Network mechanisms of memory storage in the balanced cortex

Alessandro Barri

Doctoral thesis

Université Paris Descartes

Presented for the award of

Doctor of Université Paris Descartes

Network Mechanisms of Memory Storage in the Balanced Cortex

Alessandro Barri

8th of December 2014

Committee

David DiGregorio (examiner)	Institut Pasteur, Paris
Robert Gütig	MPI für Exp. Medizin, Göttingen
David Hansel (doctoral advisor)	Université Paris Descartes
German Mato (examiner)	Centro Atómico Bariloche
Gianluigi Mongillo	Université Paris Descartes
Israel Nelken	The Hebrew University, Jerusalem

To my parents.

Abstract

It is generally maintained that one of cortex' functions is the storage of a large number of memories. In this picture, the physical substrate of memories is thought to be realised in pattern and strengths of synaptic connections among cortical neurons. Memory recall is associated with neuronal activity that is shaped by this connectivity. In this framework, active memories are represented by attractors in the space of neural activity.
Electrical activity in cortical neurones *in vivo* exhibits prominent temporal irregularity. A standard way to account for this phenomenon is to postulate that recurrent synaptic excitation and inhibition as well as external inputs are balanced. In the common view, however, these balanced networks do not easily support the coexistence of multiple attractors. This is problematic in view of memory function.
Recently, theoretical studies showed that balanced networks with synapses that exhibit short-term plasticity (STP) are able to maintain multiple stable states. In order to investigate whether experimentally obtained synaptic parameters are consistent with model predictions, we developed a new methodology that is capable to quantify both response variability and STP at the same synapse in an integrated and statistically-principled way. This approach yields higher parameter precision than standard procedures and allows for the use of more efficient stimulation protocols.
However, the findings with respect to STP parameters do not allow to make conclusive statements about the validity of synaptic theories of balanced working memory.
In the second part of this thesis an alternative theory of cortical memory storage is developed. The theory is based on the assumptions that memories are stored in attractor networks, and that memories are not represented by network states differing in their average activity levels, but by micro-states sharing the same global statistics. Different memories differ with respect to their spatial distributions of firing rates. From this the main result is derived: the balanced state is a necessary condition for extensive memory storage. Furthermore, we analytically calculate memory storage capacities of rate neurone networks. Remarkably, it can be shown that crucial properties of neuronal activity and physiology that are consistent with experimental observations are directly predicted by the theory if optimal memory storage capacity is required.

Contents

Chapter 1

Introduction

Memories are undoubtedly central when it comes to understanding the brain. There is an abundant amount of both experimental and theoretical studies concerned with this subject. At this, memory has been investigated on different levels of biological organisation, from the molecular to the psychological and even beyond. Here, we consider mechanisms of memory that span from the level of synapses to the level of neural networks.
Our general approach here is reductionist, assuming that complex biological phenomena can ultimately be explained by uncovering a set of simple, elementary principles. We hope to make a small contribution to this endeavour.
This first chapter is dedicated to introduce the general concepts and the set of problems we want to concern ourselves with. At first we will give a short introduction to memory function and to working memory. Closely linked to this we will discuss crucial properties of the physical substrate of memory activation: cortical activity. We proceed by introducing balanced networks and will briefly review models that try to explain memory function in this framework. Finally, motivated by work which relates memory function to synaptic short-term plasticity, we will give a short introduction to this topic. In Chapter 2 we will review the most important models of short-term plasticity and demonstrate that they can be subsumed in a common framework that is based on the notion of the release site. We show that this framework enables us to devise synapse models that capture both the dynamics and the stochasticity of synaptic transmission. The advantages of using this new methodology are presented in chapter 3 together with an analysis of experimental recordings from synaptic connections. The obtained parameters estimates do not permit us to make conclusive statements about the validity of aforementioned synaptic memory models. We thus propose an alternative theory of memory storage in chapter 4, with which we can predict crucial properties of cortical activity from optimality considerations. Chapter 5 is devoted to a general discussion.

1.1 Memory, working memory and cortical activity

In this section we want to give a short introduction to the organisation of memory function in cortex. We will follow the work of Fuster [see e.g. Fuster (1995, 1997, 2009)].

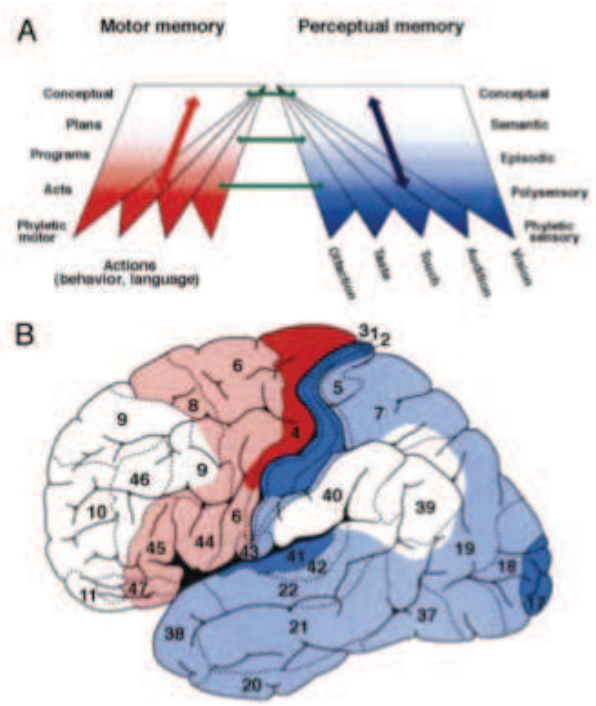

Figure 1.1: **Organisation of memory in the cortex. A)**: Hierarchical organisation of motor and perceptual memory from less (dark colours) to more abstract (white). **B)**: Memory locations in the cortex; same colour code as in **A**; numbers denote the Brodmann areas. Figure adapted from Fuster (1997).

According to Fuster, memory can be subdivided in two classes: perceptual and motor memory. Perceptual memory is linked to sensory experience in the broadest sense, including episodic and declarative memory. Motor memory refers to basic acts but also to more complex motor programs, goal directed behaviour and plans.
At this, both memory types are organised in a hierarchical fashion. This hierarchy ranges from very primitive memories up to very complex, abstract representations. On the lowest level we find so-called phyletic memories. These are thought to be innate, that is, present before learning. Fuster calls them 'memories of the species'. Phyletic perceptual memory is thought to correspond to basic sensory experiences, while phyletic motor memory represents information linked to elementary motor acts, like the contraction of specific muscles and muscle groups. Ascending the hierarchy, memories gain in their level of abstraction; at the top, we find representations of semantic concepts, faces, ethical principles etc.. At any stage of the hierarchy, higher level memories build upon lower level memories. Figure 1.1 shows a scheme of the described organisation. As indicated by the arrows between different levels and between the two memory pillars, these organisational principles do not imply a strict separation between these entities. On the contrary, interconnections are vast.

The two memory hierarchies are to a significant extend reflected by anatomy. Figure 1.1B shows how the memory areas extend over the cortex. Starting from phyletic memory areas close to the central sulcus, motor memory expands into frontal cortex and perceptual memory to posterior regions. This gradient from simple to complex is also mirrored by phylogenetical and ontogenetical developmental processes; the prefrontal cortex, for instance, is one of the evolutionary newest cortical regions and one of the last to mature.
So far we have considered memories in rather abstract terms. But what is their physical substrate? It is generally maintained that memories can be identified with the way cortical neuronal networks are organised. This idea dates back to Hebb, who coined the notion of cell assemblies [Hebb (1968)]. In this view, memories are formed by sets of neurones that are recurrently interconnected. When a memory is activated, or recalled, the neurones in that network excite one another, thereby maintaining the memory-circuit as a whole in an activated state. Learning of these assemblies happens by strengthening of synaptic connections between neurones that are simultaneously activated. By virtue of this associative process, existing networks can be in principle endlessly extended in an huge variety of ways.
We want to emphasise an important issue at this point. According to the Hebbian framework, memories are ultimately identical to the synaptic configuration of cortical networks. Once formed, memories are physically present in the synaptic structure, independent of neuronal activity. The fact that a memory is stored in a network becomes apparent when this memory is retrieved: in that case, the synaptic structure enables the network to display the particular neuronal activity that represents that memory. In this view, the synaptic configuration is identical to what is commonly known as long-term memory, while the effect it causes - the corresponding neuronal activity - can be identified as short-term memory, or, using the vocabulary of Fuster, 'active memory'. Importantly, this implies that short-term/active and long-term memory are not physically separated entities but are co-located in the same cortical networks.
This suggests a strategy for the study of memory function: employing experimental paradigms that require active memory can be useful to uncover general principles of memory organisation and memory dynamics. Indeed, in what follows we will focus on a specific form of activated memory, the so-called working memory, which has been a fruitful and intensively studied research topic in the last decades.

1.1.1 Working memory

Working memory (WM) is the ability to temporarily retain information that is relevant for a motor or cognitive task in the near future [Baddeley (1992)]. The memorised information can be of discrete nature, like an image or a telephone number as well as a continuous quantity, like the position of a dot on a screen, or the strength of a mechanical vibration. WM is generally considered to be of utmost importance for nonroutine behaviour, like the creative use of language, the solving of a puzzle, the handling of an equation and so on.
A cortical region that is thought to play a key role in WM is the prefrontal cortex (PFC). For instance, studies in which the PFC of monkeys was temporarily cooled found that

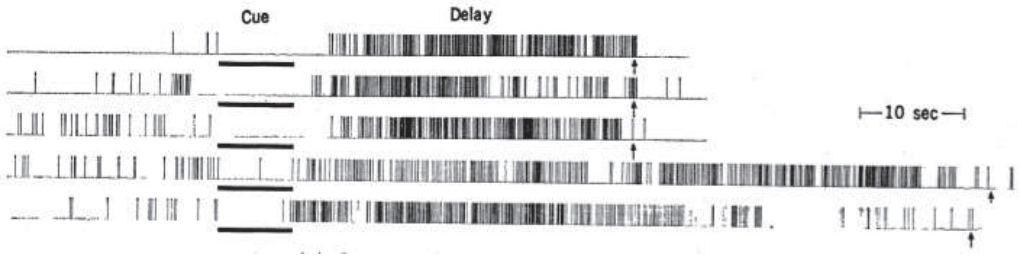

Figure 1.2: **Persistent firing of a memory cell.** On and offset of delay period activity in the same PFC cell in different trials. Figure adapted from Fuster et al. (1971).

the behavioural performance in tasks that require WM is significantly impaired in that condition [Bauer and Fuster (1976); Fuster et al. (1985)]. We refer at this point to reviews of anatomy and function of the prefrontal cortex than can be found, for example in Fuster (1988) and Miller and Cohen (2001).

In order to expose the neuronal mechanisms that underlie WM many researchers have over the last decades recorded PFC neurones in animals performing so-called delayed response tasks. One of the first of a long series of studies was carried out by Fuster and Alexander [Fuster et al. (1971)]. In their setup, a monkey was shown how a piece of apple was placed under one of two objects, which were subsequently concealed for a delay period of several tens of seconds. After this period, the objects were revealed and the monkey had to choose the correct object to obtain a reward. In simultaneous extracellular recordings from PFC the authors found neurons that exhibited an elevated firing-rate during the delay period. This is shown in figure 1.2. Two things are crucial here: first, the neurones started elevated firing only after cue presentation, and second, returned to baseline only after the end of the delay period (marked by the arrows). This suggests that this elevated activity is contingent upon the necessity of active memory maintenance.

The notion is now widely held that this type of 'persistent activity' is the hallmark of so-called 'memory cells' [Fuster (1995); Goldman-Rakic (1995)]. These cells are thought to belong to memory networks that encode task-relevant information by means of their elevated activity. Memory cells have been found in many different versions of the delayed response task, across species [e.g. Kesner et al. (1996)] and in various cortical areas [e.g. Miyashita and Chang (1988); Miller et al. (1996)]. They can be activated by different sensory modalities [e.g. Romo et al. (1999)] and can code for discrete items and also continuous quantities [e.g. Funahashi et al. (1989)].

A visual form of so-called parametric WM can be studied with the oculomotor delayed-response task (ODR task) [see e.g. Funahashi et al. (1989)]. In the classical version of this setting a monkey has to memorise a visual cue presented on a screen. After presentation of the cue, that can appear in one of eight possible locations around a fixation point, the monkey needs to retain a spatial memory of the cue's position. Finally, the monkey gets rewarded if he performs a saccade to the correct target. In Funahashi et al.

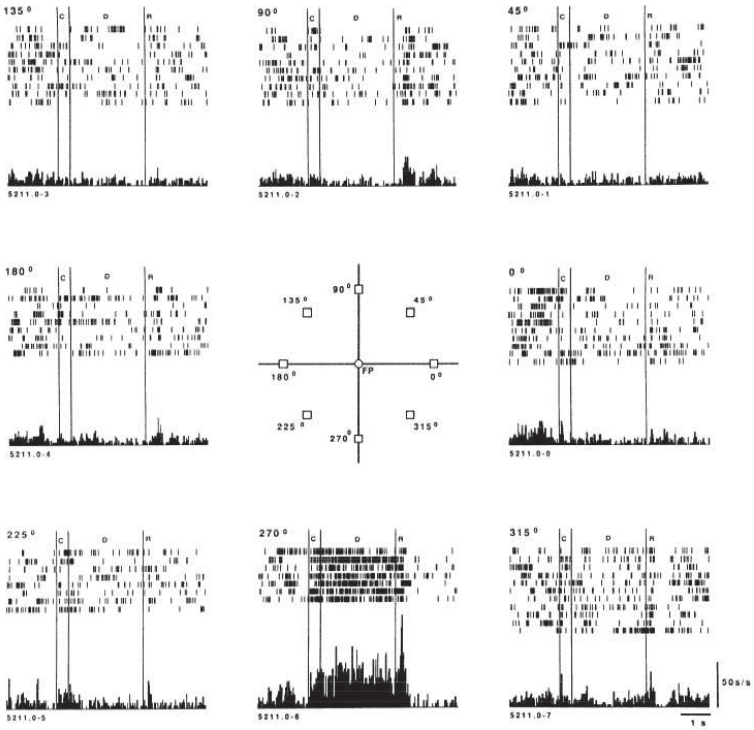

Figure 1.3: **Selective persistent activity.** Middle: the eight possible positions at which a cue can appear during the ODR task. Other panels: activities of the same memory cell in the eight different cue conditions; vertical lines represent onset/offset of cue and delay period. Adapted from Funahashi et al. (1989).

(1989) neurones in PFC were recorded during the task. Figure 1.3 shows recordings from an exemplary neurone that exhibits persistent activity during the delay period that is specific to one out of eight directions. This example thereby also illustrates that memory can feature strong selectivity to certain stimuli.
We have seen that the general organisation of memory in the cortex can potentially be revealed by the study of active memory. In the context of WM, active memory involves the elevated activation of PFC neurones, but also neurones from other cortical areas. It is therefore worthwhile to consider general properties of cortical neuronal activity.

1.1.2 Properties of cortical activity

Here, we want to shortly summarise some well established findings about cortical activity. This list focuses on those statistics of neuronal activity that are relevant to our work.
An ubiquitous feature of electrical activity in cortical neurones *in vivo* is their prominent temporal irregularity [Softky and Koch (1993); Bair et al. (1994)]. This irregularity differs in function of the cortical area [Shinomoto et al. (2009)]. The least variables neurones are found in the motor cortex, while neurones in prefrontal cortex tend to have

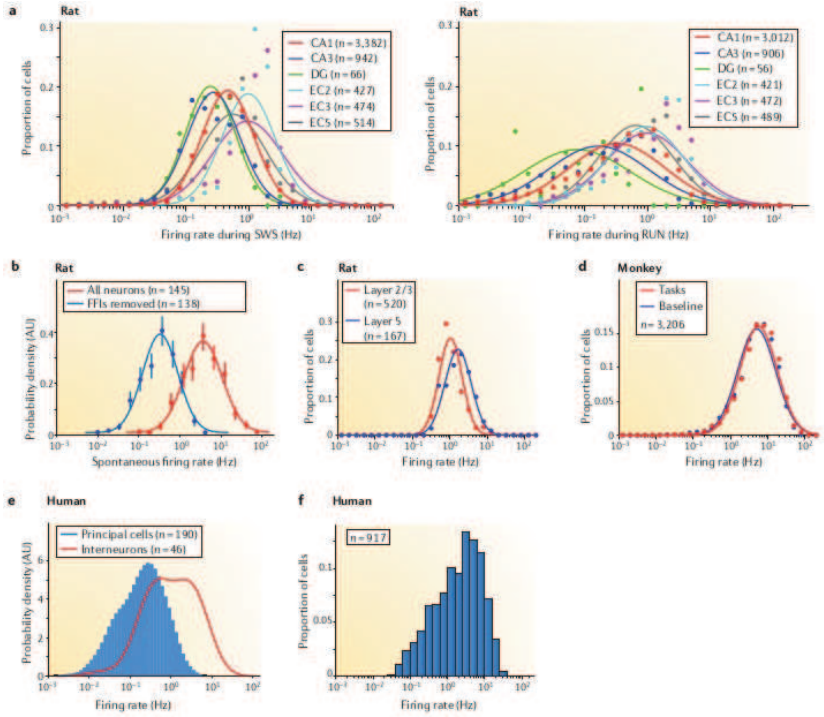

Figure 1.4: **Lognormal distribution of firing rates in the cortex. a)**: Cells from rat hippocampus and entorhinal cortex during slow-wave-sleep (left) and exploration (right). **b)**: Cells from awake rat A1. **c)**: PFC cells from exploring rats. **d)**: Cells from lateral intraparietal and parietal reach region areas of the macaque cortex during a baseline condition and during performance of a reaching task. **e)**: Human middle temporal gyrus cells recorded during sleep. **f)**: Neurons from multiple cortical areas of several human patients during various tasks. Figure and caption modified from Buzsáki and Mizuseki (2014).

inter-spike-interval distributions with coefficients of variation of 1 or higher. In an analysis of PFC neurones recorded in monkeys performing ODR tasks, several studies showed that neurones fire irregularly both during delay period persistent activity and in baseline conditions [Shinomoto et al. (1999); Compte et al. (2003); Shafi et al. (2007)].
Another universal property of all cortical regions is associated with the distributions of firing-rates. Numerous recent studies have found that firing-rates of most neurones are small, but that a few exhibit high activities. These observations have been made, for instance, in parietal and prefrontal cortex of monkeys [Shafi et al. (2007)] and in rat barrel cortex [O'Connor et al. (2010)]. When statistical models are fitted to the firing-rate distributions, log-normal distributions emerge as the best descriptions in most cases, as has been reported in rat A1 [Hromádka et al. (2008)], in rat V1 [Song et al. (2005)] and in hippocampus and entorhinal cortex of both awake and sleeping rats [Mizuseki and Buzsáki (2013)]. In two recent review articles, more examples are given [Wohrer et al.

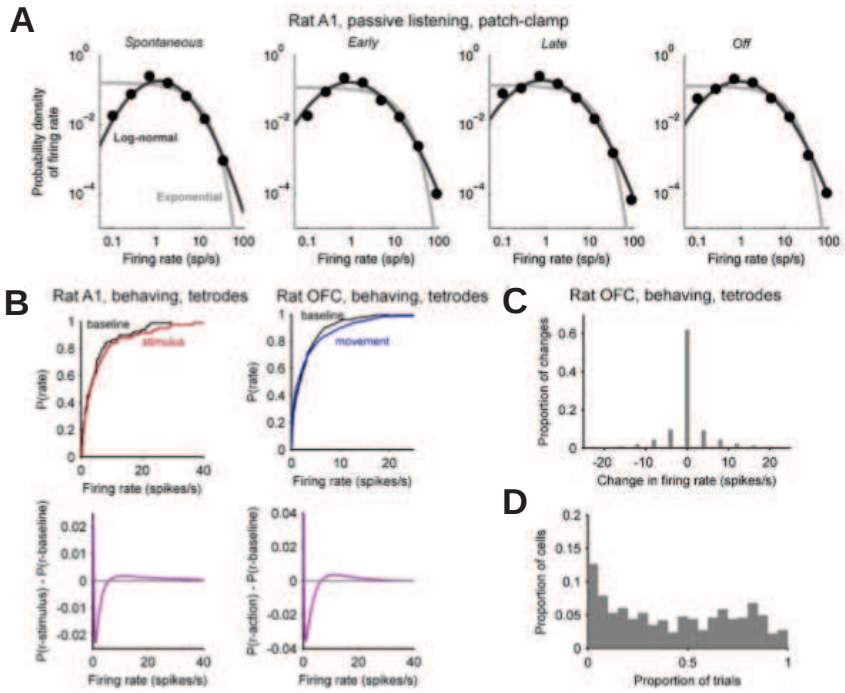

Figure 1.5: **Global firing-rate distributions change only little in function of the behavioural states. A)**: Firing-rates from rat A1 cells before (spontaneous), during (early, late) and after (off) a acoustic pip. **B)**: Stimulus-evoked changes in rat A1 and movement-evoked changes in rat OFC. Top: the empirical cumulative probability distributions for firing-rates during baseline and task. Bottom: difference in probability distribution functions for the best-fitting models to spontaneous activity, stimulus-evoked activity or action-related activity. **C)**: The distribution of rate changes in rat OFC between baseline and movement for every sampled firing rate (every neuron in every trial). **D)**: The distribution of the proportion of trials on which each neuron showed a difference in rate between baseline and movement for cells in rat OFC; the median proportion was 0.42. Figure and caption modified from Wohrer et al. (2012).

(2012); Buzsáki and Mizuseki (2014)]. Figure 1.4 shows several instances of firing-rate distributions across species, layers, neurone-types and behavioural conditions.
A particularly striking feature of cortical activity can be noticed in panel d. Shown are two distribution of firing rates from lateral intraparietal and parietal areas in a macaque monkey. The red curve shows the neuronal activity during a reaching task, the blue curve the same neuronal population when the monkey is idle. The difference in the overall activity between the two conditions is very small, indicating that the global state of these cortical regions does not change in function of the behavioural state.
Panels A and B of figure 1.5 show further evidence supporting this view. This permanence of the overall firing-rate distribution is, however, not due to the fact that no

changes of activity on the single neurone level occur. As can be seen from panels C and D, a substantial fraction of neurones indeed undergoes a modulation of their firing rates. These findings are consistent with the view that different states of cortical networks/areas correspond to different spatial distributions of neuronal activity with the same global statistical properties. At this, it seems to be irrelevant whether the animal is passively perceiving a stimulus or actively engaged in a motor behaviour. We have to note, however, that these remarks have to be restricted to states in which the animal is awake. As can be seen from figure 1.4a, changes in firing-rate statistics induced by sleep seem to be substantial.
Finally, it is well known that the average firing rates of inhibitory neurones are significantly larger than firing rates of excitatory ones [Beloozerova et al. (2003); Mitchell et al. (2007); Fujisawa et al. (2008); Gentet et al. (2010)]. This general feature can also be identified in figure 1.4, panels b and e.

1.1.3 The balanced state: a theory of cortical activity?

A framework that can potentially explain a wide range of properties of cortical activity is the theory of balanced networks. The original motivation underlying the development of this theory comes from the following (allegedly) paradoxical observations: on the one hand, the large majority of neurones *in vitro* that are subjected to a constant current pulse fire regular trains of action potentials [Connors et al. (1982)]. On the other hand, as we have described before, activity of cortical neurones *in vivo* is very irregular. This is *prima facie* surprising, as each neurone receives a large number of synaptic inputs [Softky and Koch (1993); Holt et al. (1996)]. By virtue of the law of large numbers, fluctuations in the neurone's total input should therefore be much smaller that the average input and we should expect to see regular spiking, as in the *in vitro* experiments.
This apparent contradiction can be overcome in recurrent networks with excitatory and inhibitory cells with relatively strong synapses; that is, synapses that are large compared to the neurones' thresholds. In such networks, both average excitatory and inhibitory currents are large, but fluctuations of these currents are of a size similar to the threshold's. If excitation and inhibition approximately balance, both the residual average input and its variance are of comparable magnitude (see figure 1.6). In technical terms, balanced networks provide input currents to the neurones whose average and variance are of the same order, in contrast to classical networks, where fluctuations vanish when the number of synapses becomes large. It is generally maintained that such conditions set neurones in a regime where they operate just below threshold. Fluctuations can then drive neurones above threshold and shape the spiking significantly. In this fluctuation-driven regime, neurones can produce irregular outputs with coefficients of variations around 1. We will, however, indicate below that a qualification has to be made.
Balance between excitation and inhibition has been notably discussed by Shadlen and Newsome [Shadlen and Newsome (1994, 1998)]. Van Vresswijk and Sompolinsky showed that this balance arises automatically and under very general circumstances in recurrent networks with strong synapses [van Vreeswijk and Sompolinsky (1996, 1998)]. Balance is a phenomenon that emerges from the dynamics of the network as a whole. Only very

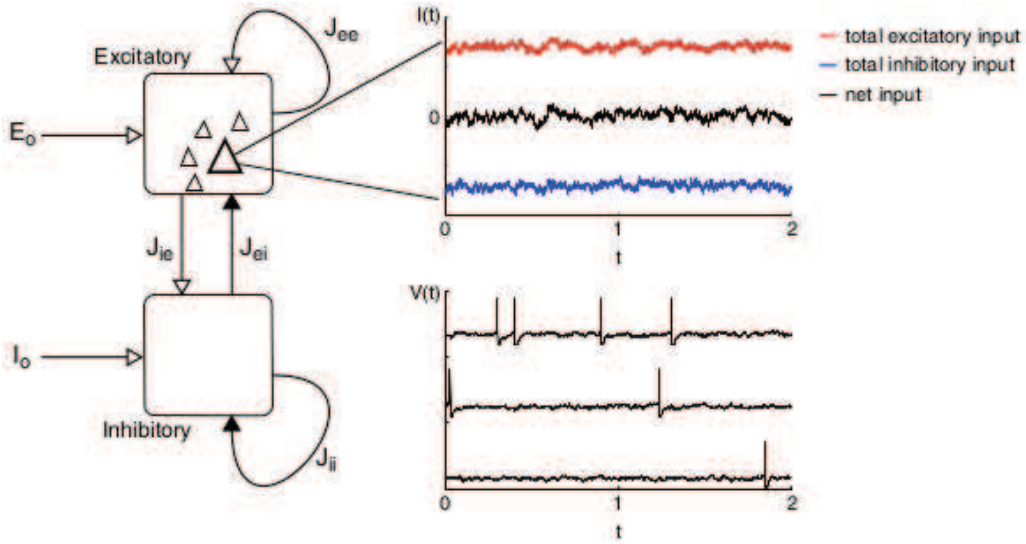

Figure 1.6: **Balanced networks can account for firing variability.** Left: Typical setup of balanced network models; E_0 and I_0 denote external inputs, the Js are the various synaptic strengths between the populations. Right: large excitatory and inhibitory inputs cancel and produce a total input of the order of the threshold at 0 (top); spiking is driven by fluctuations and thus irregular (bottom). Figure adapted from Wolf et al. (2014).

gentle requirements need to be imposed on the strength of synaptic weights and external inputs to the network in order for the balanced state to exist; no parameter fine-tuning is required. Figure 1.6 shows a schematic picture of the typical setup of balanced networks and simulated inputs and spikes.

The mechanism of dynamical balance was originally established in networks of randomly connected binary neurones, but the same principle has been demonstrated to work in various other configurations, for instance in networks of integrate-and-fire neurones [van Vreeswijk and Sompolinsky (2005); Lerchner et al. (2006); Renart et al. (2010)], conductance-based neurones [Hansel and van Vreeswijk (2012); Hansel and Mato (2013)] and rate-neurones [Roudi and Latham (2007)]. A comprehensive review of balanced networks can be found in Wolf et al. (2014).

There is indeed good experimental support for a balanced between excitation and inhibition in the cortex. It has been shown, for instance, that neurones in awake animals operate in a high-conductance state [see e.g. Destexhe et al. (2003); Shu et al. (2003); Haider et al. (2006)], indicating that at any point in time neurones receive a large number of synaptic inputs. Furthermore, it has been demonstrated *in vivo* that during spontaneous activity an increase in excitation is consistently accompanied by an increase in inhibition [Okun and Lampl (2008)]. For a review on this matter, see e.g. Isaacson and Scanziani (2011).

We have seen that the balanced state can account for the temporal variability that is characteristic of cortical neurones. Indeed, also the firing-rate distributions' high skewness can be approximately reproduced in this framework. Roxin et al. (2011) investigated a balanced network of leaky integrate-and-fire neurones (LIF). They showed that for typical LIF parameters and realistic average firing rates, synaptic inputs to neurones indeed operate in a regime not far below their thresholds. In that regime, the LIF f-I curve can be well described by an exponential function in the limit where the neuronal membrane time-constant is small. Given normally distributed synaptic inputs, the network of LIF neurones then produces log-normally distributed firing-rates. For finite membrane time-constant the LIF f-I curve is closer to a power-law function than to an exponential and the firing-rate distribution deviates from log-normality; however, it remains strongly skewed and reasonably similar to a log-normal.
Finally, we note that in the framework of balanced networks it is easily possible to obtain average inhibitory firing rates that are larger than average excitatory ones, as observed in cortex.
The balanced state can apparently provide mechanistic explanations for some crucial properties of cortical activity. However, we want to indicate at this point two important open questions. First, Lerchner et al. (2006) showed that neuronal temporal irregularity in the balanced state can vary substantially in function of synaptic strength. Coefficients of variation both significantly lower and higher than unity can be obtained. Indeed, there is no reason to assume that neurones in the balanced state automatically operate in a sub-threshold regime; in principle the level of average synaptic input could be well above threshold. In that case, although the network would still be balanced, spiking irregularity would be reduced and firing rate distributions much less skewed. Thus, while the balanced state can potentially account for a number of properties of cortical activity, it does not give an answer to the question *why* cortical networks should operate in the biologically plausible regime.
The second issue is that, in the common view, balanced networks do not easily support the coexistence of different activity states. The dynamics of these networks tend to linearise the relationship between external stimuli and the neuronal response on the population level [van Vreeswijk and Sompolinsky (1998)]. In other words, the same external input will always elicit the same network response. However, as we have seen above, in delayed response tasks sensory input is identical during fixation and delay period, but memory cells change their activity significantly. It is not straightforward to explain this phenomenon by invoking bistability between two balanced states. The putative rigidness of balanced networks thus seems to be problematic in view of memory function.
All of these issues will be revisited in chapter 4. For now, we turn our attention to a short review of some recent attractor network models of working memory.

1.1.4 Attractor models of working memory

Several strategies have been used to account for the phenomena observed during WM delay tasks. On the one hand, models based on single neurone properties have been developed to explain the bistability of memory cells. Proposed mechanism include high-

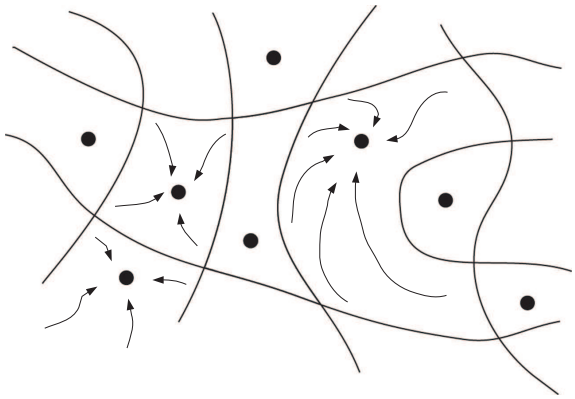

Figure 1.7: **Schematic representation of a multistable system's phase-space.** Dots represent fixed points; the solid lines confine different basins of attractions. Figure modified from Stewart (2011).

voltage activated calcium channels in combination with calcium activated cationic currents, post-spike after-depolarisation induced by neuromodulation and the non-linear current-voltage-curve of NMDA receptors [see e.g. Marder et al. (1996); Durstewitz et al. (2000a); Major and Tank (2004)]. Another line of research puts forward models of WM that explain persistent activity in terms of collective network states. Two different types of models have been developed: classical attractor networks and, more recently, networks that perform so-called reservoir computing [Jaeger (2001); Maass et al. (2002)]. Here, we will concentrate on recent attractor models of working memory. For a general review of the variety of working memory models see for example Durstewitz et al. (2000b), Wang (2001), Barbieri and Brunel (2008), Barak and Tsodyks (2014) and Mongillo (2014).
A standard way to model memory storage and retrieval mathematically is by making use of attractor networks [Amari (1977); Hopfield (1982); Amit (1992)]. The notion of 'attractor' refers to the portion of phase-space to which a dynamical system converges in the course of time. In this picture, memories correspond to attractors in the space of neuronal activity. During activation, that is during retrieval or, in the case of WM, active maintenance of a specific memory, the activity of all neurones converges to the pattern that represents that memory. Figure 1.7 shows a schematic representation of this principle. The plane on which the figure is sketched represents the phase space of the network, that is the ensemble of its firing rates. Each dot stands for a stable fixed point and the lines confine different basins of attractions. If the neuronal activity is set close to one of the attractor fixed points, for instance by a sensory input, the network's state will, when the external input is removed, evolve in time towards it (indicated by the arrows). The network will remain in that state, until some other input pushes it to another region of phase space. Since multiple attractors exist, the network can respond

in different ways to different inputs and can thereby keep track of the last one.
The attractor view elegantly provides another important property of memory: its auto-associative nature. When a memory is only partially hinted at, for instance one sees an item that vaguely resembles a memorised one, eventually this latter one is recalled or 'comes to mind'. This amounts to setting the neural activity somewhere in a basin of attraction in figure 1.7; the dynamics of the network neural then 'autocorrect' the neural activity over time, until it corresponds to the fixed-point activity.
In the attractor picture the above described dichotomy between passive and active memory - or long-term memory and short-term/working memory - arises naturally. The attractors themselves are here identified with robustly stored memories. The network dynamics can then recall, retrieve or activate these memories by entering the attractors. At this, the shape of the attractor landscape is determined by structured synaptic connectivity between the neurones. Roughly speaking, neurones that are highly active in the same memories tend to be stronger connected among each other than neurones that are not activated at the same time. This is, of course, nothing else but Hebb's principle [Hebb (1968)].
Many works in the past have employed such synaptic structuring in attractor networks with simplified binary neurones [see e.g. Amari (1977); Hopfield (1982); Tsodyks and Feigel'Man (1988); Amit (1992)]. These models yielded a good understanding of the general behaviour of networks with numerous attractors. In recent years, researches took further steps to develop more realistic models that reproduce qualitative aspects of experimental data obtained in WM tasks.
A very important study in this respect was performed by Amit and Brunel who for the first time studied analytically the associative memory properties of a network of LIF neurones [Amit and Brunel (1997), and see also Brunel (2004)]. In this model, inhibitory synaptic connectivity is unstructured, while a very simple form of Hebbian excitatory connectivity creates a certain number of non-overlapping sub-populations of excitatory neurones. Figure 1.8 illustrates this architecture. Each of the sub-populations contains a number of neurones that corresponds only to a small fraction f of the whole excitatory population. The network is able to switch from a baseline state, where all sub-populations fire at roughly the same rate, to various retrieval states, each in which one of the sub-populations fires at an elevated rate. Different sub-populations are of course activated for different inputs, that is, the network exhibits multistability between selective persistent activity states.
The neurones in both the baseline state and the various retrieval states feature realistic average firing rates, as shown in figure 1.8B. Moreover, it has been shown that the baseline state in Amit and Brunel's model corresponds to the balanced state we have introduced above [Brunel (2004)], so that pronounced temporal variability in this state is to be expected. However, a crucial shortcoming of this model is that the multistability is an effect of the network's finite size. Scaling up the number of neurones destroys the persistent state, that is all memory relevant properties are undone and only the baseline state remains.
Apart from explicitly considering scaling effects, Renart et al. (2007) modified Amit and Brunel's model by including inhibition into the network's clustering structure. With

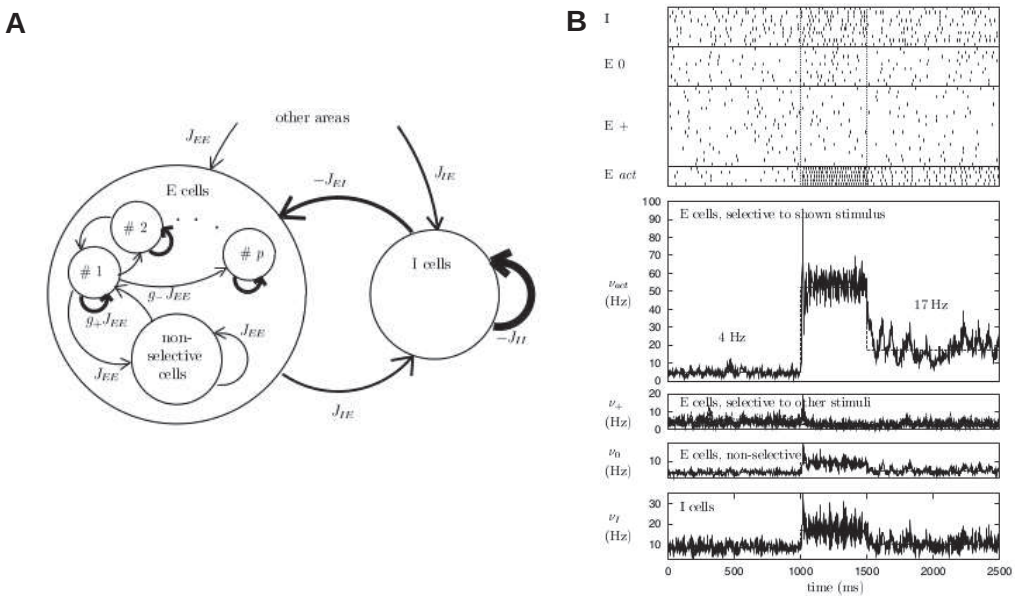

Figure 1.8: **Persistent activity model by Amit and Brunel.A)**: Scheme of the model's setup. **B)**: The model's behaviour in response to a stimulus presented between $t = 1.0 - 1.5$s. Pre-stimulus ($t = 0 - 1.0$s) and delay activity ($t = 1.5 - 2.5$s) are clearly distinct. The different panels show the activity of the different classes of neurones. Both panels modified from Brunel (2000).

this feature, persistent activation of a sub-population enhances firing of both excitatory and associated inhibitory neurones; balance can then be maintained even in the memory states. However, this model has difficulties in achieving memory states with high temporal fluctuations. While it allows robustly for multistability between baseline and very regular retrieval states, connectivity has to be fine-tuned to assure the appearance of irregular persistent activity. This fine-tuning problem increasingly aggravates when the network size is scaled up.

Another approach to solve the scaling problem was proposed in van Vreeswijk and Sompolinsky (2005). In this work, inhibition was kept unstructured as in the original model. To ensure that all network states remain balanced when the network is sized up, the authors introduced a scaling of the quantity f into their framework. As the number of neurones in the network increases, f is reduced accordingly, with the consequence that the absolute number of neurones participating in a sub-population grows slower than linear. Problematic here is again that the way f changes with the network size has to be finely tuned, implying that the multistability in this system is fragile.

Barbieri and Brunel (2007) employ a combination of two mechanisms to obtain baseline and memory states in a LIF network that are highly irregular. The first is a post-spike reset value that is close to the neuronal spiking threshold. This generally increases the probability of neurones to emit bursts of spikes, enhancing temporal irregularity. In addi-

tion, in order to avoid that neurones work in a supra-threshold regime, synapses between excitatory cells are provided with short-term depressing synapses which reduces synaptic strength of highly active afferent inputs [Tsodyks and Markram (1997)]. However, in this approach the level of fast noise seen by each neurone is not calculated self-consistently, but treated as a fixed parameter. Again, this parameter has to be tuned in order to produce the desired effects.
Finally, the work by Roudi and Latham (2007) introduces Hopfield-like attractors into a balanced network. This is done by endowing the excitatory-to-excitatory synapses with some structuring on top of random connectivity; inhibition is unstructured. The authors obtain a network where both background and memory retrieval state operate in the balanced regime. The neurones feature thus in both states a relatively high degree of temporal fluctuations, although it remains somewhat below the experimentally reported values. Although this study successfully demonstrates how a balanced network can be reliably be furnished with multiple memory patterns, it has several drawbacks. The synaptic connectivity's structured part has to be roughly an order of magnitude (or more) smaller than the unstructured part and has therefore to be finely tuned. Moreover, the number of patterns that can be stored is quite small.
We have seen that all of the above models exhibit, in one way or another, a fine-tuning problem. There is another problematic feature that these models have in common: they do not reproduce realistic firing-rate distributions. During delay activity, neurones in the persistently activated population fire - as intended - at higher rates than the rest of the network. The overall firing-rate distribution in this state is therefore bimodal. In addition, when the network is in the baseline state, the overall firing-rate distribution is significantly different from the one in the memory state. These characteristics are clearly inconsistent with the findings we have reviewed in section 1.1.2. The reason for these differences is ultimately the compartmentalisation of the network in different functional sub-populations.

A synaptic theory of balanced working memory

In this section, we consider a recent model by Hansel and Mato (2013) which manages to embed WM robustly into a balanced network, avoiding the fine-tuning problem. The authors developed a computational model that explains persistent activity and selectivity of PFC cells during ODR tasks. The mechanism employed in this work is based on short-term modulation of excitatory-to-excitatory synapses, which has been proposed in a previous publication by Mongillo et al. (2012). Synaptic short-term-plasticity (STP) is indeed a widespread phenomenon in prefrontal pyramidal-to-pyramidal synapses [Hempel et al. (2000); Wang et al. (2006)].
As we have seen, the population level responses in balanced networks depend linearly on the external inputs to the network, thereby impeding multistability among balanced states to arise. The solution suggested here relies on non-linearities introduced into the system by facilitating synapses. In function of the level of pre-synaptic activity, the release probability of synapses endowed with STP changes and therefore the effective synaptic weight is modulated. For parameters that correspond to strongly facilitating

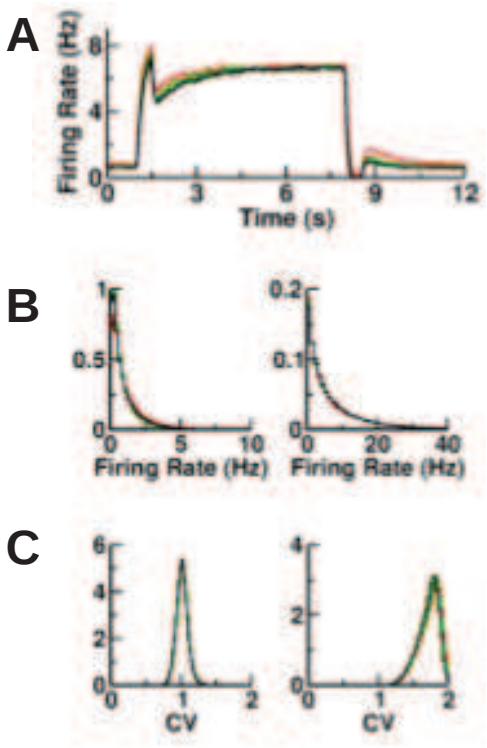

Figure 1.9: **Bistability between balanced states with facilitating synapses.A)**: Change of the average firing-rate during the delay period. **B)**: Firing-rate distribution during baseline (left) and delay period (right). **C)**: Distribution of the inter-spike-intervals during baseline (left) and delay period (right). Figure adapted from Hansel and Mato (2013).

synapses, this leads to a scenario that features a bistability between a low-activity state with unmodulated, weak synapses and a high-activity state with facilitated synapses. Figure 1.9A shows the bistable response properties of a simulated excitatory population; the average firing rate increases in response to a transient input that facilitates the synapses. Note that the two states have very different global properties, as can be seen from their firing rate distributions (panel B) and their temporal variability (panel C). This mechanism lies the basis for the model in Hansel and Mato (2013) where an additional connectivity structure similar to the classical ring-model [Ben-Yishai et al. (1995)] is imposed on the synaptic weights in order to obtain spatial selectivity. A whole series of experimental findings is successfully described by this work, as, for instance, the difference in irregularity of firing between baseline and high-activity state reported by some authors [Compte et al. (2003)] and the diversity of neuronal tuning-curves [Funahashi et al. (1989, 1990)]. Furthermore, the functioning of the model is robust with respect to

changes in several parameters.
However, two potentially problematic issues have to be raised. First, as we have seen above, experimental evidence suggests that the overall properties of the firing-rate distributions of task-relevant populations do not change. However, as can clearly be seen in figure 1.9B, this model predicts a significant change in the overall firing-rate statistics, like the others we encountered above. Second, it is not clear whether the regime of synaptic parameters required for the functioning of the model is consistent with experimental findings. The facilitation-based mechanism described above requires synaptic release probabilities that are smaller than 0.12. Analysis of prefrontal pyramidal-to-pyramidal connections, however, indicates that this value ranges between $0.15 - 0.35$ [Wang et al. (2006)].
We will revisit the first of these issues in chapter 4, but will turn our attention first to the discrepancy between required and measured release probabilities. Reasonable doubt in the statistical methodology used to obtain parameter estimates from synaptic recordings motivated us to develop a new theoretical framework for the description of synaptic transmission and the analysis of experimental data. This will be the topic of chapters 2 and 3. We have therefore to put the matter of memory aside for now and introduce the basic concepts of synaptic short-term plasticity.

1.2 Synaptic short-term plasticity

Short-term plasticity (STP) is the transient modification of post-synaptic responses at chemical synapses induced by pre-synaptic activity on time scales of hundreds of milliseconds to several seconds. Commonly two types of STP are distinguished: short-term depression and short-term facilitation. Depression, referring to the activity dependent reduction of synaptic efficacy, is thought to be primarily caused by depletion of synaptic vesicles. Facilitation, the use-dependent enhancement of synaptic transmission, is a consequence of enhanced probability of neurotransmitter release, which, in turn, depends mostly on the spike-triggered calcium-influx into the pre-synaptic terminal.
Since STP is effectively a necessary consequence of synaptic physiology, virtually all synapses display some form of STP, which is often a mixture of depression and facilitation. However, strikingly difference types of STP have been found that differ systematically in function of cell types of pre- and post-synaptic partner, cortical region and age [e.g. Reyes et al. (1998); Blackman et al. (2013)]. This suggests that rather than being a physiological byproduct, STP is used by the central nervous system and tuned for specific ends [but see Borst (2010)]. Indeed, in numerous experimental and theoretical studies STP has been implicated in a broad range of functional roles, for instance gain control [Abbott et al. (1997)], temporal filtering [Goldman et al. (2002); Klyachko and Stevens (2006); Rosenbaum et al. (2012)] and effects on the dynamics of attractor networks, as we have seen above.
The next chapter is dedicated to a review of the most important STP models. In the course of this, we will encounter a great variety of mechanism that contribute to its phenomenology. For detailed reviews of STP mechanism and the various functions STP has

been implied in see, for instance, Zucker and Regehr (2002), Abbott and Regehr (2004) and Tsodyks and Wu (2013).

Bibliography

Abbott, L. and Regehr, W. G. (2004). Synaptic computation. *Nature*, 431(7010):796–803.

Abbott, L., Varela, J., Sen, K., and Nelson, S. (1997). Synaptic depression and cortical gain control. *Science*, 275(5297):221–224.

Amari, S.-i. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological cybernetics*, 27(2):77–87.

Amit, D. J. (1992). *Modeling brain function: The world of attractor neural networks*. Cambridge University Press.

Amit, D. J. and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, 7(3):237–252.

Baddeley, A. (1992). Working memory. *Science*, 255(5044):556–559.

Bair, W., Koch, C., Newsome, W., and Britten, K. (1994). Power spectrum analysis of bursting cells in area mt in the behaving monkey. *The Journal of neuroscience*, 14(5):2870–2892.

Barak, O. and Tsodyks, M. (2014). Working models of working memory. *Current opinion in neurobiology*, 25:20–24.

Barbieri, F. and Brunel, N. (2007). Irregular persistent activity induced by synaptic excitatory feedback. *Frontiers in Computational Neuroscience*, 1:5.

Barbieri, F. and Brunel, N. (2008). Can attractor network models account for the statistics of firing during persistent activity in prefrontal cortex? *Frontiers in neuroscience*, 2:3.

Bauer, R. H. and Fuster, J. M. (1976). Delayed-matching and delayed-response deficit from cooling dorsolateral prefrontal cortex in monkeys. *Journal of comparative and physiological psychology*, 90(3):293.

Beloozerova, I. N., Sirota, M. G., and Swadlow, H. A. (2003). Activity of different classes of neurons of the motor cortex during locomotion. *The Journal of neuroscience*, 23(3):1087–1097.

Ben-Yishai, R., Bar-Or, R. L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, 92(9):3844–3848.

Blackman, A. V., Abrahamsson, T., Costa, R. P., Lalanne, T., and Sjöström, P. J. (2013). Target-cell-specific short-term plasticity in local circuits. *Frontiers in synaptic neuroscience*, 5.

Borst, J. G. G. (2010). The low synaptic release probability *in vivo*. *Trends in neurosciences*, 33(6):259–266.

Brunel, N. (2000). Persistent activity and the single-cell frequency-current curve in a cortical network model. *Network: Computation in Neural Systems*, 11(4):261–280.

Brunel, N. (2004). Network models of memory. *Methods and models in neurophysics. Paris: Les Houches*.

Buzsáki, G. and Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. *Nature Reviews Neuroscience*.

Compte, A., Constantinidis, C., Tegnér, J., Raghavachari, S., Chafee, M. V., Goldman-Rakic, P. S., and Wang, X.-J. (2003). Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *Journal of Neurophysiology*, 90(5):3441–3454.

Connors, B., Gutnick, M., and Prince, D. (1982). Electrophysiological properties of neocortical neurons in vitro. *J Neurophysiol*, 48(6):1302–1320.

Destexhe, A., Rudolph, M., and Paré, D. (2003). The high-conductance state of neocortical neurons in vivo. *Nature reviews neuroscience*, 4(9):739–751.

Durstewitz, D., Seamans, J. K., and Sejnowski, T. J. (2000a). Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *Journal of Neurophysiology*, 83(3):1733–1750.

Durstewitz, D., Seamans, J. K., and Sejnowski, T. J. (2000b). Neurocomputational models of working memory. *Nature neuroscience*, 3:1184–1191.

Fujisawa, S., Amarasingham, A., Harrison, M. T., and Buzsáki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nature neuroscience*, 11(7):823–833.

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1990). Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. *J Neurophysiol*, 63(4):814–831.

Funahashi, S., Bruce, C. J., Goldman-Rakic, P. S., et al. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol*, 61(2):331–349.

Fuster, J. M. (1988). *Prefrontal cortex*. Springer.

Fuster, J. M. (1995). *Memory in the cerebral cortex*. Cambridge, MA: MIT Press.

Fuster, J. M. (1997). Network memory. *Trends in neurosciences*, 20(10):451–459.

Fuster, J. M. (2009). Cortex and memory: emergence of a new paradigm. *Journal of Cognitive Neuroscience*, 21(11):2047–2072.

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

Fuster, J. M., Bauer, R. H., and Jervey, J. P. (1985). Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. *Brain research*, 330(2):299–307.

Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F., and Petersen, C. C. (2010). Membrane potential dynamics of gabaergic neurons in the barrel cortex of behaving mice. *Neuron*, 65(3):422–435.

Goldman, M. S., Maldonado, P., and Abbott, L. (2002). Redundancy reduction and sustained firing with stochastic depressing synapses. *The Journal of neuroscience*, 22(2):584–591.

Goldman-Rakic, P. (1995). Cellular basis of working memory. *Neuron*, 14(3):477–485.

Haider, B., Duque, A., Hasenstaub, A. R., and McCormick, D. A. (2006). Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *The Journal of neuroscience*, 26(17):4535–4545.

Hansel, D. and Mato, G. (2013). Short-term plasticity explains irregular persistent activity in working memory tasks. *The Journal of Neuroscience*, 33(1):133–149.

Hansel, D. and van Vreeswijk, C. (2012). The mechanism of orientation selectivity in primary visual cortex without a functional map. *The Journal of Neuroscience*, 32(12):4049–4064.

Hebb, D. (1968). *0.(1949) The organization of behavior*. Wiley, New York.

Hempel, C. M., Hartman, K. H., Wang, X.-J., Turrigiano, G. G., and Nelson, S. B. (2000). Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex. *Journal of neurophysiology*, 83(5):3031–3041.

Holt, G. R., Softky, W. R., Koch, C., and Douglas, R. J. (1996). Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *Journal of Neurophysiology*, 75(5):1806–1814.

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

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

Isaacson, J. S. and Scanziani, M. (2011). How inhibition shapes cortical activity. *Neuron*, 72(2):231–243.

Jaeger, H. (2001). The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148:34.

Kesner, R. P., Hunt, M. E., Williams, J. M., and Long, J. M. (1996). Prefrontal cortex and working memory for spatial response, spatial location, and visual object information in the rat. *Cerebral Cortex*, 6(2):311–318.

Klyachko, V. A. and Stevens, C. F. (2006). Excitatory and feed-forward inhibitory hippocampal synapses work synergistically as an adaptive filter of natural spike trains. *PLoS biology*, 4(7):e207.

Lerchner, A., Ursta, C., Hertz, J., Ahmadi, M., Ruffiot, P., and Enemark, S. (2006). Response variability in balanced cortical networks. *Neural computation*, 18(3):634–659.

Maass, W., Natschläger, T., and Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation*, 14(11):2531–2560.

Major, G. and Tank, D. (2004). Persistent neural activity: prevalence and mechanisms. *Current opinion in neurobiology*, 14(6):675–684.

Marder, E., Abbott, L., Turrigiano, G. G., Liu, Z., and Golowasch, J. (1996). Memory from the dynamics of intrinsic membrane currents. *Proceedings of the national academy of sciences*, 93(24):13481–13486.

Miller, E. K. and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, 24(1):167–202.

Miller, E. K., Erickson, C. A., and Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *The Journal of Neuroscience*, 16(16):5154–5167.

Mitchell, J. F., Sundberg, K. A., and Reynolds, J. H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area v4. *Neuron*, 55(1):131–141.

Miyashita, Y. and Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, 331(6151):68–70.

Mizuseki, K. and Buzsáki, G. (2013). Preconfigured, skewed distribution of firing rates in the hippocampus and entorhinal cortex. *Cell reports*, 4(5):1010–1021.

Mongillo, G. (2014). Models of working memory.

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

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

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

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. D. (2010). The asynchronous state in cortical circuits. *science*, 327(5965):587–590.

Renart, A., Moreno-Bote, R., Wang, X.-J., and Parga, N. (2007). Mean-driven and fluctuation-driven persistent activity in recurrent networks. *Neural computation*, 19(1):1–46.

Reyes, A., Lujan, R., Rozov, A., Burnashev, N., Somogyi, P., and Sakmann, B. (1998). Target-cell-specific facilitation and depression in neocortical circuits. *Nature neuroscience*, 1(4):279–285.

Romo, R., Brody, C., Hernández, A., and Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, 399:470–473.

Rosenbaum, R., Rubin, J., and Doiron, B. (2012). Short term synaptic depression imposes a frequency dependent filter on synaptic information transfer. *PLoS computational biology*, 8(6):e1002557.

Roudi, Y. and Latham, P. E. (2007). A balanced memory network. *PLoS computational biology*, 3(9):e141.

Roxin, A., Brunel, N., Hansel, D., Mongillo, G., and van Vreeswijk, C. (2011). On the distribution of firing rates in networks of cortical neurons. *The Journal of Neuroscience*, 31(45):16217–16226.

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

Shadlen, M. N. and Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *The Journal of neuroscience*, 18(10):3870–3896.

Shafi, M., Zhou, Y., Quintana, J., Chow, C., Fuster, J., and Bodner, M. (2007). Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience*, 146(3):1082–1108.

Shinomoto, S., Kim, H., Shimokawa, T., Matsuno, N., Funahashi, S., Shima, K., Fujita, I., Tamura, H., Doi, T., Kawano, K., et al. (2009). Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLoS Computational Biology*, 5(7):e1000433.

Shinomoto, S., Sakai, Y., and Funahashi, S. (1999). The ornstein-uhlenbeck process does not reproduce spiking statistics of neurons in prefrontal cortex. *Neural Computation*, 11(4):935–951.

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

Softky, W. R. and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *The Journal of Neuroscience*, 13(1):334–350.

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

Stewart, I. (2011). Sources of uncertainty in deterministic dynamics: an informal overview. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1956):4705–4729.

Tsodyks, M. and Feigel'Man, M. (1988). The enhanced storage capacity in neural networks with low activity level. *EPL (Europhysics Letters)*, 6(2):101.

Tsodyks, M. and Wu, S. (2013). Short-term synaptic plasticity. *Scholarpedia*, 8(10):3153.

Tsodyks, M. V. and Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the National Academy of Sciences*, 94(2):719–723.

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

van Vreeswijk, C. and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural computation*, 10(6):1321–1371.

van Vreeswijk, C. and Sompolinsky, H. (2005). Irregular activity in large networks of neurons. In Chow, C., Gutkin, B., Hansel, D., Meunier, C., and Dalibard, J., editors, *Les Houches Lectures LXXX on Methods and models in neurophysics*, pages 341–402, London. Elsevier.

Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in neurosciences*, 24(8):455–463.

Wang, Y., Markram, H., Goodman, P. H., Berger, T. K., Ma, J., and Goldman-Rakic, P. S. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nature neuroscience*, 9(4):534–542.

Wohrer, A., Humphries, M. D., and Machens, C. K. (2012). Population-wide distributions of neural activity during perceptual decision-making. *Progress in neurobiology*, 103:156–193.

Wolf, F., Engelken, Rainer nd Puelma-Touzel, M., Weidinger, J. D. F., and Neef, A. (2014). Dynamical models of cortical circuits. *Current opinion in neurobiology*, 25:228–236.

Zucker, R. S. and Regehr, W. G. (2002). Short-term synaptic plasticity. *Annual review of physiology*, 64(1):355–405.

Chapter 2

Phenomenological Models of Short-term Plasticity

In this section we wish to achieve two things. First, we want to review phenomenological ways to account of STP. Instead of enumerating models that we can find in the literature we want to review them with a certain perspective in mind. Our viewpoint is based on the observation that all existing phenomenological descriptions of short-term plasticity, however sophisticated, ignore a fundamental property of synaptic responses: their stochasticity. Rather, almost all modelling approaches try to explain average responses, disregarding the quantal nature of synaptic transmission. Our goal here is to show that this limitation can be overcome since virtually all phenomenological descriptions found in the literature can be embedded in a quantal release scheme. This is possible since all can be cast in a release-site formulation. We will thus go over a number of representative STP models and show how they can be interpreted as descriptions of probabilistic transitions between different release-site-states. This procedure of reformulation in a common framework will make it also quite easy to compare them.
Second, based on the notion of the release-site, we want to introduce a novel modelling framework that makes it possible to model synaptic responses in its entirety. That is, we propose an approach which describes not only average responses, but whole distributions of response sequences. Since all models reviewed here can be formulated as release-site models, it follows that they all can be transformed in a fully probabilistic version.
It seems fit at this point to make some remarks about the rationale of our interest in phenomenological models. We wish to consider those modelling work that tries to capture STP without retracing the fathomless complexity of the molecular machineries involved in synaptic transmission. Instead, we concentrate on models that indeed are biologically inspired, but deliberately omit a substantial amount of biological details. From this kind of reductive approach one obtains descriptions of synaptic responses that are mathematically relatively simple and have a small number of free parameters. This class of models can be named 'phenomenological models'. Of course, one cannot draw a neat separation between 'phenomenological' and 'realistic' descriptions, but we will concentrate on the simpler end of the spectrum.

What are reasons to use phenomenological models? First, we should not use models that are very complex when this is not justified by the available data. Especially for small central synapses, it is virtually impossible to obtain enough electro-physiological and anatomical data to constrain a model which aims at capturing molecular details. Thus, modelling simplifications have inevitably to be made; a point that will be further elucidated in the next section.
Second, it may be that a phenomenological model captures all behaviour that is functionally relevant for the specific scientific question that is studied. It could even turn out that the complex molecular machinery of the synapse *always* generates dynamically low-dimensional responses, sealing the molecular level from higher ones. A phenomenological description would then comprise all that is needed to study higher level functions. This is analogous to the integrate-and-fire neuron model which despite its simplicity reproduces a wide range of relevant phenomena.
Third, phenomenological models are simply much handier than detailed biological descriptions (see for instance Pan and Zucker (2009); Nadkarni et al. (2012)). Featuring a low-dimensional parameter- and state-space they can be treated both numerically and analytically with less effort. This is in particular important in the research of neural networks where preferably low-dimensional, efficient models of synaptic transmission have to be incorporated in large numerical simulations.

2.1 The release-site formalism

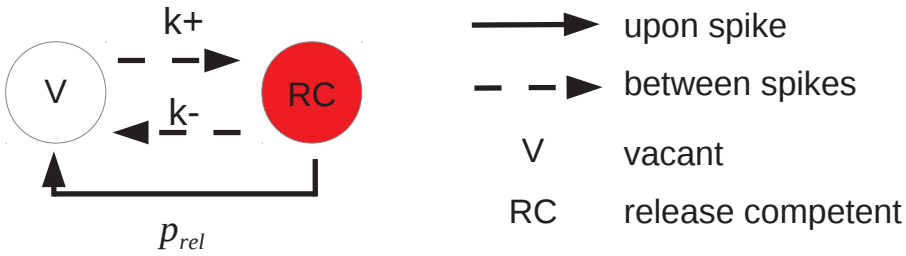

Figure 2.1

The release-site interpretation we wish to impose is tightly associated with the quantal model (Del Castillo and Katz (1954)). In its simplest form, the quantal model pictures a synapse as a collection of N identical, statistically independent sites which, upon spike, can either release one vesicle of neurotransmitter or fail to do so. Consider a single site's total probability of release p_R upon spike. Commonly, this quantity is decomposed into two parts: the probability that the site is occupied, p_{occ}, and the probability that the

site releases given that it is occupied, p_{rel}. We can write:

$$p_R = p_{rel} \cdot p_{occ}. \tag{2.1}$$

This description assumes that the release site can be in one of two states: occupied by one vesicle (release-competent) or vacant (refractory). At this, 'occupied' indeed means that all docking and priming processes involving vesicle and site are completed, and that for release to occur, no other processes are needed except those that are directly triggered by a pre-synaptic spike. On the other hand, in the refractory state a spike can never trigger release. In order to formalise the two states, we introduce a binary occupation state variable ξ, where $\xi = 1$ denotes the release-competent state and $\xi = 0$ the refractory state. Consequently, p_{occ} is the probability that $\xi = 1$.
These considerations lead to a commonly accepted single release-site scheme (see e.g. Heinemann et al. (1993); Weis et al. (1999); Wang (1999); Neher and Sakaba (2008); Neher (2010)) which we wish to use throughout our review. It is shown in figure 2.1. The possible transitions between the two occupational states comprise vesicle release as well as vesicle docking/undocking processes. At this, we have to distinguish two types of transitions: those that occur upon stimulation (upon spike), and those that occur in the time interval Δt between stimulations (in between spikes). If $\xi = 1$, the site can go to the refractory state either by release of its vesicle upon spike (with p_{rel}), or by vesicle unbinding in between spikes (with $\Delta t \cdot k^-$). If $\xi = 0$, the site either binds a vesicle in between spikes (with $\Delta t \cdot k^+$) or remains refractory. In the following we will assume (unless stated otherwise) that the vesicle supply is unlimited; transition probabilities cannot be altered by vesicle shortage. Two things are noteworthy: first, p_{occ} is now fully replaced by the rates k^+ and k^-. Second, in this scheme the release site is clearly *stochastic*.

2.1.1 Depletion

We start our review with simple models and will progressively add more details. The most basic short-term effect on synaptic efficacy is depression due to depletion of vesicles. In the release-site scheme, this corresponds to:

$$\begin{aligned} k^- &= 0 \\ k^+ &= const \\ p_{rel} &= const. \end{aligned} \tag{2.2}$$

The fact that $k^- = 0$ means that once a vesicle is docked at a release-site, it must stay there. Thus, for sufficiently long inter-spike intervals $\Delta t = t_{k+1} - t_k$, this scheme dictates that the probability of occupancy p_{occ} equals 1. As a consequence, in simple depletion models all release sites are in the release-competent state in absence of pre-synaptic activity (for example before stimulation in an *in-vitro* experiment).
Models that can be formulated in terms of equations 2.2 are for instance Liley and North (1953), Betz (1970) and Tsodyks and Markram (1997), which apply it, respectively, to the rat or frog neuromuscular junction and to synapses between rat layer 5 somatosensory

pyramidal neurons. The latter work predicts that depletion causes steady-state responses to be inversely proportional to the input frequency.

2.1.2 Activity-dependant release probability

The next more complex class of models features a release probability that depends on the activation-history of the synapse. Our scheme changes accordingly:

$$\begin{aligned} k^- &= 0 \\ k^+ &= const \\ p_{rel} &= p_{rel}(t). \end{aligned} \tag{2.3}$$

Commonly, one considers activity-dependant changes that are increasing p_{rel}; this mechanism is thought to underlie synaptic facilitation. However, processes that reduce p_{rel} in function of synaptic stimulation have also been captured in a number of models. We will consider facilitation first.

Simple facilitation

One of the most prominent and popular models of short-term plasticity has been conceived by Tsodyks and Markram (Tsodyks and Markram (1997); Markram et al. (1998)). It describes synaptic transmission that features both a depressing and facilitating component and hassub been successfully applied to a wide range of preparations (e.g. Fuhrmann et al. (2004); Loebel et al. (2009); Wang et al. (2006)). It is, due to its simplicity, often used in theoretical studies exploring the functional role of STP (Tsodyks et al. (1998); Fuhrmann et al. (2002); Mongillo et al. (2008); Pfister et al. (2010); Mongillo et al. (2012); Cortes et al. (2013); Hansel and Mato (2013)). The Tsodyks-Markram (TM) model can be accommodated in our scheme by setting:

$$\begin{aligned} k^- &= 0 \\ k^+ &= \frac{1}{\tau_D} \end{aligned} \tag{2.4}$$

where τ_D is the typical time constant of the persistence of depression. The activity-dependant dynamics of p_{rel} are given by the following differential equation:

$$\dot{p}_{rel}(t) = \frac{p_0 - p_{rel}(t)}{\tau_F} + p_0(1 - p_{rel}(t)) \sum_k \delta(t - t_k). \tag{2.5}$$

According to this, release probability increases upon spikes (occurring at times t_k), and then decays back to its baseline level, p_0, with a time constant τ_F in between spikes. We see that short-term depression arises in the TM-model from the simplest possible form, namely release-site depletion. On the other hand, short-term facilitation is solely determined by an activity-dependent increase of p_{rel}, which can be, for instance, interpreted as caused by the elevation of the residual Ca^{2+} concentration. If we set τ_F to a very small value, we obtain the purely depressing model mentioned above.

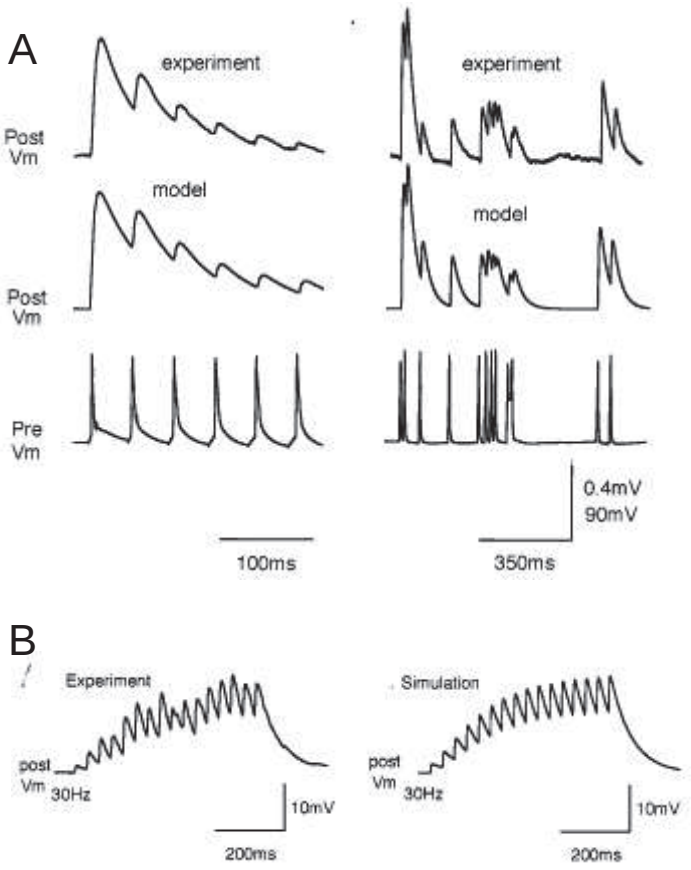

Figure 2.2: **Behaviour of the Tsodyks-Markram model. A)**: Left: Experimentally measured postsynaptic potentials (top) generated by a regular spike train (Bottom). Model fit without facilitation component in the middle. Right: The same with irregular spike train. **B)**: Data and model traces for a facilitating connection. Panel A, B and captions adapted from Tsodyks and Markram (1997) and Markram et al. (1998), respectively.

Note that in the above formulation, the increase in p_{rel} is proportional to the initial release probability p_0. In general, however, we can consider an increase Δ_p, yielding

$$\dot{p}_{rel}(t) = \frac{p_0 - p_{rel}(t)}{\tau_F} + \Delta_p(1 - p_{rel}(t)) \sum_k \delta(t - t_k), \tag{2.6}$$

as has been employed in Costa et al. (2013) and Hennig (2013). Another simple model belonging to this class has been used in Varela et al. (1997).

Bertram's facilitation

Depending on the scientific question considered, facilitation of the Tsodyks-Markram type might be too simplistic. However, the scheme given by equations 2.3 is readily compatible with more biophysically realistic forms of facilitation. Following Bertram et al. (1996), facilitation can be understood as resulting from the cooperative action of four different calcium binding sites. To trigger release, four calcium ions have to bind to the release machinery. The unbinding from the four sites occurs then at different timescales, allowing for the possibility of partially bound sites the next time an action potential triggers calcium influx. Considering its average effect, this mechanism can be expressed by setting:

$$p_{rel}(t) = F_1(t) \cdot F_2(t) \cdot F_3(t) \cdot F_4(t), \tag{2.7}$$

where

$$\dot{F}_i(t) = -\frac{F_i(t)}{\tau_{F_i}} + b_i^+ \cdot Ca(t) \cdot (1 - F_i(t)), \quad \tau_{F_i} = (b_i^+ \cdot Ca(t) + b_i^-)^{-1}. \tag{2.8}$$

At this, $F_i(t)$ denotes the probability that the ith calcium binding site is bound, b_i^+ and b_i^- are the corresponding binding and unbinding rates, and $Ca(t)$ is the calcium-concentration at the release site. The factor $1 - F_i(t)$ guarantees that the probabilities do not grow beyond unity.[1] As shown in Bertram et al. (1996), these relations capture the fourth-power relationship between calcium-concentration and release probability that has been observed by Dodge and Rahamimoff (1967) at the frog neuromuscular junction. In the model of Wang (1999) a simplified version of equations 2.7 and 2.8 is used to describe neocortical synapses between pyramidal cells and interneurons, as described by Thomson et al. (1993). The dynamics of the four gates are combined into one variable and accordingly, only one time constant is considered:

$$\dot{F}(t) = -\frac{F(t)}{\tau_F} + Ca(t) \cdot (1 - F(t)) = -\frac{F(t)}{\tau_F} + \Delta_C (1 - F(t)) \sum_k \delta(t - t_k). \tag{2.9}$$

Here, Δ_C is the instantaneous increase in F upon spike when due to calcium influx. To preserve the aforementioned fourth-power relationship the release probability is set to:

$$p_{rel}(t) = p_0 \cdot F^4(t). \tag{2.10}$$

In this formulation, $Ca(t)$ is given by a delta-pulse at spike-time. What is hence considered in this model is the effect due to a local calcium-concentration transient, while the residual calcium-concentration plays no role. A very similar version of this model has been used by Matveev and Wang (2000a), where three time-constants are considered. The explicit effect of the residual calcium-concentration has been included in a further

[1] In this formulation, the $F_i(t)$ are meant to be updated before the release-transition is performed, otherwise $p_{rel} = 0$ upon the first spike.

development of the Wang (1999) model. In Hempel et al. (2000), the form of $Ca(t)$ is changed according to:

$$Ca(t) = \Delta_C \sum_k \delta(t - t_k) + Ca_{gl}(t), \tag{2.11}$$

where the temporal development of the global calcium-concentration is given by:

$$\dot{Ca}_{gl}(t) = -\frac{Ca_{gl}(t)}{\tau_C} + \Delta_{CG} \sum_k \delta(t - t_k). \tag{2.12}$$

While the effect of the local calcium-concentration Δ_C is restricted to spike-times, the global calcium-signal only falls off on a timescale of τ_C, after being raised by $\Delta_{CG} < \Delta_C$ upon spike.

Reduction of p_{rel}

Activity dependant reduction of p_{rel} has been observed in various preparations (see e.g. Betz (1970); Wu and Borst (1999)). Possible mechanism that generally come into consideration are inactivation of calcium-channels (Forsythe et al. (1998); Xu and Wu (2005)) and the activation of pre-synaptic autoreceptors (CITE Zucker and Regehr (2002)). A simple description of these effects was included in Fuhrmann et al. (2004) and is basically an inversion of the Tsodyks-Markram facilitation prescription:

$$\dot{p}_{rel}(t) = \frac{p_0 - p_{rel}(t)}{\tau_{inac}} - \Delta_{inac} \cdot p_{rel}(t) \sum_k \delta(t - t_k). \tag{2.13}$$

A slightly different formulation is chosen in Billups et al. (2005) and Hennig (2013) to model mGluR autoreceptor activation. Here, p_{rel} is not decreased directly, but by modulating p_0:

$$\dot{p}_0(t) = \frac{p_{0,ini} - p_0(t)}{\tau_{p_0}} - \Delta_{p_0} \cdot p_0(t) \sum_k \delta(t - t_k). \tag{2.14}$$

In general, Δ_{p_0} does not need to be a constant, but can depend on the amount of released neurotransmitter. An extended version of this model that takes into account different autoreceptor states has been developed in Hennig et al. (2008).

Combining processes that modulate p_{rel}

Equation 2.7 illustrates how p_{rel} can be facilitated by different components. Obviously, processes that depress synaptic transmission by reducing p_{rel} can be added in the same way. This has been implemented in the model by Varela et al. (1997) where various (unspecified) modulatory processes are combined. For i facilitating and j depressing components we can write:

$$p_{rel}(t) = p_0 \cdot F_1(t) \cdot \ldots \cdot F_i(t) \cdot D_1(t) \ldots \cdot D_j(t). \tag{2.15}$$

with

$$\begin{aligned}\dot{F}_i(t) &= \frac{1-F_i(t)}{\tau_{F_i}} + \Delta_{F_i}\cdot(1-F_i)\sum_k \delta(t-t_k) \\ \dot{D}_j(t) &= \frac{1-D_j(t)}{\tau_{D_j}} - \Delta_{D_j}\cdot(1-D_j)\sum_k \delta(t-t_k).\end{aligned} \tag{2.16}$$

2.1.3 Activity-dependant recovery from depression

In the models discussed so far, the only time-dependant transition-probability is p_{rel}. Things become more complex when we allow for a time-dependence of k^+, that is, for an activity-dependant recovery of vesicles. Adding this process to our scheme yields:

$$\begin{aligned}k^- &= 0 \\ k^+ &= k^+(t) \\ p_{rel} &= p_{rel}(t).\end{aligned} \tag{2.17}$$

This mechanism was put into a mathematical form in Dittman and Regehr (1998) and Dittman et al. (2000). The adoption to our scheme reads:

$$\begin{aligned}k^- &= 0 \\ k^+(t) &= k_0 + \frac{k_{max}-k_0}{1+\frac{K_D}{CaX_D(t)}} \\ p_{rel}(t) &= p_0 + \frac{1-p_0}{1+\frac{K_F}{CaX_F(t)}}.\end{aligned} \tag{2.18}$$

Here, we distinguish two calcium related processes: $CaX_F(t)$ and $CaX_D(t)$ denote the concentrations of some calcium-receptor molecules that are responsible, respectively, for facilitated release (with an initial probability of p_0 and a maximal value of 1) and increased vesicle recruitment (that ranges between k_0 and k_{max}). $CaX_F(t)$ and $CaX_D(t)$ determine the increase in release probability and binding rate not in a linear way, as seen in the models so far, but by means of a Michaelis-Menten-type relation. At this, the constant K_F (K_D) denotes the value of the concentration $CaX_F(t)$ ($CaX_D(t)$) at which p_{rel} (k^+) has increased halfway between its baseline and maximum value. The values of $CaX_F(t)$ and $CaX_D(t)$ are given by:

$$\begin{aligned}\dot{CaX}_F(t) &= -\frac{CaX_F(t)}{\tau_{XF}} + \Delta_{XF}\sum_k \delta(t-t_k), \\ \dot{CaX}_D(t) &= -\frac{CaX_D(t)}{\tau_{XD}} + \Delta_{XD}\sum_k \delta(t-t_k).\end{aligned} \tag{2.19}$$

The symbols are analogous to the ones used in equation 2.12.
When compared to models with constant k^+, Dittman's scheme predicts much less depression after long stimulus trains, i.e. when the synapse reaches steady-state behaviour.

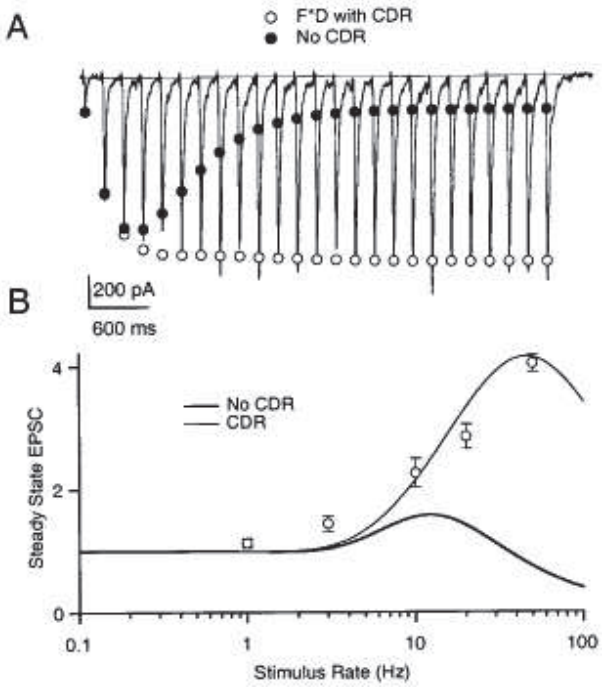

Figure 2.3: **Behaviour of Dittman's model. A)**: Trace represents a single trial of Parallel fiber to Purkinje cell EPSCs recorded during 25 stimuli at 50 Hz. Open circles are the model fit, filled circles represent model prediction with the same parameters but without calcium-dependent recovery from depression (CDR). **B)**: Steady-state EPSC size plotted against stimulus frequency for the parallel fiber synapse with (thin line) and without (thick line) CDR. Open circles represent parallel fiber data. Figure and caption adapted from Dittman et al. (2000).

This is illustrated in figure 2.3. Indeed, in Dittman et al. (2000) it was shown that the model can successfully capture average transient as well as steady state responses of climbing fibre to Purkinje cell synapses, parallel fibre to Purkinje cell synapses (see figure 2.3) and Schaffer collateral to CA1 pyramidal-neuron synapses.
A very similar model by Lee et al. (2009) reduces equations 2.19 to a single calcium variable $Ca(t)$ with baseline calcium-level Ca_0:

$$\dot{Ca}(t) \quad = \quad \frac{Ca_0 - Ca(t)}{\tau_C} + \Delta_C \sum_k \delta(t - t_k). \tag{2.20}$$

In this formulation, the dynamics of facilitation and recovery from depression are thus unified. Furthermore, k^- and k^+ are adopted without change, while p_{rel} reads:

$$p_{rel}(t) \quad = \quad p_{max} \cdot \frac{1}{1 + (\frac{K_F}{Ca(t)})^4}, \tag{2.21}$$

where we encounter anew the fourth-power relationship between calcium-concentration and release-probability found by Dodge and Rahamimoff (1967). In addition, p_{rel} is upper

bounded by a maximal release probability p_{max}. The model describes appropriately the responses of the climbing fibre to Purkinje cell synapse, the calyx of Held and synapses between neo-cortical pyramidal neurons, predicting in particular the resonance-frequency of steady-state responses.
A more phenomenological description of activity-dependent recovery from depression has been proposed in Fuhrmann et al. (2004):

$$k^+(t) = \frac{1}{\tau_D(t)}, \tag{2.22}$$

where $\tau_D(t)$ is determined by:

$$\dot{\tau}_D(t) = \frac{\tau_{D0} - \tau_D(t)}{\tau_\tau} - \Delta_\tau \cdot \tau_D(t) \sum_k \delta(t - t_k). \tag{2.23}$$

In contrast to Dittman's model, k^+ can in principle grow without bounds (as τ_D approaches zero).
Other works that model activity-dependant recovery from depression in a similar way are Hennig et al. (2008) and Yang and Xu-Friedman (2008). They describe calyx of Held responses over a wide range of stimulation frequencies.

2.1.4 Non-zero unbinding rate

Experimental evidence from several preparations indicates that vesicle binding to release-sites and/or vesicle priming is reversible[2] (Murthy and Stevens (1999); Zenisek et al. (2000); Nofal et al. (2007)). Models allowing vesicles to undock from a release-site, obey the following scheme:

$$\begin{aligned} k^- &= const \\ k^+ &= k^+(t) \\ p_{rel} &= p_{rel}(t). \end{aligned} \tag{2.24}$$

The obvious consequence of $k^- > 0$ is that, even in absence of stimulation, p_{occ} is never equal to 1. Instead, the probability of finding the release-site occupied at any time t is given by:

$$p_{occ}(t) = \frac{k^+(t)}{k^+(t) + k^-}. \tag{2.25}$$

Note that this scheme can provide for a facilitation mechanism that goes beyond an increase in p_{rel}. If k^+ is increased during activation of the synapse, p_{occ} increases, leading

[2]Note that our definition of the vacant state includes release-sites that exhibit a docked, but unprimed vesicle.

on average to an augmentation of the number of release-sites that can be activated. The time-constant associated with the occupation probability, $p_{occ}(0)$, is given by:

$$\tau_0(t) = \frac{1}{k^+(t) + k^-}. \tag{2.26}$$

Other than in the models we have seen so far, the time-constant $\tau_0(t)$ is not exclusively linked to depression but represents the typical recovery-time of the time-dependant occupation probability.

Two models that are among the first to adopt above scheme can be found in Heinemann et al. (1993) and Weis et al. (1999). While the former work was concerned with non-synaptic release (hormone-secretion from chromaffin cells), the latter was used to describe both transient and steady-state release at the calyx of Held synapse. We will thus concentrate on Weis' model. It's transition-parameters can be written as:

$$\begin{aligned} k^- &= const \\ k^+(t) &= k_0 \cdot \frac{Ca(t)}{Ca_0} \\ p_{rel} &= const, \end{aligned} \tag{2.27}$$

with the standard calcium-dynamics:

$$\dot{Ca}(t) = \frac{Ca_0 - Ca(t)}{\tau_C} + \Delta_C \sum_k \delta(t - t_k). \tag{2.28}$$

As mentioned above, since $Ca(t)$ and thus k^+ can in principle increase without bounds, facilitation can be induced, despite $p_{rel} = const$, by an increase of p_{occ}, which is, being a probability, upped bounded by unity.

In the same paper, the authors present an alternative interpretation of scheme 2.28, where the number of release-sites N is not fixed ab initio. At this, the two-states shown in figure 2.1 represent two consecutive states of vesicle maturation rather than the occupancy states of release sites. N is then the number of readily releasable vesicles. This 'vesicle-state' model allows, in principle, for an unlimited growth in N, which seems incompatible with our premise, namely that all relevant dynamics can be captured by models that use release-sites as the basic entities. However, as already pointed out in Weis et al. (1999), the release-site and vesicle-state models become mathematically equivalent if $N \cdot p_{occ} \ll N$. In this light, Weis' vesicle-state model is simply a release-site model with $p_{occ}(t) \ll 1$ for all times t.

Structurally similar models have been developed in Worden et al. (1997) and Bykhovskaia et al. (2000) where they are used to address facilitation of the lobster neuromuscular junction in response to mono-frequency stimuli. In contrast to Weis et al. (1999) they feature, however, a constant refilling rate k^+. In order to include facilitation, an additional refilling process is introduced, which comprises the occupation of a fixed number of release sites upon spike, a mechanism which we will revisit below. This process successfully explains experimental findings. Unfortunately, these works do not investigate transient, but only steady-state responses.

An example of release-site heterogeneity

The models considered so far assumed the properties of all release-sites to be identical. Heterogeneity of release-site properties is, however, a common phenomenon (see e.g. Dobrunz and Stevens (1997); Murthy et al. (1997); Branco and Staras (2009), but also Koester and Johnston (2005)). To explore the effects of heterogeneity in the simplest possible way, Trommershäuser et al. (2003) introduced two different types of release sites: one type forms a readily-releasable, the other a reluctantly-releasable pool[3]. In addition, this model includes activity-dependant recovery from depletion and facilitation that emerges from an interaction of local calcium-domains and calcium-buffers. In summary, the complexity of Trommershäuser's model sets it at the boundary between phenomenological and biophysically realistic modelling and it is the most complex one we will consider here.
Expressed in our scheme, we obtain for the transition parameters:

$$\begin{aligned} k_1^- &= const \\ k_1^+(t) &= const \\ p_{rel,1} &= \frac{1}{1+(\frac{K_F}{Ca_1(t)})^4} \end{aligned} \tag{2.29}$$

$$\begin{aligned} k_2^- &= const \\ k_2^+(t) &= k_0 + k_s \cdot \frac{Ca_{gl}(t)}{Ca_{gl,0}} \\ p_{rel,2} &= \frac{1}{1+(\frac{K_F}{Ca_2(t)})^4}. \end{aligned} \tag{2.30}$$

where the increase of k_2^+ depends on the global intracellular calcium-concentration, which is modelled as usual (see equation 2.28). The variables $Ca_1(t)$ and $Ca_2(t)$ reflect different local calcium-concentrations prevailing at the two types of sites, stemming from differences in distance to calcium-channels. Release-sites associated with pool 2 are thought to be closer to channels than release-sites from pool 1. $Ca_1(t)$ and $Ca_2(t)$ are given by:

$$Ca_i(t) = Ca_{gl}(t) + J(Ca_{out}) \cdot \alpha \cdot [\delta_{i,2} + \eta \cdot \{1 + \gamma \cdot (Ca_{gl}(t) - Ca_{gl,0})\}] \tag{2.31}$$

At this, $J(Ca_{out})$ is a constant that depends on the extracellular calcium-concentration and α, η and γ are constant parameters. While α is a measure of the distance between release-sites and calcium-channels, γ determines the strength that buffer-saturation has on facilitation (i.e. no facilitation when $\gamma = 0$). η sets, together with the Kronecker delta $\delta_{i,2}$, the exact difference in release probability between the two types of release-sites; in general, we have $p_{rel,2} > p_{rel,1}$, as intended. Figure 2.4 shows the dynamics of both

[3]Originally, only the dynamics of the readily-releasable pool are described by a release-site model, while those of the reluctantly-releasable pool are described by a vesicle-state model in the sense of Weis et al. (1999). However, by virtue of the argument given above, we can apply the release-site formalism to both pool types.

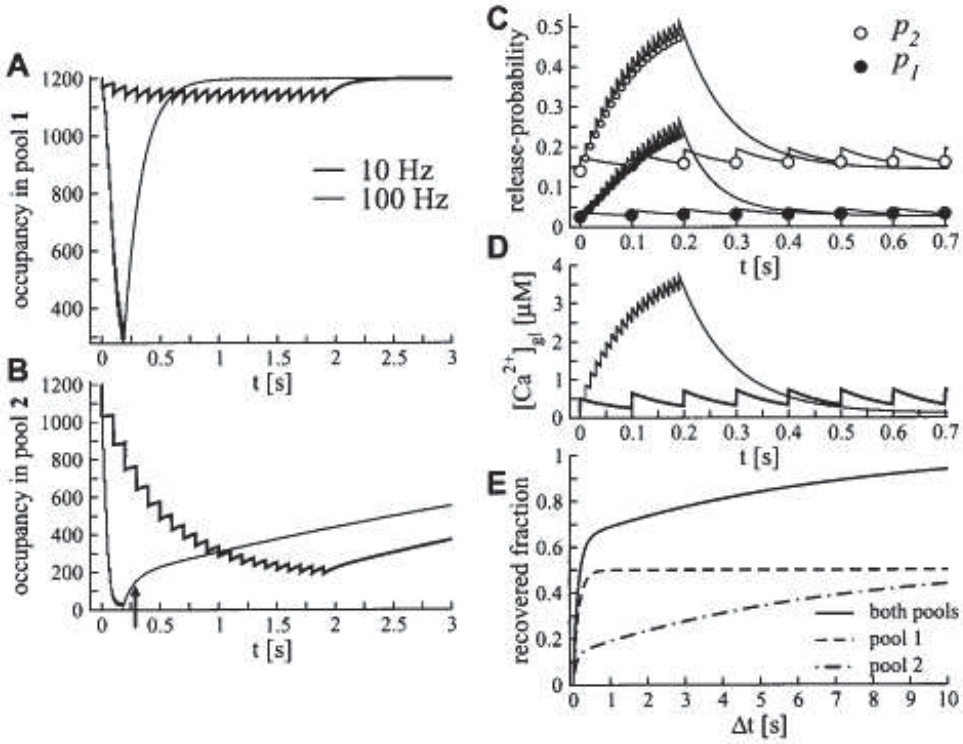

Figure 2.4: **Behaviour of Trommershäuser's model.** Model predictions Pool dynamics during and after repetitive stimulation with 10 Hz (thick black line) and 100 Hz(black line). **A)**: Occupancy in pool 1. **B)**: Occupancy in pool2. The arrow indicates the calcium-dependent recovery at the end of the stimulus train. **C)**: Facilitation of the release probabilities $p_{rel,1}$ (open circles) and $p_{rel,2}$ (solid circles). **D)**: Elevation in the global calcium concentration during repetitive stimulation. **E)**: Recovery after repetitive stimulation with 200 Hz (50 stimuli). Figure and caption adapted from Trommershäuser et al. (2003).

vesicle pools and their corresponding release probabilities for two different stimulation frequencies.

This model has been designed for the calyx of Held synapse, that has been extensively studied over decades. The detailed understanding of this synapse allows to fix a large number of the model's parameters. As a consequence, the model is biophysically realistic, but seems highly specialised to the calyx. Nevertheless, Trommershäuser's model has been used, with different degrees of modifications, to study, for example, the Drosophila neuromuscular junction with some success (Hallermann et al. (2010b,a); Weyhersmüller et al. (2011)). It is less clear, however, whether findings can be readily extended to small central synapses, where that have not been physiologically characterised as exhaustively as large synapses.

2.2 A stochastic framework of synaptic function

All models discussed so far have a crucial feature in common: they describe average responses. Suppose that a pre-synaptic spike-train consists of M spikes. Let us denote the post-synaptic response amplitude to the ith spike by R_i and a sequence of M responses by $R_{1\to M} \equiv \{R_1, \ldots, R_i, \ldots, R_M\}$. Then these deterministic models follow the pattern:

$$\text{pre-synaptic input} \longrightarrow \text{deterministic model} \longrightarrow \text{average: } \bar{R}_{1\to M}, \tag{2.32}$$

where the bar symbolises the average over trials. Dynamical changes of p_{rel}, for instance, are merely expressed as effects on the mean-response, while these models make no statements, among other things, about the trial-to-trial variability of individual responses or correlations between different responses.
In this section we want to show that, based on the release-site formalism, it is possible to construct probabilistic models of synaptic transmission. Since, as we have shown, virtually all important models of STP can be expressed in the release-site scheme, they can all be endowed with full stochasticity. Stochastic models obey:

$$\text{pre-synaptic input} \longrightarrow \text{stochastic model} \longrightarrow \text{distribution: } P(R_{1\to M}). \tag{2.33}$$

Here, $P(R_{1\to M})$ is the probability distribution of response sequences. This class of models is able to describe synaptic transmission beyond average responses, as it captures higher momenta and correlation structures of the response sequences.
We proceed by presenting the construction of stochastic models.

2.2.1 The synaptic state

We have seen that we can understand a wide number of pre-synaptic STP-mechanisms and even release-site heterogeneity by considering their effects on a single release-site. It is now straightforward to obtain a model of a full synapse by grouping N identical single sites together[4].
To see this more clearly, let us define the synaptic state S. S corresponds to the number of occupied release-sites, i.e. it is a natural number between 0 and N:

$$S = \{\#\text{ of sites with } \xi = 1\}. \tag{2.34}$$

At this, we do not keep track of the release-sites' identities; that is, all instances with 4 docked vesicles, for example, are equivalent, no matter at which sites they are docked. We can think of the synaptic state as representing the ready-releasable pool (RRP) of the synapse.
Now consider the situation where a spike train activates the synapse at times $t_{1\to M} \equiv \{t_1, \cdots, t_M\}$. Clearly, the spike train will drive the release-sites through a sequence of those transitions we have discussed at length in the previous sections and S will change

[4] If we wish to introduce heterogeneity, we simply take N_1 sites of type one, N_2 sites of type two etc.

correspondingly. Let us denote by S_k^- and S_k^+ the synaptic state S immediately before and after spike i:

$$\begin{aligned} S_i^- &= \{\#\text{ of sites with } \xi = 1 \text{ immediately before } t_i\} \\ S_i^+ &= \{\#\text{ of sites with } \xi = 1 \text{ immediately after } t_i\}. \end{aligned} \tag{2.35}$$

Since all transitions are probabilistic we can now formulate probabilities for transitions between synaptic states. At this we have to distinguish, as we have done before, between two different types: transitions upon spike and transitions in between spikes. We can write:

$$\begin{aligned} P(S_i^+ \longleftarrow S_i^-) &= P(S_i^+|S_i^-) \quad \text{probability upon spike} \\ P(S_{i+1}^- \longleftarrow S_i^+) &= P(S_{i+1}^-|S_i^+) \quad \text{probability in between spikes.} \end{aligned} \tag{2.36}$$

These probabilities depend on the one hand on the properties of the single release-sites, that is on $p_{rel}(t)$, $k^+(t)$ and k^-. On the other hand, the size of the synapse, which is correlated with the number of release sites, N, (Schikorski and Stevens (1997)), plays an important role. Furthermore, the transitions depend also in the ith interspike-interval Δt_i. We will proceed by giving concrete examples of the expression in equation 2.36, starting with in between spike transitions.

2.2.2 Transitions in between spikes

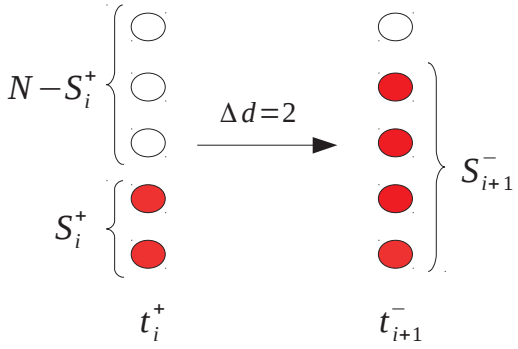

Figure 2.5

Let us first consider the case with $k^- = 0$ and $k^+ = const$. The probability that a single empty release-site is occupied during an interspike-interval Δt_i is given by:

$$l_i = 1 - e^{-\frac{\Delta t_i}{\tau_D}}, \tag{2.37}$$

where we used $k^+ = \frac{1}{\tau_D}$. If Δt_i is much longer than τ_D, this probability approaches unity. Figure 2.5 shows a graphical representation of an exemplary transition from S_i^+ to S_{i+1}^-.

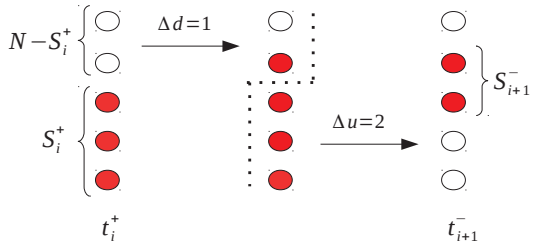

Figure 2.6

The recovery-probability of $S_{i+1}^- - S_i^+$ empty sites is given by a binomial distribution:

$$P(S_{i+1}^-|S_i^+) = \binom{N - S_i^+}{S_{i+1}^- - S_i^+} l_i^{(S_{i+1}^- - S_i^+)} (1 - l_i)^{(N - S_{i+1}^-)}. \tag{2.38}$$

This relation holds true for $S_{i+1}^- \geq S_i^+$, that is, if the number of docked sites increases or stays the same. Otherwise, we set $P(S_{i+1}^-|S_i^+) = 0$, since transitions with $k^- = 0$ can never reduce the number of docked sites.
The case with $k^- = 0$ and $k^+ = k^+(t)$ is very similar, with the difference that $k^+(t)$ has to be integrated over the interval Δt_i. Considering Dittman's model, for instance, this leads to the single-site recovery probability (see Dittman et al. (2000)):

$$\begin{aligned} l_i &= 1 - e^{-\zeta(\Delta t_i)} \\ \zeta(\Delta t_i) &= k_{max} \cdot \Delta t_i + \tau_{XD} \cdot (k_{max} - k_0) \cdot \ln \left\{ \frac{K_D + CaX_D(t_i)}{K_D \cdot e^{\frac{\Delta t_i}{\tau_{XD}}} + CaX_D(t_i)} \right\}, \end{aligned} \tag{2.39}$$

where $CaX_D(t_i)$ is a deterministic function of the input spike times, which can easily be evaluated at t_i. Equation 2.38 remains unchanged in this case.

On the contrary, allowing for $k^- > 0$ changes the situation quite a bit, as both occupied release-sites can become vacant and vacant release-sites can be docked by a vesicle during the inter-spike interval. Recall from equations 2.25 and 2.26 the definitions of the equilibrium probability of occupancy p_{occ} and the time constant τ_0 of the decay to p_{occ}. With these, we can write down the probability that a single vacant site goes into the occupied state p_+, and the probability that a single occupied site goes into the vacant

state p_-[5]:

$$\begin{aligned} p_{+,i} &= p_{occ} \cdot \left(1 - e^{-\frac{\Delta t_i}{\tau_0}}\right) \\ p_{-,i} &= (1 - p_{occ}) \cdot \left(1 - e^{-\frac{\Delta t_i}{\tau_0}}\right), \end{aligned} \tag{2.40}$$

where the index i flags the dependence on the ith inter-spike interval. As $\Delta t_i \longrightarrow \infty$ the transition probabilities converge to $p_+ = p_{occ}$ and $p_- = (1 - p_{occ})$, yielding the correct steady-state values. For $k^- = 0$, we have $p_{occ} = 1$ and expression 2.40 reverts to equation 2.37.

The transition probability from one synaptic state to the next, $P(S^-_{i+1}|S^+_i)$, has now to be decomposed into two parts, as show in figure 2.6. First, we can state the probability of recovering Δd release sites. Second, given Δd, the probability of emptying Δu release sites can be given, hereby considering only those sites which have not been updated yet. Finally, we have to sum over all possible values of Δd. In formal terms, this yields:

$$\begin{aligned} P(S^-_{i+1}|S^+_i) &= \sum_{\Delta d=\Delta d_{min}}^{\Delta d_{max}} P(S^-_{i+1}|S^+_i + \Delta d) \cdot P(S^+_i + \Delta d|S^+_i) \\ &= \sum_{\Delta d=\Delta d_{min}}^{\Delta d_{max}} \binom{N - S^+_i}{\Delta d} \cdot p_{+,i}^{\Delta d} \cdot (1 - p_{+,i})^{(N - S^+_i - \Delta d)} \times \\ &\qquad \times \binom{S^+_i}{\Delta u} \cdot p_{-,i}^{\Delta u} \cdot (1 - p_{-,i})^{(S^+_i - \Delta u)}. \end{aligned} \tag{2.41}$$

where $\Delta u = S^+_i + \Delta d - S^-_{i+1}$. The sum over Δd starts at $\Delta d_{min} = max\{S^-_{i+1} - S^+_i, 0\}$ and ends at $\Delta d_{max} = min\{S^-_{i+1}, N - S^+_i\}$. These limits guarantee that two obvious constraints are met: no more release-sites can be occupied than are empty and no more release-sites can be emptied than are occupied.

2.2.3 Transitions upon spike

Transitions upon spike are associated with the release process. We have to distinguish two scenarios. First, especially at hippocampal synapses, findings suggest that only one single vesicle can be released per active zone (Stevens and Wang (1995); Hanse and Gustafsson (2001a,b)), despite the presence of more than one release-site at the active zone. The probability of releasing a vesicle, however, does not simply correspond to the single site's p_{rel}, but seems to increase with the number of occupied release-sites (Dobrunz and Stevens (1997)). Following Kandaswamy et al. (2010), we can express this cooperation between release-sites by[6]:

$$P(S^+_i|S^-_i) = \begin{cases} 1 - (1 - p_{rel,i})^{S^-_i} & \text{if} \quad S^+_i = S^-_i - 1 \\ 0 & \text{else,} \end{cases} \tag{2.42}$$

[5] Here, we assume for simplicity $k_+ = const$.

[6] An alternative, but roughly equivalent formulation can be obtained from Wang (1999): $P(S^+_i|S^-_i) = 1 - \exp(-p_{rel,i} \cdot S^-_i)$.

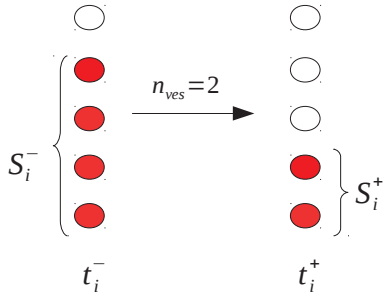

Figure 2.7

Here, $p_{rel,i}$ is a short-hand notation for $p_{rel}(t_i)$, that is, for the release-probability upon the ith spike. The condition in equation 2.42 ensures that only release of a single vesicle can occur.
In the second scenario, release sites act independently from each other and release of multiple vesicles at once is possible (Auger et al. (1998); Oertner et al. (2002); Loebel et al. (2009); Bender et al. (2009); Huang et al. (2010)). Figure 2.7 shows the transition between two synaptic states in this case. The number of released vesicles upon the ith spike is simply given by:

$$n_{ves}^i = S_i^- - S_i^+. \tag{2.43}$$

With this, the transition probabilities can be straightforwardly written as:

$$P(S_i^+|S_i^-) = \binom{S_i^-}{n_{ves}^i} \cdot p_{rel,i}^{n_{ves}^i} \cdot (1 - p_{rel,i})^{S_i^- - n_{ves}^i}, \tag{2.44}$$

where, of course, $n_{ves}^i \geq 0$ and otherwise $P(S_i^+|S_i^-) = 0$.
Note that in the second case described here - the multiple vesicle release model - sites are independent from each other (as far as the release process is concerned). This is in contrast to the single vesicle release model, where different sites exhibit a cooperative effect.

A fast refilling process upon spike

Worden et al. (1997) and Bykhovskaia et al. (2000) proposed a model where a fixed number of vesicles is 'mobilised' upon spike, which amounts to the occupation of a fixed number of release sites. In our framework we can address this process in a more general manner. Instead of deterministically increasing the number of occupied sites, we can introduce an additional transition probability, as shown in figure 2.8A. Upon spike then, vacant sites can be docked by a vesicle with a probability p_{fast}, which may or may not depend on the activation history. Concerning the sequence of the processes involved,

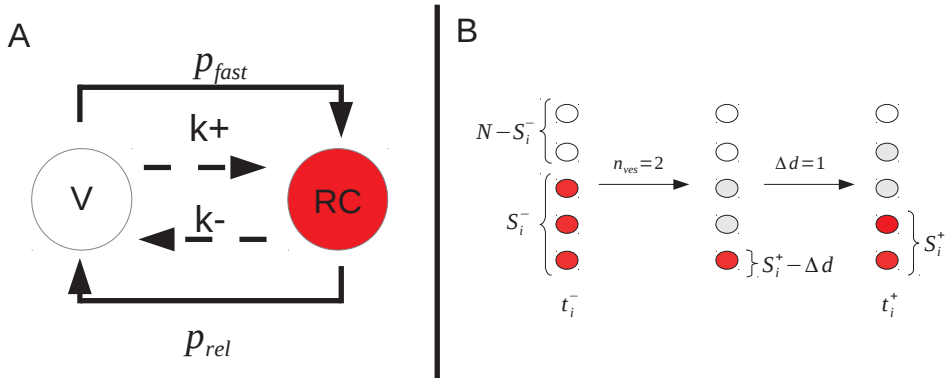

Figure 2.8

different scenarios are conceivable. For instance, fast docking could occur before release, or vice versa. Let us assume here that release occurs first, and that fast docking can only update those release-sites which were vacant *before* release. This process is sketched in figure 2.8B. Denoting the number of sites that become docked by Δd, we obtain the following state transition probabilities:

$$\begin{aligned} P(S_i^+|S_i^-) &= \sum_{\Delta d=0}^{N-S_i^-} P(S_i^+|S_i^+ - \Delta d) \cdot P(S_i^+ - \Delta d|S_i^-) \\ &= \sum_{\Delta d=0}^{N-S_i^-} \binom{N-S_i^-}{\Delta d} \cdot p_{fast,i}^{\Delta d} \cdot (1-p_{fast,i})^{N-S_i^- - \Delta d} \times \\ &\qquad \times \binom{S_i^-}{n_{ves}^i} \cdot p_{rel,i}^{n_{ves}^i} \cdot (1-p_{rel,i})^{S_i^- - n_{ves}^i}, \end{aligned} \tag{2.45}$$

where, in contrast to equation 2.44, we have $n_{ves}^i = S_i^- - S_i^+ + \Delta d$. Equation 2.45 holds if $S_i^+ \leq N - n_{ves}^i$, otherwise $P(S_i^+|S_i^-) = 0$, which reflects the fact that only sites vacant before release can be refilled.

2.2.4 Quantal response function

If we want to obtain $P(R_{1\to M})$, we need to connect the synaptic states sequences with the observable responses R_i. As we have seen in the previous section, the model provides us with the probability that one or more vesicles are released upon a given spike. In order to convert a certain number of released vesicles n_{ves}^i into R_i, we adopt the assumptions of the quantal model in its simplest form. According to the quantal model, each of the n_{ves} vesicles produces on average a response q (called the quantal size or the unitary quantal response). Disregarding saturation and post-synaptic receptor desensitisation processes, the effects of all released vesicles sum linearly. The unitary quantal response, however, exhibits some variability, which is quantified through its standard deviation,

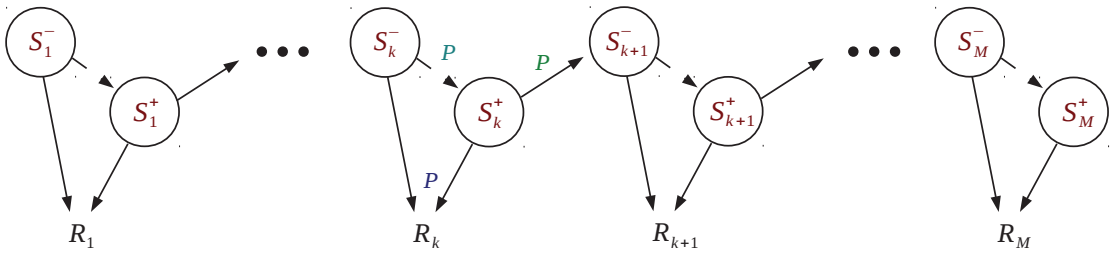

Figure 2.9

σ_q, called the quantal variability. The probability of measuring R_i given the release of n_{ves}^i vesicles can thus be described by a probability distribution with mean $n_{ves}^i \cdot q$ and variance $n_{ves}^i \cdot \sigma_q^2$.
What is the form of this distribution? Measurements of synaptic miniature events suggest that the distribution of the quantal response is not simply a Gaussian, but skewed to the right (Bekkers et al. (1990); Bhumbra and Beato (2013)). We can, for example, model this distribution with an Inverse Gaussian:

$$P(R_i|n_{ves}^i) = \frac{q^{\frac{3}{2}} \cdot n_{ves}^i}{\sqrt{2\pi\sigma_q^2 R_i^3}} \exp\left\{-\frac{q \cdot \left[R_i - q \cdot n_{ves}^i\right]^2}{2\sigma_q^2 R_i}\right\} \tag{2.46}$$

We are not limited to describe $P(R_i|n_{ves}^i)$ in this form, but can choose from a wide range of possibilities (see e.g. Stricker and Redman (1994)).

2.2.5 The total probability distribution

With the components at hand to describe the synaptic state sequence and its relation to the observable quantities we can state $P(R_{1\to M})$. The probability of observing a single sequence of responses is equal to the sum of probabilities of all possible synaptic state sequences that can contribute. The abstract scheme in figure 2.9 summarises the composition of $P(R_{1\to M})$ for the standard case, i.e. when $n_{ves}^i = S_i^- - S_i^+$. Formally, we obtain:

$$\begin{aligned} P(R_{1\to M}) &= \sum_{\text{all } S^+,S^-} P(R_{1\to M}, S_{1\to M}^-, S_{1\to M}^+) \\ &= \sum_{\text{all } S^+,S^-} P(S_1^-)\prod_{i=1}^{M} P(S_i^+|S_i^-)P(R_i|n_{ves}^i)\prod_{i=1}^{M-1} P(S_{i+1}^-|S_i^+). \end{aligned} \tag{2.47}$$

The only thing left to specify is the initial state distribution, $P(S_1^-)$. For all models with $k^- = 0$, we have $P(S_1^- = N) = 1$ and $P(S_1^-) = 0$ for all other values of S_1^-. If $k^- > 0$, $P(S_1^-)$ is simply given by a binomial distribution with probability p_{occ}.

Equation 2.47 yields a full probabilistic description of synaptic transmission, from which all momenta of the responses can be calculated.

2.2.6 Possible extensions

So far, we have disregarded a few potentially important mechanisms of synaptic transmission, especially post-synaptic processes. Here, we wish to shortly suggest how those effects could be implemented in our framework. We will do this in broad terms and without presenting complete elaborations.

Saturation

At certain synapses, receptor saturation has a significant effect on the post-synaptic responses (Tang et al. (1994); Auger et al. (1998); Foster et al. (2002)). In order to allow for saturation, we have to abandon the linear summation assumption of vesicle effects described in section 2.2.4. With regard to the response probability distribution, $P(R_i|n^i_{ves})$, it is thus no longer possible to increase its mean and variance linearly with n_{ves} . Following Auger et al. (1998) and Matveev and Wang (2000b), we can instead choose a more complicated dependency between these quantities. Let us denote by w the fraction of post-synaptic receptors that are activated by a single vesicle and by R_{max} the maximal response the synapse is capable of. Then the average response μ_R in function of the number of release vesicles is:

$$\mu_R(n^i_{ves}) = R_{max} \cdot \left(1 - (1-w)^{n^i_{ves}}\right). \tag{2.48}$$

In this setting, the quantal amplitude is given by the average response to a single vesicle, i.e. $q = R_{max} \cdot w$. We may choose a similar dependence on n_{ves} for the response variability, where analogously the variability of a single vesicle response would be given by $\sigma_q^2 = \mathrm{Var}_{max} \cdot w$, Var_{max} denoting the maximal response variability. This would account for the lowering of responses' coefficients of variation with increasing n_{ves}, a phenomenon presumed by Franks et al. (2003). We want to clearly stress, however, that the answer to the question how response variability changes under the influence of increasing saturation is less straightforward, and our proposal here is somewhat *ad-hoc*. Nonetheless, this example should make it clear that once relationships between n_{ves} and response average and variance are established, they can easily be included in the probabilistic framework.

Desensitisation

Receptor desensitisation is another post-synaptic phenomenon that reduces the efficacy of released neurotransmitter. While saturation causes sub-linear summation of single vesicle effects, desensitisation temporarily inactivates receptors in function of the amount of neurotransmitter released. A description of this process requires q to be treated not as a constant, but a as function of release-history. The simplest way to include desensitisation in our framework is to treat it deterministically. This would come down to decreasing

q upon spike by an amount that is proportional to the *average* number of vesicles released at that time, with subsequent recovery (Brenowitz and Trussell (2001); Hennig et al. (2008)). Resorting to a deterministic description, however, may appear somewhat inconsistent with our probabilistic approach.
Alternatively, we can follow Yang and Xu-Friedman (2008) and consider the desensitisation-dynamics to be caused by the presence of neurotransmitter in the synaptic cleft that is gradually cleared. In this case, the synaptic state can be extended to include an additional state variable that keeps track of extracellular neurotransmitter concentration, this latter determining the values of q and σ_q. At any point in time then, the synaptic state contains information about the number of docked sites and the updated values of the quantal amplitude and quantal variance, from which $P(R_i|n^i_{ves})$ can be calculated. Admittedly, adopting this scheme comes at the cost: the state-space is larger, which renders all applications of the model computationally more expensive. We can conclude, however, that there is no conceptual barrier for embedding stochastic desensitisation in our framework.

Including more vesicle pools

Synaptic responses to long stimulus trains (on the timescales of minutes) may require the modelling of additional vesicles pools (Wu and Betz (1998); Hallermann et al. (2010b); Kandaswamy et al. (2010)). At the bottom of this requirement is the observation that, after prolonged stimulation, the recovery from depression cannot be described by a single exponential time-course. In these cases, depletion of a single RRP is not a sufficient explanation for depression, and the dynamics of a second pool have to be included. The common view is that beside the RRP two other pools can be identified in most synapses: a recycling pool and a reserve pool (Zucker and Regehr (2002); Rizzoli and Betz (2005)). The recycling pool is in fast equilibrium with the RRP and is slowly refilled by the reserve pool. If the recycling pool is partially depleted after prolonged stimulation it can to some extend refill the RRP. Full replenishment, however, will be dictated by the slow processes which mediate recovery of the recycling pool.
One way to include recycling pool contributions in our framework is a deterministic treatment. We can introduce an additional variable which keeps track of the recycling pool's fractional filling level. The refilling rate of the RRP, that is, vesicle binding rate to the release site, $k_+(t)$, can then be considered a function this variable, so that $k_+(t)$ is zero when the recycling pool is empty and maximal when the recycling pool completely refilled. In other words, $k_+(t)$ is endowed with an activity-dependant *decreasing* component.
Certainly, a full stochastic description of the recycling pool can be devised. This will comprise the inclusion of a recycling pool state in the synaptic state S and the definition of corresponding transitions. However, we expect that the impact of the stochastic transitions between recycling pool and RRP on the observable response variability will be quite insulated. In view of the implementation costs, the discussed deterministic pool dynamics may turn out to be sufficient.

2.3 Conclusion

We have considered two things in this chapter. First, we have reviewed phenomenological models of STP that range from simple to moderately complex. We showed that the description with a simple release-site scheme is sufficient to accommodate a wide range of mechanisms and, by virtue of this very fact, a wide range of different synapse types. The release-site view is useful to structure our view on the rich diversity of models as it helps to trace descriptive similarities and differences. Second, we showed that thanks to the common release-site structure, all phenomenological models of STP can be transformed into a fully stochastic version and have demonstrate the basic recipe to perform this step. The probabilistic modelling framework permits us to go beyond average response perspective and ask: "Given a model, what is the probability to observe a certain response sequence?"
We stress that the assumptions underlying any stochastic model devised within our framework do not go beyond those of the quantal model and the underlying deterministic STP model. In general, the responses of model synapses with large N show less fluctuations; synapses that are built of independent release sites are indeed self-averaging for increasing N[7]. In these cases a deterministic description seems fit. Most central synapses, however, are small and response fluctuations are significant. We propose that probabilistic modelling is much more appropriate in these cases. The next chapter will be dedicated to exactly this topic. There we will show the superior performance of this type of modelling over standard average-response models when confronted with real-data applications.

[7] In the case described by equation 2.42 the release-sites are not independent as they exhibit a cooperative effect on the overall release probability. Nevertheless, for large N this release probability is close to unity, causing the responses of this model to be deterministic.

Bibliography

Auger, C., Kondo, S., and Marty, A. (1998). Multivesicular release at single functional synaptic sites in cerebellar stellate and basket cells. *The Journal of neuroscience*, 18(12):4532–4547.

Bekkers, J., Richerson, G., and Stevens, C. (1990). Origin of variability in quantal size in cultured hippocampal neurons and hippocampal slices. *Proceedings of the National Academy of Sciences*, 87(14):5359–5362.

Bender, V. A., Pugh, J. R., and Jahr, C. E. (2009). Presynaptically expressed long-term potentiation increases multivesicular release at parallel fiber synapses. *The Journal of Neuroscience*, 29(35):10974–10978.

Bertram, R., Sherman, A., and Stanley, E. F. (1996). Single-domain/bound calcium hypothesis of transmitter release and facilitation. *Journal of Neurophysiology*, 75(5).

Betz, W. (1970). Depression of transmitter release at the neuromuscular junction of the frog. *The Journal of Physiology*, 206(3):629.

Bhumbra, G. S. and Beato, M. (2013). Reliable evaluation of the quantal determinants of synaptic efficacy using bayesian analysis. *Journal of neurophysiology*, 109(2):603–620.

Billups, B., Graham, B. P., Wong, A. Y., and Forsythe, I. D. (2005). Unmasking group iii metabotropic glutamate autoreceptor function at excitatory synapses in the rat cns. *The Journal of physiology*, 565(3):885–896.

Branco, T. and Staras, K. (2009). The probability of neurotransmitter release: variability and feedback control at single synapses. *Nature Reviews Neuroscience*, 10(5):373–383.

Brenowitz, S. and Trussell, L. O. (2001). Minimizing synaptic depression by control of release probability. *The Journal of Neuroscience*, 21(6):1857–1867.

Bykhovskaia, M., Worden, M. K., and Hackett, J. T. (2000). Stochastic modeling of facilitated neurosecretion. *Journal of computational neuroscience*, 8(2):113–126.

Cortes, J. M., Desroches, M., Rodrigues, S., Veltz, R., Muñoz, M. A., and Sejnowski, T. J. (2013). Short-term synaptic plasticity in the deterministic tsodyks–markram model leads to unpredictable network dynamics. *Proceedings of the National Academy of Sciences*, 110(41):16610–16615.

Costa, R. P., Sjöström, P. J., and van Rossum, M. C. (2013). Probabilistic inference of synaptic dynamics in neocortical microcircuits. *BMC Neuroscience*, 14(Suppl 1):P403.

Del Castillo, J. and Katz, B. (1954). Quantal components of the end-plate potential. *The Journal of physiology*, 124(3):560–573.

Dittman, J. S., Kreitzer, A. C., and Regehr, W. G. (2000). Interplay between facilitation, depression, and residual calcium at three presynaptic terminals. *The Journal of Neuroscience*, 20(4):1374–1385.

Dittman, J. S. and Regehr, W. G. (1998). Calcium dependence and recovery kinetics of presynaptic depression at the climbing fiber to purkinje cell synapse. *The Journal of neuroscience*, 18(16):6147–6162.

Dobrunz, L. E. and Stevens, C. F. (1997). Heterogeneity of release probability, facilitation, and depletion at central synapses. *Neuron*, 18(6):995–1008.

Dodge, F. and Rahamimoff, R. (1967). Co-operative action of calcium ions in transmitter release at the neuromuscular junction. *The Journal of physiology*, 193(2):419–432.

Forsythe, I. D., Tsujimoto, T., Barnes-Davies, M., Cuttle, M. F., and Takahashi, T. (1998). Inactivation of presynaptic calcium current contributes to synaptic depression at a fast central synapse. *Neuron*, 20(4):797–807.

Foster, K. A., Kreitzer, A. C., and Regehr, W. G. (2002). Interaction of postsynaptic receptor saturation with presynaptic mechanisms produces a reliable synapse. *Neuron*, 36(6):1115–1126.

Franks, K. M., Stevens, C. F., and Sejnowski, T. J. (2003). Independent sources of quantal variability at single glutamatergic synapses. *The Journal of neuroscience*, 23(8):3186–3195.

Fuhrmann, G., Cowan, A., Segev, I., Tsodyks, M., and Stricker, C. (2004). Multiple mechanisms govern the dynamics of depression at neocortical synapses of young rats. *The Journal of physiology*, 557(2):415–438.

Fuhrmann, G., Segev, I., Markram, H., and Tsodyks, M. (2002). Coding of temporal information by activity-dependent synapses. *Journal of neurophysiology*, 87(1):140–148.

Hallermann, S., Fejtova, A., Schmidt, H., Weyhersmüller, A., Silver, R. A., Gundelfinger, E. D., and Eilers, J. (2010a). Bassoon speeds vesicle reloading at a central excitatory synapse. *Neuron*, 68(4):710–723.

Hallermann, S., Heckmann, M., and Kittel, R. J. (2010b). Mechanisms of short-term plasticity at neuromuscular active zones of drosophila. *HFSP journal*, 4(2):72–84.

Hanse, E. and Gustafsson, B. (2001a). Paired-pulse plasticity at the single release site level: an experimental and computational study. *The Journal of Neuroscience*, 21(21):8362–8369.

Hanse, E. and Gustafsson, B. (2001b). Quantal variability at glutamatergic synapses in area ca1 of the rat neonatal hippocampus. *The Journal of physiology*, 531(2):467–480.

Hansel, D. and Mato, G. (2013). Short-term plasticity explains irregular persistent activity in working memory tasks. *The Journal of Neuroscience*, 33(1):133–149.

Heinemann, C., von Rüden, L., Chow, R. H., and Neher, E. (1993). A two-step model of secretion control in neuroendocrine cells. *Pflügers Archiv*, 424(2):105–112.

Hempel, C. M., Hartman, K. H., Wang, X.-J., Turrigiano, G. G., and Nelson, S. B. (2000). Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex. *Journal of neurophysiology*, 83(5):3031–3041.

Hennig, M. H. (2013). Theoretical models of synaptic short term plasticity. *Frontiers in computational neuroscience*, 7.

Hennig, M. H., Postlethwaite, M., Forsythe, I. D., and Graham, B. P. (2008). Interactions between multiple sources of short-term plasticity during evoked and spontaneous activity at the rat calyx of held. *The Journal of physiology*, 586(13):3129–3146.

Huang, C.-H., Bao, J., and Sakaba, T. (2010). Multivesicular release differentiates the reliability of synaptic transmission between the visual cortex and the somatosensory cortex. *The Journal of Neuroscience*, 30(36):11994–12004.

Kandaswamy, U., Deng, P.-Y., Stevens, C. F., and Klyachko, V. A. (2010). The role of presynaptic dynamics in processing of natural spike trains in hippocampal synapses. *The Journal of Neuroscience*, 30(47):15904–15914.

Koester, H. J. and Johnston, D. (2005). Target cell-dependent normalization of transmitter release at neocortical synapses. *Science*, 308(5723):863–866.

Lee, C.-C. J., Anton, M., Poon, C.-S., and McRae, G. J. (2009). A kinetic model unifying presynaptic short-term facilitation and depression. *Journal of computational neuroscience*, 26(3):459–473.

Liley, A. and North, K. (1953). An electrical investigation of effects of repetitive stimulation on mammalian neuromuscular junction. *J Neurophysiol*, 16(5):509–527.

Loebel, A., Silberberg, G., Helbig, D., Markram, H., Tsodyks, M., and Richardson, M. J. (2009). Multiquantal release underlies the distribution of synaptic efficacies in the neocortex. *Frontiers in computational neuroscience*, 3.

Markram, H., Wang, Y., and Tsodyks, M. (1998). Differential signaling via the same axon of neocortical pyramidal neurons. *Proceedings of the National Academy of Sciences*, 95(9):5323–5328.

Matveev, V. and Wang, X.-J. (2000a). Differential short-term synaptic plasticity and transmission of complex spike trains: to depress or to facilitate? *Cerebral Cortex*, 10(11):1143–1153.

Matveev, V. and Wang, X.-J. (2000b). Implications of all-or-none synaptic transmission and short-term depression beyond vesicle depletion: a computational study. *The Journal of Neuroscience*, 20(4):1575–1588.

Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, 319(5869):1543–1546.

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

Murthy, V. N., Sejnowski, T. J., and Stevens, C. F. (1997). Heterogeneous release properties of visualized individual hippocampal synapses. *Neuron*, 18(4):599–612.

Murthy, V. N. and Stevens, C. F. (1999). Reversal of synaptic vesicle docking at central synapses. *Nature neuroscience*, 2(6):503–507.

Nadkarni, S., Bartol, T. M., Stevens, C. F., Sejnowski, T. J., and Levine, H. (2012). Short-term plasticity constrains spatial organization of a hippocampal presynaptic terminal. *Proceedings of the National Academy of Sciences*, 109(36):14657–14662.

Neher, E. (2010). What is rate-limiting during sustained synaptic activity: vesicle supply or the availability of release sites. *Frontiers in synaptic neuroscience*, 2.

Neher, E. and Sakaba, T. (2008). Multiple roles of calcium ions in the regulation of neurotransmitter release. *Neuron*, 59(6):861–872.

Nofal, S., Becherer, U., Hof, D., Matti, U., and Rettig, J. (2007). Primed vesicles can be distinguished from docked vesicles by analyzing their mobility. *The Journal of neuroscience*, 27(6):1386–1395.

Oertner, T. G., Sabatini, B. L., Nimchinsky, E. A., and Svoboda, K. (2002). Facilitation at single synapses probed with optical quantal analysis. *Nature neuroscience*, 5(7):657–664.

Pan, B. and Zucker, R. S. (2009). A general model of synaptic transmission and short-term plasticity. *Neuron*, 62(4):539–554.

Pfister, J.-P., Dayan, P., and Lengyel, M. (2010). Synapses with short-term plasticity are optimal estimators of presynaptic membrane potentials. *Nature neuroscience*, 13(10):1271–1275.

Rizzoli, S. O. and Betz, W. J. (2005). Synaptic vesicle pools. *Nature Reviews Neuroscience*, 6(1):57–69.

Schikorski, T. and Stevens, C. F. (1997). Quantitative ultrastructural analysis of hippocampal excitatory synapses. *The Journal of neuroscience*, 17(15):5858–5867.

Stevens, C. F. and Wang, Y. (1995). Facilitation and depression at single central synapses. *Neuron*, 14(4):795–802.

Stricker, C. and Redman, S. (1994). Statistical models of synaptic transmission evaluated using the expectation-maximization algorithm. *Biophysical journal*, 67(2):656–670.

Tang, C.-M., Margulis, M., Shi, Q.-Y., and Fielding, A. (1994). Saturation of postsynaptic glutamate receptors after quantal release of transmitter. *Neuron*, 13(6):1385–1393.

Thomson, A., Deuchars, J., and West, D. (1993). Single axon excitatory postsynaptic potentials in neocortical interneurons exhibit pronounced paired pulse facilitation. *Neuroscience*, 54(2):347–360.

Trommershäuser, J., Schneggenburger, R., Zippelius, A., and Neher, E. (2003). Heterogeneous presynaptic release probabilities: functional relevance for short-term plasticity. *Biophysical journal*, 84(3):1563–1579.

Tsodyks, M., Pawelzik, K., and Markram, H. (1998). Neural networks with dynamic synapses. *Neural computation*, 10(4):821–835.

Tsodyks, M. V. and Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the National Academy of Sciences*, 94(2):719–723.

Varela, J. A., Sen, K., Gibson, J., Fost, J., Abbott, L., and Nelson, S. B. (1997). A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *The Journal of neuroscience*, 17(20):7926–7940.

Wang, X.-J. (1999). Fast burst firing and short-term synaptic plasticity: a model of neocortical chattering neurons. *Neuroscience*, 89(2):347–362.

Wang, Y., Markram, H., Goodman, P. H., Berger, T. K., Ma, J., and Goldman-Rakic, P. S. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nature neuroscience*, 9(4):534–542.

Weis, S., Schneggenburger, R., and Neher, E. (1999). Properties of a model of ca^{++}-dependent vesicle pool dynamics and short term synaptic depression. *Biophysical Journal*, 77(5):2418–2429.

Weyhersmüller, A., Hallermann, S., Wagner, N., and Eilers, J. (2011). Rapid active zone remodeling during synaptic plasticity. *The Journal of Neuroscience*, 31(16):6041–6052.

Worden, M. K., Bykhovskaia, M., and Hackett, J. T. (1997). Facilitation at the lobster neuromuscular junction: a stimulus-dependent mobilization model. *Journal of neurophysiology*, 78(1):417–428.

Wu, L.-G. and Betz, W. J. (1998). Kinetics of synaptic depression and vesicle recycling after tetanic stimulation of frog motor nerve terminals. *Biophysical journal*, 74(6):3003–3009.

Wu, L.-G. and Borst, J. G. G. (1999). The reduced release probability of releasable vesicles during recovery from short-term synaptic depression. *Neuron*, 23(4):821–832.

Xu, J. and Wu, L.-G. (2005). The decrease in the presynaptic calcium current is a major cause of short-term depression at a calyx-type synapse. *Neuron*, 46(4):633–645.

Yang, H. and Xu-Friedman, M. A. (2008). Relative roles of different mechanisms of depression at the mouse endbulb of held. *Journal of neurophysiology*, 99(5):2510–2521.

Zenisek, D., Steyer, J., and Almers, W. (2000). Transport, capture and exocytosis of single synaptic vesicles at active zones. *Nature*, 406(6798):849–854.

Zucker, R. S. and Regehr, W. G. (2002). Short-term synaptic plasticity. *Annual review of physiology*, 64(1):355–405.

Chapter 3

Quantifying Short-Term Plasticity and Variability at Chemical Synapses: A Generative-Model Approach

Alessandro Barri, Yun Wang, David Hansel, Gianluigi Mongillo
to be submitted to PLOS Computational Biology

3.1 Introduction

A distinctive feature of chemical transmission is the rapid, transient modification of the (post-)synaptic response as a result of the repetitive pre-synaptic activation [Zucker and Regehr (2002); Fioravante and Regehr (2011); Regehr (2012)]. Such short-term plasticity (STP) has been suggested to endow chemical synapses, and consequently neuronal circuits, with important computational capabilities [Abbott and Regehr (2004); Tsodyks and Wu (2013)]. A full understanding of 'synaptic computations' clearly requires simple, yet qualitatively and quantitatively accurate, models of STP.

Quantitative investigation of STP relies to a significant extent on phenomenological descriptions. In such descriptions, the synaptic response is modulated by different dynamical variables that describe facilitating or depressing processes taking place at the synapse. Upon spikes, these variables increase/decrease by discrete amounts while, in between spikes, they decay back to their baseline levels. A notable example is the Tsodyks-Markram (TM) model, where synaptic transmission is described in terms of two dynamical variables and four free parameters, which can be estimated from experimental data. In their providing a compact (i.e., with few free parameters), low-dimensional description of STP, phenomenological models such as the TM model have been highly instrumental in effectively classifying different STP patterns [Tsodyks and Markram (1997); Markram et al. (1998); Hempel et al. (2000); Wang et al. (2006)], in uncovering the under-

lying mechanisms of synaptic transmission [Dittman and Regehr (1998); Dittman et al. (2000); Hennig et al. (2008); Kandaswamy et al. (2010)], and in exploring theoretically the functional/computational consequences of STP [see e.g. Buonomano (2000); Goldman et al. (2002); Mongillo et al. (2008); Pfister et al. (2010); Rosenbaum et al. (2012)]. Phenomenological models, however, either only describe the *average* synaptic responses or, where the model is stochastic, it is the average model responses which are fitted to the trial-averaged experimental responses. In either case, the trial-to-trial variability and the within-trial correlation of the synaptic responses are neglected (but see Loebel et al. (2009); Scheuss and Neher (2001); Hallermann et al. (2010)).

Stochasticity of synaptic responses is, indeed, another distinctive feature of chemical transmission, which is especially apparent at central synapses. The quantitative analysis of fluctuations in the synaptic responses is based on the notion of 'quantal' release, that is, neurotransmitter is secreted in discrete units (quanta) rather than in continuous quantities. In its simplest instantiation, the quantal model describes the synapse as a collection of identical, independent release sites each of which, upon spike, can release at most one quantum in a probabilistic way. The response to the release of multiple quanta from different sites is the sum of the responses to a single quantum (the unitary quantal response). The model has three free parameters - the number of release sites, the probability of release and the unitary quantal response - which can be estimated from sufficiently long recordings of synaptic responses. Methods to estimate quantal parameters are tailored for steady-state conditions, and their extension to dynamical conditions has proven difficult [Loebel et al. (2009); Scheuss and Neher (2001); Hallermann et al. (2010)].

Explicitly taking into account the variability of the synaptic responses while fitting phenomenological models of STP to experimental data appears highly desirable. Models will be more constrained, as they have to provide a good description not only for the average responses but also for the variability as well as for the correlation between different responses. As a result, more precise estimates of the parameters are expected. Data-constrained modelling of the stochasticity of synaptic responses in dynamical conditions would be extremely helpful in distinguishing different putative mechanisms underlying STP. Different mechanisms could, in fact, predict the same average responses but differ on higher order statistics as, for instance, the coefficient of variation of the responses or the correlation between consecutive responses. Moreover, this would lead to compact (in the spirit of the TM model) models of *stochastic* STP, together with the biologically relevant ranges for the accompanying parameters, to be used in theoretical investigations.

Currently, there is no methodology to quantify responses' variability and STP at the same synapse, and from the same set of recordings, in an integrated and statistically principled way. Here, we provide such a methodology. We use a generative-model approach to build a parametric, probabilistic model of the synaptic response to patterns of pre-synaptic activation. Point-estimates of the model parameters are then obtained by maximum-likelihood estimation. We demonstrate two main advantages of our approach over conventional techniques. First, we simultaneously estimate both quantal and dynamical parameters from the same recordings, consisting of synaptic responses to pre-synaptic spike trains of varying rates at Layer 5 pyramidal-to-pyramidal connections in the ferret

medial pre-frontal cortex [Wang et al. (2006)]. The parameter estimates obtained with our method are consistent with those derived by standard procedures. Second, and most importantly, since the estimation procedure does not rely on trial-averaged quantities, the repetition of identical stimulations becomes unnecessary. Parameters can be estimated from single traces. It is thus possible to devise alternative stimulation protocols and analyze their impact on parameter estimation by the use of theoretical tools. Specifically, by using Fisher Information Matrix theory one can design 'optimal' stimulation protocols (e.g., protocols which minimize the variance of the parameter estimates) for any given synaptic model. As an example, we show that Poisson spike trains yield better parameter estimates than periodic spike trains with the same rate.

3.2 Overview

3.2.1 Stochastic models of repetitive synaptic transmission

The basic unit of modeling in our approach is the release site. The release site can be thought of as being in one of two states: release-competent or refractory. The release-competent state represents a situation where calcium entry upon spike is able to trigger release, that is all the docking and priming processes needed for neurotransmitter exocytosis are completed. Hereafter we refer to all these processes simply as docking. The refractory state represents the complementary situation where, e.g., the vesicle is far from the release site and/or priming is not yet completed. In this case, the spike is unable to trigger release. We introduce a binary variable s to describe the state of the release site, where $s = 1$ denotes the release-competent state and $s = 0$ the refractory state. The transitions between the states describe release upon spike as well as vesicle docking/undocking processes in between spikes (see figure 3.1A). In between spikes, a refractory site becomes release-competent (i.e., $s = 0 \to 1$) with a probability per unit time k_s^+ (docking rate), while a release-competent site becomes refractory with a probability per unit time k_s^- (undocking rate). Release occurs only upon spike, if the site is release-competent, with probability p_{rel}. The site then becomes refractory (i.e., $s = 1 \to 0$). Thus, the state variable s evolves stochastically according to

$$P[s(t+\Delta t) = 1] = P[s(t) = 0] \cdot k_s^+ \cdot \Delta t + P[s(t) = 1] \cdot (1 - k_s^- \cdot \Delta t - p_{rel} \cdot \sum_k \delta(t - t_k)) \quad (3.1)$$

where $P[s(t) = 1]$ is the probability that, at time t, the site is release-competent, and the sum is over all spike times t_k. Note that the transition $s = 1 \to 0$ can occur either as a result of undocking processes or as a result of spike-triggered release. In this latter case, a post-synaptic response is observed. The post-synaptic response to the release of a vesiscle is variable, and we denote q its average (quantal size) and σ_q^2 its variance (quantal noise).

A large class of models of synaptic transmission upon repetitive pre-synaptic activation can be described in this framework. Different pre-synaptic mechanisms of STP can be straightforwardly modeled by making the docking/undocking rates k_s^+ and k_s^-, as well

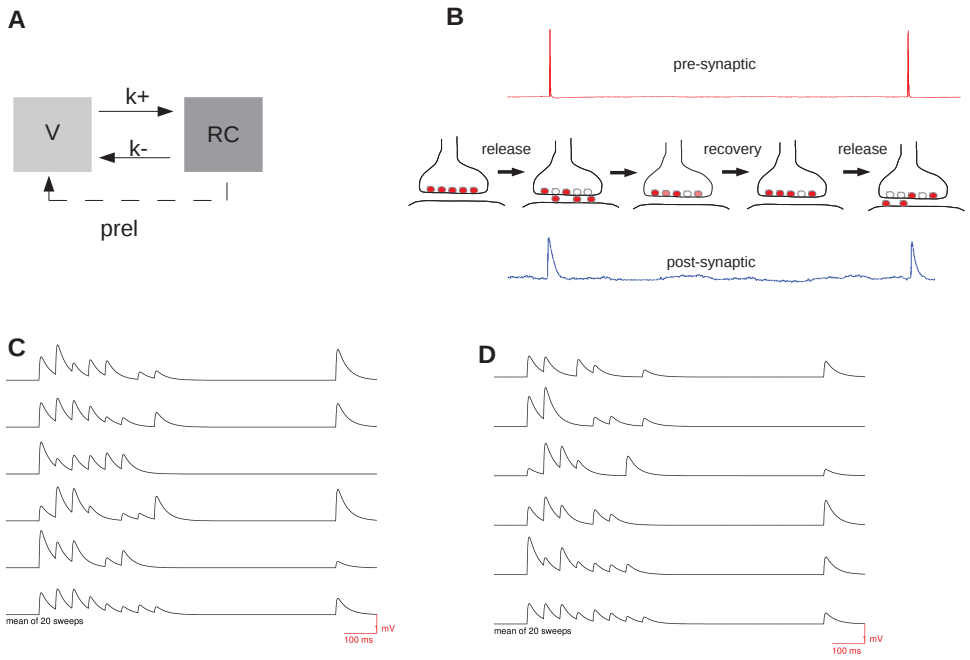

Figure 3.1: **The generative model. A)**: Scheme of the single release site model; solid lines: transitions in between spikes, dashed line: transition upon spike; V: vacant state, RC: release competent state. **B)**: Scheme of the synaptic states and transitions between them. **C)**: single synthetic traces and average for facilitation dominated synaptic connection ($N = 10$, $q = 0.15mV$, $\sigma_q = 0.03mV$, $U = 0.3$, $\tau_D = 195ms$, $\tau_F = 570ms$, $\sigma_{noise} = 0.03mV$). **D)**: single synthetic traces and average for depression dominated synaptic connection ($N = 10$, $q = 0.15mV$, $\sigma_q = 0.03mV$, $U = 0.25$, $\tau_D = 670ms$, $\tau_F = 15ms$, $\sigma_{noise} = 0.03mV$).

as the single-site release probability p_{rel} dependent on the activation history. Thus, we write

$$k_s^+ \equiv k_s^+(\xi_1, \cdots, \xi_M); \;\; k_s^- \equiv k_s^-(\xi_1, \cdots, \xi_M); \;\; p_{rel} \equiv p_{rel}(\xi_1, \cdots, \xi_M) \tag{3.2}$$

where $\xi_1, \cdots, \xi_M$ are additional variables describing the state of the release site. The dynamics of these modulatory variables are typically described by first-order differential equations, where the variables increase/decrease by discrete amounts upon spike, while they decay back to their baseline levels in between spikes. Some examples are in order at this point. The TM model can be obtained by setting

$$k_s^+ = \frac{1}{\tau_D}; \;\; k_s^- = 0 \tag{3.3}$$

where τ_D is the time constant for depression. To model facilitation, an additional state variable is needed, ξ, which evolves according to

$$\dot{\xi} = -\xi/\tau_F + p_0(1-\xi) \cdot \sum_k \delta(t - t_k) \tag{3.4}$$

The probability of release is then given by

$$p_{rel} = p_0 + (1 - p_0) \cdot \xi \tag{3.5}$$

that is, the release probability increases upon spikes (occurring at times t_k), and then decays back to its baseline level, p_0, with a time constant τ_F in between spikes. Equations 3.4-3.5 are a simplified description of the effects of calcium influx into the synaptic terminal, and its subsequent removal, on the probability of release [Bertram et al. (1996); Dittman and Regehr (1998); Neher and Sakaba (2008)]. More biophysical detail can be easily added, e.g., along the lines of Dittman et al. (2000), Trommershäuser et al. (2003) and Kandaswamy et al. (2010).

Modulatory variables need not be continous nor their dynamics deterministic. For instance, one can take into account the existence of different synaptic vesicle pools by introducing state variables which keep track of the number of vesicles available in each pool [Rizzoli and Betz (2005)]. In the simplest scheme, one would consider the recycling pool and the readily releasable pool. The docking rate would be proportional to the number of vesicles in the recycling pool, ξ_{rp} (which is now a discrete variable), i.e., $k_s^+ = \xi_{rp}/\tau_D$. A docking event reduces the number of vesicles in the recycling pool (i.e., $\xi_{rp} \leftarrow \xi_{rp} - 1$), and no further docking is possible as long as the site is release-competent. The docked vesicle belongs then to the readily releasable pool. The recycling pool is refilled at exponentially distributed random times, with average refilling time τ_{rp}, until some maximum is reached, $\xi_{rp}^{(max)}$, after which no further increase in the size of the pool is possible. Biophysical detail comes, however, at the cost of increasing both the number of state variables needed to describe single-site dynamics and the number of parameters. As compared to the TM model, for instance, this scheme requires one additional state variable, ξ_{rp}, and two more parameters, $\xi_{rp}^{(max)}$ and τ_{rp}.

Post-synaptic mechanisms, as for instance receptor desensitization, can be similarly described by making q and σ_q^2 dependent on the release history (see chapter 2).

A synaptic connection is thought of as a collection of identical, statistical independent release sites. At each time, thus, the state of the connection is defined by the number of release-competent sites (a discrete variable, which we denote S, ranging from 0 to N, where N is the number of release sites) and by the current values of all the modulatory variables (e.g., ξ in the TM model). The knowledge of the synaptic state immediately before a spike allows one to compute the probability of observing a given post-synaptic response, once one specifies the distribution of the responses to a single vesicle. Following the quantal model, we assume that the total response is simply the sum of the single quantal responses. For instance, if a single vesicle causes a Gaussian distributed post-synaptic response with mean q and variance σ_q^2, the release of three vesicles will produce a

response which is also Gaussian distributed with mean $3q$ and variance $3\sigma_q^2$. The number of vesicles released, n_{rel}, is binomially distributed according to

$$P(n_{rel} = n) = \binom{S}{n} \cdot p_{rel}^n \cdot (1 - p_{rel})^{(S-n)} \tag{3.6}$$

where S and p_{rel} are the values of the correspondng variables immediately before the spike, and $P(n_{rel} = n) = 0$ for $n > S$.

3.2.2 Parameter estimation from experimental data

Here we describe how, once a specific model is chosen, to estimate *all* the model's free parameters from a *single* set of experimental data. Before doing that, it is instructive to shortly review current, state-of-the-art methodologies.

Quantal analysis aims at estimating the so-called *quantal parameters*, that is the number of release sites N, the average quantal response q and its variance σ_q^2, and the probability of release p_{rel}. For this, the synapse is stimulated at very low rates, while collecting the corresponding post-synaptic responses. Once a sufficiently large number of responses (typically, indeed, a very large number) is collected, and a parametric model for the distribution of responses to unitary quantal events has been chosen, the quantal parameters can be estimated by using Equation 3.6 (with $S = N$) [Del Castillo and Katz (1954)]. An alternative method, the so-called mean-variance analysis, allows one to reduce the number of response needed, as compared to the *classical* quantal analysis, at the cost of manipulating the probability of release by changing extra-cellular calcium concentration [Silver (2003)]. As, in either cases, the rate of stimulation is purposefully chosen so as to allow the synapse to recover its initial state, quantal analysis can not provide any information about the dynamical parameters (e.g., τ_D and τ_F in the TM model). The extension of quantal analysis to dynamical conditions has proven difficult Scheuss and Neher (2001).

To extract information about the dynamical parameters, the synapse is stimulated at high rates (the inter-spike interval must be of the order of the underlying time constants). Standard protocols include pair-pulse stimulation [e.g. Thomson et al. (1993)], long spike trains [e.g. Varela et al. (1997); Dittman et al. (2000)] and short spike trains followed by a recovery spike [e.g. Tsodyks and Markram (1997); Markram et al. (1998); Wang et al. (2006)]. Typically the spike trains are periodic, but this is not necessarily the case [Varela et al. (1997)]. Post-synaptic responses, however, exhibit strong variability which is dealt with by averaging over multiple repetitions of the same pattern of stimulation. The dynamical parameters are then estimated by least-squares fitting the average model response to the average experimental response. Unfortunately, while allowing for least-squares fitting, the averaging procedure also destroys important information contained in the trial-to-trial fluctuations and the within-trial correlation between responses (as we show below). This has important consequences. Clearly, one is unable to estimate the quantal parameters in this way (but see Loebel et al. (2009)). More significantly, the precision of the least-squares estimates is fundamentally limited by the precision of the empirical average response. In view of the large variability exhibited by central synapses,

achieving a reliable estimate of the average response would appear to require quite a large number of repetitions. Furthermore, it is rather cumbersome to quantify the effects of the estimation errors on the average response on the estimation of the model's parameters.

The limitations described above can be easily overcome by noticing that, once the sources of stochasticity are explicitly described, the model allows one to compute the probability of observing any given post-synaptic response as a function of the stimulation protocol and of the model parameters. The model parameters can then be adjusted so that the distribution of the model responses approximates, as closely as possible, the distribution of the experimental responses. More technically, this requires to minimize the Kullback-Leibler divergence (i.e., a measure of similarity between two distribution functions) between the *experimental* and the *model* distribution of synaptic responses. This can be achieved by determining the model parameters by maximum-likelihood estimation.

Let us consider a given pattern of stimulation $t_{1\to M} \equiv \{t_1, t_2, \cdots, t_M\}$, where t_k denotes the time of k-th pre-synaptic spike. A sequence of responses $R_{1\to M} \equiv \{R(t_1), R(t_2), \cdots, R(t_M)\}$ will correspondingly be observed. The likelihood function is defined as

$$L(\boldsymbol{\theta}|R_{1\to M}) \equiv P(R_{1\to M}|\boldsymbol{\theta}) \tag{3.7}$$

that is, the probability of observing the actually-observed sequence of responses as a function of the model parameters, which we denote $\boldsymbol{\theta}$. We want to maximize $L(\boldsymbol{\theta}|R_{1\to M})$ with respect to $\boldsymbol{\theta}$. The direct maximization of Equation 3.7, however, turns out to be impractical because the likelihood function can not be expressed conveniently in an analytical form, although it can efficiently be evaluated numerically (as we show in the Methods section). On the other hand, the joint probability of the observed responses *and* the underlying (hidden) sequence of synaptic states responsible for their generation is easily written down. We denote $U^-_{1\to M} \equiv \{U(t_1^-), U(t_2^-), \cdots, U(t_M^-)\}$ and $U^+_{1\to M} \equiv \{U(t_1^+), U(t_2^+), \cdots, U(t_M^+)\}$ the synaptic states immediately before and after the corresponding spikes, respectively. The joint probability of the responses $R_{1\to M}$, and synaptic states $U^-_{1\to M}$ and $U^+_{1\to M}$, $P(R_{1\to M}, U^-_{1\to M}, U^+_{1\to M}|\boldsymbol{\theta})$ reads

$$P(R_{1\to M}, U^-_{1\to M}, U^+_{1\to M}|\boldsymbol{\theta}) = P(U_1^-|\boldsymbol{\theta}) \prod_{k=1}^{M} P(U_k^+|U_k^-, \boldsymbol{\theta}) P(R_k|U_k^+, U_k^-, \boldsymbol{\theta}) \prod_{k=1}^{M-1} P(U_{k+1}^-|U_k^+, \boldsymbol{\theta}) \tag{3.8}$$

The conditional probabilities appearing in the above equation have a straightforward interpretation: $P(U_1^-|\boldsymbol{\theta})$ is the steady distribution of synaptic states in absence of stimulation; $P(U_k^+|U_k^-, \boldsymbol{\theta})$ is the probability that the synaptic state changes from U_k^- to U_k^+ upon spike; $P(R_k|U_k^+, U_k^-, \boldsymbol{\theta})$ is the probability of observing a post-synaptic response R_k when the synaptic states changes from U_k^- to U_k^+; $P(U_{k+1}^-|U_k^+, \boldsymbol{\theta})$ is the probability that the synaptic state changes from U_k^+ to U_{k+1}^- during the time interval $t_{k+1} - t_k$ in absence of spikes. Once the model has been specified, these conditional probabilities are easily computed.

As $P(R_{1\to M}, U^-_{1\to M}, U^+_{1\to M}|\boldsymbol{\theta})$ conveniently factorizes, the maximum-likelihood estimation of the model parameters can be efficiently carried out by using the Expectation-Maximization (EM) algorithm [Rabiner (1989)]. Instead of maximizing the likelihood function in Equation 3.7, one maximizes the so-called auxiliary function with respect to $\boldsymbol{\theta}$. The auxiliary function, $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(old)})$, is defined as

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(old)}) = \sum_{U^-_{1\to M}, U^+_{1\to M}} P(U^-_{1\to M}, U^+_{1\to M}|R_{1\to M}, \boldsymbol{\theta}^{(old)}) \log \left[P(R_{1\to M}, U^-_{1\to M}, U^+_{1\to M}|\boldsymbol{\theta})\right] \tag{3.9}$$

where $\boldsymbol{\theta}^{(old)}$ is an initial guess for the parameters, and the sum is over all possible sequences of synaptic states. This is the so-called E-step, as the evaluation of the auxiliary function requires the computation of the expectation of $\log \left[P(R_{1\to M}, U^-_{1\to M}, U^+_{1\to M}|\boldsymbol{\theta})\right]$ over the distribution of synaptic states, conditional on the observed responses and on the current parameters estimate. In the M-step, a new parameters estimate, $\boldsymbol{\theta}^{(new)}$, is obtained

$$\boldsymbol{\theta}^{(new)} = \arg\max_{\boldsymbol{\theta}} \; Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(old)}) \tag{3.10}$$

so that $L(\boldsymbol{\theta}^{(new)}|R_{1\to M}) \geq L(\boldsymbol{\theta}^{(old)}|R_{1\to M})$. By iterating the E- and M-step, one can improve the inital guess until eventually a fixed point is reached (i.e., $\boldsymbol{\theta}^{(new)} = \boldsymbol{\theta}^{(old)}$), which corresponds to a (local) maximum of the likelihood function $L(\boldsymbol{\theta}|R_{1\to M})$. In the Methods section, we show how all the quantities needed to carry out EM can be efficiently computed, and we obtain explicit re-estimation formulas for the parameters in the case of the TM model. Although we have explictly worked out only the case of the TM model, for reasons to be explained shortly, the techniques and algorithms described in Methods are straightforwardly applicable to any model which admits a factorization of the joint probability of the responses and synaptic states as in Equation 3.8.

3.3 Results

In order to demonstrate the feasibility of our approach, and the advantages it entails, we choose the specific instantiation of the release site and release probability dynamics that corresponds to the *stochastic* TM model. Several considerations motivated this choice: (i) the TM model has a small number of free parameters, thus minimizing potential problems of overfitting, and yet is able to describe very diverse STP patterns; (ii) The data set we presently analyse have been previously analysed with the TM model, which allows for direct comparison with parameter estimates, obtained by least-squares fit, reported in Wang et al. (2006); (iii) The *stochastic* TM model has been widely employed in theoretical investigations of functional/computational implications of synaptic variability, thus it appeared relevant to asses to which extent the model captures variability in real synapses.

The general structure of the model is illustrated in Fig. 3.1**B**. It has 6 free parameters: the number of release sites N, the quantal size q, the quantal variability σ_q, the initial

release probability p_0 and the time constants for docking and facilitation, τ_D and τ_F respectively. The model contains three sources of variability: both release and docking are stochastic processes, and the post-synaptic response exhibits fluctuations at parity of number of vesicles released. These sources of variability are all well documented experimentally [e.g. Faber et al. (1992); Bekkers et al. (1990); Franks et al. (2003); Branco and Staras (2009); Ribrault et al. (2011)]. To illustrate the *generative* sufficiency of the model, that is its ability to qualitatively reproduce the recorded traces when probed with the experimental protocol, we show the resulting synthetic single traces, together with the trial-averaged trace, for a representative facilitating connection in Fig. 3.1**C**, and for a representative depressing connection in Fig. 3.1**D**. The variability of the single traces is substantial and comparable to the one observed in the experimental data (see *Response variability*).

The data set we analysed consisted of dual whole-cell patch clamp recordings between synaptically connected layer 5 pyramidal cells in the medial pre-frontal cortex of adult ferrets (1.5-3 months) (see Wang et al. (2006) for details about electrophysiology and the experimental set-up). The stimulation protocol consisted of a regular train of 5-8 spikes, at varying frequencies, followed by a recovery spike. Post-synaptic responses were recorded in the current-clamp mode. The inter-spike interval of the train, T, ranged between 14.3ms and 200ms (ν = 5 - 70Hz), with most of the recordings carried out at T = 50ms, and the interval for the recovery spike was correspondingly determined as $T_{rec} = T + 500$ms. Patterns of transmission at pre-frontal cortex synapses were found to be especially complex, as compared to patterns in primary sensory cortices [e.g. Varela et al. (1997); Tsodyks and Markram (1997); Markram et al. (1998)], exhibiting both depressing and facilitating components over a wide range of time scales [Wang et al. (2006)].

3.3.1 Response variability

We began by analysing response variability, which was not previously done, and found that synaptic responses exhibited strong variability. For purpose of illustration, we show in Fig. 3.2 sample voltage traces for one facilitating (**A**) and one depressing (**B**) connection, together with the corresponding trial-averaged traces. In both cases, the large variability across the different repetitions is immediately evident. The variability is, indeed, so strong that the facilitating/depressing nature of the transmission is largely concealed in the single traces, while it becomes readily apparent in the trial-averaged traces.

To quantify variability for each connection, we extracted from the corresponding single-trial voltage traces the peak excitatory post-synaptic response (pEPSP) corresponding to each pre-synaptic spike. The procedure is detailed in *Methods*. We then computed, for each connection and for each response, the associated coefficient of variation (CV). The results of this analysis are shown in Fig. 3.2**D**, where we report the histograms of the CV across the population of synaptic connections separately for each response. The average of the initial response ranged between 0.11mV and 3.43mV (0.70 ± 0.61mV; $n = 69$), while the associated CV ranged between 0.21 and 1.58 (0.59 ± 0.27; $n = 69$).

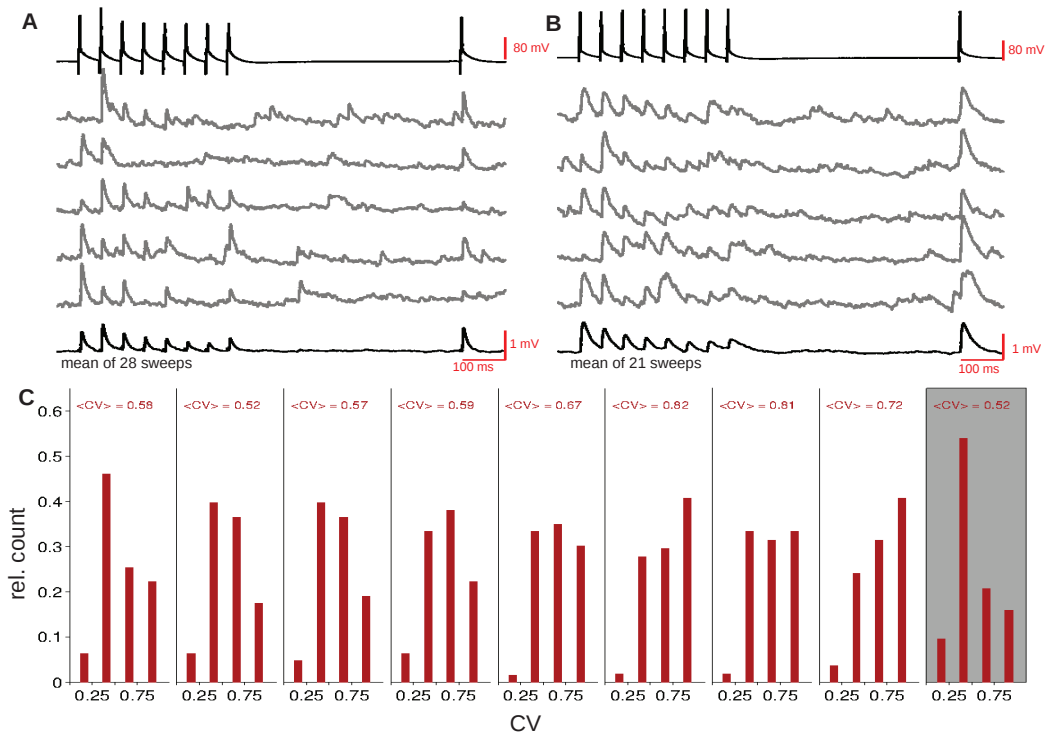

Figure 3.2: **Variability of synaptic transmission. A)**: Input protocol (first line), single sweeps (grey) and average response (last line) for a facilitation dominated synaptic connection. **B)**: Input protocol (first line), single sweeps (grey) and average response (last line) for a depression dominated synaptic connection. **C)**: Histograms of the CVs for each response index; grey shaded panel: recovery response. Data from Wang et al. (2006).

The average values for the synaptic efficacy (i.e., the trial-averaged pEPSP to the first spike in the train) and for the CV, as well as the corresponding ranges, are fully consistent with previous studies [Thomson and Deuchars (1997); Markram et al. (1997)]. We took this as an indication of the reliability of the method we used for isolating synaptic responses within the single-trial voltage traces.

Synaptic unreliability remained high all along the stimulation, and it even increased for late responses, as can be seen in Fig. 3.2**D**. This is a consequence of the increasing probability of failure due to vesicles depletion. The population-averaged CV was smallest for the second and the recovery response. This is a consequence of the increasing probability of release occurring at facilitating synapses while release-ready vesicles are still abundant (i.e., before depression builds up). The second and the recovery response were, in fact, the most facilitated responses on average.

One immediate consequence of such high levels of variability is that, to achieve a relatively accurate estimate of the average response, one needs a large number of repetitions.

For a *true* CV of 0.3, an estimate of the *true* average response within 10% relative precision would require about 80 repetitions, while an estimate within 5% relative precision would require more than 300 repetitions. Given the CVs estimated from the data, these figures should be considered as the *minimal* number of repetitions needed to estimate the average response with reasonable accuracy.

The above considerations clearly illustrate the inadequacy of least-squares fitting techniques for estimating synaptic parameters at unreliable connections. On one hand, parameters estimate obtained with a small number of repetitions (e.g., 20-30 trials) could be grossly imprecise, as a result of the poor estimate of the average responses. On the other hand, increasing the number of repetitions to improve the accuracy of the empirical averages would lead to a very inefficient use of the experimental data. Out of hundreds of responses, in fact, we would just be distilling one number, the average response, to be used in the fitting procedure. Note that high levels of variability (i.e., high CVs) are the rule, rather than the exception, for central chemical synapses.

3.3.2 Maximum-likelihood estimation of the synaptic parameters

To estimate the synaptic parameters $\boldsymbol{\theta} = \{N, q, \sigma_q, p_0, \tau_D, \tau_F\}$ for each connection we proceeded as follows. For a fixed value of N (note that N takes on only integer values, while the other parameters are continuous), the maximum-likelihood estimate of the remaining parameters can be straightforwardly obtained by using the EM algorithm described in *Methods*. We repeatedly applied the EM algorithm while varying N between 1 and 100, and obtained the parameters which maximizes the likelihood for each N. We then selected the value of N (and of the corresponding parameters) for which the likelihood was maximal, as the maximum-likelihood estimate.

The above procedure is illustrated in Fig. 3.3**A** for a sample connection. In the left panel, we plot the log-likelihood as a function of N while, in the right panels, we plot the values of the parameters which maximize the log-likelihood for the corresponding N. As can be seen, the log-likelihood exhibits a clear maximum at $N = 17$. The values of remaining parameters can be read from the corresponding curves. They are: $q = 0.18$mV, $\sigma_q = 0.06$mV, $p_0 = 0.27$, $\tau_D = 202$ms and $\tau_F = 449$ms. Using the estimated parameters, we then generated 500 synthetic experiments in which the model is probed with the same stimulation protocol, and for the same number of trials, as in the real experiment. We then computed the average responses and the associated CVs for each experiments (28 trials), and from these the corresponding grand-averages together with 95% confidence intervals. The results are shown in Fig. 3.3**B**. In the top panel, we report the experimental (black curve) and the model average responses (red curve - error bars represent 95% confidence interval). For comparison, in the same panel we also report the least-squares fit to the experimental average responses (blue curve). In the bottom panel, we report the experimental (black curve) and the synthetic CVs (red curve - error bars represent 95% confidence interval). As can be seen, both for the average responses and for the CVs, the experimental data are well reproduced by the model. It is important to stress that the model parameters have not been selected by the estimation procedure to reproduce the average responses, as it is the case for the least-squares fit, nor the CVs

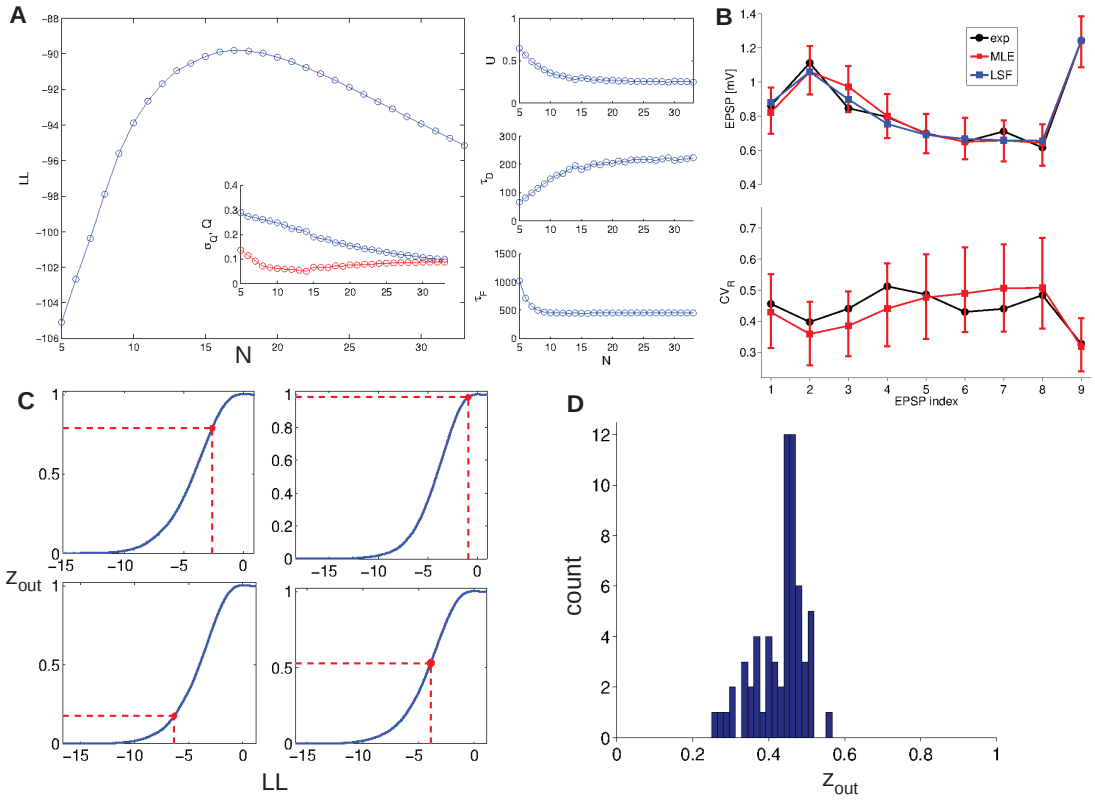

Figure 3.3: **MLE fit to experimental data. A)**: Log-likelihood and model parameters as functions of N. Inset, blue: q; red: σ_q. **B)**: pEPSPs and CVs versus response index of a single exemplary synaptic connection; black: experimental data, blue: LS-fit, red: MLE-fit; error bar denote 95 % confidence intervals of the model prediction. **C)**: Generative log-likelihood cumulative distributions for 4 different instances of the leave-one-out procedure (blue curve). Log-likelihood and z_{out} of the corresponding test trial (red dot). **D)**: Distribution of z_{out} values over the entire data set.

but rather to maximize the probability of the trains of responses observed in single trials.

In 6 out of 69 cases the estimation procedure returned values for one or more parameters that were judged problematic. In 4 cases, the estimation procedure returned values for one or both the time constants (i.e., τ_D and τ_F) that were several orders of magnitude larger than the longest time scale at which the synaptic connections were probed. In the remaining 2 cases, the procedure returned very small values for q (i.e., far outside the reported physiological range) because the maximum of the log-likelihood was achieved for large values of N. These connections were excluded from further analysis.

Next, we checked whether the model was overfitting the data by using a leave-one-out cross-validation procedure. We estimated the parameters as described above while leaving out one set of responses (i.e., one trial). With the parameters thus obtained, we computed the probability that the model would generate a set of responses with a log-

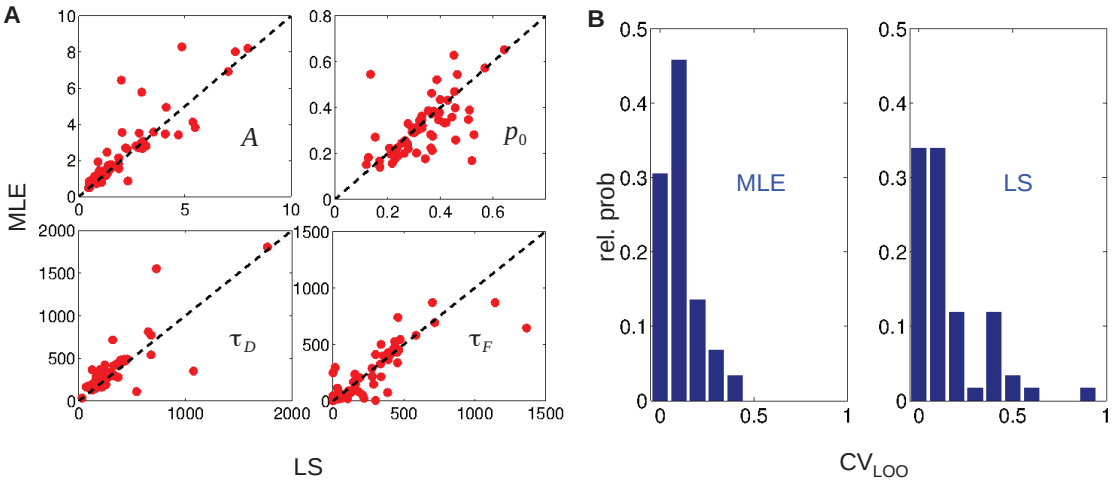

Figure 3.4: **Comparison of LS and MLE estimates. A)**: Average estimates obtained with leave-one-out procedure; results obtained with MLE plotted versus results obtained with LS for each synaptic connection. **B)**: Distributions of CV of the estimator obtained from leave-one-out procedure for each method.

likelihood equal or smaller than the log-likelihood of the set of responses that was actually left out. We denote this probability by z_{out}. This procedure is illustrated in Fig. 3.3**C** for the same connection as in Fig. 3.3**A**. In each sub-panel we show (i) the cumulative distribution of the log-likelihood for a set of responses generated from the model, where the parameters are estimated by leaving out one trial (blue curve); (ii) the value of the log-likelihood of the set of responses left out during the estimation procedure; (iii) the corresponding value of z_{out}. For sets of responses generated by the model, one expects z_{out} to be uniformly distributed between 0 and 1. For each connection, we thus computed the average z_{out} across all trials. The corresponding distribution across the data set is shown in Fig. 3.3**D**. As can be seen, most of the values are between 0.4 and 0.5. For only 5 out of 63 connections the obtained z_{out} where significantly different from the uniform distribution (Kolmogorov-Smirnov test; $p = 0.01$). For 3 of these connections, less than 25 trials were available. When only connections with 30 trials or more were checked, for no one (out of 10) the resulting z_{out} showed statistically significant deviation from the uniform distribution. We concluded that, for the large majority of the connections, there were no overfitting problems, while for the remaining it is very likely that we simply lacked statistical power to asses overfitting.

3.3.3 Maximum-likelihood vs. least-squares estimation

We estimated the synaptic parameters by least-square fitting using, as above, a leave-one-out procedure. The purpose of this analysis was to compare the accuracy as well as the stability of the two estimation procedures. Notice that the two methods are expected

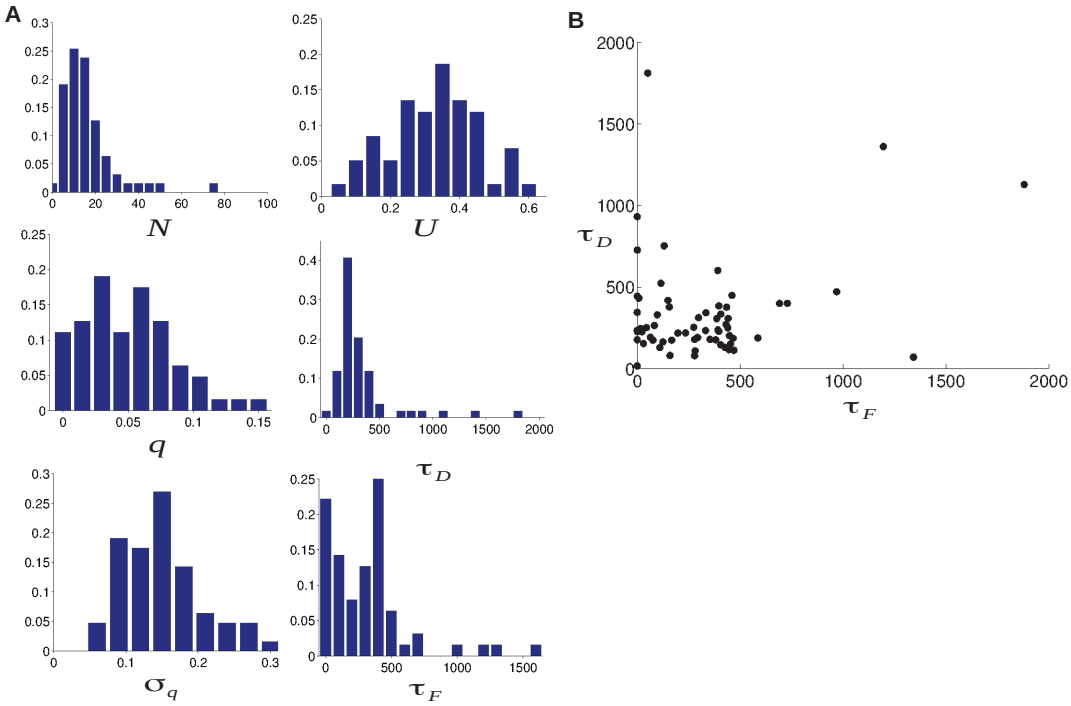

Figure 3.5: **Population data. A)**: Distributions of the synaptic parameters across the population. **B)**: Plot of τ_D versus τ_F; each dot represents one synaptic connection.

to return the same estimates (for $A = N \cdot q$, p_0, τ_D and τ_F). This is confirmed in Fig. 3.4**A**, where we plot the estimates obtained with our method *vs.* the estimates obtained with the least-square fitting procedure. As can be seen, the estimates obtained with the two methods are highly correlated, with the large majority of points lying very close to the diagonal.

The leave-one-out procedure allowed us to evaluate the average coefficient of variation of the estimates with the two procedures, which we took as a quantitative measure of the accuracy and stability. The corresponding distributions are reported in Fig. 3.4**B** (left panel: maximum-likelihood; right panel: least-square fit). As can be seen, there is a tendency for the maximum-likelihood procedure to exhibit larger accuracy and stability. We could, however, not detect any statistically significant difference between the two distributions (Kolmogorov-Smirnov test; $p = 0.01$).

We concluded that our method exhibited an accuracy and stability certainly not worse than the least-square fitting procedure, while allowing one to estimate 2 more parameters.

3.3.4 Population analysis

We compared the parameters estimates obtained with our method with the corresponding estimates, obtained by least-squares fit, reported in Wang et al. (2006). When averaging

across the population (denoted by angular brackets) we obtained for the absolute synaptic efficacy $\langle A \rangle \equiv \langle N \cdot q \rangle = 2.20 \pm 1.72$ mV (range: 0.48 mV - 7.97 mV), for the initial release probability $\langle p_0 \rangle = 0.33 \pm 0.13$ (range: 0.05 - 0.73), for the time constant of the docking process $\langle \tau_D \rangle = 335 \pm 306$ ms (range: 16 ms - 1800 ms), and for the time constant of facilitation $\langle \tau_F \rangle = 321 \pm 340$ ms (range: 0 ms - 1900 ms). The corresponding distributions are shown in Fig. 3.5**A**. The population averaged values reported in Wang et al. (2006) are: $\langle A \rangle = 3.46 \pm 2.79$ mV; $\langle p_0 \rangle = 0.27 \pm 0.15$; $\langle \tau_D \rangle = 396 \pm 163$ ms; $\langle \tau_F \rangle = 292 \pm 240$ ms. They all are in excellent agreement with our estimates. Also, consistently with the results reported in Wang et al. (2006), we found both strongly facilitating (i.e., $\tau_F \gg \tau_D$) and strongly depressing (i.e., $\tau_F \ll \tau_D$) connections in the synaptic population (see Fig. 3.5**B**).

With our method, we were also able to estimate the quantal parameters. When averaging across the population we obtained for the quantal size $\langle q \rangle = 0.15 \pm 0.06$ mV (range: 0.06 - 0.32 mV), for the quantal noise $\langle \sigma_q \rangle = 0.05 \pm 0.05$ mV (range: 0.00 - 0.31 mV), and for the number of release sites $\langle N \rangle = 15 \pm 12$ (range: 2 - 77). The corresponding distributions are showed in Fig. 3.5**A**. The estimate of $\langle q \rangle$ obtained with our method is in excellent agreement with those obtained with other techniques in intra layer 5 connections in rat somato-sensory cortex ($\langle q \rangle = 0.13 \pm 0.04$ mV; $n = 20$; [Loebel et al. (2009)]) and at layer 4 spiny stellar cell connections to layer 2/3 pyramidal cells in the rat barrel cortex ($\langle q \rangle = 0.15 \pm 0.02$ mV; $n = 32$; [Silver et al. (2003)]). Hardingham and collaborators Hardingham et al. (2010) found a somewhat larger value in rat intra layer 5 connections ($\langle q \rangle = 0.211mV \pm 0.065mV$; $n = 20$), which however lies still inside the error bar range. Finally, Thomson and West Thomson and West (1993) report quantal sizes ranging from $q = 0.179mV \pm 0.017mV$ to $q = 0.382mV \pm 0.093mV$ for intra layer 2/3 and intra layer 4 connections in the rat somato-sensory/motor cortex (cfr. Fig. 3.5**A**). In order to compare the estimated value of σ_q we calculated the average quantal coefficient of variation, $\langle CV_q \rangle = \langle \frac{<\sigma_q>}{<q>} \rangle = 0.38 \pm 0.28$ (range: 0.00 - 1.21). This value is similar to the value reported by Silver and collaborators ($\langle CV_q \rangle = 0.43 \pm 0.06$; $n = 32$; Silver et al. (2003)). The number of release-sites obtained by our analysis ranges from 2 to 77. We are aware of only one study that estimated the number of release sites N in dynamical conditions [Loebel et al. (2009)]. The range for N reported in that study was 7-170, with $\langle N \rangle = 53 \pm 42$.

We next searched for correlations between the different synaptic parameters. We found a strong correlation between the number of release sites N and the synaptic strength, as measured by the largest average pEPSP encountered in the train of responses ($R = 0.76$, $p < 10^{-12}$). This result is in line with the results of Loebel et al. (2009) for purely depressing synapses. We also found a weak, but statistically significant correlation between the quantal amplitude q and the synaptic strength ($R = 0.25$, $p < 0.05$), in line with the results of Hardingham et al. (2010). We were unable to detect other correlations.

It should be noted that the data used for this study were not recorded for the purpose of performing quantal analysis, that is, in many cases only a small number of repetitions are available (range: 8-43 trials). Nonetheless, the maximum-likelihood estimation procedure returned quantal parameters which were in good agreement with the estimates

obtained in similar preparations using especially-tailored estimation procedures.

3.3.5 Poisson input protocol outperforms repetitive input protocols

Estimating synaptic parameters by least-squares fit requires the experimenter to repeat the stimulation identically a certain number of times (this number increasing with the unreliability of the synaptic connection - see *Response variability*), in order to estimate the average post-synaptic responses. This is not necessary with our method. The question thus naturally arises as to whether varying the stimulation from trial to trial might lead to improved estimates.

To investigate this issue we ran numerical experiments in which we compared the accuracy of the estimates obtained in the two conditions (i.e., repetitive *vs.* non-repetitive stimulation). In the repetitive condition, we used the actual experimental protocol. In the non-repetitive condition, for each trial we generated a train of 9 spikes where the inter-spike intervals are randomly and independently drawn from an exponential distribution with the same average inter-spike interval as in the repetitive condition. An additional 500ms are added to the last inter-spike interval (see Fig. 3.6**A**). Notice that both the number of responses obtained and the recording time (on average) are matched in the two conditions.

The parameters of the model were set to the corresponding population-averaged values as estimated from the physiological recordings (see *Population analysis*). Using the model, we ran an experiment of 20 trials for each condition, and estimated the parameters using our method and, in the repetitive condition, also using the least-squares fit procedure. The frequencies of stimulation were chosen to be the same as in the real experiment. At parity of condition, we obtained different estimates for each experiment, due to the stochasticity of the model. In the non-repetitive condition, the variability of the stimulation across trials is an additional source of stochasticity. Thus, to quantify the accuracy of the estimates, we computed for each parameter j ($j = 1 \cdots N_{par}$, where N_{par} is the number of estimated parameters) the corresponding standard deviation of the relative error across 500 experiments in the same condition. Finally, to obtain just a single number, we averaged over the parameters, i.e.,

$$\epsilon = \frac{1}{N_{par}} \sum_{j=1}^{N_{par}} \sqrt{\frac{\langle \hat{\theta}_j^2 \rangle - \theta_j^2}{\theta_j^2}} \equiv \frac{1}{N_{par}} \sum_{j=1}^{N_{par}} \epsilon_j \tag{3.11}$$

where the sum over j runs over all estimated parameters ($N_{par} = 6$ for our method, $N_{par} = 4$ for the least-squares fit), $\hat{\theta}_j$ is the estimate of the parameter j, whose true value is θ_j, obtained in one experiment, and the angular brackets denote average over the experiments. The larger the value of ϵ the less accurate are, on average, the estimates. Notice that we have used the *true* values of the parameters in Eq. 3.11 because both maximum-likelihood and least-squares fit are expected to return unbiased estimates. We have nevertheless checked that in our numerical experiments this was indeed the case (data not shown).

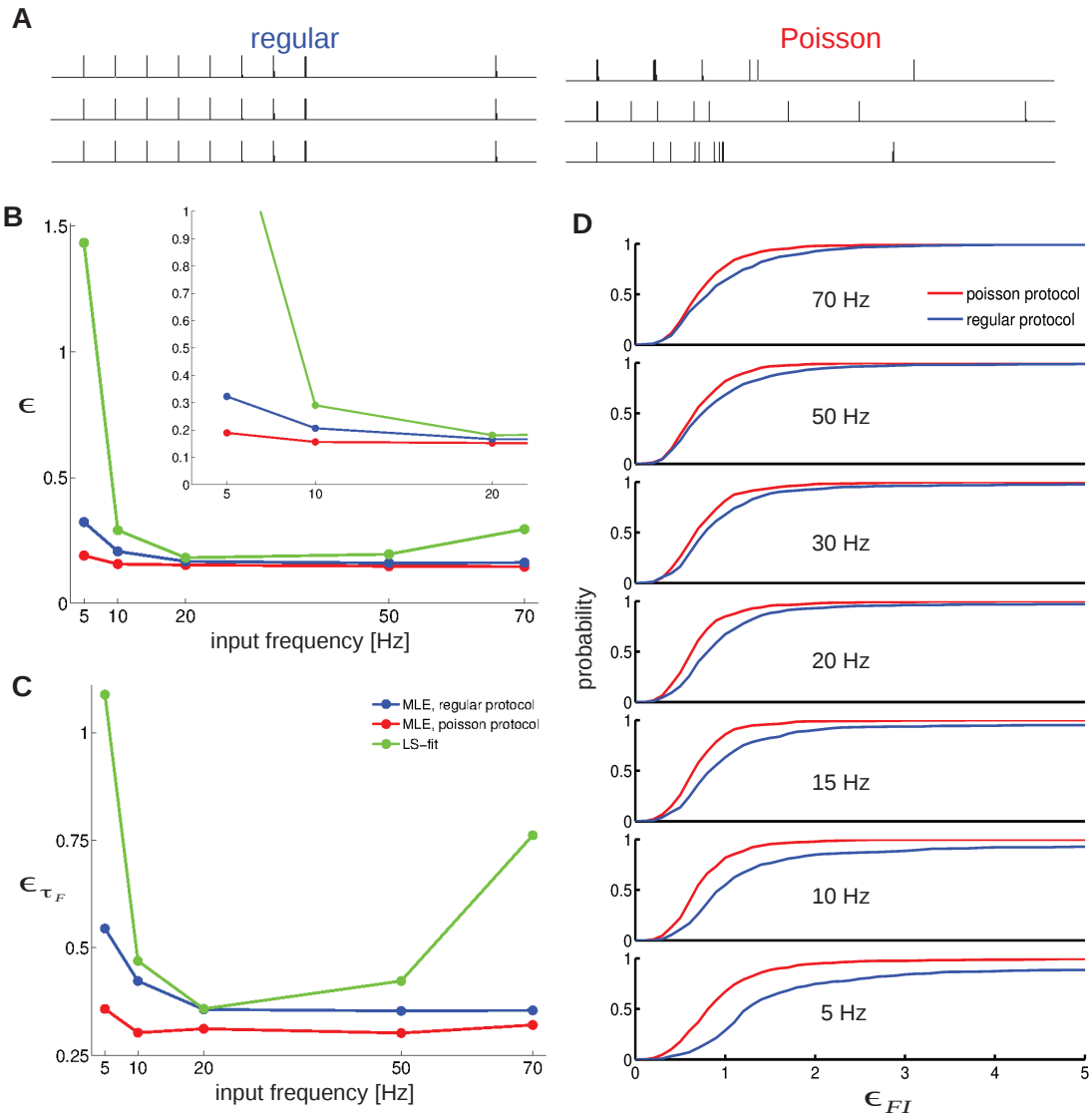

Figure 3.6: ***Regular protocol* versus *Poisson protocol*.** **A)**: Scheme of the two used input protocols. **B)**: ϵ versus input-train frequency; blue: *regular protocol*, red: *Poisson protocol*, green: *least-squares fit*; inset: zoom on the low frequency range. **C)**: Relative error of the worst estimate, ϵ_{τ_F}, versus input-train frequency. **D)**: Cumulative distributions of relative errors obtained from

In Fig. 3.6**B** we plot ϵ as a function of the frequency of stimulation for the repetitive (blue curve: maximum-likelihood estimation; green curve: least-square fit) and the non-repetitive condition (red curve). In the inset we also plot, with the same color code, a

zoomed version.

As can be seen, the estimates obtained in the non-repetitive condition are more precise than those obtained in the repetitive condition. The overall improvement of ϵ seems moderate, but is greatest in the physiologically relevant region of low input frequencies. However, as shown in Fig. 3.6**C**, the relative error of the worst estimate (τ_F) is significantly reduced in the non-repetitive condition, across all frequencies. Estimates obtained with the least-squares fit are inferior for all stimulation frequencies.

Finally, we verified that the superior performance of the Poisson protocol with respect to the regular protocol does not depend on the concrete parameter choice. To do so, we calculated the fisher-information of the parameter estimates (see *Methods*) in function of the input protocols for 500 parameter sets that were drawn from the experimentally obtained distributions. As explained in *Methods*, the fisher-information can be used to obtain a lower bound on the parameter estimates. From this a relative error ϵ_{FI} can be obtained, similar to expression 3.11.

In Fig. 3.6**D** we plot the cumulative distributions of ϵ_{FI} for the regular (blue curve) and Poisson protocol (red curve) for various input frequencies. The Poisson protocol yields on average lower ϵ_{FI} than the regular protocol in all conditions. This effect is more prominent for small input frequencies, where the ϵ_{FI} obtained from the regular protocol remarkably display a long tail. The difference between the protocols becomes smaller for higher input frequencies, but the cumulative distributions remain, however, significantly different in all conditions (two-sample Kolmogorov-Smirnov test, $p < 10^{-}4$).

3.4 Discussion

We have developed a general statistical framework that integrates the features of quantal and STP models in a uniform scheme. By applying standard machine learning techniques, it can be used to extract synaptic parameters and to quantify simultaneously dynamic and statistical properties of synaptic transmission. We demonstrated with the analysis of experimental recordings that this is feasible in practice while the obtained results are consistent with previously reported ones. Furthermore, the approach provides as an essential novelty a freedom in the choice of the input protocol, which can be utilized to find more efficient experimental setups.

In this work, the two components of the stochastic framework, the binomial quantal model and the TM model, haven been chosen in part for their simplicity and illustrative clearness. However, our framework is general. It imposes no restrictions on more extensive modeling, e.g. the inclusion of dynamics with more time constants, of post-synaptic effects or of more than one vesicle pool. As long as single sequences of synaptic responses can be formulated as being caused by a discrete Markov chain of synaptic states, a likelihood function can be derived as shown in the example at hand.

The analysis of the double cell recordings presented in this work is intended to be a feasibility study. The estimated synaptic parameters are in reasonable agreement with results obtained by state-of-the-art analysis. This is in spite of the fact that the recordings were originally not intended for quantal analysis, that is, in many cases only a small number of

repetitions were available. Furthermore, the pre-processing applied to the data was kept very basic, e.g. no explicit identification of release failures was performed. An inclusion of this will further increase the precision of the estimates.

3.4.1 Parameter estimates

For some connections we observed that the dynamical parameters obtained with the deterministic TM model deviated significantly from the maximum likelihood estimates. These connections also exhibited particularly variable responses, suggesting that the neglect of higher order statistics, as it is done when fitting to trial averaged data, is not justified in these cases. Fits to synthetic data indeed show that the uncertainty of the estimators is systematically higher for the deterministic model compared to the full stochastic model. This is attributable to the superiority of the MLE compared to the LS fit. It should be stressed that this difference vanishes in the limit where an infinite amount of data is available. However, if we consider the cases which are relevant in practice, we find that the MLE approach always yields an improvement in the precision of parameter estimates.

3.4.2 Free choice of input patterns

The generative model approach grants a novel freedom in the choice of the pre-synaptic input since it renders trial averaging unnecessary. As we have shown, this can be used to design input protocols which, for a given amount of data and a given model, allow a preciser estimation of model parameters as compared to the classical input protocol. Equivalently, with a superior experimental design it becomes possible to obtain the same estimator quality with shorter measurements. In addition, our method makes it possible to estimate synaptic parameters from single input spike trains of arbitrary length. It becomes thus feasible to economize the intervals between single trials used to restore the synapse's initial state. In principle, this gained time again can be used to record more responses from the same connection or from a higher number of different connections.
Note that the notion of 'input protocol quality' is associated with a given model: for different models there may be different optimal input protocols, all of which can be characterized with the techniques outlined in this work. However, independent of the particular model used, the free input choice allows to use physiologically more realistic inputs, like Poisson trains or *in vivo* recorded trains. Extraction of synaptic parameters from realistic input patterns will remove possible artificial contributions stemming from the regularity of the standard input protocol.

3.4.3 Further benefits

The probabilistic nature of the modeling framework provides us with a likelihood measure which makes it possible to compare the explanatory power of different stochastic models. In the simplest case, when two models with equal numbers of free parameters are fitted to a given data set, the model which returns a higher likelihood value is preferable. This allows e.g. to compare models which make different mechanistic assumptions,

which could given some insight into the microscopic processes involved in STP. Model comparison with the full stochastic framework is in particular useful in cases where different models produce responses with the same average traces but different statistics.
A straightforward extension of our formalism is the inclusion of prior distributions. If some prior knowledge on the synaptic parameters is available, this information can be implemented by simply multiplying the prior distributions of the parameters to the likelihood function. The corresponding optimization is known as maximum a posteriori probability (MAP) estimation. Also for MAP estimation efficient EM re-estimation formulas can be found.

3.5 Methods

3.5.1 The stochastic Tsodyks-Markram model

Assuming the identity of all release sites and the settings of the dynamic transitions as outlined above in equations 3.3, 3.4 and 3.5, we obtain a stochastic model that exhibits the same average response as the Tsodyks-Markram model. This model contains 6 free parameters: the number of release sites N, the quantal size q, the quantal variability σ_q, the initial release probability p_0 and the time constants of depression and facilitation τ_D and τ_F. In the following, we will show in detail how the transition probabilities can be calculated, how the parameters can be estimated via EM and how the Fisher-information-matrix can be determined.

3.5.2 Likelihood of a response sequence

We compute the probability of observing a sequence of post-synaptic responses $R_{1\to M} \equiv \{R_1, \cdots, R_M\}$ in correspondence to a train of pre-synaptic spikes occurring at times $t_{1\to M} \equiv \{t_1, \cdots, t_M\}$. In the case of the stochastic TM model, the release probability $p_{rel,k}$ is a deterministic function of the input protocol and can consequently be calculated *ab initio*. The only state variable upon which the observed responses will depend are thus the occupation states of the synapse immediately before and after each spike. We therefore set $U^-_{1\to M} = S^-_{1\to M}$ and $U^+_{1\to M} = S^+_{1\to M}$.
Note that the difference $S^-_k - S^+_k$ is the number of vesicles released upon the k-th spike. The occupation states are not directly observable nor they are a deterministic function of the spike times as the release probability. Thus, we need to compute first the joint probability of a sequence of responses and a sequence of occupation states, $P(R_{1\to M}, S^-_{1\to M}, S^+_{1\to M}|t_{1\to M}, \boldsymbol{\theta})$, and then marginalise over $S^-_{1\to M}$ and $S^+_{1\to M}$ to obtain $P(R_{1\to M}|t_{1\to M}, \boldsymbol{\theta})$. To enlighten the notation, we drop hereafter the dependence of the various probabilities on the spike times (given) and on the synaptic parameters (assumed constant). We start by rewriting $P(R_{1\to M}, S^-_{1\to M}, S^+_{1\to M})$ as

$$P(R_{1\to M}, S^-_{1\to M}, S^+_{1\to M}) = P(S^-_1) \prod_{k=1}^{M} P(S^+_k|S^-_k) P(R_k|S^+_k, S^-_k) \prod_{k=1}^{M-1} P(S^-_{k+1}|S^+_k). \quad (3.12)$$

The conditional probabilities appearing in the above equation are easily computed. We consider first $P(S_k^+|S_k^-)$, which is the probability that the occupation state changes from S_k^- to S_k^+ upon the k-th spike. It is given by

$$P(S_k^+|S_k^-) = \binom{S_k^-}{S_k^- - S_k^+} p_{rel,k}^{(S_k^- - S_k^+)} (1 - p_{rel,k})^{S_k^+}, \tag{3.13}$$

if $S_k^+ \leq S_k^-$, and it is 0 otherwise. Before the k-th spike there are S_k^- release-competent sites which can independently release with probability $p_{rel,k}$. On the other hand, the number of release-competent sites cannot increase upon spike. The probability of release $p_{rel,k}$ can be computed recursively from equations 3.4 and 3.5

$$p_{rel,k+1} = p_0 + p_{rel,k} \cdot (1 - p_0) \cdot \exp\left(-\frac{\Delta_k}{\tau_F}\right) \tag{3.14}$$

where $\Delta_k \equiv t_{k+1} - t_k$ is the k-th interspike interval. The probability of observing a response R_k when the occupation state changes from S_k^- to S_k^+, $P(R_k|S_k^+, S_k^-)$, represents the quantal model part. According to the quantal model, each vesicle produces on average a response q (quantal size or unitary quantal response), and the effects of all released vesicles sum linearly. The unitary quantal response, however, exhibits some variability, which is quantified through its standard deviation, σ_q, called the quantal variability. Measurements of synaptic miniature events suggest that the distribution of the quantal response is not a Gaussian, but skewed to the right [Bekkers et al. (1990); Bhumbra and Beato (2013)]. That is, with a small probability a single quantum triggers a relatively large response. In the following, we thus assume that fluctuations around the unitary quantal response can be described by an Inverse Gaussian distribution. As a consequence, the post-synaptic response to the release of $S_k^- - S_k^+$ vesicles is also given by an Inverse Gaussian, with mean $(S_k^- - S_k^+) \cdot q$ and variance $(S_k^- - S_k^+) \cdot \sigma_q^2$. We can write:

$$P(R_k|S_k^+, S_k^-) = \frac{q^{\frac{3}{2}} \cdot (S_k^- - S_k^+)}{\sqrt{2\pi\sigma_q^2 R^3}} \exp\left\{-\frac{q \cdot \left[R_k - q \cdot (S_k^- - S_k^+)\right]^2}{2\sigma_q^2 R}\right\} \tag{3.15}$$

To fully describe the statistics of the post-synaptic responses as measured in the experiment, we have to introduce an *instrumental* Gaussian noise with variance σ_{noise}^2, independent of the number of vesicles released, which sums linearly to the quantal noise σ_q. The presence of such noise is evident from the recordings, and makes it difficult or, better, arbitrary to distinguish between failures and responses which are just small. Its introduction in the model allows one to deal with failures in a 'soft' way (i.e., what is the probability of this response being a failure), by avoiding the need of 'hard' classification (i.e., it is or not a failure) of small responses. Formally, the effect of instrumental noise can be expressed by a convolution of expression 3.15 with a Gaussian distribution with zero mean and variance σ_{noise}^2:

$$P(R_k|S_k^+, S_k^-, \sigma_{noise}^2) = \int_{-\infty}^{+\infty} P(x|S_k^+, S_k^-) \cdot P(R_k|x, \sigma_{noise}^2) dx. \tag{3.16}$$

In general, this expression has to be evaluated numerically. Although, for reasons of clarity, we will omit instrumental noise in the below derivations, it was included in all analysis of the experimental data.
Finally, the probability that the occupation state changes from S_k^+ to S_{k+1}^- during the k-th interspike interval, $P(S_{k+1}^-|S_k^+)$, is given by

$$P(S_{k+1}^-|S_k^+) = \binom{N - S_k^+}{S_{k+1}^- - S_k^+} l_k^{(S_{k+1}^- - S_k^+)} (1 - l_k)^{(N - S_{k+1}^-)} \tag{3.17}$$

if $S_{k+1}^- \geq S_k^+$, and it is 0 otherwise. After the k-th spike there are $N - S_k^+$ refractory sites which can independently become release-competent within the time interval Δ_k with probability $l_k = 1 - e^{-\Delta_k/\tau_D}$. On the other hand, the number of release-competent sites cannot decrease in between spikes.
In the case of more than one sequence (say, N_t sequences), the total likelihood is just the product of the likelihoods of the individual sequences:

$$P(R_{1\to M}^{1\to N_t}) = \prod_{i=1}^{N_t} P(R_{1\to M}^i). \tag{3.18}$$

3.5.3 Expectation-Maximization

A series of authors have developed EM methods designed for the estimation of quantal parameters only [Kullmann (1989); Stricker and Redman (1994)]. Here, we extend these algorithms significantly by including the dynamical STP parameters.
With the definition of $S_{1\to M}^-$ and $S_{1\to M}^+$ we can write:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{old}) = \sum_{S_{1\to M}^-} \sum_{S_{1\to M}^+} P(S_{1\to M}^-, S_{1\to M}^+ | R_{1\to M}, \boldsymbol{\theta}_{old}) \log \left[P(R_{1\to M}, S_{1\to M}^-, S_{1\to M}^+ | \boldsymbol{\theta}) \right] \tag{3.19}$$

The condition for the EM maximization step can now be written as:

$$\frac{\partial}{\partial \theta_j} Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{old}) = 0 \tag{3.20}$$

From this, we can derive re-estimation formulae for all continuous free parameters. These have to be solved at every step of the EM algorithm in function of the estimates at the previous step, $\boldsymbol{\theta}_{old}$. Using Equations 3.12, 3.13, 3.15 and 3.17 in equation 3.20 (and omitting the instrumental noise), we obtain the following re-estimation formulae (in an implicit form) for the model parameters

$$q_{new}: \quad \sum_{k=1}^{M} \left\langle R_k - q_{new} \cdot (S_k^- - S_k^+) \right\rangle = 0 \tag{3.21}$$

$$\sigma_{q,new}: \quad \sum_{k=1}^{M} \left\langle 1 - \frac{q_{new}}{\sigma_{q,new}^2 R_k} \cdot (R_k - q_{new} \cdot (S_k^- - S_k^+))^2 \right\rangle = 0 \tag{3.22}$$

$$\tau_{D,new}: \quad \sum_{k=1}^{M-1} \frac{\partial l_k}{\partial \tau_D} \left[\left\langle \frac{S_{k+1}^-}{l_k(1-l_k)} \right\rangle - \left\langle \frac{S_k^+}{l_k} \right\rangle - \frac{N}{1-l_k} \right] = 0 \tag{3.23}$$

$$U_{new}: \quad \sum_{k=1}^{M} \frac{\partial p_k}{\partial U} \left[\left\langle \frac{S_k^-}{p_k} \right\rangle - \left\langle \frac{S_k^+}{p_k(1-p_k)} \right\rangle \right] = 0 \tag{3.24}$$

$$\tau_{F,new}: \quad \sum_{k=1}^{M} \frac{\partial p_k}{\partial \tau_F} \left[\left\langle \frac{S_k^-}{p_k} \right\rangle - \left\langle \frac{S_k^+}{p_k(1-p_k)} \right\rangle \right] = 0. \tag{3.25}$$

Here, the brackets $\langle \cdot \rangle$ denote the average over all possible sequences of occupation states, weighted by the posterior distribution $P(S_{1\to M}^-, S_{1\to M}^+ | R_{1\to M}, \boldsymbol{\theta}_{old})$. This average has the following convenient property:

$$\begin{aligned} \left\langle g(S_k^-, S_k^+) \right\rangle &= \sum_{S_{1\to M}^-} \sum_{S_{1\to M}^+} g(S_k^-, S_k^+) \cdot P(S_{1\to M}^-, S_{1\to M}^+ | R_{1\to M}, \boldsymbol{\theta}_{old}) \\ &= \sum_{S_k^-, S_k^+} g(S_k^-, S_k^+) \cdot P(S_k^-, S_k^+ | R_{1\to M}, \boldsymbol{\theta}_{old}). \end{aligned} \tag{3.26}$$

This holds for any arbitrary function g.
Note that the conditions for U_{new} and $\tau_{F,new}$ involve p_k and its derivatives, which have to be calculated with the new estimates of the parameters. Thus, the estimates U_{new} and $\tau_{F,new}$ depend on each other and have to be found simultaneously, by numerical means. The same issue arises between q_{new} and $\sigma_{q,new}$ when instrumental noise is included.

3.5.4 Forward-Backward formalism

The computation of $P(R_{1\to M})$ by marginalising over $S_{1\to M}^-$ and $S_{1\to M}^+$ in Equation 3.12 is impractical. We develop for our case a forward-backward procedure in analogy with the one used with Hidden Markov Models [Rabiner (1989)]. We define two forward variables as

$$\alpha_k^-(S) = P(S_k^- = S, R_{1\to k-1}) \tag{3.27}$$

$$\alpha_k^+(S) = P(S_k^+ = S, R_{1\to k}) \tag{3.28}$$

These variables can be evaluated recursively as follows

$$\alpha_1^-(S) = P(S_1^- = S) \tag{3.29}$$

$$\alpha_k^+(S) = \sum_{S_k^-=0}^{N} \alpha_k^-(S) P(S_k^+ = S | S_k^-) P(R_k | S_k^+ = S, S_k^-) \tag{3.30}$$

$$\alpha_{k+1}^-(S) = \sum_{S_k^+=0}^{N} \alpha_k^+(S) P(S_{k+1}^- = S | S_k^+) \tag{3.31}$$

Note that $\alpha_M^+(S) = P(S_M^+ = S, R_{1\to M})$ and thus $\sum_{S=0}^N \alpha_M^+(S) = P(R_{1\to M})$. Similarly, we can define two backward variables as

$$\begin{aligned} \beta_k^-(S) &= P(R_{k\to M}|S_k^- = S) && (3.32)\\ \beta_k^+(S) &= P(R_{k+1\to M}|S_k^+ = S) && (3.33) \end{aligned}$$

that can be evaluated recursively as follows

$$\begin{aligned} \beta_1^+(S) &= 1 && (3.34)\\ \beta_k^-(S) &= \sum_{S_k^+=0}^{N} \beta_k^+(S) P(S_k^+|S_k^- = S) P(R_k|S_k^+, S_k^- = S) && (3.35)\\ \beta_{k-1}^+(S) &= \sum_{S_k^-=0}^{N} \beta_k^-(S) P(S_k^-|S_{k-1}^+ = S) && (3.36) \end{aligned}$$

and $\sum_{S=0}^N \beta_1^-(S) P(S_1^- = S) = P(R_{1\to M})$.
From this follows that the conditional distribution of the states in equation 3.26 can be easily computed from:

$$P\left(S_k^-, S_k^+ | R_{1\to M}, \boldsymbol{\theta}\right) = \frac{\beta_k^+(S_k^+) \cdot P(R_k|S_k^-, S_k^+) \cdot P(S_k^+|S_k^-) \cdot \alpha_k^-(S_k^-)}{P(R_{1\to M})}. \quad (3.37)$$

3.5.5 Fisher information matrix

In function of the stimulation protocol, the synaptic responses can be expected to carry more or less information about the model parameters. For instance, assume that one evokes a pre-synaptic spike train with inter-spike intervals that are much larger than the synaptic time constants. Clearly, the produced post-synaptic responses will not be able to resolve them. To quantify the amount of information a specific stimulus protocol gives us about a specific model, we compute the Fisher-Information Matrix (FIM) of the generative model:

$$\mathcal{I}(\boldsymbol{\theta})_{j,k} = E\left[\left(\frac{\partial}{\partial\theta_j}\log\left\{P(R_{1\to M}|t_{1\to M}, \boldsymbol{\theta})\right\}\right) \cdot \left(\frac{\partial}{\partial\theta_k}\log\left\{P(R_{1\to M}|t_{1\to M}, \boldsymbol{\theta})\right\}\right) |\boldsymbol{\theta}\right] \quad (3.38)$$

where we have re-introduced the likelihood's dependencies on $t_{1\to M}$ and $\boldsymbol{\theta}$. j and k are the indices of the model parameters and $E[\cdot]$ denotes the expectation value over the responses R. The diagonal elements of the inverse FIM are lower bounds on the variances of the parameter estimates:

$$Var(\theta_{jj}) \geq [\mathcal{I}(\boldsymbol{\theta})]_{jj}^{-1} \quad (3.39)$$

This relation is known as the Cramér-Rao bound [Radhakrishna Rao (1945); Cramér (1999)]. For a given model, the optimal stimulation protocol $t_{1\to M}^{opt}$ is the one that

minimises $\sum_j Var(\theta_{jj})$.
The calculation of the derivatives in equation 3.38 can be carried out efficiently by use of the forward-variables $\alpha^{\pm}$. We can write:

$$\frac{\partial}{\partial\theta_j}\log\left[P\left(R_{1\to M}|t_{1\to M},\boldsymbol{\theta}\right)\right] = \frac{1}{P\left(R_{1\to M}|t_{1\to M},\boldsymbol{\theta}\right)}\cdot\sum_{S=0}^{N}\frac{\partial\alpha_M^+(S)}{\partial\theta_j}. \tag{3.40}$$

Since α_k^+ depends on α_k^- and vice versa, we obtain recursive formulae for their respective derivatives:

$$\begin{aligned}\frac{\partial}{\partial\theta_j}\alpha_k^+(S) &= \sum_{S_k^-=0}^{N}\alpha_k^-(S_k^-)\cdot P(S_k^+=S|S_k^-)\cdot\left\{\frac{\partial P(R_k|S_k^-,S_k^+=S)}{\partial\theta_j}\right\}\\ &+\sum_{S_k^-=0}^{N}\alpha_k^-(S_k^-)\cdot\left\{\frac{P(S_k^+=S|S_k^-)}{\partial\theta_j}\right\}\cdot P(R_k|S_k^-,S_k^+=S)\\ &+\sum_{S_k^-=0}^{N}\left\{\frac{\partial\alpha_k^-(S_k^-)}{\partial\theta_j}\right\}\cdot P(S_k^+=S|S_k^-)\cdot P(R_k|S_k^-,S_k^+=S),\end{aligned} \tag{3.41}$$

$$\begin{aligned}\frac{\partial}{\partial\theta_j}\alpha_k^-(S) &= \sum_{S_{k-1}^+=0}^{N}\alpha_{k-1}^+(S_{k-1}^+)\cdot\left\{\frac{\partial P(S_k^-=S|S_{k-1}^+)}{\partial\theta_j}\right\}\\ &+\sum_{S_{k-1}^+=0}^{N}\left\{\frac{\partial\alpha_{k-1}^+(S_{k-1}^+)}{\partial\theta_j}\right\}\cdot P(S_k^-=S|S_{k-1}^+).\end{aligned} \tag{3.42}$$

The derivatives of equations 3.13, 3.15 and 3.17 appearing above are easily computed.

3.5.6 Preprocessing

Traces which featured a clear directed deviation from baseline during recording or which featured abrupt changes in the baseline were excluded by visual judgement and not used in the further analysis.
After subtraction of baseline the single noisy traces were smoothed by using a rectangular window of 2 ms size. From the average of the smoothed traces we determined each neuron's membrane time constant τ_m by fitting an exponential decay to the falling edge of the recovery response and, if possible, also to the first and averaged over these. Subsequently we deconvolved the smoothed single voltage traces $V(t)$ using the following relation [Richardson and Silberberg (2008)]:

$$R\cdot I(t) = \tau_m\cdot\frac{dV(t)}{dt} + V(t) \tag{3.43}$$

From this, we obtained the quantities $R\cdot I(t)$ which are identical to the synaptic currents $I(t)$ up to a proportionality factor R (input resistance). The $I(t)$ feature well separated

response peaks which we cut out in the following way: first, for a given data set, we determined the nominal positions of the response peaks from the average current trace. Around the expected positions, we cut out a certain time window of each single current trace (called 'crops' in the following). For input frequencies of less than 50 Hz the window extended from -10 ms to $+10$ ms relative to the expected position, at around 50 Hz from -8 ms to $+8$ ms and at around 70 Hz from -6 ms to $+6$ ms. Reconvolution of the crops yielded fully separated EPSPs whose peaks were obtained by searching for the maximum response in a time window of 6 ms length staring after the expected current peak. This window was adjusted by eye to compensate for variability in rise-times between different connections. After the position of the maximum was found, we averaged over a symmetric window of 1 ms size around this point. This value in turn was normalised by subtracting a baseline value obtained by averaging over a 5 ms window starting 0.5 ms after the crop's onset.

Finally, we estimated σ_{noise}, which is defined as the standard deviation of the fluctuations of 1 ms-window averages. We took for each crop of a data set the average of two subsequent 1 ms windows (inside the baseline interval used for normalisation), obtaining thus a distribution of $N_t \cdot M \cdot 2$ noise values per data set. σ_{noise} was then determined by computing the square-root of the distribution's standard deviation.

Bibliography

Abbott, L. and Regehr, W. G. (2004). Synaptic computation. *Nature*, 431(7010):796–803.

Bekkers, J., Richerson, G., and Stevens, C. (1990). Origin of variability in quantal size in cultured hippocampal neurons and hippocampal slices. *Proceedings of the National Academy of Sciences*, 87(14):5359–5362.

Bertram, R., Sherman, A., and Stanley, E. F. (1996). Single-domain/bound calcium hypothesis of transmitter release and facilitation. *Journal of Neurophysiology*, 75(5).

Bhumbra, G. S. and Beato, M. (2013). Reliable evaluation of the quantal determinants of synaptic efficacy using bayesian analysis. *Journal of neurophysiology*, 109(2):603–620.

Branco, T. and Staras, K. (2009). The probability of neurotransmitter release: variability and feedback control at single synapses. *Nature Reviews Neuroscience*, 10(5):373–383.

Buonomano, D. V. (2000). Decoding temporal information: a model based on short-term synaptic plasticity. *The Journal of Neuroscience*, 20(3):1129–1141.

Cramér, H. (1999). *Mathematical methods of statistics*, volume 9. Princeton university press.

Del Castillo, J. and Katz, B. (1954). Quantal components of the end-plate potential. *The Journal of physiology*, 124(3):560–573.

Dittman, J. S., Kreitzer, A. C., and Regehr, W. G. (2000). Interplay between facilitation, depression, and residual calcium at three presynaptic terminals. *The Journal of Neuroscience*, 20(4):1374–1385.

Dittman, J. S. and Regehr, W. G. (1998). Calcium dependence and recovery kinetics of presynaptic depression at the climbing fiber to purkinje cell synapse. *The Journal of neuroscience*, 18(16):6147–6162.

Faber, D. S., Young, W. S., Legendre, P., and Korn, H. (1992). Intrinsic quantal variability due to stochastic properties of receptor-transmitter interactions. *Science*, 258(5087):1494–1498.

Fioravante, D. and Regehr, W. G. (2011). Short-term forms of presynaptic plasticity. *Current opinion in neurobiology*, 21(2):269–274.

Franks, K. M., Stevens, C. F., and Sejnowski, T. J. (2003). Independent sources of quantal variability at single glutamatergic synapses. *The Journal of neuroscience*, 23(8):3186–3195.

Goldman, M. S., Maldonado, P., and Abbott, L. (2002). Redundancy reduction and sustained firing with stochastic depressing synapses. *The Journal of neuroscience*, 22(2):584–591.

Hallermann, S., Heckmann, M., and Kittel, R. J. (2010). Mechanisms of short-term plasticity at neuromuscular active zones of drosophila. *HFSP journal*, 4(2):72–84.

Hardingham, N. R., Read, J. C., Trevelyan, A. J., Nelson, J. C., Jack, J. J. B., and Bannister, N. J. (2010). Quantal analysis reveals a functional correlation between presynaptic and postsynaptic efficacy in excitatory connections from rat neocortex. *The Journal of Neuroscience*, 30(4):1441–1451.

Hempel, C. M., Hartman, K. H., Wang, X.-J., Turrigiano, G. G., and Nelson, S. B. (2000). Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex. *Journal of neurophysiology*, 83(5):3031–3041.

Hennig, M. H., Postlethwaite, M., Forsythe, I. D., and Graham, B. P. (2008). Interactions between multiple sources of short-term plasticity during evoked and spontaneous activity at the rat calyx of held. *The Journal of physiology*, 586(13):3129–3146.

Kandaswamy, U., Deng, P.-Y., Stevens, C. F., and Klyachko, V. A. (2010). The role of presynaptic dynamics in processing of natural spike trains in hippocampal synapses. *The Journal of Neuroscience*, 30(47):15904–15914.

Kullmann, D. (1989). Applications of the expectation-maximization algorithm to quantal analysis of postsynaptic potentials. *Journal of neuroscience methods*, 30(3):231–245.

Loebel, A., Silberberg, G., Helbig, D., Markram, H., Tsodyks, M., and Richardson, M. J. (2009). Multiquantal release underlies the distribution of synaptic efficacies in the neocortex. *Frontiers in computational neuroscience*, 3.

Markram, H., Lübke, J., Frotscher, M., Roth, A., and Sakmann, B. (1997). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *The Journal of physiology*, 500(Pt 2):409.

Markram, H., Wang, Y., and Tsodyks, M. (1998). Differential signaling via the same axon of neocortical pyramidal neurons. *Proceedings of the National Academy of Sciences*, 95(9):5323–5328.

Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, 319(5869):1543–1546.

Neher, E. and Sakaba, T. (2008). Multiple roles of calcium ions in the regulation of neurotransmitter release. *Neuron*, 59(6):861–872.

Pfister, J.-P., Dayan, P., and Lengyel, M. (2010). Synapses with short-term plasticity are optimal estimators of presynaptic membrane potentials. *Nature neuroscience*, 13(10):1271–1275.

Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Radhakrishna Rao, C. (1945). Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37(3):81–91.

Regehr, W. G. (2012). Short-term presynaptic plasticity. *Cold Spring Harbor perspectives in biology*, 4(7):a005702.

Ribrault, C., Sekimoto, K., and Triller, A. (2011). From the stochasticity of molecular processes to the variability of synaptic transmission. *Nature Reviews Neuroscience*, 12(7):375–387.

Richardson, M. J. and Silberberg, G. (2008). Measurement and analysis of postsynaptic potentials using a novel voltage-deconvolution method. *Journal of neurophysiology*, 99(2):1020–1031.

Rizzoli, S. O. and Betz, W. J. (2005). Synaptic vesicle pools. *Nature Reviews Neuroscience*, 6(1):57–69.

Rosenbaum, R., Rubin, J., and Doiron, B. (2012). Short term synaptic depression imposes a frequency dependent filter on synaptic information transfer. *PLoS computational biology*, 8(6):e1002557.

Scheuss, V. and Neher, E. (2001). Estimating synaptic parameters from mean, variance, and covariance in trains of synaptic responses. *Biophysical journal*, 81(4):1970–1989.

Silver, R. A. (2003). Estimation of nonuniform quantal parameters with multiple-probability fluctuation analysis: theory, application and limitations. *Journal of neuroscience methods*, 130(2):127–141.

Silver, R. A., Lübke, J., Sakmann, B., and Feldmeyer, D. (2003). High-probability uniquantal transmission at excitatory synapses in barrel cortex. *Science*, 302(5652):1981–1984.

Stricker, C. and Redman, S. (1994). Statistical models of synaptic transmission evaluated using the expectation-maximization algorithm. *Biophysical journal*, 67(2):656–670.

Thomson, A., Deuchars, J., and West, D. (1993). Single axon excitatory postsynaptic potentials in neocortical interneurons exhibit pronounced paired pulse facilitation. *Neuroscience*, 54(2):347–360.

Thomson, A. and West, D. (1993). Fluctuations in pyramid-pyramid excitatory postsynaptic potentials modified by presynaptic firing pattern and postsynaptic membrane potential using paired intracellular recordings in rat neocortex. *Neuroscience*, 54(2):329–346.

Thomson, A. M. and Deuchars, J. (1997). Synaptic interactions in neocortical local circuits: dual intracellular recordings in vitro. *Cerebral Cortex*, 7(6):510–522.

Trommershäuser, J., Schneggenburger, R., Zippelius, A., and Neher, E. (2003). Heterogeneous presynaptic release probabilities: functional relevance for short-term plasticity. *Biophysical journal*, 84(3):1563–1579.

Tsodyks, M. and Wu, S. (2013). Short-term synaptic plasticity. *Scholarpedia*, 8(10):3153.

Tsodyks, M. V. and Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the National Academy of Sciences*, 94(2):719–723.

Varela, J. A., Sen, K., Gibson, J., Fost, J., Abbott, L., and Nelson, S. B. (1997). A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *The Journal of neuroscience*, 17(20):7926–7940.

Wang, Y., Markram, H., Goodman, P. H., Berger, T. K., Ma, J., and Goldman-Rakic, P. S. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nature neuroscience*, 9(4):534–542.

Zucker, R. S. and Regehr, W. G. (2002). Short-term synaptic plasticity. *Annual review of physiology*, 64(1):355–405.

A Comment on the Synaptic Parameters

The work presented in the last chapter was initially motivated by the need to better quantify synaptic parameters. As we have seen, the model by Hansel and Mato (2013), based on the mechanism exposed in Mongillo et al. (2012), relies on relatively small values of the initial release probability, p_0. Recall that the original estimates of short-term plasticity properties in pre-frontal cortex in Wang et al. (2006), however, indicate that $p_0 = 0.15 - 0.35$. This lies outside the range where multi-stability can be achieved by the proposed mechanism.

We wanted to investigate the possibility that experimental estimates of synaptic parameters obtained with classical methods are not reliable enough to make conclusive statements about synaptic properties. As we have shown this is indeed the case. However, although parameters of single synaptic connections can vary substantially in function of the analysis method used (see figure 3.5), the global distribution of parameters we obtain is quite similar to what has been found before. Particularly with regard to p_0 we cannot observe significant differences with respect to the results of Wang et al. (2006).

Is the model Hansel and Mato (2013) thereby refuted? To answer this question, other factors have to be taken into account. It has, for instance, been pointed out by Borst (2010) that synapses studied *in vivo* exhibit a substantially lower release probability than synapses recorded *in vitro*. According to this reference, a whole series of factors may distort synaptic properties in experimental conditions, the most prominent being the different extracellular calcium concentrations in slices and in *in vivo*. Furthermore, changes in p_0 are accompanied by changes in the phenomenology of STP. Other factor like slice temperature have been shown to play an important role as well [Klyachko and Stevens (2006)].

All of this points to the very general issue whether available synaptic recordings provide STP parameter estimates that are meaningful enough for modelling work. This question is beyond the scope of our work. For us remains at this point only the somewhat sobering conclusion that we cannot make any definite statement about the validity of STP based models of balanced working memory.

The next chapter will therefore be dedicated to the development of an alternative theory where STP plays no role. Indeed, we will rely on static synapses only, but will make use of correlations between neuronal activity and synaptic connections.

Bibliography

Borst, J. G. G. (2010). The low synaptic release probability *in vivo*. *Trends in neurosciences*, 33(6):259–266.

Hansel, D. and Mato, G. (2013). Short-term plasticity explains irregular persistent activity in working memory tasks. *The Journal of Neuroscience*, 33(1):133–149.

Klyachko, V. A. and Stevens, C. F. (2006). Temperature-dependent shift of balance among the components of short-term plasticity in hippocampal synapses. *The Journal of neuroscience*, 26(26):6945–6957.

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

Wang, Y., Markram, H., Goodman, P. H., Berger, T. K., Ma, J., and Goldman-Rakic, P. S. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nature neuroscience*, 9(4):534–542.

Chapter 4

Multistability in the Balanced State

It is generally maintained that one of cortex' functions is the storage of a large number of memories [Goldman-Rakic (1987); Fuster (1995)]. In this picture, the physical substrate of memories is thought to be realised in the way cortical neural networks are structured. A common notion is, for instance, that non-random synaptic connectivity is the holder of the mnemonic information [Hebb (1968); Hopfield (1982)]. Memories then, are 'present' in the network independently from whether neurones are active or not. Neural activity itself is associated with the retrieval process: activation of different memories - their recall - is indicated by different patterns of neuronal activity. The presence of stored memories becomes apparent in their shaping of the possible activity states of the network.
A standard way to model memory storage and retrieval mathematically is by making use of attractors [see e.g. Amari (1977); Hopfield (1982); Amit (1992)]. The notion of 'attractor' refers to the portion of phase-space to which a dynamical system converges in the course of time. In this picture, memories correspond to attractors in the space of neuronal activity. During activation - retrieval - of a specific memory, the activity of all neurones converges to the pattern that represents that memory. The attractor view elegantly provides another important property of memory: its auto-associative nature. When a memory is only partially hinted at, for instance one sees an item that vaguely resembles a memorised one, eventually this latter one is recalled or 'comes to mind'. Analogously, setting neural activity sufficiently close to an attractor will cause it to converge to this very attractor.
Electrical activity in cortical neurones *in vivo* exhibits prominent temporal irregularity [Softky and Koch (1993); Bair et al. (1994); Shinomoto et al. (2009)]. A standard way to account for this phenomenon is to postulate that recurrent synaptic excitation and inhibition as well as external inputs are balanced [Shadlen and Newsome (1994); van Vreeswijk and Sompolinsky (1996, 1998); Shadlen and Newsome (1998)]. It can be shown that, when neurones receive relatively strong synaptic inputs, recurrent networks adjust automatically to a working point where inhibitory and excitatory currents approximately cancel. In this balanced state, neurones can potentially operate near threshold, where

their spiking is driven by fluctuations around the average input, causing aforementioned temporal irregularity. The network's self-regulation occurs dynamically, and under very general conditions; no fine-tuning is required.
Balanced networks have been widely studied and employed to explain a broad range of experimental observations beyond temporal irregularity, like persistent delay activity related to working memory [e.g. van Vreeswijk and Sompolinsky (2005); Hansel and Mato (2013)] and the emergence of selectivity from random connectivity [Hansel and van Vreeswijk (2012)]. However, in the common view, balanced networks do not easily support the coexistence of different activity states, as the dynamics of these networks tend to linearise the relationship between external stimuli and the neuronal response on the population level. This is problematic from the perspective of a memory framework based on attractors, as networks that store multiple memories need to feature multistability among attractors.
In this work we set out to show that the common belief that balanced networks are inadequate for memory storage is erroneous. In fact, we highlight that it is not necessary to invoke neuronal or synaptic non-linearities to create multistability, but that simple learning of synaptic weights as in perceptrons suffices. Quite on the contrary to the common view then, we demonstrate that the balanced state is necessary for extensive pattern storage in neural networks. We show further that by demanding that pattern storage is optimal, a series of experimentally observed properties of neural networks can be predicted.

4.1 Extensive memory storage requires balances

To demonstrate the new functional role of the balanced state, we begin by explicitly stating the experimental findings we will make use of and the further assumptions we make about the nature of cortical activity.

- At any time, each neurone receives a large number of excitatory and inhibitory inputs [Matsumura et al. (1988); Destexhe et al. (2003); Shu et al. (2003); Haider et al. (2006)]. We assume that these inputs sum linearly.
- Cortical activity is asynchronous [see e.g. Destexhe et al. (2003)].
- An activity pattern is represented by the ensemble of all neuronal firing rates in the network. We assume that it is sufficient to consider neurones as rate-units.
- Growing experimental evidence supports the idea that the global operating state of cortical areas does change only little, if at all, in function of the behavioural task (in awake animals) [see the reviews in Wohrer et al. (2012); Buzsáki and Mizuseki (2014)]. We assume therefore that different network states share the same global statistics. As a consequence, we consider also retrieval activity of different memory patterns being identically distributed.
- For simplicity, we assume that all firing-rate patterns are uncorrelated.

Let us consider a very simple neural network: a single inhibitory population of N neurones that are recurrently and (potentially) all-to-all connected. Apart from the same external excitatory input h_{ext}, all neurones receive inhibitory currents which are proportional to the respective firing-rates of their pre-synaptic partners in the local population. A current elicited in neurone i by pre-synaptic neurone j, can be written as the firing-rate ν_j of neurone j, weighted by the synaptic weight between the neurones J_{ij}. The dynamics of the total input h_i seen by to neurone i can then be written:

$$\tau_m \dot{h}_i = -h_i + \left[h_{ext} - \sum_{j=1}^{N} J_{ij}\nu_j \right]. \tag{4.1}$$

Here, τ_m is the neuronal membrane time constant, which, for simplicity, we consider to be identical across the population. The relationship between inputs and firing-rates in governed by the neuronal transduction function ϕ:

$$\nu_i = \phi(h_i). \tag{4.2}$$

At this point, we do not need to specify ϕ further apart from demanding that it should be a monotonically increasing function and lower bounded by zero, as firing-rates cannot be negative.

According to our assumptions, we consider a pattern to be given by the entirety of all firing rates in the network. The necessary condition that a given activity pattern is indeed stored in the network is that this pattern is a fixed point of the network's dynamics. This is the basic idea underlying attractor models of memory. In formal term this condition can be written by setting $\dot{h}_i = 0$ in the above dynamical equation. We obtain:

$$\begin{aligned} h_i &= h_{ext} - \sum_{j=1}^{N} J_{ij}\nu_j \\ &= h_{ext} - \sum_{j=1}^{N} J_{ij}\phi(h_j), \end{aligned} \tag{4.3}$$

that is, at the fixed point the inputs produce just the right firing-rates to sustain themselves and the system does not move from the state in which it is.

The above condition is straightforwardly extensible to the storage of more than one pattern. Suppose P firing-rate patterns are to be stored. We introduce the index $\mu = 1, .., P$ to distinguish between them. We can write:

$$h_i^\mu = h_{ext} - \sum_{j=1}^{N} J_{ij}\nu_j^\mu. \tag{4.4}$$

In order to satisfy these P conditions we have to adjust the J_{ij}; that is, the synaptic weights have to be learnt. It seems natural to expect that as N gets larger the number of patterns P that the network can store should, in principle, grow: increasing N, and

thus the number of pre-synaptic inputs, more adjustable parameters become available. If P indeed grows with N, pattern storage of the network would be extensive: the number of memories the network can accommodate grows with the network's size. However, a further qualification on the J_{ij} has to be made.
To elucidate this, note that, regardless of how exactly the weights are learnt, it is crucial that the neurones must be able to distinguish between different patterns. Thus, if we consider two pattern μ and ξ, the corresponding h_i^μ and h_i^ξ must be sufficiently distinct. To investigate this quantitatively, it suffices to consider a single neurone. Dropping the index i in equation 4.4, the difference between two inputs is:

$$h^\mu - h^\xi = \sum_{j=1}^{N} J_j \left(\nu_j^\xi - \nu_j^\mu \right). \tag{4.5}$$

Recall the assumption that all patterns obey the same statistics and thereby have identical average firing-rate and average input. We can therefore use the expressions $h^\mu = \bar{h} + \delta h^\mu$ and $\nu_j^\mu = \bar{\nu} + \delta\nu_j^\mu$, where the overlined terms represent the averages and the δ-terms denote fluctuations around these averages. We obtain:

$$\delta h^\mu - \delta h^\xi = \sum_{j=1}^{N} J_j \left(\delta\nu_j^\xi - \delta\nu_j^\mu \right), \tag{4.6}$$

and can observe that the average quantities drop out. This equation tells us that the information about the identity of a pattern available to a single neurone resides in fluctuations around the average synaptic input. In order to obtain the typical difference between two patterns we square both sides of equation 4.6 and average over patterns. If we set aside for a moment possible correlations between the J_j and the ν_j^μ, this yields:

$$\begin{aligned}
\mathrm{E}\left[\left(\delta h^\mu - \delta h^\xi \right)^2 \right] &= \mathrm{E}\left[\left(\sum_{j=1}^{N} J_j \left(\delta\nu_j^\xi - \delta\nu_j^\mu \right) \right)^2 \right] \\
\Leftrightarrow 2\sigma_h^2 &= \sum_{j=1}^{N} J_j^2 \, \mathrm{E}\left[\left(\delta\nu_j^\xi - \delta\nu_j^\mu \right)^2 \right] + \sum_{j \neq k}^{N} J_j J_k \, \mathrm{E}\left[\left(\delta\nu_j^\xi - \delta\nu_j^\mu \right) \left(\delta\nu_k^\xi - \delta\nu_k^\mu \right) \right] \\
\Leftrightarrow 2\sigma_h^2 &= 2\sigma_\nu^2 \sum_{j=1}^{N} J_j^2
\end{aligned} \tag{4.7}$$

where we have used the fact that correlations among neurons for each pattern are zero. σ_h^2 and σ_ν^2 denote the variances of inputs and firing-rates, respectively. In our theory, these quantities should be finite, as they are observable properties of the system we want to model and their magnitude should not depend on the number of neurones in the network. Note, however, that the way in which equation 4.7 is written, poses a problem in this respect: we have a sum of N terms on the right-hand side. To compensate, we clearly need the J_j to scale with N in some way, in the sense that their magnitude has

to be reduced as N grows.
The obvious choice here is to scale the synaptic weights as $\frac{1}{\sqrt{N}}$, as it eliminates all dependence on N. With this scaling, finite differences between firing-rate patterns lead to finite differences in the synaptic inputs. If we were to adopt for instance a scaling $\frac{1}{N}$, all patterns would look increasingly uniform to the single neurone with growing N.
The $\frac{1}{\sqrt{N}}$-scaling thus enables the single neurone to distinguish between patterns (or network-states) independent of the network size. However, this seems to come at a cost as now the average input to each neurone appears to grow with N. This can be seen by writing equation 4.4 with all synapses scaled accordingly:

$$h_i^\mu = \sqrt{N} h_{ext} - \sum_{j=1}^{N} \frac{J_{ij}}{\sqrt{N}} \nu_j^\mu. \tag{4.8}$$

At this, the factor $\sqrt{N}$ in front of h_{ext} reflects the fact that the synapses delivering the external input into the network are scaled in the same way as the recurrent synapses. The average input that each neurone sees is:

$$\overline{h} = \mathrm{E}[h_i^\mu] = \sqrt{N} \left(h_{ext} - \overline{J} \cdot \overline{\nu} \right), \tag{4.9}$$

Naively, the above equations tell us that the input's average is not independent of the network's size, but grows with N. From this it would seem that we have gained nothing with the $\frac{1}{\sqrt{N}}$-scaling: finite differences from pattern to pattern would still be vanishingly small compared to the mean. However, it has been shown in two important papers by Van Vreeswijk and Sompolinsky [van Vreeswijk and Sompolinsky (1996, 1998)] that the dynamics of a network as in equation 4.8 are guaranteed to produce a cancellation between recurrent inhibition and external input to leading order. This balance occurs automatically, without the need to fine-tune parameters. In our case, terms of order $\sqrt{N}$ in equation 4.9 sum up to zero, and $\overline{h}$ is given by contributions of order unity. The effect of the balance can be expressed by:

$$h_{ext} - \overline{J} \cdot \overline{\nu} = 0. \tag{4.10}$$

This equation tells us that (in the limit where N is very large) once the external input is fixed, the average firing rate of the activity patterns unambiguously determines the average synaptic strength in a network storing those patterns. We will make use of this relationship below.
Let us summarise the first important result of this section. We started by considering storage of identically distributed patterns in a neuronal network. The information about the network's state, or equivalently, the information about the activated pattern accessible to each neurone resides in the deviations from the average input. To be able to store an extensive number of patterns, that is, a number proportional to the network size, the size of these fluctuation has to be independent of N. This, in turn, requires a synaptic scaling that leads to the existence of a balanced state. We can conclude thus that extensive memory storage implies balance. In other words, balance is a necessary

condition for extensive memory storage.
The finding that both average input and input fluctuations remain finite in balanced networks, even in the limit of infinite N, is well established and was historically indeed the main motivation for the development of this theory. The novelty of our results is to show that this property is of exceeding importance in the context of memory storage. In fact, the balanced state gains a purpose other than being a mechanistic explanation for irregular neuronal activity.
The model of memory storage we propose here is the following. The network's macro-state, determined by the dynamics of the balance, characterises the global firing-rate statistics and thereby defines the network's working point. The activity patterns themselves are represented by the various micro-states that the network can enter. We will show in the following that the network's working point determines the number of activity patterns that can be sustained. Remarkably, demanding that this number is optimal predicts crucial properties of neuronal activity and physiology that are consistent with experimental observations.

4.2 Critical capacity of one inhibitory population

In order to study the relationship between storage capacity and firing-rate statistics, let us consider the single inhibitory population from the previous section, which is indeed the simplest possible network of rate units that can exhibit a balanced state. Recall the condition for the storage of activity patterns:

$$h_i^\mu = \sqrt{N} h_{ext} - \sum_{j=1}^{N} \frac{J_{ij}}{\sqrt{N}} \nu_j^\mu, \tag{4.11}$$

with

$$\nu_i^\mu = \phi(h_i^\mu). \tag{4.12}$$

Let us add one additional constraint on the synaptic weights. It is well established that neurones obey a principle known as Dale's Law [Dale (1935)]. It states that all the synapses of a neurone release the same type of neurotransmitter. In our setting this translates to the requirement that all synaptic connections of our inhibitory neurones have the same sign. We thus demand that the J_{ij} are non-negative. Furthermore, we ban autapses, that is we set $J_{ii} = 0$.
We have seen in the previous section that, if we fix the external inputs h_{ext}, adjusting the synaptic weights J_{ij} is the only way by which equation 4.11 can be satisfied for multiple patterns. Here, we want to ask how many activity patterns can be stored, and on which network parameters this number depends.
Let us, following the classical work in statistical physics, define the maximal number of pattern that can be learnt without errors (in the limit of infinite N) as P_c, where the subscript 'c' stands for 'critical'. The critical storage capacity is then given by P_c

normalised by the network size, N:

$$\alpha_c = \frac{P_c}{N}. \tag{4.13}$$

If we can store only a finite number of patterns in the limit of infinite N, the critical capacity will equal zero. If, however, storage capacity is extensive, and P_c grows, say, linearly with N, α_c is a finite quantity[1]. At any capacity level below α_c, that is, when a number $P < P_c$ of patterns is learnt, many solutions to the synaptic weight learning problem exist since condition 4.11 is underdetermined. Exactly at critical capacity, only one solution exists. These considerations constitute the basic principle of an elegant mathematical approach devised by Gardner [Gardner (1988)], with which - in the limit of infinite N - the volume of all possible J_{ij} that solve conditions like 4.11 can be calculated. P_c, and thus α_c, are determined by the value for which the volume collapses to a single point, which, in turn, depends on the statistics of the activity patterns one wishes to store.

Two crucial points about Gardner's approach are in particular worthy to be emphasised. First, this technique does not make any assumptions on the process that governs learning of the synaptic connections. Instead, it makes statements about the general properties of networks storing patterns in their synaptic structure. It offers thus a way to study the final outcome of learning, not learning itself. Second, central to Gardner's approach is the observation that, in the large N-limit, all neurones become statistically independent, making it sufficient to consider the weight-space of a single neurone.

Clopath and Brunel (2013) applied Gardner's calculation, which was originally conceived for binary units, to the case of a neurone that receives a continuous input signal (a firing-rate) and produces a continuous output signal. Furthermore, the weights in their calculation are required to be non-negative, respecting thereby Dale's Law. This corresponds exactly to the situation outlined here, and we can straightforwardly make use of their findings.

Before we state their results, however, we want to stress that some information about the synaptic weights is already provided by the balance equation 4.10. We see that the average synaptic strength is given by:

$$\overline{J} = \frac{h_{ext}}{\overline{\nu}}. \tag{4.14}$$

With this, the critical capacity from Clopath and Brunel (2013) reads:

$$\alpha_c = H(B), \tag{4.15}$$

where B is determined by the following equation:

$$\frac{B}{G(B) - B \cdot H(B)} = \frac{\sigma_h^2}{\sigma_\nu^2 \cdot \overline{J}^2}, \tag{4.16}$$

[1]The exotic cases that can arise in the theory of spin-glasses where P_c can grow exponentially with N can not arise in our theory.

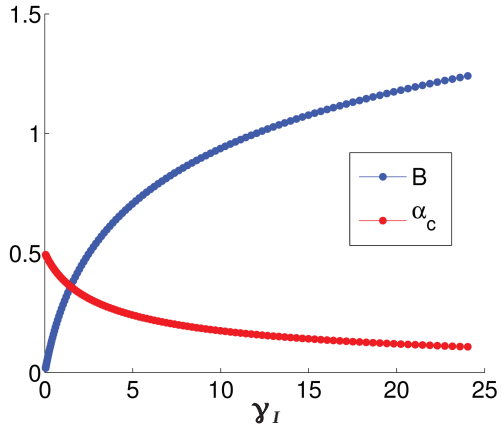

Figure 4.1: **The parameter B and critical capacity in function of γ_I**: The maximal value $\alpha_c = 0.5$ is reached when $\gamma_I = B = 0$.

and the functions $H(B)$ and $G(B)$ are defined as:

$$H(B) = \frac{1}{2}\left[1 - \text{erf}\left(-\frac{B}{\sqrt{2}}\right)\right], \qquad G(B) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{B^2}{2}\right). \tag{4.17}$$

Let us further write:

$$\gamma_I := \frac{\sigma_h^2}{\sigma_\nu^2 \cdot \overline{J}^2} = \frac{\sigma_h^2 \cdot \overline{\nu}^2}{\sigma_\nu^2 \cdot h_{ext}^2}. \tag{4.18}$$

We can see from equation 4.15 and the definition of $H(B)$ that the critical capacity becomes maximal when $B = 0$, yielding $\alpha_c = 0.5$. B itself depends on the parameters of the network and the statistics of activity patterns that appear on the right-hand side of equation 4.18. Clearly, $B = 0$ when $\gamma_I = 0$. Figure 4.1 shows a more detailed picture of the behaviour of α_c and B in function of γ_I. The most important point here is that α_c is a monotonically decreasing function of γ_I.
The novelty in our work is that we wish to interpret equations 4.15 - 4.18 in the context of a recurrent network. Clopath and Brunel (2013) considered a feed-forward structure (a Purkinje-cell receiving input from the granule-cell synapses) where the statistics of ν and h can be treated as independent. This is not true anymore in a recurrent network. By virtue of equation 4.12, we see that the statistics of inputs and firing rates are tightly linked by the neuronal transduction function. We can now proceed by asking how the properties of ϕ and the choice of the firing-rate statistics (or, equivalently, the input statistics) change the critical capacity of the network.

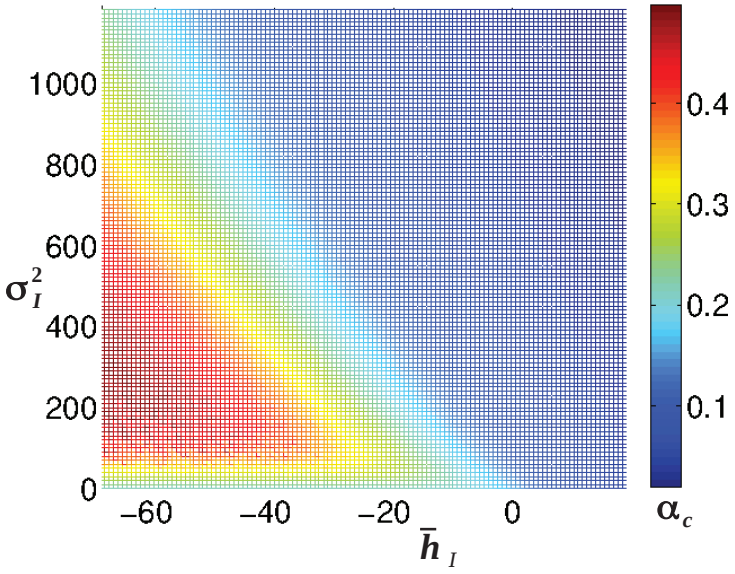

Figure 4.2: **Critical capacity in the one-population scenario.** α_c is given in function of the input-pattern statistics. Parameters are: $h_{ext} = 2$, $\beta = 5$, $v_0 = 4$

4.2.1 Critical capacity under the constraint of self-consistency

In order to get a grip on the self-consistent relationships between firing-rates, inputs and the neuronal transduction function, we need a starting point. Recall some of the assumptions made in section 4.1: each neurone receives a large number of synaptic inputs, its pre-synaptic weights are only weakly correlated and firing rates in the network are in general asynchronous. From this we should expect that the inputs to each neurone are distributed normally, both across patterns and across the population. There exists indeed good experimental support for this assumption [Destexhe et al. (2003); Carandini (2004)]. Therefore, instead of considering memory patterns defined in terms of firing rates, let us study the network's critical capacity for Gaussian input patterns as we vary the average input $\overline{h}$ and its variance σ_h^2.
Clearly, α_c will depend not only on $\overline{h}$ and σ_h^2 but also in a crucial manner on the neuronal transduction function. In the rest of this work we will use the following, biologically plausible transduction function:

$$\phi(h) = \nu_0 \cdot \log\left(1 + \exp\left(\frac{h}{\beta}\right)\right). \tag{4.19}$$

For small inputs, ϕ exhibits an exponential shape, while in the limit of large inputs it becomes linear (see figure 4.3, red curves). This behaviour mimics roughly the FI-curve of the integrate-and-fire neurone [Brunel and Sergi (1998); Roxin et al. (2011)].
In order to analyse the dependence of the critical capacity on the pattern statistics, we

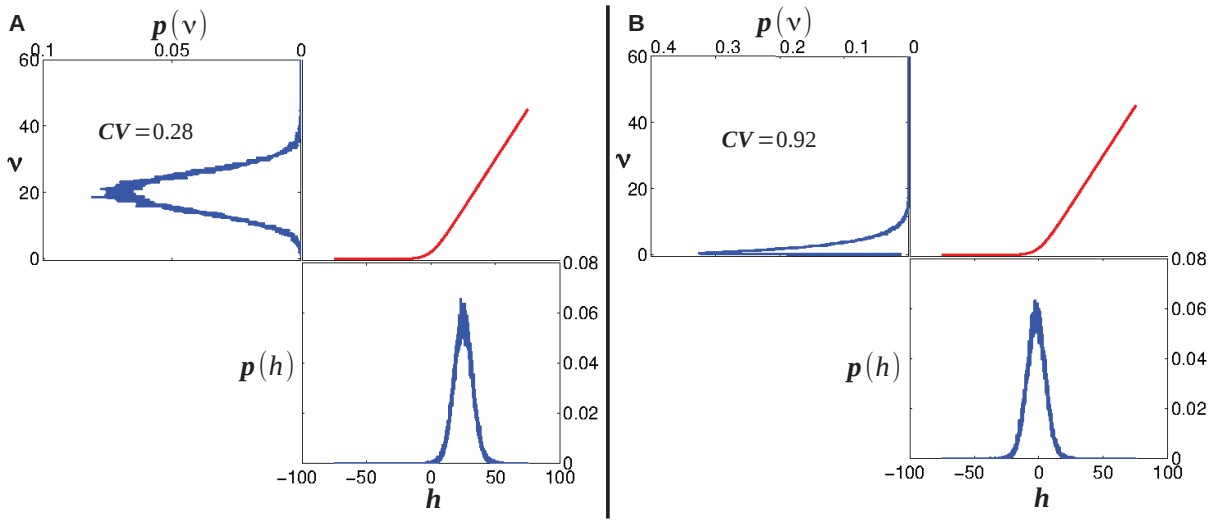

Figure 4.3: **Non-linear transformation of the input distribution.** Blue curves show histograms of inputs and firing-rates, red curve ϕ. **A)** Inputs located mostly in the linear part of ϕ, with $\overline{h} = 25$, $\sigma_h^2 = 50$, result in firing-rate distribution with $\overline{\nu} = 20$Hz, $\sigma_\nu^2 = 31\text{Hz}^2$. **B)** Inputs located mostly in the exponential part of ϕ, with $\overline{h} = -2$, $\sigma_h^2 = 50$, result in firing-rate distribution with $\overline{\nu} = 2.9$Hz, $\sigma_\nu^2 = 6.9\text{Hz}^2$. Parameters in all panels: $\beta = 5$, $\nu_0 = 4$Hz

calculate α_c in function of $\overline{h}$ and σ_h^2. At this, we fix $h_{ext} = 2$, $\beta = 5$ and $\nu_0 = 4$. The ingredient to calculate γ_I are obtained by generating normal input distributions for each combination of $\overline{h}$ and σ_h^2, passing them through ϕ and calculating the corresponding firing-rate statistics. Finally, α_c is obtained by solving equation 4.16 numerically. The result of this procedure is shown in figure 4.2, where α_c is displayed in colour-code on the $\overline{h}$-σ_h^2 plane. The optimal region can be found at negative $\overline{h}$ values; its size grows with decreasing $\overline{h}$.

To understand the preference for small $\overline{h}$ values, recall that α_c grows with decreasing γ_I. Fixing the external input h_{ext}, equation 4.18 reveals that the important issue for achieving high critical capacity is the following. For a given variance of the inputs h_i, the ratio $\frac{\overline{\nu}^2}{\sigma_\nu^2}$ should be as small as possible; or, stated otherwise, the firing-rate distribution's coefficient of variation should be as large as possible.

The relationship between σ_h^2 and $\frac{\overline{\nu}^2}{\sigma_\nu^2}$ depends on the properties of the transduction function. Figure 4.3 shows that input patterns that reside in the exponential part of ϕ - that is, patterns with small $\overline{h}$ - generate a firing rate distribution with higher CV than the same input patterns shifted to the linear part. For this reason, large values of α_c are found on the left hand side of figure 4.2.

Let us turn to the dependence of α_c on σ_h^2. It is instructive to consider two limiting cases. First, if $\overline{h}$ is sufficiently small, all h_i will settle in the exponential part of ϕ. We

can then write:

$$\begin{aligned}
\overline{\nu} &= \frac{1}{N}\sum_{i=1}^{N}\phi(h_i) \\
&\overset{N\to\infty}{=} \int_{-\infty}^{+\infty} D\eta\ \phi(\overline{h}+\sigma_h\cdot\eta) \\
&\approx \int_{-\infty}^{+\infty} D\eta\ \nu_0\cdot\exp\left(\left(\overline{h}+\sigma_h\cdot\eta\right)/\beta\right) \\
&= \nu_0\cdot\exp\left(\frac{\overline{h}}{\beta}+\frac{1}{2}\frac{\sigma_h^2}{\beta^2}\right)
\end{aligned} \tag{4.20}$$

and

$$\begin{aligned}
\overline{\nu^2} &= \int_{-\infty}^{+\infty} D\eta\ \phi^2(\overline{h}+\sigma_h\cdot\eta) \\
&\approx \nu_0^2\cdot\exp\left(2\frac{\overline{h}}{\beta}+2\frac{\sigma_h^2}{\beta^2}\right),
\end{aligned} \tag{4.21}$$

where $D\eta$ is the Gaussian measure. From this we can deduce

$$\sigma_\nu^2 = \overline{\nu^2}\cdot\left[\exp\left(\frac{\sigma_h^2}{\beta^2}\right)-1\right]. \tag{4.22}$$

which yields

$$\gamma_I = \frac{\sigma_h^2}{h_{ext}^2}\cdot\left[\exp\left(\frac{\sigma_h^2}{\beta^2}\right)-1\right]^{-1}. \tag{4.23}$$

Note that the ratio $\frac{\overline{\nu}^2}{\sigma_\nu^2}$ does not depend on $\overline{h}$ anymore. Thus, γ_I converges to a finite value for any given σ_h^2 as $\overline{h}$ becomes very negative. When $\sigma_h^2\to 0$, γ_I further simplifies:

$$\gamma_I = \frac{\beta^2}{h_{ext}^2}. \tag{4.24}$$

As long as the h_i stay in the exponential part of ϕ, α_c increases with σ_h^2, since γ_I is a decreasing function of the input variance. This trend inverts when the h_i reach the linear part of ϕ. At this point, $\overline{\nu}^2$ starts to grow faster than $\frac{\sigma_\nu^2}{\sigma_h^2}$ and γ_I is inflated. This behaviour can be recognised on the left part of figure 4.2. For each slice of small constant $\overline{h}$ we can see the non-monotonic progression of α_c.
The capacity's other limiting behaviour can be seen on the right part of figure 4.2, where α_c simply falls off with increasing σ_h^2. This can be understood from the fact that for rather large $\overline{h}$, the h_i settle in the linear part of ϕ. Here, the ratio $\frac{\overline{\nu}^2}{\sigma_\nu^2}$ remains constant for growing σ_h^2. In that limit, α_c thus becomes a strictly decreasing function of σ_h^2.

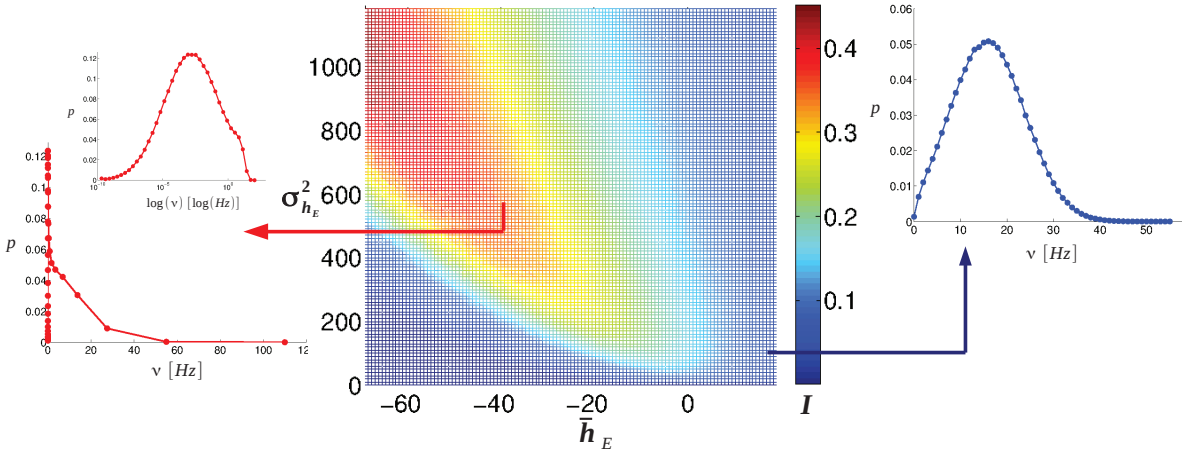

Figure 4.4: **Information measure of network performance. Middle**: I is given in function of the input-pattern statistics. **Left**: Exemplary firing-rate distribution for inputs in optimal region with $\overline{h} = -39$, $\sigma_h^2 = 600$; inset: semi-log plot of same distribution. **Right**: Exemplary firing-rate distribution for inputs in non-optimal region with $\overline{h} = 20$, $\sigma_h^2 = 100$. Parameters: $h_{ext} = 2\text{Hz}$,$\beta = 5$, $v_0 = 4$

Retrievability

So far we have considered the capability of the network to store patterns. However, patterns should also be reliably retrievable in order to assure efficient memory function. We have seen in the previous section that capacity is linked to the network's ability to distinguish between patterns. Retrievability, on the other hand, is associated with the network's ability to distinguish between patterns and noise; in other words, it depends on a signal-to-noise ratio.

It should be clear that if we assume the total inputs h_i to each neurone to be subjected to some form of additive noise σ_N^2, activity patterns with $\sigma_h^2 < \sigma_N^2$ should not be retrievable; the noise dominates the total input and different activity patterns become indistinguishable. This is true regardless of the value of α_c, indicating that α_c alone can not be a good measure of the network's performance in this limit. Indeed, this problem was already pointed out in Clopath and Brunel (2013) and remedied by introducing an information-theoretical measure. Likewise, let us quantify the network's performance by the following variable:

$$I = \alpha_c \log_2 \left(1 + \frac{\sigma_h^2}{\sigma_N^2}\right) /2. \tag{4.25}$$

where σ_N^2 is the variance of Gaussian white noise with zero mean that is added to the h. I represents the mutual information between firing-rates and inputs, multiplied by α_c; it grows with the number of storable patterns and the signal-to-noise ratio of the input.

A large source of noise in neuronal networks comes from temporal fluctuations in the

neuronal firing. Since our framework does not feature spiking neurones but rate units, there is no rigorous way to include this kind of noise. We can, however, make the reasonable assumption that the spike generation of each neurone obeys Poisson statistics. With this, we can roughly estimate that σ_N^2 should depend on the average firing rate of the population:

$$\sigma_N^2 \sim \overline{J^2} \cdot \overline{\nu}, \tag{4.26}$$

where $\overline{J^2}$ is the second moment of synaptic weights distribution. The middle panel of figure 4.4 shows I in function of $\overline{h}$ and σ_h^2. It can be seen that the optimal region is shifted to larger σ_h^2 values, as expected. Our previous finding, however, that the system's optimal performance is linked to an operating point in the expansive region of the transduction function is still valid.

To conclude, the first result of this section is that in order to maximise storage capacity, the network should be set up such that synaptic inputs to its neurones are matched to the expansive region of the neuronal transduction function. In fact, this indicates a functional role of the neuronal non-linearity. Moreover, recall our assumption that ϕ approximates the IF neurone well. The expansive regime in the IF transduction function corresponds to an operating regime where the neurone is close to threshold and spiking is fluctuation driven [Roxin et al. (2011)]. This indicates that the irregularity of neuronal spiking in cortex may be linked to the requirement of optimising capacity. As we have seen, this remains true even if we include the deleterious effect of this irregularity.

Another interesting observation to be made concerns the shape of the firing rate distributions, which is tightly linked to the neuronal non-linearity just mentioned. As critical capacity grows to the extent to which the inputs sit in the expansive region of the transduction function, large values of α_c imply that firing-rate distributions should be right-skewed. In the concrete case discussed here, ϕ provides an exponential non-linearity for small inputs, and firing-rate distributions become thus increasingly log-normal as α_c or I increase. The observation that firing-rate distributions should be skewed makes sense intuitively, since high capacity requires distributions with large CV but firing rates cannot become negative. The left and right panels of figure 4.4 illustrate this relationship. As can be seen, the firing-rate distribution in the optimal region is indeed strongly skewed and close to log-normal, with some small deviations at high firing rates. We come to our second conclusion that the skewness of firing rate distributions that indeed seem to be a ubiquitous feature of cortical activity [Song et al. (2005); Shafi et al. (2007); Hromádka et al. (2008); O'Connor et al. (2010); Wohrer et al. (2012); Buzsáki and Mizuseki (2014)] can be interpreted in our theory as signatures of a system that is optimised for memory storage and retrieval.

4.3 Two populations: excitation and inhibition

In this section we extend our consideration to the more realistic case of networks comprising two neuronal populations, one inhibitory, one excitatory. By making the conservative assumption that learning affects merely the excitatory to excitatory connections we set

out to study the effect of unstructured inhibition on the network's learning capabilities. Let the excitatory population comprise N_E neurones with firing rates ν_j^E and the excitatory population comprise N_I neurones with firing rates ν_j^I. The total inputs to excitatory and inhibitory neurones for different patterns μ can be written respectively as:

$$\begin{aligned} h_i^{E,\mu} &= \sqrt{N_{ext}}\tilde{h}_{ext}^E - \sum_{j=1}^{N_I} \frac{J_{ij}^{EI}}{\sqrt{N_I}}\nu_j^{I,\mu} + \sum_{j=1}^{N_E} \frac{J_{ij}^{EE}}{\sqrt{N_E}}\nu_j^{E,\mu} \\ h_i^{I,\mu} &= \sqrt{N_{ext}}\tilde{h}_{ext}^I - \sum_{j=1}^{N_I} \frac{J_{ij}^{II}}{\sqrt{N_I}}\nu_j^{I,\mu} + \sum_{j=1}^{N_E} \frac{J_{ij}^{IE}}{\sqrt{N_E}}\nu_j^{E,\mu}, \end{aligned} \tag{4.27}$$

where N_{ext} denotes the number of external input connections. The various J-symbols represent the four different types of synaptic weights. The relation between input and firing rate still remains

$$\nu_i^I = \phi(h_i^I), \qquad \nu_i^E = \phi(h_i^E). \tag{4.28}$$

Throughout the rest of this section, we use ϕ with the same parameters for both populations.

In the above equations we allow for the case where the numbers of external, recurrent inhibitory and recurrent excitatory synaptic inputs can be different. To clarify further derivations it is useful to express all these numbers in terms of N_E. Without loss of generality, we can absorb the fraction $\frac{N_{ext}}{N_E}$ by rewriting the external inputs. Additionally, by introducing $c = \frac{N_I}{N_E}$ we obtain:

$$h_i^{E,\mu} = \sqrt{N_E}h_{ext}^E - \sum_{j=1}^{N_I} \frac{J_{ij}^{EI}}{\sqrt{c \cdot N_E}}\nu_j^{I,\mu} + \sum_{j=1}^{N_E} \frac{J_{ij}^{EE}}{\sqrt{N_E}}\nu_j^{E,\mu} \tag{4.29}$$

$$h_i^{I,\mu} = \sqrt{N_E}h_{ext}^I - \sum_{j=1}^{N_I} \frac{J_{ij}^{II}}{\sqrt{c \cdot N_E}}\nu_j^{I,\mu} + \sum_{j=1}^{N_E} \frac{J_{ij}^{IE}}{\sqrt{N_E}}\nu_j^{E,\mu} \tag{4.30}$$

As in the previous section, we want to study the dependence of the critical capacity α_c on the statistics of the firing-rate and input distributions. For this, first note the following. As before, the average synaptic strength is given by requiring that equation 4.29 is balanced. We have:

$$\overline{J}_{EE} = \frac{h_{eff}}{\overline{\nu}_E} = \frac{\sqrt{c} \cdot \overline{J}_{EI} \cdot \overline{\nu}_I - h_{ext}^E}{\overline{\nu}_E}. \tag{4.31}$$

Note the important difference here, that the average excitatory-to-excitatory synaptic strength now also depends on the inhibitory feedback.

Furthermore, suppose we choose activity patterns only in the excitatory population with fixed $\overline{\nu}_E$ and $\sigma_{\nu_E}^2$. Given J_{ij}^{II} and J_{ij}^{IE}, the corresponding patterns in the inhibitory neurones are then unambiguously determined by equation 4.30. Since the synaptic input weights to the inhibitory population are not affected by learning, they are uncorrelated

with the firing-rates. Thus, we can also obtain average and variance of the inhibitory firing-rates directly from equation 4.30. In the limit $N_E \to \infty$, we get:

$$\overline{\nu_I} = \frac{h^I_{ext} + \overline{J}_{IE} \cdot \overline{\nu_E}}{\overline{J}_{II} \cdot \sqrt{c}}. \tag{4.32}$$

We see that equation 4.29 is different to its analogue, equation 4.11, in that the effective external input $\sqrt{N_E} h^E_{ext} - \sum_{j=1}^{N_I} \frac{J^{EI}_{ij}}{\sqrt{c \cdot N_E}} \nu^{I,\mu}_j$, against which the J^{EE}_{ij} have to be adjusted, changes from pattern to pattern. From this we can expect that inhibitory rates' quenched fluctuations will influence critical capacity.
Given also the statistics of J^{EI}_{ij}, we can use equation 4.29 to determine the space of the J^{EE}_{ij} weights that satisfy the conditions for the excitatory pattern's storage. The calculation, reported in detail in the appendix, is a generalisation of the one in Clopath and Brunel (2013). It yields:

$$\gamma_{EE} = \frac{\sigma^2_{eff}}{\sigma^2_{\nu_E} \cdot \overline{J}^2_{EE}} = \frac{\sigma^2_{eff} \cdot \overline{\nu}^2_E}{\sigma^2_{\nu_E} \cdot h^2_{eff}} \tag{4.33}$$

where, σ^2_{eff} is given by:

$$\sigma^2_{eff} \quad = \quad \sigma^2_{h_E} + \overline{J^2}_{EI} \cdot \sigma^2_{\nu_I}. \tag{4.34}$$

From the above equations, and the fact that the network has to be in the balanced state, we can readily derive two important constraints on the activity patterns. The first requirement is that $\overline{J}_{EE}$ has to be positive. From equation 4.31 we see that it is necessary that

$$\overline{\nu}_I > \frac{h^E_{ext}}{\overline{J}_{EI}} \cdot \sqrt{\frac{N_E}{N_I}}, \tag{4.35}$$

or, equivalently, by using equation 4.32

$$\overline{\nu}_E > h^E_{ext} \cdot \frac{\overline{J}_{II}}{\overline{J}_{EI} \cdot \overline{J}_{IE}} - \frac{h^I_{ext}}{\overline{J}_{IE}}. \tag{4.36}$$

To demand the positivity of the excitatory-to-excitatory connections thus sets a lower bound on the average firing rates. In addition, given the well known fact that $N_E > N_I$ [Gentet et al. (2010)], we see from here that it is quite natural that $\overline{\nu_I} > \overline{\nu_E}$.
The second constraint derives from the requirement of having a balanced state. Necessary conditions for this to happen are that the following inequalities are satisfied:

$$\frac{h^E_{ext}}{h^I_{ext}} > \frac{\overline{J}_{EI}}{\overline{J}_{II}} > \frac{\overline{J}_{EE}}{\overline{J}_{IE}} \tag{4.37}$$

However, we saw that in the limit $N_E \to \infty$, $\overline{J}_{EE}$ is given by equation 4.31. Rearranging this expression using equation 4.32 we can obtain:

$$\frac{\overline{J}_{EE}}{\overline{J}_{IE}} = \frac{\overline{J}_{EI}}{\overline{J}_{II}} \left(\frac{h^I_{ext}}{\overline{J}_{EI} \cdot \overline{\nu}_E} - \frac{h^E_{ext}}{\overline{\nu}_E} \cdot \frac{\overline{J}_{II}}{\overline{J}_{EI} \cdot \overline{J}_{IE}} + 1 \right). \tag{4.38}$$

The second inequality in 4.37 can only be violated when the term in the brackets is larger than unity. It can be easily seen that this implies $\frac{h^E_{ext}}{h^I_{ext}} < \frac{\overline{J}_{EI}}{\overline{J}_{II}}$, that is, an unbalanced solution is only possible when the network parameters reside already outside the allowed region. Thus, learning of activity patterns as it is proposed here cannot result in an unbalanced network state. Note that by setting $h^I_{ext} = 0$, that is imposing that the inhibitory population receives no input from outside the local network, the balance-constraint is automatically satisfied and the one from equation 4.36 is more likely to be satisfied.
The model outlined so far is a general description of a two population network that is able to learn memory patterns by adjustment of its excitatory-to-excitatory connections. However, the number of free parameters is quite big, as we need to specify the statistics of all synaptic populations (except the J^{EE}_{ij} of course). In order to achieve our primary goal - a better understanding of the role of inhibition - we thus consider a reduced model in the next section.

4.3.1 A reduced model

In general, the inhibitory firing-rate variance depends on the statistics of the excitation. This can be seen by calculating the variance of the input to the inhibitory neurones. Starting from equation in 4.30, we can write:

$$\begin{aligned} \sigma^2_{h_I} &= \mathrm{Var}\left(h^{I,\mu}_i\right) = \frac{1}{c \cdot N_E} \sum_{j=1}^{N_I} \mathrm{Var}\left(J^{II}_{ij} \nu^{I,\mu}_j\right) + \frac{1}{N_E} \sum_{j=1}^{N_E} \mathrm{Var}\left(J^{IE}_{ij} \nu^{E,\mu}_j\right) \\ &= \overline{J^2}_{II} \cdot \mathrm{Var}\left(\nu^{I,\mu}_j\right) + \overline{\nu}^2_I \cdot \mathrm{Var}\left(J^{II}_{ij}\right) + \overline{J^2}_{IE} \cdot \mathrm{Var}\left(\nu^{E,\mu}_j\right) + \overline{\nu}^2_E \cdot \mathrm{Var}\left(J^{IE}_{ij}\right). \end{aligned} \tag{4.39}$$

Here, we want to neglect the above equation and assume that we can adjust $\sigma^2_{\nu_I}$ independently so that we can study the effect of the inhibitory quenched noise on α_c explicitly. The requirement that the inhibitory population should be balanced is still valid, however, so that $\overline{\nu}_I$ remains determined by equation 4.32.
We start the derivation of a reduced model by expressing the effect of the recurrent inhibition purely in terms of an external current. This current is now not constant anymore, but will depend on the pattern-index. We can write:

$$h^{E,\mu}_i = \sqrt{N_E} \cdot h^\mu_{red} + \sum_{j=1}^{N_E} \frac{J^{EE}_{ij}}{\sqrt{N_E}} \nu^{E,\mu}_j, \tag{4.40}$$

with

$$h^\mu_{red} = h^E_{ext} - \sum_{j=1}^{N_I} \frac{J^{EI}_{ij}}{c \cdot N_E} \nu^{I,\mu}_j. \tag{4.41}$$

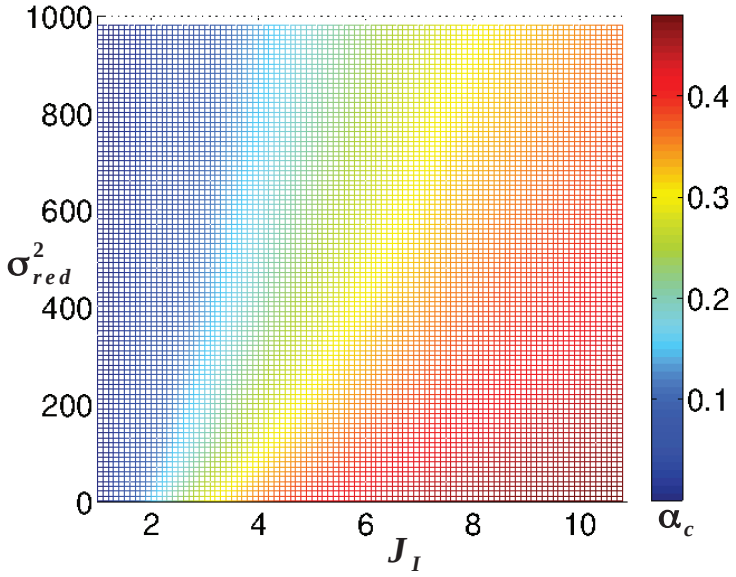

Figure 4.5: **Critical capacity in the reduced model** α_c is given in function of inhibitory feedback strength and inhibitory quenched noise. Parameters are: $\tilde{h}_{red} = 2$, $\beta = 5$, $v_0 = 4$

Since $\nu_j^{I,\mu}$ and J_{ij}^{EI} are in general uncorrelated, the distribution of h_{red}^{μ} across patterns is Gaussian. Its average is

$$\overline{h}_{red} := \mathrm{E}[h_{red}^{\mu}] = h_{ext}^{E} - \frac{\overline{J}_{EI}}{\overline{J}_{II}} \left(h_{ext}^{I} + \overline{J}_{IE} \cdot \overline{\nu}_E \right), \tag{4.42}$$

where we have used equation 4.32. The variance reads:

$$\frac{\sigma_{red}^2}{\sqrt{N_E}} := \mathrm{Var}[h_{red}^{\mu}] = \frac{\overline{J^2}_{EI}}{\sqrt{N_E}} \sigma_{\nu_I}^2. \tag{4.43}$$

Using a Gaussian random variable η^{μ} with zero mean and a variance of one, the total excitatory inputs of the reduced system can eventually be written as:

$$\begin{aligned} h_i^{E,\mu} &= \sqrt{N_E} \cdot \overline{h}_{red} + \sigma_{red} \cdot \eta^{\mu} + \sum_{j=1}^{N_E} \frac{J_{ij}^{EE}}{\sqrt{N_E}} \nu_j^{E,\mu} \\ &= \sqrt{N_E} \left(\tilde{h}_{red} - J_I \cdot \overline{\nu}_E \right) + \sigma_{red} \cdot \eta^{\mu} + \sum_{j=1}^{N_E} \frac{J_{ij}^{EE}}{\sqrt{N_E}} \nu_j^{E,\mu}. \end{aligned} \tag{4.44}$$

Again, by invoking the balance argument we can obtain the average synaptic weight:

$$\overline{J}_{EE} = \frac{J_I \cdot \overline{\nu}_E - \tilde{h}_{red}}{\overline{\nu}_E}. \tag{4.45}$$

The behaviour of the reduced network is governed by three effective parameters, one of which is σ^2_{red}, the inhibitory quenched noise variance. The other two are

$$\tilde{h}_{red} = h^E_{ext} - \frac{\overline{J}_{EI}}{\overline{J}_{II}} \cdot h^I_{ext}, \qquad J_I = \frac{\overline{J}_{EI} \cdot \overline{J}_{IE}}{\overline{J}_{II}}. \tag{4.46}$$

$\tilde{h}_{red}$ denotes the constant input seen by the network, independent of the average firing rate. In contrast, J_I represents the strength of the $\overline{\nu}_E$ dependent part of the inhibitory input; that is, it measures the effectiveness of inhibitory feedback.
It is straightforward to derive the expression for γ for this system:

$$\gamma_{red} = \frac{\sigma^2_h \cdot \overline{\nu}^2_E}{\sigma^2_{\nu_E} \cdot (J_I \cdot \overline{\nu}_E - \tilde{h}_{red})^2} + \frac{\sigma^2_{red} \cdot \overline{\nu}^2_E}{\sigma^2_{\nu_E} \cdot (J_I \cdot \overline{\nu}_E - \tilde{h}_{red})^2}. \tag{4.47}$$

The first term is similar to what we know from the case with one population. The second one quantifies the effect of quenched noise associated with the inhibitory population. Clearly, as γ_{red} grows with σ^2_{eff}, large inhibitory quenched noise is detrimental to capacity. Intuitively, the learning of each excitatory pattern now also has to compensate for the change in the inhibitory population. Thus the fraction of J_{EE} weight-space that has to be invested per pattern is larger and maximal storage capacity is reached for a smaller number of patterns.
By contrast, α_c increases with growing J_I. Recall that a large value of J_I augments the average excitatory-to-excitatory synaptic strength, $\overline{J}_{EE}$, thereby increasing the weight space. This implies that strong inhibitory feedback, for instance through high inhibitory firing rates, is beneficial to the network's memory storage ability. Analogously, decreasing $\tilde{h}_{red}$, has a similar impact on α_c, where, however, $\tilde{h}_{red} \geq 0$ must be obeyed.
In order to visualise the discussed effects we compute α_c in function of J_I and σ^2_{eff}. Figure 4.5 shows this for excitatory input patterns drawn from a Gaussian distribution with $\overline{h}_E = -7$ and $\sigma^2_{h_E} = 100$ and transduction function parameters set to $\beta = 5$ and $\nu_0 = 4$. Furthermore we kept $\tilde{h}_{red} = 2$. The global maximum of α_c over the shown parameters is clearly achieved for large J_I and small σ^2_{eff}.
The conclusion we can draw from our considerations is the following. In order to obtain high critical capacity in a two population network where only excitatory-to-excitatory connections can be learnt, inhibition has to have a strong, but uniform effect on the excitatory population. The fact that in cortex inhibitory firing rates are found to be larger than excitatory ones [Gentet et al. (2010); Beloozerova et al. (2003); Mitchell et al. (2007); Fujisawa et al. (2008)] is consistent with our optimality argument. However, variability associated with inhibition is generally large; the variance of inhibitory firing rates is much larger than that of excitation [see e.g. Gentet et al. (2010)] and inhibitory synaptic weights seem to be no less variable than excitatory ones [Chapeton et al. (2012); Levy and Reyes (2012); Avermann et al. (2012)]. This is not consistent with the prediction of our model. We thus may conclude that cortical memory storage does not only involve learning of the excitatory connections but relies significantly on inhibitory plasticity, whose importance is supported by broad evidence [see e.g. Dorrn et al. (2010); Kullmann et al. (2012)].

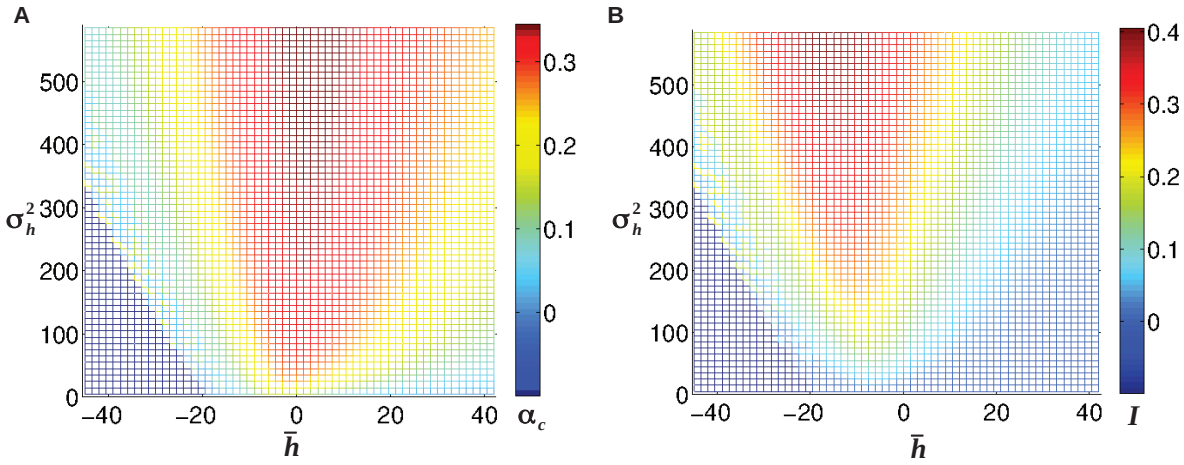

Figure 4.6: **Critical capacity and information measure for two populations with plastic E-to-E connections. A)**: α_c is given in function of the input-pattern statistics. **B)**: Same as A) for I. Parameters are: $h^E_{ext} = 0.2$, $h^I_{ext} = 0$, $\overline{J}_{EI} = 0.6$, $\overline{J}_{IE} = 0.8$, $\overline{J}_{II} = 0.2$, $\mathrm{Var}[J_{EI}] = \mathrm{Var}[J_{IE}] = \mathrm{Var}[J_{EE}] = 0.3$, $\beta = 5$, $\upsilon_0 = 4$Hz; dark blue regions in both plots indicate input region for which the network is not balanced.

4.3.2 The full system revisited

A crucial difference between the one population model in section 4.2 an the two population model still remains to be explored, namely the effect of the average firing rates on the critical capacity. In the case of one population, γ_I is very sensitive to changes in the average rate of the patterns. By contrast, ν_E^2 enters the numerator of γ_{EE} but as well the denominator. Naively, this suggests that critical capacity increases with growing $\overline{\nu}_E$. Furthermore, in the limit of very large $\overline{\nu}_E$, α_c should become independent of the average firing-rates. Indeed, this is exactly the behaviour of the reduced model from the previous section. However, deriving the reduced model we neglected the fact that the inhibitory quenched noise depends on the statistics of the excitatory patterns via equation 4.39.
In order to include this dependence we have to come back to the full system. Figure 4.6A shows α_c in function of $\overline{h}_E$ and $\sigma^2_{h_E}$. As before, inputs to the excitatory population are drawn from a Gaussian distribution with the respective parameters. The corresponding statistics for the inhibition were calculated self-consistently from equation 4.30. All parameters used for this figure are given in the caption. The main change with respect to the one-population network is the obvious shift of the optimal region (compare to figure 4.2). Instead of extending to very negative average inputs, the optimal region is now confined to intermediate values of $\overline{h}_E$. As can be seen, for every value of $\sigma^2_{h_E}$ exists an optimal $\overline{h}_E$. This behaviour can be understood best by separately considering the two contributions to γ_{EE}, namely the term proportional to $\sigma^2_{h_E}$ and the term proportional to

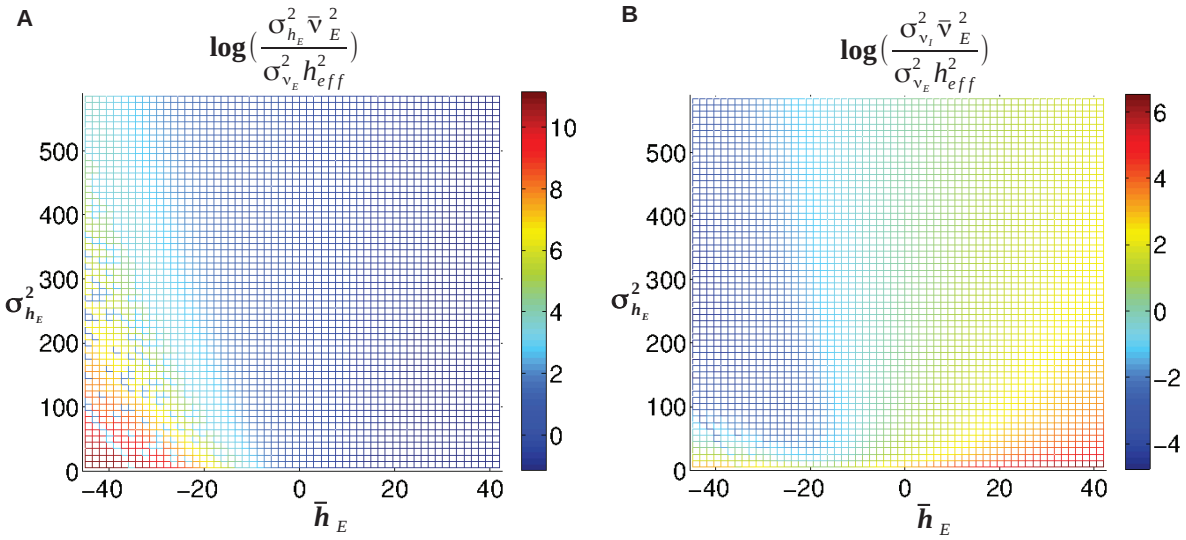

Figure 4.7: **Contributions of the two terms in equation 4.48. A)**: Term proportional to σ_h^2. **B)**: Term proportional to σ_{eff}^2. Parameters as in figure 4.6; note the log-scale.

σ_{eff}^2:

$$\gamma_{EE} = \frac{\sigma_h^2 \cdot \overline{\nu}_E^2}{\sigma_{\nu_E}^2 \cdot h_{eff}^2} + \frac{\sigma_{\nu_I}^2 \cdot \overline{J^2}_{IE} \cdot \overline{\nu}_E^2}{\sigma_{\nu_E}^2 \cdot h_{eff}^2}. \tag{4.48}$$

As we have seen above (section 4.2.1), the ratio $\frac{\overline{\nu}_E^2}{\sigma_{\nu_E}^2}$ becomes independent of the average input for very negative $\overline{h}_E$ as the inputs enter the exponential part of ϕ. In this limit, since $\frac{\overline{\nu}_E^2}{\sigma_{\nu_E}^2}$ is finite, the first term of γ_{EE} is dominated by the asymptotic behaviour of h_{eff}^2. As $\overline{h}_E$ and thus $\overline{\nu}_E$ decrease, h_{eff}^2 approaches zero and the first term of γ_{EE} diverges, as shown in figure 4.7, panel A.

By contrast, the second term does not diverge, as $\sigma_{\nu_I}^2$ goes to zero as well. Instead, it diverges in the opposite limit, that is, for large positive $\overline{h}_E$. Note that in this limit the neurones of both population operate in the linear part of ϕ and the variances of the firing rates depend linearly on the input firing rates. Hence, it can be straightforwardly seen from equation 4.39 that $\sigma_{\nu_I}^2$ (and $\sigma_{h_I}^2$) are a quadratic function of $\overline{\nu}_E$. This relationship underlies the divergence of γ_{EE} for large $\overline{h}_E$ shown in panel B of figure 4.7.

Finally, as in the single population model, the reduction of $\frac{\sigma_{h_E}^2}{\sigma_{\nu_E}^2}$ with growing $\sigma_{h_E}^2$ is responsible for the increase of α_c and thus the broadening of the optimal region in figure 4.6.

The retrievability of the two population network can be investigated similarly as in the previous case. As before, the strength of the signal is given by $\sigma_{h_E}^2$. However, the noise consists now of two contributions, corresponding to the temporal variances of both

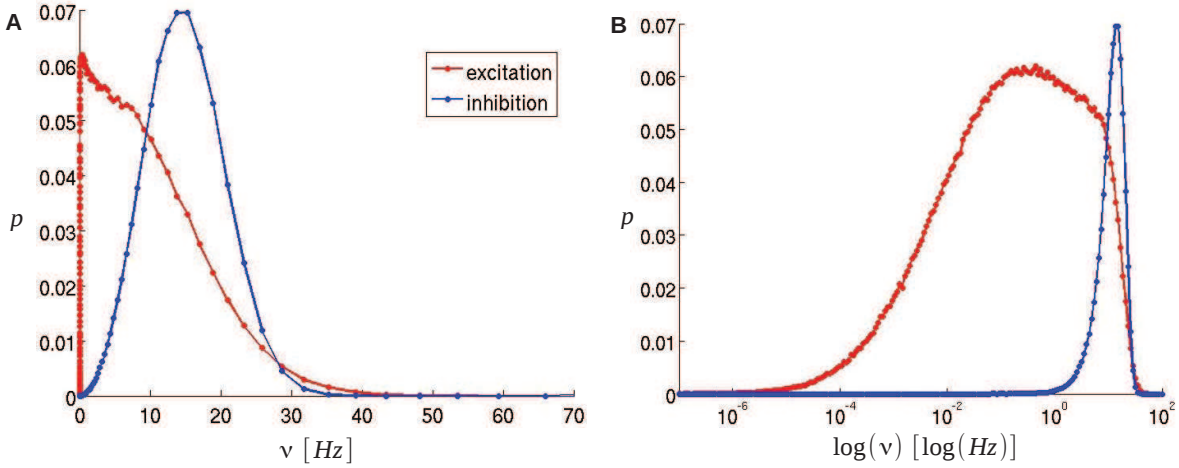

Figure 4.8: **Firing-rate distributions in the two population scenario. A)**: Exemplary firing-rate distributions of excitatory (red) and the inhibitory (blue) population for excitatory inputs in optimal region with $\overline{h}_E = -15$, $\sigma^2_{h_E} = 295$; corresponding inhibitory inputs are $\overline{h}_I = 14.9$, $\sigma^2_{h_I} = 62.0$; firing rates: $\overline{\nu}_E = 1.95$Hz, $\sigma^2_{\nu_E} = 16.45\text{Hz}^2$, $\overline{\nu}_I = 12.43$Hz, $\sigma^2_{\nu_I} = 32.29\text{Hz}^2$. **B)**: Semi-log plot of the same distributions. Parameters as in figure 4.6.

neuronal populations. We can write:

$$\sigma^2_N \sim \overline{J^2}_{EE} \cdot \overline{\nu}_E + \overline{J^2}_{EI} \cdot \overline{\nu}_I. \tag{4.49}$$

The quantity I calculated with the above σ^2_N is shown in figure 4.6B. Contrary to the single population case, the optimal region is shifted to smaller values of $\overline{h}_E$, that is towards the exponential part of ϕ. The reason for this effect lies in the contribution of the inhibition to σ^2_N, since large inhibitory firing rates create noise, but add nothing to the signal.

Concluding, we stress that given the large number adjustable free parameters, any analysis has to focus on the general, qualitative behaviour of the network and we must refrain from making quantitative statements. But whereas position and width of the optimal region may shift in function of the chosen parameters, the asymptotic behaviour exposed here holds in general. We can thus still make the following general remarks.

First, optimal storage capacity and the information measure I are subject to three constraints: $\overline{\nu}_E$ should not be too small, otherwise h_{eff} will diverge; $\sigma^2_{\nu_I}$ should not be too big; and the ratio $\frac{\sigma^2_{h_E}}{\sigma^2_{\nu_E}}$ should remain small. The first of these causes the biggest qualitative difference with respect to the single population model, where optimal critical capacity is achieved for very small inputs, that is, small average firing rates. By contrast, the two-population network prefers larger average firing rates. However, requiring that patterns are retrievable reduces this difference between the two cases.

Second, in the two population scenario, the distribution of inputs in the optimal re-

gion move closer to the linear part of the transduction function. This effect is small for the excitatory population, so that the excitatory firing-rate distribution remains highly skewed and almost log-normally distributed, with some deviations at high rates. For the inhibitory population, however, this shift is significant and skewness is highly reduced. Examples of firing-rate distributions from the optimal input region are shown in figure 4.8. We see that inhibitory activity is almost normally shaped. This phenomenon is in contrast to what is observed experimentally [see e.g. Buzsáki and Mizuseki (2014)]. As in the previous section we see that learning that is limited to the excitatory-to-excitatory synapses leads to predictions for properties of inhibitory neurones that are not consistent with experimental data. It is in fact intriguing that the population endowed with plasticity attains physiologically realistic properties through learning and the one without plasticity does not. The conclusion of this section is thus in line with the previously stated one: plastic inhibition seems to be a crucial ingredient in cortical learning.

4.4 Outlook: numerical simulations

We considered throughout this chapter the problem of learning a certain number of patterns in a recurrent neural network by adjusting the synaptic weights. Recall that the requirement for a solution to the learning problem, equation 4.11, demands that patterns are fixed-points of the network dynamics. Our theory makes strong statements about the existence of solutions to this problem, but not about the stability of these solutions. To test whether the memories we want to impose on the synaptic structure really lead to a network that is multi-stable we can resort to numerical simulations.
In this section we want to give a short overview over some interesting observation made when running numerical simulations of our attractor networks. We concentrate here on simulations of the single inhibitory population; the issues we want to point out are similar in the two population case. It is important to stress that the results described here need further thorough investigation and are thus rather preliminary.

Weight learning

The first step to be taken is the adjustment of the weight matrix such that equation 4.11 is satisfied. As we have no local learning rule, we have to make use of a global cost (or error) function which we can minimise. Let us denote the desired target input, that is, the input patterns we wish to learn, by $\tilde{h}_i^\mu$. Given any weight matrix we can then calculate the input patterns that the network actually will produce. The squared difference of target inputs $\tilde{h}_i^\mu$ and actual inputs h_i^μ defines an error in function of the matrix J as follows:

$$\varepsilon_i(J) = \frac{1}{2}\sum_\mu \left(\tilde{h}_i^\mu - h_i^\mu\right)^2 . \tag{4.50}$$

This error can be minimised via gradient descend methods. At this, the weights J_{ij} are updated repeatedly according to the following rule:

$$J_{ij} \leftarrow (J_{ij} - \eta \cdot \Delta J_{ij}) \cdot \theta\left(J_{ij} - \eta \cdot \Delta J_{ij}\right), \tag{4.51}$$

where η is the learning rate and $\theta(\cdot)$ denotes the Heaviside step-function. $\theta(\cdot)$ enforces that the synaptic weights remain non-negative. Furthermore we have:

$$\begin{aligned} \Delta J_{ij} &= \left[\bigtriangledown \varepsilon_i(J)\right]_j \\ &= \frac{1}{\sqrt{N}} \sum_{\mu} \left(\tilde{h}_i^{\mu} - h_i^{\mu}\right) \cdot \nu_j^{\mu}. \end{aligned} \tag{4.52}$$

The above equations represent the basic framework for weight learning. In our implementation, we additionally introduce an acceleration of the gradient method due to Nesterov [Nesterov et al. (2007)].
Apart from η, which has to be fixed empirically, there are two more parameters to be set for the learning: first, the error tolerance at which the weight updating is considered sufficient. In all examples shown in this section, we deemed learning sufficient when each neurone's individual error satisfied $\varepsilon_i < 10^{-6}$. Second, we have to define an initial distribution of the J_{ij}. Note that, while at a storage level that corresponds to α_c there is only one solution to the learning problem, at all levels below α_c the space of possible weights is finite. Therefore, more than one solution exists and we expect the initial weight distribution J_{ij}^{ini} to have an effect on the final weight distribution. However, we observed no significant effect of J_{ij}^{ini} on the attractor phenomenology. In the following, we choose J_{ij}^{ini} to be log-normally distributed with $\mathrm{E}[J_{ij}^{ini}] = \mathrm{Var}[J_{ij}^{ini}] = 1.0$.

Attractor phenomenology

In order to understand whether dynamically stable fixed points can be reliably learnt in our scheme, we generated and studied several instances of networks with different parameters, pattern statistics and sizes. Recall that the equations that govern the dynamics of the network (and that we used to perform the simulations) are:

$$\tau_m \dot{h}_i = -h_i + \left[\sqrt{N} h_{ext} - \sum_{j=1}^{N} \frac{J_{ij}}{\sqrt{N}} \nu_j\right], \tag{4.53}$$

$$\nu_i = \phi(h_i), \tag{4.54}$$

with

$$\phi(h_i) = \nu_0 \cdot \log\left(1 + \exp\left(\frac{h_i}{\beta}\right)\right). \tag{4.55}$$

We find that for a rather wide range of transduction function parameters and input pattern statistics the learnt attractors are indeed stable.
To demonstrate an example, we set up a network with $N = 2000$ neurones and the parameters $h_{ext} = 2\mathrm{Hz}$, $\beta = 5$, $v_0 = 4$ and $\tau_m = 0.02\mathrm{ms}$. We then performed learning of Gaussian input patterns with $\overline{h} = -13$, $\sigma_h^2 = 180$. The theoretical critical capacity of this network is $\alpha_c \approx 0.16$. As networks of finite size cannot reach capacity levels of α_c we limited ourselves to introduce merely $P = 100$ patterns, which corresponds to $\alpha = 0.05$.

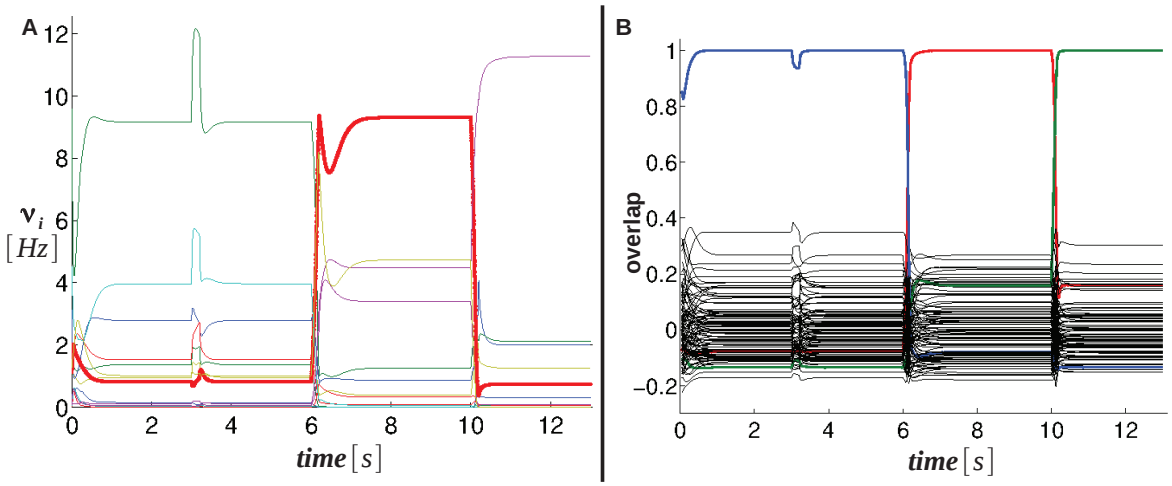

Figure 4.9: **Attractor dynamics in a simulated balanced network. A)**: Temporal development of 15 out of 2000 firing rates; for detailed description see the main text. **B)**: Temporal development of overlaps between target firing rate patterns and the networks state; the coloured lines represent overlaps with pattern states that are realised during the simulation. All input patterns were drawn from a Gaussian distribution with $\overline{h} = -13$, $\sigma_h^2 = 180$, resulting in firing rate patterns with $\overline{\nu} = 1.51$ Hz, $\sigma_\nu^2 = 8.72$ Hz^2; network parameters were: $h_{ext} = 2$Hz, $\beta = 5$, $v_0 = 4$Hz, $\tau_m = 0.02$ms

Figure 4.9A illustrates an exemplary simulation over several seconds. Firing rates of 15 randomly drawn neurones are shown. Initially, the inputs of 150 neurones are clamped for 200ms to values corresponding to attractor number 1; the rest of the network's rates and inputs are set the random initial values. This short clamping of less than 10% of the population suffices to drive the network into attractor state 1. Beginning at $t = 3$s the network is perturbed with global external input for 200ms. The fact that the network returns quickly to the attractor state after the perturbation further shows that this state is stable. Finally, at $t = 6$s and $t = 10$s the population receives external inputs that are proportional to the differences between its attractor states before and after the inputs. As can be seen, the network is able to switches reliably between states.
In order to check that the attractors we observe are indeed the states that were intended, we can calculate the overlap between the network's state at each point in time and the 100 patterns injected into the learning algorithm. At this, the overlap is defined as the correlation coefficient between network state and pattern. Panel B shows the overlap's temporal development for the same simulation as in panel A. The coloured lines indicate overlaps with the 3 attractor states that are effectively realised; these are indeed unity. Note that the relatively large overlap among patterns is a finite size effect; in the limit of large N we expect these values to approach zero.
We want to stress that the temporal sequence shown in 4.9A bears some resemblance to a classical working memory delayed response task, if we think of the phase between $t = 6$s and $t = 10$s as the delay period [see, for instance Fuster (1995); Funahashi et al. (1989)].

An experimenter who records the neurone indicated by the thick red line, would describe it as exhibiting persistence activity, coding for the memory in question. According to our theory however, *all* neurones in the population code for that memory. The emergence of a few prominent high rates originates from the strong skewness of the firing-rate distribution.
Although in most cases the learning of the synaptic weights produced the intended network behaviour as shown in the example above, we encountered cases where unexpected phenomena arose. We can distinguish between two cases. First, for certain parameter configurations, especially when the gain of the transduction function is too large, the network dynamics can become chaotic. It is not clear *a priori* whether it makes sense to consider pattern storage in this regime, that is, whether the chaotic nature of the dynamics can be consistent with multiple separate attractors. Thus chaos is potentially deleterious to pattern storage in the sense discussed throughout this chapter. As our theory for critical storage capacity cannot distinguish between chaotic and non-chaotic regimes we may have to introduce a further bound on α_c based on dynamical considerations.
The second issue concerns the appearance of stable network states that are different from those imposed during learning. Under some circumstances an additional stable state with equal average firing rate but smaller variance can arise alongside the stable pattern states. Interestingly, this low variance state has an intermediate overlap with all other states in the network. In other instances, the learnt patterns are not stable. In this case there appears a new stable attractor in the vicinity of each learnt pattern. These states are highly correlated with their learnt, unstable counterparts, but have higher variances. So far, it is not clear under which conditions these states appear. Interestingly, their occurrence is higher when we use a purely exponential transduction-function. This could indicate either an effect of very high rates or numerical problems. Without doubt, these issues have to be investigated thoroughly in future research.

Stability eigenvalues

A final intriguing observation we want to report is linked to the eigenvalue spectrum of the attractors. Given the weight-matrix J after learning, we can numerically calculate a stability matrix S^μ separately for each attractor state. It is given by:

$$S^\mu = I_{N\times N} + \frac{J}{\sqrt{N}} \cdot \phi'(h^{\mu *}), \tag{4.56}$$

where $I_{N\times N}$ denotes the identity matrix in N dimensions and $\phi'(h^{\mu *})$ is the first derivative of the transduction function, evaluated at the μth attractor state $h^{\mu *}$. The eigenvalues of S^μ provide us with information about the stability of pattern μ.
Figure 4.10 visualises the eigenvalues that are associated with the second attractor from figure 4.9 (between $t = 6$s and $t = 10$s). The phenomenology shown here is typical. First of all, note that the real parts of all eigenvalues are negative; the attractor is therefore stable. The very negative real eigenvalue in panel A is linked to the non-negativity of $S\mu$ [Frobenius (1912)], which can be seen from the fact that all synaptic weights are

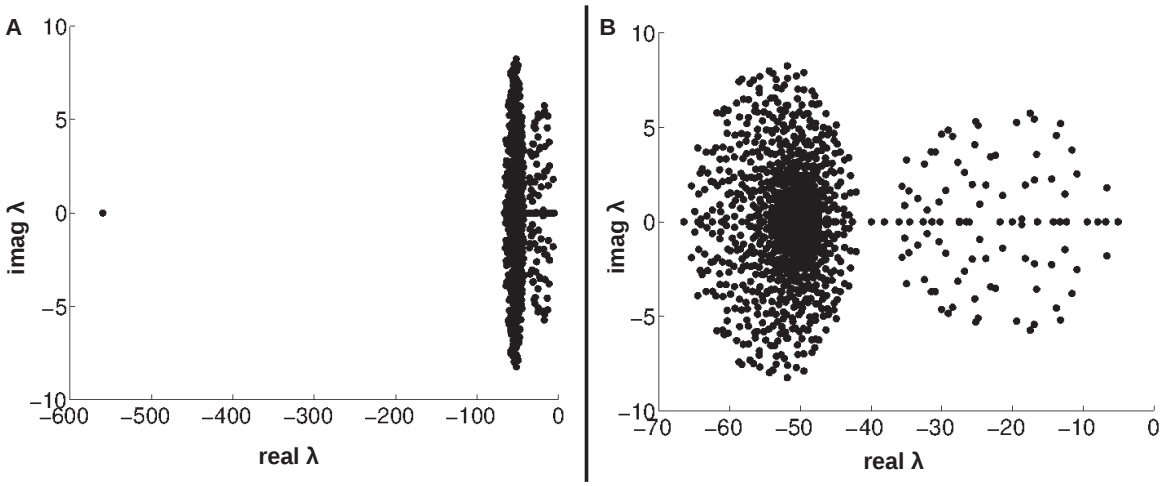

Figure 4.10: **Eigenvalues from stability analysis in the complex plane. A)**: Full view of the eigenvalues of the second attractor state from 4.9; all eigenvalues have strictly negative real parts. **B)**: Zoomed view; the disk on the right contains $P - 1 = 99$ eigenvalues. Same parameters as in figure 4.9.

non-negative and ϕ' is positive, since ϕ is a monotonically increasing function. Panel B is a zoomed version of panel A. Two disk-like structures can be spotted: the disk on the left stems from the random, unstructured part of the synaptic weight matrix J. It can be shown that the distribution of eigenvalues of a random matrix with independently and identically distributed entries uniformly covers a circular disk in the complex plane [Tao et al. (2010)]. The fact that the uniform distribution of eigenvalues inside this disk is distorted in our case can presumably be accounted for by correlations between the weights and the firing rates. The disk on the right features an interesting property: the number of eigenvalues in this structure corresponds to $P - 1$, that is 99 in the example at hand. This observation was reliably made in all investigated cases. We thus may conclude that this second disk is tied to the structured part of the weight-matrix.
Finally, we can report that we repeatedly observed the following relationship between the eigenvalue spectrum and the stability of the attractors states. On the one hand, it seems that an attractor's destabilisation in favour of low or high variance states is associated with one or a few eigenvalues from the second disk crossing the real axis. On the other hand, the transition to chaos appears to be governed by the crossing of eigenvalues from the first disk. We stress that these interpretations are not definitive and the mentioned phenomena need to be further investigated in a more systematic and extensive way.

4.5 Conclusion

In this chapter, we have derived three important results. First, starting from experimental findings and reasonable assumptions about cortical activity we showed that neural

networks that store an extensive number of memory patterns are necessarily balanced. The balanced state is needed to keep the quenched fluctuations in the average input, which are indicative of the network's state, finite. Second, requiring that memory storage capacity and retrievability are optimal sets balanced networks in a working point where they display a series of properties that are reminiscent of cortical activity. These include a highly skewed firing-rate distributions, a pronounced temporal spiking variability and generally low firing rates. Thus we suggest that these phenomena can be interpreted as signatures of networks that are optimised for memory function. Third, when we consider networks with both excitatory and inhibitory neurones, the requirement of optimality predicts features of inhibitory activity that are presumably incompatible with experimental findings. This suggests that modification of inhibitory connections is a crucial ingredient in cortical plasticity, as is indeed reported.

For a general discussion of these results the we refer to the next chapter.

Appendix

The volume of weight space

Let us consider the inputs to an excitatory neurone, given by equation 4.29. To make the notation handier, we suppress the 'E' index when writing the excitatory firing rates and inputs. Likewise, we write $N_E = N$ and $J^{EE} = J$. In the steady state, each excitatory pattern ν_j^μ has to satisfy:

$$h^\mu = \sqrt{N} h_{ext}^E - \sum_{j=1}^{N_I} \frac{J_j^{EI}}{\sqrt{c \cdot N}} \nu_j^{I,\mu} + \sum_{j=1}^{N} \frac{J_j}{\sqrt{N}} \nu_j^\mu. \tag{4.57}$$

For convenience we define:

$$z^\mu = h^\mu - \left[\sqrt{N} h_{ext}^E - \sum_{j=1}^{N_I} \frac{J_j^{EI}}{\sqrt{c \cdot N}} \nu_j^{I,\mu} + \sum_{j=1}^{N} \frac{J_j}{\sqrt{N}} \nu_j^\mu \right]. \tag{4.58}$$

The constraint that the network should contain P patterns can be written formally as:

$$\prod_{\mu=1}^{P} \delta(z^\mu) = 1. \tag{4.59}$$

In order to compute the average logarithm of the weight-space's volume, we make use of the replica method. We can write:

$$\begin{aligned}
\langle V^n \rangle &= \left\langle \int \left(\prod_{j,\alpha} dJ_j^\alpha \right) \prod_{\mu,\alpha} \delta(z_\alpha^\mu) \right\rangle \\
&= \left\langle \int \left(\prod_{j,\alpha} dJ_j^\alpha \right) \prod_{\mu,\alpha} \int_{-\infty}^{+\infty} \frac{dx_\alpha^\mu}{2\pi} \exp(i x_\alpha^\mu \cdot z_\alpha^\mu) \right\rangle \\
&= \int \left(\prod_{j,\alpha} dJ_j^\alpha \right) \prod_{\mu} \int_{-\infty}^{+\infty} \left(\prod_{\alpha} \frac{dx_\alpha^\mu}{2\pi} \right) \left\langle \exp(i \sum_\alpha x_\alpha^\mu \cdot z_\alpha^\mu) \right\rangle \qquad (4.60)
\end{aligned}$$

Next, we expand the exponential in the above equation. We will make use of the following order-parameters:

$$\frac{1}{N} \sum_j \left(J_j^\alpha \right)^2 = Q^\alpha \qquad (4.61)$$

$$\frac{1}{N} \sum_j J_j^\alpha J_j^\beta = q^{\alpha\beta} \qquad (4.62)$$

$$\frac{1}{N} \sum_j J_j^\alpha = \overline{J} + \frac{M^\alpha}{\sqrt{N}} = \frac{h_{eff}}{\overline{\nu}} + \frac{M^\alpha}{\sqrt{N}}, \qquad (4.63)$$

where, as before we have $h_{eff} = \sqrt{c} \cdot \overline{J}_{EI} \cdot \overline{\nu}_I - h_{ext}^E$. We start the expansion by writing:

$$\left\langle \exp(i \sum_\alpha x_\alpha^\mu \cdot z_\alpha^\mu) \right\rangle = 1 + i \sum_\alpha x_\alpha^\mu \cdot \langle z_\alpha^\mu \rangle + \frac{i^2}{2} \sum_{\alpha,\beta} x_\alpha^\mu x_\beta^\mu \cdot \left\langle z_\alpha^\mu z_\beta^\mu \right\rangle + ... \qquad (4.64)$$

Let us calculate the first two moments of z_α^μ separately:

$$\langle z_\alpha^\mu \rangle = \underbrace{\langle h^\mu \rangle}_{\overline{h}} - \underbrace{\langle \nu^\mu \rangle}_{\overline{\nu}} \cdot M^\alpha \qquad (4.65)$$

$$\left\langle z_\alpha^\mu z_\beta^\mu \right\rangle = \underbrace{\sigma_h^2 + \overline{J^2}_{EI} \cdot \sigma_{\nu_I}^2}_{\sigma_{eff}^2} + (\overline{h} - \overline{\nu} M^\alpha) \cdot (\overline{h} - \overline{\nu} M^\beta) + q^{\alpha\beta} \sigma_\nu^2 \qquad (4.66)$$

Thus the expansion of the exponential can be re-summed as:

$$\begin{aligned}
&\left\langle \exp(i \sum_\alpha x_\alpha^\mu \cdot z_\alpha^\mu) \right\rangle \\
&= 1 + i \sum_\alpha x_\alpha^\mu \underbrace{\left(\overline{h} - \overline{\nu} \cdot M^\alpha \right)}_{b^\alpha} - \frac{1}{2} \sum_{\alpha,\beta} x_\alpha^\mu x_\beta^\mu \left(\sigma_{eff}^2 + (\overline{h} - \overline{\nu} M^\alpha) \cdot (\overline{h} - \overline{\nu} M^\beta) + q^{\alpha\beta} \sigma_\nu^2 \right) + ... \\
&= 1 + i \sum_\alpha x_\alpha^\mu \cdot b^\alpha - \frac{1}{2} \sum_\alpha (x_\alpha^\mu)^2 \left((b^\alpha)^2 + \sigma_{eff}^2 + Q^\alpha \sigma_\nu^2 \right) - \frac{1}{2} \sum_{\alpha \neq \beta} x_\alpha^\mu x_\beta^\mu \left(b^\alpha b^\beta + \underbrace{\sigma_{eff}^2 + q^{\alpha\beta} \sigma_\nu^2}_{c^{\alpha\beta}} \right) + ... \\
&= \exp\left(i \sum_\alpha x_\alpha^\mu \cdot b^\alpha - \frac{1}{2} \sum_\alpha (x_\alpha^\mu)^2 \cdot (\sigma_{eff}^2 + Q^\alpha \cdot \sigma_\nu^2) - \frac{1}{2} \sum_{\alpha \neq \beta} x_\alpha^\mu x_\beta^\mu \cdot c^{\alpha\beta} \right) \qquad (4.67)
\end{aligned}$$

From this point on, the calculation is identical for the two cases of one population and two populations. We introduce conjugate momenta $\bar{M}^\alpha$, $\bar{Q}^\alpha$ and $\bar{q}^{\alpha\beta}$:

$$1 = \int \frac{dM^\alpha d\bar{M}^\alpha}{2\pi/\sqrt{N}} \exp\left(-i\bar{M}^\alpha\left[\sum_j J_j^\alpha - N\cdot\overline{J} - \sqrt{N}\cdot M^\alpha\right]\right) \tag{4.68}$$

$$1 = \int \frac{dQ^\alpha d\bar{Q}^\alpha}{2\pi/N} \exp\left(i\bar{Q}^\alpha\left[\sum_j (J_j^\alpha)^2 - N\cdot Q^\alpha\right]\right) \tag{4.69}$$

$$1 = \int \frac{dq^{\alpha\beta} d\bar{q}^{\alpha\beta}}{2\pi/N} \exp\left(i\bar{q}^{\alpha\beta}\left[\sum_j J_j^\alpha J_j^\beta - N\cdot q^{\alpha\beta}\right]\right) \tag{4.70}$$

Rearranging and factorising the products $\prod_{\mu,j}$ yields for $\langle V^n\rangle$:

$$\begin{aligned}
\langle V^n\rangle \quad &\propto \quad \prod_\alpha \int dM^\alpha d\bar{M}^\alpha dQ^\alpha d\bar{Q}^\alpha \prod_{\alpha<\beta}\int dq^{\alpha\beta}d\bar{q}^{\alpha\beta}\times \\
&\times \quad \exp\left(N\left[-i\sum_{\alpha<\beta} q^{\alpha\beta}\bar{q}^{\alpha\beta} - i\sum_\alpha Q^\alpha\bar{Q}^\alpha + i\sum_\alpha \bar{M}^\alpha\overline{J}\right]\right)\times \\
&\times \quad \exp\left(N\ln\left\{\prod_\alpha\int_0^\infty dJ^\alpha \exp\left(-i\sum_\alpha \bar{M}^\alpha J^\alpha + \sum_\alpha \bar{Q}^\alpha\cdot(J^\alpha)^2 + \frac{1}{2}\sum_{\alpha\neq\beta}\bar{q}^{\alpha\beta}J^\alpha J^\beta\right)\right\}\right)\times \\
&\times \quad \exp\left(P\ln\left\{\prod_\alpha\int_{-\infty}^\infty \frac{dx_\alpha}{2\pi}\exp\left(i\sum_\alpha x_\alpha b^\alpha - \frac{1}{2}\sum_\alpha (x_\alpha)^2(\sigma_{eff}^2 + Q^\alpha\cdot\sigma_\nu^2) - \frac{1}{2}\sum_{\alpha\neq\beta}x_\alpha x_\beta c^{\alpha\beta}\right)\right\}\right)
\end{aligned} \tag{4.71}$$

Now, we assume replica symmetry and redefine $\hat{M} = i\bar{M}$, $\hat{Q} = i\bar{Q}$ and $\hat{q} = i\bar{q}$. Let us consider equation 4.71 line by line. In the limit $n\to 0$, the first line of equation 4.71 becomes:

$$\exp\left(N\left[-\frac{1}{2}(n-1)n\cdot q\hat{q} - n\cdot Q\hat{Q} + n\cdot\hat{M}\overline{J}\right]\right) = \exp\left(Nn\left[\frac{1}{2}q\hat{q} - Q\hat{Q} + \hat{M}\overline{J}\right]\right) \tag{4.72}$$

The second line of equation 4.71 yields:

$$\begin{aligned}
&\exp\left(N\ln\left\{\prod_\alpha\int_0^\infty dJ^\alpha\exp\left(-\sum_\alpha\hat{M}J^\alpha + \sum_\alpha(\hat{Q}-\frac{\hat{q}}{2})(J^\alpha)^2 + \frac{1}{2}\sum_{\alpha,\beta}\hat{q}J^\alpha J^\beta\right)\right\}\right) \\
&\overset{HST}{=} \exp\left(N\ln\left\{\prod_\alpha\int_0^\infty dJ^\alpha\int_{-\infty}^{+\infty}\frac{dt}{i\sqrt{2\pi\hat{q}}}\exp\left(-\sum_\alpha J^\alpha(\hat{M}+it) + \sum_\alpha(\hat{Q}-\frac{\hat{q}}{2})(J^\alpha)^2 + \frac{t^2}{2\hat{q}}\right)\right\}\right) \\
&= \exp\left(N\ln\left\{\prod_\alpha\int_0^\infty dJ^\alpha\int_{-\infty}^{+\infty}\frac{du}{\sqrt{2\pi}}\exp\left(\sum_\alpha J^\alpha(\sqrt{\hat{q}}u - \hat{M}) + \sum_\alpha(\hat{Q}-\frac{\hat{q}}{2})(J^\alpha)^2 - \frac{u^2}{2\hat{q}}\right)\right\}\right) \\
&= \exp\left(N\ln\left\{\int_{-\infty}^{+\infty}Du\left[\int_0^\infty dJ\exp\left(J(\sqrt{\hat{q}}u-\hat{M}) + (\hat{Q}-\frac{\hat{q}}{2})J^2\right)\right]^n\right\}\right) \\
&\overset{n\to 0}{=} \exp\left(Nn\int_{-\infty}^{+\infty}Du\ln\left\{\int_0^\infty dJ\exp\left(J(\sqrt{\hat{q}}u-\hat{M}) + (\hat{Q}-\frac{\hat{q}}{2})J^2\right)\right\}\right)
\end{aligned} \tag{4.73}$$

The third line of equation 4.71 gives:

$$
\begin{aligned}
&\exp\left(P\ln\left\{\prod_\alpha\int_{-\infty}^{\infty}\frac{dx_\alpha}{2\pi}\exp\left(i\sum_\alpha x_\alpha b-\frac{1}{2}\sum_\alpha(x_\alpha)^2(Q-q)\sigma_\nu^2-\frac{1}{2}\sum_{\alpha,\beta}x_\alpha x_\beta c\right)\right\}\right)\\
&\overset{HST}{=}\exp\left(P\ln\left\{\prod_\alpha\int_{-\infty}^{\infty}\frac{dx_\alpha}{2\pi}\int_{-\infty}^{\infty}\frac{dt}{\sqrt{2\pi c}}\exp\left(i\sum_\alpha x_\alpha(b+t)-\frac{1}{2}\sum_\alpha(x_\alpha)^2(Q-q)\sigma_\nu^2-\frac{t^2}{2c}\right)\right\}\right)\\
&=\exp\left(P\ln\left\{\int_{-\infty}^{\infty}Du\prod_\alpha\int_{-\infty}^{\infty}\frac{dx_\alpha}{2\pi}\exp\left(i\sum_\alpha x_\alpha(b+\sqrt{c}u)-\frac{1}{2}\sum_\alpha(x_\alpha)^2(Q-q)\sigma_\nu^2\right)\right\}\right)\\
&=\exp\left(P\ln\left\{\int_{-\infty}^{\infty}Du\frac{1}{(2\pi)^n}\left(\sqrt{\frac{2\pi}{(Q-q)\sigma_\nu^2}}\right)^n\exp\left(-\frac{1}{2}\sum_\alpha\frac{(b+\sqrt{c}u)^2}{(Q-q)\sigma_\nu^2}\right)\right\}\right)\\
&=\exp\left(P\ln\left\{\int_{-\infty}^{\infty}Du\left[\frac{1}{\sqrt{2\pi}}\sqrt{\frac{1}{(Q-q)\sigma_\nu^2}}\exp\left(-\frac{1}{2}\frac{(b+\sqrt{c}u)^2}{(Q-q)\sigma_\nu^2}\right)\right]^n\right\}\right)\\
&\overset{n\to0}{=}\exp\left(Pn\int_{-\infty}^{\infty}Du\ln\left\{\frac{1}{\sqrt{2\pi}}\sqrt{\frac{1}{(Q-q)\sigma_\nu^2}}\exp\left(-\frac{1}{2}\frac{(b+\sqrt{c}u)^2}{(Q-q)\sigma_\nu^2}\right)\right\}\right)\\
&=\exp\left(Pn\int_{-\infty}^{\infty}Du\left(-\ln(\sigma_\nu)-\frac{1}{2}\ln(2\pi)-\frac{1}{2}\ln(Q-q)-\frac{1}{2}\frac{(b+\sqrt{c}u)^2}{(Q-q)\sigma_\nu^2}\right)\right)\\
&=\exp\left(Pn\left(-\ln(\sigma_\nu)-\frac{1}{2}\ln(2\pi)-\frac{1}{2}\ln(Q-q)-\frac{1}{2}\frac{b^2+c}{(Q-q)\sigma_\nu^2}\right)\right)
\end{aligned}
\tag{4.74}
$$

We can disregard all parts which do not depend on the order-parameters, since they do not affect the saddle point. We get:

$$
\langle V^n\rangle \quad\propto\quad \int dM d\bar{M} dQ d\bar{Q} dq d\bar{q}\exp(NnF) \tag{4.75}
$$

with

$$
\begin{aligned}
F \quad=\quad & \frac{1}{2}q\hat{q}-Q\hat{Q}+\hat{M}\overline{J}\\
&+\int_{-\infty}^{+\infty}Du\ln\left\{\int_0^\infty dJ\exp\left(J(\sqrt{\hat{q}}u-\hat{M})+(\hat{Q}-\frac{\hat{q}}{2})J^2\right)\right\}\\
&+\alpha\left(-\frac{1}{2}\ln(Q-q)-\frac{1}{2}\frac{(\overline{h}-\overline{\nu}\cdot M)^2+\sigma_{eff}^2+q\sigma_\nu^2}{(Q-q)\sigma_\nu^2}\right)
\end{aligned}
\tag{4.76}
$$

Saddle point equations and critical capacity

The value of $\langle V^n\rangle$ is dominated by the extremum of F. To obtain the values of the order parameters at the extremum, we set the derivatives of F with respect to the order-parameters to zero. At first:

$$
\begin{aligned}
\frac{\partial F}{\partial M} \quad&=\quad \overline{\nu}^2 M-\overline{\nu}\overline{h}\overset{!}{=}0\\
\Leftrightarrow M \quad&=\quad \frac{\overline{h}}{\overline{\nu}}
\end{aligned}
\tag{4.77}
$$

With this result for M follows that:

$$\begin{aligned}\frac{\partial F}{\partial Q} &= -\hat{Q} + \alpha\left(-\frac{1}{2(Q-q)} + \frac{\sigma_{eff}^2 + q\sigma_\nu^2}{2\sigma_\nu^2(Q-q)^2}\right) \stackrel{!}{=} 0 \\ \Leftrightarrow \hat{Q} &= \alpha\frac{\sigma_{eff}^2 + q\sigma_\nu^2 - \sigma_\nu^2(Q-q)}{2\sigma_\nu^2(Q-q)^2}, \end{aligned} \tag{4.78}$$

and

$$\begin{aligned}\frac{\partial F}{\partial q} &= -\frac{1}{2}\hat{q} + \alpha\left(-\frac{1}{2(Q-q)} - \frac{1}{2(Q-q)} - \frac{\sigma_{eff}^2 + q\sigma_\nu^2}{2\sigma_\nu^2(Q-q)^2}\right) \stackrel{!}{=} 0 \\ \Leftrightarrow \hat{q} &= \alpha\frac{\sigma_{eff}^2 + q\sigma_\nu^2}{\sigma_\nu^2(Q-q)^2}. \end{aligned} \tag{4.79}$$

As can be seen from the above equations, $\hat{Q}$ and $\hat{q}$ diverge in the limit $q \to Q$. Following Brunel et al. (2004), we rewrite the quantities $\hat{Q}$ and $\hat{q}$ in order to highlight their divergent behaviour:

$$\hat{q} = \frac{C}{(Q-q)^2} \tag{4.80}$$

$$\hat{q} - 2\hat{Q} = \frac{\alpha}{Q-q}, \tag{4.81}$$

with $C = \frac{\alpha(\sigma_{eff}^2 + q\sigma_\nu^2)}{\sigma_\nu^2}$.
Before evaluating the derivatives of F with respect to the conjugated order parameters $\hat{q}$, $\hat{Q}$ and $\hat{M}$, it is advantageous to rewrite the integral over J appearing in F (equation 4.76):

$$\begin{aligned}&\int_0^\infty dJ \exp\left(J(\sqrt{\hat{q}}u - \hat{M}) + (\hat{Q} - \frac{\hat{q}}{2})J^2\right) \\ =& \int_0^\infty dJ \exp\left(-\frac{1}{2}\left\{(\hat{q} - 2\hat{Q})\cdot J^2 + 2(\hat{M} - \sqrt{\hat{q}}u)\cdot J\right\}\right) \\ =& \int_0^\infty dJ \exp\left(-\frac{1}{2}\left\{\sqrt{\hat{q} - 2\hat{Q}}\cdot J + \frac{\hat{M} - \sqrt{\hat{q}}u}{\sqrt{\hat{q} - 2\hat{Q}}}\right\}^2\right)\exp\left(\frac{(\hat{M} - \sqrt{\hat{q}}u)^2}{2(\hat{q} - 2\hat{Q})}\right)\end{aligned}$$

With the choices

$$\kappa = \frac{\hat{M} - \sqrt{\hat{q}}u}{\sqrt{\hat{q} - 2\hat{Q}}}, \qquad z = \sqrt{\hat{q} - 2\hat{Q}}\cdot J + \kappa, \tag{4.82}$$

we obtain:

$$\begin{aligned}&\int_0^\infty dJ \exp\left(J(\sqrt{\hat{q}}u - \hat{M}) + (\hat{Q} - \frac{\hat{q}}{2})J^2\right) \\ =& \int_\kappa^\infty \frac{dz}{\sqrt{\hat{q} - 2\hat{Q}}}\exp\left(-\frac{z^2}{2}\right)\exp\left(\frac{\kappa^2}{2}\right) \\ =& \frac{1}{\sqrt{\hat{q} - 2\hat{Q}}}\exp\left(\frac{\kappa^2}{2}\right)\sqrt{\frac{\pi}{2}}\,\mathrm{erfc}\left(\frac{\kappa}{\sqrt{2}}\right)\end{aligned} \tag{4.83}$$

Now the derivative of F with respect to $\hat{M}$ can be written as:

$$\begin{aligned}
\frac{\partial F}{\partial \hat{M}} &= \overline{J} + \frac{\partial}{\partial \hat{M}} \int_{-\infty}^{\infty} Du \left(\frac{\kappa^2}{2} - \frac{1}{2} \ln(\hat{q} - 2\hat{Q}) + \ln \left\{ \sqrt{\frac{\pi}{2}} \operatorname{erfc}\left(\frac{\kappa}{\sqrt{2}} \right) \right\} \right) \\
&= \overline{J} + \int_{-\infty}^{\infty} Du \left(\kappa \cdot \frac{\partial \kappa}{\partial \hat{M}} - \exp\left(-\frac{\kappa^2}{2} \right) \frac{\partial \kappa}{\partial \hat{M}} \cdot \left[\sqrt{\frac{\pi}{2}} \operatorname{erfc}\left(\frac{\kappa}{\sqrt{2}} \right) \right]^{-1} \right) \\
&= \overline{J} + \hat{M} \cdot \frac{Q - q}{\alpha} - \sqrt{\frac{Q - q}{\alpha}} \int_{-\infty}^{\infty} Du \exp\left(-\frac{\kappa^2}{2} \right) \cdot \left[\sqrt{\frac{\pi}{2}} \operatorname{erfc}\left(\frac{\kappa}{\sqrt{2}} \right) \right]^{-1} \qquad (4.84)
\end{aligned}$$

where we used equation 4.81 and the fact that $\int_{-\infty}^{\infty} Du \cdot u = 0$. From equations 4.80 - 4.82 we see that κ diverges in the limit $q \to Q$, while its sign is given by the value of u as:

$$\begin{aligned}
u > \frac{\hat{M}}{\sqrt{\hat{q}}} &\Rightarrow \kappa \to -\infty \\
u < \frac{\hat{M}}{\sqrt{\hat{q}}} &\Rightarrow \kappa \to +\infty \qquad (4.85)
\end{aligned}$$

Thus, since $\operatorname{erfc}\left(\frac{\kappa}{\sqrt{2}} \right)^{-1} \to +\infty$ for $\kappa \to +\infty$, the integral in equation 4.84 is dominated by values $u < \frac{\hat{M}}{\sqrt{\hat{q}}}$. This implies that we can expand $\sqrt{\frac{\pi}{2}} \operatorname{erfc}\left(\frac{\kappa}{\sqrt{2}} \right) \approx \exp\left(\frac{-\kappa^2}{2} \right) \frac{1}{\kappa}$. With $B = \frac{\hat{M}}{\sqrt{\hat{q}}}$ we can write:

$$\begin{aligned}
\frac{\partial F}{\partial \hat{M}} &= \overline{J} + \hat{M} \cdot \frac{Q - q}{\alpha_c} - \frac{Q - q}{\alpha_c} \int_{-\infty}^{B} Du \cdot (\hat{M} - \sqrt{\hat{q}} u) \\
&= \overline{J} + \frac{Q - q}{\alpha_c} \sqrt{\hat{q}} \left(\frac{\hat{M}}{\sqrt{\hat{q}}} - \int_{-\infty}^{B} Du \cdot \left(\frac{\hat{M}}{\sqrt{\hat{q}}} - u \right) \right) \\
&= \overline{J} + \frac{Q - q}{\alpha_c} \sqrt{\hat{q}} \left(B - \int_{-\infty}^{B} Du \cdot (B - u) \right) \\
&= \overline{J} + \frac{Q - q}{\alpha_c} \sqrt{\hat{q}} \left(B \cdot \underbrace{\frac{1}{2} \left[1 - \operatorname{erf}\left(\frac{B}{\sqrt{2}} \right) \right]}_{H(B)} - \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{B^2}{2} \right)}_{G(B)} \right) \stackrel{!}{=} 0 \\
&\Leftrightarrow \quad \overline{J} \cdot \alpha_c = \sqrt{C} \left(G(B) - B \cdot H(B) \right) \\
&\Leftrightarrow \quad C = \frac{\overline{J}^2 \cdot \alpha_c^2}{G(B) - B \cdot H(B)} \qquad (4.86)
\end{aligned}$$

With $\frac{\partial \kappa}{\partial \hat{Q}} = \frac{\hat{M}-\sqrt{\hat{q}}u}{\alpha^{\frac{3}{2}}}(Q-q)^{\frac{3}{2}}$, we obtain for $\frac{\partial F}{\partial \hat{Q}}$:

$$\begin{aligned}
\frac{\partial F}{\partial \hat{Q}} &= -Q + \int_{-\infty}^{\infty} Du \left(\kappa \cdot \frac{\partial \kappa}{\partial \hat{Q}} + \frac{1}{\hat{q}-2\hat{Q}} - \exp\left(-\frac{\kappa^2}{2}\right) \frac{\partial \kappa}{\partial \hat{Q}} \cdot \left[\sqrt{\frac{\pi}{2}} \operatorname{erfc}\left(\frac{\kappa}{\sqrt{2}}\right) \right]^{-1} \right) \\
&= -Q + \frac{Q-q}{\alpha_c} + \frac{(Q-q)^2}{\alpha_c^2} \left\{ \hat{M}^2 + \hat{q} - \int_{-\infty}^{B} Du \left(\hat{M}^2 - 2\hat{M}\sqrt{\hat{q}}u + \hat{q}u^2 \right) \right\} \\
&= -Q + \frac{Q-q}{\alpha_c} + \frac{(Q-q)^2}{\alpha_c^2} \hat{q} \left\{ \frac{\hat{M}^2}{\hat{q}} + 1 - \int_{-\infty}^{B} Du \left(\frac{\hat{M}^2}{\hat{q}} - 2\frac{\hat{M}}{\sqrt{\hat{q}}}u + u^2 \right) \right\} \\
&= -Q + \frac{Q-q}{\alpha_c} + \frac{(Q-q)^2}{\alpha_c^2} \hat{q} \left\{ B^2 + 1 - \int_{-\infty}^{B} Du \left(B^2 - 2Bu + u^2 \right) \right\} \\
&= -Q + \frac{Q-q}{\alpha_c} + \frac{C}{\alpha_c^2} \left\{ (B^2+1) \cdot H(B) - B \cdot G(B) \right\} \stackrel{!}{=} 0
\end{aligned} \tag{4.87}$$

Analogously, $\frac{\partial F}{\partial \hat{q}}$ yields:

$$\frac{\partial F}{\partial \hat{Q}} = q + \frac{Q-q}{\alpha_c}(H(B)-1) - \frac{C}{\alpha_c^2} \left\{ (B^2+1) \cdot H(B) - B \cdot G(B) \right\} \stackrel{!}{=} 0 \tag{4.88}$$

Adding equations 4.87 and 4.88 we obtain:

$$\alpha_c = H(B) \tag{4.89}$$

On the other hand, subtracting equation 4.87 from 4.88 yields:

$$\begin{aligned}
& Q + q + \frac{Q-q}{\alpha_c}(H(B)-2) - 2\frac{C}{\alpha_c^2} \left\{ (B^2+1) \cdot H(B) - B \cdot G(B) \right\} = 0 \\
\stackrel{q \to Q}{\Rightarrow} \quad & Q = \frac{C}{\alpha_c^2} \left\{ (B^2+1) \cdot H(B) - B \cdot G(B) \right\}
\end{aligned} \tag{4.90}$$

Since $C = \frac{\alpha(\sigma_{eff}^2 + Q\sigma_\nu^2)}{\sigma_\nu^2}$, we can solve the above equations for C:

$$C = \frac{\alpha_c^2 \sigma_{eff}^2}{\sigma_\nu^2 \cdot (\alpha_c - (B^2+1) \cdot H(B) + B \cdot G(B))} \tag{4.91}$$

Finally, by comparing this with expression 4.86 and using $\alpha_c = H(B)$, we obtain:

$$\frac{B}{G(B) - B \cdot H(B)} = \frac{\sigma_{eff}^2}{\sigma_\nu^2 \cdot \overline{J}^2} = \frac{\sigma_{eff}^2 \cdot \overline{\nu}^2}{\sigma_\nu^2 \cdot h_{ext}^2} \tag{4.92}$$

This equation defines the parameter B in function of the statistics of the firing-rate patterns ν_i^μ and the inputs h_i^μ.

Bibliography

Amari, S.-i. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological cybernetics*, 27(2):77–87.

Amit, D. J. (1992). *Modeling brain function: The world of attractor neural networks*. Cambridge University Press.

Avermann, M., Tomm, C., Mateo, C., Gerstner, W., and Petersen, C. C. (2012). Microcircuits of excitatory and inhibitory neurons in layer 2/3 of mouse barrel cortex. *Journal of neurophysiology*, 107(11):3116–3134.

Bair, W., Koch, C., Newsome, W., and Britten, K. (1994). Power spectrum analysis of bursting cells in area mt in the behaving monkey. *The Journal of neuroscience*, 14(5):2870–2892.

Beloozerova, I. N., Sirota, M. G., and Swadlow, H. A. (2003). Activity of different classes of neurons of the motor cortex during locomotion. *The Journal of neuroscience*, 23(3):1087–1097.

Brunel, N., Hakim, V., Isope, P., Nadal, J.-P., and Barbour, B. (2004). Optimal information storage and the distribution of synaptic weights: perceptron versus purkinje cell. *Neuron*, 43(5):745–757.

Brunel, N. and Sergi, S. (1998). Firing frequency of leaky intergrate-and-fire neurons with synaptic current dynamics. *Journal of theoretical Biology*, 195(1):87–95.

Buzsáki, G. and Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. *Nature Reviews Neuroscience*.

Carandini, M. (2004). Amplification of trial-to-trial response variability by neurons in visual cortex. *PLoS biology*, 2(9):e264.

Chapeton, J., Fares, T., LaSota, D., and Stepanyants, A. (2012). Efficient associative memory storage in cortical circuits of inhibitory and excitatory neurons. *Proceedings of the National Academy of Sciences*, 109(51):E3614–E3622.

Clopath, C. and Brunel, N. (2013). Optimal properties of analog perceptrons with excitatory weights. *PLoS computational biology*, 9(2):e1002919.

Dale, H. (1935). Pharmacology and nerve-endings. *Journal of the Royal Society of Medicine*, 28(3):319–332.

Destexhe, A., Rudolph, M., and Paré, D. (2003). The high-conductance state of neocortical neurons in vivo. *Nature reviews neuroscience*, 4(9):739–751.

Dorrn, A. L., Yuan, K., Barker, A. J., Schreiner, C. E., and Froemke, R. C. (2010). Developmental sensory experience balances cortical excitation and inhibition. *Nature*, 465(7300):932–936.

Frobenius, F. G. (1912). *Über Matrizen aus nicht negativen Elementen*. Königliche Akademie der Wissenschaften.

Fujisawa, S., Amarasingham, A., Harrison, M. T., and Buzsáki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nature neuroscience*, 11(7):823–833.

Funahashi, S., Bruce, C. J., Goldman-Rakic, P. S., et al. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol*, 61(2):331–349.

Fuster, J. M. (1995). *Memory in the cerebral cortex*. Cambridge, MA: MIT Press.

Gardner, E. (1988). The space of interactions in neural network models. *Journal of physics A: Mathematical and general*, 21(1):257.

Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F., and Petersen, C. C. (2010). Membrane potential dynamics of gabaergic neurons in the barrel cortex of behaving mice. *Neuron*, 65(3):422–435.

Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. *Comprehensive Physiology*.

Haider, B., Duque, A., Hasenstaub, A. R., and McCormick, D. A. (2006). Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *The Journal of neuroscience*, 26(17):4535–4545.

Hansel, D. and Mato, G. (2013). Short-term plasticity explains irregular persistent activity in working memory tasks. *The Journal of Neuroscience*, 33(1):133–149.

Hansel, D. and van Vreeswijk, C. (2012). The mechanism of orientation selectivity in primary visual cortex without a functional map. *The Journal of Neuroscience*, 32(12):4049–4064.

Hebb, D. (1968). *0.(1949) The organization of behavior*. Wiley, New York.

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

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

Kullmann, D. M., Moreau, A. W., Bakiri, Y., and Nicholson, E. (2012). Plasticity of inhibition. *Neuron*, 75(6):951–962.

Levy, R. B. and Reyes, A. D. (2012). Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex. *The Journal of Neuroscience*, 32(16):5609–5619.

Matsumura, M., Cope, T., and Fetz, E. (1988). Sustained excitatory synaptic input to motor cortex neurons in awake animals revealed by intracellular recording of membrane potentials. *Experimental brain research*, 70(3):463–469.

Mitchell, J. F., Sundberg, K. A., and Reynolds, J. H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area v4. *Neuron*, 55(1):131–141.

Nesterov, Y. et al. (2007). Gradient methods for minimizing composite objective function.

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

Roxin, A., Brunel, N., Hansel, D., Mongillo, G., and van Vreeswijk, C. (2011). On the distribution of firing rates in networks of cortical neurons. *The Journal of Neuroscience*, 31(45):16217–16226.

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

Shadlen, M. N. and Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *The Journal of neuroscience*, 18(10):3870–3896.

Shafi, M., Zhou, Y., Quintana, J., Chow, C., Fuster, J., and Bodner, M. (2007). Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience*, 146(3):1082–1108.

Shinomoto, S., Kim, H., Shimokawa, T., Matsuno, N., Funahashi, S., Shima, K., Fujita, I., Tamura, H., Doi, T., Kawano, K., et al. (2009). Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLoS Computational Biology*, 5(7):e1000433.

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

Softky, W. R. and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *The Journal of Neuroscience*, 13(1):334–350.

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

Tao, T., Vu, V., Krishnapur, M., et al. (2010). Random matrices: universality of esds and the circular law. *The Annals of Probability*, 38(5):2023–2065.

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

van Vreeswijk, C. and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural computation*, 10(6):1321–1371.

van Vreeswijk, C. and Sompolinsky, H. (2005). Irregular activity in large networks of neurons. In Chow, C., Gutkin, B., Hansel, D., Meunier, C., and Dalibard, J., editors, *Les Houches Lectures LXXX on Methods and models in neurophysics*, pages 341–402, London. Elsevier.

Wohrer, A., Humphries, M. D., and Machens, C. K. (2012). Population-wide distributions of neural activity during perceptual decision-making. *Progress in neurobiology*, 103:156–193.

Chapter 5

Discussion

The work presented here comprises two main results. On the one hand, we developed a new methodology for the analysis of electrophysiological recordings of single-synapse transmission. On the other hand, we devised a network theory of memory storage that strongly links the balanced state with the network's ability to have multiple stable states, at the same time integrating a series of properties of cortical activity. In this chapter we will first discuss implications of these results separately and will then explore where connections can be made.

Probabilistic analysis of synaptic transmission

Central synapses are small and noisy, and, as far as cortex is concerned, this is the rule, not the exception. We have seen that the high variability of synaptic transmission entails significant difficulties when we want to assess synaptic properties by means of electrophysiology. Here, we devised a statistically sound method to fully quantify the quantal and dynamic nature synaptic transmission. We have shown in chapter 3 that our approach makes it possible to extract more information from experimental recordings precisely because we harness the variability present in those. As a consequence, we were able to extract a bigger number of parameters from the same data sets. At this, the quality of the estimates was at least as good, if not better, as the one obtained with the standard least-squares fit.
The most important novelty introduced by our method is certainly the possibility to use any input protocol for synaptic analysis. The basic finding we presented in chapter 3 is that non-repeated Poisson input trains yield better parameter estimates than regular trains. This is, however, almost secondary compared to the fact that with our method synaptic parameters can be estimated from *realistic* spike trains that were, for instance, obtained during *in vivo* recordings [see e.g. Klyachko and Stevens (2006)]. In this way synapses can be studied in more realistic environments, and phenomenological parameters, which necessarily remain to a considerable degree descriptive, would be endowed with more significance. Beyond that, our method could also be used to analyse synaptic responses recorded *in vivo*, for example from certain thalamo-cortical connections which are known to be mono-synaptic [Ganmor et al. (2010)].

It would be interesting to adapt our approach to optical data. It should be possible to probabilistically model the dependence between the amount of neurotransmitter that is released at a synapse and the luminosity of, for instance, some voltage sensitive dye. In this way the same analysis presented here could be transferred to different experimental techniques.
Finally, the method could be extended to cope with recordings that involve more than one synaptic connection. If, for instance, the recorded neurones receive inputs from a pre-synaptic population that can be reliably stimulated, our approach could be used to estimate not single parameters of an individual synapse, but the distributions of parameters of the synaptic population involved. The estimation method we presented in chapter 3 can straightforwardly be adapted to this situation.

A theory of cortical memory storage

Our most important finding is that extensive memory storage implies balance. We saw that the balanced state is needed to keep the quenched fluctuations in the average input, which are indicative of the network's state, finite. This reasoning mirrors the argument made in Hansel and van Vreeswijk (2012), where the balanced state is invoked to account for the selectivity of V1 neurones in rodents. In rodents, no functional map of orientation selectivity is present, suggesting that synaptic connectivity does not (or does not strongly) correlate with function. But even for the case where connectivity is random, Hansel and Van Vreeswijk showed that selectivity arises in the balanced state in response to very weakly tuned inputs. The mechanism is the same as in our case: the balanced network is sensitive to quenched fluctuations in the external inputs, which contain the relevant information. A significant difference is that in our theory the quenched input fluctuations derive from the synaptic connectivity and are thus generated by the network itself. By adjusting the synaptic weights, quenched fluctuations can be controlled and memory patterns can be learnt.
The requirement of optimal memory storage capacity and retrievability leads to predictions that are consistent with many properties of cortical activity. By virtue of a similar mechanism as in Roxin et al. (2011), we obtain highly skewed firing-rate distributions while we generally have low firing rates. Furthermore, the balanced state provides a pronounced temporal spiking variability. We want to stress that all previous accounts of these phenomena are mechanistic, but not functional. That is, they explain how certain properties of cortical activity can be generated, but not *why* they are the way they are. Our theory proposes a functional explanation. Indeed, it suggests that crucial properties of cortical activity can be interpreted as signatures of networks that are optimised for memory storage.
The study of networks with both excitatory and inhibitory neurones where only excitatory-to-excitatory synapses are plastic indicates that the average inhibitory firing-rate should be larger than the excitatory one, which is indeed consistent with data. However, the deteriorating effect of quenched inhibitory noise, which experimental work suggests to be rather large, indicates that the missing inhibitory plasticity has an important impact on the statistics of cortical activity. Indeed, more and more evidence for inhibitory plasticity

has become available in recent years [Dorrn et al. (2010); Kullmann et al. (2012)].
The memory model we propose does not feature a baseline state. Consistent with experimental results, the global state - or macrostate - of the network does not change. This state sets the system in a working point that determines it memory storage properties, in particular the number of microstates that the system can maintain.
In recent years, some studies have put forth evidence that persistent activity of memory cells is much less stable than previously assumed [see e.g. Romo et al. (1999, 2002)]. A large variety of trajectories has been reported. In view of this, some authors have challenged the classical attractor picture and developed WM models based on liquid state machines [Ganguli et al. (2008); Sussillo and Abbott (2009); Barak et al. (2010)]. In these networks, information about a stimulus is stored in the transient dynamics.
If we decide to agree with the claim that persistent activity is not well described by classical attractor networks, we can resort to the following conclusions. On the one hand, we may say that our model is rather a model for the storage and recall of long-term memory than for WM (which includes also the storage of unknown items). Indeed, a recent interesting study Barak et al. (2013) compared different types of memory networks, showing that different strategies of memory storage may corresponds to different degrees of proficiency in a task associated with that memory. According to the authors, liquid-state machines seem to be better fitted explaining experimental findings related to WM of novel stimuli. The attractor network, on the other hand, featuring the highest degree of synaptic structuring, seems to best represent optimal performance of highly trained animals. In this view, our model would be most adequate to explain storage and retrieval of stable long-term memories.
Another option we need to consider is that reported variable trajectories can still be a sign of attractor dynamic. If the working point of our attractor network is set in the vicinity of the bifurcation to the chaotic regime, transient responses can become quite long before reaching the steady-state. This could completely account for the mentioned phenomenology. Interestingly, increasing the gain of the neuronal transduction function both increases critical capacity of our network and draws it closer to the chaotic bifurcation.
Alternatively the variability in firing-rate trajectories could be explained if we set up our networks with correlated patterns and include a source of noise. As shown in Mongillo et al. (2003), this can cause the network activity to switch between similar attractors during delay activity, thereby causing seemingly non-stationary persistent activity patterns.
We have not addressed so far another crucial feature of the theory. As shown in Brunel et al. (2004) (and likewise used in Chapeton et al. (2012) and Clopath and Brunel (2013)), the replica-formalism can also make predictions about the distribution of synaptic weights. At critical capacity, this distribution features a large peak at zero and the positive part of a Gaussian that has a negative mean. The fraction of silent synapses S is given by the simple relation $S = 1 - \alpha_c$. This implies that for the typical capacity values we find in the region of optimal information, 80% - 90% of all connections are silent. Thus, the theory also predicts the well reported experimental finding that cortical connectivity is sparse.

However, the theoretical weight distribution at critical capacity features a tail that falls off according to a Gaussian curve. This is not consistent with experimental findings that suggest that cortical synaptic weights, like firing-rates, are distributed log-normally [see e.g. Song et al. (2005); Levy and Reyes (2012); Avermann et al. (2012); Buzsáki and Mizuseki (2014)]. It has been argued that this discrepancy could be mitigated by finite size effects [Chapeton et al. (2012)], nonlinear summation of synaptic inputs, or the fact that memory storage remains subcritical [Barbour et al. (2007)].
Subcriticality has another advantage. At critical capacity, there is only one solution to the weight learning problem. Thus, it can be expected that, in this situation, any perturbation of the synaptic weights can severely limit the network's reliability. If learning remains subcritical, the only effect of noise might be to move the synaptic weights such that they stay inside the finite volume of solutions. It should be in principle straightforward to investigate in numerical simulations how the network's robustness depends on the magnitude of the noise and the number of stored patterns.
Directly relevant to this issue is the recent experimental finding that spine-sizes of excitatory cortical neurones - which are a good proxy for synaptic efficacy - continuously change their size [Loewenstein et al. (2011)]. These spine dynamics, which include appearance of new spines and disappearance of existing spines, occur on the timescale of days. It would be interesting to study the quantitative effects of these changes on the dynamics of our attractor networks for different capacity levels.
A general strategy to gain a better understanding of the questions linked to the synaptic weights could be to use the cavity method. This approach can be employed to obtain the same results as the replica method [Mézard (1989)], but allows to get a better intuitive understanding of the variables involved. Of particular interest for us is that the cavity method involves thermal noise on the synaptic weights, which may make it possible to relate this approach to the above problems.

A possible connection

A potential connection between our work on synaptic transmission and the memory model is, quite naturally, the learning process. In a very interesting article, Seung showed that noisy synapses can in principle emulate a stochastic gradient descent [Seung (2003)]. As synapses have a finite release probability, release failures in response to pre-synaptic spikes will occur. In function of whether release was successful or not and depending on a reward signal, the synapses lower or raise their release probability and thereby their total synaptic efficacy. In this framework, networks are actually capable to learn temporally structured responses by utilising the synapses' STP.
Our results from chapter 3, however, indicate that synaptic efficacy is correlated with the number of release sites, N, and not with release probability. It could be worth investigating a mechanism that is analog to Seung's, but involves the increase or decrease of N. Note that a change in the number of release sites has a similar joint effect on synaptic reliability and efficacy as the modulation of the release probability; increasing N both augments synaptic strength and reduces the occurrence of failures. A crucial difference between such a scheme and the one by Seung is that in the latter synapses can,

in principle, become deterministic and that their dynamic range is limited. If N is the adjustable quantity, synapses may become very reliable, but would nevertheless remain noisy; differences in synaptic efficacy could become very large. This would be indeed biologically more plausible.

Bibliography

Avermann, M., Tomm, C., Mateo, C., Gerstner, W., and Petersen, C. C. (2012). Microcircuits of excitatory and inhibitory neurons in layer 2/3 of mouse barrel cortex. *Journal of neurophysiology*, 107(11):3116–3134.

Barak, O., Sussillo, D., Romo, R., Tsodyks, M., and Abbott, L. (2013). From fixed points to chaos: Three models of delayed discrimination. *Progress in neurobiology*, 103:214–222.

Barak, O., Tsodyks, M., and Romo, R. (2010). Neuronal population coding of parametric working memory. *The Journal of Neuroscience*, 30(28):9424–9430.

Barbour, B., Brunel, N., Hakim, V., and Nadal, J.-P. (2007). What can we learn from synaptic weight distributions? *TRENDS in Neurosciences*, 30(12):622–629.

Brunel, N., Hakim, V., Isope, P., Nadal, J.-P., and Barbour, B. (2004). Optimal information storage and the distribution of synaptic weights: perceptron versus purkinje cell. *Neuron*, 43(5):745–757.

Buzsáki, G. and Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. *Nature Reviews Neuroscience*.

Chapeton, J., Fares, T., LaSota, D., and Stepanyants, A. (2012). Efficient associative memory storage in cortical circuits of inhibitory and excitatory neurons. *Proceedings of the National Academy of Sciences*, 109(51):E3614–E3622.

Clopath, C. and Brunel, N. (2013). Optimal properties of analog perceptrons with excitatory weights. *PLoS computational biology*, 9(2):e1002919.

Dorrn, A. L., Yuan, K., Barker, A. J., Schreiner, C. E., and Froemke, R. C. (2010). Developmental sensory experience balances cortical excitation and inhibition. *Nature*, 465(7300):932–936.

Ganguli, S., Huh, D., and Sompolinsky, H. (2008). Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences*, 105(48):18970–18975.

Ganmor, E., Katz, Y., and Lampl, I. (2010). Intensity-dependent adaptation of cortical and thalamic neurons is controlled by brainstem circuits of the sensory pathway. *Neuron*, 66(2):273–286.

Hansel, D. and van Vreeswijk, C. (2012). The mechanism of orientation selectivity in primary visual cortex without a functional map. *The Journal of Neuroscience*, 32(12):4049–4064.

Klyachko, V. A. and Stevens, C. F. (2006). Excitatory and feed-forward inhibitory hippocampal synapses work synergistically as an adaptive filter of natural spike trains. *PLoS biology*, 4(7):e207.

Kullmann, D. M., Moreau, A. W., Bakiri, Y., and Nicholson, E. (2012). Plasticity of inhibition. *Neuron*, 75(6):951–962.

Levy, R. B. and Reyes, A. D. (2012). Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex. *The Journal of Neuroscience*, 32(16):5609–5619.

Loewenstein, Y., Kuras, A., and Rumpel, S. (2011). Multiplicative dynamics underlie the emergence of the log-normal distribution of spine sizes in the neocortex in vivo. *The Journal of Neuroscience*, 31(26):9481–9488.

Mézard, M. (1989). The space of interactions in neural networks: Gardner's computation with the cavity method. *Journal of Physics A: Mathematical and General*, 22(12):2181.

Mongillo, G., Amit, D. J., and Brunel, N. (2003). Retrospective and prospective persistent activity induced by hebbian learning in a recurrent cortical network. *European Journal of Neuroscience*, 18(7):2011–2024.

Romo, R., Brody, C., Hernández, A., and Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, 399:470–473.

Romo, R., Hernández, A., Zainos, A., Lemus, L., and Brody, C. D. (2002). Neuronal correlates of decision-making in secondary somatosensory cortex. *Nature neuroscience*, 5(11):1217–1225.

Roxin, A., Brunel, N., Hansel, D., Mongillo, G., and van Vreeswijk, C. (2011). On the distribution of firing rates in networks of cortical neurons. *The Journal of Neuroscience*, 31(45):16217–16226.

Seung, H. S. (2003). Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. *Neuron*, 40(6):1063–1073.

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

Sussillo, D. and Abbott, L. F. (2009). Generating coherent patterns of activity from chaotic neural networks. *Neuron*, 63(4):544–557.

List of Figures