A unifying functional approach towards synaptic long-term plasticity

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

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

von
Daniel Krieg
aus Frankfurt am Main

Frankfurt (2012)
(D 30)

vom Fachbereich Physik der

Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Joachim Stroth

Gutachter: Prof. Dr. Jochen Triesch, Prof. Dr. Matthias Kaschube

Datum der Disputation: 27. Mai 2013

"I can live with doubt, and uncertainty, and not knowing.
I think it's much more interesting to live not knowing
than to have answers which might be wrong."

Richard P. Feynman

Danksagungen

Ich möchte mich bei allen Personen bedanken, die zur Entstehung dieser Dissertation beigetragen haben. Der größte Dank gilt dabei meinem Betreuer und Doktorvater Herrn Prof. Dr. Jochen Triesch, der mich trotz meiner geringen Vorkenntnisse in den Neurowissenschaften in seine Arbeitsgruppe aufnahm. Sein mir entgegengebrachtes Vertrauen war für mich stets eine große Bestätigung. Ich danke ihm sowohl für den Freiraum, den er mir bei der Verfolgung eigener wissenschaftlicher Ideen gewährte, als auch für die verlässliche Unterstützung durch seinen Rat, die konstruktive Kritik und die inspirierenden Diskussionen. Ich werde diese drei Jahre auch aufgrund der angenehmen und kollegialen Atmosphäre in sehr positiver Erinnerung halten.

Für die bereitwillige Übernahme des Zweitgutachtens bedanke ich mich herzlich bei Herrn Prof. Dr. Matthias Kaschube.

Desweiteren gilt mein Dank meinen Kollegen des Bereichs Neurowissenschaften am FIAS für den kreativen, wissenschaftlichen Gedankenaustausch über die fachspezifischen Grenzen hinweg und die positive, angenehme Atmosphäre. Insbesondere danke ich Thomas Weisswange für die Diskussionen und das Feedback zum Hauptprojekt meiner Arbeit und Marc Henniges für die Korrekturlesung der Dissertation.

Und nicht zuletzt bin ich meiner Frau Daniela für ihre konstante moralische Unterstützung zutiefst dankbar. Ihr Verständnis und ihre Geduld haben mir besonders in den schwierigeren Phasen der Arbeit sehr geholfen.

Zusammenfassung

Das Gehirn ist die wohl komplexeste Struktur auf Erden, die der Mensch erforscht. Es besteht aus einem riesigen Netzwerk von Nervenzellen, welches in der Lage ist eingehende sensorische Informationen zu verarbeiten um daraus eine sinnvolle Repräsentation der Umgebung zu erstellen. Außerdem koordiniert es die Aktionen des Organismus, um mit der Umgebung zu interagieren. Das Gehirn hat die bemerkenswerte Fähigkeit sowohl Informationen zu speichern als auch sich ständig an ändernde Bedingungen anzupassen, und zwar über die gesamte Lebensdauer. Dies ist essentiell für Mensch oder Tier, um sich zu entwickeln und zu lernen.

Die Entwicklung eines menschlichen Kindes in den ersten Jahren und die Fähigkeiten, die es währenddessen erwirbt, sind bisher von keinem Computer-Algorithmus erreichbar. Die Grundlage für diesen lebenslangen Lernprozess ist die Plastizität des Gehirns, welche das riesige Netzwerk von Neuronen ständig anpasst und neu verbindet. Dieses Phänomen der neuronalen Plastizität beschäftigt die Neurowissenschaften und anderen Disziplinen bereits über mehrere Jahrzehnte. Dabei induzieren lokale, selbstorganisierte Mechanismen Veränderungen an den synaptischen Verbindungen und der intrinsischen Erregbarkeit jedes Neurons und optimieren dadurch das Verhalten des Organismus als Ganzes.

Dabei beschreibt die intrinsische Plastizität die ständige Anpassung der Erregbarkeit eines Neurons um einen ausbalancierten, homöostatischen Arbeitsbereich zu gewährleisten. Aber besonders die synaptische Plastizität, welche die Änderungen in der Stärke bestehender Verbindungen bezeichnet, wurde unter vielen verschiedenen Bedingungen erforscht und erwies sich mit jeder neuen Studie als immer vielschichtiger. Sie wird durch ein komplexes Zusammenspiel von biophysikalischen Mechanismen induziert und hängt von verschiedenen Faktoren wie der Frequenz der Aktionspotentiale, deren Timing und dem Membranpotential ab und zeigt außerdem eine metaplastische Abhängigkeit von vergangenen Ereignissen. Letztlich beeinflusst die synaptische Plastizität die Signalverarbeitung und Berechnung einzelner Neuronen und der neuronalen Netzwerke.

Der Schwerpunkt dieser Arbeit ist es das Verständnis der biologischen Mechanismen und deren Folgen, die zu den beobachteten Plastizitätsphänomenen

führen, durch eine stärker vereinheitlichte Theorie voranzutreiben. Ein solcher einheitlicher Ansatz kann Einblicke in die folgenden drei verschiedenen Dimensionen liefern, welche für synaptische Plastizität relevant sind.

Synaptische Plastizität betrifft unterschiedliche Ebenen theoretischer Abstraktion. Das komplexen Verhalten von Zellen und ihrer Wechselwirkungen in lebenden Organismen zu verstehen erfordert Untersuchungen auf verschiedenen Ebenen der Abstraktion. Marr [2010] schlug vor, drei Ebenen der theoretischen Analyse zu unterscheiden: Funktionelles Ziel, Algorithmus und Implementierung. Mechanistische Modelle über die zugrunde liegenden Mechanismen folgen einem Bottom-up Ansatz. Sie beschreiben die Implementierung der Plastizität und bilden eine Grundlage für Modellierungsstudien. Jedoch lassen sich deren Auswirkungen auf das Verhalten des Netzwerkes, jenseits einzelner Synapsen, nur schwer verstehen. Eine einfache Möglichkeit, um diese Auswirkungen auf das Netzwerk zu untersuchen, bieten phänomenologische Modelle die nur die resultierenden Phänomene und Abhängigkeiten beschreiben. Aber dieser Ansatz ist von Natur aus begrenzt, denn er kann nicht zu einem tieferen Verständnis der komplexen Wechselwirkungen zwischen den verschiedenen Mechanismen führen. Daher ist es wichtig über diese phänomenologischen Beschreibungen hinauszugehen. Hier kann ein Top-Down Ansatz helfen, mögliche funktionale Ziele des Systems zu betrachten. In dieser Hinsicht können die funktionalen Auswirkungen der synaptischen Plastizität als das Ziel betrachtet werden, welches auf Synapsen-, Zell- oder Netzwerkebene erreicht werden soll. Der wichtigste Schritt ist dann die Ableitung einer geeigneten, biologisch plausiblen Lernregel (Algorithmus) und die Identifizierung der zugrunde liegenden biologischen Mechanismen als mögliche Implementierung dieser Lernregel.

Synaptische Plastizität betrifft unterschiedliche Skalen der neuronalen Funktion und Organisation. Die neuronalen Netzwerke des Gehirns können auf unterschiedlichen Skalen untersucht werden: von intra- und subzellulären Mechanismen, über die Signalverarbeitung und Konnektivität von einzelnen Neuronen bis hin zu dem Informationsfluss zwischen großen Netzwerken oder ganzen Hirnarealen. Das Phänomen der synaptischen Plastizität liegt genau in der Mitte zwischen dem großen und kleinen Maßstab. Die Untersuchung der Mech-

anismen betrifft den kleinen Maßstab der subzellulären Reaktionen und Wechselwirkungen von Ionen, Proteinen und anderen Molekülen. Und während die Effekte der synaptischen Plastizität direkte Auswirkungen auf die Signalverarbeitung einer einzelnen Zelle haben, beeinflussen sie dadurch die Berechnung im Netzwerk und betreffen daher auch Fragen in Bezug auf die funktionalen Auswirkungen.

Synaptische Plastizität betrifft unterschiedliche Faktoren der neuronalen Verarbeitung. In den letzten sechs Jahrzehnten wurde synaptische Plastizität experimentell und theoretisch untersucht. Seitdem wurde entdeckt, dass die Verbindungsstärke beziehungsweise deren Änderung viele Abhängigkeiten besitzt: die Aktivität des prä- und postsynaptischen Neurons, das postsynaptische Membranpotential, das relative Timing zwischen prä- und postsynaptischen Aktionspotentialen, den Ort der Synapse, etc. Diese verschiedenen Phänomene werden meist unabhängig voneinander durch verschiedene experimentelle Protokolle betrachtet. Allerdings untersuchen diese Protokolle nur unterschiedliche Aspekte eines gemeinsamen biophysikalischen Mechanismus. Eine allgemeine Theorie von synaptischer Plastizität erfordert daher die Beschreibungen all dieser Abhängigkeiten aus einem Ansatz heraus.

Einige der größten theoretischen Fortschritte in der Wissenschaft wurden durch Vereinheitlichungen von scheinbar unterschiedliche Phänomene erreicht, wie zum Beispiel Newtons Vereinheitlichung der Gesetze des freien Falls mit den Gesetzen der Planetenbewegung in einem einzigen Gesetz der Schwerkraft. Solche Vereinigungen führen zu einem besseren und abstrakteren Verständnis der beobachteten Phänomene. Bislang fehlt solch ein einheitlicher Ansatz für synaptische Plastizität. Es wird experimentell und theoretisch als eine Mischung aus verschiedenen Phänomene untersucht und modelliert und keines der existierenden Modelle ist in der Lage, die Lücke in der theoretischen Abstraktion zu überbrücken. Die Hauptmotivation dieser Arbeit ist es daher, einen neuen Ansatz zu liefern, der ein einheitlicheres Verständnis von synaptischer Plastizität ermöglicht. Eine Theorie basierend auf einem funktionalen Ziel kann scheinbar unterschiedliche Abhängigkeiten vereinheitlichen und schafft damit eine Verbindung zwischen den verschiedenen Skalen und Abstraktionen. In den Kapiteln 3 und 4 stelle ich zwei funktionale Ziele für neuronale Plas-

tizität auf, leite Lernregeln aus diesen ab und analysiere deren Konsequenzen und Vorhersagen.

Kapitel 3 untersucht die Unterscheidbarkeit der Populationsaktivität in Netzwerken als funktionales Ziel für neuronale Plastizität. Die Hypothese ist dabei, dass gerade in rekurrenten aber auch in vorwärtsgekoppelten Netzwerken die Populationsaktivität als Repräsentation der Eingangssignale optimiert werden kann, wenn ähnliche Eingangssignale eine möglichst unterschiedliche Repräsentation haben und dadurch für die nachfolgende Verarbeitung besser unterscheidbar sind. Das funktionale Ziel ist daher diese Unterscheidbarkeit durch Veränderungen an den Verbindungsstärke und der Erregbarkeit der Neuronen mithilfe von lokalen selbst-organisierten Lernregeln zu maximieren. Ich zeige, dass ausgehenden von diesem Ziel die am häufigsten verwendeten Lernregeln für intrinsische sowie synaptische Plastizität in künstlichen neuronalen Netzwerken abgeleitet werden können [Krieg et al., 2010].

Die synaptische Lernregel entspricht dabei dem Standardmodell für timingabhängige Plastizität (STDP) sowohl für ein zeitdiskretes Neuron mit Ratenkodierung als auch für einen zeitkontinuierliches, spikendes Neuron. Der funktionale Ansatz führt dabei zu einem zusätzlichen, modulierender Faktor, welcher die synaptische Plastizität stabilisiert. Das heißt, die synaptischen Verbindungsstärke divergiert nicht und erfordert keine zusätzliche Normalisierung. Die synaptischen Verbindungen regulieren sich selbst, um eine balanciertes, homöostatisches Eingangssignal zu erzeugen. Dies kann daher als Mechanismus zur synaptischen Skalierung interpretiert werden. Auch dieser Stabilisierungeffekt tritt dabei sowohl für zeitdiskrete als auch zeitkontinuierliche Neuronen auf.

Das gleiche funktionale Ziel kann auch die Erregbarkeit eines Neurons optimieren. Die Lernregel für intrinsischen Plastizität, welche daraus abgeleitet wird, ist ähnlich zu früheren informationstheoretischen Ansätzen [Bell and Sejnowski, 1995]. Eine Erweiterung, basierend auf metabolischen Beschränkungen und Störanfälligkeit eines neuronalen Codes, führt zusammen mit dem ersten Ziel zu einer realistischeren Verteilung der Feuerrate und einer zusätzlichen Lernregel für inhibitorische Neuronen. Insofern erlaubt Kapitel 3 eine Reihe von Standard-Lernenregeln für künstliche neuronale Netze [Bell and Sejnowski, 1995; Triesch, 2005; Babadi and Abbott, 2010; Vogels et al., 2011] aus

einem gemeinsamen funktionalen Ziel abzuleiten.

Inspiriert von den weitreichenden Konsequenzen, welche aus einer so einfachen Beschreibung wie in Kapitel 3 folgen, erweiterte ich den Ansatz auf ein komplexeres, biophysikalisches Neuronenmodell in Kapitel 4 [Krieg and Triesch, 2011a,b, 2012 submitted]. Die dort abgeleitete Lernregel ist in der Lage, direkte Vergleiche mit den experimentellen Ergebnissen zu synaptischer Plastizität zu machen. Das Ziel ist eine spärliche, stark asymmetrische Verteilung der synaptischen Stärke wie sie auch bereits mehrfach experimentell gefunden wurde. Exzitatorische Verbindungen folgen in etwa einer Log-normal Verteilung. Das heißt, es gibt viele schwache Verbindungen, jedoch einige wenige, die sehr stark sind, und diese Asymmetrie kann durch das statistische Maß der Schiefe charakterisiert werden. Da nicht nur die Generierung von Aktionspotentialen hohe energetische Kosten für das presynaptische Neuron verursacht, sondern auch deren Effekte auf das empfangende, postsynaptische Neuron, kann eine solche schiefe Verteilung der synaptische Stärke zu einer energieeffizienten Signalübertragung beitragen. Denn anstatt die Signale von vielen unkorrelierten, mittelstarken Synapsen zu integrieren, wird das Neuron von wenigen starken Verbindungen zum Erzeugen eines Aktionspotentials angeregt.

Das funktionale Ziel ist daher die Maximierung der Schiefe dieser Verteilung durch lokale, synaptische Lernregeln. Aus diesem funktionalen Ansatz können alle wichtigen Phänomene der synaptischen Plastizität erklärt werden. Simulationen der Lernregel in einem realistischen Neuronmodell mit voller Morphologie erklären die Daten von timing-, raten- und spannungsabhängigen Plastizitätsprotokollen. Die Lernregel hat auch eine intrinsische Abhängigkeit von der Position der Synapse auf dem Dendritenbaum, welche mit den experimentellen Ergebnissen übereinstimmt. Darüber hinaus kann die Lernregel ohne zusätzliche Annahmen ein Phänomene der sogenannten Metaplastizität erklären. Diese beschreibt, dass die Plastizitätsmechanismen selbst plastisch sind und zum Beispiel von der vorherigen neuronalen Aktivität abhängen können. Dabei erklärt der Ansatz nicht nur eine bekannte Form der raten-basierten Metaplastizität, sondern sagt auch eine neue Form der Metaplastizität voraus, welche die timing-abhängige Plastizität beeinflusst.

Andere Ansätze für funktionale Ziele der synaptischen Plastizität wurden

bereits zuvor postuliert. Jedoch waren diese zum Einen auf ein einzelnes Phänomen beschränkt und lieferten keine einheitliche Sicht auf die experimentellen Phänomene. Zum Anderen führten sie nicht zu biologisch plausiblen Lernregeln und konnten somit auch keine Verbindung zu den biologischen Mechanismen herstellen. Somit besteht der Beitrag der vorliegende Arbeit aus zwei neuartigen Vereinheitlichungen für synaptische Plastizität: Erstens zeigt sie, dass die verschiedenen Phänomene der synaptischen Plastizität als Folge eines einzigen funktionalen Ziels verstanden werden können. Und zweitens überbrückt der Ansatz die Lücke zwischen der funktionalen und mechanistische Beschreibungsweise. Das vorgeschlagene funktionale Ziel führt zu einer Lernregel mit biophysikalischer Formulierung, welche mit etablierten Theorien der biologischen Mechanismen in Verbindung gebracht werden kann. Außerdem kann das Ziel einer spärlichen Verteilung der synaptischen Stärke als Beitrag zu einer energieeffizienten synaptischen Signalübertragung und optimierten Kodierung interpretiert werden.

In der vorliegenden Arbeit konnte ich zeigen, dass es möglich ist die verschiedenen Mechanismen der neuronalen Plastizität in künstlichen Netzwerken auf ein gemeinsames funktionales Ziel abzubilden. Speziell für synaptische Plastizität lassen sich alle wichtigen experimentellen Phänomene aus einem funktionalen Ziel erklären und die biophysikalischen Mechanismen als Implementierung interpretieren. Daher lässt sich vermuten, dass die Mechanismen der neuronalen Plastizität, welche die wohl wichtigste Eigenschaft des Gehirns darstellen, sich gemeinsam unter evolutionär relevanten Einflüssen entwickelt haben, um verschiedene funktionale Aspekte der neuronalen Signalverarbeitung und Berechnung durch lokale Selbstorganisation zu optimieren.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis investigates a functional view on the plastic changes at the synaptic connections between neurons. The hypothesis is that the biological mechanisms of this synaptic plasticity have evolved following some functional goal. The aim of this thesis is to propose a specific and intuitive functional goal and to derive a learning rule from it. The hypothesis will be supported by comparing its theoretical predictions to experimental data and biophysical mechanisms.

1.1 Motivation

The brain is the most complex structure known to mankind. It consists of a vast network of nerve cells that is able to process incoming sensory information in order to make sense of the world. It coordinates the actions of the organism in order to interact with the environment. The brain has the remarkable ability to memorize and store information as well as to adapt to changing conditions throughout lifetime. This is essential for a human or animal to develop and learn.

The development of a human child and the abilities it acquires over only a few years are still unmatched by any computer algorithm. The basis for this life-long learning process is the plasticity of the brain, which constantly adapts and rewires the huge network of neurons. This phenomenon of neural plasticity has attracted the attention of neuroscience and other disciplines for several decades already. It is driven by local self-organized mechanisms changing the

synaptic connections and the intrinsic excitability of each neuron. Thereby, neural plasticity optimizes the overall behavior of the organism.

Intrinsic plasticity continuously adapts the excitability to maintain a homeostatic set point. But especially synaptic plasticity, which describes the changes in the strength of existing connections, proved to be more and more intricate as the studies progressed. It is induced by a complex interaction of biophysical mechanisms, depends on various factors such as firing rate, spike-timing, and membrane potential, exhibits metaplastic dependence on the context, and, ultimately, influences the computation of the neuron and the network it is embedded in.

The main focus of this thesis is to further the understanding of the biological mechanisms and their consequences leading to the observed phenomena by proposing a more unified theory. Such a unified approach can provide insights into the following three different dimensions which are relevant to synaptic plasticity:

Synaptic plasticity regards different levels of theoretical abstractions. Understanding the complex behavior of cells and their interactions arising in living systems requires studies at different levels of abstraction. Marr [2010] proposed to distinguish three levels of theoretical analysis: computational goal, algorithm, and implementation. In order to get a coherent image of the mechanisms as well as the functional consequences of synaptic plasticity, it is important to model it at these different levels.

Mechanistic models about the underlying mechanisms can be considered the most grounded approach. They start from the biological implementation to identify the biophysical mechanisms. This provides the basis for computational modeling studies simulating the process in detail. But it is hard to assess the effects of these mechanisms beyond single synapses.

The intermediate algorithmic level is best described by phenomenological models which only model the observed phenomena. They provide a simple way to study the effects on the network scale. But they are inherently limited, since they can not provide a deeper understanding of the complex interactions between the diverse mechanisms. Therefore, it is important to go beyond those models.

A top-down approach can help to evaluate possible functional goals the system wants to achieve. In that respect, the functional implications of synaptic plasticity can be regarded as the computational goal to be achieved by the network. The important step is then to connect the different levels by deriving a suitable, biologically plausible learning rule (algorithm) and identifying the underlying biological mechanisms (implementation).

Synaptic plasticity regards different scales of neuronal function and organization. The neural networks of the brain are studied at different scales: from intra- and sub-cellular mechanisms over computation and connectivity of single neurons to the information flow between large networks or whole brain areas. The phenomenon of synaptic plasticity lies right in the middle between the larger and the smaller scale. The study of its mechanisms is connected to processes at smaller scale, namely sub-cellular reactions and cascades of interactions between ions, proteins and other molecules. And while the effects of synaptic plasticity directly affect the computation of a single cell, they also influence the computation of the larger network touching upon questions regarding the functional implications.

Synaptic plasticity regards different quantities of neuronal computation. It has been studied experimentally and theoretically over the last six decades. Since then the synaptic connection strength and its change have been found to have many dependencies: the activity of the pre- and postsynaptic neuron, the postsynaptic membrane potential, the relative timing between pre- and postsynaptic spikes, the location of the synapse, etc. The observations of these different phenomena are achieved by using different experimental protocols. But they only probe different aspects of a common biophysical mechanism.

1.1.1 Objectives

Some of the greatest theoretical advances in science have been unifications of seemingly different phenomena such as Newton's unification of the laws of free fall and the laws of planetary motion into a single law of gravity. Such unifications lead to a more fundamental and abstract understanding of the

observed phenomena. So far, synaptic plasticity is missing such a unified approach. It is experimentally probed and theoretically modeled as a mixture of distinct phenomena. And none of the existing models are able to bridge the gap in theoretical abstraction and spatial scale. *The main motivation of this thesis is, therefore, to provide a new approach which allows for a more unified understanding of synaptic plasticity.* An understanding in terms of a computational goal can unify seemingly different dependencies and create a connection across the different scales and levels of abstraction.

The hypothesis of this thesis it that synaptic plasticity has evolved to optimize one or more computational goals. To be considered a reasonable computational goal for synaptic plasticity, it has to fulfill three requirements: First, it has to reproduce and unify experimental findings. Since it is probable that the brain has many different computational goals to achieve, the proposed goal does not need to explain every possible phenomena but at least a major subset. Second, it has to bridge all levels of theoretical analysis. That means, it should lead to a learning rule that is biologically plausible and show connections to the underlying biological mechanisms. Third, it should have a natural interpretation and optimize quantities that are evolutionary relevant.

1.2 Contributions

In the main Chapters 3 and 4, I introduce computational goals for neural plasticity, propose learning rules derived from them, and analyze their consequences and predictions.

1.2.1 Separability objective for artificial neural networks

Chapter 3 introduces an objective function for neural plasticity in terms of the separability of the population activity. From this single objective, commonly used learning rules for intrinsic plasticity as well as synaptic plasticity can be derived [Krieg et al., 2010].

The synaptic learning rule resembles the standard models for spike-timing dependent plasticity (STDP) for a discrete-time rate coding neurons as well as for a continuous-time spiking neuron model. The objective function approach introduces an additional modulating factor, which can be shown to

stabilize the synaptic learning, i.e. it does not diverge and requires no additional normalization procedure. The synaptic strengths will adjust themselves to provide a balanced homeostatic input, thereby, directly accounting for the phenomenon of synaptic scaling. The same objective is also useful to optimize the excitability of a neuron. This leads to an intrinsic plasticity rule which is similar to previous information theoretic approaches [Bell and Sejnowski, 1995].

A second objective function regarding metabolic constraints and noise robustness of a neural rate code nicely integrates with the separability objective. It leads to a more realistic distribution of firing rates and additionally introduces a learning rule for inhibitory interneurons. Thus, Chapter 3 proposes a simple way to consistently derive a set of standard learning rules for artificial neuronal networks [Bell and Sejnowski, 1995; Triesch, 2005; Babadi and Abbott, 2010; Vogels et al., 2011] from a common computational goal.

1.2.2 Sparseness objective for synaptic plasticity

Motivated by the findings for the learning rules of artificial neural networks in Chapter 3, I extended the approach to a more complex conductance-based neuron model in Chapter 4 [Krieg and Triesch, 2011a,b, 2012 submitted]. The resulting learning rule is able to make direct comparisons to the experimental findings on synaptic plasticity.

The objective is a sparse distribution of synaptic strength and from this single computational idea all major phenomena of synaptic plasticity can be explained. Simulating the learning rule in a realistic neuron model with full morphology fitted the data from spike-timing-, rate-, and voltage-dependent plasticity protocols. It also has an intrinsic dependence on the synaptic location which agrees with experimental findings. Furthermore, the learning rule naturally accounts for a phenomenon called metaplasticity, where the plasticity mechanisms themselves are plastic and are modulated by, e.g. recent history of neural activity. In this respect, the approach not only explains the findings on activity-based metaplasticity, but also predicts a new form of metaplasticity which will modify the spike-timing dependent plasticity.

While other ideas for computational goals of synaptic plasticity have been postulated, they either were restricted to a single phenomenon and did not

provide a unified view on the experimental findings. Or they did not lead to biologically plausible learning rules and, therefore, also lacked a connection to the biological mechanisms. Thus, this thesis contributes two novel unifications for synaptic plasticity: First, it shows that the different phenomena of synaptic plasticity arise as a consequence of the same computational goal following different experimental protocols. And second, it bridges the gap between the computational and mechanistic view. The proposed computational goal leads to a learning rule in biophysical terms which can be related to established theories of the biological mechanisms. Finally, the objective of a sparse distribution of synaptic strength can be interpreted as contributing to an energy-efficient synaptic signaling and an optimized coding.

Chapter 2

Background and context

The brain contains a large variety of cell types with the main distinction between neuronal cells and glial cells. The complex network of neurons is the basis for the computations in the brain and the plastic changes of this network are key to adaptation and learning. This background chapter gives an introduction to neural processing, neural plasticity, and neural computation and identifies open questions and problems. This sets the stage for the questions addressed in this thesis.

Section 2.1 introduces the general anatomy and functionality of neurons. The main focus lies on a detailed description of the resting potential, the Hodgkin-Huxley model for action potential generation, and the mechanisms of neural communication via synaptic transmission as modeled in Chapter 4. This is complemented with a short description of simplified neuron models usually employed in network simulations and relevant for Chapter 3.

Section 2.2 details the phenomenon of neural plasticity, which is the central theme of this thesis. After a short introduction to intrinsic plasticity guided by information theory, I describe the various experimental findings on synaptic plasticity, how they have been modeled, and which aspects have not been considered so far.

Section 2.3 discusses the constraints which the neural system faces and how they influence its computation. I introduce the principle of energy efficiency as the major guideline in this thesis. The section ends on the topic of sparseness, which has emerged as a central theme in neural function, and provides the necessary tools to address the problem of energy-efficient neural information

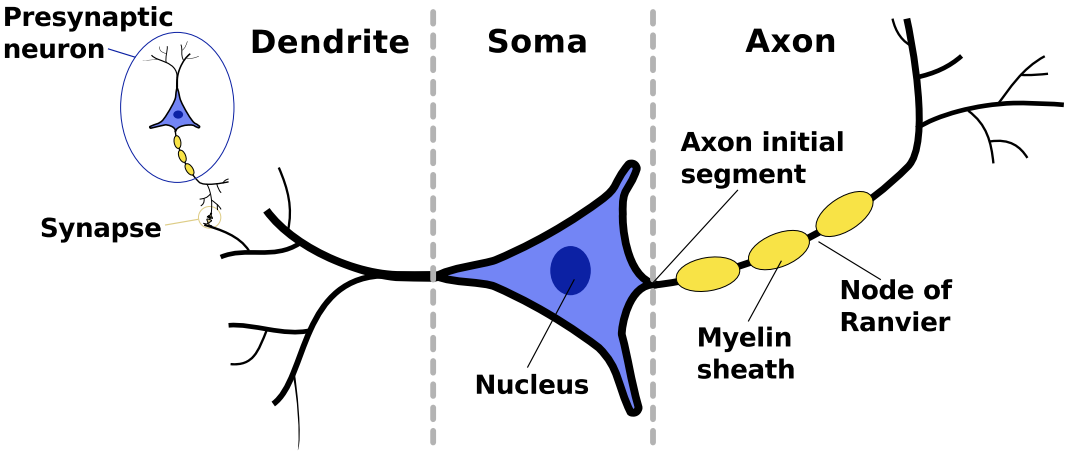

Figure 2.1 Neuronal morphology. The soma is the cell body and contains the nucleus. It has two tree-like extensions: the axon transmits the signal to other neurons. It connects to their dendrites via a synapse.

transmission in Chapter 4.

2.1 Neural processing and communication

Neurons are specialized cells that are excitable and that are able to transmit that excitation to other neurons. They are enclosed by a membrane and have a cell body, called soma, containing the nucleus. The cell body has two types of tree-like extensions: the dendrites and the axon. The cell receives input from other cells at the dendrites, which is then integrated at the soma. The axon is often surrounded by a myelin sheath with equidistant interruptions known as 'Nodes of Ranvier'. The signals from the soma travel down the axon which is connected to the dendrites of other neurons via synapses. The general morphology is shown in Figure 2.1. Details about the synapse structure are described in Section 2.1.1 and Figure 2.5.

Following the textbook by Aidley [1998], I discuss the conductance-based neuron model and the Hodgkin-Huxley model of action potential generation. With these models, the neural dynamics can be described in detail and simulated in a computer. Such simulations are employed in Chapter 4 to compare the results of my model to experimental findings. After that, I describe the simplified rate-based neuron model, which is employed in the derivations and simulations of Chapter 3.

2.1.1 Conductance-based models

Neurons are electrical units and can be described by an electrical circuit consisting of active and passive elements. One can distinguish different types of neurons, which differ in three-dimensional structure, excitability, and effect on other neurons. Despite their differences, the dynamics and the signal transmission work in very similar ways. The majority of neuron types does not communicate with a continuous signal, but with short voltage pulses called action potentials or spikes. One benefit of these spikes is that they can travel over longer distances without suffering from attenutation. They are strong deflections from the resting potential, all-or-none events and always nearly identical in shape. Spikes will travel along the axon where they are transmitted via synapses to other neurons. This transmission leads to synaptic currents which are integrated at the soma of the receiving neuron. Finally, if that input signal exceeds a threshold, the receiving neuron itself emits a spike.

Resting potential

In their resting state (in the absence of excitation), neurons exhibit a potential difference (voltage) between the intracellular and extracellular medium. This is due to active ion pumps transporting ions into or out of the cell across the membrane, thereby generating a concentration gradient for different ion species. The Na^+/K^+-ATPase pump requires one ATP molecule to exchange 2 potassium ions from the outside with 3 sodium ions from the inside of the cell. Under physiological conditions the pump generates a concentration ratio on the order of 10, i.e. the intracellular potassium concentration $[K^+]_i$ is ten times larger than the extracellular concentration $[K^+]_o$.

Beside the active pumps, the membrane contains passive ion channels which let specific ions diffuse into and out of the cell. Due to this permeability, the ions will follow their concentration gradient until a dynamic equilibrium with no net ion flux through the membrane is reached. At resting, the membrane is mainly permeable to potassium. Potassium ions flow out of the cell due to the gradient and create an increasing electrostatic potential which leads to an opposite force. When both forces are equal in strength they neutralize each other and result in a dynamic equilibrium.

The equilibrium potential of an ion species X depends on its concentration ratio and can be calculated by the Nernst equation:

$$E_X = \frac{RT}{zF} \ln \frac{[X]_\mathrm{o}}{[X]_\mathrm{i}} \tag{2.1.1a}$$

$$= \frac{61.54\,\mathrm{mV}}{z} \log_{10} \frac{[X]_\mathrm{o}}{[X]_\mathrm{i}} \quad \text{(at body temperature 310.15°K)}, \tag{2.1.1b}$$

where R is the universal gas constant, T is the temperature in Kelvin, F is the Faraday constant, and z is the number of elementary charges of the ion species. Since $[\mathrm{K}^+]_\mathrm{i}$ is larger than $[\mathrm{K}^+]_\mathrm{o}$, the equilibrium potential E_K of potassium is negative and around $-92\,\mathrm{mV}$ while E_Na is positive and around $+64\,\mathrm{mV}$ [Wright, 2004]. For a membrane being only permeable to potassium, the resting potential would be equal to E_K. The Goldman equation, an extension of the Nernst equation, can take into account that the membrane is also slightly permeable to sodium and chloride:

$$E_r = \frac{RT}{F} \ln \left(\frac{P_\mathrm{Na} [\mathrm{Na}^+]_\mathrm{o} + P_\mathrm{K} [\mathrm{K}^+]_\mathrm{o} + P_\mathrm{Cl} [\mathrm{Cl}^-]_\mathrm{i}}{P_\mathrm{Na} [\mathrm{Na}^+]_\mathrm{i} + P_\mathrm{K} [\mathrm{K}^+]_\mathrm{i} + P_\mathrm{Cl} [\mathrm{Cl}^-]_\mathrm{o}} \right) \tag{2.1.2a}$$

$$= \frac{P_\mathrm{K}}{P_{tot}} E_\mathrm{K} + \frac{P_\mathrm{Na}}{P_{tot}} E_\mathrm{Na} + \frac{P_\mathrm{Cl}}{P_{tot}} E_\mathrm{Cl} \tag{2.1.2b}$$

with the permeabilities P_X for different ions X. The resting potential E_r is a weighted average of the equilibrium potentials of the permeable ions. Given that potassium has the largest permeability, the resting potential is dominated by E_K and is usually around $-70\,\mathrm{mV}$.

A patch of membrane surrounding a short piece of dendrite can be described by an electrical circuit diagram. The concentration gradients established by the active pumps act like a battery providing a voltage equal to the equilibrium potential E_X and the permeabilities P_X are represented by a resistor with a given electrical conductance g_X (the inverse of the resistance R_X) which can be voltage-dependent. The membrane itself corresponds to a capacitor with capacitance C_m. Figure 2.2 depicts the circuit diagram containing potassium, sodium, and a leak current representing chloride and other ions. The ionic currents passing the membrane are given by Ohm's law as $I_X = g_X (u - E_X)$. The currents are proportional to the difference between the ions equilibrium potential E_X and the current membrane potential u. Since the ion flow will reverse its direction if this difference changes its sign, the equilibrium potential is also called reversal potential.

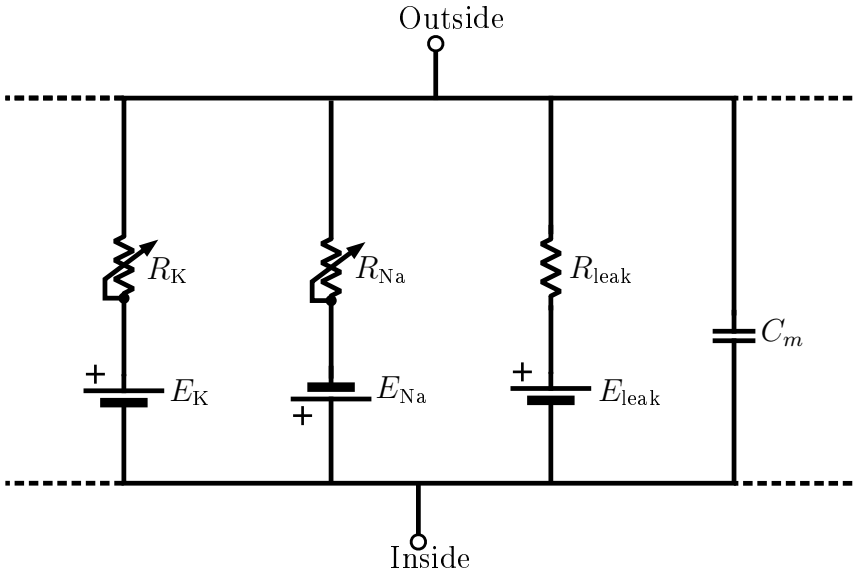

Figure 2.2 Electrical circuit diagram for a patch of membrane. The membrane, separating the intra- and extracellular space, has a specific capacity C_m and contains voltage-dependent ion channels with resistances R_{K} and R_{Na}. The remaining ion channels are represented by a constant resistance R_{leak}. The potential differences E_{ion} are generated by active ion pumps.

Following Figure 2.2 a current I_{membrane} applied to the patch of the membrane can be split in two parts: a capacity current I_C charging the membrane and the ionic currents passing the membrane through the channels:

$$I_{\mathrm{membrane}} = I_C + \sum_{\mathrm{ion}} I_{\mathrm{ion}} \tag{2.1.3a}$$

$$= I_C + g_{\mathrm{K}}\,(u - E_{\mathrm{K}}) + g_{\mathrm{Na}}\,(u - E_{\mathrm{Na}}) + g_{\mathrm{leak}}\,(u - E_{\mathrm{leak}})\,. \tag{2.1.3b}$$

The change in the membrane potential can then be calculated as $C_m \dfrac{\partial u}{\partial t} = I_C = -\sum_{\mathrm{ion}} I_{\mathrm{ion}} + I_{\mathrm{membrane}}$. The membrane current is due to the transverse flow of ions along the membrane, i.e. the propagation of excitation along the dendrite. In the absence of any membrane currents ($I_{\mathrm{membrane}} = 0$) the equilibrium point of the membrane potential is the resting potential E_R, where the ionic currents cancel each other such that their sum is zero ($\sum_{\mathrm{ion}} I_{\mathrm{ion}}$).

The equilibrium, however, is instable, since the membrane is not only permeable to potassium but also to sodium. Thus, there is always an effective ion flux of sodium inwards and potassium outwards. These fluxes will slowly run down the concentration gradients, which is prevented by the constant activity

of the ion pumps.

Action potential generation

Action potentials are very fast voltage spikes where the membrane potential jumps about 100 mV up and down in a few milliseconds. To generate these fast events the cell needs the stored potential energy in the concentration gradients. The gradient acts like a tensioned spring which is released by voltage-dependent ion channels. These channels are primarily located in the soma, at the nodes of Ranvier in the axon, and in the axon initial segment where the action potential is initiated. Hodgkin and Huxley [1952] proposed a mathematical description for changes of the sodium and potassium conductances as a function of the membrane potential. Their analysis revealed how the interplay between these two ions leads to the sharp rise and decay of the voltage during an action potential.

The conductance time course in the Hodgkin-Huxley model is described with three auxiliary variables n, m and h:

$$g_{\mathrm{K}} = \bar{g}_{\mathrm{K}} n^4 \tag{2.1.4a}$$

$$g_{\mathrm{Na}} = \bar{g}_{\mathrm{Na}} m^3 h. \tag{2.1.4b}$$

All three variables follow the differential equation

$$\frac{\partial x}{\partial t} = \alpha_x (1 - x) - \beta_x x \tag{2.1.5a}$$

$$= \frac{x_\infty - x}{\tau_x} \qquad \text{with} \quad x_\infty = \frac{\alpha_x}{\alpha_x + \beta_x}, \quad \tau_x = \frac{1}{\alpha_x + \beta_x} \tag{2.1.5b}$$

with nonlinear voltage-dependencies for $\alpha_{n,m,h}$ and $\beta_{n,m,h}$. The potassium variable n has a sigmoidal shape as function of the membrane potential u and, thus, the potassium conductance increases with u. The sodium variable m shows a similar behavior, while h has the opposite dependence and decreases with increasing voltage. Therefore, m is called activation variable and h inactivation variable.

The time course of the action potential arises due to the different timescales of sodium and potassium (cf. Fig. 2.3). At resting potential the sodium time constants τ_m and τ_h are much smaller than the potassium time constant τ_n. On depolarization of the membrane potential the sodium current rises more

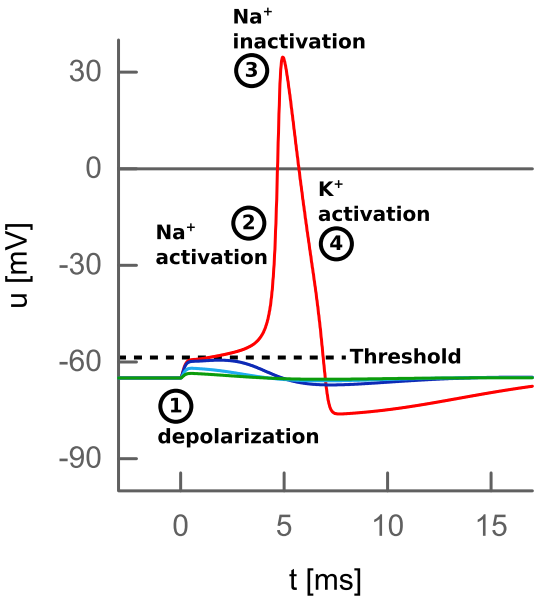

Figure 2.3
Time course of an action potential. If the depolarization from the resting potential crosses a threshold, the activation and inactivation of the sodium and potassium channels lead to a strong nonlinear voltage event.

quickly than the potassium current leading to a further depolarization ①. The potassium current can compensate for the delay and counterbalance the sodium current, if the depolarization remains below a certain threshold.

In the case that the membrane potential crosses this threshold, the positive feedback by the sodium activation leads to very strong uprise of the voltage close to the reversal potential of sodium ②. The sodium channels are now being inactivated ③ and the delayed strong activation of potassium leads to a quick repolarization ④. The shape of this nonlinear voltage spike lasts only $2-3$ ms and has a largely invariant shape. Action potentials are therefore regarded as unitary events with no information apart from their timing.

Action potentials are initiated at the axon initial segment where the density of active ion channels is largest. It is then propagated down the axon since the rise of the voltage spreads along the membrane and leads to an activation of the sodium channels in the vicinity. This kind of propagation along the axon would be relatively slow and require a lot of ions to be exchanged. So most of the axons are found to have a myelin sheath (cf. Fig. 2.1) which does not allow the flow of ions through this part of the membrane. This leads to a saltatory propagation where the action potential is only regenerated in the gaps between the myelin sheath, i.e. the Nodes of Ranvier. After the action potential has been generated, the ionic concentration gradients of sodium and potassium need to be restored by the ion pumps.

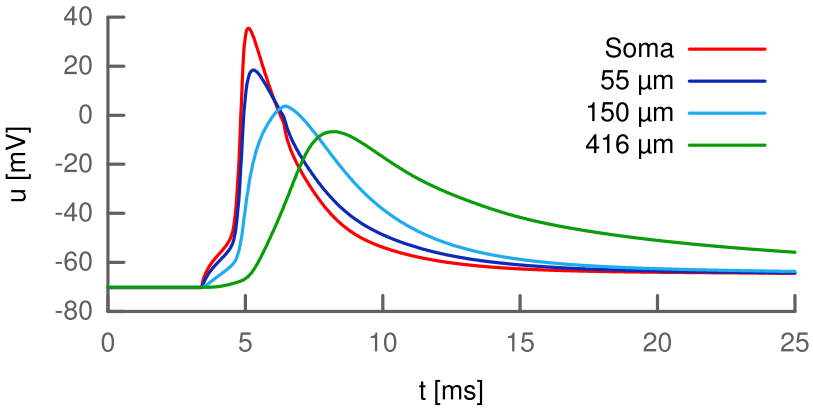

Figure 2.4 Simulated backpropagating action potential at different distances from the soma. As the action potential propagates through the dendritic tree it becomes smaller in amplitude and broader in time.

At the same time, the action potential will also propagate in the direction of the dendritic tree, which is called backpropagating action potential (bAP). The traveling distance and the strength of the bAP will depend on the density of active ion channels in the dendrites. This is a property which depends mainly on the type of neuron, but dendrites do not have such a high density of active ion channels as the axon. Thus, in general the bAP becomes smaller in amplitude and also broad in time as it travels through the dendritic tree (cf. Fig. 2.4). The bAP will, for example, influence the synaptic NMDA receptors (cf. next section).

Synaptic transmission

Every neuron has several thousand connections receiving input from other neurons. At these connections, the axon of another neuron makes a contact to the dendrite. This contact is called synapse (Fig. 2.1): the axonal side is called presynaptic, the dendritic side postsynaptic.

Apart from rare electrical synapses where the cells make a direct contact and which are mainly found in the retina and the cerebral cortex, the most abundant ones are chemical synapses. Their terminal endings are separated by a small synaptic cleft. The presynaptic terminal releases neurotransmitter into the cleft which diffuses to the postsynaptic side where it binds to neurotransmitter receptors incorporated in the postsynaptic membrane. While

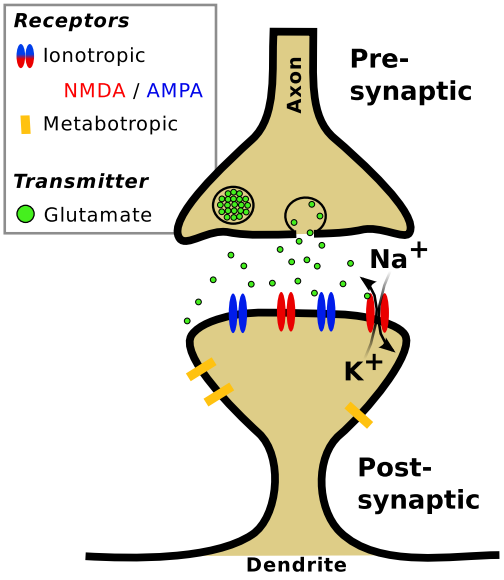

Figure 2.5
A chemical, glutamatergic synapse. The neurotransmitter released from the presynaptic terminal binds to the postsynaptic receptor which opens an ion channel.

there is an ongoing release happening spontaneously, a presynaptic action potential arriving at the synaptic terminal leads to an induced release. However, synapses do not show perfect reliability since the induced release can fail in a stochastic fashion.

While each receptor type is specific to a given transmitter, the receptors can be grouped into two broad categories: ionotropic and metabotropic. Ionotropic receptors form an ion channel pore which allows the flow of ions in or out of the cell, thus, having a direct effect on the postsynaptic membrane potential. In contrast, metabotropic receptors are coupled to intracellular proteins and initiate signal transduction mechanisms upon the binding of their agonist. In both groups, there are excitatory and inhibitory receptors. While excitatory receptors depolarize the postsynaptic cell leading to a more positive membrane potential, receptors with an inhibitory effect bring the membrane potential back to its resting value or even hyperpolarize the cell.

The most common ionotropic receptors are the glutamate-sensitive NMDA and AMPA receptors and the *gamma-aminobutyric acid* (GABA)-sensitive $GABA_A$ receptors. The glutamate receptors are named after their specific agonists *N-Methyl-D-aspartic acid* (NMDA) and *2-amino-3-(5-methyl-3-oxo-1,2-oxazol-4-yl)propanoic acid* (AMPA). The effect on the postsynaptic membrane potential depends on which ion species is admitted to flow through the receptor. NMDA and AMPA receptors are permeable for $Na^+$and $K^+$and NMDA receptors additionally permeable to Ca^{2+}. These channels have a reversal po-

tential of around 0 mV and, therefore, the NMDA and AMPA currents depolarize the postsynaptic cell and are excitatory. In contrast, $GABA_A$ receptors are only permeable to Cl^-, which has a reversal potential below the resting potential, and are inhibitory.

The channels are modeled as ionic conductances $g_{\text{syn}}(t)$. The current I_{syn} flowing through the channels is proportional to the difference between membrane and reversal potential:

$$I_{\text{syn}}(t) = g_{\text{syn}}(t)\,(u - E_{\text{syn}})\,. \tag{2.1.6a}$$

The time course of the conductance depends on the receptor type: AMPA receptors open and close on a very short timescale, usually a few milliseconds, while NMDA receptors open a bit slower but can remain open up to several hundred milliseconds. The time course is usually fitted by a sum of two (or three) exponentials. NMDA receptors exhibit an additional voltage-dependence, since they can be blocked by extracellular magnesium ions. This block is gradually released by a depolarization of the cell, which can be mediated by the bAP.

The maximal conductance determines the impact of this synapse on the postsynaptic cell. It is referred to as the synaptic strength or synaptic efficacy. The conductance depends on released amount of neurotransmitter, the total number of receptors N located in the postsynaptic membrane, and the individual conductances g_0 of each receptor. All quantities can change, thereby changing the synaptic strength. This process is called synaptic plasticity and it is a fundamental mechanism of learning in the brain. I will discuss it in more depth in Section 2.2.3.

The excitatory/inhibitory postsynaptic current (EPSC/IPSC) leads to depolarization/hyperpolarization which will be reversed by leak currents and ion pumps. The resulting deflection of the voltage depends on the membrane capacitance and conductances and it is called excitatory/inhibitory postsynaptic potential (EPSP/IPSP).

The overall dynamics of the membrane potential depend on the ionic current through the voltage-dependent channels and through the synaptic ligand-gated ion pores:

$$C_m \frac{\partial u}{\partial t} = \sum_{\text{ion}} I_{\text{ion}} + I_{\text{syn}}. \tag{2.1.7a}$$

2.1.2 Simple spiking and rate-based models

The conductance-based model of a neuron, introduced in the previous section, is very detailed and accurate. It allows to understand and predict many properties of single neurons. But it is difficult to analyze and simulate due to the nonlinear differential equations and spatial extent of a real neuron. In the case of studying a network of neurons and its computational properties, it is often desirable to employ a simpler neuron model. The most common simplification is to neglect the complex three-dimensional structure of the cell and assume a point-like neuron. All inputs, thus, arrive directly at the soma and all spatial aspects of the neuronal dynamics are not considered.

Simple spiking models like the leaky-integrate-and-fire (LIF) model drop the complicated nonlinear dynamics of the Hodgkin-Huxley mechanism for spike generation. Only the subthreshold dynamics remains and the spike time is just defined by the time of threshold crossing combined with a reset of the membrane potential. For the LIF neuron the subthreshold membrane potential is the leaky integration of the input current I:

$$C_m \frac{\partial}{\partial t} u = -g_{\text{leak}} \, (u - u_R) + I$$

with the resting potential u_R.

In the rate model of neurons the spike time is entirely disregarded. That means the output signal of a neuron is described by the number of spikes in a fixed time interval, i.e. its firing rate or activity r. This rate is calculated by two stages: a linear filter determines how the membrane potential u integrates the inputs. The output of this linear filter is then passed through a nonlinear transformation to yield the rate.

The rate model can be employed either in discrete time or continuous time. In discrete time the whole dynamics of the membrane potential is ignored. The output y of a neuron is only defined at discrete points in time $t \in \mathbb{N}$. y_t corresponds to the rate r, that means it is the average number of spikes in the time interval $[t-1, t]$. Thus, the input x_t from one to another neuron is just y_t and the total input is a weighted sum:

$$I_t = \sum_j w_j x_t^{(j)} = w x_t.$$

Since this input is an average and thus constant in the time interval, the membrane potential is assumed to arrive at the equilibrium value $u_t = u_R + \frac{\tau}{C} I_t$. The depolarization $\hat{u} = u_t - u_R$ is proportional to the input I_t. Absorbing the proportionality constant into the nonlinear transformation f, the output/rate is calculated as

$$y_t = r = f\left(w x_t\right). \tag{2.1.8a}$$

A popular choice for the transformation, also known as the activation function, is the logistic sigmoid function

$$f(u) = \frac{1}{1 + \exp\left[-a\left(u - b\right)\right]}. \tag{2.1.9a}$$

Explicit spiking can be recovered in the linear-nonlinear rate model by going to continuous time. Here, the rate is not an average over a finite interval but interpreted as an instantaneous rate, i.e. the probability of spiking in an infinitesimal interval. Given this instantaneous rate, the spike times are drawn from a stochastic process. The usual choice is a Poisson process which leads to the LNP (linear-nonlinear-Poisson) model.

2.2 Neural plasticity

The computation in a neural network is determined by the connectivity between the neurons and their individual excitabilities. The fundamental ability of the brain is to adapt the computations by modifying both, connectivity and excitability. These modifications are called plasticity and are the basis for learning and adaptation [Hebb, 1949; Kandel, 1997; Abbott and Nelson, 2000].

The changes in the excitability are called intrinsic plasticity. Structural plasticity describes the morphological changes and the creation/removal of synaptic connections, while the changes in synaptic strength of established connections are called synaptic plasticity. Synaptic plasticity is the main focus of this thesis, to be discussed in Section 3.2.1 and Chapter 4, but Section 3.2.2 will also address intrinsic plasticity.

2.2.1 Homeostatic plasticity

An overarching concept is the so-called homeostatic plasticity. The function of neural homeostasis is to maintain the stability of neural function during development as well as in adulthood [Turrigiano and Nelson, 2004]. This is done by tuning the processing of a neuron towards a specific set point. If the environment changes (e.g. stronger/weaker input) the balances have to be restored by modifying synaptic efficacies and intrinsic excitability. The homeostatic mechanisms influence and regulate the relevant plasticity mechanisms to achieve this balance.

2.2.2 Intrinsic plasticity

Intrinsic plasticity refers to changes in the input-output relationship of a neuron. This sets and adapts the working regime of the neuron, which can include e.g. spiking threshold and excitability [Sjöström et al., 2008]. The mechanistic description of this process depends on the employed neuron model. For simple rate model neurons, the input-output relationship is described by the nonlinear activation function relating input current to output firing rate. In spiking models one main parameter for the intrinsic plasticity is the spiking threshold, e.g. in LIF neuron models. For more detailed Hodgkin-Huxley-type models, the input-output relationship is described by the $f - I$ curve relating input current and output firing frequency. It depends on the dynamics of the voltage-gated ion channels generating the action potential.

By adjusting these intrinsic parameters, a neuron is able to keep the firing rate in a homeostatic range. On average, its activity is not too high or too low. This idea is connected to approaches from information theory. A firing rate distribution which is restricted to only high or low firing rates does not use its full range and can not efficiently transmit information. This is reflected in the theoretical approaches towards intrinsic plasticity. Several models have proposed that a neuron should adapt its input-output relationship in order to maximize their mutual information [Bell and Sejnowski, 1995; Stemmler and Koch, 1999; Triesch, 2007]. That means the entropy of the distribution of output firing rates should be maximal for the current distribution of the input.

Given that a neuron has some maximum firing rate $f_{\max}$ the distribution

with maximal entropy on the interval $f \in [0, f_{\max}]$ is the uniform distribution. The learning rule by Bell and Sejnowski [1995] adapts the parameters of the activation function such that the output firing rate is uniformly distributed for the current input distribution. On the other hand, a neuron has limited amount of metabolic resources. That means, the number of spikes it can emit within some fixed time period is limited and the neuron is, therefore, restricted to a maximal average firing rate as described in Section 2.3.1. Incorporating this fact, Triesch [2007] modified the learning rule by demanding a fixed average firing rate instead of a maximum rate. The maximum entropy output distribution under this constraints is the exponential distribution. This is similar to the experimentally observed firing rate distributions [Hromádka et al., 2008].

In a more detailed model, Stemmler and Koch [1999] applied the idea of maximizing mutual information to a spiking Hodgkin-Huxley neuron. They adapted the gating kinetics of the voltage-dependent ion channels. This lead to an $f - I$ curve which fitted the statistics of the input current.

2.2.3 Synaptic plasticity

Synaptic plasticity refers to ongoing changes in strength of established synaptic connections [Sjöström et al., 2008]. The phenomenon is observed on two different timescales. Short-term synaptic plasticity fluctuates on the order of seconds to minutes and is restricted to the presynaptic side. It is connected with the depletion of neurotransmitter storages and changes in the release probability of vesicles [Fioravante and Regehr, 2011].

Long-term synaptic plasticity describes changes which are stable and persist over hours or longer. The change in synaptic strength following a given stimulation protocol is measured as the increase or decrease of the EPSP amplitude. These changes are mainly associated with a change in the number of receptors in the postsynaptic membrane as well as changes in the maximal conductance of each single receptor [Song and Huganir, 2002]. But also presynaptic mechanisms have been found to influence the long-term changes in synaptic strength [Rodríguez-Moreno et al., 2011].

During the last fifty years there have been many advances in understanding the dependence of long-term plasticity on the pre- and postsynaptic conditions.

Initially pioneered by Hebb [1949] in theoretical terms, the concept of long-term potentiation (LTP) as a result of correlated activity has been refined and extended. Since then a rich body of experimental findings and theoretical models for synaptic plasticity under different protocols has accumulated.

Hebbian plasticity and STDP

One of the most quoted hypotheses in neuroscience is Hebb's postulate which basically states that the synaptic connection between two neurons should be strengthened if the presynaptic cell "repeatedly or persistently takes part in firing" the postsynaptic cell [Hebb, 1949]. With this, he established a theoretical foundation for long-term synaptic plasticity. The inherent timing or causal relation in Hebb's postulate, however, is not present in the usual formulations of Hebbian plasticity. The postulate is often simplified such that those neurons which "fire together, wire together", thereby neglecting the causal relationship. The resulting Hebbian learning rule is only sensitive to correlations of pre- and postsynaptic activity x and y, respectively. The change in the synaptic weight is defined as

$$\Delta w := \eta x y, \tag{2.2.1a}$$

where $\eta > 0$ is a learning rate.

More than twenty years later, Bliss and Lomo [1973] provided the experimental confirmation of long-term potentiation (LTP) induced by correlated pre- and postsynaptic activity. They showed that a synaptic connection got stronger if both the pre- and postsynaptic neurons concurrently showed a high firing rate, and that this increased strength persisted over several hours even in the absence of ongoing correlated activity. The opposite mechanism of long-term depression (LTD) was discovered by Lynch et al. [1977] in a different brain structure. It is now established that both processes, LTP and LTD, can happen at the same synapse meaning they show bidirectional plasticity.

Different bidirectional extensions to the LTP-only standard Hebbian learning rule were proposed which contained an explicit regime for depression. The covariance rule [Sejnowski, 1977]

$$\Delta w := \eta \left(x - \langle x \rangle\right) \left(y - \langle y \rangle\right) \tag{2.2.2a}$$

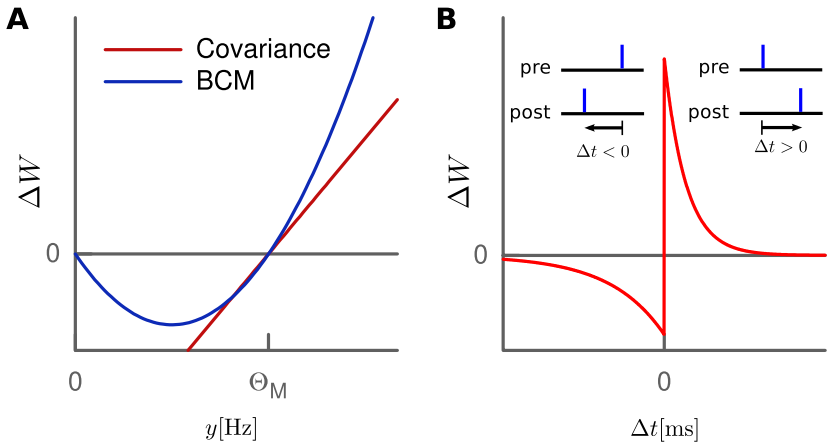

Figure 2.6 Bidirectional synaptic plasticity. **A**: The covariance rule (red) and the BCM rule (blue) predict LTD for low postsynaptic activity y and LTP for high activity. **B**: For STDP the amount and sign of the weight change depends on the relative timing between pre- and postsynaptic spike: LTP for pre-post and LTD for post-pre.

kept the linear dependence, while Bienenstock, Cooper, and Munro [1982] (BCM) proposed a theory with a quadratic dependence on the postsynaptic activity (Fig. 2.6A) as

$$\Delta w := \eta y \left(y - \Theta_M\right) x. \tag{2.2.3a}$$

Dudek and Bear [1992] showed that a presynaptic high-frequency stimulation, leading to a strong activity at the postsynaptic side, resulted in LTP. A low-frequency stimulation, insufficient to excite the postsynaptic cell, leads to slight LTD or no plasticity as predicted by the BCM theory. They also found a smooth crossover from the LTD to the LTP regime with some threshold frequency producing no change at all. This rate-dependent plasticity more closely resembled the causal nature of the full Hebbian postulate.

A similar U-shaped dependence of synaptic plasticity on the postsynaptic voltage was found in experiments [Artola et al., 1990; Ngezahayo et al., 2000]. They showed that postsynaptic firing was not necessary for synaptic plasticity. Rather, pairing presynaptic stimulation with postsynaptic depolarization was sufficient. Similar to the rate-dependent plasticity, a small depolarization leads to LTD while a larger one leads to LTP, also with a smooth crossover and a threshold voltage.

The discovery of spike timing-dependent plasticity (STDP) [Gerstner et al., 1996; Markram, 1997] finally established the importance of the relative timing between pre- and postsynaptic spikes (Fig. 2.6B). In line with the Hebbian postulate, causal pairs of spikes, with the pre- coming before the postsynaptic spike, lead to LTP ($\Delta t > 0$). The reversed order of acausal pairs ($\Delta t < 0$) results in LTD. The plasticity is strongest when the relative delay Δt is small, while there is rapid crossover between the LTP and the LTD regime. This is usually modeled with two exponentials [van Rossum et al., 2000]:

$$\Delta w := \eta \begin{cases} +A_+ \exp\left[-\dfrac{\Delta t}{\tau_+}\right] & : \Delta t > 0 \\ -A_- \exp\left[+\dfrac{\Delta t}{\tau_-}\right] & : \Delta t < 0 \end{cases} \tag{2.2.4}$$

Due to this causality, STDP was initially assumed to be a fundamental property of synaptic plasticity. More recently, however, experiments have described non-linear interactions between spike triplets and very different forms of STDP windows with depression only, potentiation only or even reversed timing requirements (see [Shouval, 2010] for a review).

Metaplasticity

Another important feature of the BCM theory is the assumption that the threshold Θ_M should be modifiable. In line with a homeostatic regulation, the threshold was set to a long-term average of the postsynaptic activity. A high activity would increase the threshold and make the induction of LTD more likely, thereby reducing the overall input and postsynaptic firing.

This hypothesis was experimentally tested by Kirkwood and Rioult [1996]. They raised kittens in a darkened environment and thereby artificially reduced the excitation to cells in the visual cortex. As a result, the threshold for the induction of LTP was shifted to lower frequencies compared to kittens under normal conditions. Complementary, Wang and Wagner [1999] primed the postsynaptic cell with a high activity stimulation and observed that the threshold shifted to higher frequencies. Thus, changing input statistics is counterbalanced by changing the regimes of LTP and LTD. While input deprivation makes LTP more likely, a strong activity of the postsynaptic cell is balanced by making LTD more likely. This homeostatic effect, that depends on condi-

tions in the recent past, is called metaplasticity due to the fact that it is a plasticity of synaptic plasticity [Abraham, 2008].

A similar homeostatic effect has been found by Ngezahayo et al. [2000] in voltage-clamp experiments. They measured the dependence of synaptic plasticity on the postsynaptic membrane potential similar to Artola et al. [1990]. Additionally, they found the voltage threshold, separating the LTD from the LTP regime to depend on the strength of the probed synaptic pathway. This metaplasticity again showed a homeostatic behavior since a strong synapse had a larger LTD regime and a previously weakened synapse a larger LTP regime.

Underlying mechanisms

The classical plasticity and metaplasticity protocols are used to probe the dependence of synaptic changes on different factors like firing rate, spike timing and membrane potential. But those dependencies are no independent phenomena but just different facets of one common underlying mechanism. The molecular processes which lead to the observed change in synaptic efficacies remained unclear for long time and are still not fully understood.

The idea that the intracellular calcium concentration is a key component in these processes was pioneered by Lisman [1988; 1989]. The important quantity is a molecular complex called calcium/calmodulin-dependent protein kinase II (CaMKII). It consists of four subunits which can be phosphorylated. In the initial state (off-state) this process is calcium-dependent, but once three subunits are phosphorylated, a calcium-independent autophosphorylation can keep the complex in a stable, fully phosphorylated state (on-state). With this, CaMKII can act as a bistable switch storing information about the synaptic strength (cf. Fig. 2.7).

The stability depends on the level of calcium as shown in Figure 2.7B. An intermediate level promotes an activation of Phosphatase I (PPI) which in turn dephosporylates CaMKII bringing it to the off-state. A high level of calcium triggers the switch from the off- to the on-state and also inactivates PPI through an interaction with Inhibitor I. If the level of calcium is too low, it does not trigger any reactions and CaMKII stays in its current stable state.

Ultimately, CaMKII in the on-state influences the synaptic strength by phosphorylating existing AMPA receptors (increasing their conductance) and

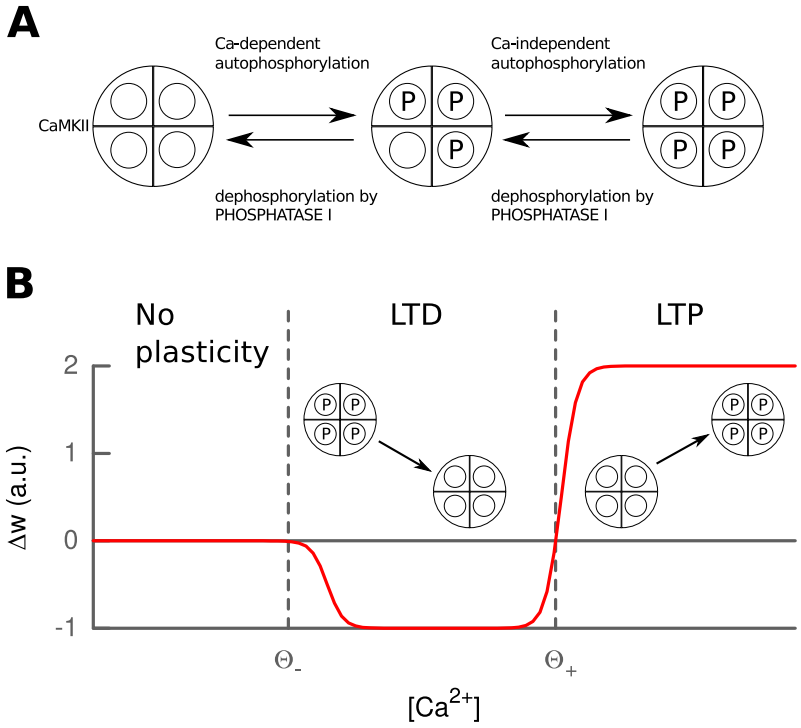

Figure 2.7 Autophosphorylation of CaMKII and its effect on synaptic plasticity. Upper: Phosphorylation of the first three CaMKII subunits is calcium-dependent, while the fourth phosphorylation happens independently of calcium. Lower: The calcium concentration determines the amount of desphosphorylation. Subsequently, this affects the induction of plasticity.

promoting the insertion of new receptors in the membrane. Calcium is a good candidate for sensing correlation between pre- and postsynaptic activity, since the calcium level is affected by presynaptic input via the postsynaptic NMDA receptors and postsynaptic depolarization via voltage-dependent calcium channels. This leads to the formulation of the so-called calcium control hypothesis for synaptic plasticity. It describes the process of LTD and LTP by two thresholds for the calcium concentration. If it is below Θ_-, no plasticity is induced, between Θ_- and Θ_+ LTD occurs, and above Θ_+, the synapse gets potentiated (cf. Fig. 2.7B). This form is seen again in the voltage-dependent [Artola and Singer, 1993] and the rate-dependent plasticity [Bienenstock et al., 1982]. It has also been shown to explain STDP [Shouval, 2010].

Shortcomings of previous models

The large body of theoretical models can be roughly grouped into three categories: mechanistic, phenomenological, and functional. Every category has its benefits and drawbacks.

Phenomenological studies start directly with the observed plasticity phenomena and their dependencies. These are described by simplified models useful to study the impact of different plasticity phenomena on network behavior, synaptic weight stability, and learning [van Rossum et al., 2000; Pfister and Gerstner, 2006; Clopath and Gerstner, 2010]. But they make no (or only abstract) reference to the underlying mechanisms and are limited to the dependencies explicitly put into the model. Further, they are not capable of assessing functional goals of synaptic plasticity.

The mechanistic models are bottom-up approaches. Based on known biophysical mechanisms in the cell, they try to explain the resulting phenomena of synaptic plasticity [Shouval et al., 2002; Graupner and Brunel, 2012]. The most promising theories are based on the calcium control hypothesis as described above. This type of approach is able to create very realistic models. On the other hand, they are difficult to analyze analytically in order to predict the behavior on the higher level. That means, it is difficult to extract possible functional goals starting from these basic mechanisms.

Functional approaches are top-down. Starting from a computational perspective, they derive an algorithm that can describe the observed phenomena. In this view these phenomena are just a consequence of achieving the functional goal and the biological mechanisms correspond to the implementation of the proposed algorithm. The difficulty lies in choosing a reasonable goal from the large set of possibilities. It needs to reproduce most of the observed dependencies of synaptic plasticity and ultimately lead to an implementation in accordance with the biological mechanisms. This is where all previous models fall short [Toyoizumi et al., 2005; Sprekeler et al., 2007; Pool and Mato, 2011].

To get a unified understanding of synaptic plasticity, a theory has to bridge all levels of abstractions. While phenomenological ones are useful, they are not capable of providing this unification. It can either be done by working upwards from the basic mechanisms or downwards from some functional goal. Since both approaches are difficult and error-prone, a full theory providing

a unified understanding probably only arises in a fruitful interaction between these two. In Chapter 4, this thesis tries to take a first step from the functional side all the way down to the biological mechanisms.

2.3 Neural computation: constraints and function

Neurons and their connections are the underlying basis of the computations in the brain. Apart from some modulatory signals, each cell is largely independent in terms of its elementary function. It is only influenced by other cells through its synaptic connections embedding it in a neural network. The overall computation of such a network of neurons is the result of the interaction between two processes: input-output mapping and connectivity. The goal of learning is to optimize the computation by adapting both processes. To analyze the arising computation and its plasticity mechanisms it is important to understand the constraints of the neural system.

2.3.1 Constraints

The neural system needs to perform its computations under various constraints and in order to work in an optimal fashion, these constraints should be considered by the brain.

Probably the most fundamental constraint is the limited amount of energy for the brain [Laughlin and Sejnowski, 2003]. Like every cell in the body, a neuron needs nutritions, oxygen, and energy for the vegetative metabolism of its cell body. But their ability to receive and transmit excitation incurs additional energy consumption. Neurons need to develop and maintain their extensive dendritic and axonal trees in order to make synaptic contacts with other neurons. The larger and longer these extensions are, the more energy they require. One reason behind the arborescent structure of dendrites may thus be the need to minimize the wiring length from the soma to the synaptic contacts [Cuntz et al., 2010].

Also establishing and maintaining the resting potential is a process that induces high metabolic costs. It requires the ongoing activity of the Na^+/K^+-

ATPase pump. The costs for neural processing and communication have been estimated to make up half of the energy consumption in human cortex [Lennie, 2003].

For the maintenance of the resting potential three regimes can be distinguished. Even in the absence of synaptic input the Na^+/K^+-ATPase pump is active to maintain the resting potential, since the membrane is not ideal and shows a permeability to ions other than potassium. Additional activity of the ion pump is required to counterbalance the ionic currents from synaptic input in the subthreshold regime. Finally, a suprathreshold input leads to an action potential which requires further pumping activity. The baseline activity is due to a static property of the membrane and can not be optimized. The metabolic costs of the sub- and suprathreshold regimes, however, depend on quantities which can be adapted. This can be used as a guiding principle to discover functional principles of the neural system.

The metabolic costs in the suprathreshold regime depend on the output of a neuron, i.e. the shape of the action potential and the distribution of firing rates. Hasenstaub et al. [2010] proposed that the shape of the action potential is optimized to be energy-efficient. They have shown that a short, thin spike requires more energy, but allows for a higher maximal frequency, leading to a trade-off. Depending on their type, neurons are optimized in their action potential shape according to these requirements.

The most widespread idea regards energy efficient coding, i.e. how can the neural system represent information in an optimal fashion. From an information theoretic viewpoint, the information capacity C corresponds to the entropy of the firing probability p in some short time interval. The average energy expenditure follows roughly as

$$E = (1-p)\,E_{\text{rest}} + pE_{\text{AP}} \tag{2.3.1a}$$

$$\propto 1 + p\,(r-1) \tag{2.3.1b}$$

where $r = E_{\text{AP}}/E_{\text{rest}}$ is the energy ratio of firing compared to resting.

A neuron should transmit as much information while spending as little energy as possible. This amounts to maximizing the ratio $\frac{C}{E}$. For an inexpensive action potential, i.e. $r \approx 1$, the energy would be independent of p. The optimal value p^* is then one half, which also maximizes the entropy. However,

given the high metabolic costs of action potentials, r is generally large and the required energy increases strongly with p. The observed low average firing frequency in real neurons, thus, optimizes the information per energy [Levy and Baxter, 1996]. These considerations are connected to the idea of a sparse neural code, where the rare, expensive firing is compensated by representing the input with a large population of neurons. I will describe the ideas and formulations for sparseness in the following section. It is, however, important to realize that not only the generation of an action potential consumes energy.

The metabolic costs in the subthreshold regime include the synaptic transmission and the postsynaptic effects of a received action potential. Both require much energy and, actually, the EPSPs make up more than fifty percent of the total energy per action potential [Lennie, 2003]. The distribution of synaptic inputs and their total strengths determine the required energy. Thus, these quantities are an important point for optimizing the energy efficiency of the neuron. Levy and Baxter [2002] already found that a nonzero probability of synaptic failure is optimal with respect to energy-efficient information transmission given the limited information capacity of the axon. How these postsynaptic costs should impact the synaptic strength and its distribution has, to the best of my knowledge, so far not been considered. In Section 4 I will show that an optimization in terms of sparseness applied to the synaptic efficacies of a neuron introduces a new functional goal for synaptic plasticity which unifies different observations as well as different levels of analysis.

2.3.2 Sparseness

For random variables, sparseness means that the variable has very small values (close to zero) most of the time and only rarely takes on large non-zero values. This is not an absolute statement and needs a baseline for comparison. Usually the distribution of the random variable X is compared to a Gaussian with the same variance. If the probability density function (PDF) of X has more probability mass around zero and in the tails (i.e. it is more peaked) than the Gaussian, it is considered to be sparse (cf. Fig. 2.8). These distributions are also called super-Gaussian.

For the distribution of neural firing rates, sparseness has two distinct aspects. First, the firing rates of each neuron can have a sparse distribution,

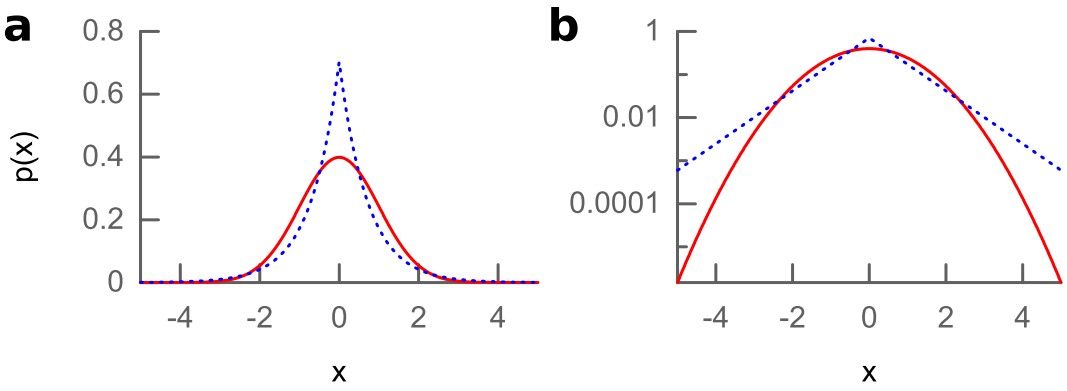

Figure 2.8 Comparison of Gaussian (solid red) and Laplacian (dotted blue) PDF in **(a)** normal and **(b)** semi-log space. The Laplacian has a large peak at zero and heavier tails and is called super-Gaussian.

which is called "lifetime sparseness". The second aspect concerns the distribution of the response from a population of neurons. Here, the input should have a sparse representation such that only few neurons are strongly active, which is called "population sparseness".

Sparse coding

A lifetime-sparse distribution of firing rates accounts for the fact that spikes are metabolically expensive. The distribution should be optimized to convey as much information while being restricted to an average amount of energy consumption. Since entropy is a measure for the information capacity, the distribution should have the maximum entropy given specific constraints. Such a distribution can be written as the Gibbs distribution

$$p(x) = \frac{1}{Z\left(\lambda_1, .., \lambda_k\right)} \exp\left[\sum_i^k \lambda_i f_i(x)\right] \tag{2.3.2a}$$

where the f_i encode the constraints in terms of expectation values such that $\mathbb{E}\left[f_i(x)\right] = \alpha_i$ and the λ_i are the Lagrange multipliers to be determined by the constraints.

The required energy depends mainly linearly on the firing rate r plus some additional baseline energy ρ for the resting potential and subthreshold fluctuations [Laughlin and Sejnowski, 2003]:

$$E_r = a \cdot r + \rho \tag{2.3.3a}$$

On average every cell has a fixed amount of energy μ_E to supply the baseline and the spiking. The constraint on the average energy translates to an average firing rate $\mu_r = \mathbb{E}[r] = \frac{\mu_E - \rho}{a}$. For a fixed mean and a domain of $r \in [0, \infty[$ the maximum entropy distribution is the exponential distribution. With the additional constraint of a maximal firing rate $r_{\max}$, the distribution is a truncated exponential. In this case it is still a decaying exponential if $\mu_r \ll r_{\max}/2$ which is granted in neurons where $\mu_r < 10$ Hz and $r_{\max}$ is on the order of 100 Hz.

Experimental firing rate distributions resemble exponential distributions but are much better fitted with the lognormal distribution [Hromádka et al., 2008] with the PDF as

$$\ln \mathcal{N}_{\mu,\sigma}(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]. \tag{2.3.4a}$$

Both distributions are considered to be sparse. Interestingly, the lognormal distribution is the maximum entropy distribution given a fixed mean and variance of $\log X$. The deviation from the exponential distribution at very low firing rates can be explained by taking noise into account. While low firing rates should be preferred in neural coding due to their low energy requirements, they are more prone to the influence of noise [Tsubo et al., 2012].

The population sparseness depends on the energy ratio of spiking to resting and the required representational capacity of the population. For inexpensive spiking comparable to resting, on average half of the neurons should be active, since this increases the capacity due to the combinatorics and the population size can be small [Laughlin, 2001]. However, given the high costs of spiking, it is more efficient to employ a large number of neurons but let only few be active.

In their seminal work, Olshausen and Field [1996] applied the idea of sparseness to visual input as arriving in the primary visual cortex. Here, neurons are described by linear basis functions $\phi_i(x, y)$ coding for a given pattern in the input $I(x, y)$. In biological terms, this basis function is analogous to the receptive field and describes the connectivity from the input stage to the individual neuron. The aim is to reconstruct the input $I(x, y)$ from the neural

responses a_i by linear superposition of the basis functions:

$$I(x,y) \approx \hat{I}(x,y) = \sum_i a_i \phi_i(x,y). \tag{2.3.5a}$$

Just minimizing the squared reconstruction error

$$E_{\text{input}} = \sum_{x,y} \left| I(x,y) - \hat{I}(x,y) \right|^2 \tag{2.3.6a}$$

is a simple optimization problem. Olshausen and Field [1996] also included a sparseness energy term penalizing high responses:

$$E_{\text{sparse}} = \sum_i S(a_i), \tag{2.3.7a}$$

where they employed different functional forms of the penalty function S like $|a_i|$ or $-\exp[-a_i^2]$. Minimizing the total error $E = E_{\text{input}} + \lambda \cdot E_{\text{sparse}}$ amounts to finding a representation approximating the input with as few active neurons as possible.

The activities which minimize the total energy and the actual minimal value strongly depend on the basis functions (or receptive fields) of the neurons. The important step of Olshausen and Field [1996] was to further optimize the total energy by adapting the basis functions to better represent the input. After convergence, the resulting basis functions resembled localized, oriented edge-filters similar to the receptive fields found in primary visual cortex [Hubel and Wiesel, 1968].

Sparse synaptic efficacies

The distribution of synaptic efficacies is well fitted by a lognormal distribution [Song et al., 2005; Loewenstein et al., 2011]. Their distribution is highly skewed and sparse meaning that most of the synapses are rather weak but a few ones are an order of magnitude stronger than the mean. Song et al. [2005] found that the strongest 17% of the synapses contributed 50% of the total synaptic strength.

The lognormal distribution also fits well with neural firing rates as described above. This agrees with the sparse coding model and has been connected to the energy efficiency of the suprathreshold regime. Similarly, the sparse distribution of the synaptic efficacies can be connected to an energy-efficient

subthreshold behavior. In Chapter 4, I describe how the synaptic efficacies influence the metabolic costs and apply this functional idea to optimize their distribution.

Independent component analysis and sparseness measures

The classical sparse coding approach with its energy formulation punishes high neural activities with a certain sparseness function. The choice between different functions can be quite arbitrary and it is not clear how to define and quantify sparseness of a single neuron. Furthermore, given a population of neurons exhibiting a distribution of activities, how sparse is this distribution? As initially introduced, a distribution is considered sparse if it is stronger peaked at zero and has heavier tails compared to a Gaussian with the same variance.

This comparison to a Gaussian is not arbitrary and resembles the connection of sparse coding to the concept of independent component analysis (ICA) [Comon, 1994; Hornillo-Mellado et al., 2005]. ICA aims to find the independent components by searching for non-Gaussian projections of the signal. Bell and Sejnowski [1997] applied ICA to natural images and found not only similar receptive field structures (Gabor-like edge filters) but also that these independent components were sparsely distributed.

The deviation of a distribution from a Gaussian can be quantified with higher-order moments. Popular choices are the third and fourth normalized moment called skewness S and kurtosis K, respectively [Hyvärinen and Oja, 2000; Blais et al., 1998]. They are defined as the n-th central moment divided by the n-th power of the standard deviation σ:

$$S = \frac{\langle x - \bar{x} \rangle^3}{\sigma^3} \tag{2.3.8a}$$

$$K = \frac{\langle x - \bar{x} \rangle^4}{\sigma^4} \tag{2.3.8b}$$

While this does not imply that skewness and kurtosis are equivalent to sparseness, these measures are generally useful to quantify and optimize the sparseness of a distribution.

Chapter 3

Separability objective for neural plasticity

The most remarkable functionality of the brain is its plasticity which provides the organism with a mechanism to adapt and learn. To understand the capability of neural networks to learn is a fundamental problem which concerns computational neuroscience, machine learning as well as engineering applications. The self-organized learning rules employed in many approaches are inspired by the experimental findings in real neurons. These rules for simplified artificial neural networks, however, are not more than a phenomenological ad-hoc description of the observed processes. This chapter will introduce a principled derivation of plasticity rules for artificial neural networks based on an objective function.

Section 3.1 gives a short, general introduction to the different topologies of artificial neural networks and the special approach of reservoir computing which is a class of recurrent networks which are efficient to train. Section 3.1.2 reviews a recent finding on self-organized learning rules applied to such recurrent networks and how those rules can improve the internal representation. In Section 3.2, I describe the main part of the objective function in terms of separability of the internal network state [Krieg et al., 2010], which was motivated by the work introduced in Section 3.1.2. I will derive synaptic and intrinsic plasticity rules from the objective function, analyze their effects and stability properties, and compare them to previously proposed plasticity rules. Section 3.3 extends the approach with an objective regarding energy consump-

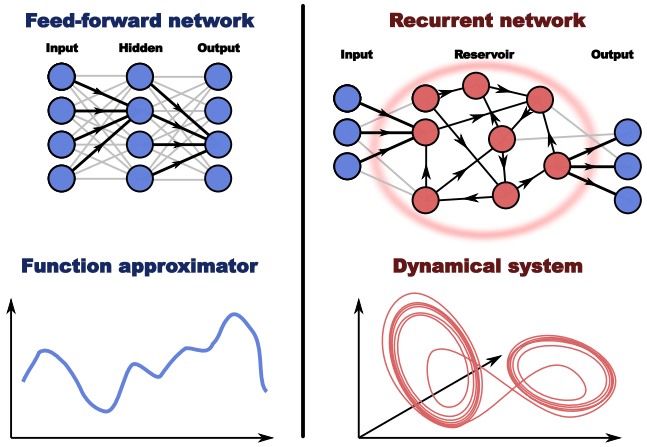

Figure 3.1
Two possible topologies of an artificial neural network. The feed-forward network can be a universal function approximator, while the recurrent network resembles a dynamical system.

tion and noise robustness. Section 3.4 demonstrates the derived learning rules applied to a recurrent network and shows the performance and convergence properties.

3.1 Artificial neural networks

The first artificial neural networks have been concerned with feed-forward structures having one or more layers of neurons. Each layer consisted of a population of neurons receiving input from a previous layer. The goal was to transform the initial input at the first layer into a suitable output at the last layer (cf. Fig. 3.1). This was achieved by adapting the connection weights in the feed-forward network. In a supervised learning paradigm the input as well as the desired output were given to the learning algorithm. The error at the output was then backpropagated through the layers providing a measure for how to change the connections. With such a learning scheme, the network could be tuned to provide a desired input-output mapping. In fact, purely feed-forward networks were shown to be universal function approximators.

More biologically realistic are networks with recurrent connections. They allow the information about previous times to be retained in the network. A recurrent neural networks (RNN) is basically a dynamical system (cf. Fig. 3.1) with the possibility of fixed points and complex attractors, but they also have their disadvantages: they can exhibit chaotic behavior which makes them complicated to predict. Thus, RNNs are hard to train and optimize with a supervised learning approach, since the backpropagated errors also enter the

recurrent loops [Werbos, 1990; Jaeger, 2002a].

3.1.1 Reservoir computing

The problem of the loops for supervised training can be circumvented by the reservoir computing approach [Jaeger et al., 2007]. While there are different variants for the structure of the network known as Echo State Network [Jaeger, 2001] and Liquid State Machine [Maass et al., 2002], it contains a recurrently connected network called the reservoir which receives the input and projects to an output layer (Fig. 3.1).

The reservoir usually contains a large number of neurons. It transforms the input in a high-dimensional, nonlinear, time-dependent space. The idea of this so-called kernel trick is that in such a space the desired information or dynamics for the output can be extracted from a linear subspace. The reservoir can be seen as providing a sufficiently large set of nonlinear transformations. The output layer linearly combines the results of these transformations to achieve the desired output.

In standard reservoir computing, the input and recurrent reservoir connections are taken as random but fixed, while the connections from the reservoir to the output are adapted via supervised learning [Lukosevicius and Jaeger, 2009]. This approach allows to combine the advantages of a recurrent network with supervised learning: a dynamical system can be used to represent the data, while still being easy to train. It can be applied, for instance, to systems classification and time-series prediction [Jaeger, 2002b; Jaeger and Haas, 2004].

3.1.2 Self-organization in the reservoir

While the training of the output connections in a reservoir computing approach is straightforward, the initialization of the recurrent reservoir needs to be tuned to the problem at hand. This usually requires some ad-hoc heuristics. Furthermore, the network is not flexible enough to adapt to changing input statistics. The reservoir should be able to find suitable representations for the input by itself. This requires local unsupervised learning rules which can lead to a self-organization of the reservoir.

Lazar et al. [2009] proposed such a self-organizing recurrent network using a combination of three plasticity rules in the reservoir: STDP, synaptic scaling, and intrinsic plasticity. All three are local learning rules inspired by biologically observed ones. They employed a simple counting task which requires the reservoir to retain information about previous inputs. The input pattern consisted of a sequence of three symbols $A, B, .., B, C$ or $C, D, .., D, E$ with n repetitions of B and D. Each symbol targeted a different subset of the reservoir. The task of the output layer was to predict the next input symbol and the connections from the reservoir to the output were trained by supervised learning.

For the output to be able to perform its task with a good performance, the reservoir needs to have a discernible representation between the k-th and the $(k+1)$-th B or D. Thus, the performance depends strongly on the dynamics of the reservoir. If upon repetition of B or D the reservoir state approaches a fixed point in its high-dimensional space, the representations will become more and more similar and the output will not be able to distinguish them above a given number of repetitions. If, on the other hand, the reservoir enters a sufficiently large limit cycle upon the repetition of input, the supervised training of the output layer can find a suitable projection to the output layer.

Lazar et al. [2009] showed that their self-organizing recurrent network outperforms a network with a static reservoir. The improved performance is due to a clear separation of the reservoir states for different input repetition numbers which the authors demonstrate in PCA (principal component analysis) space. That means, the reservoir state for the fourth B is easily separable from the reservoir state representing the fifth B, while in a static reservoir the states do not form discernible clusters. This is due to the combination of the different learning rules which alter the recurrent reservoir connectivity. Due to its 'causal' structure, STDP allows the dynamics to adapt to the sequence structure of the input, though STDP alone is not sufficient. The interaction with the homeostatic mechanism of intrinsic plasticity and the weight stabilization due to synaptic scaling are necessary for a good performance of the network.

3.2 Objective function: Separability

In order to create a grounded connection between the plasticity mechanisms and the increased separability, I will show that an IP as well as an STDP learning rule can be derived from a functional goal of maximizing separability.

But apart from the artificial counting task employed in [Lazar et al., 2009] as described, what are the benefits of an increased distance between network states? From an information theoretic point of view, a neuron should use its whole dynamic range of firing rates in order to increase the entropy [Bell and Sejnowski, 1995]. The same applies to a population of n neurons spanning an n-dimensional phase space. If all states are very similar to each other, the representational capacity is limited. Increasing the separability increases the used phase space volume which should be as large as possible to maximize the amount of transmittable information.

The separation of patterns is also important for encoding memories. Input patterns need to be stored (learning) and retrieved (recalling). A memory needs to accomplish two opposing tasks: during retrieval, the network should be able to find the stored pattern given an incomplete version of it as input. This is known as pattern completion. In the learning phase, it should separate similar input patterns i.e. enlarge their differences. This enhances the stability of the retrieval by reducing the interference between similar patterns and it is known as pattern separation. Both processes of pattern completion and separation are assumed to happen in different structures of hippocampus [Leutgeb and Leutgeb, 2007; Treves et al., 2008; Yassa and Stark, 2011]. The process of pattern separation in the hippocampus goes back to the seminal modeling work of Marr [1971]. It is also referred to as orthogonalization since it reduces the dot product between representations of two different inputs.

In line with this argument, for the functional goal of separability I consider the quadratic difference between the population activity vector at subsequent times: $(y_2 - y_1)^2 = y_2^2 + y_1^2 - 2y_2 \cdot y_1$. Maximizing the distance between these vectors minimizes their dot product given a bounded activity. To apply the formalism also to continuous time, I will formulate the objective function using the time-derivative as a generalized distance.

The response (firing rate) of a neuron i in a population is defined as a

nonlinear transformation f_i of the "membrane potential" u_i: $y_i(t) = f_i(u_i(t))$. For simplicity, I will drop the index i during the derivations and use vector and matrix notation:

$$F(t) := \text{diag}\left[f_i\left(u_i(t)\right)\right] \tag{3.2.1a}$$

$$F'(t) \equiv F^{(u)} := \text{diag}\left[\frac{\partial f_i}{\partial u_i}\right] \tag{3.2.1b}$$

The weight matrix W separates the neurons into two populations: the presynaptic and the postsynaptic population. Their responses will be labeled x and y respectively to avoid confusion:

$$x := y_{\text{pre}} \tag{3.2.2a}$$

$$y := y_{\text{post}} \tag{3.2.2b}$$

While in feed-forward architectures the populations are distinct, in the feedback loop of a recurrent network they are equivalent.

The membrane potential u of the postsynaptic neurons follows the differential equation

$$\dot{u} = -\frac{u}{\tau} + Wx$$

The objective function reads

$$O_{\text{sep}} := |\dot{y}|^2 \tag{3.2.3a}$$

$$= \left|F^{(u)}\dot{u}\right|^2 \tag{3.2.3b}$$

Discrete-time rate-coding

In the discrete-time case the reservoir state is given by the activities $y(t - n\Delta_t)$ at discrete time-points $t - n\Delta_t$ with

$$y(t + \Delta_t) = f(u(t)) \tag{3.2.4a}$$

$$u(t) = Wx(t) \tag{3.2.4b}$$

Here, the time-derivative $\dot{y}$ is not directly available. The resulting learning rules for the discrete-time case depend on the approximation of $\dot{y}$.

A simple linear approximation is the most straightforward approach. This can be formulated in terms of the activity y as

$$\dot{y}(t) \approx \frac{y(t + \Delta_t) - y(t)}{\Delta_t} \tag{3.2.5}$$

which is the discrete analog to (3.2.3a). Or in terms of the membrane potential u as

$$\dot{y}(t) \approx F'(t)\frac{u(t+\Delta_t) - u(t)}{\Delta_t} \tag{3.2.6}$$

which is the discrete analog to (3.2.3b).

But instead of linearly approximating $\dot{y}$, three subsequent points in time ($y(t+\Delta_t)$, $y(t)$, and $y(t-\Delta_t)$) can be used to construct a second-order approximation. Since these points are equidistant in time, it is easy to show that a quadratic function constrainted by these three points has a derivative of

$$\dot{y}(t) \approx \frac{y(t+\Delta_t) - y(t-\Delta_t)}{2\Delta_t} \tag{3.2.7}$$

or when approximating the membrane potential

$$\dot{y}(t) \approx F'(t)\frac{u(t+\Delta_t) - u(t-\Delta_t)}{2\Delta_t}. \tag{3.2.8}$$

Both expressions are very similar to their equivalents for the linear approximation. They only differ in the time index of the negative part, which gives a slightly different learning rule. It will lead to the same results but it has a more direct relation to commonly used learning rules. The learning rules for synaptic and intrinsic plasticity are derived in the following sections via a stochastic gradient ascent on the objective function with respect to the relevant parameters.

Activation function

The logistic sigmoid function is a common choice for the activation function f. It is defined as

$$f_{\text{sig}}(z) := (1 + \exp[-z])^{-1} \tag{3.2.9a}$$

$$\frac{\partial f_{\text{sig}}}{\partial z} = f_{\text{sig}}\,(1 - f_{\text{sig}}) \tag{3.2.9b}$$

$$\frac{\partial^2 f_{\text{sig}}}{\partial z^2} = \frac{\partial f_{\text{sig}}}{\partial z}\,(1 - 2f_{\text{sig}}) \tag{3.2.9c}$$

The argument

$$z := a\,(u - b) \tag{3.2.10}$$

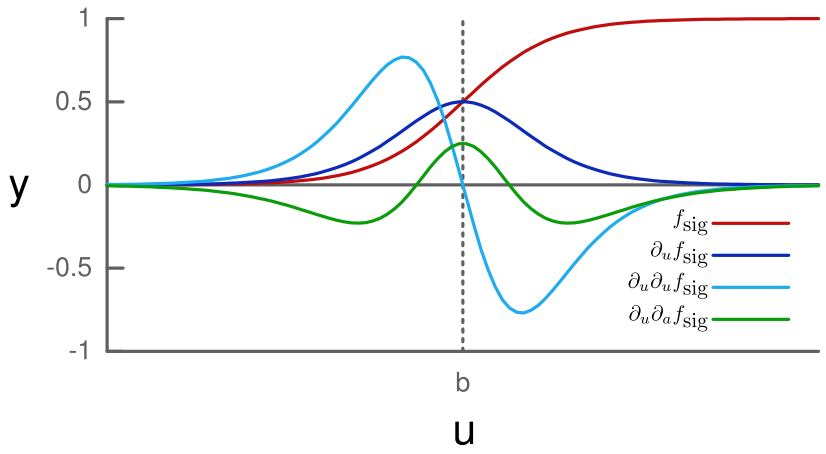

Figure 3.2 Sigmoid activation function $f = sig\,(a(u-b))$ with gain a and threshold b. The activation function (red) and its first derivative (dark blue) are always positive, while the second derivatives (light blue and green) have domains with negative values.

is a linear transformation of the "membrane potential" u introducing a gain a and a threshold b which modify the slope and the position of the sigmoid. The definition of z differs from authors where the threshold b changes sign and absorbs the gain (i.e. $z = a\,u + b$). The choice in this work for the form in Eqn. (3.2.10) is more convenient for the interpretation of b as the position of the middle point at $u = b \Leftrightarrow z = 0$ where the activation reaches its half-maximum $f_{\mathrm{sig}}(0) = 0.5$.

The derivatives with respect to u are just scaled by the gain:

$$f' = \frac{\partial f_{\mathrm{sig}}}{\partial u} = a\frac{\partial f_{\mathrm{sig}}}{\partial z} \tag{3.2.11a}$$

$$f'' = \frac{\partial^2 f_{\mathrm{sig}}}{\partial u^2} = a^2\frac{\partial^2 f_{\mathrm{sig}}}{\partial z^2} \tag{3.2.11b}$$

They are shown in Fig. 3.2. Its Taylor expansion to first order around the middle point $z = 0$ is given by:

$$\begin{aligned} f_{\mathrm{sig}}(z) \approx & f_{\mathrm{sig}}(0) + z\ \left. f'_{\mathrm{sig}}\right|_{z=0}. \\ = & \frac{1}{2} + \frac{a\,(u-b)}{4} \end{aligned} \tag{3.2.12}$$

These identities are used in the following to assess the stability of the learning rules.

3.2.1 Synaptic plasticity

The synaptic learning rule maximizing the separability in the network is found by taking the derivative of the objective function with respect to the weight matrix W. The change of the weight matrix is defined as

$$\begin{aligned}\dot{W} :=& \eta \frac{\partial O_{\text{sep}}}{\partial W} \\ =& 2\eta\ F' F' \dot{u} x^T && \text{(3.2.13a)} \\ =& 2\eta\ F' \dot{y} x^T && \text{(3.2.13b)}\end{aligned}$$

or using entry-wise notation

$$\begin{aligned}\dot{W}_{ij} :=& 2\eta\ f_i^2\left(u_i\right) \dot{u}_i x_j && \text{(3.2.14a)} \\ =& 2\eta\ f_i\left(u_i\right) \dot{y}_i x_j && \text{(3.2.14b)}\end{aligned}$$

The weight change of the connection W_{ij} is proportional to the presynaptic activity x_j and the time-derivative of the postsynaptic activity y_i. In contrast to standard Hebbian learning, which is proportional to x and y, this is called differential-hebbian learning [Kosko, 1986]. This learning rule leads to an STDP rule for discrete-time rate-coding neurons as well as for continuous-time Poisson spiking neurons.

Discrete time

The values of y at t and $t - \Delta_t$ are considered already observed and, therefore, constant. They do not change when W is varied and, hence, do not contribute to the derivative. Using the quadratic approximation from (3.2.7), the discrete time version for the weight change is given by

$$\begin{aligned}\Delta W(t) =& \eta\ \frac{\partial}{\partial W}\ \left(\frac{y(t+\Delta_t) - y(t-\Delta_t)}{2\Delta_t}\right)^2 && \text{(3.2.15a)} \\ =& \frac{\eta}{\Delta_t}\ F'(t+\Delta_t)\ \frac{y(t+\Delta_t) - y(t-\Delta_t)}{2\Delta_t} x(t)^T && \text{(3.2.15b)} \\ \propto& \underbrace{F'(t+\Delta_t)\ y(t+\Delta_t) x(t)^T}_{\Delta W_+(t)} - \underbrace{F'(t+\Delta_t)\ y(t-\Delta_t) x(t)^T}_{\Delta W_-(t)}. && \text{(3.2.15c)}\end{aligned}$$

Using the linear approximation would lead to a very similar rule differing only in the time index of y in the negative part. The general behavior would be the same, but using the quadratic approximation allows for a connection to

biologically inspired rules. Rearranging terms reveals the asymmetric structure of the learning rule

$$\begin{aligned}\Delta\tilde{W}(t+\Delta_t) :=& \Delta W_+(t-\Delta_t) - \Delta W_-(t) \\ \propto& + F'(t)\,\underbrace{y(t)x(t-\Delta_t)^T}_{\text{pre-post}} - F'(t+\Delta_t)\,\underbrace{y(t-\Delta_t)x(t)^T}_{\text{post-pre}}\end{aligned}$$

which is commonly used as the discrete-time analog of STDP. In contrast to the standard discrete-time STDP rule, the one derived here has an additional factor F'. It is modulated by the derivative of the activation function. Thus, the weight change will be almost zero for a saturated or silent postsynaptic neuron. This adds an inherent stability to the learning rule.

Stability. Although containing hebbian and anti-hebbian terms, the standard STDP rule suffers from unbounded growth due to its causal nature [Babadi and Abbott, 2010]. The modulation term in the modified STDP rule (3.2.15b) stabilizes the learning. That means, it has a stable fixed point and does not require explicit weight normalization or scaling.

To derive an analytic expression for the fixed point of the weight dynamics, I consider the simple case of one input and one output neuron and an independent, identically distributed (iid) Gaussian input $u = wx$ with $x \sim \mathcal{N}(x_0, \varepsilon^2)$. Assuming a small learning rate η the weight can be considered as fixed for calculating the mean weight change.

In the Hebbian part $\Delta W_+(t)$ of Eqn. (3.2.15c) all terms depend on the same input sample $x(t)$, since $f'(t+\Delta_t) = f'|_{u=u(t+\Delta_t)}$ and $y(t+\Delta_t) = f\left(u(t+\Delta_t)\right)$ with $u(t+\Delta_t) = W\,x(t)$. For the anti-Hebbian part $\Delta W_-(t)$, $y(t-\Delta_t)$ it is independent of the other two terms $F'(t+\Delta_t)$ and $x(t)$ (given an iid input).

If the input variance ε^2 is small, the terms y and f' can be approximated

by a first-order Taylor expansion around the mean input wx_0:

$$\begin{aligned}\langle \Delta w \rangle \propto &+ \int \left[f' + w(x-x_0)f''\right]\left[f + w(x-x_0)f'\right] x\,\mathcal{N}\left(x; x_0, \varepsilon^2\right)\,\mathrm{d}x \\ &- \left(\int \left[f' + w(x-x_0)f''\right] x\,\mathcal{N}\left(x; x_0, \varepsilon^2\right)\,\mathrm{d}x \right. \\ &\left. \int \left[f + w(x-x_0)f'\right]\mathcal{N}\left(x; x_0, \varepsilon^2\right)\,\mathrm{d}x\right) \qquad (3.2.16a) \\ = &+ \left(f f' x_0 + \varepsilon^2 w(f'^2 + f f'') + x_0 w^2 \varepsilon^2 f' f''\right) \\ &- \left(f f' x_0 + \varepsilon^2 w f f''\right) \qquad (3.2.16b) \\ = &+ \varepsilon^2 w f' \left(f' + w x_0 f''\right) \qquad (3.2.16c)\end{aligned}$$

where f' and f'' are to be evaluated at $u = wx_0$. If the mean weight change is zero for some finite value of w, the weight has a fixed point. Such a finite w can always be found for a bounded, monotonically increasing activation function f with a derivative f' that decays sufficiently fast (i.e. exponentially).

The sigmoid function fulfills the conditions of a bounded, monotonic increase and an exponential behavior of f'. The stable fixed point for this activation function is the solution to

$$0 = 1 + awx_0\left(1 - 2f_{\mathrm{sig}}\right), \qquad (3.2.17)$$

which follows from the derivatives of the sigmoid in (3.2.9). With the Taylor expansion from (3.2.12) this simplifies to

$$0 = 1 - \frac{1}{2}a^2\, wx_0\left(wx_0 - b\right), \qquad (3.2.18)$$

which has the solution

$$w^* = \frac{b}{2x_0}\left(1 \pm \sqrt{1 + \frac{8}{a^2 b^2}}\right). \qquad (3.2.19)$$

Fig. 3.3 shows the output distribution and the evolution of the weight for a threshold of $b = 1$ and different values of the gain. In this case the input distribution is Gaussian with a mean $x_0 = 0.5$ and a standard deviation $\sigma = 0.2$. For high values of the gain a (right panel) the weight converges to the theoretical value of $w^* = \dfrac{b}{x_0} = 2$ from eqn. (3.2.19) after 50,000 iterations (learning rate $\eta = 1e-3$). The rate of convergence increases with the value of

the gain a, since the modulation factor f' of the learning rule has a maximum of $a/4$.

Excluding the instable negative solution, the stable fixed point w^* is approximately $\frac{b}{x_0}$ for sufficiently large a or b. In this case, for a unimodal symmetric distribution of x as assumed in (3.2.16c), the weight approaches a value such that the mean input $u = wx_0$ equals the threshold b. The distribution of u will be centered at the middle point of the sigmoid activation function, where it behaves linearly, and 'avoids' the regions of the output saturation. This leads to a symmetric output distribution centered at 0.5 having maximal variance for a fixed gain. If the variance $(w^*\sigma)^2$ of the input fits with the large gain, the output distribution becomes flat. But if the gain is too large, the activation function becomes too steep and the output pdf exhibits two peaks at 0 and 1 (cf. Fig. 3.3, top-right).

If the gain becomes smaller such that the second term in the square root can not be neglected, the stable point for w becomes larger as given by Eqn. (3.2.19). The input variance increases for a larger w, while the input stays mainly in the linear region due to the small slope (gain) of the activation function. The output distribution will be tilted towards 1 (cf. Fig. 3.3, top-left).

Continuous-time

Now, I will apply the differential-hebbian learning rule from Eq. 3.2.13a to a spiking neuron. The activities $x(t)$ and $y(t)$ now correspond to an average firing rate. The synapse is assumed to estimate these with an exponential average over spike times t_i with (possibly) different timescale for pre- and postsynaptic averaging:

$$x_j(t) := \sum_{t_j} \exp\left[-\frac{t-t_j}{\tau_{\text{pre}}}\right] \frac{\Theta(t-t_j)}{\tau_{\text{pre}}} \tag{3.2.20a}$$

$$y_i(t) := \sum_{t_i} \exp\left[-\frac{t-t_i}{\tau_{\text{post}}}\right] \frac{\Theta(t-t_i)}{\tau_{\text{post}}} \tag{3.2.20b}$$

$$\Rightarrow \dot{y}_i = \sum_{t_i} \exp\left[-\frac{t-t_i}{\tau_{\text{post}}}\right] \left(\frac{\delta(t-t_i)}{\tau_{\text{post}}} - \frac{\Theta(t-t_i)}{\tau_{\text{post}}^2}\right) \tag{3.2.20c}$$

The activation function from the previous section related the membrane potential to the output activity. When considering explicit spike generation

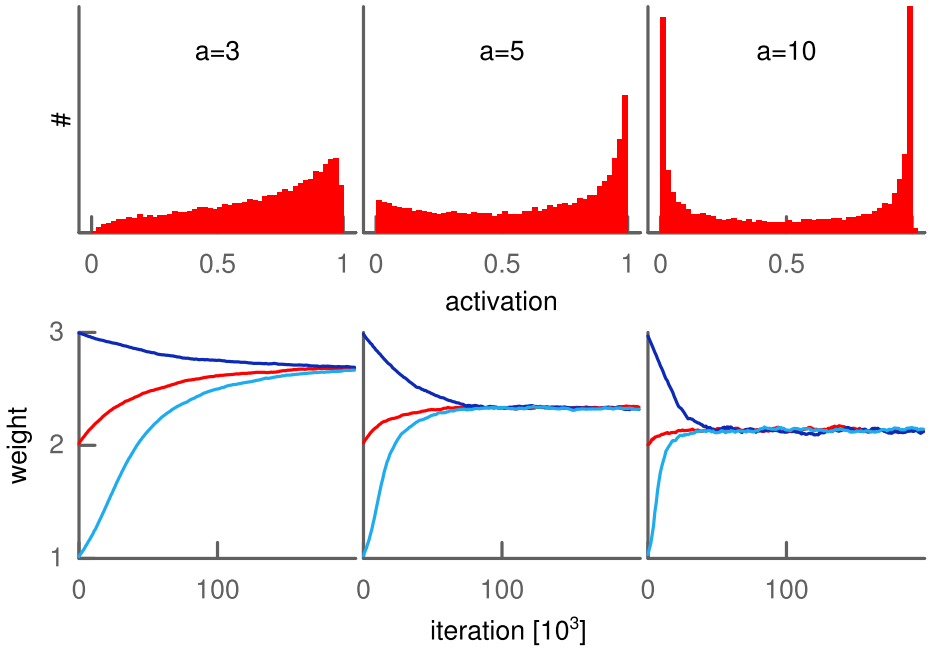

Figure 3.3 Discrete-time STDP from the separability objective in the case of one input and one output. Input with mean $x_0 = 0.5$ and standard deviation $\sigma = 0.2$, sigmoid activation function with threshold $b = 1$ and varying gain a (left, middle, and right). **Top:** distribution of outputs y after the weight has converged. **Bottom:** evolution of the weight for different initial values.

as a threshold process in terms of the membrane potential, the activation function describes the relation between the presynaptic input current I and the output firing frequency f. While this $f - I$ curve depends on the type of neuron [Connors and Gutnick, 1990], the relationship is mainly linear above some threshold and saturates for stronger input currents [Nowak et al., 2003; Tateno et al., 2004].

The total synaptic change is obtained as

$$\Delta W_{ij} = \int_{-\infty}^{\infty} \dot{W}_{ij} \, \mathrm{d}t = \eta \int_{-\infty}^{\infty} 2 f_i' \dot{y}_i x_j \, \mathrm{d}t \tag{3.2.21}$$

In a first step to simplify the derivations, I will assume a linear activation function without saturation or a maximal firing rate. Then the constant modulation of f' can be neglected.

The activity is a sum over all spike times and this will lead to a linear interaction between all pre- and postsynaptic spikes known as all-to-all STDP. Since the contributions are linear, looking at the case of one pre- and one

postsynaptic spike is sufficient.

Like in the standard STDP protocol the relevant parameter is the time delay $\Delta t := t_{\text{post}} - t_{\text{pre}}$.

- Post- before Pre-Spike: $t_{\text{post}} < t_{\text{pre}}$ $(\Delta t < 0)$

$$\begin{aligned}
\Delta W_{ij} =& \underbrace{\int_{-\infty}^{t_{\text{post}}} \dot{W}_{ij}\,\mathrm{d}t}_{=0} + \underbrace{\int_{t_{\text{post}}}^{t_{\text{pre}}} \dot{W}_{ij}\,\mathrm{d}t}_{=0} + \int_{t_{\text{pre}}}^{\infty} \dot{W}_{ij}\,\mathrm{d}t \\
=& 2\eta \int_{t_{\text{pre}}}^{\infty} \exp\left[-\frac{t-t_{\text{pre}}}{\tau_{\text{pre}}}\right] \exp\left[-\frac{t-t_{\text{post}}}{\tau_{\text{post}}}\right] \left(\underbrace{\frac{\delta\left(t-t_{\text{post}}\right)}{\tau_{\text{post}}}}_{=0} - \frac{1}{\tau_{\text{post}}^2} \right) \mathrm{d}t \\
=& \frac{2\eta}{\tau_{\text{post}}} \frac{\tau_{\text{pre}}}{\tau_{\text{pre}}+\tau_{\text{post}}} \left[\exp\left[-\frac{\left(\tau_{\text{pre}}+\tau_{\text{post}}\right) t - \tau_{\text{post}} t_{\text{pre}} - \tau_{\text{pre}} t_{\text{post}}}{\tau_{\text{pre}}\tau_{\text{post}}} \right] \right]_{t_{\text{pre}}}^{\infty} \\
=& -\frac{2\eta}{\tau_{\text{post}}} \frac{\tau_{\text{pre}}}{\tau_{\text{pre}}+\tau_{\text{post}}} \exp\left[+\frac{\Delta t}{\tau_{\text{post}}}\right]
\end{aligned}$$

- Pre- before Post-Spike: $t_{\text{post}} > t_{\text{pre}}$ $(\Delta t > 0)$

$$\begin{aligned}
\Delta W_{ij} =& 2\eta \int_{t_{\text{post}}}^{\infty} \exp\left[-\frac{t-t_{\text{pre}}}{\tau_{\text{pre}}}\right] \exp\left[-\frac{t-t_{\text{post}}}{\tau_{\text{post}}}\right] \left(\underbrace{\frac{\delta\left(t-t_{\text{post}}\right)}{\tau_{\text{post}}}}_{t=t_{\text{post}}} - \frac{1}{\tau_{\text{post}}^2} \right) \mathrm{d}t \\
=& \frac{2\eta}{\tau_{\text{post}}} \exp\left[-\frac{\Delta t}{\tau_{\text{pre}}}\right] - \frac{2\eta}{\tau_{\text{post}}} \frac{\tau_{\text{pre}}}{\tau_{\text{pre}}+\tau_{\text{post}}} \exp\left[-\frac{\Delta t}{\tau_{\text{pre}}}\right] \\
=& \frac{2\eta}{\tau_{\text{post}}} \frac{\tau_{\text{post}}}{\tau_{\text{pre}}+\tau_{\text{post}}} \exp\left[-\frac{\Delta t}{\tau_{\text{pre}}}\right]
\end{aligned}$$

Thus, the differential-hebbian learning, derived from the separability objective, leads to the standard model of the STDP pairing protocol in the form of two exponentials (cf. Fig. 3.4):

$$\Delta W_{ij} \propto +\tau_{\text{post}}\,\Theta\left(\Delta t\right) \exp\left[-\frac{\Delta t}{\tau_{\text{pre}}}\right] - \tau_{\text{pre}}\,\Theta\left(-\Delta t\right) \exp\left[+\frac{\Delta t}{\tau_{\text{post}}}\right]$$

Stability. The total integral under the STDP windows is zero irrespective of τ_{pre} and τ_{post}. But this does not imply that the weights will be stable. Like in the discrete-time case, the causal part of this plasticity rule leads to

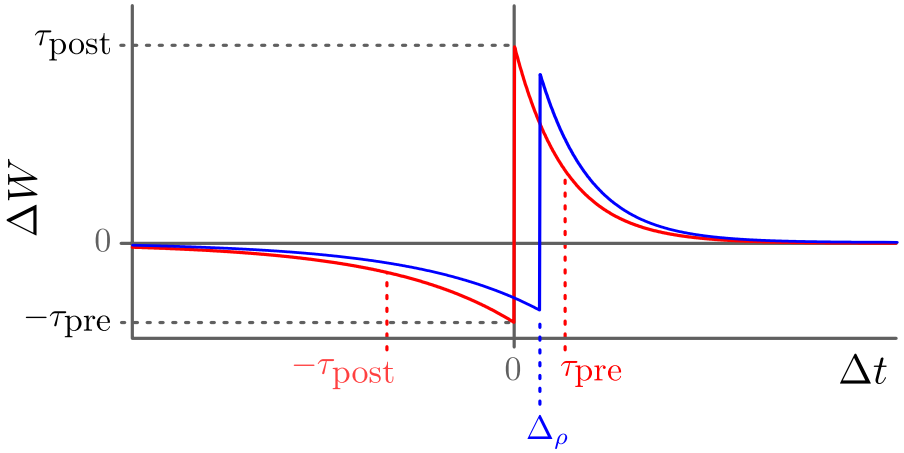

Figure 3.4 Continuous-time STDP windows from the separability objective for one pre/post spike pair. $\tau_{\text{pre/post}}$ are the time constants for the exponential averaging of pre-/postsynaptic activity from their spike trains. $\Delta t > 0$ corresponds to pre-post, $\Delta t < 0$ to post-pre condition. Red line: Standard STDP window. Blue line: STDP window shifted by Δ_ρ due to saturation in the activation function.

a potentiation loop and results in unbounded growth or saturated weights Babadi and Abbott [2010].

I will show that, like in the discrete-time case, the learning is stabilized by the modulation through f', i.e. the derivative of the activation function. A stable weight distribution can be attained by considering a bounded activation function. The actual learning rule will depend on the functional form of the activation function $y = f(x)$ relating input and output. For analytical tractability I assume a linear activation function but introduce a maximal output firing rate such that $f(x)$ is constant above some threshold $\hat{x} = \frac{\rho}{\tau_{\text{pre}}}$. In this case the derivative of f, which modulates the learning, is a Θ function as

$$f'(x) = \Theta\left(\hat{x} - x\right). \tag{3.2.22}$$

Again, I consider the STDP pairing case of one pre- and one postsynaptic spike with exponential averaging like in (3.2.20). Due to the threshold, the product $f'(x)\,x(t)$, which enters the learning rule in (3.2.21), is zero for $x > \hat{x}$. Rewriting the presynaptic activity x as

$$\begin{aligned} x(t) :=& \frac{1}{\tau_{\text{pre}}} \exp\left[-\frac{t - t_{\text{pre}}}{\tau_{\text{pre}}}\right] \Theta\left(t - t_{\text{pre}}\right) \\ =& \hat{x} \exp\left[-\frac{t - t_{\text{pre}} - \Delta_\rho}{\tau_{\text{pre}}}\right] \Theta\left(t - t_{\text{pre}}\right). \end{aligned} \tag{3.2.23}$$

one can see that this threshold corresponds to a time $t_\rho := t_{\text{pre}} + \Delta_\rho$ with $\Delta_\rho := -\tau_{\text{pre}} \ln \rho$.

The Θ function from f' over $\hat{x}$ can be rewritten into a Θ function over t_ρ:

$$\begin{aligned} f'x(t) =& \Theta(\hat{x} - x)\ \hat{x} \exp\left[-\frac{t - t_{\text{pre}} + \tau_{\text{pre}} \ln \rho}{\tau_{\text{pre}}}\right] \Theta(t - t_{\text{pre}}) \\ =& \Theta(t - t_\rho)\ \hat{x} \exp\left[-\frac{t - t_\rho}{\tau_{\text{pre}}}\right] \end{aligned} \tag{3.2.24}$$

where the last Θ function can be neglected since $t_\rho > t_{\text{pre}}$ for $\rho < 1$.

The effect on the learning rule is a time-shifted STDP window. For a purely linear activation function, the crossover from LTD to LTP happens at $\Delta t = 0$. Incorporating a simple threshold saturation results in a shift to the right, such that a pairing with $\Delta t < \Delta_\rho$ results in LTD (cf. Fig. 3.4). That means, a pre-post pairing with a short time delay leads to strong depression instead of strong potentiation.

Such a time-shifted form of STDP has been proposed and examined by Babadi and Abbott [2010]. They show that such an STDP rule can compensate the causal bump in the LTP regime. Instead of having weights saturated at zero and a maximal value, the time-shifted STDP gives rise to a stable distribution of weights. While Babadi and Abbott based their approach on a purely phenomenological argument, the current work can actually provide a justification for such a time-shift. When the STDP rule is derived from the separability objective, it intrinsically accounts for the time-shift through the nonlinear, saturating activation function.

3.2.2 Intrinsic plasticity

Intrinsic plasticity (IP) modifies the input-output mapping of each neuron individually. For a rate based neuron i this is the activation function f_i. Again, the learning rule is found by taking the derivative of the objective function with respect to the parameters of f_i. Given an activation function with some parameter p

$$\dot{p} := \eta_{\text{IP}} \frac{\partial O_{\text{sep}}}{\partial p}. \tag{3.2.25}$$

Thus

$$\dot{p} \propto \dot{u}^T F^{(u,p)} F^{(u)} \dot{u} \tag{3.2.26a}$$

$$= \dot{y}^T \left[F^{(u)}\right]^{-1} F^{(u,p)} \dot{y} \tag{3.2.26b}$$

where $F^{(u,p)}$ is the diagonal matrix of the second derivatives with respect to u and the parameter p.

For the common choice of the logistic sigmoid function, two parameters can be optimized: the gain a and the threshold b. Since $\dot{y}^2$ and $F^{(u)}$ are both always positive, the sign of the change are determined by the second derivatives of f_i (cf. Fig. 3.2).

$$\begin{aligned} \dot{b}_i \propto & \dot{y}_i^2 \left[\frac{\partial f_i}{\partial u_i}\right]^{-1} \frac{\partial^2 f_i}{\partial u_i \, \partial b_i} \\ = & - \dot{y}_i^2 a_i \left(1 - 2y_i\right) \end{aligned} \tag{3.2.27}$$

$$\begin{aligned} \dot{a}_i \propto & \dot{y}_i^2 \left[\frac{\partial f_i}{\partial u_i}\right]^{-1} \frac{\partial^2 f_i}{\partial u_i \, \partial a_i} \\ = & \dot{y}_i^2 \left[\frac{1}{a_i} + \left(u_i - b_i\right)\left(1 - 2y_i\right)\right] \end{aligned} \tag{3.2.28}$$

These learning rules derived from the objective function of separability are similar to the ones proposed by Bell and Sejnowski [1995] for maximizing mutual information. This makes intuitive sense, since the separability objective also aims at increasing the employed phase space volume. In turn, a larger phase space increases the entropy/mutual information.

So while there is a connection between these to approaches, they are, however, not the same. The difference is that, here, the learning rate of a and b depends on the actual output dynamics since it scales with $\dot{y}^2$. It is easy to show that the separability objective in the discrete-time case is equivalent to maximizing the variance of the output PDF assuming iid samples at subse-

quent times:

$$\begin{aligned}\left\langle (y_t - y_{t-1})^2 \right\rangle =& \int y_t^2\, p(y_t)\, \mathrm{d}y_t + \int y_{t-1}^2\, p(y_{t-1})\, \mathrm{d}y_{t-1} \\ & - 2 \int y_t y_{t-1}\, p(y_{t-1})\, p(y_{t-1})\, \mathrm{d}y_t\, \mathrm{d}y_{t-1} \\ =& 2 \left\langle y^2 \right\rangle - 2 \left\langle y \right\rangle^2 \\ =& 2\sigma_y^2. \end{aligned} \tag{3.2.29}$$

The uniform distribution, which is the maximum entropy PDF for a bounded output with no constraints, has indeed the maximal variance of $1/12$ for the interval $[0, 1]$ among the unimodal, symmetric distributions. There are, however, multimodal distributions with a larger variance which is reflected in the output PDF of Figure 3.6. In any case, a flat uniform distribution is not reachable by either learning rule given the limited range of transformations provided by the sigmoid activation function and its two parameters.

Stability

Like for the weights, the stability of the sigmoid parameters can be assessed by calculating the root of the average change in a and b. To transfer the learning rule into discrete time, I take the quadratic approximation from (3.2.7) as before. Assuming an iid input, the factor $\dot{y} = \dfrac{y(t + \Delta_t) - y(t - \Delta_t)}{2\Delta_t}$ is independent of the main expression which depends on $y(t)$. Thus, $\dot{y}^2$ contributes just a constant factor to the average change and can be neglected. Again, a Taylor expansion of the sigmoid around $z = 0$ leads to an analytical expression for the fixed points in the case of a Gaussian distributed input $u \sim \mathcal{N}(u_0, \sigma_u^2)$:

$$\langle \Delta b \rangle \propto - \int p(u)\; a\,(1 - 2f(u))\; \mathrm{d}u \tag{3.2.30a}$$

$$\approx \frac{a^2}{2} \int p(u)\; (u - b)\; \mathrm{d}u \tag{3.2.30b}$$

$$= \frac{a^2}{2}\,(u_0 - b) \tag{3.2.30c}$$

The threshold b has a stable fixed point at the mean membrane potential u_0, thereby making the average output $\langle y \rangle = 0.5$.

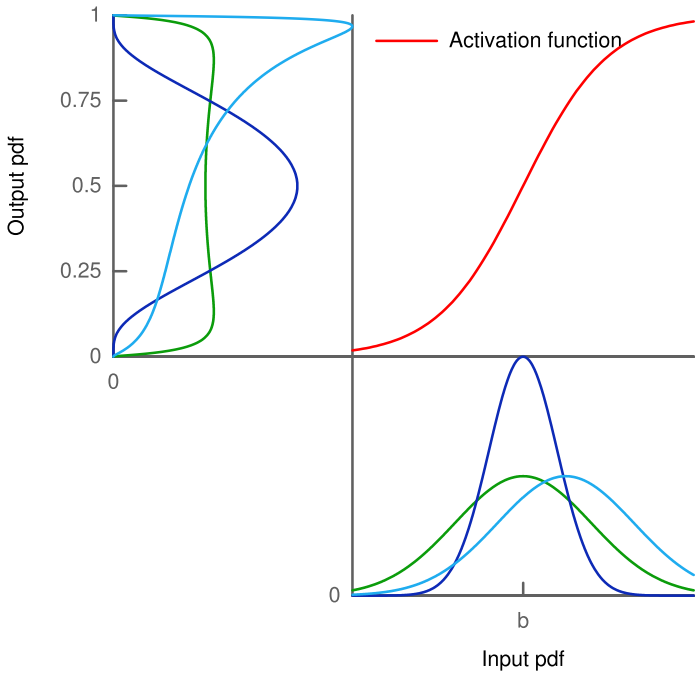

Figure 3.5
Gaussian input applied to a sigmoid activation function. The output pdf is flat if the parameters of the sigmoid fit to the mean and variance of the Gaussian (green line).

Likewise, the mean change in the gain is

$$\langle \Delta a \rangle \propto \int p(u) \left[\frac{1}{a} + (u-b)\left(1 - 2f(u)\right) \right] \mathrm{d}u \tag{3.2.31a}$$

$$\approx \frac{1}{a} - \frac{a}{2} \int p(u) \ (u-b)^2 \ \mathrm{d}u \tag{3.2.31b}$$

$$= \frac{1}{a} - \frac{a}{2}\sigma_u^2 \tag{3.2.31c}$$

by setting the threshold to its stable point $b = u_0$. The gain a has a stable fixed point at $\frac{\sqrt{2}}{\sigma_u}$. Thus, a adapts to the variance of the input pdf.

Both adaptations together can change the activation function such that it matches the input pdf. This makes the output pdf more flat and maximizes the separability/phase space as well as the entropy. Figure 3.5 shows the PDF of the output for a Gaussian-distributed input transformed by a sigmoid activation function. The output PDF is mainly flat if the mean of the Gaussian input matches the steepest point at b and the variance fits with the slope a. If either the mean or the variance is changed, the output pdf becomes less flat.

The stability of threshold and weight depends on each other. Unfortunately, they do not have a common stable point. By inserting the stable point equations of a and b into the equation for the stable point of w from (3.2.19), the

"stable" point is found to continuously shift as

$$w^* = w^* \frac{1}{2}\left(1 + \sqrt{1 + \frac{4\sigma_x^2}{x_0^2}}\right) \tag{3.2.32a}$$

$$\approx w^* \left(1 + \frac{\sigma_x^2}{x_0^2}\right). \tag{3.2.32b}$$

Thereby, the weight increases towards its new stable point which in turn increases the stable point of the threshold and so forth. That means, w and b would grow without bounds when trained together. Only in the limit when the coefficient of variation $\frac{\sigma_x}{x_0}$ goes to zero, the threshold and the weight have matching stable fixed points. In this case the gain goes to infinity resulting in a Heaviside activation function. Thus, for the network simulations in Section 3.4, with concurrent intrinsic and synaptic plasticity, the threshold is not adapted.

The effect of the learning rule can be seen in Fig. 3.6. The solid histograms in the upper row represent the distribution of the output after the gain and threshold have converged. The distribution is mainly flat irrespective of the standard deviations of the input (left, middle, and right panel). Theses differences in the input are balanced by the stable fixed point of the gain which is inversely proportional to the standard deviation $\sigma_u = w\sigma_x$ as shown in the lower row. Since the rate of convergence scales with $\dot{y}^2$, it initially depends on the standard deviation of the input which influences $\dot{y}$ while the gain is far from the fixed point. Therefore, the gain converges faster for $\sigma_x = 0.28$ (right) than for $\sigma_x = 0.14$ (left). In all three cases the threshold b converges to 1, which is a stable fixed point counterbalancing the mean input wx_0.

3.3 Objective function: Energy and noise

The objective function from Eqn. (3.2.3a) introduced in the previous section leads the neuron to employ its whole phase-space for efficiently transmitting information. As shown, this objective results in an intrinsic plasticity learning rule which is similar to maximizing mutual information [Bell and Sejnowski, 1995], i.e. the entropy of the output given the input distribution. The learned flat output distribution corresponds to the maximum-entropy distribution of a bounded output which is uniform.

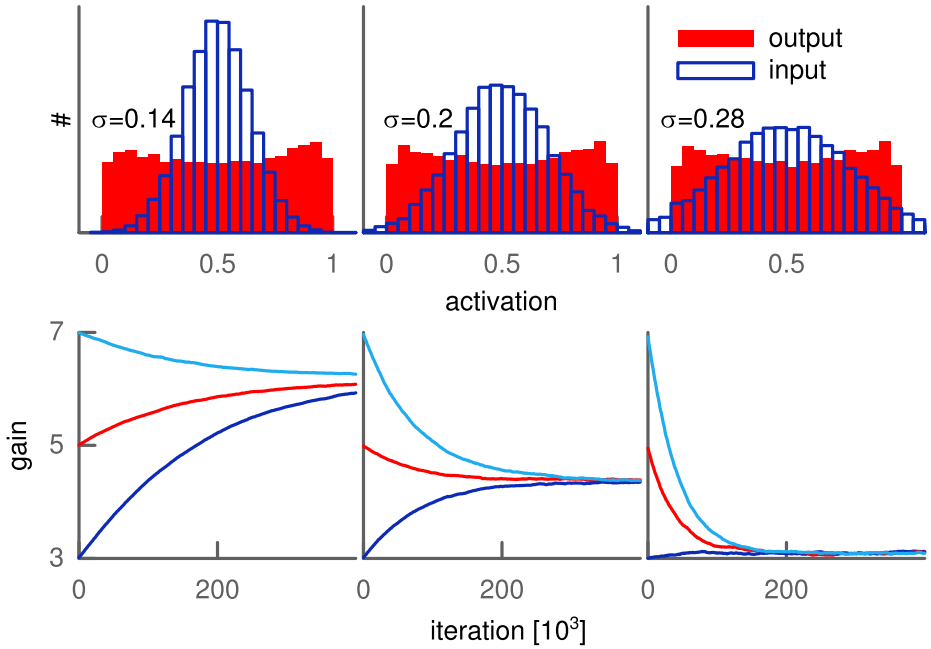

Figure 3.6 Intrinsic plasticity from the separability objective in the case of one input and one output. A fixed weight $w = 2$ and an input with mean $x_0 = 0.5$ and varying standard deviations σ_x (left, middle, and right). **Top:** distribution of inputs x and outputs y after the gain has converged. **Bottom:** evolution of the gain for different initial values. The threshold converged quickly towards $wx_0 = 1$ (not shown).

Triesch [2005] extended the mutual information approach by requiring a low mean firing rate given that spikes are metabolically expensive as discussed in Sections 2.2.2 and 2.3.1. This leads to an exponential distribution of the outputs. These extended intrinsic plasticity rules can easily be derived in the objective function framework by adding another objective linearly penalizing the output:

$$O_{\text{energy}} := -\frac{y}{\mu}, \tag{3.3.1a}$$

where the parameter μ regulates the mean firing rate. This objective introduces additional terms for the intrinsic plasticity rules equivalent[1] to the ones found

[1] note the difference in the definition of the threshold for the sigmoid activation function: $sig\,(a(u-b))$ vs. $sig\,(au+b)$ as used in [Triesch, 2005; Bell and Sejnowski, 1995]

by Triesch [2005]:

$$\frac{\partial O_{\text{energy}}}{\partial b_i} = a_i \frac{f'}{\mu} = \frac{a_i}{\mu} y\,(1-y) \tag{3.3.2a}$$

$$\frac{\partial O_{\text{energy}}}{\partial a_i} = -\,(u-b_i)\,\frac{f'}{\mu} = -\,\frac{u-b_i}{\mu} y\,(1-y) \tag{3.3.2b}$$

The combined objective $O_{\text{sep}} + \lambda O_{\text{energy}}$ yields an intrinsic plasticity rule which leads to an exponential distribution.

The problem with this combined objective function is that it is not stable for constant input. If the variance of the input goes to zero, $\dot{y}$ and the separability objective also go to zero. The remaining learning rules in (3.3.2b) and (3.3.2a) for the energy objective are, however, only stable at $y = 0$, and thus $a, b \to \infty$.

To avoid very high rates due to their energy consumption, I quadratically penalize them. This fits with the initially introduced connection between the separability objective and the orthogonalization of neural representations [Marr, 1971]:

$$-2y_2 \cdot y_1 = (y_2 - y_1)^2 - \left(y_2^2 + y_1^2\right) \tag{3.3.3a}$$

Decreasing the dot product increases the orthogonality, which can be achieved by increasing the squared difference between different outputs (separability) while decreasing the sum of the squared outputs (energy). Thus, the combination of the energy objective and the separability objective is equivalent to increasing the orthogonality.

Furthermore, firing rate distributions have been found to have a lognormal distribution [Hromádka et al., 2008]. That means, very low firing rates are less probable than an exponential distribution would suggest. A possible reason for a neuron to avoid low firing rates is their sensitivity to noise [Tsubo et al., 2012]. I will account for this effect by linearly rewarding higher firing rates.

The resulting objective creates a trade-off between energy requirements and noise stability:

$$O_{\text{energy-noise}} := +\frac{y}{\mu} - \frac{1}{2}\left(\frac{y}{\mu}\right)^2 \tag{3.3.4a}$$

3.3.1 Intrinsic plasticity

The additional terms from the new objective for the intrinsic plasticity of a sigmoid activation function are given by

$$\frac{\partial O_{\text{energy-noise}}}{\partial b_i} = a_i \frac{f'}{\mu}\left(\frac{y}{\mu} - 1\right) = \frac{a_i}{\mu} y\,(1-y)\left(\frac{y}{\mu} - 1\right) \tag{3.3.5a}$$

$$\frac{\partial O_{\text{energy-noise}}}{\partial a_i} = -\left(u - b_i\right)\frac{f'}{\mu}\left(\frac{y}{\mu} - 1\right) = -\frac{u - b_i}{\mu} y\,(1-y)\left(\frac{y}{\mu} - 1\right). \tag{3.3.5b}$$

This objective has a continuum of stable fixed points. That means, given a constant input u every configuration of a and b such that $y = \mu$ is stable.

The combined objective

$$O_{\text{all}} := O_{\text{sep}} + \lambda O_{\text{energy}} \tag{3.3.6a}$$

balances the two objectives and finds an intermediate solution where the output is broadly distributed but low values around μ are more likely. The balance between those two depends on the value of λ (cf. Fig. 3.7).

3.3.2 Synaptic plasticity

Plasticity of excitatory synapses

The new objective can also be optimized by adapting the weight:

$$\frac{\partial O_{\text{energy-noise}}}{\partial W} = \frac{1}{\mu} F^{(u)}\left(1 - \frac{y}{\mu}\right)\frac{\partial u}{\partial W}^{T} \tag{3.3.7a}$$

For the discrete-time case, where the membrane potential depends linear on the weight, it directly follows that $\dfrac{\partial u}{\partial W}$ is equal to the presynaptic activity x. In a continuous-time formulation however, the membrane potential does not directly depend on the weight. Only its change $\dot{u}$ is linear in W. In order to still derive a learning rule, I consider the membrane potential at some future time $t_0 + \delta_t$, which can be found by integration:

$$u(t_0 + \delta_t) = u(t_0) + \int_0^{\delta_t} \dot{u}(t_0 + t')\,\mathrm{d}t'. \tag{3.3.8a}$$

Thus,

$$\left.\frac{\partial u}{\partial W}\right|_{t+\delta_t} = \int_0^{\delta_t} x(t + t')\,\mathrm{d}t' \overset{\delta_t \to 0}{\to} \delta_t x \tag{3.3.9a}$$

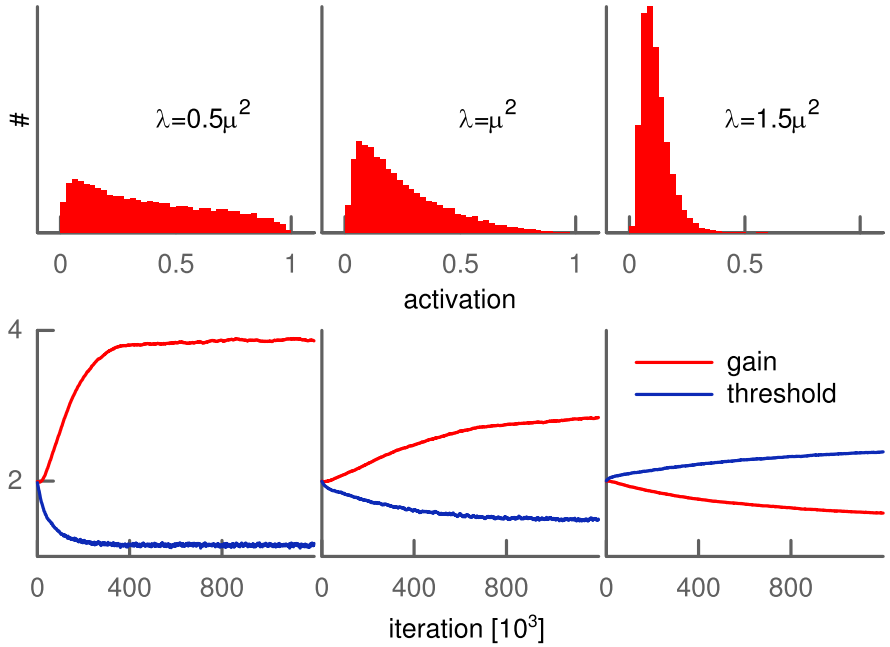

Figure 3.7 Intrinsic plasticity for the combined objective $O_{\text{sep}} + \lambda O_{\text{energy}}$ for varying values of λ (left, middle, right). *Top:* distribution of outputs y after the gain and threshold have converged. **Bottom:** evolution of the gain and threshold.

The derivative is proportional to the time step δ_t. For the discrete-time case, this factor can be set to 1 and one recovers the result from above, while for the continuous-time formulation δ_t goes to zero. Thus, the energy-noise objective can be neglected in continuous-time when it is considered together with the separability objective.

Plasticity of inhibitory synapses

While timing-dependence of synaptic plasticity at excitatory synapses is well established, synaptic plasticity at inhibitory synapses displays a larger variety, and many different forms of STDP have been observed [Lamsa et al., 2010; Maffei, 2011].

I include the influence of input at inhibitory synapses as

$$\dot{u} = -\frac{u}{\tau} + W_{\text{exc}} x_{\text{exc}} - W_{\text{inh}} x_{\text{inh}}.$$

Due to the minus sign, the separability objective in continuous-time leads to the standard STDP windows, but inverted in time. That means, pre-post leads

to LTD while post-pre to LTP as it has been observed by Bell et al. [1997].

On the other hand, an inhibitory interneuron might be considered as only being responsible to balance excitation and introduce competition. Therefore, I leave out the separability objective and only optimize the energy objective for inhibitory synapses:

$$\Delta W^{\text{inh}} = \int_{-\infty}^{\infty} \dot{W}_{ij}\,\mathrm{d}t \tag{3.3.10a}$$

$$= -\eta\lambda\delta_t \int_{-\infty}^{\infty} \left(1 - \frac{y_i}{\mu}\right) \frac{x_j}{\mu}\,\mathrm{d}t \tag{3.3.10b}$$

$$\propto -\frac{1}{\mu} + \frac{1}{\mu^2} \int_{-\infty}^{\infty} y_i x_i\,\mathrm{d}t. \tag{3.3.10c}$$

For the pairing case, one presynaptic inhibitory spike at t_{pre} and one postsynaptic excitatory spike at t_{post} are paired:

$$\int_{-\infty}^{\infty} y_i x_i\,\mathrm{d}t = \underbrace{\int_{-\infty}^{t_{\min}} y_i x_j\,\mathrm{d}t}_{=0} + \underbrace{\int_{t_{\min}}^{t_{\max}} y_i x_j\,\mathrm{d}t}_{=0} + \int_{t_{\max}}^{\infty} y_i x_j\,\mathrm{d}t \tag{3.3.11a}$$

$$= \frac{1}{\tau_{\text{pre}}\tau_{\text{post}}} \int_{t_{\max}}^{\infty} \exp\left[-\frac{t - t_{\text{pre}}}{\tau_{\text{pre}}}\right] \exp\left[-\frac{t - t_{\text{post}}}{\tau_{\text{post}}}\right] \mathrm{d}t \tag{3.3.11b}$$

$$= -\frac{1}{\tau_{\text{pre}} + \tau_{\text{post}}} \left[\exp\left[-\frac{(\tau_{\text{pre}} + \tau_{\text{post}})\,t - \tau_{\text{post}} t_{\text{pre}} - \tau_{\text{pre}} t_{\text{post}}}{\tau_{\text{pre}}\tau_{\text{post}}}\right]\right]_{t_{\max}}^{\infty} \tag{3.3.11c}$$

with $t_{\max} = \max(t_{\text{pre}}, t_{\text{post}})$ and $t_{\min} = \min(t_{\text{pre}}, t_{\text{post}})$.

- Pre- before Post-Spike: $t_{\max} \equiv t_{\text{post}}$:

$$\int_{-\infty}^{\infty} y_i x_i\,\mathrm{d}t = \frac{1}{\tau_{\text{pre}} + \tau_{\text{post}}} \exp\left[-\frac{\Delta t}{\tau_{\text{pre}}}\right] \tag{3.3.12a}$$

- Post- before Pre-Spike: $t_{\max} \equiv t_{\text{pre}}$:

$$\int_{-\infty}^{\infty} y_i x_i\,\mathrm{d}t = \frac{1}{\tau_{\text{pre}} + \tau_{\text{post}}} \exp\left[+\frac{\Delta t}{\tau_{\text{post}}}\right] \tag{3.3.13a}$$

The overall weight change of the inhibitory synapse is

$$\Delta W^{\text{inh}} \propto -\frac{1}{\mu} + \frac{1}{\mu^2\tau} \left(\Theta(\Delta t) \exp\left[-\frac{|\Delta t|}{\tau_{\text{pre}}}\right] + \Theta(-\Delta t) \exp\left[-\frac{|\Delta t|}{\tau_{\text{post}}}\right]\right) \tag{3.3.14a}$$

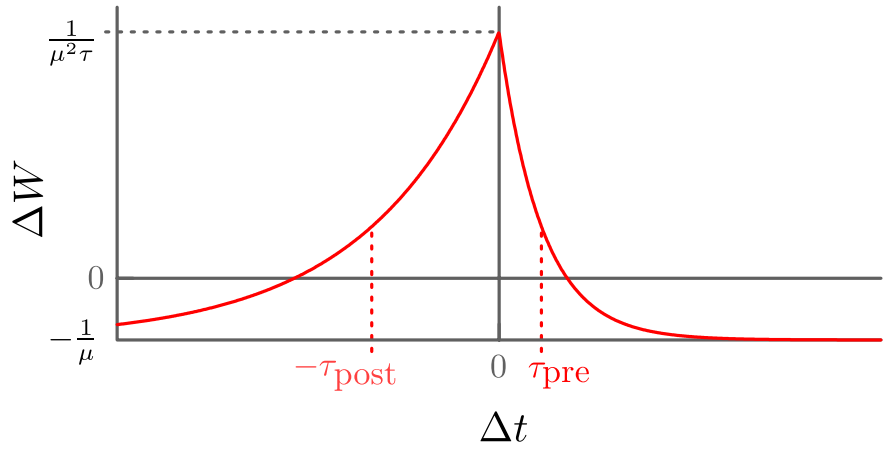

Figure 3.8 STDP for an inhibitory synapse from the energy objective. It contains a constant depression term $(-\frac{1}{\mu})$ for every presynaptic spike and an STDP term with potentiation only.

with $\tau = \tau_{\text{pre}} + \tau_{\text{post}}$. Every presynaptic spike leads to a depression of the inhibitory synapse with an additional potentiation if the presynaptic spike is close in time to the postsynaptic spike, i.e. $|\Delta t|$ is small. This is shown in Figure 3.8.

Such an inhibitory STDP rule was proposed and successfully applied in a feedforward inhibition network by Vogels et al. [2011]. They show that such a learning rule can achieve a balance between excitation and inhibition. Their work was motivated by experimental findings regarding the role of inhibitory plasticity in restoring such balance. While their learning rule is based on a purely phenomenological approach, this STDP rule for inhibitory plasticity directly follows from the energy/noise objective as proposed above. Given that both excitatory and inhibition plasticity rules can be derived from the same objective, this thesis provides a new way to analyze the interaction between those two.

3.4 Network simulations

The focus of this Chapter is on deriving common learning rules for artificial neural networks from an objective function. Here, I demonstrate the stability of the derived learning rules within a recurrent network and their impact on the network structure. The simulated network consisted of a reservoir with 40

excitatory neurons x_r and an input I of size 10:

$$x_r(t+1) = f_{a,b}\left(W_{\text{ff}}\, I(t) + W_{\text{rec}}\, x_r(t)\right). \tag{3.4.1a}$$

The feedforward weights W_{ff}, projecting from the input into the reservoir, were fully connected. The recurrent weights W_{rec} had a connectivity of 25%, i.e. every neuron had on average 10 incoming/outgoing connections. All weights were drawn uniformly at random from the interval $[0, 1]$. The activation function $f_{a,b}$ was a logistic sigmoid and all gains a were initially set to 1. The thresholds b were fixed at 6 and not adapted during learning as discussed in Section 3.2.2.

The learning rules were used according to the combined objective $O_{\text{sep}} + \lambda\, O_{\text{energy}}$ with $\lambda = 0.5\mu^2$ and $\mu = 0.1$. The learning rates were larger than in the stability analysis to speed up the convergence: $\eta_{\text{IP}} = 10^{-2}$ and $\eta_{\text{STDP}} = 10^{-1}$.

3.4.1 Sequence prediction

The reservoir received an input sequence and its performance was evaluated on its ability to predict the next input. One sequence consisted of 10 steps where the n-th input was active during the n-th step. The inactive inputs had a baseline activity with a mean of $A_0 = 0.2$. The mean of the active input was a random variable. It had a reliability of P_{rel} meaning that with probability $1 - P_{\text{rel}}$ it was equal to A_0 and, thus, indistinguishable from the background:

$$I_j^* = \begin{cases} k \cdot A_0 & : j \text{ is active} \\ A_0 & : j \text{ should be active, but failed with probability } 1 - P_{\text{rel}} \\ A_0 & : j \text{ is inactive} \end{cases} \tag{3.4.2a}$$

with the signal-to-noise parameter k. The inputs were iid Gaussian distributed as $I = \mathcal{N}\left(I^*, \sigma^2\right)$ with standard deviation $\sigma = 0.14$. The two parameters k and P_{rel} were varied to change the task difficulty.

Performance

The prediction performance was evaluated by training a supervised readout W_{out} on the reservoir state x_{r} over 20,000 iterations of the input sequence. It

was optimized with a quadratic loss function against the noiseless mean input I^*:

$$W^*_{\text{out}} = \min_{W_{\text{out}}} |I^* - W_{\text{out}} x_{\text{r}}|^2 \tag{3.4.3a}$$

The index with the maximum output of the readout was taken as the predicted next active input. The number of steps for which the input was not correctly predicted was averaged over 20,000 iterations. This error rate E of the network after self-organized learning was compared to a reference value E_0 from a supervised readout trained only on the input. The normalized performance measure was $1 - \frac{E}{E_0}$.

Figure 3.9 shows the normalized performance for different values of the reliability P_{rel} as a function of the signal-to-noise parameter k. For an untrained network (blue diamonds) the performance was at the level of the reference which was only based on the input. Thus, the supervised readout could not make use of the recurrent information in the reservoir state. I assume that this is due to the initial gain of the activation function which does not fit with the input distribution. As shown in Figures 3.10 and 3.11, the gain quickly drops by a factor of 2 when it is adapted. Surprisingly, the performance of the untrained network was even below zero (worse than the reference) for an input with 100% reliability (upper panel) and a large signal-to-noise ratio. This might be due to the performance of the reference, which can achieve a quite high prediction rate for this case since hardly any recurrent information is needed. Thus, the untrained recurrent network might actually loose information due to saturation of the poorly adapted activation function.

When the network was trained with IP only (IP: black circles), the performance increased strongly. It seems that for such a simple task retaining some recurrent information by avoiding a saturation with IP in the reservoir was already very beneficial. The additional adaptation of the feedforward weights (IP+FF: green triangles) lead to an improvement compared to IP alone. Training also the recurrent weights (IP+FF+REC: red squares) concurrently with IP and feedforward weights, the network was able to increase its performance further for high and intermediate reliabilities. When the input sequence was hard to detect (i.e. for low reliability and/or low signal-to-noise ratio) the additional training of the recurrent weights actually decreased the performance

compared to IP+FF. But the performance for IP+FF+REC was always better or equal than IP alone.

The performance shown in Figure 3.9 is the median (50% have a better performance) with the error bars corresponding to lower (25%) and upper quartile (75%) for $n = 48$ random initializations. In general, the distribution of the performance over different initial conditions was much broader when the weights were trained as compared to IP alone. Among the fully trained networks (IP+FF+REC), there were some very good and some very poor ones. Especially in the case of 100% reliability and $k = 3$, the best fully trained network had no prediction errors over 20,000 iterations. In contrast, the best network trained with IP alone had 2.6% wrong predictions. On the other hand, the worst fully trained network made 10.5% wrong predictions compared to only 3.5% for the worst network trained with IP alone. The average performance of the fully trained network, therefore, turned out to be worse at least for some parameters, due to those few initial conditions converging to very suboptimal states.

Convergence

All simulations were run until the feedforward as well as the recurrent weights converged to stable values (cf. Fig. 3.10 and 3.11). The convergence duration depended on the parameters. For a clear, distinguishable sequence (high P_{rel} and/or high k) the weights converged fast, while for a sequence that was hard to detect (low P_{rel} and/or low k) the convergence was slower. This was also reflected in the relative strength of feedforward vs. recurrent weights.

Strong recurrent weights evolved for a clear input sequence leading to a good performance. This is shown exemplary for one initial condition in Fig. 3.10. In this case ($P_{\text{rel}} = 0.85$, $k = 4$) the recurrent weights were dominating and only four feedforward input weights remained. Although the feedforward weights were suboptimal since not every input was connected to the network, the performance is still close to optimal and better than IP-only or IP+FF. Thus, the reservoir has adapted to the sequence such that it was able to predict several steps ahead and compensated the missing feedforward connections.

On the other hand, when the sequence was hard to detect, the network was dominated by the feedforward weights. Fig. 3.11 shows an example for such a

case ($P_{\text{rel}} = 0.85$, $k = 2$). In contrast to the other case of high signal-to-noise ratio, the feedforward weights represented all possible inputs: every neuron in the reservoir was connected to exactly one input. But since the learning rule was not able to detect the temporal pattern of the sequence, all recurrent weights eventually became zero. Thus, there was no recurrent information about previous states which decreased the performance. The performance was worse than for the IP+FF case, but still comparable to the IP-only case.

Effect of network size

The introduced network has a very small ratio of reservoir to input neurons (40 : 10) compared to previous works on sequence prediction and memory capacity, e.g. [Lazar et al., 2009] considered up to several hundred excitatory neurons in the reservoir for six inputs. Increasing the reservoir results in a higher-dimensional phase space with more possibilities for separating the representations. Thus, it can expected that the performance of the network improves with the number of reservoir neurons. The learning rules can establish more non-interfering pathways in the recurrent reservoir and the information of previous inputs can be retained more easily. This helps predicting the next input in the case of low reliability or high noise.

Enlarging the reservoir from 40 to 100 neurons (while keeping the in-/out-degree of each one at 10) strongly improved the average performance for various conditions. For example, the performance for $P_{\text{rel}} = 0.70$ and $k = 4$ increased from about 60% to over 90% for the fully trained network. Also in the reliable but noisy case of $P_{\text{rel}} = 1.0$ and $k = 2$, the network was able to improve its performance.

3.4.2 Separability and orthogonalization

The learning rules employed in the network simulations, as derived in the Section 3.2, are based on the functional goal of maximizing the separability between the reservoir states. And, as described in Section 3.3, with the additional energy objective, the learning rule is trying to minimize the dot product making the reservoir states more orthogonal. Figure 3.12 shows that these goals are achieved by the learning rule.

The upper panel shows the squared distance between y_t and $y_{t-2\Delta_t}$ averaged over the dimensions of y and many input presentations as a function of training time. The maximum possible value is 1 but this would correspond to an unfavorable optimum where the reservoir activation keeps switching between a few states. The untrained network was badly tuned to the input and responded with only slight variability. The squared distance between different states in time was basically zero. Adapting only the activation function with IP also only slightly increased the average distance to about $5 \cdot 10^{-4}$ (see inset). Training the feedforward weights lead an increase to about 0.08 after $200,000$ iterations. Additionally training the the recurrent weights provided a further increase to an average squared distance larger than 0.2 after $400,000$ iterations. The results for the one-step distance $(y_t - y_{t-\Delta_t})$ were very similar and only slightly smaller (not shown).

The lower panel shows the average angle between the resevoir activation vectors at y_t and $y_{t-2\Delta_t}$ as a function of training time. The untrained network showed an average angle of 15°. Suprisingly, the intrinsic plasticity decreased the angle to 7.5° after $20,000$ iterations. This also lead to an initial decrease when training the weights. However, after $200,000$ iterations the adaption of the feedforward weights had increased the average angle to about 50°. Also training the recurrent weights raised this value to about 58°. Again, results for the one-step angle $\angle(y_t, y_{t-\Delta_t})$ were very similar. Thus, the synaptic plasticity mechanisms were able to strongly increase the average angle between sucessive reservoir states and, thereby, made the representation of the input more orthogonal.

3.5 Discussion

Neurons show a wide range of different plasticity mechanisms shaping their computational properties. Many learning rules for artificial neurons have been inspired by experimental findings like intrinsic plasticity, STDP, and synaptic scaling. They are employed in the simulation of neural networks in different forms and combinations as a means of self-organized learning. Each learning rule is usually thought of as optimizing a specific computation of the network. And while it can be already complicated to analyze the effect of a single

learning rule, their combinations are even harder to study.

In this Chapter, I introduced an objective function for neural plasticity in terms of the separability of the population activity. I have shown that, from this single objective, learning rules for intrinsic plasticity as well as synaptic plasticity of STDP type can be derived. Both rules have stable fixed points for the standard sigmoid activation function. Thus, explicitly adding a mechanism for synaptic scaling with an ad-hoc normalization constant is not necessary, since the synaptic plasticity is modulated by the derivative of the activation function. Thereby, the weights will adjust to lead to a balanced level of input. The homeostatic effect of this scaling contributes to the improved separability by keeping the neurons from being either inactive or fully active.

This separability objective can be integrated with considerations regarding the energy consumption of spiking and the noise robustness of a low firing rate. By simply adding a second objective function punishing very high and very low firing rates, I extended the approach to account for the exponential-like distribution of firing rates. This second objective additionally lead to a learning rule for inhibitory interneurons. Importantly, the combination of separability and energy objective reflects the goal of orthogonalization as described by Marr [1971].

Thus, this Chapter proposed a simple way to consistently derive a set of standard learning rules for artificial neural networks [Bell and Sejnowski, 1995; Triesch, 2005; Babadi and Abbott, 2010; Vogels et al., 2011] from a common objective function. Even more, the objective is applicable to discrete-time rate coding neurons as well as continuous-time spiking neurons. The effect of the learning rules was shown in a sequence prediction task for a recurrent network. The weight matrices of feedforward input as well as recurrent feedback evolved to a stable and sparse configuration demonstrating the intrinsic stability of the synaptic learning rule. The performance compared to an untrained network improved strongly after applying intrinsic plasticity and even more with an additional training of the feedforward synaptic connections. Concurrently training the recurrent weights, the network was able to raise the performance to 100% for some cases.

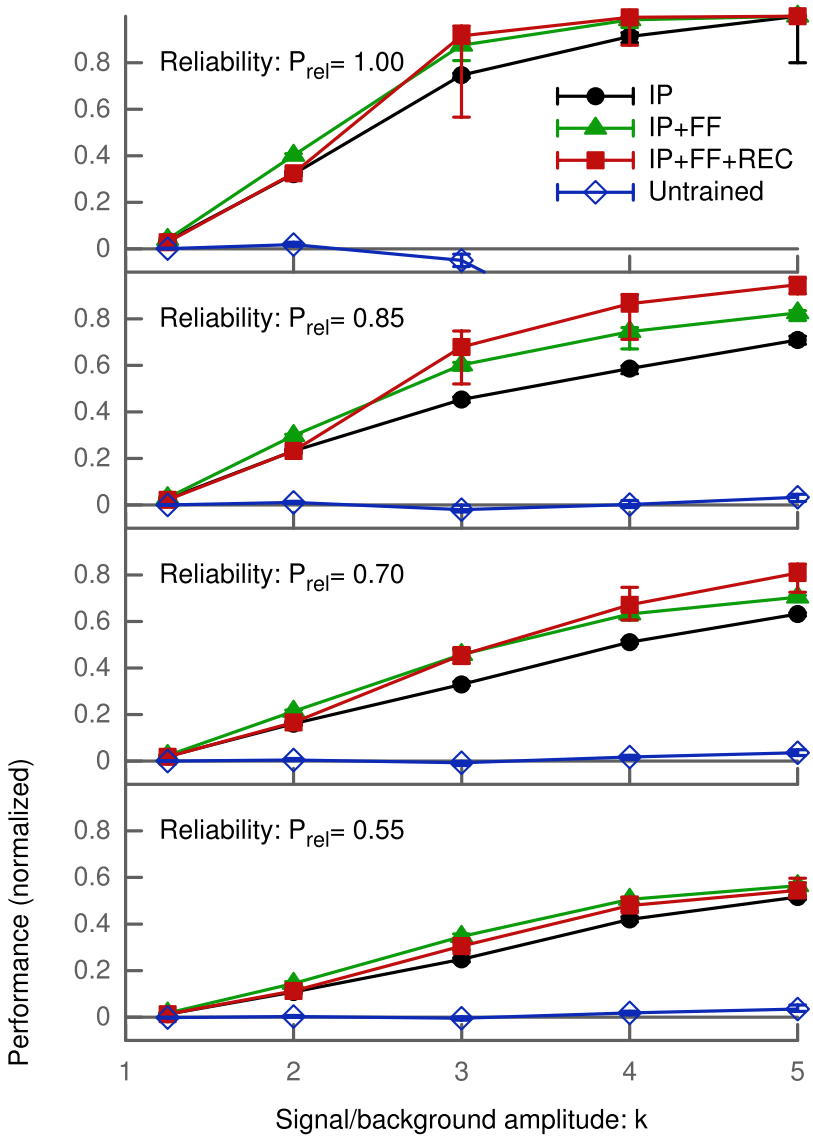

Figure 3.9 Prediction performance of a self-organized recurrent network for a cyclic input sequence. Three combinations of plasticity mechanisms (IP only, IP+feedforward weights, IP+feedforward+reservoir weights) are compared to the untrained network (blue diamonds). A performance of one means perfect prediction (no errors); for a performance of zero the number of prediction errors given by the reservoir state is equal to the number of errors based purely on the input. The task difficulty increases along two dimensions: from top to bottom panel (decreasing reliability P_{rel}) and from right to left (decreasing signal-to-noise parameter k). Shown is the median and the error bars correspond to lower and upper quartile for $n = 48$ trials.

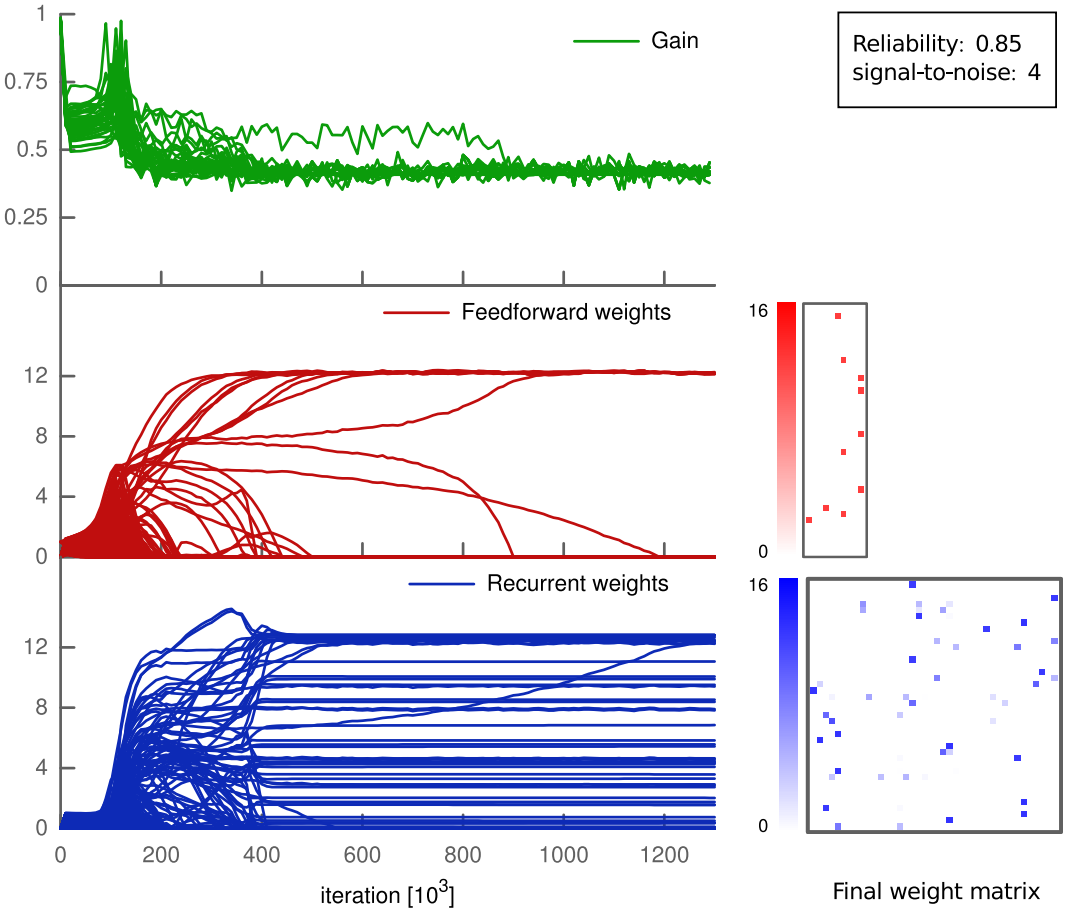

Figure 3.10 Evolution of weights for a high signal-to-noise ratio of 4. **Left:** The values of feedforward (red) and recurrent (blue) weights and gains (green) during learning. After $1.2 \cdot 10^6$ iterations, all weights and gains converged to stable values. **Right:** Final weight matrices of feedforward and recurrent weights after convergence.

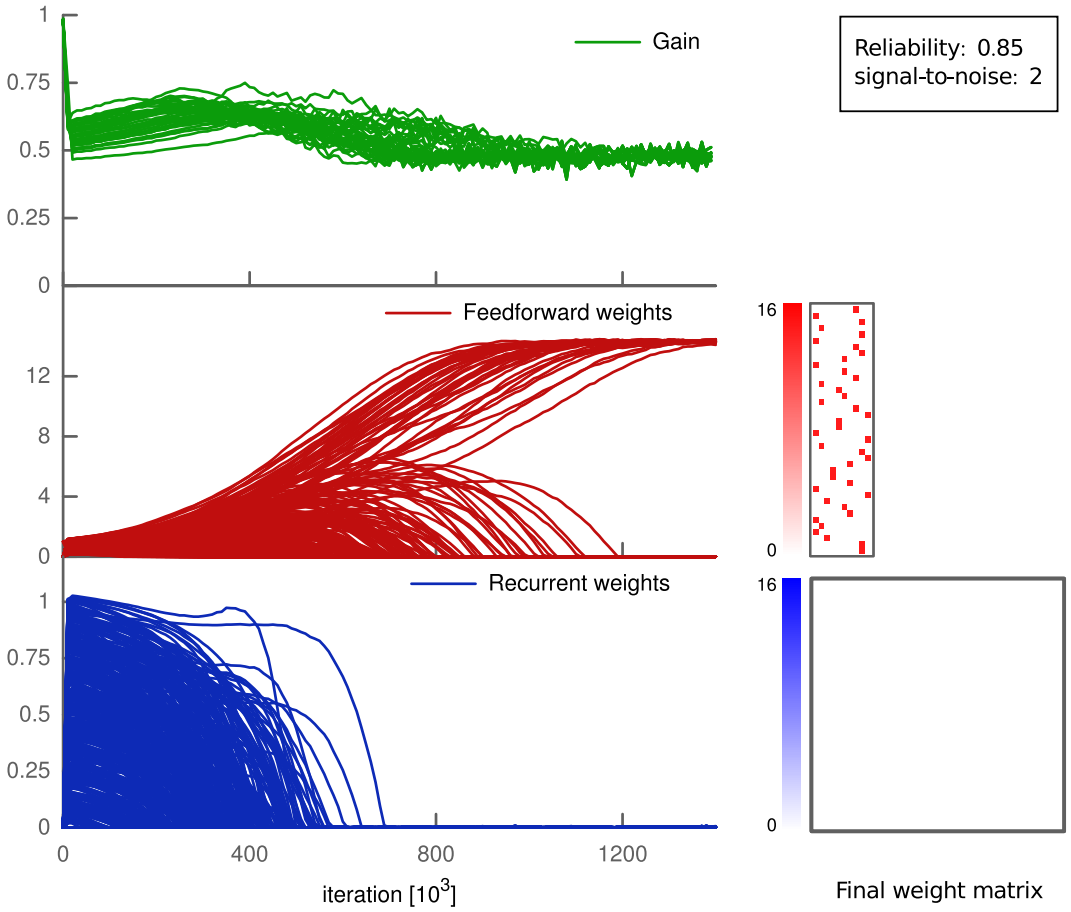

Figure 3.11 Evolution of weights and gain for a low signal-to-noise ratio of 2. **Left:** The values of feedforward (red) and recurrent (blue) weights and gains (green) during learning. After $1.2 \cdot 10^6$ iterations, all weights and gains converged to stable values. **Right:** Final weight matrices of feedforward and recurrent weights after convergence.

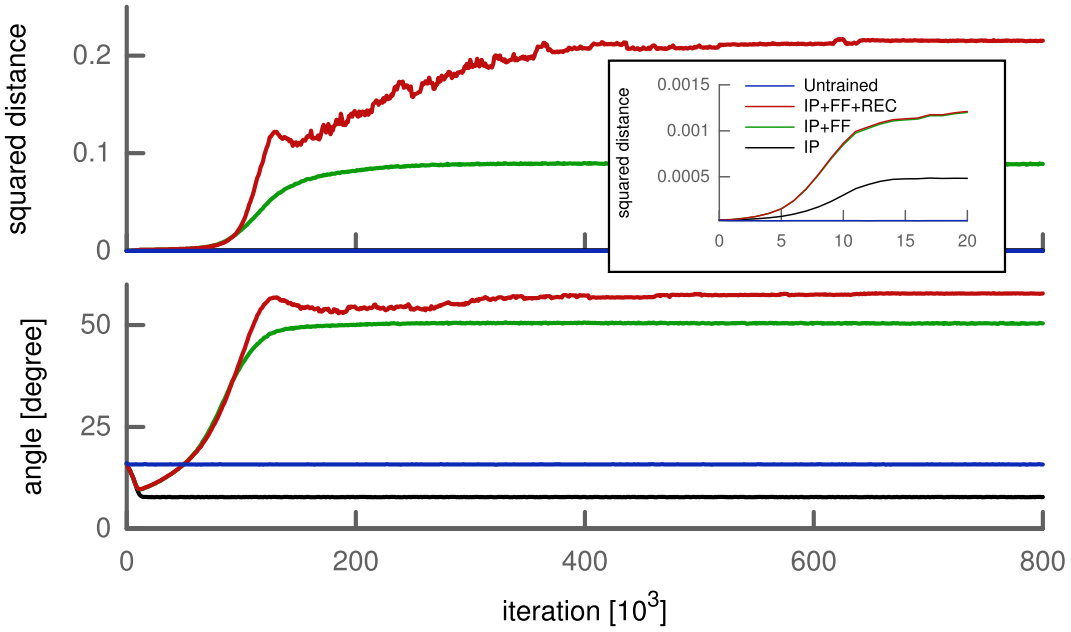

Figure 3.12 Evolution of separability and orthogonality of reservoir states for y_t and $y_{t-\Delta_t}$ as a function of training time. **Upper:** The average squared distance is basically zero in the untrained network (blue) and only increase slightly when using IP (black) (see inset). Training feedforward (green) and recurrent weights (red) leads to a strong increase in the distance. **Lower:** Average angle between the reservoir state vectors. Surprisingly, IP decreased the angle below the value of the untrained network. Training of the weights strongly increased the angle and, therefore, the orthogonality.

Chapter 4

Sparseness objective for synaptic plasticity

In the previous chapter I have shown that a gradient ascent on a very simple objective function leads to various algorithms, which are similar or equivalent to known and widely used plasticity rules for artificial neurons and neural networks. Inspired by these findings, I will apply a modified objective function to more a realistic conductance-based neuron in this chapter [Krieg and Triesch, 2011a,b, 2012 submitted].

Section 4.1 will detail the derivations of the proposed plasticity rule. It introduces the objective function which is based on the sparseness of the distribution of synaptic efficacies. This is based on the idea of minimizing metabolic costs as discussed in Sections 2.3.1 and 2.3.2.

In Section 4.2 the learning rule is simulated in a cell with full morphology using the simulator NEURON. The resulting synaptic changes are studied following the usual experimental protocols and compared to experimental data. The model reproduces results from spike-timing-, rate- and voltage-dependent plasticity and even metaplasticity, thus providing a unifying account of these diverse induction protocols. It also leads to a new prediction regarding the metaplasticity of STDP.

Section 4.3 establishes a connection between the proposed plasticity rule and the biophysical mechanisms causing the plasticity in the real cell. The energy efficiency and further consequence and benefits of the objective function will be discussed.

4.1 Objective function

As introduced in Section 2.2.3, synaptic long-term plasticity describes the bidirectional modifications of synaptic strength. It has a complex dependence on various factors. Among them are direct factors such as correlated pre- and postsynaptic firing rates [Bliss and Lomo, 1973], postsynaptic membrane potential [Artola et al., 1990; Artola and Singer, 1993; Ngezahayo et al., 2000], precise timing of pre- and postsynaptic spikes [Gerstner et al., 1996; Markram, 1997], repetition frequency of such timing [Sjöström et al., 2001], synaptic location [Froemke et al., 2005; Sjöström and Häusser, 2006], as well as indirect (or meta-) factors such as previous postsynaptic activity [Bienenstock et al., 1982; Wang and Wagner, 1999] and initial synaptic strength [Ngezahayo et al., 2000].

Theories and models for many of these factors (especially the timing dependence) have been studied on different levels of abstractions including biophysical [Lisman, 1989; Artola and Singer, 1993; Shouval et al., 2002], phenomenological [Pfister and Gerstner, 2006; Clopath and Gerstner, 2010; El Boustani et al., 2012], and functional [Toyoizumi et al., 2005; Sprekeler et al., 2007; Pool and Mato, 2011]. Nevertheless, synaptic long-term plasticity is induced by a single, complex molecular machinery. Depending on the experimental protocol only different realizations of this process are probed. But so far, most theories only address a specific induction protocol and none of the existing theories bridges the gap between the biophysical and the functional level.

Here, I present a unifying theory that describes long-term synaptic plasticity from a single objective function based on sparseness. I propose that the primary goal of synaptic long-term plasticity is a sparse distribution of synaptic strength and that STDP and other induction protocols can be understood as a consequence of this objective. The synaptic learning rule resulting from this functional goal is formulated in biophysical quantities and relies only on local information. It reproduces results from spike-timing-, rate- and voltage-dependent induction protocols and even metaplasticity, thus providing a unifying account of synaptic long-term plasticity.

While it has been known for years that the distribution of excitatory synaptic strengths is highly skewed [Song et al., 2005; Loewenstein et al., 2011] and

that models of STDP can produce such distributions [van Rossum et al., 2000; Leen and Friel, 2012], here I postulate sparseness as the *objective* of long-term plasticity. I argue that this is beneficial because it reduces metabolic costs and increases coding efficiency. This approach suggests that sparseness may play an even more important role for neural coding than previously thought [Olshausen and Field, 1996; Hromádka et al., 2008].

4.1.1 Approximations

This work rests on the hypothesis that synaptic long-term plasticity is actively maximizing the sparseness of the distribution of synaptic efficacies. While this goal appears rather simple, achieving it is not trivial because a single synapse is severely constrained by its computational abilities and the locality of signals. The aim is a local, computationally simple, and biologically plausible learning rule which explains the experimental data and fits with the mechanistic ideas about synaptic plasticity. To derive such a plasticity rule, constraints and approximations have to be considered which the synapse might be using.

Proxy and sparseness measure

The goal is a plasticity rule which, when applied locally at every synapse, increases the sparseness of the overall distribution of synaptic efficacies. This requires every single synapse to have a way of estimating this sparseness without having knowledge about the full distribution itself. It needs a proxy to gather information about the strengths of other synapses. The derivations in this work will build on a simple, but reasonable candidate: the membrane potential u.

A major assumption is that the limited computational capacity provided by the biophysical mechanisms in the neuron is not sufficient to compute the complex dependencies of the membrane potential and its distribution on the synaptic inputs. A simple approximation in terms of a first order expansion of the voltage dynamics makes the dependence explicit. The time course of the membrane potential is assumed to be linear with additive Gaussian channel noise [Steinmetz et al., 2000] over a short time period δ. Thus, I assume a

Wiener process X_δ with a drift term $\dot{u}(t_0)$:

$$u(t_0 + \delta) \approx u(t_0) + \dot{u}(t_0)\delta + \sigma X_\delta. \tag{4.1.1}$$

The conditional probability density over this time interval is found to be Gaussian as

$$p_\delta(u|t_0 + t) = \mathcal{N}\left(u; \mu(t), \sigma^2 t\right), \tag{4.1.2}$$

with $\mu(t) = u(t_0) + t\dot{u}(t_0)$. The velocity of the membrane potential $\dot{u}$ depends linearly on the total synaptic current $I_{\text{tot}} = \sum_k w_k I_k^{(0)}$.

Next, the measure for the sparseness needs to be specified. A straightforward choice are the statistical measures of skewness S or kurtosis K which are frequently used in learning algorithms such as independent component analysis [Hyvärinen and Oja, 2000] and have been used to derive a BCM-type of learning rule [Blais et al., 1998]. In my approach, both measures lead to a very similar learning rule. I will focus on skewness in the following. To calculate these measures the probability density function, or at least its moments, need to be estimated.

Skewness relation between membrane potential and synaptic efficacies

Given the proxy variable u, I will argue that, in a rough approximation, the skewness of the membrane potential is proportional to the skewness of the synaptic efficacies. The objective, thus, reduces to maximizing the skewness of the membrane potential.

The duration of EPSCs is short (decay times are usually < 10 ms [Burgard and Hablitz, 1993; Takahashi et al., 1995]) compared to the low average firing rate of cortical neurons (reported values of < 0.1 Hz [Margrie et al., 2002] up to $5 - 10$ Hz [Zhu and Connors, 1999; Hromádka et al., 2008]). I may, therefore, neglect the finite values of the EPSCs and just describe the synapses as active or inactive. At every instant in time a small number n of all synapses is "active" meaning a presynaptic spike has arrived. The total synaptic current simplifies to a sum over active synaptic efficacies times some 'average' current $\bar{I}$ as $I_{\text{tot}} \approx \sum_{k \in \text{active}} w_k \bar{I}$. By neglecting correlations between different inputs,

the total synaptic current becomes equivalent to the sum of iid samples w_k drawn from the distribution of synaptic efficacies W.

For given state in phase space $(u, \dot{u})$, the distribution of the membrane potential depends only on the weights w_k of the "active" inputs due to the first-order expansion in (4.1.1). It follows from basic statistics that the skewness of the membrane potential is linear in the skewness of W, since the strengths w_k can be regarded as iid samples from W. In the long-term average, the skewness values are related as $S_u \propto \frac{S_W}{\sqrt{n}}$. Here, n is the average number of simultaneously active inputs.

However, this would only be true for a point-like neuron with no spatial extent. For a neuron with an extended dendritic tree, the membrane potential at a given synapse does not depend on all synaptic currents arriving at that neuron. EPCSs are attenuated as they travel along the tree and the skewness of the local membrane potential only reflects the skewness of the efficacies in some neighborhood of the synapse. While this sounds like a severe limitation, it allows the neuron to separate parts of the dendritic tree. Each branch could have an individual distribution of synaptic strengths and, thereby, provide a preprocessing specific to its inputs. This is reminiscent of the idea of dendritic computation [Branco and Häusser, 2010] and the experimental findings on 'branch strength potentiation' [Losonczy et al., 2008].

Thus, in order to maximize the sparseness of synaptic strength, the aim of the learning rule at each synapse is to maximize the skewness of the membrane potential which is taken as a proxy. It is defined as the third normalized central moment. Since the overall time course of the membrane potential is divided into a sequence of short time bins δ, the moments need to be averaged over δ and an ensemble of bins Σ_δ:

$$S_u = \frac{\langle \langle \hat{u}^3 \rangle_\delta \rangle_{\Sigma_\delta}}{\langle \langle \hat{u}^2 \rangle_\delta \rangle_{\Sigma_\delta}^{3/2}}. \tag{4.1.3}$$

Stochastic gradient ascent and moments

The objective function in form of the skewness is maximized with a gradient ascent. That means, the derivative of the objective with respect to the synaptic

efficacy is used as the learning rule:

$$\dot{w} \propto \frac{\partial S_u}{\partial w} \tag{4.1.4}$$

Using the full expression of the skewness has two implausible requirements: first, the synapse needs to store information about the membrane potential distribution across time and second, it must perform the derivative on this complex expression. A stochastic gradient ascent avoids both problems. Here, the synapse does not estimate the distribution and expectation values of u over an extended period of time. Rather, the average over the bin ensemble Σ_δ is replaced by an instantaneous "sample":

$$S_u^* = \frac{\langle \hat{u}^3 \rangle_\delta}{\langle \hat{u}^2 \rangle_\delta^{3/2}}, \tag{4.1.5}$$

which is the average over the short time period δ. This leads to a random walk on the objective function, which on average converges towards a (local) optimum.

The skewness is measured relative to the mean membrane potential $\bar{u}$ which defines the reference point for the distribution and its moments. The requirement for the stochastic gradient ascent to converge is that $\bar{u}$ is not averaged within each "sample" bin. It needs to be estimated over a longer period of time:

$$\bar{u} = \langle u \rangle_{\tau_{\bar{u}} \gg \delta} \tag{4.1.6}$$

This long-term averaging of $\bar{u}$ is important. I will take it to be a low-pass filtered version of the voltage dynamics and compute it via an exponentially weighted average

$$\frac{\partial \bar{u}}{\partial t} = \frac{u - \bar{u}}{\tau_{\bar{u}}}. \tag{4.1.7}$$

To summarize the introduced approximations: the sparseness of the distribution of synaptic efficacies within a certain neighborhood (e.g. the dendritic branch), measured with the normalized higher-order moment of skewness, is estimated using the local membrane potential as a proxy. Independent presynaptic firing has been assumed to linearly relate the skewness of both distributions. For biological plausibility, the sparseness of the membrane potential

is maximized in a computationally simple way by applying a stochastic gradient ascent. The stochastic "samples" are averaged over a very small time interval during which the dynamics of the voltage can be approximated by a first-order expansion. The mean membrane potential, however, is averaged over an extended time period.

Calculating the "sample" skewness from (4.1.5) amounts to evaluating the moments of the proxy variable over δ. The moments of the mean-free membrane potential $\hat{u} = u - \bar{u}$ can be done analytically due to the Wiener approximation from (4.1.2). The n-th moment is found as

$$\begin{aligned}
\langle \hat{u}^n \rangle_\delta &= \int \mathrm{d}u \, \frac{1}{\delta} \int_0^\delta \mathrm{d}t \; p(u|t+t_0) \, (u-\bar{u})^n \\
&= \frac{1}{\delta} \int_0^\delta \mathrm{d}t \; \int \mathrm{d}u \, p(u|t+t_0) \, (u-\bar{u})^n \\
&= \frac{1}{\delta} \int_0^\delta \mathrm{d}t \; \underbrace{\int \mathrm{d}u \, \mathcal{N}\left(u; \mu(t) - \bar{u}, \sigma^2 t\right) u^n}_{=M_n}, \qquad (4.1.8)
\end{aligned}$$

where M_n is just the n-th raw moment of a Gaussian, since the order of integration can be interchanged according to Fubini's theorem. The moments $\langle \hat{u}^n \rangle_\delta$ follow as

$$\left\langle \hat{u}^2 \right\rangle_\delta = \hat{u}^2 + \frac{1}{2} \left(\sigma^2 + 2\hat{u}\dot{u}\right) \delta + \frac{1}{3} \dot{u}^2 \delta^2 \qquad (4.1.9a)$$

$$\left\langle \hat{u}^3 \right\rangle_\delta = \hat{u}^3 + \frac{3}{2} \left(\hat{u}\sigma^2 + \hat{u}^2\dot{u}\right) \delta + \left(\dot{u}\sigma^2 + \hat{u}\dot{u}^2\right) \delta^2 + \mathcal{O}\left(\delta^3\right) \qquad (4.1.9b)$$

$$\left\langle \hat{u}^4 \right\rangle_\delta = \hat{u}^4 + \left(3\hat{u}^2\sigma^2 + 2\hat{u}^3\dot{u}\right) \delta + \left(\sigma^4 + 4\hat{u}\dot{u}\sigma^2 + 2\hat{u}^2\dot{u}^2\right) \delta^2 + \mathcal{O}\left(\delta^3\right). \qquad (4.1.9c)$$

4.1.2 Stochastic gradient ascent

The synaptic efficacy (or weight) w is a single number characterizing the strength of the synapse. It is the result of a combination of different factors, e.g. released neurotransmitter, number of receptors, and single channel conductance of the receptors. Thus, the weight is a function of different factors x_i: $w = w(x_1, .., x_m)$. These are the final variables inducing plasticity which,

in order to maximize the skewness, follow the differential equation

$$\dot{x}_i := \eta_i \frac{\partial S_u^*}{\partial x_i} \tag{4.1.10}$$

$$= \eta_i \frac{\partial w}{\partial x_i} \frac{\partial \dot{u}}{\partial w} \frac{\partial S_u^*}{\partial \dot{u}}, \tag{4.1.11}$$

with learning rates η_i. The gradient decomposes into three terms since the skewness only depends on $\dot{u}$ which depends on w which in turn depends on x_i.

The first term $\frac{\partial w}{\partial x_i}$ depends on how the weight and its regulating mechanisms/variables are modeled. I will restrict myself to the common postsynaptic factors of number and single channel conductance of the receptors. For a glutamatergic synapse, there are two main types R of receptors (NMDA and AMPA) each having an average maximal single channel conductance g_R and a total number N_R. The synaptic current of a given receptor type R is given by

$$I_R(t) = \underbrace{N_R\, g_R}_{w_R}\ \underbrace{g_0^R(t)\,(E_R - u)}_{I_R^{(0)}(t)}, \tag{4.1.12}$$

with the resulting weight w_R and the normalized current of a single receptor $I_R^{(0)}(t)$.

Each quantity will be optimized simultaneously with individual learning rates:

$$\dot{N}_R := \eta_N \frac{\partial w_R}{\partial N_R} \frac{\partial \dot{u}}{\partial w_R} \frac{\partial S_u^*}{\partial \dot{u}}, = \eta_N g\ \Omega \tag{4.1.13}$$

$$\dot{g}_R := \eta_g \frac{\partial w_R}{\partial g_R} \frac{\partial \dot{u}}{\partial w_R} \frac{\partial S_u^*}{\partial \dot{u}}. = \eta_g N\ \Omega \tag{4.1.14}$$

The common expression $\Omega = \frac{\partial \dot{u}}{\partial w} \frac{\partial S_u^*}{\partial \dot{u}}$ does not depend on N and g. The coupling of the differential equations depends on the learning rates η_g and η_N. They determine the behavior of the weight change

$$\dot{w}_R = g_R \dot{N}_R + N_R \dot{g}_R. \tag{4.1.15}$$

Here, the learning rates in all simulations were set such that the weight change was multiplicative:

$$\dot{w}_R = 2\sqrt{\eta_g \eta_N} w_R \Omega. \tag{4.1.16}$$

An additive learning rule can be obtained with a different setting of the learning rates but leads to qualitatively similar results. The derivations for the dependence of w_R on the learning rates is described in Appendix A.2.

The second term $\frac{\partial \dot{u}}{\partial w}$ is simply linear in the input current for a standard conductance-based neuron model. The derivative of the membrane potential depends linearly on the synaptic current $I_{\text{syn}}(t) = \sum_R w_R I_R^{(0)}(t)$ and thus

$$\frac{\partial \dot{u}}{\partial w_R} = \frac{I_R^{(0)}(t)}{C}, \tag{4.1.17}$$

where $I_R^{(0)}(t)$ is the single channel current through a receptor of type R and C is the membrane capacitance. While the current through a single channel $I_R^{(0)}$ is a quantity that can not be determined by a synapse, the overall plasticity rule depends on the total synaptic current for the multiplicative case from (4.1.16).

The third term establishes the connection to the sparseness measure and is therefore the one which is fundamental to the approach. Using the identities from (4.1.9a)–(4.1.9c), I take the derivative of S_u^* with respect to $\dot{u}$ and expand to first order in the small time interval δ:

$$\frac{\partial S_u^*}{\partial \dot{u}} \propto \frac{\left(\frac{1}{4}\dot{u}\hat{u} - \frac{1}{2}\sigma^2\right)|\hat{u}^3| + \mathcal{O}(\delta)}{\hat{u}^6 + \mathcal{O}(\delta)}. \tag{4.1.18}$$

For $\hat{u} \to 0$ the higher-order terms in δ can not be neglected. Instead of using the full gradient expression, I introduced a simple regularization parameter γ. It is a constant phenomenological parameter which represents the higher-order terms and thereby prevents the expression from diverging:

$$\frac{\partial S_u^*}{\partial \dot{u}} \approx \frac{1}{4}\frac{(\dot{u}\hat{u} - 2\sigma^2)|\hat{u}^3|}{\hat{u}^6 + \gamma^6} \tag{4.1.19}$$

$$\xrightarrow{\hat{u} \gg \gamma} \frac{1}{4}\frac{\dot{u}\hat{u} - 2\sigma^2}{|\hat{u}^3|}. \tag{4.1.20}$$

Differential Hebbian learning

The final plasticity rule with all three terms combined, for skewness as the objective function, is

$$\dot{w}_R := \eta \frac{1}{4}\frac{I_R(t)}{C}\frac{(\dot{u}\hat{u} - 2\sigma^2)|\hat{u}^3|}{\hat{u}^6 + \gamma^6}. \tag{4.1.21}$$

Using kurtosis as the sparseness measure results in the similar expression

$$\dot{w}_R := \eta \frac{2}{3}\frac{I_R(t)}{C}\frac{(\dot{u}\hat{u} - 1.5\sigma^2)\hat{u}^5}{\hat{u}^8 + \gamma^8}. \tag{4.1.22}$$

Here, all remaining constant factors and parameters were combined into one parameter: the learning rate $\eta = 2\sqrt{\eta_g \eta_N}\,\delta^2$ has now the dimension of seconds.

The differential equations exhibit a strongly nonlinear dependence on the mean-free membrane potential $\hat{u}$. In both cases the functional form of this dependence is similar. For low depolarizations the noise parameter σ dominates the numerator and the expression is negative. For a positive $\dot{u}$, the numerator will be zero for some intermediate value of $\hat{u}$ and become positive for larger values. The whole expression, dominated by the denominator, approaches zero again for sufficiently large depolarizations. This dependence of the plasticity on the membrane potential resembles the experimental findings on voltage-dependent plasticity as I will show in the next section.

The other important functional dependence is on the correlation between the time derivative of the membrane potential $\dot{u}$ and the synaptic input current I_R. Neglecting the small noise parameter σ, the plasticity rule is proportional to their product as

$$\dot{w}_R \propto \dot{u} I_R. \tag{4.1.23}$$

Thus, it belongs to the class of so-called differential Hebbian rules where the weight change depends on the correlation between presynaptic activity and the derivative of postsynaptic activity [Kosko, 1986]. In contrast to previous work on differential Hebbian learning [Saudargiene et al., 2004; Kolodziejski et al., 2009], this thesis provides a grounded derivation of such a learning rule from an objective function.

4.2 Simulations

In this section the validity of the proposed synaptic plasticity rule has been evaluated by comparing its results and predictions to experimental data. Different dependencies of synaptic plasticity have been probed by applying the standard stimulation protocols as described in the following.

4.2.1 Methods

I used full morphological simulations in NEURON [Carnevale and Hines, 2006]. All simulations were done with a layer 5 pyramidal neuron [Mainen et al., 1995]

Description	Symbol	Value
Learning rate	η	1.5 s
Regularization for $\bar{u}$	γ	10 mV
Noise level	σ^2	0.036 $\mathrm{mV}^2/\mathrm{ms}$
Time constant for $\bar{u}$	$\tau_{\bar{u}}$	30 s

Table 4.1 The phenomenological model parameters.

provided on ModelDB [Hines et al., 2004] (accession number 8210).

The plasticity rule depends on several parameters falling into two categories: The phenomenological model parameters, like the learning rate, were fixed for all experiments. The values are listed in Table 4.1.

The physiological parameters comprise the morphology of the dendritic tree, the synaptic location on this tree, the spine geometry, the initial strength of the synapse, and the contributions of the different receptors types (AMPA and NMDA) as well as their gating kinetics. Those parameters are also relevant in the experimental preparations and can vary between different cell types, brain area, etc.

The dependence of the resulting plasticity on the exact morphology of the tree is beyond the scope of this work. Although the spine geometry is found to depend on the dendritic location [Berard et al., 1981; Jones and Powell, 1969] and influence plasticity [Yuste and Bonhoeffer, 2001], it was also taken to be fixed. The spine consisted of a neck (1 µm long, 0.1 µm thick) and the head (0.6 µm long, 0.3 µm thick) as used by Koch and Poggio [1983].

The time course of the synaptic conductances of a single AMPA/NMDA receptor were modeled by the sum of three exponentials $\hat{g}(t)$ normalized to have a maximum of 1: one exponential for the rising phase (time constant τ_r), two for the decaying phase (time constants τ_{d_1}, τ_{d_2}), and their relative strength λ. The parameter values are listed in Table 4.2.

$$g_0^{\mathrm{AMPA}}(t) = \hat{g}_{\mathrm{AMPA}}(t) \tag{4.2.1a}$$

$$g_0^{\mathrm{NMDA}}(t,u) = \hat{g}_{\mathrm{NMDA}}(t) \left(1 + \frac{[\mathrm{Mg}^{2+}]_o}{\beta} \exp\left[-\alpha u\right]\right)^{-1} \tag{4.2.1b}$$

The additional term accounts for the voltage-dependent Mg^{2+} block of the

Parameter	AMPA [Spruston et al., 1995]	NMDA [Kinney et al., 1994]
τ_r	0.55 ms	4.05 ms
τ_{d_1}	2.0 ms	27.6 ms
τ_{d_1}	8.0 ms	147.4 ms
λ	0.8	0.5
N	2.00	1.36
$\hat{g}(t) = N\Big(\lambda \exp\left[-t/\tau_{d_1}\right] + (1-\lambda)\, \exp\left[-t/\tau_{d_2}\right] - \exp\left[-t/\tau_r\right]\Big)$		

Table 4.2 The parameters for the conductance time course $\hat{g}(t)$ of AMPA and NMDA.

NMDA receptor with $\alpha = 0.062\,\mathrm{mV}^{-1}$, $\beta = 3.57\,\mathrm{mM}$, and the external magnesium concentration $[\mathrm{Mg}^{2+}]_o = 1.2\,\mathrm{mM}$ [Gabbiani et al., 1994]. The postsynaptic action potentials were elicited by current injection of 2 mA for 3 ms at the soma.

Weight and weight change

The total synaptic conductance of a synapse was a weighted sum of both receptor types:

$$g_{\mathrm{total}}(t) = w_{\mathrm{AMPA}} g_0^{\mathrm{AMPA}}(t) + w_{\mathrm{NMDA}} g_0^{\mathrm{NMDA}}(t) \tag{4.2.2a}$$

Both weights w_{R} independently followed the differential Equation (4.1.21) and were changed continuously. The strength of a synapse was measured as the EPSP peak amplitude at the soma. The increase/decrease of synaptic strength was calculated as the relative change of this peak amplitude.

The initial value of the NMDA weight was taken to be independent of the distance to soma with $w_{\mathrm{NMDA}} = 500\,\mathrm{pS}$. The initial AMPA weight w_{AMPA} was increased with distance from the soma as motivated by experimental findings [Andrasfalvy and Magee, 2001]. The resulting total synaptic strength was on the order of $0.1 - 0.2\,\mathrm{mV}$ [Magee and Cook, 2000].

If not explicitly measured by the experiment, the distance of the synaptic location to the soma as well as the weight w_{AMPA} were adjusted to fit the experimental setup and data (see simulations for details).

4.2.2 Weight distribution and skewness

The learning rule was derived from the goal of maximizing skewness. Therefore, I assessed the effect of the learning rule on the distribution of synaptic weights and its skewness. 100 synapses located at 200 µm from the soma received independent presynaptic Poisson spike trains at 3 Hz. The distribution was estimated over the population of 100 synapses for several trials. In each trial every synaptic efficacy was initially set to $w_0 = 0.16$ mV and new random spike trains were sampled over a given duration. I did not employ any saturation effects or hard bounds, therefore, some synapses grew strong enough such that they triggered a postsynaptic spike. Since their strength was measured at the soma (see Methods), this spike strongly biased the measured strength of these synapse. While in principle the computer simulation would also allow to neglect this spike contribution by measuring the strength at the synaptic spine, I applied the same way of measurement throughout this chapter to be in line with usual experimental procedures. To prevent the disortion of the skewness analysis by these nonlinear effects, I excluded all synapses with a strength larger than 10 mV.

In contrast to all other simulations with a multiplicative weight change, this was done also for the additive weight change. The reason for this was that any multiplicative weight change following some distribution ultimately leads to a lognormal distribution according to the central limit theorem and Gibrat's law. And such a lognormal distribution will always exhibit a positive skewness. Thus, the effect of the learning rule might not be distinguishable from random weight changes. On the other hand, for an additive random weight change following some distribution the synaptic efficacies would become Gaussian due to the central limit theorem. A lognormal distribution even for additive changes shows that the objective of an increased skewness is actually achieved by the learning rule.

Figure 4.1 shows the cumulative density function of synaptic weights for two different simulation durations ($T = 30\,\mathrm{s}, 300\,\mathrm{s}$) in log-scale. The cumulative density function, rather than the probability density function, was chosen to eliminate statistical problems with the choice of the histogram bin sizes. The weights from the multiplicative learning rule are well fitted by a lognormal distribution for both durations as expected. But also the ad-

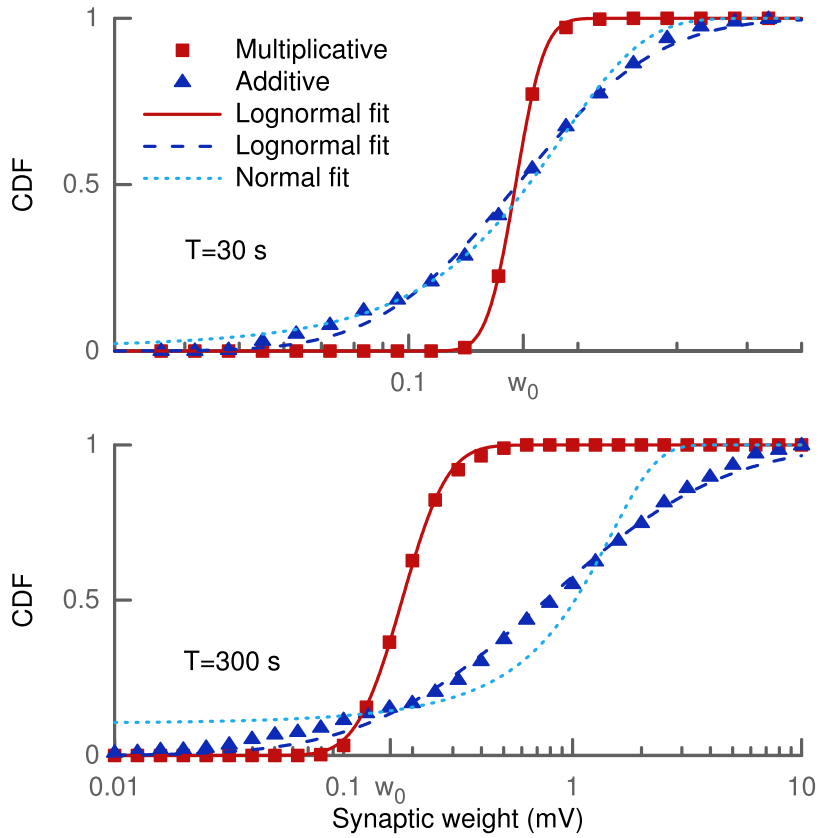

Figure 4.1 Synaptic weight CDF for the additive (blue triangle) and multiplicative (red square) learning rule at two different durations (upper and lower panel). The weights for the 100 synapses are initialized at $w_0 = 0.16\,\mathrm{mV}$. Both cumulative densities are fitted with a lognormal (straight and dashed line). Additionally, a Gaussian CDF (dotted line) is fitted to the weights from the additive learning rule.

ditive learning rule leads to a cumulative density function which follows a lognormal distribution. For the short simulation of 30 s (upper panel) the Gaussian fit is only slightly worse in terms of the root mean square of the residuals (RMSR) ($RMSR_{\mathrm{lognormal}} = 0.0209$, $RMSR_{\mathrm{Gaussian}} = 0.0238$), but it becomes worse as the weight change progresses. The weights at 300 s show a slight deviation from the lognormal cumulative density function at low values while the Gaussian does not fit the data very well ($RMSR_{\mathrm{lognormal}} = 0.0227$, $RMSR_{\mathrm{Gaussian}} = 0.0814$).

Figure 4.2 shows the skewness and the mean of the weight distribution as a function of simulation duration. Both types of learning rules show an increase of skewness from the initial delta peak, which is not skewed, to values of around

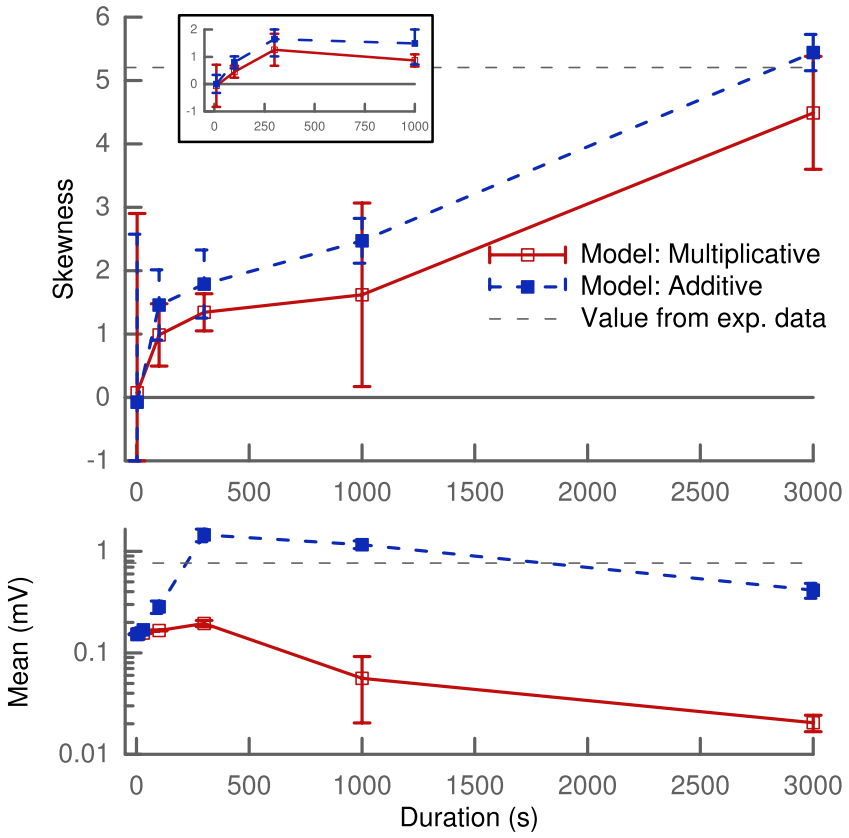

Figure 4.2 Skewness (upper) and mean (lower panel) of synaptic weight distribution. The horizontal dashed lines correspond to the values from Song et al. [2005] calculated from their lognormal fits. Results shown are calculated for one population of 100 synapses with the mean and variance of skewness averaged over several trials.

4 after 5 minutes of simulation. These values are comparable to the value of 5.2 extracted from a lognormal fit to experimental data [Song et al., 2005]. For additive learning the mean synaptic weight increases from 0.16 mV to values around 1 mV. On the other, it drops by a factor of about 10 in the case of multiplicative learning. This is, however, due to the exclusion of very strong synapses with weights larger then 10 mV as discussed above.

4.2.3 STDP: spike pairings

First, I tested if the approach can explain the basic phenomenon of STDP. The main protocol for studying STDP is by pairing pre- and postsynaptic action potentials at different time delays Δ at low repetition frequency. For

a negative delay the postsynaptic spike precedes the presynaptic spike (post-pre) and for a positive delay the order is reversed (pre-post). I simulated 7 spike pairings at 1 Hz for two different synapses at 200 µm ($w_{\mathrm{AMPA}} = 200$ pS) and 400 µm ($w_{\mathrm{AMPA}} = 250$ pS) distance from the soma. The results were in good agreement with the experimental data [Bi and Poo, 1998; Zhang et al., 1998] apart from the LTP at very short delays, which is predicted too strong (Fig. 4.3). This could be accounted for by saturation mechanisms which I have not included so far.

Interestingly, the STDP window depended on the synaptic location. This is related to the broadening of the backpropagating action potential (bAP) along the dendritic tree. The change in the timescale of the bAP decay affects the interaction of post-pre pairs. The width of the LTD window in the present model is therefore increasing with distance from the soma, while the LTP window is relatively independent of the synaptic location. I show this by way of example for two synapses at 200 µm and 400 µm (Fig. 4.3). This feature of the model was also found experimentally [Froemke et al., 2005].

4.2.4 STDP: frequency dependence

The standard STDP protocol consists of spike pairings repeated at low frequencies. Increasing this repetition frequency leads to interactions between spikes of different pairs. The LTP and LTD of these pairings did not add linearly as shown by experimental findings [Sjöström et al., 2001]. Figure 4.4 shows the simulation results for 100 pairings at positive (+10ms) and negative (-10ms) delays as a function of the repetition frequency compared to the experimental data. The modeled synapse was located proximally at 150 µm and only contained NMDA receptors ($w_{\mathrm{AMPA}} = 0$ pS). This is comparable to the experimental conditions where the measured excitatory postsynaptic potential only showed one slow component with decay times of about 50 ms [Sjöström et al., 2001]. At low frequencies below 5 Hz, the model predicted LTD for negative pairings but hardly any LTP for positive pairings. I retrieved the missing LTP, which is in contrast to the results of Figure 4.3, for more distal synapses containing AMPA receptors. LTP for positive pairings increased with frequency and the LTD for negative pairings was converted to LTP for frequencies above 30 Hz. At 50 Hz, where both protocols became

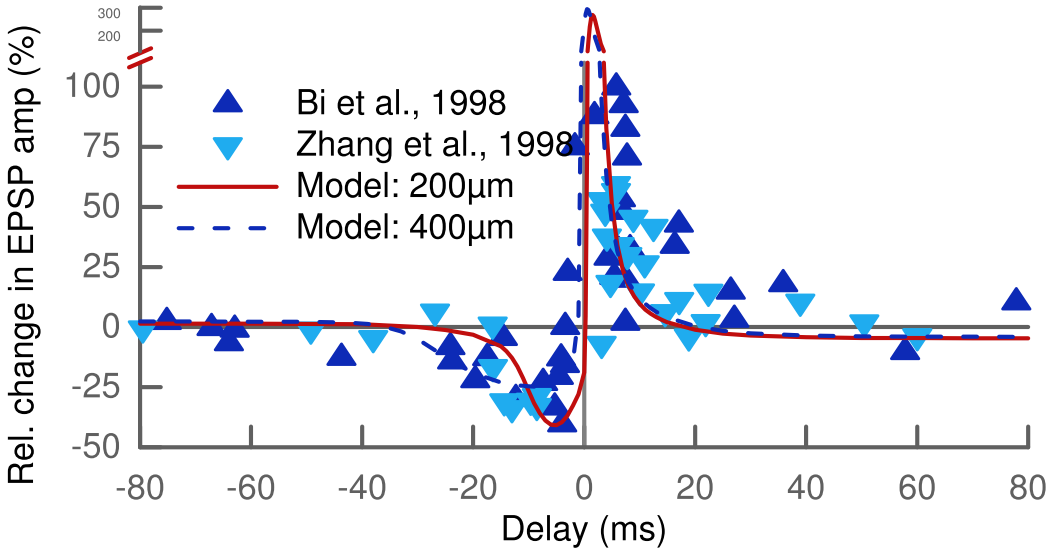

Figure 4.3 Change of synaptic strength in an STDP pairing protocol. The experimental data [Bi and Poo, 1998; Zhang et al., 1998] (triangles) show LTD for post-pre pairs and LTP for pre-post pairs with delays up to 30-40 ms. The results of the model are shown for two different synaptic locations on the apical dendrite: 200 µm (solid line) and 400 µm (dashed line) from the soma. The timescale of the LTD window increases with distance while the LTP window shows no dependence on the location.

equivalent, robust LTP was predicted by the model matching the experimental data [Sjöström et al., 2001] (Fig. 4.4). Thus, the model can account for the nonlinear integration of LTP and LTD when looking beyond isolated spike pairs.

Like for the standard STDP protocol, I observed a dependence on the synaptic location for this frequency-dependent STDP protocol. The robust potentiation at 50 Hz for ± 10 ms pairings was only measured at proximal synapses. It converted to a slight depression for very distal locations more than 700 µm from the soma. This LTP-LTD switch was observed for NMDA-only synapses as well as for synapses with a distance-dependent increase in AMPA receptors (Fig. 4.5). These results of the model are qualitatively similar to experimental findings [Sjöström and Häusser, 2006].

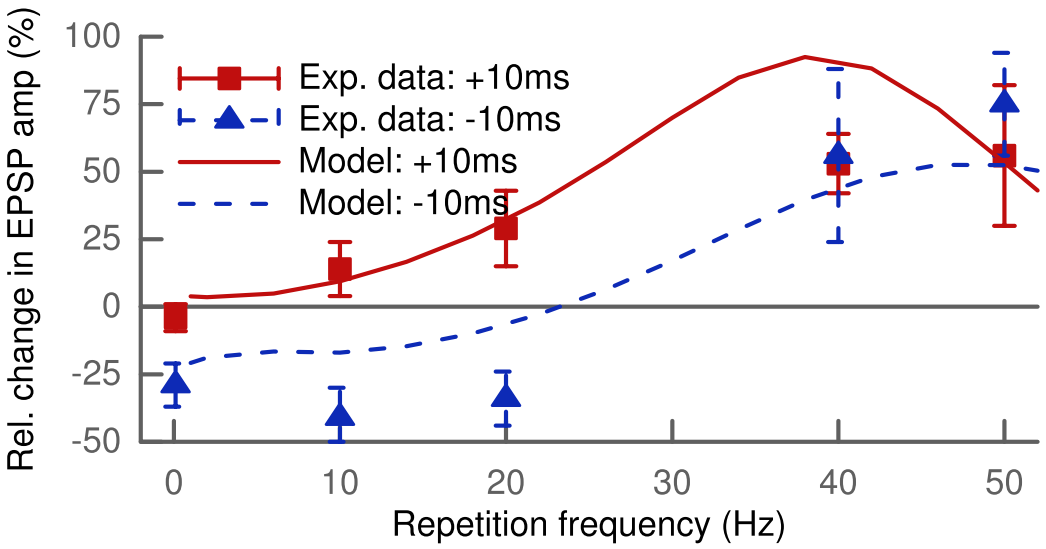

Figure 4.4 STDP pairing depends non-linearly on the repetition frequency. Delays were fixed at +10ms (squares and solid line) or -10ms (triangles and dashed line). The experimental data [Sjöström et al., 2001] show LTP for pre-post at low frequencies up to 20 Hz which increases with frequency, while post-pre pairing leads to LTD with roughly constant amplitude. At frequencies above 30 Hz LTP for pre-post starts to saturate, but post-pre pairings exhibit a crossover from LTD to LTP. At 50 Hz both result in strong LTP, since pre-post and post-pre protocols become equivalent for $\Delta = \pm 10$ms. The model predicts the same behavior for a synapse at 150 µm containing only NMDA receptors ($w_{\mathrm{AMPA}} = 0$ pS).

4.2.5 Rate-dependent plasticity and metaplasticity

Depression and potentiation can also be induced without a controlled timing of pre- and postsynaptic spikes. I simulated this rate-dependent plasticity by independent pre- and postsynaptic Poisson spike trains lasting 30 seconds. The simulated synapse was the same as in the STDP pairing protocol (200 µm distance from the soma with $w_{\mathrm{AMPA}} = 200$ pS). The presynaptic frequency was fixed at 10 Hz and I probed the induced plasticity as a function of the postsynaptic conditioning frequency. The model predicted LTD at low conditioning frequencies and LTP above a given frequency threshold (Fig. 4.6). Furthermore, increasing the presynaptic frequency increased the amplitude of the observed change in synaptic strength (not shown). These results agree with the predictions of the BCM theory [Bienenstock et al., 1982] and also fit

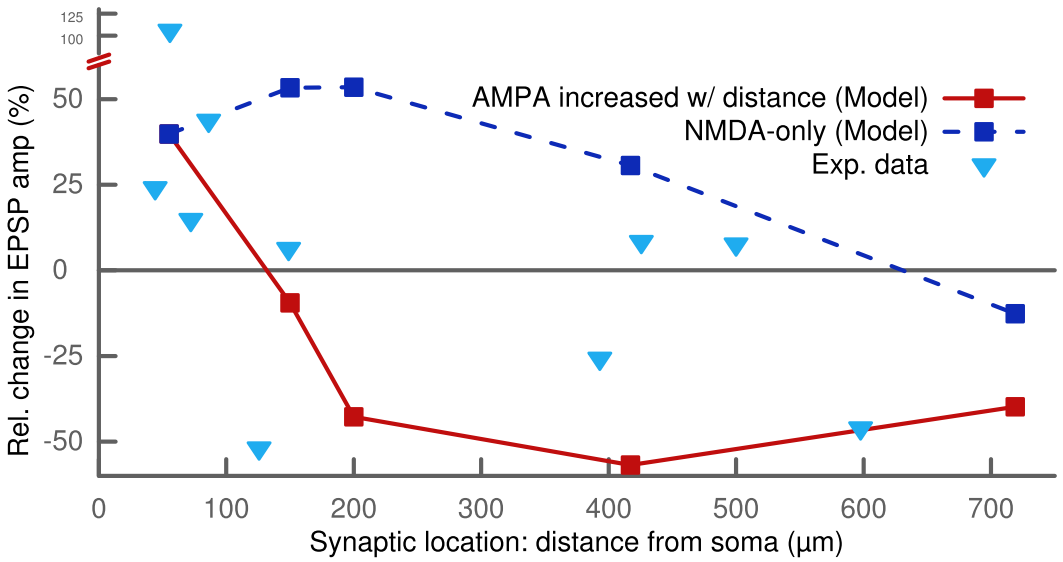

Figure 4.5 STDP pairings with a delay of $+10$ ms at 50 Hz for different synaptic locations. Proximal synapses (< 200 µm) show strong LTP. In the case of no AMPA receptors (dashed line), the potentiation decreases with distance starting at 200 µm and eventually converts to LTD above 700 µm from the soma. Increasing the number of AMPA receptors with distance (solid line) abolishes the potentiation and converts it to LTD. Proximal synapses with a distance to the soma smaller than 200 µm were assumed to be dominated by NMDA receptors. The modeled synapse at 55 µm did not contain any AMPA receptors (see Methods). The experimental data (triangles) show a similar switch from LTP to LTD as a function of distance [Sjöström and Häusser, 2006].

the data of its experimental test [Wang and Wagner, 1999].

Another prediction of the BCM theory is the so-called metaplasticity, where the history of postsynaptic activity influences the observed plasticity. Because of the learning rule's dependence on the average memrane potential, I hypothesized that the model would also show a form of metaplasticity. To test this hypothesis, the cell is primed with a 100 Hz stimulus [Wang and Wagner, 1999] which I applied for 3 seconds prior to the conditioning. The high postsynaptic activity increased the average membrane potential $\bar{u}$ at the synapse due to the bAP. This led to a larger amount of LTD (Fig. 4.6) which replicated experimental findings [Wang and Wagner, 1999]. This homeostatic effect is of heterosynaptic nature, since the bAP influences $\bar{u}$ at all synapses.

So far, metaplasticity has only been studied for the above mentioned rate-

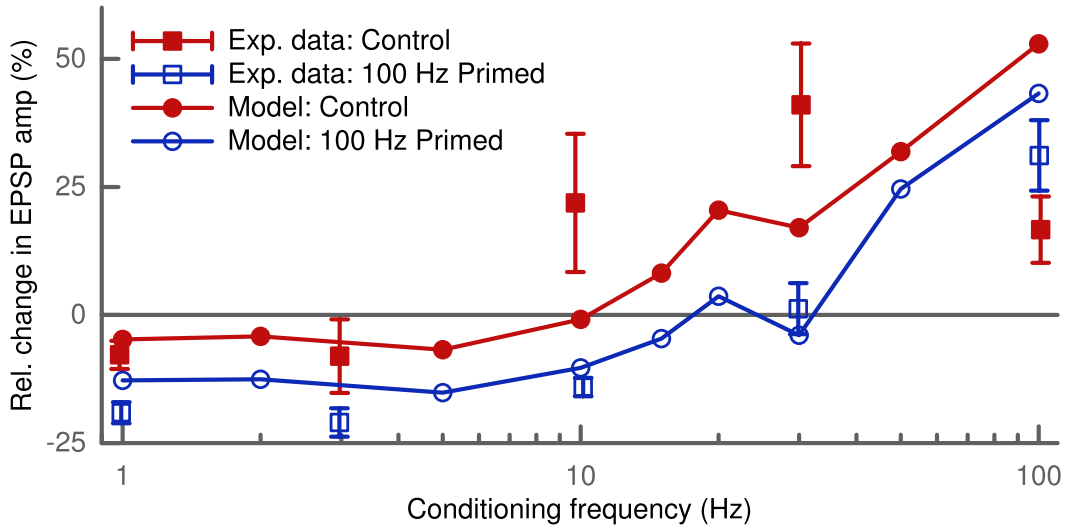

Figure 4.6 BCM-like metaplasticity depends on recent postsynaptic activity. The experimental data [Wang and Wagner, 1999] in the control conditions (filled squares) reveal slight LTD at low conditioning frequency (< 5 Hz) and strong LTP above. After priming the postsynaptic cell with a 100 Hz stimulus (empty squares) the LTD window becomes larger (up to 30 Hz) and stronger (amplitude doubles). In the model the control case (filled circles) corresponds to an average postsynaptic activity history of 10 Hz, while in the priming case (empty circles) the neuron was stimulated with 100 Hz for 3 seconds before the conditioning. Upon priming the model displays the same leftward shift of LTD/LTP crossover point and an increase in LTD amplitude.

dependent protocol. The level of abstraction used in the model does not allow to make novel predictions in a quantitative way. But combining STDP with a priming protocol for metaplasticity results in a clear qualitative prediction. The model predicts that the width of the LTD window in an STDP protocol will be smaller if it is preceded by strong priming (Fig. 4.7). This prediction is a fundamental consequence of the learning rule given the influence of the priming, namely raising the mean membrane potential $\bar{u}$. It is also not depending on the choice of skewness as the sparseness measure, but will be valid also for kurtosis or any other higher moment.

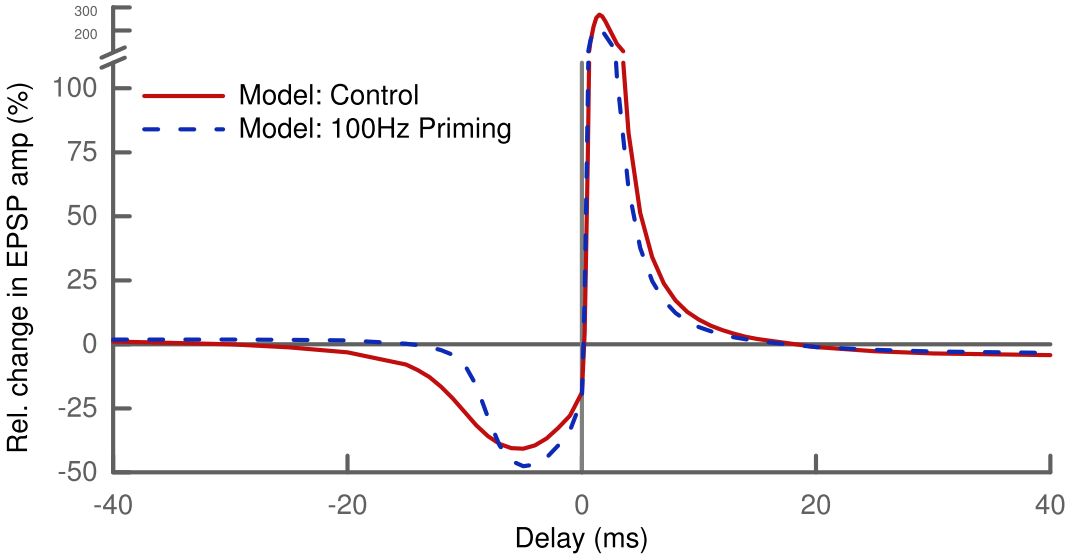

Figure 4.7 STDP pairing window depends on recent postsynaptic activity. Priming the postsynaptic cell with a 100 Hz stimulus (dashed line) alters the amplitude of LTD and LTP slightly. Most notably, the priming decreases the width of the LTD window compared to control conditions (solid line). The decay time to 5% of the maximum LTD decreases from 17 ms to 7 ms.

4.2.6 Voltage-dependent plasticity

Finally, I tested if the model can also explain results from an induction protocol based on clamping postsynaptic membrane voltage. To this end, I paired a presynaptic stimulation with postsynaptic depolarization. I clamped the soma of the postsynaptic cell at different voltages. A distal synapse (700 µm) with only NMDA receptors was stimulated 100 times at a frequency of 2 Hz following the protocol from [Ngezahayo et al., 2000]. No change is induced when clamping the membrane voltage around the resting potential, while low depolarization lead to LTD and higher depolarization resulted in LTP (Fig. 4.8).

The general trend of the results is in line with experimental findings [Artola et al., 1990; Ngezahayo et al., 2000; Feldman, 2000]. But the amount of plasticity was very sensitive to the synaptic location. In the current formulation of the model, plasticity can only be generated for synapses which are either far from the clamping position and/or have a spine with a small isolating neck. There, the influence of the voltage clamp is weaker and $\dot{u}$ is not absolutely zero but can slightly fluctuate. Also the effect of AMPA receptors in the synapse

was quite pronounced. Increasing the number of AMPA receptors led to more LTP (not shown).

In these voltage clamp simulations I also looked at the influence of the phenomenological parameter σ which I had taken as fixed so far. σ was introduced to account for the influence of noise of the voltage dynamics on its sparseness. This noise is partially originating from the stochasticity of ion channels and synaptic receptors. For simplicity, I scaled σ linearly with the synaptic strength. One could also take a linear dependence on the square root of the synaptic strength, e.g. by assuming the variance arise through independent channel noise. However, the qualitative behavior would be the same. Interestingly, this homosynaptic scaling of σ had a homeostatic effect. The LTD window of the depressed synapse was decreased, while it increased for the potentiated synapse (Fig. 4.8). A similar shift in the voltage-dependent plasticity was observed experimentally and also shown to be of homosynaptic nature [Ngezahayo et al., 2000].

4.3 Discussion

The phenomenon of synaptic long-term plasticity is driven by a single complex mechanism and the diverse experimental induction protocols are only probing different aspects of this process. I showed that induction of spike-timing-, rate-, and voltage-dependent plasticity as well as metaplasticity all follow from the single computational idea of a sparse distribution of synaptic strengths. The approach has an intrinsic dependence on the synaptic location which agrees with the experimental findings. I therefore propose the maximization of sparseness as a central goal of synaptic plasticity. The model also leads to a novel prediction of metaplasticity for STDP.

This approach provides a two-fold unification compared to existing models of synaptic plasticity. Detailed models [Lisman, 1989; Artola and Singer, 1993; Shouval et al., 2002; Graupner and Brunel, 2012] addressing the biophysical mechanisms can not make conclusions regarding functional goals or consequences. Previous functional models [Toyoizumi et al., 2005; Sprekeler et al., 2007; Pool and Mato, 2011], however, can not be connected to the biophysical mechanisms, since their learning rule is formulated in abstract ways.

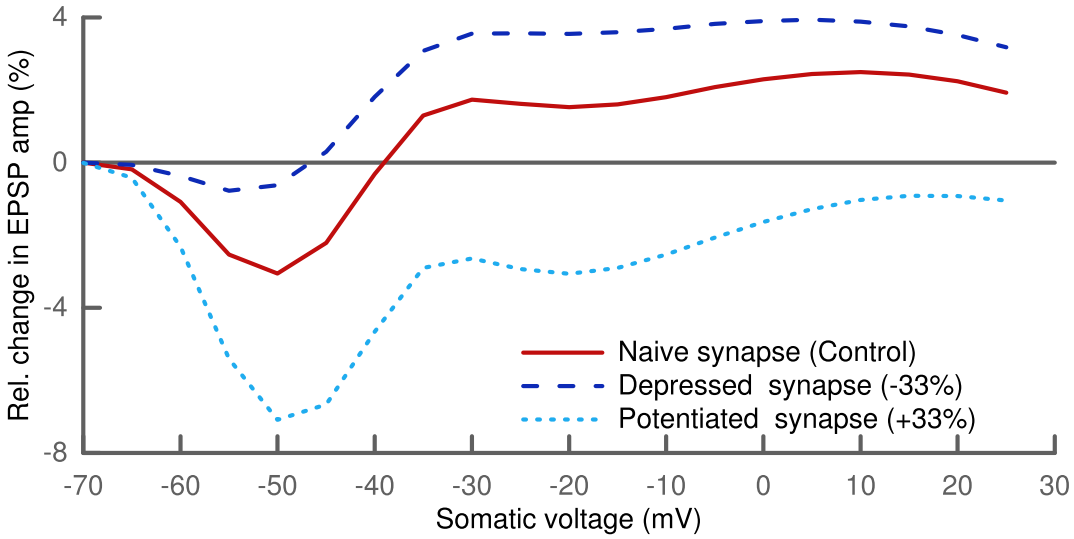

Figure 4.8 Synaptic plasticity depends on postsynaptic depolarization. Simulation of a distal synapse (700 µm from soma) with only NMDA receptors. Pairing presynaptic spikes with postsynaptic voltage clamping at the soma leads to no change for slight depolarization, LTD for intermediate and LTP for strong depolarization (solid line). I scaled the parameter σ linearly with the initial synaptic strength for the depressed and potentiated synapses (see Text). Depressed synapses (dashed line) have a smaller LTD window and potentiated synapses (dotted line) a larger one.

Also they have only addressed one specific induction protocol, namely STDP for spike pairs. In contrast, the functional model I proposed explains the findings from different induction protocols as emerging from one learning rule and it provides a connection to the underlying biophysics by proposing a concrete, biologically plausible learning rule.

Plasticity protocols

A direct consequence of the differential-Hebbian learning rule is the asymmetric shape of the STDP window. Under conditions of an STDP protocol the time derivative of the membrane potential $\dot{u}$ will mainly depend on the bAP from the soma. Thus, plasticity is driven by an interaction of bAP and excitatory postsynaptic current [Markram, 1997]. The location dependence of the STDP window in the model is due to the broadening of the bAP along the dendritic tree. For a pre-post pairing the synaptic current mainly coincides

with the rise of the bAP where $\dot{u}$ is positive, while in the post-pre case it only correlates with the decay where $\dot{u}$ is negative. Such differential-Hebbian learning rules for STDP have already been proposed by Saudargiene et al. [2004]. Their model was motivated by reinforcement learning algorithms [Kolodziejski et al., 2009]. But they did not reproduce other plasticity phenomena than STDP pairing. My approach provides a functional justification for the use of differential Hebbian rules in STDP and reinforcement learning, since this type of learning arises naturally from the computational goal of sparse synaptic strength.

Previous models have proposed a specific triplet interaction for STDP [Pfister and Gerstner, 2006] to explain the results for the different repetition frequencies (Fig. 4.4). The present model does not need any additional mechanism to account for these findings. The robust LTP at high frequencies is due to different timescales of LTP and LTD in the model. Induction of LTP is short but strong due to the rising phase of the bAP, while induction of LTD is weaker but develops over a longer time. Thus, LTD needs more time to build up but its induction is cut off at high frequencies by the bAP of the next pairing repetition. So the contributions for potentiation will dominate the depression and effectively result in LTP. The vanishing of LTP at low frequencies arises due to the kinetics of the NMDA receptors. For a pre-post pairing the synapse is typically near the resting potential on the arrival of the presynaptic spike. Then the NMDA channels are not fully relieved from their Mg^{2+}-block at the bAP arrival and the synaptic current is low. At increasing frequencies the bAP from the previous pairing will provide increasing depolarization. This will gradually relieve the Mg^{2+}-block and lead to stronger synaptic currents. For a post-pre pairing on the other hand, the cell is still depolarized from the bAP. Then the NMDA channel is already unblocked at low frequencies.

The concept of metaplasticity from the BCM theory [Bienenstock et al., 1982] has a direct analog in my model and does not have to be put in explicitly. For BCM, the firing rate threshold regulating the LTD-LTP crossover is a function of the recent postsynaptic activation in order to maintain a homeostatic balance. In the derived learning rule, the average membrane potential $\bar{u}$ naturally accounts for this regulation.

The induction of potentiation and depression fits qualitatively with the

results from voltage clamp experiments. But the plasticity in the learning rule is linear in $\dot{u}$ which is zero at the location of the clamp. Only synapses which are sufficiently far, where the clamp effect is weaker, can experience a visible amount of plasticity. I hope that this shortcoming of the model in the current formulation can be resolved by extending it with the calcium concentration as a proxy for the distribution of synaptic strength (see Future work).

Sparseness objective

I have postulated that the goal of synaptic plasticity is the maximization of sparseness of the synaptic weights. There exist many different ideas regarding the advantages of sparseness and sparse coding (see [Földiak and Young, 1995; Olshausen and Field, 2004] for a review). The use of sparse representations in the brain in terms of the firing rate of neurons has been successfully applied to models of receptive field learning [Olshausen and Field, 1996] and confirmed by experimental findings [Hromádka et al., 2008]. Maybe the most appealing interpretation is in terms of energy efficiency. The brain needs to fulfill its computations with a limited energy budget. On the level of a single neuron it tries to maximize the information transmission for some average firing rate [Attwell and Laughlin, 2001]. Thus, sparse coding can be interpreted as a mechanism for energy-efficient coding given that the generation of action potentials is metabolically costly.

But besides the output of a neuron, also the input in terms of the synaptic efficacies is log-normal distributed [Song et al., 2005; Loewenstein et al., 2011]. Such a sparse distribution of synaptic efficacies is energy-efficient since also the postsynaptic effects of a presynaptic action potential induce metabolic costs [Attwell and Laughlin, 2001]. An excitatory postsynaptic current raises the postsynaptic membrane potential which subsequently has to be brought back to the resting state by restoring ionic concentrations. Given a sparse distribution of synaptic strength, most synapses induce an excitatory postsynaptic potential (EPSP) with low amplitude and low energy consumption, while few synapses induce an EPSP with a high amplitude and high energy consumption. The energy for synaptic transmission will be mainly spent in the few strong synapses, but these have a high probability of inducing a postsynaptic action potential. Thus, the objective of a sparse synaptic strength leads to an

energy-efficient synaptic signaling which may be a relevant factor in nervous system evolution.

A second relevant view on the benefits of a sparse synaptic strength is the need for tuning and input specificity. For a neuron to transmit relevant information, it should be tuned to specific inputs and avoid an unspecific Gaussian distributed input. But, according to the central limit theorem, a sum of n independent random variables always approaches a Gaussian regardless of their distribution as n becomes large. Interestingly, a lognormal weight distribution slows down the convergence towards the Gaussian due to its large skewness (see Text S2).

There are many different possible mechanisms that could achieve the proposed computational goal of sparseness. The present model is a result of the optimization via a stochastic gradient ascent, the specific choice of the sparseness measure, and the use of the membrane potential as a proxy for the synaptic weight distribution. The stochastic gradient ascent is necessary to keep the problem computationally simple. For the sparseness measure I have chosen skewness as a higher-order statistical moment. Similarly, one could apply kurtosis or any higher standardized cumulant which would result in largely similar learning rule. An interesting difference between the cumulants is the dependence on the mean-free potential $\hat{u}$. A learning rule based on skewness (or any odd cumulant) will be negative if $\hat{u}$ is negative, while for kurtosis (or any even cumulant) it does not depend on the sign of $\hat{u}$. Thus, in the context of metaplasticity, a learning rule based on kurtosis does not lead to stronger depression after a 100 Hz priming. These higher-order moments are usually considered instable since they can be very sensitive to outliers. I believe, however, that this is not a problem in my approach since the neuron is a bounded dynamical system. The voltage 'samples' driving the synaptic changes are continuously distributed and have no outliers.

Biological realization and future work

I have discussed the computational and algorithmic level of the approach. But the learning rule can also be related to mechanistic theories of synaptic plasticity. Its nonlinear dependence on the membrane potential fits with the Artola-Bröcher-Singer (ABS) theory [Artola and Singer, 1993] and is quite comparable

to the common U-shaped Ω function characterizing the dependence of synaptic changes on the intracellular calcium concentration $[Ca^{2+}]_i$ [Shouval et al., 2002; Graupner and Brunel, 2012]. There, the voltage dependence of synaptic plasticity has consistently been connected to the surge of $[Ca^{2+}]_i$. In this so-called calcium control hypothesis [Lisman, 1989] it is suggested that different $[Ca^{2+}]_i$ levels have opposing influence on the Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII). This affects synaptic strength, since CaMKII regulates the phosphorylation level of AMPA receptors which determines channel conductance.

Relating the membrane potential in the learning rule with the calcium concentration also fits with the question how the synapse should estimate $\dot{u}$. It was hypothesized that the calcium-binding protein calmodulin, due to its kinetics, could serve as differentiator of $[Ca^{2+}]_i$ [Rao and Sejnowski, 2001]. However, an implementation of the learning rule might not necessarily employ a continuous time derivative at every moment in time. A simple estimate of the correlations between the synaptic current and the time derivative of the voltage (or calcium concentration) could be sufficient. In this respect, it was shown that the amplitude of synaptic change in an STDP protocol is strongly correlated with the initial rate of increase of $[Ca^{2+}]$ over the stimulation period [Aihara et al., 2007].

The strong dependence of the learning rule on the membrane potential is due to the choice of u as a proxy to estimate the synaptic weight distribution. It was chosen since its dependence on the presynaptic input and the synaptic efficacies are easy to formulate. Given the important role of calcium in mechanistic models, an important extension to the approach would be to take the intracellular calcium concentration as a proxy. Formulating the objective in this quantity would allow for a very direct connection to the established mechanistic theory of the calcium control hypothesis. The derivation of a learning rule in this case would involve the dependencies of the calcium concentration on presynaptic input, membrane potential, and release from intracellular stores.

With this extension, also the dependence on NMDA and AMPA receptor strength could be formulated more realistically. So far I assumed that the weight change of each receptor type only depends on its own synaptic current. But synaptic plasticity is calcium-dependent and thus mainly me-

diated by NMDA receptors. Normal AMPA receptors only indirectly affect the calcium concentration due to voltage-dependent calcium channels and the voltage-dependent magnesium block of NMDA receptors. But also the role of calcium-permeable AMPA receptors is just emerging [Man, 2011]. All of these dependencies can be incorporated into the current approach by extending it with a calcium concentration proxy.

A second important step towards a biologically realistic model is to take into account the actual mechanisms which contribute to the weight change. Up to now I have used the simplifying assumption that the number and single-channel conductances of the NMDA- and AMPA-type glutamate receptors are independently adapted as continuous variables. Incorporating expression, saturation, and maintenance of plasticity into the approach could be done with an explicit model of the discrete conductance/phosphorylation states and their transitions.

Also spine neck geometry can influence the results of the learning rule. A strengthened synapse would not only have a large EPSP amplitude at the soma, but also locally in the spine head. This leads to a stronger potentiation since the synaptic current correlates more strongly with the local EPSP. Such a positive feedback could be avoided by regulating the spine neck resistance to achieve a "standardized" EPSP in the spine head [Gulledge et al., 2012].

Finally, it has been shown recently that the metaplasticity threshold is actually independent of postsynaptic action potentials but rather depends on calcium [Hulme et al., 2012]. Since the present model in the current formulation intrinsically accounts for BCM-like metaplasticity as a function of the mean membrane potential, I expect that this influence of calcium be accounted for when using the calcium proxy extension.

Conclusions

In the light of these findings, I propose that synaptic plasticity is arising from the single computational goal of sparse signaling. David Marr proposed to distinguish three levels of theoretical analysis: computational goal, algorithm, and implementation [Marr, 2010]. The proposal of a sparse synaptic strength (computational goal) leads to a differential-Hebbian learning rule (algorithm) which can be mapped onto the biological mechanisms (implementa-

tion), thereby, covering all three levels of analysis. The principle of sparseness would then be even more fundamental to neuronal computation than previously assumed. Given the interpretation in terms of energy efficiency it can be considered a fundamental factor that has shaped the mechanisms of neuronal plasticity during nervous system evolution.

Chapter 5

Conclusions

Neural plasticity is the basis for learning and memory and it allows humans and animals to acquire their cognitive and motor skills. These abilities are still unmatched by any computer algorithm. Therefore, the mechanisms and effects of neural plasticity have been of great interest to neuroscience and other disciplines over the last decades.

The large body of experimental findings and theoretical models especially for synaptic plasticity has contributed a lot to our understanding. The current picture of synaptic plasticity stretches across different dimensions. A multitude of induction protocols are used to study it. Furthermore, synaptic plasticity also interacts with other plasticity mechanisms such as intrinsic plasticity. Thus, in order to further the understanding of the influence of these different dimensions and interactions I argue that it is important to develop a unified view on synaptic plasticity. This requires a theory that is able to describe the seemingly different phenomena within a single model. Furthermore, this theory must not be limited to one level of abstraction but needs to create a connection between the mechanistic and computational level. This thesis approached the problem from the computational side. I proposed objective functions in Chapters 3 and 4 and derived learning rules for neural plasticity from them. In both cases, one single objective was able to account for all major plasticity phenomena. And while the formulations of the objective function differed, both lead to a differential-Hebbian learning rule for synaptic plasticity.

In Chapter 3 the formulations and derivations were done in the context of

simplified artificial neural networks. The proposed separability objective unifies all major forms of neural plasticity and is a step forward in understanding their interactions. It allowed to derive intrinsic plasticity rules as well as synaptic plasticity rules from the same objective function. The objective function and resulting synaptic plasticity rules were applicable to discrete-rate as well as continuous-time spiking neuron models. Furthermore, learning rules for inhibitory as well as excitatory synapses were derived which were also intrinsically stable displaying a form of synaptic scaling.

Usually different learning rules, which are inspired by biological phenomena but only approximately model them, are combined to achieve different optimizations in a neural network. But it is hard to tell if and how they influence each other. The proposed objective function shows that the commonly employed learning rules are actually optimizing a common goal and might not be as distinct and independent from each other as usually thought. This is quite plausible given that plasticity mechanisms have probably evolved over a long time period. They have been shaped in order to work together providing a stable, self-organizing network capable of learning and adaptation.

The focus of Chapter 4 was on a more detailed functional description of synaptic plasticity. The context, compared to Chapter 3, was on the finer level of single neurons and single synapses. The derived learning rule was based on a single objective function and formulated in biophysical quantities. Its simulation results were directly comparable to the experimental data. It explained the findings from all major induction protocols for synaptic plasticity and also captured their location dependence as well as the phenomenon of metaplasticity. Furthermore, it predicted a new form of metaplasticity for STDP.

The objective of a sparse distribution of synaptic strength proposed in Chapter 4 unifies all major induction protocols for synaptic plasticity and bridges the gap between mechanistic and functional levels by providing a biologically plausible learning rule. These different phenomena are arising from a single set of biophysical mechanisms and I have shown that these mechanisms might be optimizing a higher computational goal in terms of energy-efficient signaling

as well as optimized coding.

In sum, I was able to show that it is possible to map different learning rules of neural plasticity in artificial networks onto a common functional goal. Specifically, the sparseness objective for synaptic plasticity can explain all major experimental phenomena and the biophysical mechanisms can be viewed as its implementation. Therefore, it can be conjectured that the mechanisms of neuronal plasticity, which is the most important feature of the brain, have been shaped by common, evolutionary relevant goals in order to optimize various functional aspects of neural computation by local self-organization.

Having a comprehensive theory of neural plasticity, which is not restricted to single phenomena but allows a more unifying description, can provide new insights for brain-inspired algorithms in related disciplines such as computer vision or machine learning. This work shows that studying or employing only a few learning rules, resembling a subset of the observed biological processes, misses important interactions and, therefore, might not provide sufficient insight into the computations of the brain. However, a deeper understanding of the relations between the different forms of plasticity is even more important for developing and improving clinical applications. In contrast to computer algorithms, the arising interactions can not be neglected during pharmacological or other forms of treatment. For example, new forms of brain stimulation such as TMS (transcranial magnetic stimulation) as used in therapy effect a wide spectrum of neuronal properties and it is hard predict all of their consequences. Improved therapeutic protocols and treatments tailored to individual patients rely on a comprehensive understanding of the effects and side-effects.

Theoretical models and simulations are one important part to achieve that goal. While this thesis provided a new view on how neural plasticity can be unified from a functional aspect, the link to mechanistic models is obvious but still too weak. Removing some of the simplifying assumptions in this work would allow for more quantitative predictions. Especially, in contrast to the artificial, standardized protocols used for STDP and other stimulations, the model needs to be able to predict synaptic plasticity also for in-vivo stimulation patterns. Such an improved learning rule derived from functional properties could provide enough realism and predictive power while still being efficient

to simulate compared to models based on complex biological processes. This trade-off between the degree of detail and the complexity of simulations does not only concern research on computational neuroscience but can also affect the clinical application when it comes to individualized treatment.

Appendix A

Mathematical derivations and definitions

A.1 Information theory in a nutshell

Since the brain is a complex device for receiving, processing, and transmitting information, many proposals for its functional principles rely on the framework of information theory [see MacKay, 2003; Hyvärinen et al., 2009, for an introduction]. Even though this theory has been developed in the context of serial and discrete information processing, its concepts have successfully been applied to the massively parallel and event-based neural computation. Central to information theory is the concept of Shannon entropy defined as

$$H(X) = \mathbb{E}_X\left[I(x)\right] = -\sum_{x\in\mathbb{X}} p(x)\log p(x) \tag{A.1.1a}$$

where $\mathbb{X}$ is the set of all messages. X is a random variable over that set with a probability $p(x)$ for an instance (message) x. The entropy is the expectation of the self-information $I(x)$ of each message, i.e. the information gain upon receiving message x. In the context of signal transmission a similar measure called mutual information

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) && \text{(A.1.2a)}\\ &= H(Y) - H(Y|X) && \text{(A.1.2b)}\\ &= H(X) + H(Y) - H(X,Y) && \text{(A.1.2c)} \end{aligned}$$

is defined which measure the information gain for a random variable X given a second random variable Y (or vice versa since the definition is symmetric). If a value of x is completely known after observing y, the conditional entropy $H(X|Y)$ is zero and the mutual information attains its maximum $H(X)$. Over an unreliable (noisy) channel, however, the information loss decreases the mutual information.

Neural communication can be cast in that framework by interpreting it as sending a signal (action potential, firing rate) over an unreliable channel (synapse) and passing it through a nonlinearity (excitability, activation function). The computation in a neural network can be optimized by maximizing the mutual information between each connected pair of neurons. Based on this objective several neural learning mechanism have been proposed for the intrinsic excitability [Bell and Sejnowski, 1995; Stemmler and Koch, 1999; Triesch, 2007] as well as the synaptic connections [Toyoizumi et al., 2005; Pool and Mato, 2011]. Interestingly, the concept of mutual information is closely related to the influential approaches of sparse coding and independent component analysis (ICA).

A.2 Multiplicative vs. additive learning rule

The synaptic weight w_R is modeled as a product of receptor number N_R and single channel conductance g_R. I will omit the receptor type R in the following derivations. Both quantities are optimized simultaneously with individual learning rates following the gradient Equation (4.1.11):

$$\dot{N} := \eta_N \frac{\partial w}{\partial N} \frac{\partial \dot{u}}{\partial w} \frac{\partial S_u^*}{\partial \dot{u}} = \eta_N g \Omega \tag{A.2.1}$$

$$\dot{g} := \eta_g \frac{\partial w}{\partial g} \frac{\partial \dot{u}}{\partial w} \frac{\partial S_u^*}{\partial \dot{u}} = \eta_g N \Omega \tag{A.2.2}$$

The common expression $\Omega = \frac{\partial \dot{u}}{\partial w} \frac{\partial S_u^*}{\partial \dot{u}}$ does not depend on N and g. These two coupled differential equations can be decoupled as

$$\dot{X}_\pm = \lambda_\pm X_\pm \Omega \tag{A.2.3}$$

by introducing $X_\pm = \frac{g}{\sqrt{\eta_g}} \pm \frac{N}{\sqrt{\eta_N}}$ and $\lambda_\pm = \pm\sqrt{\eta_g \eta_N}$. Observing that

$$\frac{\partial (X_+ X_-)}{\partial t} = (\lambda_+ + \lambda_-) X_+ X_- \Omega \equiv 0 \tag{A.2.4a}$$

it follows that N and g are constrained on a hyperbola with "radius"

$$r^2 := \frac{g^2}{\eta_g} - \frac{N^2}{\eta_N} \equiv \text{const.} \tag{A.2.5a}$$

With this constraint the original differential equations can be decoupled as

$$\dot{N} = \lambda\Omega\sqrt{N^2 + \eta_N r^2} \tag{A.2.6}$$

$$\dot{g} = \lambda\Omega\sqrt{g^2 - \eta_g r^2} \tag{A.2.7}$$

with $\lambda := \lambda_+$. The overall synaptic weight w evolves as

$$\dot{w} = g\dot{N} + N\dot{g} = \lambda\left(\sqrt{w^2 + \eta_N g^2 r^2} + \sqrt{w^2 - \eta_g N^2 r^2}\right)\Omega. \tag{A.2.8}$$

For the case of $r^2 = 0$ the hyperbola degenerates to a linear relationship between N and g and the learning rule will be purely multiplicative. Reintroducing the dependence on the receptor type R, it reads

$$\dot{w}_R = 2\lambda w_R \Omega \tag{A.2.9a}$$

In the case $r^2 \to \pm\infty$ only one variable is changed and the other is kept fixed. This leads to a purely additive learning rule. For example keeping g fixed ($\eta_g = 0$) and only updating N will lead to

$$\dot{w}_R = \eta_N g_R^2 \Omega. \tag{A.2.10a}$$

For finite values of r^2 the learning rule will tend to be multiplicative for large w_R and gradually become additive for smaller w_R.

A.3 Effect of skewness on the central limit theorem

The central limit theorem states that the distribution of the sample mean of n i.i.d. random variables converges to a normal distribution as n goes to infinity:

$$\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right) \xrightarrow{d} \mathcal{N}(0,\ \sigma^2)$$

with $E(X) = \mu$ and $E[(X-\mu)^2] = \mathrm{Var}(X) = \sigma^2$.

The Barry-Esseen theorem [Esseen, 1956] attempts to quantify the rate of convergence in the central limit theorem. It provides an upper bound on the absolute difference between Φ (the cdf of the standard normal distribution) and F_n (the cdf of the normalized sample mean of the i.i.d random variables). It states that for all x and n there exists a positive constant C such that

$$|F_n(x) - \Phi(x)| \leq \frac{C}{\sqrt{n}} \frac{\rho}{\sigma^3}$$

with the third absolute moment $E\left[|X - \mu|^3\right] = \rho < \infty$.

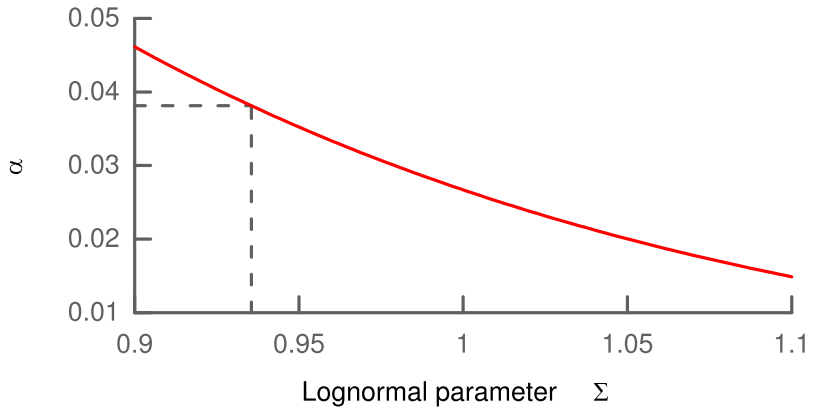

Figure A.1 The dependence of α on the lognormal parameter Σ. The dashed line indicates the value from the fit to experimental data ($\Sigma = 0.9355$).

While the upper bound of this difference decreases with the number of samples n, it depends linear on the normalized moment $\beta = \frac{\rho}{\sigma^3}$. By increasing β the difference to the Gaussian distribution can be increased. Since ρ is the third absolute central moment, a trivial lower bound for β is always the skewness S which is the third central moment divided by σ^3, i.e. $S \leq \beta$.

Synaptic strength is observed to approximately follow a lognormal distribution $\ln \mathcal{N}(M, \Sigma^2)$ [Song et al., 2005; Loewenstein et al., 2011]. Interestingly, for lognormal distribution one also finds that there is always a value α such that $S(1 + \alpha)$ is an upper bound on β. More concrete, $\beta < S(1 + \alpha)$ where $\alpha \propto \exp[-3\Sigma^2]$ for $\Sigma \to \infty$. That means,

$$S < \beta < S(1 + \alpha)$$

and the lower bound will become tight for sufficiently large Σ. The skewness is than a reasonable approximation to β.

Specifically, for a log-normal fit to experimental data with $\Sigma = 0.9355$ [Song et al., 2005] the bound is already very close with $\alpha < 0.04$ (cf. Fig A.1). Thus, maximizing the skewness of synaptic strength will maximize β and the upper bound in the Barry-Esseen theorem. For example, doubling the skewness in a lognormal distribution (by increasing its variance) doubles the upper bound. This would require four times as many inputs to compensate for. Thus, for a highly skewed weight distribution, the total synaptic current will only slowly converge to a Gaussian distribution and retain its specificity even for a large number of synaptic inputs.

Appendix B

Glossary

Abbreviation	Expression
ABS theory	Artola-Bröcher-Singer theory
AMPA	2-amino-3-(5-methyl-3-oxo-1,2-oxazol-4-yl)propanoic acid
bAP	backpropagating action potential
BCM theory	Bienenstock-Cooper-Munroe theory
CaMKII	Ca^{2+}/calmodulin-dependent protein kinase II
EPSC/IPSC	excitatory/inhibitory postsynaptic current
EPSP/IPSP	excitatory/inhibitory postsynaptic potential
GABA	gamma-aminobutyric acid
ICA	independent component analysis
iid	independent and identically distributed
IP	intrinsic plasticity
LIF neuron	leaky-integrate-and-fire neuron
LTD	long-term depression
LTP	long-term potentiation
NMDA	N-Methyl-D-aspartic acid
PCA	principal component analysis
PDF	probability density function
PPI	phosphatase I
RNN	recurrent neural networks
STDP	spike timing-dependent plasticity

Table B.1 Abbreviations

Bibliography

L. F. Abbott and S. B. Nelson. Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3(november):1178–83, Nov. 2000.

W. C. Abraham. Metaplasticity: tuning synapses and networks for plasticity. *Nature Reviews. Neuroscience*, 9(5):387, May 2008.

D. J. Aidley. *The Physiology of Excitable Cells*. Cambridge University Press, 4 edition, 1998.

T. Aihara, Y. Abiru, Y. Yamazaki, H. Watanabe, Y. Fukushima, and M. Tsukada. The relation between spike-timing dependent plasticity and Ca2+ dynamics in the hippocampal CA1 network. *Neuroscience*, 145(1): 80–7, Mar. 2007.

B. K. Andrasfalvy and J. C. Magee. Distance-dependent increase in AMPA receptor number in the dendrites of adult hippocampal CA1 pyramidal neurons. *The Journal of Neuroscience*, 21(23):9151–9, Dec. 2001.

A. Artola and W. Singer. Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends in Neurosciences*, 16(11), 1993.

A. Artola, S. Bröcher, and W. Singer. Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature*, 347(6 September 1990):69–72, 1990.

D. Attwell and S. B. Laughlin. An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow and Metabolism*, 21 (10):1133–45, Oct. 2001.

B. Babadi and L. F. Abbott. Intrinsic Stability of Temporally Shifted Spike-Timing Dependent Plasticity. *PLoS Computational Biology*, 6(11):e1000961, Nov. 2010.

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

A. J. Bell and T. J. Sejnowski. The "independent components" of natural scenes are edge filters. *Vision Research*, 37(23):3327–38, Dec. 1997.

C. C. Bell, V. Z. Han, Y. Sugawara, and K. Grant. Synaptic plasticity in a cerebellum-like structure depends on temporal order. *Nature*, 387(15 May 1997):278–81, May 1997.

D. R. Berard, J. W. Burgess, and R. G. Coss. Plasticity of dendritic spine formation: A state dependent stochastic process. *International Journal of Neuroscience*, 13:93–98, Jan. 1981.

G.-q. Bi and M.-m. Poo. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, 18(24):10464–72, Dec. 1998.

E. Bienenstock, L. N. Cooper, and P. Munro. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience*, 2(1):32, 1982.

B. Blais, N. Intrator, H. Shouval, and L. N. Cooper. Receptive field formation in natural scene environments: comparison of single-cell learning rules. *Neural Computation*, 10(7):1797–1813, 1998.

T. Bliss and T. Lomo. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology*, 232(2):331, 1973.

T. Branco and M. Häusser. The single dendritic branch as a fundamental functional unit in the nervous system. *Current Opinion in Neurobiology*, pages 494–502, Aug. 2010.

E. C. Burgard and J. J. Hablitz. NMDA receptor-mediated components of miniature excitatory synaptic currents in developing rat neocortex. *Journal of Neurophysiology*, 70(5):1841–52, Nov. 1993.

N. T. Carnevale and M. L. Hines. *The NEURON Book*. Cambridge University Press, Cambridge, UK, 2006.

C. Clopath and W. Gerstner. Voltage and spike timing interact in STDP - a unified model. *Frontiers in Synaptic Neuroscience*, 2(July):1–11, 2010.

P. Comon. Independent Component Analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

B. Connors and M. Gutnick. Intrinsic firing patterns of diverse neocortical neurons. *Trends in Neurosciences*, 13(3):99–104, 1990.

H. Cuntz, F. Forstner, A. Borst, and M. Häusser. One Rule to Grow Them All: A General Theory of Neuronal Branching and Its Practical Application. *PLoS Computational Biology*, 6(8):e1000877, Aug. 2010.

S. M. Dudek and M. F. Bear. Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade. *Neurobiology*, 89(May):4363–4367, 1992.

S. El Boustani, P. Yger, Y. Fregnac, and a. Destexhe. Stable Learning in Stochastic Network States. *The Journal of Neuroscience*, 32(1):194–214, Jan. 2012.

C. G. Esseen. A moment inequality with an application to the central limit theorem. *Scandinavian Actuarial Journal*, 1956(2):160–70, 1956.

D. E. Feldman. Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex. *Neuron*, 27(1):45–56, July 2000.

D. Fioravante and W. G. Regehr. Short-term forms of presynaptic plasticity. *Current Opinion in Neurobiology*, 21(2):269–74, Apr. 2011.

P. Földiak and M. P. Young. Sparse Coding in the Primate Cortex. In Michael A. Arbib, editor, *Handbook of Brain Theory and Neural Networks*, pages 895–898. 1995.

R. C. Froemke, M.-m. Poo, and Y. Dan. Spike-timing-dependent synaptic plasticity depends on dendritic location. *Nature*, 434(March 2005):221–225, 2005.

F. Gabbiani, J. Midtgaard, and T. Knöpfel. Synaptic integration in a model of cerebellar granule cells. *Journal of Neurophysiology*, 72(2), 1994.

W. Gerstner, R. Kempter, J. L. van Hemmen, and H. Wagner. A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 383(6595):76–78, 1996.

M. Graupner and N. Brunel. Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *Proceedings of the National Academy of Sciences of the United States of America*, pages 1–6, Feb. 2012.

A. T. Gulledge, N. T. Carnevale, and G. J. Stuart. Electrical Advantages of Dendritic Spines. *PLoS ONE*, 7(4):e36007, Apr. 2012.

A. Hasenstaub, S. Otte, E. Callaway, and T. J. Sejnowski. Metabolic cost as a unifying principle governing neuronal biophysics. *Proceedings of the National Academy of Sciences of the United States of America*, 107(27): 12329–34, July 2010.

D. Hebb. *The organization of behavior*. Wiley & Sons, New York, 1949.

M. L. Hines, T. Morse, M. Migliore, N. T. Carnevale, and G. M. Shepherd. ModelDB: A Database to Support Computational Neuroscience. *Journal of Computational Neuroscience*, 17(1):7–11, 2004.

A. Hodgkin and A. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117(4):500–544, 1952.

S. Hornillo-Mellado, R. Martín-Clemente, and J. M. Górriz-Sáez. Connections Between ICA and Sparse Coding. *Computational Intelligence and Bioinspired Systems*, 3512:1035–1042, 2005.

T. Hromádka, M. R. DeWeese, and A. M. Zador. Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biology*, 6(1), 2008.

D. Hubel and T. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195(1):215, 1968.

S. R. Hulme, O. D. Jones, D. R. Ireland, and W. C. Abraham. Calcium-Dependent But Action Potential-Independent BCM-Like Metaplasticity in the Hippocampus. *Journal of Neuroscience*, 32(20):6785–6794, May 2012.

A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4-5):411–30, 2000.

A. Hyvärinen, J. Hurri, and P. Hoyer. *Natural Image Statistics: A probabilistic approach to Early Computational Vision.* Springer, 2009.

H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. Technical report, Fraunhofer Institute for Autonomous Intelligent Systems, 2001.

H. Jaeger. A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach. Technical report, Fraunhofer Institute for Autonomous Intelligent Systems, 2002a.

H. Jaeger. Adaptive nonlinear system identification with echo state networks. In *NIPS*, 2002b.

H. Jaeger and H. Haas. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, Apr. 2004.

H. Jaeger, W. Maass, and J. Principe. Special issue on echo state networks and liquid state machines. *Neural Networks*, 20(3):287–289, Apr. 2007.

E. G. Jones and T. P. Powell. Morphological variations in the dendritic spines of the neocortex. *Journal of Cell Science*, 5(2):509–29, Sept. 1969.

E. R. Kandel. Genes, synapses, and long-term memory. *Journal of Cellular Physiology*, 173(2):124–5, Nov. 1997.

G. A. Kinney, B. W. Peterson, and N. T. Slater. The synaptic activation of N-methyl-D-aspartate receptors in the rat medial vestibular nucleus. *Journal of Neurophysiology*, 72(4), 1994.

A. Kirkwood and M. Rioult. Experience-dependent modification of synaptic plasticity in visual cortex. *Nature*, 1996.

C. Koch and T. Poggio. A theoretical analysis of electrical properties of spines. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 218(1213):455, 1983.

C. Kolodziejski, B. Porr, and F. Wörgötter. On the asymptotic equivalence between differential Hebbian and temporal difference learning. *Neural Computation*, 21(4):1173–202, Apr. 2009.

B. Kosko. Differential Hebbian learning. In *AIP Conference Proceedings*, volume 151, pages 277–282. AIP, 1986.

D. Krieg and J. Triesch. How metabolic constraints shape neuronal adaptation: A unifying objective function for synaptic plasticity. In *Conference Abstract: Bernstein Conference on Computational Neuroscience*, volume 5. Frontiers in Computational Neuroscience, 2011a.

D. Krieg and J. Triesch. STDP explained? Connecting function and biophysics. In *Cosyne Abstracts*, 2011b.

D. Krieg and J. Triesch. A unifying theory of synaptic long-term plasticity based on a sparse distribution of synaptic strength. Submitted to *PLoS Computational Biology*, July 2012 under review.

D. Krieg, A. Perianez, and J. Triesch. An objective function for STDP: increasing the separability in self-organized recurrent neural networks. In *Conference Abstract: Bernstein Conference on Computational Neuroscience*. Frontiers in Computational Neuroscience, 2010.

K. Lamsa, D. Kullmann, and M. Woodin. Spike-timing dependent plasticity in inhibitory circuits. *Frontiers in Synaptic Neuroscience*, 2(June):1–8, 2010.

S. B. Laughlin. Energy as a constraint on the coding and processing of. *Current Opinion in Neurobiology*, 11(4):475–480, 2001.

S. B. Laughlin and T. J. Sejnowski. Communication in neuronal networks. *Science*, 301(5641):1870–4, Sept. 2003.

A. Lazar, G. Pipa, and J. Triesch. SORN: a self-organizing recurrent neural network. *Frontiers in Computational Neuroscience*, 3(October):23, Jan. 2009.

T. K. Leen and R. Friel. Stochastic Perturbation Methods for Spike-Timing-Dependent Plasticity. *Neural Computation*, 24(5):1109–1146, 2012.

P. Lennie. The cost of cortical computation. *Current Biology*, 13(6):493–497, 2003.

S. Leutgeb and J. K. Leutgeb. Pattern separation, pattern completion, and new neuronal codes within a continuous CA3 map. *Learning & Memory*, 14 (11):745–57, Nov. 2007.

W. B. Levy and R. A. Baxter. Energy efficient neural codes. *Neural Computation*, 8(3):531–43, Apr. 1996.

W. B. Levy and R. A. Baxter. Energy-efficient neuronal computation via quantal synaptic failures. *The Journal of Neuroscience*, 22(11):4746–55, June 2002.

J. E. Lisman. A mechanism for the Hebb and the anti-Hebb processes underlying learning and memory. *Proceedings of the National Academy of Sciences of the United States of America*, 86(23):9574–8, Dec. 1989.

J. E. Lisman and M. A. Goldring. Feasibility of long-term storage of graded information by the Ca2+/calmodulin-dependent protein kinase molecules of the postsynaptic density. *Proceedings of the National Academy of Sciences of the United States of America*, 85(14):5320–4, July 1988.

Y. Loewenstein, A. Kuras, and S. Rumpel. Multiplicative dynamics underlie the emergence of the log-normal distribution of spine sizes in the neocortex in vivo. *The Journal of Neuroscience*, 31(26):9481–8, June 2011.

A. Losonczy, J. K. Makara, and J. C. Magee. Compartmentalized dendritic plasticity and input feature storage in neurons. *Nature*, 452(7186):436–41, Mar. 2008.

M. Lukosevicius and H. Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.

G. Lynch, T. Dunwiddie, and V. Gribkoff. Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. *Nature*, 266(April):737–739, 1977.

W. Maass, T. Natschläger, and H. Markram. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural computation*, 14(11):2531–60, Nov. 2002.

D. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 6 edition, 2003.

A. Maffei. The many forms and functions of long term plasticity at GABAergic synapses. *Neural Plasticity*, 2011:254724, Jan. 2011.

J. C. Magee and E. P. Cook. Somatic EPSP amplitude is independent of synapse location in hippocampal pyramidal neurons. *Nature Neuroscience*, 3(9):895–903, Sept. 2000.

Z. Mainen, J. Joerges, J. Huguenard, and T. J. Sejnowski. A model of spike initiation in neocortical pyramidal neurons. *Neuron*, 15(6):1427–1439, 1995.

H.-Y. Man. GluA2-lacking, calcium-permeable AMPA receptors–inducers of plasticity? *Current Opinion in Neurobiology*, 21(2):291–8, Apr. 2011.

T. W. Margrie, M. Brecht, and B. Sakmann. In vivo, low-resistance, whole-cell recordings from neurons in the anaesthetized and awake mammalian brain. *European Journal of Physiology*, 444(4):491–8, July 2002.

H. Markram. Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs. *Science*, 275(5297):213–215, Jan. 1997.

D. Marr. Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 262(841):23–81, 1971.

D. Marr. *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. MIT Press, 2010.

A. Ngezahayo, M. Schachner, and A. Artola. Synaptic activity modulates the induction of bidirectional synaptic changes in adult mouse hippocampus. *The Journal of Neuroscience*, 20(7):2451–8, Apr. 2000.

L. G. Nowak, R. Azouz, M. V. Sanchez-Vives, C. M. Gray, and D. a. McCormick. Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. *Journal of neurophysiology*, 89 (3):1541–66, Mar. 2003.

B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(June): 607–609, 1996.

B. A. Olshausen and D. J. Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4):481–7, Aug. 2004.

J.-P. Pfister and W. Gerstner. Triplets of spikes in a model of spike timing-dependent plasticity. *The Journal of Neuroscience*, 26(38):9673–82, Sept. 2006.

R. R. Pool and G. Mato. Spike-timing-dependent plasticity and reliability optimization: the role of neuron dynamics. *Neural Computation*, 23(7): 1768–89, July 2011.

R. P. N. Rao and T. J. Sejnowski. Spike-timing-dependent Hebbian plasticity as temporal difference learning. *Neural Computation*, 13(10):2221–37, Oct. 2001.

A. Rodríguez-Moreno, M. M. Kohl, J. E. Reeve, T. R. Eaton, H. A. Collins, H. L. Anderson, and O. Paulsen. Presynaptic Induction and Expression of Timing-Dependent Long-Term Depression Demonstrated by Compartment-Specific Photorelease of a Use-Dependent NMDA Receptor Antagonist. *The Journal of Neuroscience*, 31(23):8564–9, June 2011.

A. Saudargiene, B. Porr, and F. Wörgötter. How the shape of pre- and post-synaptic signals can influence STDP: a biophysical model. *Neural Computation*, 16(3):595–625, Mar. 2004.

T. J. Sejnowski. Storing covariance with nonlinearly interacting neurons. *Journal of Mathematical Biology*, 4(4):303–21, Oct. 1977.

H. Z. Shouval. Spike timing dependent plasticity: a consequence of more fundamental learning rules. *Frontiers in Computational Neuroscience*, 4 (July):1–13, 2010.

H. Z. Shouval, M. F. Bear, and L. N. Cooper. A unified model of NMDA receptor-dependent bidirectional synaptic plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16):10831–6, Aug. 2002.

P. J. Sjöström and M. Häusser. A cooperative switch determines the sign of synaptic plasticity in distal dendrites of neocortical pyramidal neurons. *Neuron*, 51(2):227–38, July 2006.

P. J. Sjöström, G. G. Turrigiano, and S. B. Nelson. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, 32(6):1149–64, Dec. 2001.

P. J. Sjöström, E. Rancz, A. Roth, and M. Häusser. Dendritic excitability and synaptic plasticity. *Physiological Reviews*, 88(2):769–840, 2008.

I. Song and R. L. Huganir. Regulation of AMPA receptors during synaptic plasticity. *Trends in Neurosciences*, 25(11):578–88, Nov. 2002.

S. Song, P. J. Sjöström, M. Reigl, S. Nelson, and D. B. Chklovskii. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biology*, 3(3):e68, Mar. 2005.

H. Sprekeler, C. Michaelis, and L. Wiskott. Slowness: an objective for spike-timing-dependent plasticity? *PLoS Computational Biology*, 3(6):1137–1148, June 2007.

N. Spruston, P. Jonas, and B. Sakmann. Dendritic glutamate receptor channels in rat hippocampal CA3 and CA1 pyramidal neurons. *Journal of Physiology*, 482 (Pt 2):325–52, Jan. 1995.

P. N. Steinmetz, a. Manwani, C. Koch, M. London, and I. Segev. Subthreshold voltage noise due to channel fluctuations in active neuronal membranes. *Journal of Computational Neuroscience*, 9(2):133–48, 2000.

M. Stemmler and C. Koch. How voltage-dependent conductances can adapt to maximize the information encoded by neuronal firing rate. *Nature Neuroscience*, 2(6):521–7, June 1999.

M. Takahashi, Y. Kovalchuk, and D. Attwell. Pre- and postsynaptic determinants of EPSC waveform at cerebellar climbing fiber and parallel fiber to Purkinje cell synapses. *The Journal of Neuroscience*, 15(8):5693–702, Aug. 1995.

T. Tateno, a. Harsch, and H. P. C. Robinson. Threshold firing frequency-current relationships of neurons in rat somatosensory cortex: type 1 and type 2 dynamics. *Journal of neurophysiology*, 92(4):2283–94, Oct. 2004.

T. Toyoizumi, J.-P. Pfister, K. Aihara, and W. Gerstner. Spike-timing dependent plasticity and mutual information maximization for a spiking neuron model. In *NIPS*, 2005.

A. Treves, A. Tashiro, M. E. Witter, and E. I. Moser. What is the mammalian dentate gyrus good for? *Neuroscience*, 154(4):1155–72, July 2008.

J. Triesch. A gradient rule for the plasticity of a neuron's intrinsic excitability. In *Proceedings of the International Conference on Artificial Neural Networks*, 2005.

J. Triesch. Synergies between intrinsic and synaptic plasticity mechanisms. *Neural Computation*, 2007.

Y. Tsubo, Y. Isomura, and T. Fukai. Power-Law Inter-Spike Interval Distributions Infer a Conditional Maximization of Entropy in Cortical Neurons. *PLoS Computational Biology*, 8(4):e1002461, Apr. 2012.

G. G. Turrigiano and S. B. Nelson. Homeostatic plasticity in the developing nervous system. *Nature Reviews. Neuroscience*, 5(2):97–107, Feb. 2004.

M. C. van Rossum, G.-q. Bi, and G. G. Turrigiano. Stable Hebbian learning from spike timing-dependent plasticity. *The Journal of Neuroscience*, 20 (23):8812–21, Dec. 2000.

T. P. Vogels, H. Sprekeler, F. Zenke, C. Clopath, and W. Gerstner. Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science*, 334(6062):1569–73, Dec. 2011.

H. Wang and J. J. Wagner. Priming-induced shift in synaptic plasticity in the rat hippocampus. *Journal of Neurophysiology*, 82(4):2024–8, Oct. 1999.

P. Werbos. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560, 1990.

S. H. Wright. Generation of resting membrane potential. *Advances in Physiology Education*, 28(1-4):139–42, Dec. 2004.

M. A. Yassa and C. E. L. Stark. Pattern separation in the hippocampus. *Trends in Neurosciences*, 34(10):515–25, Oct. 2011.

R. Yuste and T. Bonhoeffer. Morphological changes in dendritic spines associated with long-term synaptic plasticity. *Annual Review of Neuroscience*, 24:1071–89, 2001.

L. I. Zhang, H. W. Tao, and M.-m. Poo. A critical window for cooperation and competition among developing retinotectal synapses. *Nature*, 395 (September):37–44, 1998.

J. J. Zhu and B. W. Connors. Intrinsic firing patterns and whisker-evoked synaptic responses of neurons in the rat barrel cortex. *Journal of neurophysiology*, 81:1171–1183, 1999.