framework of Burak and Fiete (2012) to frozen perturbations. For general parameters of short-term plasticity, we plot in Fig. 4.21 the drift term Eq. (4.56) and the value of the normalization constant S in Fig. 4.21 for a fixed perturbation. For evaluations including the facilitation time constant τ_u see Section 4.4.6.2.

As before for the diffusion strength (see Section 4.6.2.1), we find that increasing the diffusion time constant τ_x leads to increased drifts (Fig. 4.21). Although the rest of the non-normalized summands become small (Fig. 4.21B), vanishing values of the normalization constant S lead to this behavior (Fig. 4.21A). A similar argument as in Section 4.6.2.1 shows that this is a general feature of the normalization constant S, which becomes small as τ_x approaches a critical, firing rate dependent, value $\tau_{x,c}$ (cf. Eq. (4.54)).

4.6.3 Spiking network model

Spiking simulations are based on a variation of a popular ring-attractor model of visuospatial working memory (Compte et al., 2000) (and used with variations in Renart et al. (2003); Murray et al. (2012); Wei et al. (2012); Pereira and Wang (2015); Almeida et al. (2015); see also Section 3.6.2 in the previous chapter). The recurrent excitatory connections of the original network model have been simplified, to allow for faster simulation as well as analytical derivations of the recurrent synaptic activation. The implementation details are given below, however the major changes are: 1) all recurrent excitatory conductances are voltage independent; 2) a popular model of synaptic short-term plasticity via facilitation and depression (Markram et al., 1998; Tsodyks et al., 1998; Mongillo et al., 2008) is used to dynamically regulate the weights of the incoming spike-trains 3) recurrent excitatory conductances are computed as linear filters of the weighted incoming spike trains instead of the second-order kinetics for NMDA saturation used in Compte et al. (2000). The inclusion of short-term depression effectively leads to a saturation of the recurrent excitatory transmission at high firing rates (see below).

4.6.3.1 Neuron model

Neurons are modeled by leaky integrate-and-fire dynamics with conductance based synaptic transmission (Compte et al., 2000; Brunel and Wang, 2001b). The network consists of recurrently connected populations of N_E excitatory and N_I inhibitory neurons, both additionally receiving external spiking input with spike times generated by N_{ext} independent, homogeneous Poisson processes, with rates ν_{ext}. We assume that external excitatory inputs are mediated by fast AMPA receptors, while, for simplicity, recurrent excitatory currents are mediated only by slower NMDA channels. While some of the definitions below are similar to those of Section 3.6.2, they are restated here for completeness.

The dynamics of neurons in both excitatory and inhibitory populations are governed by the following system of differential equations indexed by $i \in \{0, ..., N_{E/I} - 1\}$:

$$\begin{aligned} C_m \dot{V}_i(t) &= -I_i^{\text{L}}(t) - I_i^{\text{Ext}}(t) - I_i^{\text{I}}(t) - I_i^{\text{E}}(t), \\ I_i^P &= g_P \, s_i^P(t) \, (V_i(t) - V_P), \end{aligned} \tag{4.57}$$

where $P \in$ {L,Ext,I,E}, V denotes voltages (membrane potential) and I denotes currents. Here, C_{m} is the membrane capacitance and $V_{\text{L}}, V_{\text{E}}, V_{\text{I}}$ are the reversal potentials for leak, excitatory currents, and inhibitory currents, respectively. The parameters g_P for $P \in$ {L,Ext,I,E} are fixed scales for leak (L), external input (Ext) and recurrent excitatory (E) and inhibitory (I) synaptic conductances, which are dynamically gated by the unit-less gating variables $s_i^P(t)$. These gating variables are described in detail below, however we set the leak conductance gating variable to $s_i^{\text{L}} = 1$. For excitatory neurons, we refer to the excitatory and inhibitory conductance scales by $g_{\text{EE}} \equiv g_{\text{E}}$ and $g_{\text{EI}} \equiv g_{\text{I}}$, respectively. Similarly, for inhibitory neurons, we refer to the excitatory and inhibitory conductance scales by $g_{\mathrm{IE}} \equiv g_{\mathrm{E}}$ and $g_{\mathrm{II}} \equiv g_{\mathrm{I}}$, respectively.

The model neuron dynamics (Eq. 4.57) are integrated until their voltage reaches a threshold V_{thr}. At any such time, the respective neuron emits a spike and its membrane potential is reset to the value V_{res}. After each spike, voltages are clamped to V_{res} for a refractory period of τ_{ref}. See Table D.1 for parameter values used in simulations.

4.6.3.2 Synaptic gating variables and short-term plasticity

The unit-less synaptic gating variables $s_i^P(t)$ for $P \in \{\mathrm{Ext,I}\}$ (external and inhibitory currents) are exponential traces of the spike trains of all presynaptic neurons j with firing times t_j:

$$\dot{s}_i^P(t) = -\frac{s_i^P(t)}{\tau_P} + \sum_{j\in\mathrm{pre}(P)} w_{ij}^P \sum_{t_j} \delta\left(t - t_j\right), \tag{4.58}$$

where pre(P) indicates all neurons presynaptic to the neuron i for the the connection type P. The factors w_{ij}^P are unit-less synaptic efficacies for the connection from neuron j to neuron i. For the excitatory gating variables of inhibitory neurons s_i^{IE} (IE denotes connections from E to I neurons) we also use the linear model of Eq. (4.58) with time constant $\tau_{IE} = \tau_E$.

For excitatory to excitatory conductances, we use a well established model of synaptic short-term plasticity (STP) (Tsodyks et al., 1998; Markram et al., 1998; Mongillo et al., 2008) which provides dynamic scaling of synaptic efficacies depending on presynaptic firing. This yields two additional dynamical variables, the facilitating synaptic efficacy $u_j(t)$, as well as the fraction of available synaptic resources $x_j(t)$ of the outgoing connections of a presynaptic neuron j, which are implemented according to the following differential equation:

$$\begin{aligned}
\dot{u}_j &= -\frac{1}{\tau_u}\left(u_j - U\right) + U\left(1 - u_j^-\right)\sum_{t_j}\delta(t - t_j), \\
\dot{x}_j &= -\frac{1}{\tau_x}\left(x_j - 1\right) - x_j^- u_j^- \sum_{t_j}\delta(t - t_j).
\end{aligned} \tag{4.59}$$

Here, the indices u_j^- and x_j^- indicate that for the incremental update of the variables upon spike arrival, the values of the respective variables immediately before the spike arrival are used (Mongillo et al., 2008). The dynamics of recurrent excitatory-to-excitatory transmission with STP are then given by gating variables that linearly filter the incoming spikes, which are scaled by facilitation and depression:

$$s_i^{\mathrm{EE}}(t) = \sum_{j\in\mathrm{pre(EE)}} w_{ij}^{\mathrm{EE}} s_j \tag{4.60}$$

$$\dot{s}_j = -\frac{s_j}{\tau_{\mathrm{E}}} + \sum_{t_j}\delta\left(t - t_j\right) u_j^-(t) x_j^-(t). \tag{4.61}$$

Here, pre(EE) indicates all excitatory neurons that make synaptic connections to the neuron i.

This makes the system Eqs. (4.59) and (4.61) a spiking variant of the rate-based dynamics of Eq. (4.19), and s_i^{EE} a variable related to the input J_i (cf. Eq. (4.18)). In Section 4.6.4 we will make this link explicit.

4.6.3.3 Network connectivity

All connections except for the recurrent excitatory connections are all-to-all and uniform, with unit-less connection strengths set to $w_{ij}^{\mathrm{I}} = w_{ij}^{\mathrm{ext}} = 1$ and for inhibitory neurons additionally $w_{ij}^{\mathrm{E}} = 1$. The recurrent excitatory connections are distance dependent. Each neuron of the excitatory population with index $i \in \{0, \dots, N_E - 1\}$ is assigned an angular position $\theta_i = i \cdot \frac{2\pi}{N_E} \in [0, 2\pi)$. Recurrent excitatory connection weights w_{ij}^{EE} from neuron j to neuron i are then given by the Gaussian function $w^{\mathrm{EE}}(\theta)$ as

$$w_{ij}^{\mathrm{EE}} = w^{\mathrm{EE}}(\theta_i - \theta_j) = w_0 + (w_+ - w_0) \exp\left(- \left[\min\left(\left|\theta_i - \theta_j\right|, 2\pi - \left|\theta_i - \theta_j\right| \right) \right]^2 \frac{1}{2\sigma_w^2} \right).$$

Additionally, for each neuron we keep the integral over all recurrent connection weights normalized, resulting in the normalization condition $\frac{1}{2\pi} \int_{-\pi}^{\pi} d\varphi \, w^{\mathrm{EE}}(\varphi) = 1$. This normalization ensures that varying the maximum weight w_+ will not change the total recurrent excitatory input if all excitatory neurons fire at the same rate. Here, we choose w_+ as a free parameter constraining the baseline connection weight to:

$$w_0 = \frac{w_+ \sigma_w \operatorname{erf}\left(\frac{\pi}{\sqrt{2}\sigma_w} \right) - \sqrt{2\pi}}{\sigma_w \operatorname{erf}\left(\frac{\pi}{\sqrt{2}\sigma_w} \right) - \sqrt{2\pi}}.$$

4.6.4 Firing rate approximation of the spiking network

In this section we perform a firing rate approximation for spiking neurons, similar to that of Section 3.6.3 in the last chapter, allowing us to relate the theory of Section 4.6.1 to the spiking network. The spiking dynamics here are simpler, since we do *not* have saturating second-order NMDA kinetics and all conductances are voltage independent. Nevertheless, the calculation of the mean steady-state input J_i will require the approximation of the steady states of short-term plasticity, in dependence on the mean firing rate of neurons (Section 4.6.4.1). This allows us to calculate the prediction of mean firing rates F (Section 4.6.4.2) in dependence of the steady-state input J_i, as well as derivatives of firing rates with respect to perturbations (Section 4.6.4.3), which appear in Eq. (4.50), Eq. (4.56) and related terms.

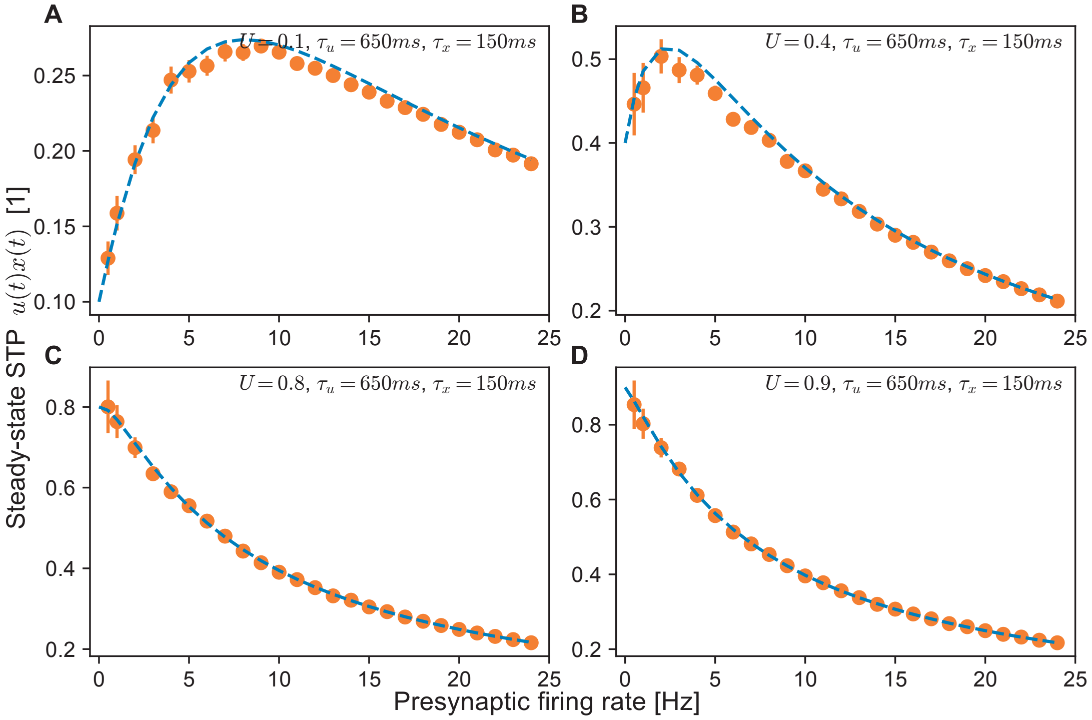


Figure 4.22 – **Steady-state approximation of short-term plasticity.** Analytical predictions $\langle u_j x_j \rangle$ as functions of the presynaptic firing rate ν_j (cf. Eq. (4.66)) compared to simulations by the temporal average of synaptic activation $\frac{\langle s_j \rangle_t}{\tau_E \nu_j}$. Presynaptic neurons fire with spike spike counts distributed as Poisson processes with mean rates ν_j. After an initial transient of $2s$, which was discarded, the temporal mean $\langle s_j \rangle_t$ averaged over $3s$ simulation time was recorded. Dots and error bars are mean and 95% confidence of the mean of $\frac{\langle s_j \rangle_t}{\tau_E \nu_j}$ recorded over 300 realizations of presynaptic Poisson spike trains.

4.6.4.1 Mean synaptic activations

Here we calculate mean synaptic activations, which will be used in Section 4.6.6 to approximate the bump shapes in networks of spiking neurons. We first replace the synaptic activation variables $s^P(V,t)$ for $P \in \{\mathrm{I}, \mathrm{ext}\}$ by their expectation values under input with Poisson statistics. We assume that the inhibitory population fires at rates ν_I. For the linear synapses this yields

$$\left\langle s^{\mathrm{ext}} \right\rangle = \tau_{\mathrm{ext}} N_{\mathrm{ext}} \nu_{\mathrm{ext}}, \tag{4.62}$$

$$\left\langle s^{\mathrm{I}} \right\rangle = \tau_{\mathrm{I}} N_I \nu_{\mathrm{I}}. \tag{4.63}$$

For the recurrent excitatory-to-excitatory synapses with short-term plasticity, we set the differential equations (4.59) to zero, and also average them over the Poisson statistics. Akin to the “mean-field” model of Tsodyks et al. (1998), we average the steady-state values of facilitation and depression separately over the Poisson statistics. This implicitly assumes that facilitation and depression are statistically independent, with respect to the distributions of spike times – while this is not strictly true, the approximations work well, as has been previously reported (Tsodyks et al., 1998). This allows a fairly straightforward evaluation of the mean steady-state value of the combined facilitation and depression variables $\langle u_j x_j \rangle$, under the assumption that the neuron j fires at a mean rate ν_j with Poisson statistics. This yields rate approximations of the steady-state values similar to those obtained in Eq. (4.21) and Eq. (4.22):

$$\langle u_j \rangle = \frac{U\left(\tau_u \nu_j + 1\right)}{U\tau_u \nu_j + 1} \tag{4.64}$$

$$\langle x_j \rangle = \frac{U\tau_u \nu_j + 1}{U\nu_j(\tau_u + \tau_x + \nu_j \tau_u \tau_x) + 1} \tag{4.65}$$

$$\langle u_j x_j \rangle = \langle u_j \rangle \langle x_j \rangle = \frac{U\left(\nu_j \tau_u + 1\right)}{U\nu_j(\tau_u + \tau_x + \nu_j \tau_u \tau_x) + 1}. \tag{4.66}$$

We now assume that the excitatory population of N_E neurons fires at the steady-state rates ϕ_i $(0 \leq i < N)$. To calculate the synaptic activation of excitatory-to-excitatory connections $\langle s_i^{\mathrm{EE}} \rangle$, we set Eq. (4.60) to zero, and average over Poisson statistics (again neglecting correlations), which yields

$$\langle s_j \rangle = \tau_E \langle u_j x_j \rangle \phi_j$$

and $\langle s_i^{\mathrm{EE}} \rangle = \sum_j w_{ij}^{\mathrm{EE}} \tau_E \langle u_j x_j \rangle \phi_j$. In Fig. 4.22 we evaluated the analytical approximation $\langle s_j \rangle$ against simulations, by comparing it to the temporal average $\langle s \rangle_t$ recorded from spiking simulations under homogeneous Poisson input with different firing rates.

We then define the normalized steady-state input J_i by:

$$J_i \equiv \frac{1}{N_{\mathrm{E}}} \langle s_i^{\mathrm{EE}} \rangle = \frac{1}{N_{\mathrm{E}}} \sum_j w_{ij}^{\mathrm{EE}} \langle s_j \rangle. \tag{4.67}$$

This form of the steady-state input Eq. (4.67) will be used to relate the theory of Section 4.6.1 to the spiking network. The additional factor $1/N_E$ is introduced to make the scaling of the excitatory-to-excitatory conductance with the size of the excitatory population N_{E} explicit, which will be used in Section 4.6.5.4. To see this, we assume that the excitatory conductance scale of excitatory neurons g_{EE} is scaled such that the total conductance is invariant under changes of N_{E} (Gerstner et al., 2014): $g_{\mathrm{EE}} = \tilde{g}_{\mathrm{EE}}/N_E$, for some fixed value $\tilde{g}_{\mathrm{EE}}$. This yields the total excitatory-to-excitatory conductance $g_{\mathrm{EE}} s_i^{\mathrm{EE}} = \tilde{g}_{\mathrm{EE}} J_i$ with J_i as introduced above, where the scaling with N_{E} is now shifted to the input variable J_i.

Finally, for the synaptic activation of excitatory to inhibitory connections, we get the mean activations:

$$\left\langle s^{\mathrm{IE}} \right\rangle = \tau_{\mathrm{E}} \sum_j \phi_j . \tag{4.68}$$

4.6.4.2 Firing rate prediction

In this section, we reiterate the approach of calculating the mean firing rates of neurons in dependence of the mean synaptic activations calculated in the last section. The resulting expressions are very similar to Section 3.6.3.3, and follow Brunel and Wang (2001b) to reduce the differential equations of Eq. (4.57) to a dimensionless form. Since the main difference consists in the absence of the voltage dependent NMDA conductance, we can set the two associated parameters $\beta \to 0, \gamma \to 0$ in Eq. (3.25), to arrive at:

$$\begin{aligned}
\tau_i \dot{V}_i &= -(V_i - V_L) + \mu_i + \sigma_i \sqrt{\tau_i}\eta_i(t) \\
S_i &= 1 + T_I \nu_I + T_{\mathrm{ext}} \nu_{\mathrm{ext}} + T_E J_i && (4.69) \\
\mu_i S_i &= (V_I - V_L)\, T_I \nu_I + (V_E - V_L)\, T_{\mathrm{ext}} \nu_{\mathrm{ext}} + (V_E - V_L)\, T_E J_i && (4.70) \\
\sigma_i &= \frac{g_{\mathrm{ext}}}{C_m} (\langle V \rangle - V_E)\, \tau_{\mathrm{ext}} \sqrt{\tau_i N_{\mathrm{ext}} \nu_{\mathrm{ext}}}. && (4.71) \\
\tau_i &= \frac{C_m}{g_L S_i} \\
\left\langle \eta_i(t) \right\rangle &= 0 \\
\left\langle \eta_i(t) \eta_i(t') \right\rangle &= \frac{1}{\tau_{\mathrm{ext}}} \exp\!\left(-\frac{\left|t - t'\right|}{\tau_{\mathrm{ext}}}\right) && (4.72)
\end{aligned}$$

where $T_{\mathrm{ext}} = N_{\mathrm{ext}} \tau_{\mathrm{ext}} \frac{g_{\mathrm{ext}}}{g_L}$, $T_I = N_I \tau_I \frac{g_I}{g_L}$ are effective timescales of external and inhibitory inputs, and $T_E = N_E \frac{g_E}{g_L}$ is a dimensionless scale for the excitatory conductance. As in Section 3.6.3.3, μ_i is the bias of the membrane potential due to synaptic inputs, and σ_i measures the scale of fluctuations in the membrane potential due to random spike arrival approximated by the Gaussian process η_i.

Again, the mean firing rates F and mean voltages $\langle V_i \rangle$ of populations of neurons governed by this type of differential equation can be approximated by (Brunel and Wang, 2001b) (we restate here only the central result, see Section 3.6.3.3 for the full expressions):

$$\begin{aligned}
F\left[\mu_i, \sigma_i, \tau_i\right] &= \left(\tau_{\mathrm{ref}} + \sqrt{\pi}\tau_i \int_{\beta(\mu_i,\sigma_i)}^{\alpha(\mu_i,\sigma_i)} du \exp(u^2)\, [1 + \mathrm{erf}(u)] \right)^{-1}, && (4.73) \\
\langle V_i \rangle &= \mu_i + V_L - (V_{\mathrm{thr}} - V_{\mathrm{reset}})\, \phi_i \tau_i . && (4.74)
\end{aligned}$$

4.6.4.3 Derivatives of the rate prediction

Here we calculate derivatives of the input-output relation (Eq. (4.73)) that will be used below in Section 4.6.5.

The expressions for drift and diffusion (see Section 4.4.3) contain the derivative $\phi_i' = \frac{dF}{dJ}\Big|_{J_i}$ of the input-output relation (Eq. (4.73)) with respect to the recurrent excitatory input J_i. Note, that F depends on J_i through all three arguments μ_i, σ_i and τ_i. First, we define $X(u) \equiv \exp(u^2)\,[1 + \mathrm{erf}(u)]$, and the shorthand $F_i = F\left[\mu_i, \sigma_i, \tau_i\right]$. The derivative can then be readily evaluated as (to shorten the notation in the following, we skip noting the evaluation points for derivatives in the following):

$$\begin{aligned}
\frac{dF}{dJ_i} &= -F_i \frac{T_E}{S_i}\left(F_i \tau_\mathrm{ref} - 1\right) + \sqrt{\pi} F_i^2 \tau_i \left[X(\beta)\frac{d\beta}{dJ_i} - X(\alpha)\frac{d\alpha}{dJ_i}\right], \\
\frac{d\alpha/\beta}{dJ_i} &= \left(\frac{\partial \alpha/\beta}{\partial \mu_i} + \frac{\partial \alpha/\beta}{\partial \sigma_i}\frac{\partial \sigma_i}{\partial \langle V_i\rangle}\right)\frac{T_E}{S_i}\left(-\mu_i + (V_E - V_L)\right) \\
&\quad - \left(\frac{\partial \alpha/\beta}{\partial \sigma_i}\left[\frac{\partial \sigma_i}{\partial \tau_i} - \frac{\partial \sigma_i}{\partial \langle V_i\rangle}\left(V_\mathrm{thr} - V_\mathrm{reset}\right)\phi_i\right] + \frac{\partial \alpha/\beta}{\partial \tau_i}\right)\frac{T_E}{S_i}\tau_i .
\end{aligned} \tag{4.75}$$

where α/β stands as a placeholder for either function, and the expressions for α and β are given in Eq. (3.21) and Eq. (3.22), respectively.

A second expression involving the derivative of Eq. (4.73) is $\frac{d\phi_{0,i}}{d\Delta_i^\mathrm{L}}$ which appears in the theory when estimating firing rate perturbations caused by frozen heterogeneities in the leak potentials of excitatory neurons (see Eq. (4.80)). The resulting derivatives are almost similar, which can be seen by the fact that replacing $V_\mathrm{L} \to V_\mathrm{L} + \Delta_i^\mathrm{L}$ in Eq. (4.57) only leads to an additional term Δ_i^L in Eq. (4.70). Thus, for neuron i the derivative can be evaluated to

$$\begin{aligned}
\frac{dF}{d\Delta_i^\mathrm{L}} &= \sqrt{\pi} F_i^2 \tau_i \left[X(\beta)\frac{\partial \beta}{\partial \Delta_i^\mathrm{L}} - X(\alpha)\frac{\partial \alpha}{\partial \Delta_i^\mathrm{L}}\right], \\
\frac{d\alpha/\beta}{d\Delta_i^\mathrm{L}} &= \left(\frac{\partial \alpha/\beta}{\partial \mu_i} + \frac{\partial \alpha/\beta}{\partial \sigma_i}\frac{\partial \sigma_i}{\partial \langle V_i\rangle}\right)\frac{1}{S_i}.
\end{aligned} \tag{4.76}$$

In practice, given a vector $\phi_{i,0}$ of firing rates in the attractor state, as well as the mean firing rate of inhibitory neurons ν_I, we evaluate the right hand side of Eq. (4.75) and Eq. (4.76) by replacing $F_i \to \phi_{i,0}$. This allows efficiently calculating the derivatives without having to perform any numerical integration. The two terms will be exactly equal if $\phi_{0,i}$ is a self-consistent solution of Eq. (4.73) for firing rates of the excitatory neurons across the network. For the theoretical estimate of $\phi_{0,i}$ (see Section 4.6.6), which is found by minimizing this self-consistency at a few points along the ring, this will be exactly satisfied at these few points only, and approximatively at all others. In the Results part of the study, we used numerical estimates of $\phi_{i,0}$ and ν_I that were directly measured from simulations. Since these were very close to the theoretical predictions for all networks we investigated, the replacement $F_i \to \phi_{i,0}$ above yields similar results.

4.6.5 Frozen noise and expected squared field magnitude

In this section we describe two sources of frozen noise that introduce drift into the system by Eq. (4.56): sparse connectivity and heterogeneous leak potentials. Additionally, we derive the expected squared magnitude for the drift fields for both of these sources. While the derivations are generally valid under the framework introduced in Section 4.6.1, we introduce frozen noise in the setting of the spiking network of Section 4.6.3, which contains excitatory and inhibitory neurons. We will also use the normalized input defined above (Eq. (4.67)), which ensures proper scaling of the system size in Section 4.6.5.4.

4.6.5.1 Sparse and heterogeneous connectivity

Introducing sparse connectivity, we replace the recurrent weights in Eq. (4.67) by:

$$w_{ij}^{\mathrm{EE}} \quad \rightarrow \quad \left[w_{ij}^{\mathrm{EE}} + \Delta_{ij}^{w}\right] \frac{p_{ij}}{p}. \tag{4.77}$$

Here, $p_{ij} \in \{0, 1\}$ are Bernoulli variables, with $P(p_{ij} = 1) = p$, where the connectivity parameter $p \in (0, 1]$ controls the overall sparsity of recurrent excitatory connections. For $p = 1$ the entire network is all-to-all connected. Additionally, we have studied noisy synaptic heterogeneities $\Delta_{ij}^{w} = \eta_{ij}\sigma_w$ added to the recurrent weights (as in Itskov et al. (2011b)), where $\left\{\eta_{ij} | 1 \leq i, j \leq N_E\right\}$ are independent, normally distributed random variables with zero mean and unit variance. We did not investigate this type of heterogeneity in the main text, since increasing σ_w lead to a loss of the attractor state before creating large enough directed drifts to be comparable to the other sources of frozen noise considered here – most of the small effects were "hidden" behind diffusive displacement (Laing and Longtin, 2001). Nevertheless, we included this case in the analysis here for completeness.

Let the center position of the bump be $\varphi_k = k\frac{2\pi}{N} - \pi$ (for $0 \leq k < N$). Subject to the perturbed weights, the recurrent steady-state excitatory input $J_i(\varphi_k)$ Eq. (4.67) to any excitatory neuron can be written as the unperturbed input $J_{0,i}(\varphi_k)$ plus an additional input $J_i^{\mathrm{struct}}(\varphi_k)$ arising from the perturbed connectivity. Note that the synaptic steady-state activations $s_{0,j}(\varphi_k)$

change with varying bump centers – in the following, we denote $s_{0,j}^{k} \equiv s_{0,j}(\varphi_k)$:

$$\begin{aligned}
J_i(\varphi_k) &= \frac{1}{N_E}\sum_j \left[w_{ij}^{\mathrm{EE}} + \Delta_{ij}^{w}\right]\frac{p_{ij}}{p} s_{0,j}^{k} \\
&= \frac{1}{N_E}\frac{1}{p}\left[\sum_j \left[w_{ij}^{\mathrm{EE}} + \Delta_{ij}^{w}\right] s_{0,j}^{k} - \sum_j \left[w_{ij}^{\mathrm{EE}} + \Delta_{ij}^{w}\right](1-p_{ij})\, s_{0,j}^{k}\right] \\
&= \frac{1}{N_E}\sum_j w_{ij}^{\mathrm{EE}} s_{0,j}^{k} + \frac{1}{N_E}\frac{1}{p}\left[(1-p)\sum_j w_{ij}^{\mathrm{EE}} s_{0,j}^{k} + \sum_j \Delta_{ij}^{w} s_{0,j}^{k} - \sum_j \left[w_{ij}^{\mathrm{EE}} + \Delta_{ij}^{w}\right](1-p_{ij})\, s_{0,j}^{k}\right] \\
&= \underbrace{\frac{1}{N_E}\sum_j w_{ij}^{\mathrm{EE}} s_{0,j}^{k}}_{J_{0,i}(\varphi_k)} + \underbrace{\frac{1}{N_E}\frac{1}{p}\left[\sum_j \left[w_{ij}^{\mathrm{EE}} + \Delta_{ij}^{w}\right] p_{ij} s_{0,j}^{k} - p\sum_j w_{ij}^{\mathrm{EE}} s_{0,j}^{k}\right]}_{J_i^{\mathrm{struct}}(\varphi_k)}.
\end{aligned}$$

Note that $J_{0,i}(\varphi_k)$ is an index-shifted version of the steady-state input: $J_{0,i}(\varphi_k) = J_{0,i-k}$. However, such a relation does not hold for $J_i^{\mathrm{struct}}(\varphi_k)$, since the random numbers p_{ij} will change the resulting value for varying center positions.

We calculate the firing rate perturbations $\delta\phi_i(\varphi_k)$ resulting from the additional input by a linear expansion around the steady-state firing rates $\phi_{0,i}(\varphi_k) \to \phi_{0,i}(\varphi_k) + \delta\phi_i(\varphi_k)$. These evaluate to:

$$\begin{aligned}
\delta\phi_i(\varphi_k) &= \left.\frac{dF}{dJ}\right|_{J_{0,i}}(\varphi_k)\cdot J_i^{\mathrm{struct}}(\varphi_k) \\
&= \phi_{i,0}'(\varphi_k)\cdot J_i^{\mathrm{struct}}(\varphi_k). \qquad (4.78)
\end{aligned}$$

See Section 4.6.4.3 for the derivation of the function $\frac{dF}{dJ}(J_{0,i})$ for the spiking network used in the main text.

In the sum of Eq. (4.56), we used the firing rate profile $\vec{\phi}_0$ centered at φ_0 to calculate the drift for varying center positions. To accommodate the shifted indices resulting from moving center positions, we re-index the summands to yields the perturbations $\phi_{0,i} \to \phi_{0,i} + \Delta\phi_i(\varphi_k)$ used there:

$$\Delta\phi_i(\varphi_k) = \phi_{i,0}' \cdot J_{i+k}^{\mathrm{struct}}(\varphi_k).$$

4.6.5.2 Heterogeneous leak reversal potentials

We further investigated random distributions of the leak reversal potential V_{L}. These are implemented by the substitution

$$V_{\mathrm{L}} \quad \to \quad V_{\mathrm{L}} + \Delta_i^{\mathrm{L}}, \qquad (4.79)$$

where the Δ_i^{L} are independent normally distributed variables with zero mean, i.e. $\left\langle \Delta_i^{\mathrm{L}} \right\rangle = 0mV, \left\langle \Delta_i^{\mathrm{L}} \Delta_j^{\mathrm{L}} \right\rangle = \sigma_{\mathrm{L}}^2 \delta_{ij}$. The parameter σ_{L} controls the standard deviation of these random variables, and thus the noise level of the leak heterogeneities.

Let $\varphi_k = k\frac{2\pi}{N} - \pi$ for $0 \leq k < N$ be the center position of the bump. First, note that the heterogeneities Δ_i^L do not depend on the center position φ_k, since they are single neuron properties. As in the last section, we calculate the firing rate perturbations $\delta\phi_i(\varphi_k)$ resulting from the additional input by a linear expansion around the steady-state firing rates $\phi_{0,i}(\varphi_k) \to \phi_{0,i}(\varphi_k) + \delta\phi_i(\varphi_k)$:

$$\begin{aligned}\delta\phi_i(\varphi_k) &= \frac{dF_i}{d\Delta_i^{\mathrm{L}}}(J_i(\varphi_k)) \cdot \Delta_i^{\mathrm{L}} \\ &\equiv \frac{d\phi_{0,i}}{d\Delta_i^{\mathrm{L}}}(\varphi_k) \cdot \Delta_i^{\mathrm{L}}.\end{aligned} \tag{4.80}$$

Here, $\frac{dF_i}{d\Delta_i^{\mathrm{L}}}(J_i(\varphi_k))$ is the derivative of the input-output relation of neuron i in a bump centered at φ_k, with respect to the leak perturbation. We introduced $\frac{d\phi_{0,i}}{d\Delta^L}(\varphi_k)$ as a shorthand notation for this derivative, since it is evaluated at the steady-state input $J_{i,0}(\varphi_k)$. For the spiking network of the main text, this is derived in Section 4.6.4.3.

In the sum of Eq. (4.56), we keep the firing rate profile $\vec{\phi}_0$ centered at φ_0 while calculating the drift for varying center positions. As in the last section, we re-index the sum to yield the perturbations $\phi_{0,i} \to \phi_{0,i} + \Delta\phi_i(\varphi_k)$ used there:

$$\Delta\phi_i(\varphi_k) = \frac{d\phi_{0,i}}{d\Delta_i^{\mathrm{L}}} \cdot \Delta_{i+k}^{\mathrm{L}}.$$

4.6.5.3 Squared field magnitude

Using the equation of the drift field in Eq. (4.56), and the firing rate perturbations above (Eq. (4.78) and Eq. (4.80)) it is easy to see that for any center position φ the expected drift field averaged over the noise parameters is 0, since all single firing rate perturbations vanish in expectation. In the following we calculate the variance of the drift field averaged over noise realizations, which turns out to be additive with respect to the two noise sources.

We begin by calculating the correlations between frozen noises caused by sparse connectivity and leak heterogeneities. For the Bernoulli distributed variables p_{ij} it holds that $\langle p_{ij} \rangle = p, \langle p_{ij} p_{lk} \rangle = \delta_{il}\delta_{jk} p + (1 - \delta_{il}\delta_{jk}) p^2$. For the other independent random variables it holds that $\langle \Delta_i^L \rangle = 0mV, \left\langle \left(\Delta_i^L\right)^2 \right\rangle = \sigma_L^2, \left\langle \Delta_{ij}^w \right\rangle = 0, \left\langle \left(\Delta_{ij}^w\right)^2 \right\rangle = \sigma_w^2$. Again, the weight heterogeneities Δ_{ij}^w are only included for completeness – all analyses of the main text assume that $\sigma_w = 0$.

For the correlations between the perturbations we then know that (for brevity, we omit the dependence on the center position φ):

$$\begin{aligned}
\langle J_i^{\text{struct}} \Delta_i^L \rangle &= 0 \\
\langle J_i^{\text{struct}} J_l^{\text{struct}} \rangle &= \frac{1}{N_E^2} \left\langle \left(\sum_k s_{0,k} w_{lk}^{\text{EE}} \left(\frac{p_{lk}}{p} - 1 \right) + \frac{1}{p} \sum_k s_{0,k} \Delta_{lk}^w p_{lk} \right) \right. \\
&\qquad \left. \left(\sum_k s_{0,k} w_{lk}^{\text{EE}} \left(\frac{p_{lk}}{p} - 1 \right) + \frac{1}{p} \sum_k s_{0,k} \Delta_{lk}^w p_{lk} \right) \right\rangle \\
&= \frac{1}{N_E^2} \left(\sum_{j,k} s_{0,j} w_{ij}^{\text{EE}} s_{0,k} w_{lk}^{\text{EE}} \left\langle \left(\frac{p_{ij}}{p} - 1 \right) \left(\frac{p_{lk}}{p} - 1 \right) \right\rangle + \frac{1}{p^2} \sum_{j,k} s_{0,j} s_{0,k} \left\langle \Delta_{ij}^w \Delta_{lk}^w \right\rangle \langle p_{ij} p_{lk} \rangle \right) \\
&= \frac{1}{N_E^2} \left(\sum_{j,k} s_{0,j} w_{ij}^{\text{EE}} s_{0,k} w_{lk}^{\text{EE}} \left(\frac{1}{p^2} \langle p_{ij} p_{lk} \rangle - 1 \right) + \frac{1}{p} \sum_j s_{0,j}^2 \sigma_w^2 \delta_{il} \right) \\
&= \frac{1}{N_E^2} \left(\sum_j s_{0,j}^2 \left(w_{ij}^{\text{EE}} \right)^2 \left(\frac{1}{p} - 1 \right) + \frac{1}{p} \sum_j s_{0,j}^2 \sigma_w^2 \right) \delta_{il}.
\end{aligned}$$

Starting from Eq. (4.56), we now use as a firing rate perturbation the sum of firing rate perturbations from both Eq. (4.78) and Eq. (4.80). We further denote the pre-factor $C_i = \frac{dJ_{0,i}}{d\varphi} \frac{1+\tau_u \phi_{0,i}(U\tau_u \phi_{0,i}+2)}{(U\phi_{0,i}(\tau_u \tau_x \phi_{0,i}+\tau_u+\tau_x)+1)^2}$, and explicitly state the dependence on the bump center position φ. The expected squared field averaged over ensemble of frozen noises is then:

$$\begin{aligned}
\left\langle A(\varphi)^2 \right\rangle_{\text{frozen}} &= \frac{1}{S^2} \sum_{i,j} C_i(\varphi) C_j(\varphi) \cdot \\
&\left\langle \left(\phi'_{0,i}(\varphi) J_i^{\text{struct}}(\varphi) + \frac{d\phi_{0,i}}{d\Delta_L}(\varphi) \Delta_i^L \right) \left(\phi'_{0,j}(\varphi) J_j^{\text{struct}}(\varphi) + \frac{d\phi_j}{d\Delta_L}(\varphi) \Delta_j^L \right) \right\rangle_{\text{frozen}} \\
&= \frac{1}{S^2} \sum_{i,j} C_i(\varphi) C_j(\varphi) \cdot \\
&\left(\left\langle \phi'_{0,i}(\varphi) J_i^{\text{struct}}(\varphi) \phi'_{0,j}(\varphi) J_j^{\text{struct}}(\varphi) \right\rangle_{\text{frozen}} + \left\langle \frac{d\phi_{0,i}}{d\Delta_L}(\varphi) \Delta_i^L \frac{d\phi_{0,j}}{d\Delta_L}(\varphi) \Delta_j^L \right\rangle_{\text{frozen}} \right) \\
&= \frac{1}{S^2} \sum_i C_i^2(\varphi) \cdot \\
&\left(\frac{\left(\phi'_{0,i}(\varphi) \right)^2}{N_E^2} \left[\left(\frac{1}{p} - 1 \right) \sum_j \left(s_{0,j}(\varphi) \right)^2 \left(w_{ij}^{\text{EE}} \right)^2 + \frac{1}{p} \sum_j s_{0,j}^2(\varphi) \sigma_w^2 \right] + \left(\frac{d\phi_{0,i}}{d\Delta_L}(\varphi) \right)^2 \sigma_L^2 \right).
\end{aligned} \tag{4.81}$$

One can see directly that the two last terms are invariant under shifts of the bump center φ, since these introduce symmetric shifts of the indexes i. Similarly, it is easy to see that the first term is also invariant. Let φ' be shifted to the right by one index from φ . It then holds that:

$$\begin{aligned}
&\sum_i C_i^2(\varphi')\left(\frac{\left(\phi'_{0,i}(\varphi')\right)^2}{N_E^2}\left(\frac{1}{p}-1\right)\sum_j \left(s_{0,j}(\varphi')\right)^2\left(w_{ij}^{\mathrm{EE}}\right)^2\right)\\
=&\sum_i C_{i-1}^2\left(\frac{\left(\phi'_{0,i-1}(\varphi)\right)^2}{N_E^2}\left(\frac{1}{p}-1\right)\sum_j \left(s_{0,j-1}(\varphi)\right)^2\left(w_{ij}^{\mathrm{EE}}\right)^2\right)\\
=&\sum_i C_{i-1}^2\left(\frac{\left(\phi'_{0,i-1}(\varphi)\right)^2}{N_E^2}\left(\frac{1}{p}-1\right)\sum_j \left(s_{0,j}(\varphi)\right)^2\left(w_{i-1,j}^{\mathrm{EE}}\right)^2\right).
\end{aligned}$$

The final equation holds since w_{ij}^{EE} consists of index-shifted rows of the same vector (see generally Section 4.4.1 and Section 4.6.3.3 for the spiking network weights).

In summary, $\langle A(\varphi)^2\rangle_{\text{frozen}}$ will evaluate to the same quantity $\langle A^2\rangle_{\text{frozen}}$ for all center positions φ. In the main text, we use this fact to estimate $\langle A^2\rangle_{\text{frozen}}$ from simulations, by additionally averaging over the all center positions and interchanging the ensemble and positional averages:

$$\langle A^2\rangle_{\text{frozen}} = \frac{1}{N_E}\sum_k \langle A(\varphi_k)^2\rangle_{\text{frozen}} = \left\langle \frac{1}{N_E}\sum_k A(\varphi_k)^2 \right\rangle_{\text{frozen}} .$$

Thus, we can compare the value of $\langle A^2\rangle_{\text{frozen}}$ to the mean squared drift field over all center positions, averaged over instantiations of noises.

4.6.5.4 System size scaling

To make explicit the scaling with system size, we argue as follows. Generally, sums over the discretized intervals $[-\pi,\pi)$ as they appear in Eqs. (4.52) and (4.56) will scale with the number N chosen for the discretization of the positions on the continuous ring $\varphi(i) = \frac{i}{N}2\pi - \pi$. Consider two discretizations of the ring, partitioned into N_1 and N_2 uniformly spaced bins of width $\frac{2\pi}{N_1}$ and $\frac{2\pi}{N_2}$. We can then approximate integrals over any continuous (Riemann integrable) function f on the ring by the two Riemann sums:

$$\frac{2\pi}{N_1}\sum_{i=0}^{N_1-1} f(\varphi_{1,i}) \approx \int_{-\pi}^{\pi} f(\varphi)d\varphi \approx \frac{2\pi}{N_2}\sum_{i=0}^{N_2-1} f(\varphi_{2,i}), \tag{4.82}$$

where, $i\frac{2\pi}{N_1} \leq \varphi_{1,i} < (i+1)\frac{2\pi}{N_1}$ (for N_2 and $\varphi_{2,i}$ analogously) are points in the bins (Forster, 2016).

Numerical quantities for the results of the main text have been calculated for $N_E = 800$. In the following we denote all of these quantities with an asterisk $(*)$. To generalize these results to arbitrary system size N, we replace sums over N bins in Eqs. (4.52) and (4.56) by scaled sums over N_E bins using the relation Eq. (4.82):

$$\sum_{i=0}^{N-1} \to \frac{N}{N_E} \sum_{i=0}^{N_E-1}.$$

First, we find that the normalization constant scales as $S = \frac{N}{N_E} S^*$, and thus (dots indicate the summands, which are omitted for clarity) for the diffusion strength B (cf. Eq. (4.52)):

$$\begin{aligned} B &= \frac{1}{S^2} \sum_{i=0}^{N-1} \cdots \\ &= \frac{N}{N_E} \frac{1}{S^2} \sum_{i=0}^{N_E-1} \cdots \\ &= \frac{N_E}{N} \frac{1}{(S^*)^2} \sum_{i=0}^{N_E-1} \cdots \\ &= \frac{N_E}{N} B^*. \end{aligned} \tag{4.83}$$

For the drift magnitude we turn to the expected squared drift magnitude calculated earlier (cf. Eq. (4.81)), for which we find that

$$\begin{aligned} \langle A^2 \rangle_{\text{frozen}} &= \left(\frac{N_E}{N}\right)^2 \frac{1}{(S^*)^2} \frac{N}{N_E} \sum_{i=0}^{N_E-1} C_i^2 \left(\frac{\left(\phi_i'\right)^2}{N^2} \left(\frac{1}{p} - 1\right) \frac{N}{N_E} \sum_{j=0}^{N_E-1} s_j^2 w_{ij}^2 + \left(\frac{d\phi_i}{dE_L}\right)^2 \sigma_L^2 \right) \\ &= \frac{1}{(S^*)^2} \sum_{i=0}^{N_E-1} C_i^2 \left(\frac{1}{N^2} \left(\phi_i'\right)^2 \left(\frac{1}{p} - 1\right) \sum_{j=0}^{N_E-1} s_j^2 w_{ij}^2 + \frac{N_E}{N} \left(\frac{d\phi_i}{dE_L}\right)^2 \sigma_L^2 \right). \end{aligned} \tag{4.84}$$

Note, that we could not resolve this scaling in dependence of $\langle A^2 \rangle^*_{\text{frozen}}$, since the two sources of frozen noise (connectivity and leak heterogeneity) show different scaling with N.

4.6.6 Steady-state approximation and parameter optimization

In this section we sketch the self-consistent prediction of steady-state firing rate profiles (presented in the previous chapter, Chapter 3) to the networks with facilitation and depression investigated here. The main theoretical handiwork has already been performed in Section 4.6.4.1 and Section 4.6.4.2. Here we follow the approach of Section 3.4.3, to constrain Eq. (4.73) to a low-dimensional set of self-consistent equations. This will also be used to optimize network parameters below. Following Section 3.6.3.3, we replace the network activity ϕ_j in the total input Eq. (4.67) by a parametrization

$$g(\theta_j) = g_0 + g_1 \exp\left(- \left[\frac{|\theta_j|}{g_\sigma} \right]^{g_r} \right). \tag{4.85}$$

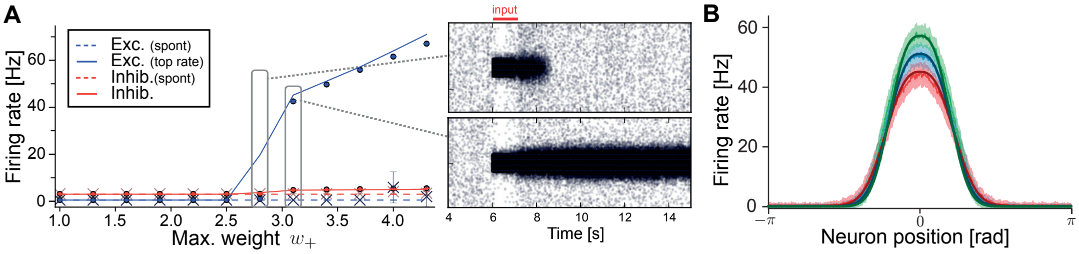


Figure 4.23 – **Prediction of bump shapes for spiking networks with short-term plasticity. A** Theoretical prediction (lines) of maximal firing rate of firing rate profile (blue, $g_0 + g_1$) and inhibitory firing rate (red) for varying parameter w_+. Dots show averaged (10 repetitions, error bars show 95% confidence of the mean) maximal mean firing rates estimated on $7s$ of delay activity, $1s$ after offset of the initial cue. Crosses show the maximal firing rates if no initial cue was given (mean of maximal rate of fit to g over 10 repetitions, error bars show 95% confidence of the mean) estimated on $4.5s$ of simulation time, with theoretical predictions (dashed lines, by setting $w_+ = 1$ in the theory). **A, right:** Raster plots of excitatory population for two example simulations at values of w_+ indicated by lines. **B** Shape of the optimized firing rates profiles (thick colored lines) compared to mean firing rate of spiking simulations (thin colored lines) (firing rates recorded from $1s$ of delay activity and 10 repetitions, we show mean $\pm 0.5 STD$). Short-term plasticity parameters were $U = 0.05, \tau_u = 650ms, \tau_x = 150ms$. Other network parameters were $g_{\mathrm{EE}} = 0.0322nS$, $g_{\mathrm{EI}} = 2.062nS$, $g_{\mathrm{IE}} = 2.567pS$, $g_{\mathrm{II}} = 1.5911nS$, and $w_\sigma = 0.45$.

Approximating sums $\frac{1}{N_E}\sum_{j=0}^{N_E-1}$ with integrals $\frac{1}{2\pi}\int_{-\pi}^{\pi} d\varphi$ we arrive at:

$$J_i \approx \frac{1}{2\pi}\int_{-\pi}^{\pi} d\varphi\, w^{\mathrm{EE}}\left(\theta_i - \varphi\right)\langle s_j\rangle(g\left(\varphi\right)) = \frac{1}{2\pi}\mathrm{input}_{\theta_i}\left[g\right].$$

We then substitute this relation in Eq. (4.73) and Eq. (4.74) to arrive at:

$$\begin{aligned} \mu_i &= \mu_i\left(\mathrm{input}_{\theta_i}\left[g\right], \nu_I, \nu_{\mathrm{ext}}\right), \\ \tau_i &= \tau_i\left(\mathrm{input}_{\theta_i}\left[g\right], \nu_I, \nu_{\mathrm{ext}}\right), \\ F(\mathrm{input}_{\theta_i}\left[g\right], \nu_I, \langle V_i\rangle) &\equiv \phi\left[\mu_i, \sigma_i, \tau_i\right]. \end{aligned} \tag{4.86}$$

Comparing Eq. (4.86) to the earlier form Eq. (3.24), we see that the equations for the voltage bias term μ_i, the effective membrane time constants τ_i, as well as the total conductance S_i (Eq. (4.69)) are all independent of the mean voltage, since here we have no voltage dependence of the NMDA conductance. This makes the equation for the mean voltage Eq. (4.74) an explicit equation, which can be used in the fluctuation scale Eq. (4.71), leaving us with only a single self-consistent equation Eq. (4.86).

4.6.6.1 Optimization for steady-state firing rates

To approximate the steady-state firing rates, we again seek to contain the parameters of the parametrization Eq. (4.85). For each point θ_i that we choose to sample from the excitatory population, the theory now yields one constraining equation Eq. (4.73). The inhibitory population, again also yields a single equation, for the free variable ν_I . This again allows us to choose 4 evaluation points to constrain the 4 free parameters of the parametrization. These are chosen as described in Section 3.6.3.1. Errors were optimized using the *optimize.root* solver described in Section 3.6.3.5.

An example for optimization results is shown in Fig. 4.23, where we predicted the bump shape under the changing parameter w_+ (compare to Section 3.4.3.2 and Fig. 3.4 in the last chapter).

4.6.6.2 Optimization for network parameters

Here, we describe the optimization procedure performed to produce approximately similar bump shapes as the parameters of short-term plasticity are varied. It should be noted that the parameters presented here were optimized at an earlier stage of this work than the results presented in Section 3.4.3.3. In particular, the optimization targets should have contained the parameter g_σ, and the state g_r should have been optimized for. The resulting procedure unfortunately required some hand-tuning (see below).

In hindsight, these oversights could have been improved. However, we stress that we have shown (see Fig. 4.4) that the parameters achieved in this way yield a set of spiking attractor networks with very similar firing rates. Further, we have validated our results in the main text by extrapolating theoretical predictions for drift and diffusion from a single bump attractor shape ($U = 1, \tau_x = 150 ms$) by changing the short-term plasticity parameters. This demonstrated that the effects of short-term plasticity on drift and diffusion were due to changes in the plasticity parameters, and not due to possible differences in the shapes of the attractor states of the networks that were simulated for each of these parameters.

We optimize networks by imposing the following targets for parameters:
$g_0 = 0.1$Hz, $g_1 = 40$Hz, $\nu_{E,\text{basal}} = 0.5$Hz, $\nu_{I,\text{basal}} = 3$Hz. For all networks we chose $w_+ = 4.0$, $g_r = 2.5$. The following parameters were then optimized: ν_I, g_σ, g_{EE} (excitatory conductance g_E on excitatory neurons); g_{IE} (excitatory conductance g_E on inhibitory neurons); g_{EI} (inhibitory conductance g_I on excitatory neurons); g_{II} (inhibitory conductance g_I on inhibitory neurons). The basal firing rates (firing rates in the uniform state of the network, prior to being cued) yielded two equations from Eq. (4.86) by setting $w_+ = 1$. This left 4 free parameters, again requiring evaluating Eq. (4.86) at 4 points as described in Section 3.6.3.1. Errors were optimized using the *optimize.root* solver described in Section 3.6.3.5. The basal firing rates were chosen to be fairly low to make the uniform state more stable (as in Barbieri and Brunel (2008)).

This procedure does not yield a fixed value for g_σ, since g_σ is optimized for and is not set as a target value. We thus iterated the following until a solution was found with $g_\sigma \approx 0.5$: a) change the width of the recurrent weights w_σ; b) optimize network parameters as described here; c) optimize the expected bump shape for the new network parameters to predict g_σ. The resulting parameter values are given in Table D.2.

4.6.7 Numerical methods

4.6.7.1 Spiking simulations

All network simulations and models were implemented in the NEST simulator (Bos et al., 2015). Neuronal dynamics are integrated by fourth-order Runge-Kutta as implemented in the GSL package (Galassi et al., 2009). The short-term plasticity model is integrated exactly, based on inter-spike intervals.

Simulation protocol In all experiments (except those involving bi-stability, see below) spiking networks were simulated for a transient initial period of $t_{\text{initial}} = 500ms$. To center the network in an attractor state at a given angle $-\pi \leq \varphi < \pi$, we gave an initial cue signal by stimulating $0.1 \cdot N_E$ neurons centered at φ by strong excitatory input mediated by additional Poisson firing onto AMPA receptors ($1s$, 2kHz) with connections scaled down by a factor of $g_{\text{signal}} = 0.5$. The external input ceased at $t = t_{\text{off}} = 1.5s$. For simulations to estimate the diffusion we simulated until $t_{\text{max}} = 15s$, yielding $13.5s$ of delay activity after the cue offset. For simulations to estimate drift we set $t_{\text{max}} = 8s$, yielding $6.5s$ of delay activity after the cue offset.

For simulations exploring the bi-stability between the uniform state and a bump state (Fig. 4.2B and Fig. 4.23), we added an additionaly input prior to the spontaneous state. We stimulated simultaneously 20 excitatory neurons around 4 equally spaced cue points each (80 neurons in total, 500ms, 1.5kHz, AMPA connections scaled by a factor $g_{\text{signal}} = 2$). This was applied to settle networks into the uniform state more stably – without this perturbation, networks sometimes approached the bump state after being uniformly initialized. In both figures, we show population activity only after this initial stimulus was applied. For Fig. 4.23, firing rates of the uniform state were measured after offset of this stimulus.

Estimation of centers and mean bump shapes To estimate centers of bump states, simulations were run until $t = t_{\text{max}}$ and spikes were recorded from the excitatory population and converted to firing rates by convolving them with an exponential kernel ($\tau = 100ms$) (Nawrot et al., 1999) and then sampled at resolution 1ms. This results in vectors of firing rates $\nu_j(t), 0 \leq j \leq N_E - 1$ for every time t. We calculated the population center $\varphi(t)$ for time t by measuring the phase of the first spatial Fourier coefficient of the firing rates. This is given by $\varphi(t) = \arg\left(\sum_j \exp(i\frac{2\pi}{N_E}j)\nu_j(t)\right) - \pi$. For all analyses below, we identify $t = 0$ to be the time $t = t_{\text{off}}$ of the initial cue.

To measure the mean bump shapes, we first rectified the vectors $\nu_j(t)$ for every t by rotating the vector until $\varphi(t) = 0$. We then sampled the rectified firing rates starting from $1s$ after cue offset at intervals of 20ms, which were used to calculate the mean firing rates. In Fig. 4.23 we display the mean of sampled firing rates from all 10 repetitions. Fig. 4.4 shows mean rates for each simulation averaged over the ~ 1000 repetitions performed in the diffusion estimation (below).

Exclusion of bump trajectories Sometimes bump trajectories would leave the attractor state and return to the uniform state. We identified these trajectories in all experiments by identifying maximal firing rates across the population that dropped below 10Hz during the delay period. The such identified repetitions were excluded from the analyses, which occurred mostly in networks with no facilitation for $\tau_x = 150ms, \tau_u = 650ms$: at $U = 1$, we excluded $222/1000$ repetitions from the diffusion estimation, while for all other $U \leq 0.8$ at most $15/1000$ were excluded. Increasing the depression time constant also lead to less stable attractor states: for $\tau_x = 200ms, \tau_u = 650ms$ and $U = 0.8$, we had to exclude $250/1000$ repetitions. During the simulations for drift estimation, we observed that frozen noise also leads to less stable bumps under weak facilitation for sparse connectivity ($p \ll 1$) and high leak variability ($\sigma_L \gg 0$).

Diffusion estimation Diffusion was estimated for each combination of network parameters by simulating 1000 repetitions (10 initial cue positions, 100 repetitions each) of $13.5s$ of delay activity. Center positions $\varphi_k(t)$ were estimated for each repetition k as described above. We then calculated for each repetition the offset relative to the position at $500ms$ by $\Delta\varphi_k(t) = \varphi_k(t - 500ms) - \varphi_k(500ms)$, effectively discarding the first $500ms$ after cue-offset. The time-dependent variance of K repetitions (excluding those repetitions in which the bump state was lost, see above) was then calculated as $V(t) = \frac{1}{K}\sum_k \Delta\varphi_k^2(t)$. The diffusion strnegth can then be estimated from the slope of a linear least-squares regression (using the Scipy method *scipy.stats.linregress* (Jones et al., 2017)) to the variance as a function of time: $V(t) \approx D_0 + D \cdot t$, where the intercept D_0 is included to account for initial transients. We estimated confidence intervals by bootstrapping (Efron and Tibshirani, 1994): sampling K elements out of the K repetitions with replacement (5000 samples) and estimating the confidence level of 0.95 by the bias corrected and accelerated bootstrap implemented in *scikits-bootstrap* (Evans, 2017). As a control, we calculated confidence intervals for D additionally by Jackknifing: after building a distribution of estimates of D on K one-left-out samples of all repetitions, the standard error of the mean can be calculated and is multiplied by 1.96 to obtain the 95% confidence interval (Abdi and Williams, 2010) – confidence intervals obtained by this method were almost indistinguishable from confidence intervals obtained by bootstrapping.

Drift estimation Drift was estimated numerically for each combination of network and frozen noise parameters by simulating 400 repetitions (20 initial cue positions, 20 repetitions each) of $6.5s$ of delay activity. Centers positions $\varphi_k(t)$ were estimated for all K repetitions (excluding those repetitions in which the bump state was lost, see above) as explained above. We then computed displacements in time by computing a set of discrete differences

$$\Delta\varphi_k = \left\{ \left(\varphi_k \left[t_0 + (j+1)dt \right] - \varphi \left[t_0 + j \cdot dt \right] \right) / dt \,\middle|\, \forall j \in \mathbb{N}_0 : t_0 + (j+1)dt \leq t_{\max} \right\},$$

where we chose $dt = 1.5s$ and $t_0 \in \{500ms, 700ms, 900ms, \ldots, 1900ms\}$. All differences are calculated with periodic boundary conditions on the circle $[-\pi, \pi)$, i.e. the maximal difference was π/dt. We then calculated a binned mean (100 bins on the ring, unless mentioned otherwise) of differences calculated for all K trajectories, to approximate the drift-fields as a function of positions on the ring.

4.6.7.2 Mutual information measure

We are estimating the mutual information between a set of initial positions $x \in [0, 2\pi)$ and associated final positions $y(x) \in [0, 2\pi)$ of the trajectories of a continuous attractor network over a fixed delay period of T. For our results, we take $T = 6.5s$. We constructed binned and normalized histograms (with bin size $n = 100$, but see below) as approximate probability distributions of initial positions $p_i = p([i-1]\frac{2\pi}{n} \leq x < i\frac{2\pi}{n})$ and all final positions $q_i = p([i-1]\frac{2\pi}{n} \leq y < i\frac{2\pi}{n})$ (with bins indexed by $1 \leq i \leq n$), as well as the bivariate probability distribution $r_{ij} = p([i-1]\frac{2\pi}{n} \leq x < i\frac{2\pi}{n}, [j-1]\frac{2\pi}{n} \leq y(x) < j\frac{2\pi}{n})$.

Using these, we can calculate the mutual information as (Latham and Roudi, 2009; Cover and Thomas, 2012) $MI = \sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij} \log_2\left(\frac{r_{ij}}{p_i q_j}\right)$. Note, that the sum effectively counts only nonzero entries of r_{ij} (trajectories that started in bin i and ended in bin j): these imply that $p_i \neq 0$ (a trajectory started in bin i) and $q_j \neq 0$ (a trajectory ended in bin j), which makes the sum well defined. Although the value of MI depends on the number of bins n, in Fig. 4.14 we always normalize MI to that of the control network ($U = 1$, no frozen noise, see Section 4.4.6.1), which leaves the resulting plot nearly invariant under a change of bin numbers.

4.6.7.3 Numerical integration of Langevin equations

Numerically integration of the homogeneous Langevin equations (Eq. (4.3)) describing drift and diffusion of bump positions $\varphi \in [-\pi, \pi)$ (with circular boundary conditions) has been implemented as a C extension in Cython (Behnel et al., 2011) to the Python language (Oliphant, 2007). Since the drift fields $A(\varphi)$ are estimated on a discretization of the interval $[-\pi, \pi)$ into N bins, we first interpolate drift fields A given as N discretized values to obtain continuous fields – interpolations are obtained using cubic splines on periodic boundary conditions using the class *gsl_interp_cspline_periodic* of the Gnu Scientific Library (Galassi et al., 2009).

For forward integration of the Langevin equation Eq. (4.3) from time $t = 0$, we start from an initial position $\varphi_0 = \varphi(t = 0)$. Given a time resolution dt (unless otherwise stated we use

$dt = 0.1s$) and a maximal time $t_{\max}$ we repeat the following operations until we reach $t = t_{\max}$:

$$\begin{aligned} t &\to t + dt, \\ \varphi &\to \varphi + dt \cdot A(\varphi) + \sqrt{dtB} \cdot r, \\ \varphi &\to ((\varphi + \pi) \mod 2\pi) - \pi. \end{aligned}$$

Here, for each iteration r is a random number drawn from a normal distribution with zero mean and unit variance ($\langle r \rangle = 0$ and $\langle r^2 \rangle = 1$). The last step is performed to implement the circular boundary conditions on $[-\pi, \pi)$.

A Secondary publications

This chapter contains additional publications that were completed during the time of my doctoral thesis, to which I did not contribute in the role of a first author. For each publication, a short summary is given, and the author contributions are provided in detail.

A.1 Synaptic patterning and the timescales of cortical dynamics

Renato Duarte, **Alexander Seeholzer**, Karl Zilles and Abigail Morrison

Published in: Current Opinion in Neurobiology, 43:156-165 (DOI: 10.1016/j.conb.2017.02.007; Duarte et al. (2017))

Summary

Neocortical circuits, as large heterogeneous recurrent networks, can potentially operate and process signals at multiple time scales, but appear to be differentially tuned to operate within certain temporal receptive windows. The modular and hierarchical organization of this selectivity mirrors anatomical and physiological relations throughout the cortex and is likely determined by the regional electrochemical composition. Being consistently patterned and actively regulated, the expression of molecules involved in synaptic transmission constitutes the most significant source of laminar and regional variability. Due to their complex kinetics and adaptability, synapses form a natural primary candidate underlying this regional temporal selectivity. The ability of cortical networks to reflect the temporal structure of the sensory environment can thus be regulated by evolutionary and experience-dependent processes.

In this review, we compile evidence suggesting that the most significant source of both laminar and regional variability in the adult cortex is the patterning of the synaptic machinery. This patterning is highly conserved across individuals, and reflects the prevalence of innate constraints. We then examine how the properties of synaptic composition and local connectivity are reflected in the emergent dynamics of large recurrent networks, their characteristic time scales, and their complex spatiotemporal activity patterns. Finally, we discuss how activity-dependent modifications, acting across multiple time scales, have the potential to tune temporal receptive fields by locally stabilizing the circuit's state space.

Author contributions

R.D., K.Z., A.M., and I wrote the review together. I contributed particularly to the section focusing on emergent dynamics and temporal receptivity of cortical circuits ("Temporal receptivity in cortical circuits"). In addition, I collaborated with R.D. on conception and layout of Figure 2.

A.2 Multi-contact cooperativity in spike-timing dependent structural plasticity stabilizes networks

Moritz Deger, **Alexander Seeholzer** and Wulfram Gerstner

Submitted (preprint published at arXiv:1609.05730 [q-bio.NC]; Deger et al. (2016))

Summary

Excitatory synaptic connections in the adult neocortex consist of multiple synaptic contacts, almost exclusively formed on dendritic spines. Changes of spine shape and volume, a correlate of synaptic strength, can be tracked in vivo for weeks. The number of potential synaptic contacts between pairs of neurons is always uni-modally distributed, with an estimated mean value significantly greater than the mean number of actual contacts observed in experiments.

By unifying STDP and dendritic spine formation and removal, we show here that STDP of single excitatory synaptic contacts provides a crucial cooperative mechanism regulating structural plasticity. A novel Hebbian mechanism that limits correlations keeps networks stable and enables us to study the cooperative network effects of structural plasticity. In a network model with our correlation-stabilizing plasticity rule, many presynaptic neurons compete to make strong synaptic connections onto a postsynaptic neuron, while the synaptic contacts comprising each connection cooperate via postsynaptic firing.

We demonstrate that this local spike-timing dependent learning rule is sufficient to explain the multi-contact configuration of synapses in adult neocortical networks, both under steady-state and lesion-induced conditions. In simulations of a simplified network model of barrel cortex, our plasticity rule reproduces whisker-trimming induced rewiring of thalamo-cortical and recurrent synaptic connectivity on realistic time scales.

Author contributions

M.D. and W.G. conceived and designed the study. M.D. conducted the simulation experiments and analyses. M.D. approached me during the project for advice on the implementability of the structural plasticity model used in the study in a parallelizable fashion as part of the NEST simulator (Bos et al., 2015). After I implemented a fully tested and working version of the plasticity rule, M.D. and I collaborated to improve it in over several iterations, increasing the speed at which simulations were able to run. I then provided additional testing to ensure proper function of the code. After large network simulations became possible using this new implementation of the model, I contributed in an advisory fashion to the design and analysis of the recurrent network simulations (Fig. 6). M.D., W.G. and I wrote and finalized the manuscript together.

A.3 Algorithmic Composition of Melodies with Deep Recurrent Neural Networks

Florian Colombo, Samuel P Muscinelli, **Alexander Seeholzer**, Johanni Brea and Wulfram Gerstner

Published in: Proceeding of the 1st Conference on Computer Simulation of Musical Creativity, Huddersfield University (DOI: 10.13140/RG.2.1.2436.5683; Colombo et al. (2016))

Summary

Algorithmic composition of music is challenging, since algorithms have to be able both extract and reproduce the long-range temporal dependencies inherent to musical pieces. Recent advances in deep recurrent neural networks show these models might be able to overcome these challenges, with the further advantage that they can be trained on any given large dataset.

Here show that multi-layer neural networks trained on a large corpus of monophonic melodies can be used as automated music composers, that are able to generate new melodies coherent with the style they have been trained on. To facilitate stable representations of long-range temporal dependencies, we use gated recurrent units which are architecturally related to long short-term memory (LSTM) neurons. Both have been shown to be particularly efficient in neuronal networks when tasked with learning time series with long-range dependencies. Our network architecture processes duration and pitch in parallel, while modeling the relation between these two properties.

After training our model to predict upcoming notes of melodies taken from a corpus of Irish folk songs, we are able to operate the network in two modes. First, the network is able to generate song continuations, when provided with the beginning of a given song, while retaining melodic and harmonic features of the initially provided song fragment in the continuation. Second, we can use the network to generate whole songs, providing it only with two initial notes as “inspiration”. While we did not extensively analyze these autonomously produced songs, they differ from songs in the training set, have coherent metrical structure, and are generally pleasant to listen to.

We have thus shown that multi-layer LSTM networks can be used to extract and reproduce the musical rules of a given body of music. In a follow-up study (see Section A.4) we investigated the ability of a variant of these networks to learn different corpora, and assessed the generated music in more detail.

Author contributions

The model was conceived by F.C., S.P.M., J.B. and me. An earlier version of the model using LSTM neurons and a joint representation of rhythm and melody was investigated by F.C. under my supervision during his Master’s thesis. The model using gated recurrent units and separation of rhythm and melody was implemented by F.C. and J.B. All authors conceived the experiments and wrote the publication.

A.4 Deep Artificial Composer: A Creative Neural Network Model for Automated Melody Generation

Florian Colombo, **Alexander Seeholzer** and Wulfram Gerstner

Published in: Computational Intelligence in Music, Sound, Art and Design: 6th International Conference, EvoMUSART 2017, Amsterdam, The Netherlands, April 19–21, 2017, Proceedings (DOI: 10.1007/978-3-319-55750-2_6; Colombo et al. (2017))

Summary

Inherent complexity and structure on long time scales make the automated composition of music a challenging problem. Here we present an extension to the model presented in the previous publication (Section A.3), which we termed the Deep Artificial Composer (DAC). This study focuses on training the model to learn distinct styles of music and an assessment of the novelty of generated songs.

Similar to the previous model, we employ multi-layer networks of gated recurrent units to learn and produce monophonic melodies. In the present study we extend the model to be trainable on more general corpora (using the MIDI format) and use deeper (4 recurrent layers) architecture. We show that this enables us to train the model simultaneously on two large corpora, with two distinct musical styles (Klezmer and Irish folk).

We analyze the DAC generated melodies with a “novelty” measure, based on similarity of motifs in musical sequences. This shows that melodies “imagined” by the DAC are as novel as melodies produced by human composers. We further use the novelty measure to show that the DAC creates melodies musically consistent with either of the musical styles it was trained on. This study shows that multi-layer LSTM networks (as used in the DAC) are promising candidates for the automated composition of convincing musical pieces of any provided style.

Author contributions

Extensions to the original model were conceived, implemented and simulated by F.C. Experiments and analyses were conceived by F.C. and A.S. All authors wrote the publication.

B Supplementary Material: ICGenealogy

	Subtype	Count
Kv	A	238
	delayed rectifier (dr)	187
	Hodgkin-Huxley (HH)	155
	m	105
	not specified	63
	slow	32
	fast	23
	inward rectifier (IR, IR2, IRK)	17
	D	15
	2	14
	leak	11
	4	10
	1	8
	RP	7
	low threshold	6
	high threshold	6
	3	5
	3.1	5
	anomalous rectifier (AR)	4
	x	4
	1.2	3
	persistent	2
	ito	2
	other	1
	transient	1
	2.1	1
	KCNQ	1
	7	1
	HVA	1
	udr	1
	LVA	1
	x1	1
	Total Models	**931**
KCa	not specified	99
	AHP	58
	BK	34
	SK	32
	C	20
	2	8
	CT	6
	HH	5
	fast	3
	slow	2
	mAHP	2
	Total Models	**269**

	Subtype	Count
Nav	not specified	235
	Hodgkin-Huxley (HH)	115
	fast	82
	persistent	78
	resurgent	8
	leak	6
	1	4
	1.6	4
	slow	4
	transient	3
	1.7	3
	1.9	3
	1.2	2
	T-type	1
	1.3	1
	P-type	1
	2	1
	low threshold	1
	1.8	1
	Total Models	**553**
Cav	T-type	109
	L-type	99
	HVA	58
	N-type	33
	R-type	33
	not specified	25
	P-type	15
	low threshold	12
	LVA	10
	high threshold	9
	HH	6
	TNL	5
	E	4
	Q	4
	NPQ	3
	leak	2
	active flux	1
	2.1	1
	P/N	1
	slow	1
	Total Models	**431**
Ih	h	171
	HCN	13
	anomalous rectifier (AR)	7
	TNC	2
	HH	1
	Total Models	**194**

Table B.1 – **Figure 2.2–Source data 1: Table of subtypes for each ion type class.** Subtypes of each of the five classes currently available in the resource (Kv, Nav, Cav, KCa, Ih), sorted by the frequency of their occurrence.

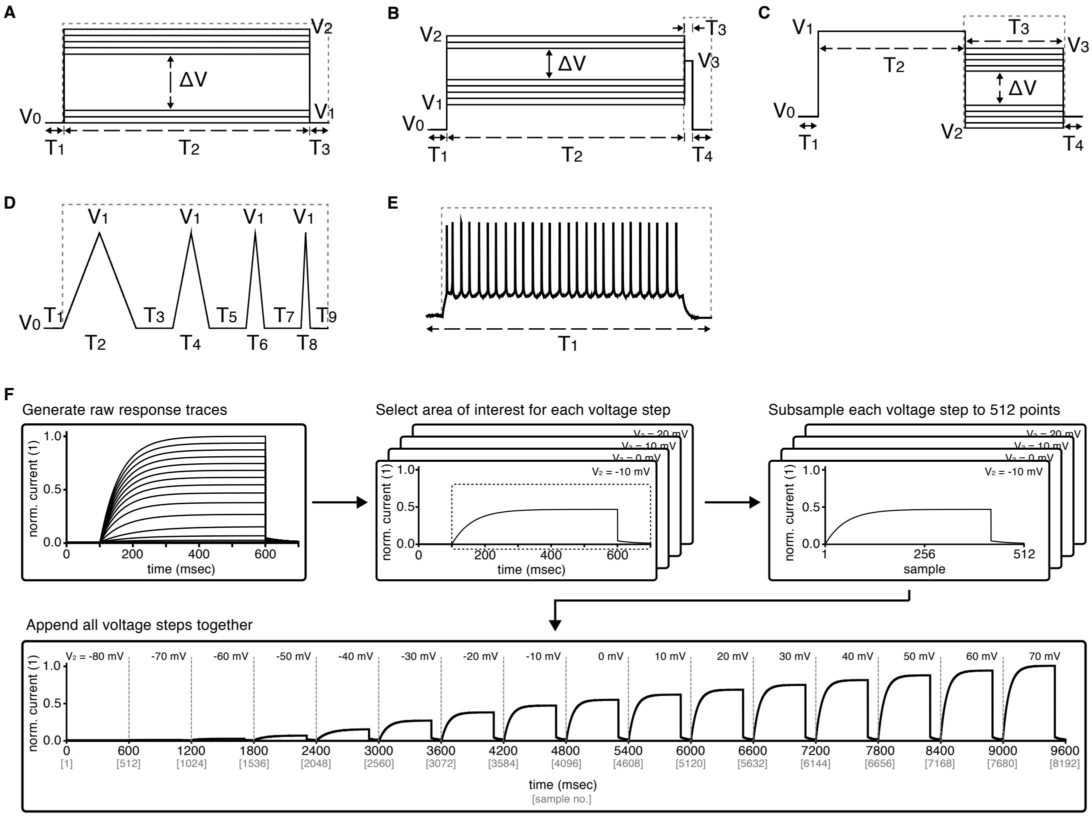

A
V2
ΔV
V0
V1
T1
T2
T3
B
V2
T3
ΔV
V3
V1
V0
T1
T2
T4
C
V1
T3
V3
T2
ΔV
V0
V2
T1
T4
D
V1
V1
V1
V1
V0
T1
T2
T3
T4
T5
T6
T7
T8
T9
E
T1
F
Generate raw response traces
norm. current (1)
1.0
0.5
0.0
0
200
400
600
time (msec)
Select area of interest for each voltage step
V2 = 20 mV
V2 = 10 mV
V2 = 0 mV
V2 = -10 mV
time (msec)
Subsample each voltage step to 512 points
1
256
512
sample
Append all voltage steps together
V2 = -80 mV
-70 mV
-60 mV
-50 mV
-40 mV
-30 mV
-20 mV
-10 mV
0 mV
10 mV
20 mV
30 mV
40 mV
50 mV
60 mV
70 mV
0 [1]
600 [512]
1200 [1024]
1800 [1536]
2400 [2048]
3000 [2560]
3600 [3072]
4200 [3584]
4800 [4096]
5400 [4608]
6000 [5120]
6600 [5632]
7200 [6144]
7800 [6656]
8400 [7168]
9000 [7680]
9600 [8192]
time (msec)
[sample no.]

Figure B.1 – **Figure 2.3–Figure supplement 1: Graphical description of the five voltage-clamp protocols used for ion channel model analysis.** Activation (**A**), inactivation (**B**), deactivation (**C**), ramp (**D**), action potential (**E**). See Table 2.2 for values of variables used in simulations. Dashed lines represent the approximate regions used for analysis. **F** Preprocessing stage of analysis – example shown for Kv *activation* protocol. Raw current response traces are cut to include only the region inside the dashed lines (as in panels *A*-*E*) for each of the (sixteen, for *activation*) voltage steps of the protocol. Then, each of the resulting traces is separately subsampled to obtain 512 data points. Finally, the subsampled traces are appended to form a final vector of size $512 * 16 = 8192$. Other protocols for Kv and the other four classes are processed similarly, with the following two comments: (1) ramp and action potential protocols do not feature multiple voltage steps, and so only contain 512 datapoints total; (2) KCa current traces are run at seven different calcium concentrations, which are processed separately for each concentration as shown here, and then appended together – e.g., KCa activation will contain $512 * 16 * 7 = 57344$ data points.

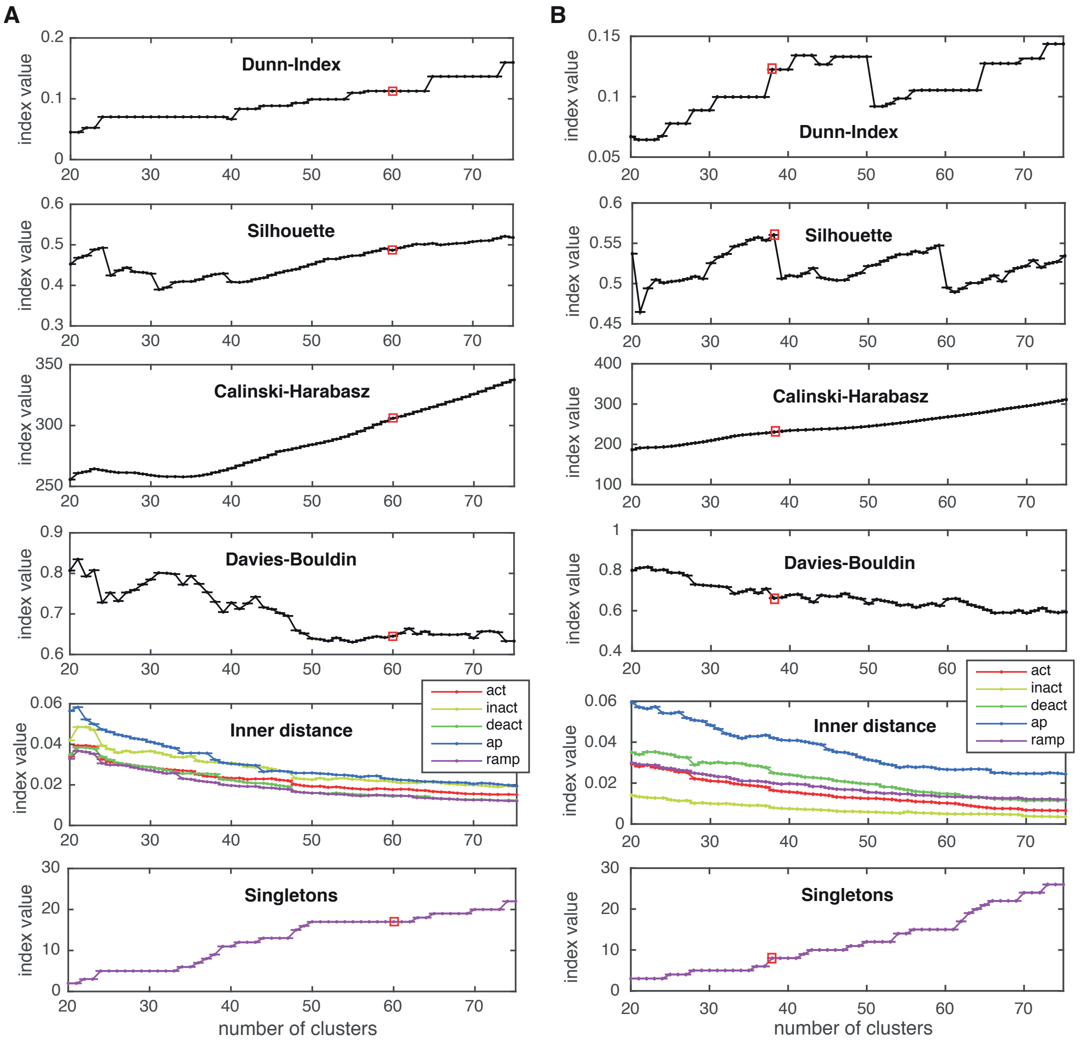


Figure B.2 – **Figure 2.3–Figure supplement 2: Cluster indexes for Kv and Nav classes.** See Methods for description. "Singletons" refers to the number of single-element clusters. **A** Cluster indexes of the Kv class. The cluster number 60 was chosen since Dunn and Silhouette indexes display a large value, Calinski-Harabasz displays a "knee" (decrease in steepness) while the Davies-Bouldin is close to a local minimum. Candidate numbers 55 (Davies-Bouldin minimum) and 59 (local maximum of Silhouette, decrease in inner distance of all conditions) were disregarded due to the lower Silhouette index, and since a splitting of a major A-type cluster into two distinct clusters was observed at 60. **B** Cluster indexes of the Nav class. The cluster number 38 was chosen since Calinski-Harabasz displays a "knee" (decrease in steepness), the Dunn index displays a sharp increase, the Silhouette index a sharp decrease for higher numbers, and the inner distance displays a slight decrease in all conditions. Chosen cluster number is indicated by a red box.

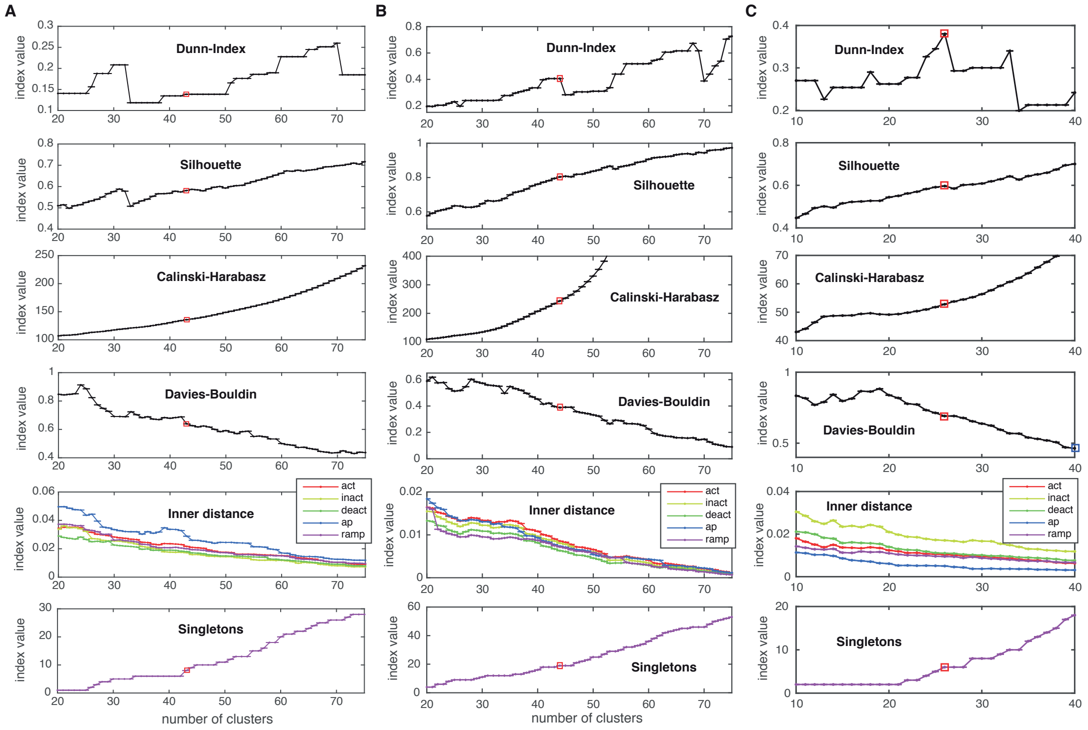


Figure B.3 – **Figure 2.3–Figure supplement 3: Cluster indexes for Cav, KCa and Ih classes.** Cluster indexes, see Methods for description. "Singletons" refers to the number of single-element clusters. **A** Cluster indexes of the Cav class. The candidate cluster number 30 (peak of Dunn-index & Silhouette, local minimum of Davies-Bouldin) was disregarded since clusters were too heterogeneous (large inner distance). The next candidate cluster number 43 was chosen instead (Dunn & Silhouette display a local increase), Calinski-Harabasz displays a slight "knee" (decrease in steepness) and Davies-Bouldin displays a sharp drop. Also, a sharp decrease in "action potential" (ap) inner distance is visible at cluster number 43. **B** Cluster indexes of the KCa class. The cluster number 44 was chosen since Calinski-Harabasz displays a slight "knee" (decrease in steepness), the Dunn index displays an increase before a sharp drop while the Silhouette increases over the candidate number of 41. Also, Davies-Bouldin, stops decreasing and the inner distance displays a decrease in all conditions. **C** Cluster indexes of the Ih class. The cluster number 26 was chosen since the Dunn index displays a clear local maximum, as does the Silhouette, while Davies-Bouldin stops decreasing. The candidate number 27 (Calinski-Harabasz displays a clear "knee") was disregarded, since the other indexes all show unfavorable changes. Chosen cluster number is indicated by a red box.

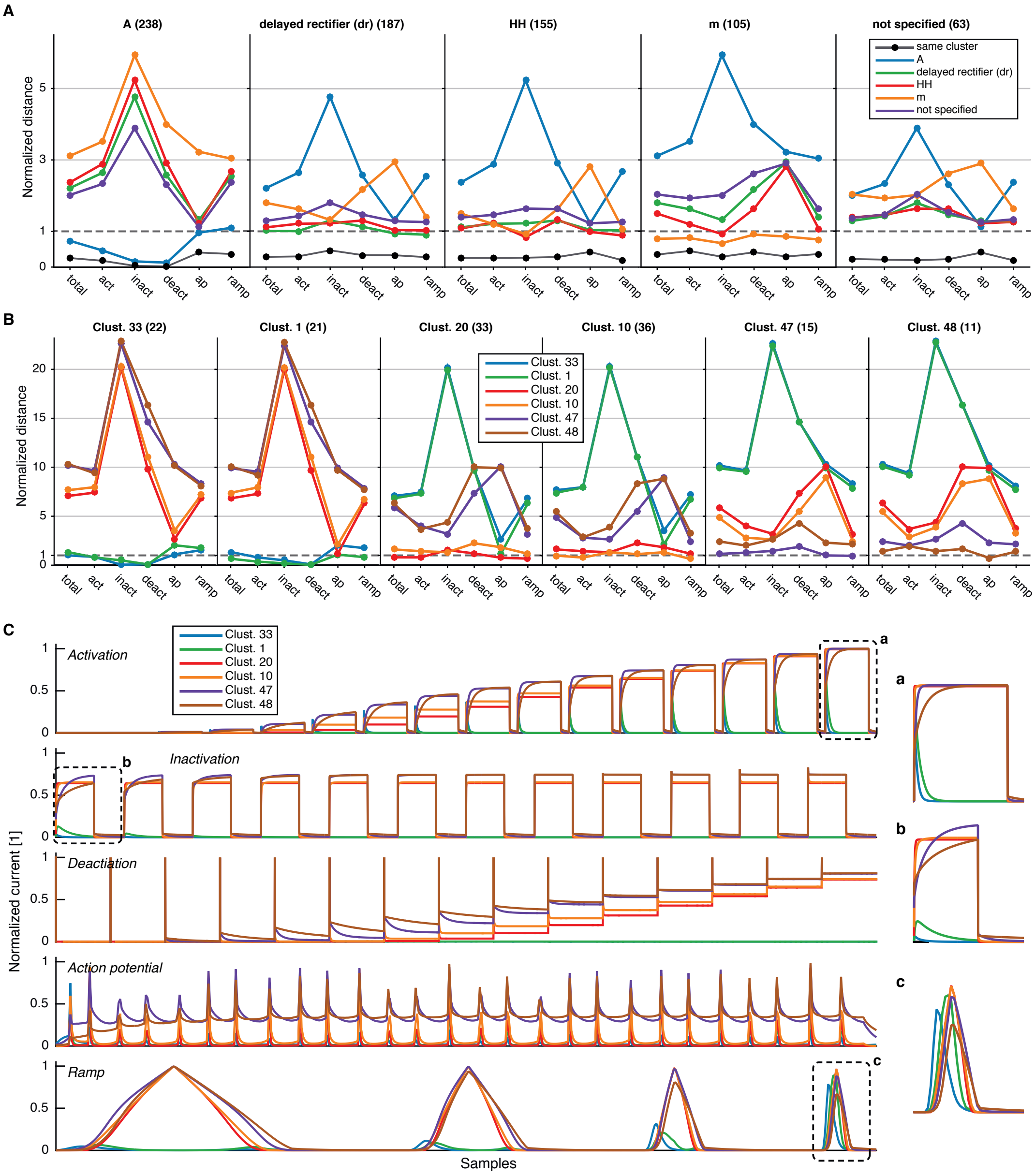

A
A (238)
delayed rectifier (dr) (187)
HH (155)
m (105)
not specified (63)
Normalized distance
0
1
3
5
total
act
inact
deact
ap
ramp
same cluster
A
delayed rectifier (dr)
HH
m
not specified
B
Clust. 33 (22)
Clust. 1 (21)
Clust. 20 (33)
Clust. 10 (36)
Clust. 47 (15)
Clust. 48 (11)
Normalized distance
0
1
5
10
15
20
Clust. 33
Clust. 1
Clust. 20
Clust. 10
Clust. 47
Clust. 48
C
Activation
Inactivation
Deactiation
Action potential
Ramp
Normalized current [1]
0
0.5
1
a
b
c
Samples

Figure B.4 – **Figure 2.3–Figure supplement 4: Comparison of intra- and inter-subtype variability with intra- and inter-cluster variability.** **A** Median pairwise distance between all channels of a given subtype, calculated either with the final score (“total”), or only each of the intermediate condition scores (here abbreviated, cf. Figure 2.3). Distance is given with respect to subtype given in each column, i.e. the first column displays the distance of channels of all other subtypes to the A subtype. Black circles display the median pairwise distance between those channels of a given subtype that are in the same clusters. All distances are normalized by the pairwise median distance averaged over all displayed subtypes. This shows that, e.g., *A* differs from other subtypes in all but the action potential protocol, while for example the *m-type* label differs from all other subtypes mostly in the action potential protocol. The clustering reduces the median variability to much lower levels than just the subtype partition, as visible in the black dots of each plot. Numbers in parentheses show the number of channels in the given subtype. **B** Same as in A, however pairwise distances are calculated for channels in the given clusters. Distance is given with respect to the cluster in each column, i.e. the first column displays the distance of channels in all clusters to channels in cluster 33. The first two clusters consist mostly of *A-type* channels (33: 80%,1: 91%), the middle two clusters comprise primarily of *HH*/*dr* channels (20: 36% HH/55% dr, 10: 39% HH/55% dr), and the right two clusters consist mostly of *m-type* channels (47: 73%,48: 73%). Numbers in parentheses correspond to the number of unique channels in the given cluster. **C** Mean current response traces of the clusters in panel B. The insets marked by a, b, c are zooms of parts of the response traces, enlarged to clarify regions of distinct current responses.

	ModelDB ID	mod name	Issue
Kv	127992	HHcn.mod	Produces abnormal oscillations for all protocols
	127992	HHcnf.mod	Produces abnormal oscillations for all protocols
	141272	hh58F1.mod	Produces abnormal current trace responses
	141272	hh58F1ss.mod	Produces abnormal current trace responses
	155601	markov.mod	Produces abnormal oscillations for all protocols
	155602	markov.mod	Produces abnormal oscillations for all protocols
	135838	GK.mod	Tab data insufficient for simulation protocol
	114637	iA.mod	Unknown Error during run(): Segmentation Violation
	140828	kir.mod	Table not specified in hoc func table - no table provided
Nav	114637	iap.mod	Produces abnormal oscillations in AP protocol
	151443	Naf_No.mod	Produces abnormal oscillations for all protocols
	153280	ch_NavPVBC.mod	Produces abnormal oscillations for all protocols
	84589	naf.mod	Produces abnormal oscillations in AP protocol
	3263	nahh.mod	Produces abnormal oscillations for all protocols
	150446	NafTraub1994sd.mod	Produces abnormal oscillations for all protocols
	127351	nax.mod	Produces abnormal oscillations for all protocols
	147867	nahh.mod	Produces abnormal oscillations for all protocols
	150239	nax.mod	Produces abnormal oscillations for all protocols
	141272	hh58F1.mod	Produces abnormal current trace responses
	141272	hh58F1ss.mod	Produces abnormal current trace responses
	155601	markov.mod	Current traces are mostly zero, does not seem to work
	155602	markov.mod	Current traces are mostly zero, does not seem to work
	127992	HHcn.mod	Current traces are mostly zero, does not seem to work
	127992	HHcnf.mod	Current traces are mostly zero, does not seem to work
	135838	GNa.mod	Tab data insufficient for simulation protocol
Cav	135838	GCa.mod	Tab data insufficient for simulation protocol
KCa	139654	kca.mod	Skipped due to mod file format (point process)
	20007	KctBG99.mod	Produces extreme large current values, unknown reason
IH	144520	isi.mod	Unknown error during run()

Table B.2 – **Figure 2.3–Source data 1: Table of omitted files.** List of omitted files for each of the five ion type classes, with a brief description of the reason for omission.

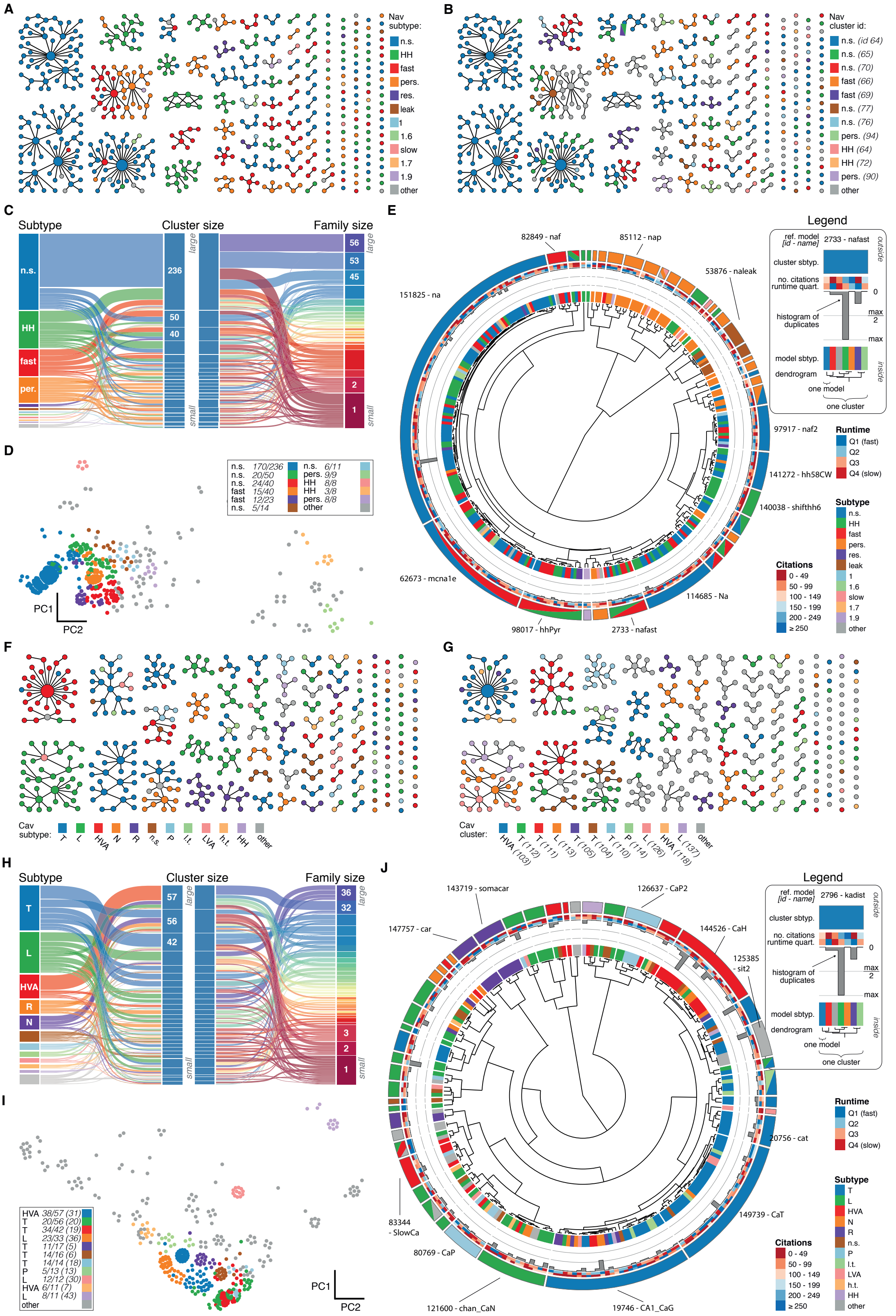

A
Nav subtype:
n.s.
HH
fast
pers.
res.
leak
1
1.6
slow
1.7
1.9
other
B
Nav cluster id:
n.s. (id 64)
n.s. (65)
n.s. (70)
fast (66)
fast (69)
n.s. (77)
n.s. (76)
pers. (94)
HH (64)
HH (72)
pers. (90)
other
C
Subtype
Cluster size
Family size
n.s.
HH
fast
per.
236
50
40
large
small
56
53
45
2
1
D
n.s. 170/236
n.s. 20/50
n.s. 24/40
fast 15/40
fast 12/23
n.s. 5/14
n.s. 6/11
pers. 9/9
HH 8/8
HH 3/8
pers. 8/8
other
PC1
PC2
E
82849 - naf
85112 - nap
53876 - naleak
151825 - na
97917 - naf2
141272 - hh58CW
140038 - shifthh6
62673 - mcna1e
114685 - Na
98017 - hhPyr
2733 - nafast
Legend
ref. model [id - name]
2733 - nafast
cluster sbtyp.
no. citations
runtime quart.
histogram of duplicates
model sbtyp.
dendrogram
one model
one cluster
outside
inside
0
max
2
max
Runtime
Q1 (fast)
Q2
Q3
Q4 (slow)
Subtype
n.s.
HH
fast
pers.
res.
leak
1
1.6
slow
1.7
1.9
other
Citations
0 - 49
50 - 99
100 - 149
150 - 199
200 - 249
≥ 250
F
Cav subtype:
T
L
HVA
N
R
n.s.
P
l.t.
LVA
h.t.
HH
other
G
Cav cluster:
HVA (103)
T (112)
T (111)
L (113)
T (105)
T (104)
T (110)
P (114)
L (126)
HVA (118)
L (137)
other
H
Subtype
Cluster size
Family size
T
L
HVA
R
N
57
56
42
large
small
36
32
3
2
1
I
HVA 38/57 (31)
T 20/56 (20)
T 34/42 (19)
L 23/33 (36)
T 11/17 (5)
T 14/16 (6)
T 14/14 (18)
P 5/13 (13)
L 12/12 (30)
HVA 6/11 (7)
L 8/11 (43)
other
PC1
PC2
J
143719 - somacar
126637 - CaP2
147757 - car
144526 - CaH
125385 - sit2
20756 - cat
149739 - CaT
83344 - SlowCa
80769 - CaP
121600 - chan_CaN
19746 - CA1_CaG
Legend
ref. model [id - name]
2796 - kadist
cluster sbtyp.
no. citations
runtime quart.
histogram of duplicates
model sbtyp.
dendrogram
one model
one cluster
outside
inside
0
max
2
max
Runtime
Q1 (fast)
Q2
Q3
Q4 (slow)
Subtype
T
L
HVA
N
R
n.s.
P
l.t.
LVA
h.t.
HH
other
Citations
0 - 49
50 - 99
100 - 149
150 - 199
200 - 249
≥ 250

Figure B.5 – **Figure 2.4–Figure supplement 1: Nav and Cav class genealogy and clustering.** **A**: Families of the Nav class, ordered from left to right by family size. Each dot represents an ion channel model and edges represent family relationships. Colors indicate the 11 most prevalent subtypes in the class, with other subtypes colored in grey. **B** Families of the Nav class, as in panel *A*. Colors indicate membership in one of the 11 largest clusters in the class (named by most prevalent subtype, right hand side), with membership to other clusters colored in grey. Cluster ID is given for easy comparison with website. **C** 'Sankey' diagram for the Nav ion type class, showing the relation between subtype, cluster identification and family identification, each ordered from top to bottom by increasing size. The 11 most common subtypes are shown, with others being grouped together (grey). Small families (size 1 to 5 members) are grouped together. **D**: Plot of Nav models in the first two principal components of score space. Colors indicate membership in one of the 11 largest clusters in the class, with membership to other clusters colored in grey. Legend indicates the most common subtype in each cluster, and the proportion of models with that subtype. Points lying very close to each other have been distributed around the original coordinate for visualization purposes. **E**:'Circos' diagram of the Nav ion type class. All unique ion channel models are displayed on a ring, organized by cluster identification. From outside to inside, each segment specifies: cluster reference model (only displayed for large clusters), cluster subtype, number of citations, runtime, number of duplicates, model subtype, as well as a dendrogram of cluster connections (black) and family relations (grey). n.s.: not specified, HH: Hodgkin-Huxley, per./pers.: persistent, res.: resurgent. **F-J**: Cav family, similar to panels *A-E*. T: T-type, L: L-type, HVA: high-voltage activated, N: N-type, R: R-type, n.s.: not specified, P: P-type, l.t.: low threshold, LVA: low-voltage activated, h.t.: high threshold.

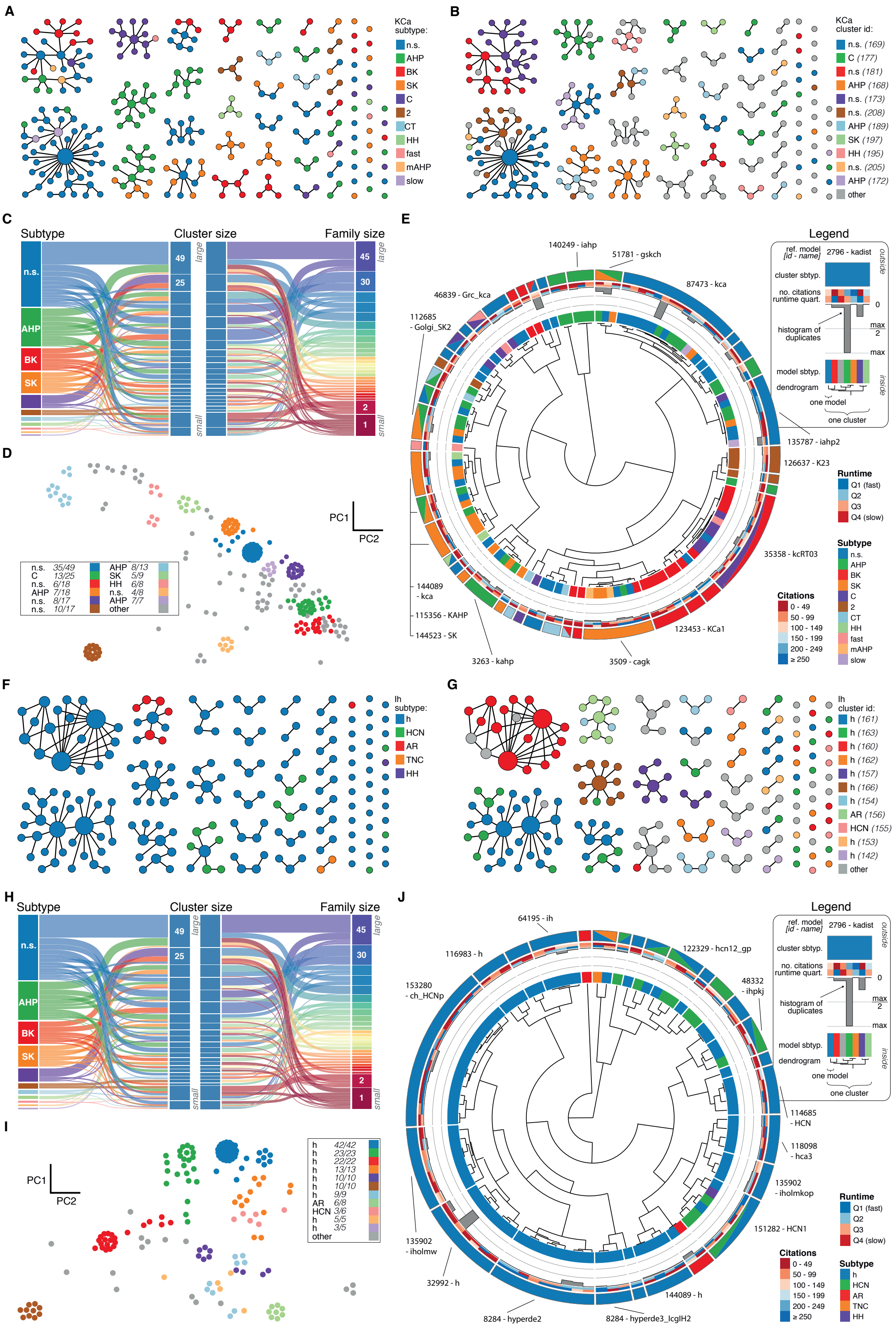

A
KCa
subtype:
n.s.
AHP
BK
SK
C
2
CT
HH
fast
mAHP
slow
B
KCa
cluster id:
n.s. (169)
C (177)
n.s (181)
AHP (168)
n.s. (173)
n.s. (208)
AHP (189)
SK (197)
HH (195)
n.s. (205)
AHP (172)
other
C
Subtype
Cluster size
Family size
n.s.
AHP
BK
SK
49
25
45
30
2
1
large
small
D
PC1
PC2
n.s. 35/49
C 13/25
n.s. 6/18
AHP 7/18
n.s. 8/17
n.s. 10/17
AHP 8/13
SK 5/9
HH 6/8
n.s. 4/8
AHP 7/7
other
E
Legend
ref. model [id - name]
2796 - kadist
cluster sbtyp.
no. citations
runtime quart.
histogram of duplicates
0
max
2
max
model sbtyp.
dendrogram
one model
one cluster
outside
inside
140249 - iahp
51781 - gskch
87473 - kca
46839 - Grc_kca
112685 - Golgi_SK2
135787 - iahp2
126637 - K23
35358 - kcRT03
123453 - KCa1
3509 - cagk
3263 - kahp
144089 - kca
115356 - KAHP
144523 - SK
Runtime
Q1 (fast)
Q2
Q3
Q4 (slow)
Subtype
n.s.
AHP
BK
SK
C
2
CT
HH
fast
mAHP
slow
Citations
0 - 49
50 - 99
100 - 149
150 - 199
200 - 249
≥ 250
F
Ih
subtype:
h
HCN
AR
TNC
HH
G
Ih
cluster id:
h (161)
h (163)
h (160)
h (162)
h (157)
h (166)
h (154)
AR (156)
HCN (155)
h (153)
h (142)
other
H
Subtype
Cluster size
Family size
n.s.
AHP
BK
SK
49
25
45
30
2
1
large
small
I
PC1
PC2
h 42/42
h 23/23
h 22/22
h 13/13
h 10/10
h 10/10
h 9/9
AR 6/8
HCN 3/6
h 5/5
h 3/5
other
J
Legend
ref. model [id - name]
2796 - kadist
cluster sbtyp.
no. citations
runtime quart.
histogram of duplicates
0
max
2
max
model sbtyp.
dendrogram
one model
one cluster
outside
inside
64195 - ih
116983 - h
122329 - hcn12_gp
48332 - ihpkj
153280 - ch_HCNp
114685 - HCN
118098 - hca3
135902 - iholmkop
151282 - HCN1
135902 - iholmw
32992 - h
144089 - h
8284 - hyperde2
8284 - hyperde3_IcgIH2
Runtime
Q1 (fast)
Q2
Q3
Q4 (slow)
Citations
0 - 49
50 - 99
100 - 149
150 - 199
200 - 249
≥ 250
Subtype
h
HCN
AR
TNC
HH

Figure B.6 – **Figure 2.4–Figure supplement 2: KCa and Ih class genealogy and clustering. A**: Families of the KCa class, ordered from left to right by family size. Each dot represents an ion channel model and edges represent family relationships. Colors indicate the 11 subtypes in the class. **B** Families of the KCa class, as in panel *A*. Colors indicate membership in one of the 11 largest clusters in the class (named by most prevalent subtype, right hand side), with membership to other clusters colored in grey. Cluster ID is given for easy comparison with website. **C** 'Sankey' diagram for the KCa ion type class, showing the relation between subtype, cluster identification and family identification, each ordered from top to bottom by increasing size. Small families (size 1 to 2 members) are grouped together. **D**: Plot of KCa models in the first two principal components of score space. Colors indicate membership in one of the 11 largest clusters in the class, with membership to other clusters colored in grey. Legend indicates the most common subtype in each cluster, and the proportion of models with that subtype. Points lying very close to each other have been distributed around the original coordinate for visualization purposes. **E**: 'Circos' diagram of the KCa ion type class. All unique ion channel models are displayed on a ring, organized by cluster identification. From outside to inside, each segment specifies: cluster reference model (only displayed for large clusters), cluster subtype, number of citations, runtime, number of duplicates, model subtype, as well as a dendrogram of cluster connections (black) and family relations (grey). n.s.: not specified, AHP: after-hyperpolarization, BK: big conductance, SK: small conductance, HH: Hodgkin-Huxley, mAHP: medium AHP. **F-J**: Ih family, similar to panels *A*-*E*. h: hyperpolarization-activated, HCN: hyperpolarization-activated cyclic nucleotide-gated, AR: anomalous rectifier, TNC: tonic nonspecific current.

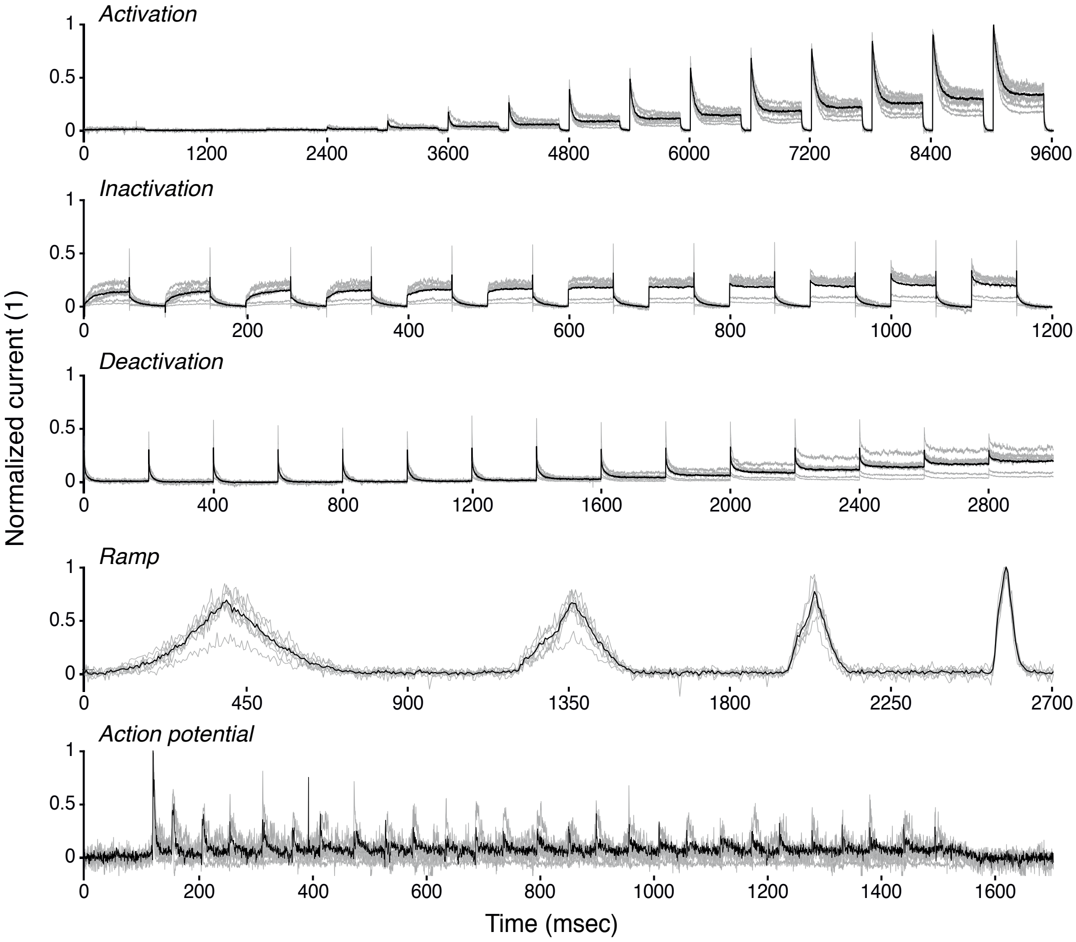


Figure B.7 – **Figure 2.6–Figure supplement 1: K+ current recordings from Drosophila Kenyon cells.** Recordings from 8 example traces of K+ current from Drosophila Kenyon cells (grey), with average (black). See Methods for a description of experiments. Response traces are arranged as detailed in Figure 2.3–Figure supplements 1 & 2.

C Supplementary Material: Efficient low-dimensional approximation

Parameter name	Parameter Symbol	Units	Excitatory neurons	Inhibitory neurons
Neuron number	$N_\bullet$	1	$N_E = 800$	$N_I = 200$
Poisson neuron number	N_{ext}	1	1000	1000
Membrane capacitance	C_m	pF	500	200
Exc. reversal potential	V_E	mV	0	0
Inh. reversal potential	V_I	mV	-70	-70
Leak reversal potential	V_L	mV	-70	-70
After spike reset potential	V_{res}	mV	-60	-60
Spiking threshold	V_{thr}	mV	-50	-50
NMDA parameter 1	β	1	0.062	0.062
NMDA parameter 2	γ	1	$1/3.57$	$1/3.57$
NMDA rise parameter	α	s	0.5	0.5
External conductance	g_{ext}	nS	2.08	1.62
Recurr. inh. conductance	g_{I}	nS	1.336	1.024
Recurr. exc. conductance	g_{E}	nS	0.381	0.292
Leak conductance	g_{L}	nS	25	20
External synaptic timescale	τ_{ext}	ms	2	2
Recurr. inh. timescale	τ_I	ms	10	10
Recurr. exc. timescale	τ_E	ms	100	100
Recurr. exc. rise timescale	$\tau_{\text{E,rise}}$	ms	2	2
Membrane time constant $\frac{C_m}{g_L}$	τ_m	ms	20	10
Refractory period	τ_{ref}	ms	2	1
Width of distance dep. weights	σ_w	rad	$\frac{18\,\text{deg}}{360\,\text{deg}} \cdot 2\pi \approx 0.31$	-
Frequency of Poisson neurons	ν_{ext}	Hz	$2.4Hz$	$2.4Hz$

Table C.1 – **Parameters for spiking simulations**. Parameter values are modified from (Compte et al., 2000) and Brunel and Wang (2001b).

System	w_0	w_1	w_σ	w_r
Sys. 0	-0.8	2.3	0.9	2.0
Sys. 1	-1.0	10.	0.2	2.0
Sys. 2	-3.0	15.0	0.5	2.0

Table C.2 – **Connectivity parameters of rate models.**

Figure	Fixed	Optimized	# Optimized	Points / Errors
Fig. 3.4	w_+, w_σ $g_{\mathrm{EE}}, g_{\mathrm{IE}}, g_{\mathrm{EI}}, g_{\mathrm{II}}$	g_0, g_1, g_σ, g_r $\langle V\rangle_{\theta_1}, \dots, \langle V\rangle_{\theta_4}$ $\nu_I, \langle V\rangle_I$	10	4 / 10
Fig. 3.6 red	$g_1 = 50Hz$ $g_\sigma = 0.6$	g_0, g_r $w_+ = 4.100, w_\sigma = 0.1899$ $g_{\mathrm{EE}} = 0.3923, g_{\mathrm{IE}} = 0.3958$ $g_{\mathrm{EI}} = 1.1611, g_{\mathrm{II}} = 0.9570$ $\langle V\rangle_{\theta_1}, \dots, \langle V\rangle_{\theta_7}$ $\nu_I, \langle V\rangle_I$	17	7 / 16
Fig. 3.6 blue	$g_1 = 20Hz$ $g_\sigma = 1.2$	g_0, g_r $w_+ = 2.423, w_\sigma = 0.4750$ $g_{\mathrm{EE}} = 0.1798, g_{\mathrm{IE}} = 0.1858$ $g_{\mathrm{EI}} = 0.7882, g_{\mathrm{II}} = 0.7632$ $\langle V\rangle_{\theta_1}, \dots, \langle V\rangle_{\theta_7}$ $\nu_I, \langle V\rangle_I$	17	7 / 16
Fig. 3.6 green	$g_1 = 25Hz$ $g_r = 8$	g_0, g_σ $w_+ = 4.4917, w_\sigma = 0.0909$ $g_{\mathrm{EE}} = 0.4397, g_{\mathrm{IE}} = 0.4742$ $g_{\mathrm{EI}} = 1.1933, g_{\mathrm{II}} = 1.1948$ $\langle V\rangle_{\theta_1}, \dots, \langle V\rangle_{\theta_7}$ $\nu_I, \langle V\rangle_I$	17	7 / 16

Table C.3 – **Parameters optimized in spiking networks.** For constraints in Fig. 3.4 see the parameter values in Table C.1. For network parameters in Fig. 3.6, we additionally give the values obtained by optimization. *Points* is the number of sampling points. *Errors* is the number errors used for optimization, this includes 2 errors for the inhibitory population, in addition to 2 errors per sampling point.

D Supplementary Material: Effects of short-term plasticity on drift and diffusion

D.1 Extension: Synaptic connections with saturating NMDA conductance

In the theory we developed above, we have assumed the dynamics of synaptic activation variables s to be linear. According to Eq. (4.4), incoming spikes are dynamically scaled by short-term facilitation and depression, while the variable s can, in principle, increase without bounds. In the networks considered so far, short-term depression took over the role of saturating the maximally achievable synaptic activations at high firing rates. For example, in a purely depressing system (set $U \to 1$) the steady-state synaptic activations $s_{0,i}$ of Eq. (4.21) are upper bounded by $\frac{\tau_s}{\tau_x}$ since then:

$$s_{0,i} = \frac{\tau_s \phi_{0,i}}{1 + \tau_x \phi_{0,i}} \tag{D.1}$$

A different method of achieving this dynamical upper bounding is by including nonlinear saturation of NMDA conductances (see e.g. Wang (1999); Compte et al. (2000)), as already introduced in Section 3.6.2.2 (Eqs. (3.15)-(3.15)). We introduce here a rate based version of these dynamics:

$$\begin{aligned} \dot{s}_i^{\text{sat}} &= -\frac{s_i^{\text{sat}}}{\tau_s} + \alpha r_i \left(1 - s_i^{\text{sat}}\right), \\ \dot{r}_i &= -\frac{r_i}{\tau_{\text{rise}}} + \phi_i, \\ J_i &= \sum_j w_{ij} s_j^{\text{sat}}. \end{aligned} \tag{D.2}$$

Here s_j^{sat} is the synaptic activation, r_j acts as a filter on the firing rates $\phi_{0,i}$, and α is a constant with units of $\frac{1}{s}$. We call this “logistic”, since the term $(1 - s_i^{\text{sat}})$ on the right hand in Eq. (D.2) is reminiscent of the upper bound of a logistic differential equation (see e.g. Strogatz (2000)). It is this term that causes the upper bounding of synaptic activations at high rates. In fact, the steady-state of the synaptic dynamics of Eq. (D.2) can be shown to be almost equivalent to that of Eq. (D.1) with depressive synapses only. Solving the dynamics of Eq. (D.2) for 0, we arrive at

$$s_{0,i}^{\text{sat}} = \frac{\alpha \tau_s \tau_{\text{rise}} \phi_{0,i}}{1 + \alpha \tau_s \tau_{\text{rise}} \phi_{0,i}}.$$

Thus, if we identify $\alpha = \frac{1}{\tau_{\text{rise}}}$, we arrive at a saturation behavior similar to Eq. (D.1) (with $\tau_s = \tau_x$).

We were interested in the effects that including logistic saturation of NMDA conductances had on drift and diffusion, especially in interplay with short-term plasticity. To approach this, we included short-term plasticity according to Eq. (4.4) in the equations of Eq. (D.2), by scaling the firing rates by the short-term plasticity variables (all other equations stay the same).

$$\dot{r}_i = -\frac{r_i}{\tau_{\text{rise}}} + \phi_i u_i x_i.$$

Similar to our approach in the system with only facilitation and depression, we linearize these dynamics including the logistic NMDA saturation (cf. Eq. (4.23)). This yields the following linear stochastic differential equation:

$$\begin{pmatrix} \dot{\delta \mathbf{r}} \\ \dot{\delta \mathbf{s}} \\ \dot{\delta \mathbf{u}} \\ \dot{\delta \mathbf{x}} \end{pmatrix} = \begin{pmatrix} -\frac{1}{\tau_{\text{rise}}}\mathbb{1} & D\left(\mathbf{u}_0 \cdot \mathbf{x}_0 \cdot \phi_0'\right) W & D\left(\phi_0 \cdot \mathbf{x}_0\right) & D\left(\phi_0 \cdot \mathbf{u}_0\right) \\ D(1-\mathbf{s}_0)\alpha & -\frac{1}{\tau_s} - D(\mathbf{r}_0)\alpha & 0 & 0 \\ 0 & UD\left((1-\mathbf{u}_0)\cdot \phi_0'\right) W & -\frac{1}{\tau_u}\mathbb{1} - UD\left(\phi_0\right) & 0 \\ 0 & -D\left(\mathbf{u}_0 \cdot \mathbf{x}_0 \cdot \phi_0'\right) W & -D(\mathbf{x}_0 \cdot \phi_0) & -\frac{1}{\tau_x}\mathbb{1} - D\left(\phi_0 \cdot u_0\right) \end{pmatrix} \begin{pmatrix} \delta \mathbf{r} \\ \delta \mathbf{s} \\ \delta \mathbf{u} \\ \delta \mathbf{x} \end{pmatrix} + \begin{pmatrix} \eta \mathbf{u}_0 \mathbf{x}_0 \sqrt{\phi_0} \\ 0 \\ \eta U(1-\mathbf{u}_0)\sqrt{\phi_0} \\ -\eta \mathbf{u}_0 \mathbf{x}_0 \sqrt{\phi_0} \end{pmatrix} \tag{D.3}$$

Similar to Section 4.6.1.4, we can calculate the right eigenvector of this linear kernel for movement along the attractor manifold under the assumption that the system has a one-dimensional continuous manifold of steady states. Unfortunately, we have not managed to derive the corresponding left eigenvector for this system. Mainly, the Ansatz $\frac{d}{dt} t_i = \frac{d}{dt} \sum_j w_{ij} s_j$, which we used to derive the left eigenvector (cf. Section 4.6.1.5) does not produce a differential equation that we are able to identify with the original system. However. numerical evaluation of the left eigenspace of Eq. (D.3) for a given implementation is indeed possible, and would allow numerical predictions of the amount of drift and diffusion expected in this system.

We should remark that the steady state of the simple rate model proposed in Eq. (D.2) does not faithfully approximate the steady state of the spiking network implementation given in Eqs. (3.15)-(3.15) (see Fig. D.1, dashed red line). However, a better approximation under the

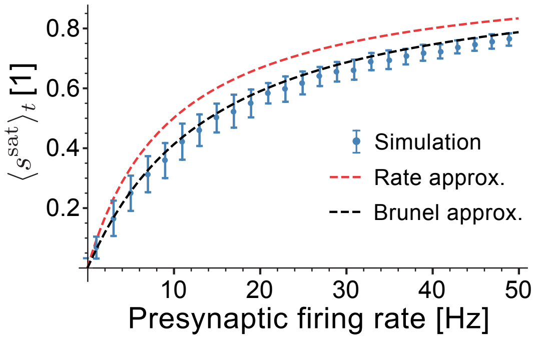


Figure D.1 – **Steady state of synaptic activation with logistic NMDA saturation.** Temporal mean of synaptic activations with logistic saturation s_j^{sat} (Eq. (3.15)), stimulated by spike trains with spike counts distributed as homogeneous Poisson processes (blue dots). The steady state of the naive rate approximation (dashed read line, Eq. (D.2)) is compared to the approximation of Brunel and Wang (2001b, p. 80) (dashed black line). Temporal means are averaged over 100 repetitions of $2s$ recorded time after allowing $1s$ for equilibration. Error bars display 1 standard deviation of the temporal means across repetitions.

assumption of Poisson statistics has been stated in Brunel and Wang (2001b, p. 80) (Fig. D.1, dashed black line; the infinite sum of Brunel and Wang (2001b, p. 80) was truncated at 10 summands), which could serve to predict the steady-state s_0, given the stable firing rates ϕ_0, more accurately. When linearized at the resulting steady-state prediction, the naive rate approximation of Eq. (D.3) could still give reasonable numerical results.

At the time of writing, it remains an open question whether the additional introduction of logistic NMDA saturation has any significant impact on the emergent diffusion and drift dynamics. We have sketched here a first starting point to incorporate these models into the frame of our theory. Preliminary simulations with facilitation and logistic NMDA saturation had indicated that the stabilizing effects of facilitation might be even more pronounced than in the case of short-term facilitation and depression considered in the earlier sections. According to the results of the main text, this is in principle to be expected, since synaptic depression increases the strength of drift and diffusion (see main text). However, we can not exclude a dynamical impact of logistic NMDA saturation on drift and diffusion along the attractor manifold, which remains to be investigated and quantified. By the considerations above, including both depression and logistic NMDA saturation introduces doubly saturating synaptic conductances – incoming spikes are scaled down at high rates, and their effect on the synaptic conductances is logistically bounded. Nevertheless, this could be a redundancy only at first glance: it could indeed be possible that NMDA saturation reduces the tendency of short-term depression to increase drift and diffusion on continuous attractor networks.

D.2 Extension: Simplified theory by projection onto Fourier modes

In Itskov et al. (2011b) a slightly simpler approach is taken to calculating the effects of short-term facilitation (for fixed depression) on a rate-based ring-attractor model. Instead of calculating the exact projection eigenvectors for the shape of the firing rate profile as we have done above, the authors approximate the drift terms by projecting a combined variable $\phi \cdot u \cdot x$ onto the dynamics of its first Fourier coefficient, and calculating the change of this Fourier coefficient under the system dynamics. In this section we perform a similar calculation for the drift and diffusion strengths, with the auxiliary variable $s \cdot u \cdot x$, for which we have calculated the linearized dynamics above in Eq. (4.23). This approach yields approximative equations that seem to capture the limit of large diffusion less well than the approach demonstrated in the main text.

Let $m(\theta,\varphi) = s(\theta,\varphi)u(\theta,\varphi)x(\theta,\varphi)$ be the auxiliary variable composed of synaptic activations s, facilitation u, and depression x, for a firing rate profile $\phi(\theta,\varphi)$ centered at an angle φ. Here, $-\pi \le \theta < \pi$ are possible positions of neurons arranged on a ring. As in Itskov et al. (2011b) we identify now the first Fourier component of the auxiliary variable m with the position of the rate profile φ – both m and ϕ are unimodal shapes centered at φ . We can calculate the first Fourier coefficient as a complex number

$$x(t) \equiv r(t)\exp(i\varphi(t)) = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta \exp(i\theta)\, m(\theta,\varphi) \tag{D.4}$$

and take the derivative with respect to time, to arrive at a differential equation of the position φ:

$$\begin{aligned}\dot{x} = \dot{r}\exp(i\varphi) + i r\dot{\varphi}\exp(i\varphi) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta \exp(i\theta)\,\dot{m}(\theta,\varphi)\\ \dot{r} + i r\dot{\varphi} = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta &\left[\cos(\theta-\varphi)\,\dot{m}(\theta,\varphi,t) + i\sin(\theta-\varphi)\,\dot{m}(\theta,\varphi,t)\right].\end{aligned}$$

We now assume for simplicity that the distribution is located at the center $\varphi = 0$ and drop this explicit index. This immediately implies (cf. Eq. (D.4) for $\varphi = 0$) that

$$r = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta \cos(\theta)\, m(\theta).$$

Further, we also assume that $\dot{r} = 0$, since the distribution of rates only shifts its phase along the attractor manifold while keeping its shape stable. Then, by comparing complex and real parts, we see that:

$$\begin{aligned}\dot{\varphi} &= \frac{1}{2\pi r}\int_{-\pi}^{\pi} d\theta \sin(\theta)\,\dot{m}(\theta,t)\\ &= \frac{1}{2\pi r}\int_{-\pi}^{\pi} d\theta\, sin(\theta)\left[\dot{s}ux + s\dot{u}x + su\dot{x}\right].\end{aligned} \tag{D.5}$$

Now, inserting the linearized dynamics of Eq. (4.23) (we replace all variables with their perturbations around the steady state, keeping only linear orders) we arrive at

$$\dot{\varphi} = \frac{1}{2\pi r}\int_{-\pi}^{\pi} d\theta \sin(\theta)\left[\left(\cdots + u_0 x_0\sqrt{\phi_0}\eta(\theta)\right)u_0 x_0 + \right. \\ \left. s_0\left(\cdots + U(1-u_0)\sqrt{\phi_0}\eta(\theta)\right)x_0 + s_0 u_0\left(\cdots - u_0 x_0\sqrt{\phi_0}\eta(\theta)\right)\right] \tag{D.6}$$

Here, $\langle\eta(\theta)\rangle = 0$ and $\langle\eta(\theta,t)\eta(\theta',t)\rangle = \delta(t-t')\delta(\theta-\theta')$, and the dots indicate the linear parts of the linearization in Eq. (4.23). We assume that the deterministic part of the linearized system has no dynamics acting on the phase φ. This allows calculating the diffusion strength of Eq. (D.6) by evaluating

$$\begin{aligned}\langle\dot{\varphi}(t)\dot{\varphi}(t+\tau)\rangle &= \frac{1}{4\pi^2 r^2}\int_{-\pi}^{\pi} d\theta \sin^2(\theta)\delta(\tau)\left[u_0^4 x_0^4\phi_0 + s_0^2 x_0^2 U^2(1-u_0)^2\phi_0 + s_0^2 u_0^4 x_0^2\phi_0 + 2\,(\text{crossterms})\right] \\ &= \frac{U^4}{4\pi^2 r^2}\int_{-\pi}^{\pi} d\theta \sin^2(\theta)\delta(\tau)\phi_0\left(\phi_0\tau_u+1\right)^2 \cdot \\ &\quad \frac{\left[\left(\phi_0\tau_u+1\right)\left(U\phi_0\tau_u+1\right) - \tau_s\phi_0\left(U^2\phi_0\tau_u + U\left(\phi_0^2\tau_u^2+\phi_0\tau_u+2\right)-1\right)\right]^2}{\left(U\phi_0\tau_u+1\right)^2\left(U\phi_0\left(\tau_u+\phi_0\tau_u\tau_x+\tau_x\right)+1\right)^4}\end{aligned} \tag{D.7}$$

For the second equation we have substituted $s_0 \to \tau_s u_0 x_0 \phi_0$ and inserted the steady-state values Eq. (4.21) for u_0 and x_0. The same substitutions can be used to evaluate the normalization constant r as

$$r = \frac{U^2}{2\pi}\int_{-\pi}^{\pi} d\theta \cos(\theta)\frac{\phi_0\tau_s\left(\phi_0\tau_u+1\right)^2}{\left(U\phi_0\left(\tau_u+\phi_0\tau_u\tau_x+\tau_x\right)+1\right)^2}. \tag{D.8}$$

Calculating a drift-field is possible by replacing the terms in the integrand of Eq. (D.6) by those obtained by introducing fixed small perturbations in the input $\phi_0 \to \phi_0 + \delta\phi$ (cf. 4.55):

$$\begin{aligned}\dot{\varphi} &= \frac{1}{2\pi r}\int_{-\pi}^{\pi} d\theta \sin(\theta)\left[\left(u_0 x_0\delta\phi\right)u_0 x_0 + s_0\left(U(1-u_0)\delta\phi\right)x_0 + s_0 u_0\left(-u_0 x_0\delta\phi\right)\right] \\ &= \frac{U^2}{2\pi r}\int_{-\pi}^{\pi} d\theta \sin(\theta)\delta\phi \\ &\quad \frac{\left(\phi_0\tau_u+1\right)\left(\phi_0\tau_s\left(U\left(\phi_0\tau_u\left(\phi_0\tau_u+U+1\right)+2\right)-1\right)-\left(\phi_0\tau_u+1\right)\left(U\phi_0\tau_u+1\right)\right)}{\left(U\phi_0\tau_u+1\right)\left(U\phi_0\left(\tau_u+\phi_0\tau_u\tau_x+\tau_x\right)+1\right)^2}\end{aligned} \tag{D.9}$$

We compared the solutions for diffusion Eq. (D.7) and drift Eq. (D.9) to those obtained from our theory in Eq. (4.3) (see Fig. D.2). To compare the diffusion coefficients, we normalized the coefficient obtained from Eq. (D.7) for $U = 1, \tau_x = 150ms$ to that obtained from our theory for the same parameters (blue line in Fig. 4.7) and varied the parameter U. We find that the dependence of the diffusion coefficient on the parameter U is similar, (Fig. D.2A), although the approximated solution consistently lies above the full theory (and simulations, Fig. 4.7A). Varying the parameter τ_x around $150ms$ for a fixed value of $U = 0.8$ (same experiment as in Fig. 4.7B), shows that the dependence on τ_x is not well captured by the approximation. Comparing the drift-field obtained by the theory in Eq. (D.9) to that obtained by Eq. (4.3) for the same parameter settings, coefficients, and perturbations $\delta\phi$, we find that the predicted fields are somewhat similar in magnitude (standard deviations: $10.2deg^2/s$ for the full theory, $9.6deg^2/s$ for the approximation), with most zero crossings closely related (Fig. D.2C, black lines). Decreasing the facilitation parameter to $U = 0.1$ for the same perturbations, we also find that the resulting fields are of reduced magnitude with respect to $U = 1$ (Fig. D.2C, blue lines), although the full theory predicts a larger decrease (standard deviations: $2.4deg^2/s$ for the full theory, $5.0deg^2/s$ for the approximation).

While these observations are preliminary, they show that the full theory predicts the magnitude of diffusive dynamics along the attractor manifold more accurately, while in general the approximate approach presented here bears some similarities to the result of the full theory. Some deviations are to be expected, since a “projection” is performed here onto a simplified approximation of the change of inputs to the system under movement along the attractor manifold: the terms $\frac{dJ}{d\varphi}$ of the full theory Eq. (D.7) correspond to the terms $\sin(\theta)$ here. Further, the first order scaling with respect to U is similar, since the U^2 scaling of the normalization constant r Eq. (D.8) cancels with that of the integrands, as before (Eq. (D.7)). Further investigation is needed to uncover the differences observed in the limit of large τ_x. Indeed, we have seen earlier (section Section 4.6.2.1) that the diffusion strength grows large mainly due to the normalization constant S, which becomes small – in Eq. (D.8) we find no such τ_s-dependent root for the integrand of the normalization Eq. (D.8).

It should be noted, that we have also tried performing this calculation on the auxiliary variable $\phi \cdot u \cdot x$, following the approach of Itskov et al. (2011b) (they take $\phi \cdot u$, for only facilitation). This was not pursued further, however: this rearrangement yields non-trivial correlations when calculating the diffusion strength, since all variables ϕ_i correlate. For example, the following terms appear:

$$\begin{aligned}\dot{\phi}_i &= \phi_i' \sum_j w_{ij} \delta \dot{s}_j = \phi_i' \sum_j w_{ij} \left(\text{linear terms} + u_{0,j} x_{0,j} \sqrt{\phi_{0,j}} \eta_j \right), \\ \langle \dot{\phi}_i \dot{\phi}_k \rangle &= \phi_i' \phi_k' \sum_j w_{ij} w_{kj} u_{0,j}^2 x_{0,j}^2 \phi_{0,j}.\end{aligned}$$

Thus, the double integral in Eq. (D.7) does not collapse to the single integral we observed here, which complicates the evaluation of the diffusion strength considerably.

D.3 Analysis of CVs in the attractor state

A previous study by Barbieri and Brunel (2007) has found that short-term depression together with a high reset of the membrane potential after spikes can lead to a high variability of

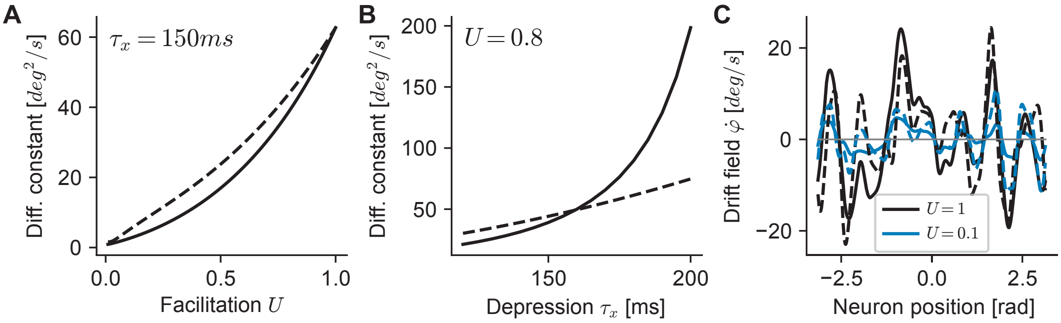


Figure D.2 – **Comparison of full theory to approximation by projection onto Fourier coefficients.** **A** Diffusion strength as a function of the facilitation variable U. Full line is from full theory (compare to Fig. 4.7A), dashed black line is from the approximation Eq. (D.7) *normalized to coincide* at $U = 1$. **B** Dependence of the diffusion on the depression time constant τ_x for $U = 0.8$. Full line is from full theory (compare to Fig. 4.7B), dashed black line is from the approximation Eq. (D.7) normalized by the same value as in panel A. **C** Drift fields extracted by the full theory (full lines) and approximation Eq. (D.9) (dashed lines) for $U = 1$ (black) and $U = 0.1$ (blue). The perturbation used to generate the field was a perturbation in the membrane leak potentials with $\sigma_L = 1mV$.

neuronal spiking in the class of spiking attractor network models that we are using (see the discussion). To investigate whether varying the parameters of short-term plasticity in our network has any implications for the spiking statistics of the network model, we measured the mean rates and CVs of excitatory neurons during the attractor state for different networks (Fig. D.3). We find similar a dependence of the CV to the mean firing rate (Fig. D.3A) as reported previously (Compte et al., 2000; Barbieri and Brunel, 2007): across all parameter settings networks display similarly reduced CVs for increased mean firing rates, leading to large CVs for neurons located in the flanks of the firing rate profile and low CVs for neurons located near the center.

However, attention should be payed when measuring the CV in the attractor network with large diffusion coefficients. Since firing rates are non-stationary in these networks, a moving center of the firing rate profile can lead to increased CVs in single units, especially at the flanks of the profile. In Fig. D.3A we show mean firing rates and CVs for excitatory neurons in a network with large diffusion coefficient ($\tau_x = 200ms$), which displayed a shift of the center towards an increased angular position during the recording period. This is visible in a reduced mean firing rate, slightly increased width, and shifted center of the mean firing rates. The non-stationarity is also visible in the measured CVs, which are elevated especially at intermediate rates (Fig. D.3A, bottom). Mistaking this non-stationarity for a high CV can be evaded by measuring the CV_2 (Compte et al., 2003), a localized version of the CV. The CV_2 measure consistently reports values similar to stationary firing (blue dots in Fig. D.3), even in networks with high diffusion coefficients. Unsurprisingly, for strong facilitation ($U \ll 1$) the CV and CV_2 measures coincide (blue dots in Fig. D.3B).

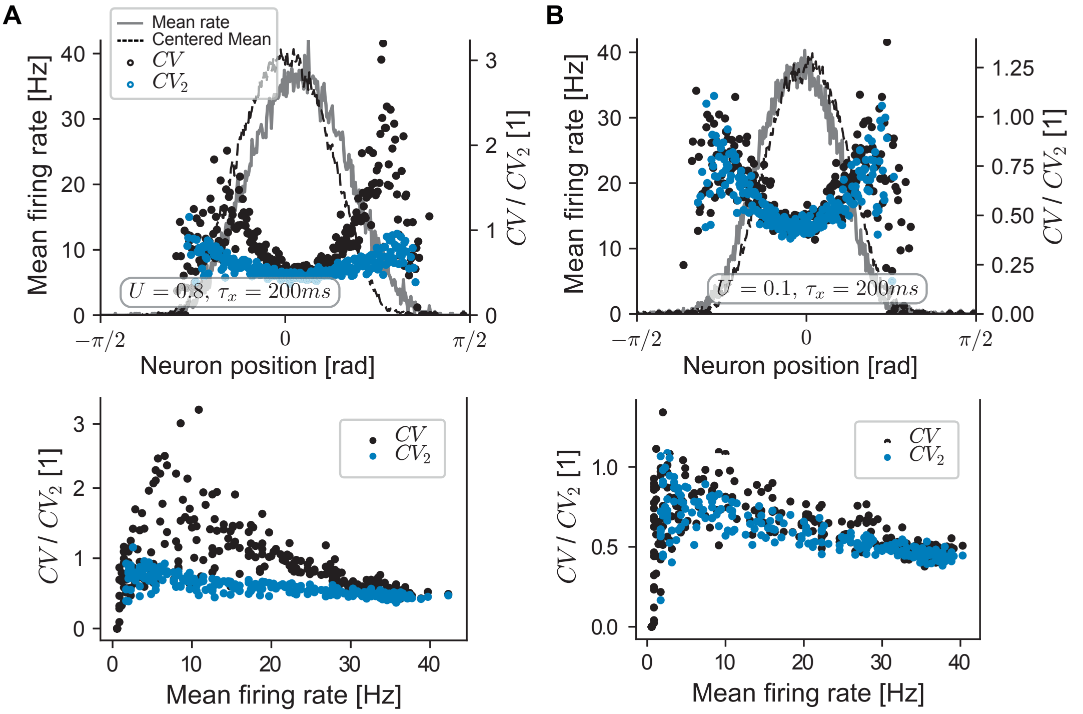


Figure D.3 – **Short-term plasticity does not affect spiking statistics.** Mean firing rate, coefficient of variation of the inter-spike interval distribution (CV), and local CV (CV_2, see main text) for two attractor networks. All measures were computed on spike-trains measured over a period of $4s$, recorded $500ms$ after offset of the external input which was centered at angle 0. **A** Networks with large diffusion coefficient ($U = 0.8$, $\tau_u = 650ms$, $\tau_x = 200ms$) that underwent non-stationary diffusion during the recording of spikes: the measured mean firing rates (gray line) differ visibly from the firing rates estimated after centering the firing rate distribution at each point in time. Due to this non-stationarity, CVs at intermediate firing rates appear elevated, while the local CV (CV_2) shows values close to stationary networks (see B). **B** The same network as in A, with strong facilitation ($U = 0.1$). Reduced diffusion leads to a nearly stationary firing rate profile, and coincident CV and CV_2 measures.

Parameter name	Parameter Symbol	Units	Excitatory neurons	Inhibitory neurons
NMDA parameter 1	β	1	0	0
NMDA parameter 2	γ	1	0	0
Synaptic timescale	τ_s	ms	100	100
Frequency of Poisson neurons	ν_{ext}	Hz	$2.6Hz$	$2.6Hz$

Table D.1 – **Parameters for spiking simulations.** All parameters are similar to Table C.1 except for those noted above and in Table D.2, where the values of recurrent conductances depending on the settings of facilitation and depression are listed.

STP U [1]	STP τ_u [ms]	STP τ_x [ms]	g_{EE} [nS]	g_{IE}[nS]	g_{EI}[nS]	g_{II}[nS]	w_σ[nS]
1	650	150	0.03489	0.004975	2.639	1.637	0.38
0.8	650	150	0.03421	0.005167	2.537	1.641	0.38
0.6	650	150	0.0328	0.005485	2.418	1.647	0.4
0.4	650	150	0.0315	0.00529	2.295	1.643	0.4
0.2	650	150	0.03356	0.005349	2.18	1.644	0.35
0.1	650	150	0.03393	0.005906	2.107	1.655	0.4
0.08	650	150	0.03441	0.005746	2.092	1.652	0.4
0.06	650	150	0.03646	0.006018	2.079	1.657	0.4
0.04	650	150	0.03771	0.005764	2.062	1.652	0.42
0.1	650	140	0.03243	0.004679	2.103	1.632	0.33
0.1	650	120	0.02972	0.004423	2.097	1.627	0.3
0.1	650	160	0.03606	0.006294	2.112	1.662	0.41
0.1	650	180	0.04054	0.005043	2.123	1.639	0.32
0.1	650	200	0.04547	0.005479	2.134	1.647	0.32
0.4	650	140	0.03094	0.005342	2.291	1.644	0.37
0.4	650	160	0.03557	0.005666	2.33	1.65	0.37
0.4	650	120	0.02882	0.005507	2.274	1.647	0.34
0.4	650	180	0.03841	0.005654	2.352	1.65	0.39
0.4	650	200	0.04315	0.005914	2.391	1.655	0.39
0.8	650	120	0.02738	0.004926	2.44	1.636	0.38
0.8	650	140	0.03171	0.005033	2.502	1.638	0.38
0.8	650	160	0.03682	0.005294	2.574	1.643	0.38
0.8	650	180	0.03829	0.005065	2.591	1.639	0.415
0.8	650	200	0.0419	0.005046	2.64	1.639	0.425
0.8	1000	150	0.03433	0.005176	2.548	1.641	0.38
0.6	1000	150	0.03371	0.005401	2.445	1.645	0.38
0.4	1000	150	0.03273	0.005514	2.324	1.647	0.38
0.2	1000	150	0.03346	0.006019	2.195	1.657	0.37
0.1	1000	150	0.03295	0.006211	2.113	1.66	0.41
0.08	1000	150	0.03292	0.006113	2.097	1.659	0.42
0.06	1000	150	0.03353	0.006124	2.08	1.659	0.43
0.04	1000	150	0.03517	0.006177	2.064	1.66	0.44

Table D.2 – **Network parameters for spiking simulations.** For all networks $w_+ = 4.0$. Recurrent conductance parameters are given for combinations of short-term plasticity parameters according to the following notation. g_{EE}: excitatory conductance g_E on excitatory neurons; g_{IE}: excitatory conductance g_E on inhibitory neurons; g_{EI}: inhibitory conductance g_I on excitatory neurons; g_{II}: inhibitory conductance g_I on inhibitory neurons.

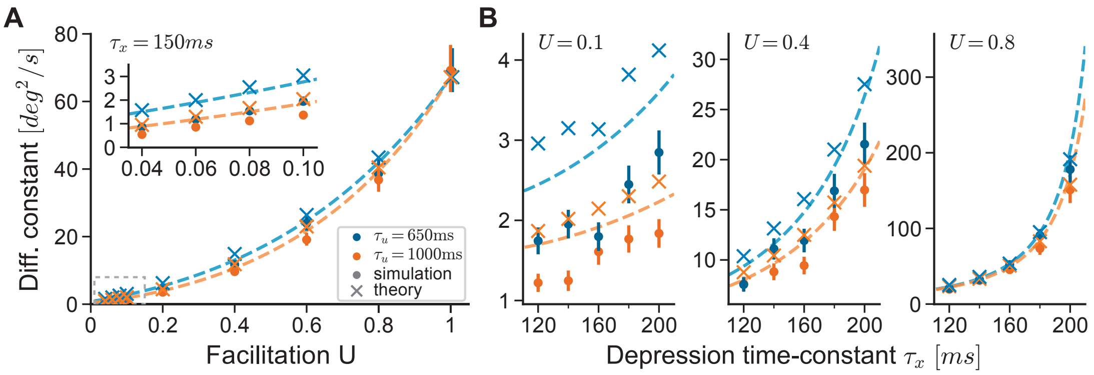


Figure D.4 – **Diffusion estimated by theory only.** Compare with Fig. 4.5 of the main text. All theoretical predictions (crosses, dashed lines) are computed by predictions of the firing rate profile using the mean-field theory presented in Section 4.6.6.

Bibliography

Abbott, L. F. (1997). Synaptic Depression and Cortical Gain Control. *Science*, 275(5297):221–224.

Abbott, L. F. and Nelson, S. B. (2000). Synaptic plasticity: Taming the beast. *Nature Neuroscience*, 3 Suppl(november):1178–83.

Abbott, L. F. and Regehr, W. G. (2004). Synaptic computation. *Nature*, 431(7010):796–803.

Abdi, H. and Williams, L. (2010). Jackknife. *Encyclopedia of research design*, pages 1–10.

Akil, H., Martone, M. E., and Van Essen, D. C. (2011). Challenges and Opportunities in Mining Neuroscience Data. *Science*, 331(6018):708–712.

Almeida, R., Barbosa, J., and Compte, A. (2015). Neural circuit basis of visuo-spatial working memory precision: A computational and behavioral study. *Journal of Neurophysiology*, 114(3):1806–1818.

Amari, S.-i. (1977). Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields. *Biological cybernetics*, 87:77–87.

Amit, D. J. (1998). Simulation in neurobiology: Theory or experiment? *Trends in Neurosciences*, 21(6):231–237.

Amit, D. J. and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral cortex*, 7:237–252.

Amit, D. J. and Mongillo, G. (2003). Spike-driven synaptic dynamics generating working memory states. *Neural computation*, 15(3):565–596.

Anwar, H., Li, X., Bucher, D., and Nadim, F. (2017). Functional roles of short-term synaptic plasticity with an emphasis on inhibition. *Current Opinion in Neurobiology*, 43:71–78.

Arnsten, A. F. T., Wang, M. J., and Paspalas, C. D. (2012). Neuromodulation of Thought: Flexibilities and Vulnerabilities in Prefrontal Cortical Network Synapses. *Neuron*, 76(1):223–239.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., and Butler, H. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, 20:25–29.

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, 63(1):1–29.

Badel, L., Lefort, S., Berger, T. K., Petersen, C. C. H., Gerstner, W., and Richardson, M. J. E. (2008). Extracting non-linear integrate-and-fire models from experimental data using dynamic I–V curves. *Biological Cybernetics*, 99(4-5):361.

Balaguer-Ballester, E., Lapish, C. C., Seamans, J. K., and Durstewitz, D. (2011). Attracting Dynamics of Frontal Cortex Ensembles during Memory-Guided Decision-Making. *PLOS Computational Biology*, 7(5):e1002057.

Barak, O. and Tsodyks, M. (2007). Persistent activity in neural networks with dynamic synapses. *PLOS Computational Biology*, 3(2):e35–e35.

Barak, O. and Tsodyks, M. (2014). Working models of working memory. *Current Opinion in Neurobiology*, 25:20–24.

Barbieri, F. and Brunel, N. (2007). Irregular Persistent Activity Induced by Synaptic Excitatory Feedback. *Frontiers in Computational Neuroscience*, 1.

Barbieri, F. and Brunel, N. (2008). Can attractor network models account for the statistics of firing during persistent activity in prefrontal cortex? *Frontiers in neuroscience*, 2(1):114–122.

Barrière, G., Tartas, M., Cazalets, J.-R., and Bertrand, S. S. (2008). Interplay between neuromodulator-induced switching of short-term plasticity at sensorimotor synapses in the neonatal rat spinal cord. *The Journal of Physiology*, 586(Pt 7):1903–1920.

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *International AAAI Conference on Weblogs and Social Media*.

Behnel, S., Bradshaw, R., Citro, C., Dalcin, L., Seljebotn, D. S., and Smith, K. (2011). Cython: The Best of Both Worlds. *Computing in Science Engineering*, 13(2):31–39.

Ben-Yishai, R., Bar-Or, R. L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, 92(9):3844–8.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35:99–109.

Bibitchkov, D., Herrmann, J. M., and Geisel, T. (2002). Pattern storage and processing in attractor networks with short-time synaptic dynamics. *Network*, 13(1):115–129.

Biel, M., Wahl-Schott, C., Michalakis, S., and Zong, X. (2009). Hyperpolarization-Activated Cation Channels: From Genes to Function. *Physiological Reviews*, 89(3):847–885.

Bos, H., Morrison, A., Peyser, A., Hahne, J., Helias, M., Kunkel, S., Ippen, T., Eppler, J. M., Schmidt, M., Seeholzer, A., Djurfeldt, M., Diaz, S., Morén, J., Deepu, R., Stocco, T., Deger, M., Michler, F., and Plesser, H. E. (2015). NEST 2.10.0. https://doi.org/10.5281/zenodo.44222.

Bostock, M., Ogievetsky, V., and Heer, J. (2011). D3 Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309.

Braitenberg, V. and Schütz, A. (1991). *Cortex: Statistics and Geometry of Neuronal Connectivity*. Springer.

Brenowitz, S., David, J., and Trussell, L. (1998). Enhancement of Synaptic Efficacy by Presynaptic GABAB Receptors. *Neuron*, 20(1):135–141.

Bressloff, P. C. and Cowan, J. D. (2003). A spherical model for orientation and spatial-frequency tuning in a cortical hypercolumn. *Phil. Trans. R. Soc. Lond. B*, 358(1438):1643–67.

Brette, R. and Gerstner, W. (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology*, 94(5):3637–42.

Brette, R., Rudolph, M., Carnevale, T., Hines, M., Beeman, D., Bower, J. M., Diesmann, M., Morrison, A., Goodman, P. H., Harris, F. C., Zirpe, M., Natschläger, T., Pecevski, D., Ermentrout, B., Djurfeldt, M., Lansner, A., Rochel, O., Vieville, T., Muller, E., Davison, A. P., El Boustani, S., and Destexhe, A. (2007). Simulation of networks of spiking neurons: A review of tools and strategies. *Journal of Computational Neuroscience*, 23(3):349–398.

Brody, C., Romo, R., and Kepecs, A. (2003a). Basic mechanisms for graded persistent activity: Discrete attractors, continuous attractors, and dynamic representations. *Current Opinion in Neurobiology*, pages 204–211.

Brody, C. D., Hernández, A., Zainos, A., and Romo, R. (2003b). Timing and Neural Encoding of Somatosensory Parametric Working Memory in Macaque Prefrontal Cortex. *Cerebral Cortex*, 13(11):1196–1207.

Brunel, N. (2000a). Dynamics of networks of randomly connected excitatory and inhibitory spiking neurons. *Journal of Physiology (Paris)*, 94(5–6):445–463.

Brunel, N. (2000b). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience*, 8(3):183–208.

Brunel, N. (2003). Dynamics and Plasticity of Stimulus-selective Persistent Activity in Cortical Network Models. *Cerebral Cortex*, 13:1151–1161.

Brunel, N. and Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural computation*, 11(7):1621–71.

Brunel, N. and Sergi, S. (1998). Firing frequency of leaky intergrate-and-fire neurons with synaptic current dynamics. *Journal of Theoretical Biology*, 195(1):87–95.

Brunel, N. and Wang, X. (2001a). Corrections for Effects of Neuromodulation in a Cortical Network Model of Object Working Memory Dominated by Recurrent Inhibition -. https://galton.uchicago.edu/~nbrunel/pdfs/corrBW01.pdf.

Brunel, N. and Wang, X. (2001b). Effects of Neuromodulation in a Cortical Network Model of Object Working Memory Dominated by Recurrent Inhibition. *Journal of Computational Neuroscience*, 11:63–85.

Brunton, B. W., Botvinick, M. M., and Brody, C. D. (2013). Rats and Humans Can Optimally Accumulate Evidence for Decision-Making. *Science*, 340(6128):95–98.

Burak, Y. (2014). Spatial coding and attractor dynamics of grid cells in the entorhinal cortex. *Current Opinion in Neurobiology*, 25:169–175.

Burak, Y. and Fiete, I. R. (2009). Accurate Path Integration in Continuous Attractor Network Models of Grid Cells. *PLOS Computational Biology*, 5(2):e1000291.

Burak, Y. and Fiete, I. R. (2012). Fundamental limits on persistent activity in networks of noisy neurons. *Proceedings of the National Academy of Sciences*, 109(43):17645–17650.

Burkitt, A. N. (2006a). A Review of the Integrate-and-fire Neuron Model: I. Homogeneous Synaptic Input. *Biological Cybernetics*, 95(1):1–19.

Burkitt, a. N. (2006b). A review of the integrate-and-fire neuron model: II. Inhomogeneous synaptic input and network properties. *Biological cybernetics*, 95(2):97–112.

Byrd, R., Lu, P., Nocedal, J., and Zhu, C. (1995). A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.

Cai, X., Liang, C. W., Muralidharan, S., Kao, J., and Tang, C. M. (2004). Unique roles of SK and Kv4. 2 potassium channels in dendritic integration. *Neuron*, 44(2):351–364.

Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1):1–27.

Camperi, M. and Wang, X. J. (1998). A model of visuospatial working memory in prefrontal cortex: Recurrent network and cellular bistability. *Journal of Computational Neuroscience*, 5(4):383–405.

Campiglio, M. and Flucher, B. E. (2015). The Role of Auxiliary Subunits for the Functional Diversity of Voltage-Gated Calcium Channels. *Journal of Cellular Physiology*, 230(9):2019–2031.

Cannon, R. C., Gleeson, P., Crook, S., Ganapathy, G., Marin, B., Piasini, E., and Silver, R. A. (2014). LEMS: A language for expressing complex biological models in concise and hierarchical form and its use in underpinning NeuroML 2. *Frontiers in Neuroinformatics*, 8.

Cano-Colino, M., Almeida, R., and Compte, A. (2013). Serotonergic modulation of spatial working memory: Predictions from a computational network model. *Frontiers in Integrative Neuroscience*, 7.

Cano-Colino, M., Almeida, R., Gomez-Cabrero, D., Artigas, F., and Compte, A. (2014). Serotonin Regulates Performance Nonmonotonically in a Spatial Working Memory Network. *Cerebral Cortex*, 24(9):2449–2463.

Cano-Colino, M. and Compte, A. (2012). A computational model for spatial working memory deficits in schizophrenia. *Pharmacopsychiatry*, 45(S 01):S49–S56.

Carnevale, N. and Hines, M. (2014). NEURON for empirically-based simulations of neurons and networks of neurons. [Online; accessed 26-May-2015: http://www.neuron.yale.edu/neuron].

Carnevale, N. T. and Hines, M. L. (2006). *The NEURON Book*. Cambridge University Press.

Carrillo-Reid, L., Yang, W., Bando, Y., Peterka, D. S., and Yuste, R. (2016). Imprinting and recalling cortical ensembles. *Science*, 353(6300):691–694.

Carter, E. and Wang, X.-J. (2007). Cannabinoid-Mediated Disinhibition and Working Memory: Dynamical Interplay of Multiple Feedback Mechanisms in a Continuous Attractor Model of Prefrontal Cortex. *Cerebral Cortex*, 17:i16–i26.

Catterall, W. (1995). Structure and function of voltage-gated ion channels. *Annual Review of Biochemistry*, 64:493–531.

Catterall, W. A., Goldin, A. L., and Waxman, S. G. (2005a). International Union of Pharmacology. XLVII. Nomenclature and Structure-Function Relationships of Voltage-Gated Sodium Channels. *Pharmacological Reviews*, 57(4):397–409.

Catterall, W. A., Perez-Reyes, E., Snutch, T. P., and Striessnig, J. (2005b). International Union of Pharmacology. XLVIII. Nomenclature and Structure-Function Relationships of Voltage-Gated Calcium Channels. *Pharmacological Reviews*, 57(4):411–425.

Celery (2017). Celery: Distributed Task Queue. http://www.celeryproject.org/.

Chafee, M. V. and Goldman-Rakic, P. S. (1998). Matching Patterns of Activity in Primate Prefrontal Area 8a and Parietal Area 7ip Neurons During a Spatial Working MemoryTask. *Journal of Neurophysiology*, 79(6):2919–2940.

Chaudhuri, R. and Fiete, I. (2016). Computational principles of memory. *Nature Neuroscience*, 19(3):394–403.

Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B., and Christen, R. (2006). TreeDyn: Towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics*, 7(1):439.

Collins, C. E., Turner, E. C., Sawyer, E. K., Reed, J. L., Young, N. A., Flaherty, D. K., and Kaas, J. H. (2016). Cortical cell and neuron density estimates in one chimpanzee hemisphere. *Proceedings of the National Academy of Sciences*, 113(3):740–745.

Colombo, F., Muscinelli, S. P., Seeholzer, A., Brea, J., and Gerstner, W. (2016). Algorithmic Composition of Melodies with Deep Recurrent Neural Networks. In *Proceedings of the 1st Conference on Computer Simulation of Musical Creativity, Huddersfield University*, Huddersfield, UK.

Colombo, F., Seeholzer, A., and Gerstner, W. (2017). Deep Artificial Composer: A Creative Neural Network Model for Automated Melody Generation. In *Computational Intelligence in Music, Sound, Art and Design*, pages 81–96. Springer, Cham.

Compte, A. (2006). Computational and in vitro studies of persistent activity: Edging towards cellular and synaptic mechanisms of working memory. *Neuroscience*, 139(1):135–151.

Compte, A., Brunel, N., Goldman-Rakic, P., and Wang, X. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, 10:910–923.

Compte, A., Constantinidis, C., Tegnér, J., Raghavachari, S., Chafee, M. V., Goldman-Rakic, P. S., and Wang, X.-J. (2003). Temporally Irregular Mnemonic Persistent Activity in Prefrontal Neurons of Monkeys During a Delayed Response Task. *Journal of Neurophysiology*, 90(5):3441–3454.

Constantinidis, C., Franowicz, M. N., and Goldman-Rakic, P. S. (2001). Coding Specificity in Cortical Microcircuits: A Multiple-Electrode Analysis of Primate Prefrontal Cortex. *Journal of Neuroscience*, 21(10):3646–3655.

Constantinidis, C. and Wang, X.-J. (2004). A neural circuit basis for spatial working memory. *The Neuroscientist*, 10(6):553–65.

Cover, T. M. and Thomas, J. A. (2012). *Elements of Information Theory*. John Wiley & Sons.

Curti, E. and Amit, D. J. (2004). Mean Field and Capacity in Realistic Networks of Spiking. *Neural Computation*, 2637:2597–2637.

Curtis, C. E. and D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, 7(9):415–423.

Curtis, C. E. and Lee, D. (2010). Beyond working memory: The role of persistent activity in decision making. *Trends in Cognitive Sciences*, 14(5):216–22.

Davies, D. L. and Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1(2):224–227.

Dayan, P. (2006). Levels of Analysis in Neural Modeling. In *Encyclopedia of Cognitive Science*. John Wiley & Sons, Ltd.

Dayan, P. and Abbott, L. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* The MIT Press.

Debanne, D., Daoudal, G., and Sourdet, V. (2003). Brain plasticity and ion channels. *Journal of Physiology (Paris)*, 27(4-6):403–414.

Dégenètais, E., Thierry, A.-M., Glowinski, J., and Gioanni, Y. (2002). Electrophysiological Properties of Pyramidal Neurons in the Rat Prefrontal Cortex: An In Vivo Intracellular Recording Study. *Cerebral Cortex*, 12(1):1–16.

Deger, M., Seeholzer, A., and Gerstner, W. (2016). Multi-contact synapses for stable networks: A spike-timing dependent model of dendritic spine plasticity and turnover. *arXiv:1609.05730 [q-bio]*.

Destexhe, A. and Huguenard, J. R. (2000). Which formalism to use for modeling voltage-dependent conductances? In *Computational Neuroscience: Realistic Modeling for Experimentalists*, pages 129–157. CRC Press Boca Raton, FL.

Destexhe, a., Mainen, Z., and Sejnowski, T. (1994). Synthesis of models for excitable membranes, synaptic transmission and neuromodulation using a common kinetic formalism. *Journal of Computational Neuroscience*, 1(3):195–230.

Destexhe, A., Mainen, Z. F., and Sejnowski, T. J. (1998). Kinetic models of synaptic transmission. *Methods in Neuronal Modeling*, 2:1–25.

Destexhe, A. and Paré, D. (1999). Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. *Journal of Neurophysiology*, 81(4):1531–1547.

Diesmann, M. and Gewaltig, M.-O. (2007). NEST (NEural Simulation Tool). *Scholarpedia*, 2(4):1430–1430.

Dittman, J. S., Kreitzer, A. C., and Regehr, W. G. (2000). Interplay between Facilitation, Depression, and Residual Calcium at Three Presynaptic Terminals. *Journal of Neuroscience*, 20(4):1374–1385.

Django (2017). The Web framework for perfectionists with deadlines | Django. https://www.djangoproject.com/.

Django-Notifications (2017). Django-notifications/django-notifications. https://github.com/django-notifications/django-notifications.

Django-REST (2017). Django REST framework. http://www.django-rest-framework.org/.

Druckmann, S., Berger, T. K., Schürmann, F., Hill, S., Markram, H., and Segev, I. (2011). Effective stimuli for constructing reliable neuron models. *PLoS Computational Biology*, 7(8).

Druckmann, S. and Chklovskii, D. B. (2012). Neuronal circuits underlying persistent representations despite time varying activity. *Current biology*, 22(22):2095–103.

Duarte, R., Seeholzer, A., Zilles, K., and Morrison, A. (2017). Synaptic patterning and the timescales of cortical dynamics. *Current Opinion in Neurobiology*, 43:156–165.

Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3(3):32–57.

Durstewitz, D., Seamans, J. K., and Sejnowski, T. J. (2000a). Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *Journal of Neurophysiology*, 83:1733–1750.

Durstewitz, D., Seamans, J. K., and Sejnowski, T. J. (2000b). Neurocomputational models of working memory. *Nature Neuroscience*, 3:1184–91.

Eby, P. J. (2003). PEP 3333 – Python Web Server Gateway Interface v1.0.1. https://www.python.org/dev/peps/pep-3333/.

Efron, B. and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. CRC Press.

Evans, C. (2017). Scikits-Bootstrap. https://github.com/cgevans/scikits-bootstrap.

Ferguson, A. R., Nielson, J. L., Cragin, M. H., Bandrowski, A. E., and Martone, M. E. (2014). Big data from small data: Data-sharing in the ’long tail’ of neuroscience. *Nature Neuroscience*, 17(11):1442–1447.

Fineberg, J. D., Ritter, D. M., and Covarrubias, M. (2012). Modeling-independent elucidation of inactivation pathways in recombinant and native A-type Kv channels. *The Journal of General Physiology*, 140(5):513–527.

Forster, O. (2016). *Analysis 1: Differential- und Integralrechnung einer Veränderlichen*. Springer-Verlag.

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

Fourcaud-Trocmé, N., Hansel, D., van Vreeswijk, C., and Brunel, N. (2003). How Spike Generation Mechanisms Determine the Neuronal Response to Fluctuating Inputs. *The Journal of Neuroscience*, 23(37):11628–11640.

Funahashi, S. (2006). Prefrontal cortex and working memory processes. *Neuroscience*, 139(1):251–61.

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey’s dorsolateral prefrontal cortex. *Journal of Neurophysiology*, pages 331–349.

Fung, C., Wong, K., and Wu, S. (2010). A moving bump in a continuous manifold: A comprehensive study of the tracking dynamics of continuous attractor neural networks. *Neural Computation*, 22(3):752–792.

Fusi, S., Miller, E. K., and Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. *Current Opinion in Neurobiology*, 37:66–74.

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Booth, M., and Rossi, F. (2009). *GNU Scientific Library Reference Manual*. 3rd edition. ISBN 0954612078.

Gardiner, C. (2009). *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Springer, 4th edition.

Gasque, G., Labarca, P., Reynaud, E., and Darszon, A. (2005). Shal and shaker differential contribution to the K+ currents in the Drosophila mushroom body neurons. *The Journal of Neuroscience*, 25(9):2348–2358.

Georgopoulos, a. P., Lurito, J. T., Petrides, M., Schwartz, a. B., and Massey, J. T. (1989). Mental rotation of the neuronal population vector. *Science*, 243(4888):234–6.

Gerstner, W. and Kistler, W. M. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press.

Gerstner, W., Kistler, W. M., Naud, R., and Paninski, L. (2014). *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. Cambridge University Press.

Gleeson, D. P., Silver, P. A., and Cantarelli, M. (2013). Open Source Brain. In Jaeger, D. and Jung, R., editors, *Encyclopedia of Computational Neuroscience*, pages 1–3. Springer New York.

Goldberg, E. M., Clark, B. D., Zagha, E., Nahmani, M., and Erisir, A. (2008). K+ channels at the axon initial segment dampen near-threshold excitability of neocortical fast-spiking GABAergic interneurons. *Neuron*, 58(3):387–400.

Goldman, M. S., Golowasch, J., Marder, E., and Abbott, L. F. (2001). Global structure, robustness, and modulation of neuronal models. *The Journal of Neuroscience*, 21(14):5229–38.

Goldman-Rakic, P. S. (1991). Cellular and circuit basis of working memory in prefrontal cortex of nonhuman primates. *Progress in Brain Research*, 85:325–336.

Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron*, 14(3):477–485.

Golowasch, J., Goldman, M. S., Abbott, L. F., and Marder, E. (2002). Failure of averaging in the construction of a conductance-based neuron model. *Journal of Neurophysiology*, 87:1129–1131.

Golub, G. H. and Loan, C. F. V. (1996). *Matrix Computations*. JHU Press.

Grafana (2017). Beautiful Metrics Dashboards, Data Visualization and Monitoring. http://grafana.org/.

Grillner, S. (2014). Megascience Efforts and the Brain. *Neuron*, 82(6):1209–1211.

Guckenheimer, J. and Holmes, P. J. (2013). *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, volume 42. Springer Science & Business Media.

Guo, Z. V., Inagaki, H. K., Daie, K., Druckmann, S., Gerfen, C. R., and Svoboda, K. (2017). Maintenance of persistent activity in a frontal thalamocortical loop. *Nature*, advance online publication.

Gupta, a. (2000). Organizing Principles for a Diversity of GABAergic Interneurons and Synapses in the Neocortex. *Science*, 287(5451):273–278.

Gutman, G. A., Chandy, K. G., Grissmer, S., Lazdunski, M., Mckinnon, D., Pardo, L. A., Robertson, G. A., Rudy, B., Sanguinetti, M. C., Stühmer, W., and Wang, X. (2005). International Union of Pharmacology. LIII. Nomenclature and Molecular Relationships of Voltage-Gated Potassium Channels. *Pharmacological Reviews*, 57(4):473–508.

Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2001). On Clustering Validation Techniques. *Journal of Intelligent Information Systems*, 17(2-3):107–145.

Hansel, D. and Mato, G. (2001). Existence and Stability of Persistent States in Large Neuronal Networks. *Physical Review Letters*, 86(18):4175–4178.

Hansel, D. and Mato, G. (2013). Short-term plasticity explains irregular persistent activity in working memory tasks. *The Journal of Neuroscience*, 33(1):133–49.

Hansel, D. and Sompolinsky, H. (1998). Modeling Feature Selectivity in Local Cortical Circuits. In Koch, C. and Segev, I., editors, *Methods in Neural Modeling. from Synapses to Networks*, pages 499–567. MIT Press.

Hay, E., Hill, S., Schürmann, F., Markram, H., and Segev, I. (2011). Models of neocortical layer 5b pyramidal cells capturing a wide range of dendritic and perisomatic active properties. *PLoS Computational Biology*, 7(7).

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Approach.* John Wiley & Sons, New York, NY, USA.

Hempel, C. M., Hartman, K. H., Wang, X.-J., Turrigiano, G. G., and Nelson, S. B. (2000). Multiple Forms of Short-Term Plasticity at Excitatory Synapses in Rat Medial Prefrontal Cortex. *Journal of Neurophysiology*, 83(5):3031–3041.

Hille, B. (2001). *Ion Channels of Excitable Membranes*, volume 507. Sinauer Sunderland, MA.

Hines, M. and Carnevale, N. (2004). Discrete event simulation in the NEURON environment. *Neurocomputing*, pages 1–6.

Hines, M. and Carnevale, N. T. (2001). NEURON: A tool for neuroscientists. *The Neuroscientist*, 7(2):123–135.

Hines, M. L. and Carnevale, N. T. (2000). Expanding NEURON's repertoire of mechanisms with NMODL. *Neural Computation*, 12(5):995–1007.

Hines, M. L., Morse, T., Migliore, M., Carnevale, N. T., and Shepherd, G. M. (2004). ModelDB: A Database to Support Computational Neuroscience. *Journal of Computational Neuroscience*, 17(1):7–11.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Hodgkin, A. L. and Huxley, A. F. (1952a). The dual effect of membrane potential on sodium conductance in the giant axon of Loligo. *The Journal of Physiology*, 116:497–506.

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

Hopfield, J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 81(10):3088–3088.

Hsu, C. L., Cembrowski, M. S., Mensh, B. D., and Spruston, N. (2015). Dendritic sodium spikes are required for long-term potentiation at distal synapses on hippocampal pyramidal neurons. *eLife*, 10.7554:eLife.06414.

ICGenealogy (2016a). ICGenealogy Gihub Repository. [Online; accessed 1-Oct-2016: https://github.com/icgenealogy].

ICGenealogy (2016b). ICGenealogy Website. [Online; accessed 1-Oct-2016: http://icg.neurotheory.ox.ac.uk].

Itskov, V., Curto, C., Pastalkova, E., and Buzsáki, G. (2011a). Cell assembly sequences arising from spike threshold adaptation keep track of time in the hippocampus. *The Journal of Neuroscience*, 31(8):2828–34.

Itskov, V., Hansel, D., and Tsodyks, M. (2011b). Short-Term Facilitation may Stabilize Parametric Working Memory Trace. *Frontiers in Computational Neuroscience*, 5(October):40–40.

Jahr, C. E. and Stevens, C. F. (1990a). A quantitative description of NMDA receptor-channel kinetic behavior. *Journal of Neuroscience*, 10(6):1830–1837.

Jahr, C. E. and Stevens, C. F. (1990b). Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics. *Journal of Neuroscience*, 10(9):3178–3182.

Jolivet, R., Schürmann, F., Berger, T. K., Naud, R., Gerstner, W., and Roth, A. (2008). The quantitative single-neuron modeling competition. *Biological Cybernetics*, 99(4-5):417.

Jones, E., Oliphant, T., and Peterson (2017). SciPy.org — SciPy.org. http://scipy.org/.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (2000). *Principles of Neural Science*. McGraw-Hill Medical, 4th edition.

Kase, D. and Imoto, K. (2012). The Role of HCN Channels on Membrane Excitability in the Nervous System. *Journal of Signal Transduction*, 2012:619747.

Kilpatrick, Z. P. and Ermentrout, B. (2012). Wandering bumps in stochastic neural fields. *arXiv:1205.3072 [math, nlin, q-bio]*.

Kilpatrick, Z. P., Ermentrout, B., and Doiron, B. (2013). Optimizing Working Memory with Heterogeneity of Recurrent Cortical Excitation. *Journal of Neuroscience*, 33(48):18999–19011.

Knierim, J. J. and Zhang, K. (2012). Attractor dynamics of spatially correlated neural activity in the limbic system. *Annual Review of Neuroscience*, 35:267–85.

Koch, C. (2004). *Biophysics of Computation: Information Processing in Single Neurons*. Oxford University Press, Oxford, UK.

Koch, C. and Buice, M. A. (2015). A Biological Imitation Game. *Cell*, 163(2):277–280.

Koch, C. and Segev, I. (2000). The role of single neurons in information processing. *Nature Neuroscience*, 3:1171–1177.

Kraft, D. (1988). A software package for sequential quadratic programming. *Forschungsbericht - Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt*.

Kreitzer, A. C. and Regehr, W. G. (2000). Modulation of transmission during trains at a cerebellar synapse. *Journal of Neuroscience*, 20(4):1348–1357.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, 19(9):1639–1645.

Kubo, Y., Adelman, J. P., Clapham, D. E., Jan, L. Y., Karschin, A., Kurachi, Y., Lazdunski, M., Nichols, C. G., Seino, S., and Vandenberg, C. A. (2005). International Union of Pharmacology. LIV. Nomenclature and Molecular Relationships of Inwardly Rectifying Potassium Channels. *Pharmacological Reviews*, 57(4):509–526.

La Camera, G., Giugliano, M., Senn, W., and Fusi, S. (2008). The response of cortical neurons to in vivo-like input current: Theory and experiment: I. Noisy inputs with stationary statistics. *Biological Cybernetics*, 99(4-5):279–301.

Laing, C. R. (2014). Derivation of a neural field model from a network of theta neurons. *Physical Review E*, 90(1):010901.

Laing, C. R. and Longtin, A. (2001). Noise-induced stabilization of bumps in systems with long-range spatial coupling. *Physica D: Nonlinear Phenomena*, 160(3):149–172.

Laing, C. R., Troy, W. C., Gutkin, B., and Ermentrout, G. B. (2002). Multiple Bumps in a Neuronal Model of Working Memory. *SIAM Journal on Applied Mathematics*, 63(1):62–97.

Latham, P. E. and Roudi, Y. (2009). Mutual information. *Scholarpedia*, 4(1):1658. revision 122173.

Leary, T., Williams, A. H., Franci, A., and Marder, E. (2014). Cell Types, Network Homeostasis, and Pathological Compensation from a Biologically Plausible Ion Channel Expression Model. *Neuron*, 82(4):809–821.

Levitan, I. B. (1994). Modulation of ion channels by protein phosphorylation and dephosphorylation. *Annual Review of Physiology*, 56:193–212.

Lewicki, M. S. (1998). A review of methods for spike sorting: The detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9(4):R53–R78.

Li, Q., Lee, J.-A., and Black, D. L. (2007). Neuronal regulation of alternative pre-mRNA splicing. *Nature Reviews Neuroscience*, 8(11):819–831.

Lim, S. and Goldman, M. S. (2013). Balanced cortical microcircuitry for maintaining information in working memory. *Nature Neuroscience*, 16(9).

Lim, S. and Goldman, M. S. (2014). Balanced Cortical Microcircuitry for Spatial Working Memory Based on Corrective Feedback Control. *Journal of Neuroscience*, 34(20):6790–6806.

Lindner, B., Gangloff, D., Longtin, A., and Lewis, J. E. (2009). Broadband Coding with Dynamic Synapses. *Journal of Neuroscience*, 29(7):2076–2087.

Lundqvist, M., Herman, P., and Lansner, A. (2011). Theta and gamma power increases and alpha/beta power decreases with memory load in an attractor network model. *Journal of Cognitive Neuroscience*, 23(10):3008–3020.

Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., and Miller, E. K. (2016). Gamma and Beta Bursts Underlie Working Memory. *Neuron*, 90(1):152–164.

Maass, W., Natschläger, T., and Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–2560.

Machens, C. K., Romo, R., and Brody, C. D. (2010). Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex. *Journal of Neuroscience*, 30(1):350–60.

Macoveanu, J., Klingberg, T., and Tegnér, J. (2006). A biophysical model of multiple-item working memory: A computational and neuroimaging study. *Neuroscience*, 141(3):1611–8.

Macoveanu, J., Klingberg, T., and Tegnér, J. (2007). Neuronal firing rates account for distractor effects on mnemonic accuracy in a visuo-spatial working memory task. *Biological Cybernetics*, 96(4):407–419.

Mainen, Z. F. and Sejnowski, T. J. (1995). Reliability of Spike Timing in Neocortical Neurons. *Science*, 268(5216):1503–1506.

Major, G. and Tank, D. (2004). Persistent neural activity: Prevalence and mechanisms. *Current Opinion in Neurobiology*, 14(6):675–684.

Marblestone, A. H., Wayne, G., and Kording, K. P. (2016). Toward an Integration of Deep Learning and Neuroscience. *Frontiers in Computational Neuroscience*, 10.

Marder, E. and Goaillard, J.-M. (2006). Variability, compensation and homeostasis in neuron and network function. *Nature Reviews Neuroscience*, 7(7):563–574.

Marin, O. (2012). Interneuron dysfunction in psychiatric disorders. *Nature Reviews Neuroscience*, 13(2):107–120.

Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., Kahou, G. A. A., Berger, T. K., Bilgili, A., Buncic, N., Chalimourda, A., Chindemi, G., Courcol, J.-D., Delalondre, F., Delattre, V., Druckmann, S., Dumusc, R., Dynes, J., Eilemann, S., Gal, E., Gevaert, M. E., Ghobril, J.-P., Gidon, A., Graham, J. W., Gupta, A., Haenel, V., Hay, E., Heinis, T., Hernando, J. B., Hines, M., Kanari, L., Keller, D., Kenyon, J., Khazen, G., Kim, Y., King, J. G., Kisvarday, Z., Kumbhar, P., Lasserre, S., Le Bé, J.-V., Magalhães, B. R. C., Merchán-Pérez, A., Meystre, J., Morrice, B. R., Muller, J., Muñoz-Céspedes, A., Muralidhar, S., Muthurasa, K., Nachbaur, D., Newton, T. H., Nolte, M., Ovcharenko, A., Palacios, J., Pastor, L., Perin, R., Ranjan, R., Riachi, I., Rodríguez, J.-R., Riquelme, J. L., Rössert, C., Sfyrakis, K., Shi, Y., Shillcock, J. C., Silberberg, G., Silva, R., Tauheed, F., Telefont, M., Toledo-Rodriguez, M., Tränkler, T., Van Geit, W., Díaz, J. V., Walker, R., Wang, Y., Zaninetta, S. M., DeFelipe, J., Hill, S. L., Segev, I., and Schürmann, F. (2015). Reconstruction and Simulation of Neocortical Microcircuitry. *Cell*, 163(2):456–492.

Markram, H., Wang, Y., and Tsodyks, M. (1998). Differential signaling via the same axon of neocortical pyramidal neurons. *Proceedings of the National Academy of Sciences*, 95(9):5323–8.

Martin, S. J., Grimwood, P. D., and Morris, R. G. (2000). Synaptic plasticity and memory: An evaluation of the hypothesis. *Annual Review of Neuroscience*, 23:649–711.

Mejias, J. F. and Torres, J. J. (2009). Maximum memory capacity on neural networks with short-term synaptic depression and facilitation. *Neural Computation*, 21(3):851–71.

Memcached (2017). Memcached - a distributed memory object caching system. https://memcached.org/.

Mensi, S., Hagens, O., Gerstner, W., and Pozzorini, C. (2016). Enhanced Sensitivity to Rapid Input Fluctuations by Nonlinear Threshold Dynamics in Neocortical Pyramidal Neurons. *PLOS Computational Biology*, 12(2):e1004761.

Mensi, S., Naud, R., Pozzorini, C., Avermann, M., Petersen, C. C. H., and Gerstner, W. (2012). Parameter extraction and classification of three cortical neuron types reveals two distinct adaptation mechanisms. *Journal of Neurophysiology*, 107(6):1756–1775.

Merdanović, E. (2017). The complete list of all time series databases for your IoT project | Erol.si. http://www.erol.si/2015/01/the-complete-list-of-all-timeseries-databases-for-your-iot-project/.

Mi, Y., Lin, X., and Wu, S. (2016). Neural Computations in a Dynamical System with Multiple Time Scales. *Frontiers in Computational Neuroscience*, 10.

Migliore, M., Hoffman, D. A., and Magee, J. C. (1999). Role of an A-type K+ conductance in the back-propagation of action potentials in the dendrites of hippocampal pyramidal neurons. *Journal of Computational Neuroscience*, 7(1):5–15.

Miller, E. K., Erickson, C. A., and Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, 16(16):5154–5167.

Miller, K. D. and Fumarola, F. (2012). Mathematical equivalence of two common forms of firing rate models of neural networks. *Neural Computation*, 24(1):25–31.

Miller, P. (2013). Stabilization of Memory States by Stochastic Facilitating Synapses. *The Journal of Mathematical Neuroscience*, 3(1):19.

Milligan, G. W. and Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2):159–179.

Misonou, H., Mohapatra, D. P., Park, E. W., Leung, V., Zhen, D., Misonou, K., Anderson, A. E., and Trimmer, J. S. (2004). Regulation of ion channel localization and phosphorylation by neuronal activity. *Nature Neuroscience*, 7(7):711–718.

Mongillo, G., Amit, D. J., and Brunel, N. (2003). Retrospective and prospective persistent activity induced by Hebbian learning in a recurrent cortical network. *European Journal of Neuroscience*, 18(7):2011–2024.

Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, 319(5869):1543–1546.

Mongillo, G., Curti, E., Romani, S., and Amit, D. J. (2005). Learning in realistic networks of spiking neurons and spike-driven plastic synapses. *European Journal of Neuroscience*, 21(11):3143–3160.

Mongillo, G., Hansel, D., and van Vreeswijk, C. (2012). Bistability and Spatiotemporal Irregularity in Neuronal Networks with Nonlinear Synaptic Transmission. *Physical Review Letters*, 108(15):158101.

Moreno-Bote, R., Rinzel, J., and Rubin, N. (2007). Noise-Induced Alternations in an Attractor Network Model of Perceptual Bistability. *Journal of Neurophysiology*, 98(3):1125–1139.

Moser, E. I., Roudi, Y., Witter, M. P., Kentros, C., Bonhoeffer, T., and Moser, M.-B. (2014). Grid cells and cortical representation. *Nature Reviews Neuroscience*, 15(7):466–481.

Murray, J. D., Anticevic, A., Gancsos, M., Ichinose, M., Corlett, P. R., Krystal, J. H., and Wang, X.-J. (2012). Linking Microcircuit Dysfunction to Cognitive Impairment: Effects of Disinhibition Associated with Schizophrenia in a Cortical Working Memory Model. *Cerebral Cortex*.

Murray, J. D., Bernacchia, A., Roy, N. A., Constantinidis, C., Romo, R., and Wang, X.-J. (2016). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proceedings of the National Academy of Sciences*, 114(2):394–399.

Murthy, M. and Turner, G. (2013). Whole-cell in vivo patch-clamp recordings in the drosophila brain. *Cold Spring Harbor Protocols*, 8(2):140–148.

MySQL (2017). MySQL. https://www.mysql.com/.

Nadim, F. and Bucher, D. (2014). Neuromodulation of neurons and synapses. *Current Opinion in Neurobiology*, 29:48–56.

Nawrot, M., Aertsen, A., and Rotter, S. (1999). Single-trial estimation of neuronal firing rates: From single-neuron spike trains to population activity. *Journal of Neuroscience Methods*, 94:81–92.

Neher, E. (1992). Correction for liquid junction potentials in patch clamp experiments. *Methods in Enzymology*, 207(1991):123–131.

Neher, E. and Sakaba, T. (2008). Multiple roles of calcium ions in the regulation of neurotransmitter release. *Neuron*, 59(6):861–872.

NeuroML (2017). NeuroML Database. https://neuroml-db.org/.

Nichols, C. G. and Lopatin, A. N. (1997). Inward Rectifier Potassium Channels. *Annual Review of Physiology*, 59(1):171–191.

Oh, M., Zhao, S., Matveev, V., and Nadim, F. (2012). Neuromodulatory changes in short-term synaptic dynamics may be mediated by two distinct mechanisms of presynaptic calcium entry. *Journal of Computational Neuroscience*, 33(3).

Oliphant, T. E. (2007). Python for Scientific Computing. *Computing in Science & Engineering*, 9(3):10–20.

OpenTSDB (2017). OpenTSDB - A Distributed, Scalable Monitoring System. http://opentsdb.net/.

O'Reilly, R. and Frank, M. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, 18(2):283–328.

Pan, B. and Zucker, R. S. (2009). A General Model of Synaptic Transmission and Short-Term Plasticity. *Neuron*, 62(4):539–554.

Parekh, R. and Ascoli, G. A. (2013). Neuronal Morphology Goes Digital: A Research Hub for Cellular and System Neuroscience. *Neuron*, 77(6):1017–1038.

Pawson, A. J., Sharman, J. L., and Benson, H. E. (2014). The IUPHAR/BPS Guide to PHARMACOLOGY: An expert-driven knowledgebase of drug targets and their ligands. *Nucleic Acids Research*, 42:D1098–106.

Perea, G., Sur, M., and Araque, A. (2014). Neuron-glia networks: Integral gear of brain function. *Frontiers in Cellular Neuroscience*, 8.

Pereira, J. and Wang, X.-J. (2015). A Tradeoff Between Accuracy and Flexibility in a Working Memory Circuit Endowed with Slow Feedback Mechanisms. *Cerebral Cortex*, 25(10):3586.

Pfister, J.-P., Dayan, P., and Lengyel, M. (2010). Synapses with short-term plasticity are optimal estimators of presynaptic membrane potentials. *Nature Neuroscience*, 13(10):1271–5.

Ploner, C. J., Gaymard, B., Rivaud, S., Agid, Y., and Pierrot-Deseilligny, C. (1998). Temporal limits of spatial working memory in humans. *European Journal of Neuroscience*, 10(2):794–797.

Podlaski, W. F., Seeholzer, A., Groschner, L. N., Miesenböck, G., Ranjan, R., and Vogels, T. P. (2017). Mapping the function of neuronal ion channels in model and experiment. *eLife*, 6:e22152.

Poirazi, P., Brannon, T., and Mel, B. (2003). Pyramidal neuron as two-layer neural network. *Neuron*, 37(6):989–99.

Powell, M. J. (1970). A hybrid method for nonlinear equations. *Numerical Methods for Nonlinear Algebraic Equations*, 7:87–114.

Pozzorini, C., Mensi, S., Hagens, O., Naud, R., Koch, C., and Gerstner, W. (2015). Automated High-Throughput Characterization of Single Neurons by Means of Simplified Spiking Models. *PLOS Computational Biology*, 11(6):e1004275.

Prinz, A. A., Billimoria, C. P., and Marder, E. (2003). Alternative to Hand-Tuning Conductance-Based Models: Construction and Analysis of Databases of Model Neurons. *Journal of Neurophysiology*, 90(6):3998–4015.

Prinz, A. A., Bucher, D., and Marder, E. (2004). Similar network activity from disparate circuit parameters. *Nature Neuroscience*, 7(12):1345–52.

Python (2017). Python.org. https://www.python.org/.

RabbitMQ (2017). RabbitMQ - Messaging that just works. https://www.rabbitmq.com/.

Rainer, G., Asaad, W., and Miller, E. (1998). Memory fields of neurons in the primate prefrontal cortex. *Proceedings of the National Academy of Sciences*, 95(December):15008–15013.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer, New York, 2 edition.

Ranjan, R., Khazen, G., Gambazzi, L., Ramaswamy, S., Hill, S. L., Schürmann, F., and Markram, H. (2011). Channelpedia: An integrative and interactive database for ion channels. *Frontiers in Neuroinformatics*, 5:36.

Rao, S. C. (1997). Integration of What and Where in the Primate Prefrontal Cortex. *Science*, 276(5313):821–824.

Rauch, A., Camera, G. L., Lüscher, H.-R., Senn, W., and Fusi, S. (2003). Neocortical Pyramidal Cells Respond as Integrate-and-Fire Neurons to In Vivo–Like Input Currents. *Journal of Neurophysiology*, 90(3):1598–1612.

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. D. (2010). The asynchronous state in cortical circuits. *Science*, 327(5965):587–90.

Renart, A., Moreno-Bote, R., Wang, X.-J., and Parga, N. (2007). Mean-driven and fluctuation-driven persistent activity in recurrent networks. *Neural Computation*, 19(1):1–46.

Renart, A., Song, P., and Wang, X.-J. (2003). Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron*, 38(3):473–85.

Rigotti, M., Rubin, B. D., D, D., Wang, X.-J., and Fusi, S. (2010). Internal Representation of Task Rules by Recurrent Dynamics: The Importance of the Diversity of Neural Responses. *Frontiers in Computational Neuroscience*, 4.

Roach, J. P., Ben-Jacob, E., Sander, L. M., and Zochowski, M. R. (2015). Formation and Dynamics of Waves in a Cortical Model of Cholinergic Modulation. *PLoS Comput Biol*, 11(8):e1004449.

Robinson, R. B. and Siegelbaum, S. A. (2003). Hyperpolarization-Activated Cation Currents: From Molecules to Physiological Function. *Annual Review of Physiology*, 65(1):453–480.

Roggeman, C., Klingberg, T., Feenstra, H. E. M., Compte, A., and Almeida, R. (2014). Trade-off between Capacity and Precision in Visuospatial Working Memory. *Journal of Cognitive Neuroscience*, 26(2):211–222.

Rolls, E. T., Dempere-Marco, L., and Deco, G. (2013). Holding Multiple Items in Short Term Memory: A Neural Mechanism. *PLOS ONE*, 8(4):e61078.

Romani, S., Amit, D. J., and Mongillo, G. (2006). Mean-field analysis of selective persistent activity in presence of short-term synaptic depression. *Journal of Computational Neuroscience*, 20(2):201–17.

Romani, S. and Tsodyks, M. (2010). Continuous attractors with morphed/correlated maps. *PLOS Computational Biology*, 6(8).

Romani, S. and Tsodyks, M. (2015). Short-term plasticity based network model of place cells dynamics. *Hippocampus*, 25(1):94–105.

Romo, R., Brody, C. D., Hernández, a., and Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, 399(6735):470–3.

Rosenbaum, R. and Doiron, B. (2014). Balanced Networks of Spiking Neurons with Spatially Dependent Recurrent Connections. *Physical Review X*, 4(2).

Rossant, C., Goodman, D. F. M., Fontaine, B., Platkiewicz, J., Magnusson, A. K., and Brette, R. (2011). Fitting Neuron Models to Spike Trains. *Frontiers in Neuroscience*, 5.

Rossant, C., Goodman, D. F. M., Platkiewicz, J., and Brette, R. (2010). Automatic fitting of spiking neuron models to electrophysiological recordings. *Frontiers in Neuroinformatics*, 4.

Rössert, C., Pozzorini, C., Chindemi, G., Davison, A. P., Eroe, C., King, J., Newton, T. H., Nolte, M., Ramaswamy, S., Reimann, M. W., Gewaltig, M.-O., Gerstner, W., Markram, H., Segev, I., and Muller, E. (2016). Automated point-neuron simplification of data-driven microcircuit models. *arXiv:1604.00087 [q-bio]*.

Roudi, Y. and Treves, A. (2008). Representing where along with what information in a model of a cortical patch. *PLOS Computational Biology*, 4(3):e1000012–e1000012.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Sakurai, A. and Katz, P. S. (2009). State-, Timing-, and Pattern-Dependent Neuromodulation of Synaptic Strength by a Serotonergic Interneuron. *Journal of Neuroscience*, 29(1):268–279.

Samsonovich, A. and McNaughton, B. L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *Journal of Neuroscience*, 17(15):5900–20.

Sankey, H. (1898). Introductory note on the thermal efficiency of steam-engines. In *Minutes of Proceedings of The Institution of Civil Engineers*, pages 278–283.

Scalaidhe, S. P. Ó., Wilson, F. A. W., and Goldman-Rakic, P. S. (1997). Areal Segregation of Face-Processing Neurons in Prefrontal Cortex. *Science*, 278(5340):1135–1138.

Schmidt, M. (2008). The Sankey Diagram in Energy and Material Flow Management. *Journal of Industrial Ecology*, 12(1):82–94.

Schulz, D. J., Temporal, S., Barry, D. M., and Garcia, M. L. (2008). Mechanisms of voltage-gated ion channel regulation: From gene expression to localization. *Cellular and Molecular Life Sciences*, 65(14):2215–31.

Seelig, J. D. and Jayaraman, V. (2015). Neural dynamics for landmark orientation and angular path integration. *Nature*, 521(7551):186–191.

Sejnowski, T. J., Churchland, P. S., and Movshon, J. A. (2014). Putting big data to good use in neuroscience. *Nature Neuroscience*, 17(11):1440–1441.

Selemon, L. D. and Goldman-Rakic, P. S. (1988). Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: Evidence for a distributed neural network subserving spatially guided behavior. *Journal of Neuroscience*, 8(11):4049–4068.

Seung, H. S. (1998). Continuous attractors and oculomotor control. *Proceedings of the National Academy of Sciences*, 11(7-8):1253–1258.

Shah, M. M., Hammond, R. S., and Hoffman, D. A. (2010). Dendritic ion channel trafficking and plasticity. *Trends in Neurosciences*, 33(7):307–316.

Shepherd Lab, Y. U. (2015). ModelDB. [Online; accessed 01-July-2015: http://senselab.med.yale.edu/ModelDB/].

Shouval, H. Z., Wang, S. S.-H., and Wittenberg, G. M. (2010). Spike timing dependent plasticity: A consequence of more fundamental learning rules. *Frontiers in Computational Neuroscience*, 4(July):1–13.

Sjöström, P. J. and Nelson, S. B. (2002). Spike timing, calcium signals and synaptic plasticity. *Current Opinion in Neurobiology*, 12(3):305–314.

Softky, W. R. and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, 13(1):334–350.

Spiridon, M. and Gerstner, W. (2001). Effect of lateral connections on the accuracy of the population code for a network of spiking neurons. *Network*, 12(4):409–21.

Sreenivasan, K. K., Curtis, C. E., and D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, 18(2):82–89.

Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework. *Trends in Cognitive Sciences*, 19(7):394–405.

Strata, P. and Harvey, R. (1999). Dale's principle. *Brain Research Bulletin*, 50(5–6):349–350.

Stringer, S. M., Trappenberg, T. P., Rolls, E. T., and de Araujo, I. E. T. (2002). Self-organizing continuous attractor networks and path integration: One-dimensional models of head direction cells. *Network*, 13(2):217–42.

Strogatz, S. (2000). *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering.* Westview Press, 1st edition.

Stuart, G. and Spruston, N. (1998). Determinants of voltage attenuation in neocortical pyramidal neuron dendrites. *The Journal of Neuroscience*, 18(10):3501–10.

Supervisor (2017). Supervisor: A Process Control System — Supervisor 3.3.1 documentation. http://supervisord.org/.

Sussillo, D., Toyoizumi, T., and Maass, W. (2007). Self-Tuning of Neural Circuits Through Short-Term Synaptic Plasticity. *Journal of Neurophysiology*, 97(6):4079–4095.

Takeda, K. and Funahashi, S. (2002). Prefrontal Task-Related Activity Representing Visual Cue Location or Saccade Direction in Spatial Working Memory Tasks. *Journal of Neurophysiology*, 87(1):567–588.

Takeuchi, T., Duszkiewicz, A. J., and Morris, R. G. M. (2014). The synaptic plasticity and memory hypothesis: Encoding, storage and persistence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1633).

Tegnér, J., Compte, A., and Wang, X.-J. (2002). The dynamical stability of reverberatory neural circuits. *Biological Cybernetics*, 87(5-6):471–481.

Traub, R., Buhl, E., Gloveli, T., and Whittington, M. (2003). Fast rhythmic bursting can be induced in layer 2/3 cortical neurons by enhancing persistent Na+ conductance or by blocking BK channels. *Journal of neurophysiology*, 89(2):909–21.

Traub, R., Contreras, D., Cunningham, M., Murray, H., LeBeau, F., Roopun, A., Bibbig, A., Wilent, W., Higley, M., and Whittington, M. (2004). Single-column thalamocortical network model exhibiting gamma oscillations, sleep spindles, and epileptogenic bursts. *Journal of Neurophysiology*, 93(4):2194–232.

Tripathy, S. J., Savitskaya, J., Burton, S. D., Urban, N. N., and Gerkin, R. C. (2014). NeuroElectro: A window to the world's neuron electrophysiology data. *Frontiers in Neuroinformatics*, 8.

Tritsch, N. X., Ding, J. B., and Sabatini, B. L. (2012). Dopaminergic neurons inhibit striatal output through non-canonical release of GABA. *Nature*, 490(7419):262–6.

Tsodyks, M. (1999). Attractor neural network models of spatial maps in hippocampus. *Hippocampus*, 9(4):481–9.

Tsodyks, M. and Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the National Academy of Sciences*, 94(January):719–723.

Tsodyks, M., Pawelzik, K., and Markram, H. (1998). Neural Networks with Dynamic Synapses. *Neural Computation*, 10(4):821–835.

Tsodyks, M. and Sejnowski, T. (1995). Associative memory and hippocampal place cells. *Neural Systems*, 6:81–86.

Tsodyks, M., Uziel, A., and Markram, H. (2000). Synchrony generation in recurrent networks with frequency-dependent synapses. *The Journal of Neuroscience*, 20(1):RC50–RC50.

Ubuntu (2017). Ubuntu 16.04.1 LTS (Xenial Xerus). http://releases.ubuntu.com/16.04/.

uWSGI (2017). The uWSGI project — uWSGI 2.0 documentation. https://uwsgi-docs.readthedocs.io/en/latest/.

van Kampen, N. G. (1992). *Stochastic Processes in Physics and Chemistry*. North Holland, 2nd edition.

Vergara, C., Latorre, R., Marrion, N. V., and Adelman, J. P. (1998). Calcium-activated potassium channels. *Current Opinion in Neurobiology*, 8(3):321–329.

Wang, H., Lam, K., Fung, C. C. A., Wong, K. Y. M., and Wu, S. (2015). Rich spectrum of neural field dynamics in the presence of short-term synaptic depression. *Physical Review E*, 92(3):032908.

Wang, M., Yang, Y., Wang, C.-J., Gamo, N. J., Jin, L. E., a Mazer, J., Morrison, J. H., Wang, X.-J., and Arnsten, A. F. T. (2013). NMDA receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex. *Neuron*, 77(4):736–49.

Wang, X. (2002). Probabilistic Decision Making by Slow Reverberation in Cortical Circuits. *Neuron*, 36:955–968.

Wang, X. J. (1999). Synaptic basis of cortical persistent activity: The importance of NMDA receptors to working memory. *The Journal of Neuroscience*, 19(21):9587–603.

Wang, X. J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neurosciences*, 24(8):455–63.

Wang, Y., Markram, H., Goodman, P. H., Berger, T. K., Ma, J., and Goldman-Rakic, P. S. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nature Neuroscience*, 9(4):534–42.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244.

Warden, M. R. and Miller, E. K. (2010). Task-Dependent Changes in Short-Term Memory in the Prefrontal Cortex. *Journal of Neuroscience*, 30(47):15801–15810.

Wei, A. D., Gutman, G. A., Aldrich, R., Chandy, K. G., Grissmer, S., and Wulff, H. (2005). International Union of Pharmacology. LII. Nomenclature and Molecular Relationships of Calcium-Activated Potassium Channels. *Pharmacological Reviews*, 57(4):463–472.

Wei, Z., Wang, X.-J., and Wang, D.-H. (2012). From distributed resources to limited slots in multiple-item working memory: A spiking network model with normalization. *Journal of Neuroscience*, 32(33):11228–40.

White, J. M., Sparks, D. L., and Stanford, T. R. (1994). Saccades to remembered target locations: An analysis of systematic and variable errors. *Vision Research*, 34(1):79–92.

Wibisono, A., Jog, V., and Loh, P.-L. (2017). Information and estimation in Fokker-Planck channels. *arXiv preprint arXiv:1702.03656*.

Willms, A. R., Baro, D. J., Harris-Warrick, R. M., and Guckenheimer, J. (1999). An improved parameter estimation method for Hodgkin-Huxley models. *Journal of Computational Neuroscience*, 6(2):145–168.

Wilson, H. R. and Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13(2):55–80.

Wimmer, K., Nykamp, D. Q., Constantinidis, C., and Compte, A. (2014). Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature Neuroscience*, 17(3):431–439.

Wong, K.-F. and Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *The Journal of Neuroscience*, 26(4):1314–28.

Wu, S., Amari, S.-I., and Nakahara, H. (2002). Population coding and decoding in a neural field: A computational study. *Neural Computation*, 14(5):999–1026.

Wu, S., Hamaguchi, K., and Amari, S.-i. (2008). Dynamics and computation of continuous attractors. *Neural Computation*, 20(4):994–1025.

Wu, S., Wong, K. Y. M., Fung, C. C. A., Mi, Y., and Zhang, W. (2016). Continuous Attractor Neural Networks: Candidate of a Canonical Model for Neural Information Representation. *F1000Research*, 5.

Yang, C. R., Seamans, J. K., and Gorelova, N. (1996). Electrophysiological and morphological properties of layers V-VI principal pyramidal cells in rat prefrontal cortex in vitro. *Journal of Neuroscience*, 16(5):1904–1921.

Yoon, K., Buice, M. A., Barry, C., Hayman, R., Burgess, N., and Fiete, I. R. (2013). Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature Neuroscience*, 16(8):1077–1084.

York, L. C. and van Rossum, M. C. W. (2009). Recurrent networks with short term synaptic depression. *Journal of Computational Neuroscience*, 27(3):607–620.

Yu, F. H. and Catterall, W. A. (2004). The VGL-chanome: A protein superfamily specialized for electrical signaling and ionic homeostasis. *Science Signaling*, 2004(253):re15.

Zador, A. (1998). Impact of Synaptic Unreliability on the Information Transmitted by Spiking Neurons. *Journal of Neurophysiology*, 79(3):1219–1229.

Zenke, F., Agnes, E. J., and Gerstner, W. (2015). Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature Communications*, 6.

Zenke, F. and Gerstner, W. (2017). Hebbian plasticity requires compensatory processes on multiple timescales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1715):20160259.

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience*, 16(6):2112–2126.

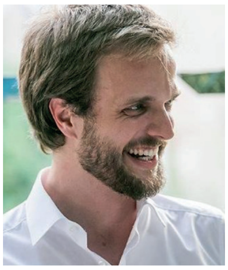
Zhou, F. and Hablitz, J. (1996). Layer I neurons of the rat neocortex. II. Voltage-dependent outward currents. *Journal of Neurophysiology*, 76(2):668–82.

Zucker, R. S. and Regehr, W. G. (2002). Short-term synaptic plasticity. *Annual Review of Physiology*, 64:355–405.

ALEXANDER SEEHOLZER

COMPUTATIONAL SCIENTIST

+41 77 908 26 76 – alex@seeholzer.de – Chemin de Champ-Rond 21, CH-1010 Lausanne
LinkedIn GitHub



Specialist in computational neuroscience, data science and mathematical modeling, **fast and creative thinker** and **great at improvising.**

EDUCATION

EPF Lausanne
Since 2012

PHD CANDIDATE

Life Sciences Section, Laboratory of Computational Neuroscience

LMU München
2003-2011

DIPLOMA IN MATHEMATICS

Grade: 1.1. Major: Algebraic Geometry, Minor: Theoretical Physics.
Diploma thesis: Computational Biology

Gymnasium Erding
2002

ABITUR (A-Level)

Head of student council for 2 years. Member of tutoring and event committees.

PROFESSIONAL EXPERIENCE

EPF Lausanne
Since 2012

DOCTORAL RESEARCHER:
Laboratory of Computational Neuroscience

Neural network models of short- and long-term memory: simulations on distributed computer clusters and theoretical analysis of network dynamics.

Technical conception and head developer of Neuroscience big-data project with field-wide positive impact: ICGenealogy (icg.neurotheory.ox.ac.uk).

Teaching and supervision for students at Master & PhD levels.
Participated in writing of several successful grant proposals (ERC, BBSRC).

The Catalyst
(voluntary work)
Since 2015

PRESIDENT, MANAGING DIRECTOR

Head of science entertainment association with around 40 members. Development and execution of science communication projects in theater and diverse media. Several active roles: team management, direction and marketing of projects, held internal trainings for acting, improvisation & rhetorics.

EPF Lausanne
2008

INTERNSHIP (6 Mo)

Laboratory of Computational Neuroscience: Data fitting methods for the prediction of single brain cell activity with simplified neuron models.

TV1.EU
2005 – 2011

STUDENT WORKER: WEB DEVELOPER

Java CMS development, full-stack web-application prototyping, development and graphics design. Held internal IT trainings and presentations for several costumers.

Freelance
2002 – 2011

WEB DEVELOPER & GRAPHICS DESIGNER

Self-marketing and customer acquisition. Planning and execution of several smaller web projects, involving UI and corporate graphics design.

TECHNICAL SKILLS

Science	Deterministic and stochastic dynamical systems: Specialized in biologically constrained as well as artificial neural networks, reaction-diffusion systems. Machine learning: dimensionality reduction, clustering, classification, deep learning.
IT-Scientific	C++, MPI, Python, Mathematica, Matlab, LaTeX
IT-Web	HTML & CSS, Java, JavaScript (node.js, d3.js), PHP, Python (Django, SQLAlchemy), MySQL, web server administration (Linux, Apache, Nginx)
IT-Media	Photoshop, Illustrator, After Effects, Premiere, DaVinci Resolve, Flash, Ableton Live, all MS-Office products
Management	Target Process (Kanban), Trello, Slack

LANGUAGES

German: Native language

French: Verbal and written working proficiency

Japanese: Basic proficiency

English: Native language

Spanish: Verbal working proficiency, limited written proficiency

AWARDS

Admittance to Advanced Course in Computational Neuroscience, Poland (2013)

Travel scholarship, Studienstiftung des deutschen Volkes (2008)

Scholarship for undergraduate students with outstanding curriculum, Studienstiftung des deutschen Volkes (2006-2010)

EXTRA-CURRICULAR ACTIVITIES

Improvisational theater: regular appearance in shows.

Open source development (github.com/flinz).

Music: Piano and several other instruments, audio production and DJ-ing.

Film making: e.g. Handmade (2015), winner of *Coup de Cœr, Haut & Court short film competition*, Lausanne.

Traveling: extended trips throughout Asia, Australia, Europe & Northern America.

PERSONAL DETAILS

33 years old, in a stable relationship with 1 child.

German, British and Canadian citizenships. Swiss B-permit.

Alexander Seeholzer: Publications

Peer-reviewed publications (Google Scholar)

2017	R. Duarte, A. Seeholzer, A. Morrison (2017). *Synaptic patterning and the timescales of cortical dynamics*. Current Opinion in Neurobiology.
	W. Podlaski*, A. Seeholzer*, L. Groschner, G. Miesenboeck, R. Ranjan, T. Vogels (2017). *Mapping the function of neuronal ion channels in model and experiment*. eLife 2017;6:e22152.
	F. Colombo, A. Seeholzer, W. Gerstner (2017). *Deep Artificial Composer: A Creative Neural Network Model for Automated Melody Generation*. EvoMUSART 2017, LNCS, Springer.
2016	M. Deger, A. Seeholzer, W. Gerster (submitted) *Multi-contact cooperativity in spike-timing dependent structural plasticity stabilizes networks*. preprint arXiv:1609.05730.
	F. Colombo, S. P. Muscinelli, A. Seeholzer, J. Brea, W. Gerstner (2016). *Algorithmic Composition of Melodies with Deep Recurrent Neural Networks*. Proc. 1st Conf. Comp. Sim. Musical Creativity
2014	A. Seeholzer, E. Frey, B. Obermayer (2014). *Periodic versus Intermittent Adaptive Cycles in Quasispecies Coevolution*. Physical Review Letters, 113(12), 128101.

Conference presentations

2016	M. Deger, A. Seeholzer, W. Gerstner (2016). *A spike timing dependent model of dendritic spine plasticity and turnover*. **Bernstein Conference 2016**.
	A. Seeholzer, W. Podlaski, L. Groschner, G. Miesenboeck, T. Vogels (2016). *ICGenealogy: Towards a Common Topology of Neuronal Ion Channel Function and Genealogy in Model and Experiment*. **BMC Neuroscience 16**.
2015	A. Seeholzer, M. Deger, W. Gerstner (2015). *A mean-field theory for slow drift-diffusion dynamics of spiking continuous attractor models*. **Bernstein Conference 2015**.
2014	A. Seeholzer, M. Deger, W. Gerstner (2014). *Stabilizing working memory in spiking networks with biologically plausible synaptic dynamics.* **BMC Neuroscience 15**.
	W. Podlaski, A. Seeholzer, R. Ranjan, H. Markram, T. Vogels, W. Gerstner (2014). *Visualizing the similarity and pedigree of neuronal ion channel models for NEURON*. **CoSyne 2014**, II-89.
2013	W. Podlaski, A. Seeholzer, R. Ranjan, H. Markram, T. Vogels, W. Gerstner (2013). *Visualizing the similarity and pedigree of NEURON ion channel models available on ModelDB.* **SFN 2013**, 678.31/MMM42

Invited oral presentations

October 2016	*A mean-field theory for drift-diffusion dynamics of spiking continuous attractor models*. Oxford Neurotheory Forum. University of Oxford.
July 2016	*ICGenealogy: Mapping the function of neuronal ion channels in model and experiment*. Neural Computation Laboratory. Korea University, Seoul.
June 2015	*Computational Neuroscience at the LCN*. INCF Swiss Node Inaugural Meeting. Centre Biotech, Geneva.
May 2015	*Visualizing the similarity and pedigree of neuronal ion channel models for NEURON*. Open Source Brain Workshop 2015. Alghero, Sardinia.
March 2015	*Spiking neural networks at the LCN*. 1st HBP Workshop: Community Simulator Development. EITN, Paris.