Freie Universität Berlin

Modelling Cortical Variability Dynamics

From Inhibitory Clustering to Context Dependent Modulation

Inaugural-Dissertation

to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the

Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by

thomas rost

from

Kingston upon Thames

2016

Die vorliegende Arbeit entstand zwischen 2011 und 2016 am Institut für Biologie der Freien Universität Berlin und dem Institut für Zoologie der Universität Köln unter der Leitung von Prof. Dr. Martin Nawrot.

1. Gutachter: Prof. Dr. Martin Nawrot

2. Gutachter: Prof. Dr. Hans-Joachim Pflüger

Disputation am 6.9.2016

Abstract

The present study is concerned with measurement and modelling of neural variability dynamics in the mammalian neocortex. The variability statistics of cortical spike trains are analysed in the framework of renewal process theory. The estimation of interval variability is refined so that it can be used to extract the underlying firing rate variance from neural activity. These methods are then applied to a cortical data set where we show that the reduction in count variability commonly observed at the onset of stimuli can be dissected into a modulated component of rate variance and an interval variability that remains constant. The reduction of count variability is commonly modelled using balanced networks of excitatory and inhibitory units where the excitatory population is divided into clusters of stronger internal connectivity. Using the mean field description of balanced networks of binary neurons, we analyse the stable fixed points of these excitatory cluster networks. We find that the firing rates of the active clusters are always close to the saturation limit. As a possible remedy we propose to subdivide the inhibitory population into clusters as well so that each excitatory cluster is selectively balanced and show in the mean field analysis that this does indeed solve the problem of rate saturation. We then transfer the concept of inhibitory and excitatory clustering of the connectivity to more realistic models of integrate and fire neurons and show that here it conserves the local balance of excitation and inhibition during cluster cycling dynamics. Thereby the interval variability seen in cortical data is conserved during high activity states of the clusters which is in contrast to previous models. Finally we analysed a data set where monkeys executed a delayed centre out reach task where varying information about the target direction was given. The depth of modulation of the count variance in the monkey's motor cortex was dependent on how much information was given during the delay period. Using a balanced network with inhibitory and excitatory clusters, we showed that the context dependent modulation of variability during the delay period can be captured by a simple probabilistic model of movement preparation.

Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Messung und Modellierung der Variabilitätsdynamik im Neokortex der Säugetiere. Die Variabilitätsstatistik von Zeitreihen kortikaler Aktionspotentiale werden aus der Perspektive der Theorie der Erneuerungsprozesse analysiert. Die Schätzung der Intervalvariabilität wurde verfeinert, so dass diese herbeigezogen werden kann um die Feuerratenvarianz neuronaler Aktivität zu berechnen. Diese Methoden wurden dann auf einem Satz neokortikaler Daten angewendet um zu zeigen, dass die oft beobachtete Reduktion der Fanofanofaktoren zu beginn der Stimuluspräsentation einer modulierten Ratenvarianz bei konstanter Intervalvariabilität zugeschrieben werden kann. Die Reduktion des Fanofaktors wird oft in neuronalen Netzwerken simuliert, in denen exzitatorische und inhibitorische Neuronen sich das Gleichgewicht halten (Balanced Networks), indem Gruppen (Cluster) von exzitatorischen Neuronen untereinander stärker verbunden werden. Mithilfe der Meanfield-Beschreibung binärer Netzwerke wurden die stabilen Fixpunkte solcher exzitatorischer Clusternetzwerke untersucht, mit dem Ergebnis, dass hier die aktiven Cluster in die Feuerratensättigung getrieben werden. Als mögliche Abhilfe wird vorgeschlagen, auch die inhibitorische Population in Cluster zu unterteilen, so dass jede exzitatorische Population von einer inhibitorischen im Gleichgewicht gehalten wird. Die Meanfieldanalyse solcher Konfigurationen ergab, dass hierdurch die Ratensättigung verhindert wird. Das Konzept der inhibitorischen und exzitatorischen Cluster wurde dann auf Modelle bestehend aus realistischeren Leckintegratorneuronen übertragen. Hier führte dies zu einem lokalen Gleichgewicht der synaptischen Ströme, so dass die im Kortex beobachtete Intervallvariabilität auch in den aktiven Clustern erhalten blieb, ein Aspekt den bisherige Modelle nicht abbilden konnten. Im letzten Schritt analysierten wir kortikale Ableitungen eines Experiments bei dem Affen zur Aufgabe hatten, am Ende eines Verzögerungsintervalls in eine bestimmte Richtung zu greifen. Die Zielrichtung wurde dabei zu Beginn des Verzögerungsintervalls unterschiedlich genau spezifiziert. Hierbei waren die gemessenen Fanofaktoren im Motorkortex für die einzelnen Konditionen unterschiedlich stark moduliert. Diese konditi-

onsabhängige Modulation der Varibilität konnte mit einem einfachen Balanced Network Model mit exzitatorischen und inhibitorischen Clustern gut abgebildet werden.

Acknowledgements

First and foremost I would like to thank my supervisor Martin Nawrot for his prolonged support. I am also indebted to all the members of the former AG-Neuroinformatik at the FU. Particularly I thank Farzad Farkhooi for many discussions which may have left me confused at first but usually started to make sense after some consideration. At this point I would also like to mention Aster, the *coffee woman* in Dahlem who ensured a constant supply of caffeine. In Cologne Moritz Deger always had something useful to say and I also thank him for proofreading parts of this thesis. Finally, I thank my friends and family, in particular Laura, who supported me throughout and helped me through the occasional period of frustration.

Whatever the Thinker thinks, the Prover will prove.

- Robert Anton Wilson

Contents

Chapter 1

Introduction

Neural responses in the mammalian neocortex are notoriously variable. Even when animal behaviour is consistent across repetitions of experimental tasks, the neuronal responses look very different each time. This variability is found on all temporal and spatial scales. To this day it remains a matter of discussion how the brain can cope with this variability or whether it might even be an essential part of neural computation.

Classically, this variability has been interpreted as noise. In this view, there is a *signal* in the neural activity which is buried under some *noise*. By averaging over trials aligned to some experimental task, the noise is cancelled out and the *trial averaged firing rate* emerges as the signal. An obvious problem of this view is that trials are a construct imposed by the experimenter which the animal is likely unaware of. Still, behaviour seems to work fine on a single trial basis where averaged *signals* are unavailable. One might argue that averages could be taken instantaneously over populations of neurons coding for the same entity. If the *noise* carried by the individual neurons is uncorrelated, it can be eliminated in that way. A potential source of this noise could be the thermodynamic or quantum mechanical randomness inherent in external stimulus modalities [Faisal et al., 2008]. But then this averaging should tend to reduce the amount of noise with increasing distance from the sensory periphery. In fact the opposite has been observed: cortical firing becomes more and more variable in higher brain areas [Kara et al., 2000].

In earlier years, researchers attempted to explain this variability using stochastic elements in detailed single neuronal models (e.g. [Stein, 1967]). Later it was shown that single neurons can precisely reproduce the same spike trains when repeatedly injected with identical current traces which resemble their natural inputs [Mainen and Sejnowski, 1995].

Arieli et al. [1996] showed that the variability in neural responses can be greatly reduced if the activity just before stimulus onset is subtracted from the signal in each trial before averaging. They hypothesized that there is some background state or *ongoing activity* which varies on a much slower time scale than the activity related to the experimental task. Churchland et al. [2010] examined the spike count variance over trials in a range of datasets and found that it was *quenched* at stimulus onset, i.e. variability is much lower when the activity is aligned with respect to some experimental event than during spontaneous activity.

Both these results indicate that variability is reduced if more factors are controlled for. Unfortunately most of the activity in the neocortex is beyond the control of the experimenter. Still, the interpretation is shifting in recent years from viewing cortical activity as noise to calling it *unexplained variance* [Masquelier, 2013]. Renart and Machens [2014] argue that nearly all neural variability can either be attributed to uncontrolled external factors or uncontrolled internal factors such as attention, mood or motivation and that more and more of it will eventually become experimentally controllable.

Consequently, neural activity may range from being completely controlled, as in the current injection experiments of Mainen and Sejnowski [1995], to consisting entirely of *unexplained variance* as is the case for spontaneous cortical activity. Most cortical datasets are situated at an intermediate level on this scale with a tendency towards more control with the development of more sophisticated experimental techniques.

An interesting development is the interpretation of spontaneous cortical activity as a representation of experience or expectation about the world. Spontaneously occurring patterns tend to resemble patterns evoked by sensory input [Kenet et al., 2003; Luczak et al., 2009] and the frequency of their occurrence can be seen as a prior in a Bayesian sense [Berkes et al., 2011].

In the present study we adopt the view that variability is not noise and ask what some of its characteristics are and how they could come about. This thesis is thematically divided into three parts.

In chapter 2 we consider in detail how the variability of spike trains can be quantified. Here we make use of the theory developed for renewal point processes, a class of random processes that generates events with certain interval distributions [Cox and Isham, 1980]. Note that this does not imply that we interpret spike trains as *noisy*. We use refined spike statistics to dissect the components of variability similar to previous methods [Shadlen and Newsome, 1998; Churchland et al., 2011]. The analyses are then applied to a dataset of extracellular recordings from monkey motor cortex.

Chapters 3 and 4 are concerned with basic mechanistic models of cortical variability. In chapter 3 we use the mean field theory developed for balanced networks of binary neurons [van Vreeswijk and Sompolinsky, 1998; Renart et al., 2010] to examine the properties of a recently proposed class of network models which can produce stimulation-induced variability reduction [Deco and Hugues, 2012; Litwin-Kumar and Doiron, 2012; Mazzucato et al., 2015]. In these models, activity *cycles* between assemblies of excitatory units, introducing an additional component of rate variance on a relatively slow time scale. We show that the high firing rates often seen in the active assemblies of such models which hamper switching between states can be overcome by introducing inhibitory clusters as well. In chapter 4 we then apply the concept to spiking network models and show that the introduction of inhibitory clusters causes a local balance in the networks so that interval variability is conserved as we have found in the cortical data set.

In chapter 5 we show that the variability in monkey motor cortex depends on the experimental condition. We then apply a similar stimulus protocol to our model and show that it can capture the context dependent variability modulation – an effect that has to our knowledge not been previously explained.

Note that no detailed discussion is given in the individual chapters. The main results are discussed in context in chapter 6.

Chapter 2

Quantifying Neural Variability

In order to gain insight into the nature and possible function of neural variability, the first step must be to become clear about what one is actually measuring. This work deals with the variability of spike trains of single neurons. Two perspectives exist on the variance of spike trains: variability of the spike count in a certain time frame and variability of the inter spike intervals (ISIs).

Count variance is usually measured across several time windows of equal length. This is only meaningful, if similar conditions exist in time windows that are compared. Usually these time windows are aligned to some repeated experimental condition so that as many factors as possible are controlled for. We will refer to such repeated measurements as trials. The time between consecutive trials of the same experimental condition can be substantial, on the order of seconds to minutes, so that count statistics across trials capture neural variability on relatively long time scales.

Interval statistics on the other hand also capture variability on very short time scales (10s of ms).

In the current chapter we will first deal with the intricacies of estimating count and interval statistics in section 2.1. In section 2.2 we will then see how interpreting neuronal firing as a doubly stochastic point process allows us to derive the trial-to-trial variance in the underlying spiking probability from the basic statistics. Finally in section 2.3 the analyses will be applied to physiological recordings from monkey motor cortex.

2.1 Measuring Count and Interval Statistics

When quantifying variability of neuronal firing, it is useful to compare the statistics to those of simplified models of spike generation to guide the interpretation of the results. The simplest model of neuronal firing that is often employed is the Poisson process. The Poisson process belongs to the family of renewal point processes where intervals between events are independently drawn from some distribution [Cox and Isham, 1980]. In the Poisson case the intervals are exponentially distributed. For ISIs longer than about $10ms$, experimentally determined ISI-histograms are well described by this model [Dayan and Abbott, 2001]. For shorter intervals however, a discrepancy arises because refractoriness causes short intervals to be less likely than the Poisson model predicts. A more flexible model for neuronal activity is the Gamma process, of which the Poisson model is a special case. It has an additional parameter that controls the regularity of firing and although it does not explicitly account for refractoriness, it can be well fitted to experimental data and has frequently been used to model neural firing (e.g. [Stein, 1967; Reich et al., 1998; Baker and Gerstein, 2001; Dayan and Abbott, 2001]).

For a Gamma process, the intervals are independent and identically distributed (iid) according to the probability density function ($p.d.f.$):

$$f_{\alpha,\rho}(x) = \begin{cases} \frac{1}{\Gamma(\alpha)} \rho^{\alpha} x^{\alpha-1} \, \mathrm{e}^{-\rho x} & , x \geq 0 \\ 0 & , x < 0 \end{cases} \tag{2.1}$$

$\Gamma(s) = \int_0^{\infty} y^{s-1} \, \mathrm{e}^{-y} \, dy$ is the Gamma function and α and ρ are denoted shape and rate parameters respectively. The resulting mean interval of the distribution is:

$$\mu = \mathrm{E}[x] = \frac{\alpha}{\rho} \tag{2.2}$$

Note that for $\alpha = 1$ equation 2.1 becomes the exponential distribution, resulting in a Poisson process.

Where simulated gamma processes are shown throughout this work, we drew ISIs according to equation 2.1. The resulting intervals were then concatenated to spike trains and binned with a resolution of $1ms$. Inhomogeneous processes, i.e. processes with varying firing rates, were generated following the time warping method [Nawrot, 2003; Brown et al., 2001]. We also ensured equilibrium conditions for all simulated processes by either including a warm-up time in the simulation for non-stationary processes or by drawing the first interval from a modified *length biased* distribution $\mathcal{F}_{\alpha,\rho}(x)$ which was modelled as the product two independently drawn random variables UY, where Y was drawn

from $f_{\alpha+1,\rho}(y)$ and U was uniformly distributed in $[0, 1]$ [Nawrot et al., 2008].

The Fano Factor (FF) was originally used to estimate the statistical fluctuations of ionization in gases subject to radiation [Fano, 1947]. In neuroscience, it is used to quantify the dispersion of spike counts across trials.

$$FF = \frac{\sigma_c^2}{\mu_c}, \tag{2.3}$$

where σ_c^2 and μ_c are the variance and mean count over repeated observations. The estimation of the FF is biased towards unity for small estimation windows Δ. This bias however quickly becomes negligible when Δ is several multiples of the mean ISI [Nawrot, 2010].

Interval statistics are usually characterised by the Coefficient of Variation (CV) of the ISI distribution.

$$CV^2 = \frac{\sigma_{ISI}^2}{\mu_{ISI}^2} \tag{2.4}$$

Here, σ_{ISI}^2 and μ_{ISI} are the variance and mean of the intervals between action potentials. Note that we have given the squared Coefficient of Variation (CV^2). The reason is that for Gamma processes, the CV^2 is equal to the inverse of the shape parameter α [Cox and Isham, 1980; Nawrot et al., 2008].

$$CV_\infty^2 = \frac{1}{\alpha} \tag{2.5}$$

The subscript ∞ here indicates that this is the theoretically predicted value, which is not necessarily what one measures from spike trains, as we will see in the following.

Estimating the CV^2 requires some caution, as modulations in firing rate increase the interval variability. This problem can be overcome by a technique called *rate unwarping* [Nawrot et al., 2008; Nawrot, 2010]. If repeated observations or trials of neural activity are available and one assumes that the action potentials in each trial were generated according to some underlying firing rate $\nu(t)$ which varies consistently in time with respect to some external reference (i.e. experimental condition), the spike trains can be projected to a new time axis t', where the firing rate is constant. [Nawrot et al., 2008] used the term operational time (ot) for this transformed time axis and defined it as the unit of the cumulative integral of the trial-averaged firing rate:

$$t'(t) = \int_0^t \nu(s)ds \tag{2.6}$$

Figure 2.1 illustrates how repeated trials of poisson activity with (arbitrarily

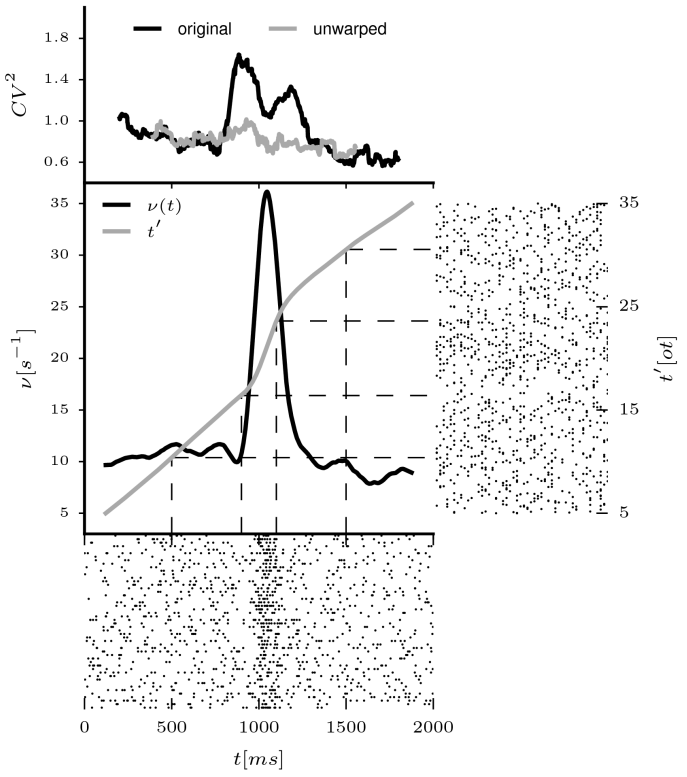

Figure 2.1: Illustration of rate-unwarped CV^2 calculation. The lower panel shows a raster plot of 50 of a Poisson process with modulated firing rate. The firing rate estimated with a triangular kernel (ν) and its cumulative integral (t') are shown in the central panel. Dashed lines indicate how t' is used to project the spike trains to a space where the average firing rate is constant (raster plot on the $r.h.s.$).

generated) time dependent rate $\nu(t)$ can be projected to the ot-axis using this technique. The resulting spike trains in ot (right-hand side ($r.h.s.$) panel) have a constant firing rate $\nu' = 1/\mu_{ISI'} = 1$. In this domain, the time resolved CV^2 can then be calculated in sliding windows of constant width expressed in units of the average transformed ISI (i.e. units of ot). The resulting statistics can then be projected back into the original *real* time domain using the inverse of t'. Note that the *back-projected* values are not equally spaced in time. Where averages over thus calculated CV^2-estimates are taken, we therefore interpolate the resulting values to a regular time grid. The effect of this technique is shown in the top panel of figure 2.1. The black line is the result of calculating the CV^2 in a sliding window of $400ms$ width in real time. In the region where the rate is modulated, this method strongly over-estimates the expected poissonian interval variability of 1. The grey trace is the result of the rate-unwarped CV^2-estimation. Note

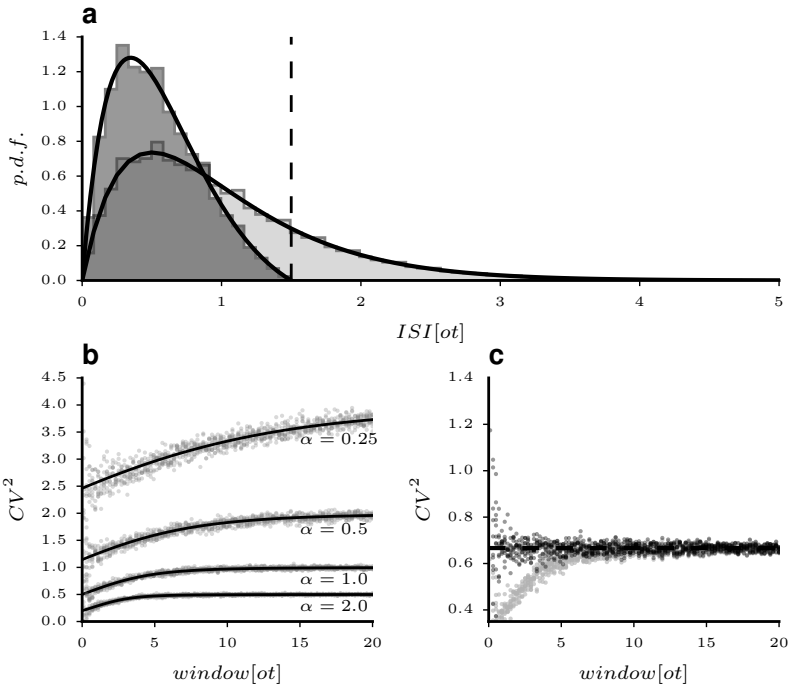

Figure 2.2: Bias in estimating the CV^2 **a**) Illustration of bias due to *right censoring*. Lighter shaded histogram shows ISI-distribution of $\sim 10^6$ intervals of a simulated gamma-process with $\alpha = 2$. The darker histogram shows the effect of a narrow observation window of $ot = 1.5$ (dashed line). Solid lines represent theoretical predictions $f(x)$ and $\hat{f}(x)$. **b**) Effect of estimation window size for different gamma-order parameters α. Dots represent measured CV^2 from simulations, solid lines are the theoretically expected values calculated from $\hat{f}(x)$. **c**) CV^2 measured from gamma-simulations with $\alpha = 0.7^{-1}$ before (lighter dots) and after bias correction using equation 2.11 (darker dots). Dashed line represents CV^2_∞. Each dot was measured from 500 trials of simulated activity.

that the line is shorter than the black one. This is due to the fact that two sliding window operations were employed in the process. One for the kernel-based firing rate estimation and another one for the CV^2 calculation in the ot domain.

Another problem with estimating the CV^2 follows from finite size estimation windows. In an estimation window of width T, only $ISIs < T$ can be observed. If the underlying process has non-zero probabilities for larger intervals, the CV^2 will be under-estimated. This effect is known as *right-censoring* [Wiener, 2003; Nawrot et al., 2008]. In figure 2.2 a it is illustrated for a Gamma process with $\alpha = 2$. The lighter histogram was computed from $\sim 10^6$ intervals of simulated activity (i.e. $T = 10^6 ot$). The same spike train was then cut into windows of $T = 1.5ot$ and the resulting intervals were plotted as the darker histogram. [Nawrot et al., 2008] have shown that the *censored* version of a distribution $f(x)$ arising from a window-width T can be well approximated by assuming that

very short intervals are virtually unaffected by the observation window, while the probability of finding longer windows falls linearly to zero at $x = T$. I.e. the probability of observing an interval $x \in [0, T]$ is proportional to $(T - x)$. The *censoring*-corrected distribution is then:

$$\hat{f}(x|T) = \begin{cases} \frac{(T-x)}{\eta} f(x) & \text{for } 0 \leq x < T \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

with $\eta = \int_0^T (T - s) f(s) ds$ being a normalisation constant.

It can be seen in figure 2.2a, that the censored distribution is well approximated by $\hat{f}(x)$ (solid line on darker histogram). For a known process order we can hence calculate the coefficient of variation that would be estimated in a window of size T as:

$$\widehat{CV}^2 = \frac{\hat{\sigma}^2}{\hat{\mu}^2} \tag{2.8}$$

with

$$\hat{\mu} = \int_0^T x \hat{f}(x) dx \tag{2.9}$$

and

$$\hat{\sigma}^2 = \int_0^T (x - \hat{\mu})^2 \hat{f}(x) dx \tag{2.10}$$

We denote the empirical estimate that we obtain from spike trians of unknown Gamma order as $CV^{2\prime}$. Figure 2.2 b shows $CV^{2\prime}$ measured from simulations and the theoretical prediction $\widehat{CV}^2$ for different Gamma orders α versus the estimation window width T. For more irregular processes (small α), the effect of right-censoring is more severe. For poisson processes ($\alpha = 1$), the bias disappears at window sizes of about $10ot$.

It can be seen that the match between simulation and theoretical prediction is quite close. Further, there is a one to one mapping for a given window size between the estimate $CV^{2\prime}$ and process order α. Since the CV^2 is calculated in the ot domain where the average ISI equals one, equation 2.7 has only one free parameter α. We can hence correct for the bias due to right-censoring by finding the underling process order $\hat{\alpha}$ as:

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \left(\widehat{CV}^2 - CV^{2\prime} \right)^2, \tag{2.11}$$

the inverse of which is our bias-corrected CV^2 estimate.

Since no closed solution exists for the cumulative Gamma density function, we solve equation 2.11 numerically. The results are shown in figure 2.2 c for simulated processes with $CV^2_\infty = 1/\alpha = 0.7$. The bias correction allows us to

obtain better estimates of the CV^2 for small estimation windows and will be used throughout this work were CV^2-estimates are given.

The CV^2 lends itself to applications to Gamma processes because of its theoretical relationship to the distribution's parameters. There are however other measures of interval variability which to some degree overcome the problem of rate-change sensitivity by considering only pairs of consecutive intervals. We will compare two of these measures to our method of calculating the CV^2. The local Coefficient of Variation (CV_2) was developed to compare the irregularity of firing of visual neurons in the anaesthetised cat to that in vitro [Holt et al., 1996]:

$$CV_2 = 2\left\langle \frac{|\tau - \tau'|}{\tau + \tau'} \right\rangle \tag{2.12}$$

Here, $\langle ... \rangle$ denotes averaging and τ and τ' are consecutive ISIs. A similar measure, the local variation (L_V) was devised to distinguish functionally different regions in monkey motor cortex based on irregularity [Shinomoto et al., 2003]:

$$L_V = 3\left\langle \frac{(\tau - \tau')^2}{(\tau + \tau')^2} \right\rangle \tag{2.13}$$

The factors in equations 2.12 and 2.13 are intended to ensure a value of unity for poisson firing.

2.2 Neuronal Activity as a Doubly Stochastic Process

In the literature, FFs of unity and above are usually reported for neocortical data (e.g. [Scholvinck et al., 2015; Nawrot et al., 2008; Churchland et al., 2010; Rickert et al., 2009]). Also, count variability is modulated during experimental trials [Rickert et al., 2009; Churchland et al., 2010] while interval variability seems to be lower and fairly constant over time (e.g. [Gur et al., 1997; Churchland et al., 2010; Ponce-Alvarez et al., 2010; Shinomoto et al., 2009]). For renewal processes, FF and CV^2 are equal given that the firing rate does not change over trials [Cox and Isham, 1980]. The observation that $FF > CV^2$ therefore suggests that the firing rate is *not* constant over trials. The nature of this firing rate variance is not clear. [Arieli et al., 1996] have introduced the term *ongoing activity* to describe variations in firing rate that take place on time scales slower than experimental trials and may or may not be related to current behavioural tasks.

In the frame work of point process theory, such rate variations can be

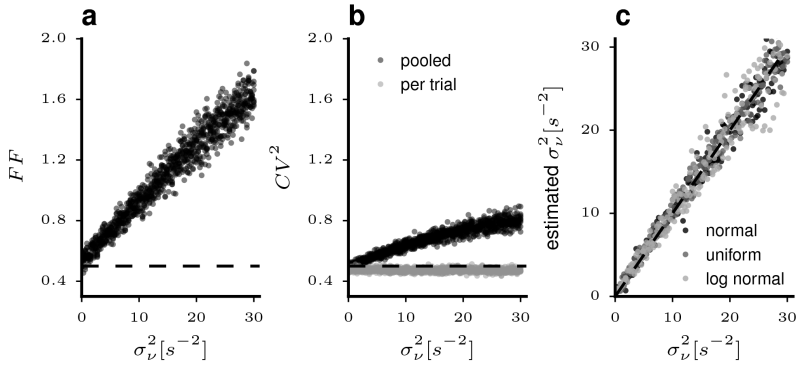

Figure 2.3: Effect of trial-to-trial rate variance σ_ν^2 on spike statistics. Dots represent statistics from 500 trials of simulated Gamma activity. Dashed lines are theoretically expected results. **a**) FF estimates in a $400ms$ window. **b**) Comparison of CV^2 on pooled intervals and evaluated per trial. Window size was $10ot$. **c**) Recovery of simulated rate variance from spike trains for different underlying firing rate distributions.

described as a doubly stochastic process [Cox and Isham, 1980; Shadlen and Newsome, 1998] where the firing rate and spike generating mechanism are seen as two independent random processes. The total count variance σ_c^2 for such a process can then be expressed as the sum of the variance of those components [Shadlen and Newsome, 1998; Churchland et al., 2011]:

$$\underbrace{\sigma_c^2}_{\text{total variance}} = \underbrace{\sigma_{\langle c \rangle}^2}_{\text{rate variance}} + \underbrace{\left\langle \sigma_{c|\nu}^2 \right\rangle}_{\text{intrinsic variance}} \tag{2.14}$$

The *intrinsic variance*, i.e. the variability in spike count c given a rate ν, corresponds to the variability of a point process with stationary firing rate. In neuronal terms it can be interpreted as the variability due to the action potential generation process and the specific connectivity a neuron receives while the *rate variance* is due to changes in network state, behaviour and other external factors.

Looking at equation 2.3, it can be seen that the total count variance σ_c^2 is contained in the definition of the Fano Factor. The FF captures the intrinsic variability as well as variance in firing rate over trials. Figure 2.3 illustrates the effect of rate variance on the elementary statistic measures. Panel a shows the FF calculated over 500 trials of simulated Gamma activity with $CV_\infty^2 = 0.5$. Rate variance has been introduced by adding a constant offset drawn from a uniform distribution to the individual trial rates. Hence, each trial was simulated with a different firing rate and the true rate variance was known. It can be seen that the FF increases linearly with σ_ν^2.

When calculated in the traditional manner, i.e. by pooling ISIs from repeated observations or trials and then applying equation 2.4, the CV^2 also increases with rate variance (darker dots in figure 2.3 b). It is however also plausible to compute the CV^2 per trial and then average over these values in a given estimation window. The effect of this can be seen in the lighter dots in figure 2.3 b. In the absence of rate variance the pooled estimates and those averaged over trial wise values are identical. However, since the offsets producing the rate variance are constant per trial, the per-trial estimates do not increase with σ_ν^2. Note that for small samples, as in the trial-wise estimation of the CV^2, it becomes important to use the unbiased estimator of the sample variance ($\sigma_s^2 = \frac{1}{n-1}\sum_i^n (x_i - \bar{x})^2$)[1]. A slight under-estimation of the theoretically expected CV^2 can still be seen in the figure. The reason for this is not entirely understood.

For renewal processes the per-trial CV^2 is hence equal to the FF at a constant rate ν. Substituting the equality

$$FF_\nu = \frac{\sigma_{c|\nu}^2}{\mu_c} = CV^2 \tag{2.15}$$

into equation 2.14 we therefore get:

$$\sigma_c^2 = \sigma_{\langle c \rangle}^2 + CV^2 \mu_c \tag{2.16}$$

Dividing by μ_c and rescaling by the width of the estimation window Δ, we obtain an expression for the absolute rate variance in units of $1/s^2$.

$$\sigma_\nu^2 = \frac{\mu_c}{\Delta^2}\left(FF - CV^2\right) \tag{2.17}$$

This rate variance is very similar to the quantity *VarCE* computed by [Churchland et al., 2011]. It is however a quantitative improvement because the point process variance has been estimated from the trial-wise CV^2 rather than by simply using the lowest possible value that avoids negative variances. In figure 2.3 c we show that the method works well in recovering the rate variance used in gamma

[1] There are several proofs for this (see e.g. [Barlow, 1989]) Intuitively this can be understood after observing that the expected value of $(x_1 - x_2)^2$ equals twice the variance of the distribution if x_1 and x_2 are independent samples. Consider $E\left[(x_1 - x_2)^2\right] = E\left[x_1^2\right] - 2E[x_1 x_2] + E\left[x_2^2\right]$ From the definition of variance it follows that $E\left[x^2\right] = \mu^2 + \sigma^2$, where μ and σ^2 are the true mean and variance of the distribution. Also, since the samples are independent, we have $E[x_1 x_2] = E[x_1]E[x_2] = \mu^2$. So $E\left[x_1^2\right] - 2E[x_1 x_2] + E\left[x_2^2\right] = \left(\mu^2 + \sigma^2\right) - 2\mu^2 + \left(\mu^2 + \sigma^2\right) = 2\sigma^2$. If we now take n pairs of samples x_a, x_b where a and b are drawn uniformly from $[1, n]$, a fraction $1/n$ of the times we have $a = b$ and the quantity $E\left[(x_1 - x_2)^2\right]$ will be zero. The remaining $1 - 1/n$ of the time it will give twice the distribution variance. We hence have to correct the estimated sample variance by a factor $\frac{n}{n-1}$.

process simulations. The performance of the method is independent of the underlying firing rate distributions.

So far we have only considered Gamma processes where the firing rate was stationary for the duration of a trial. In *vivo* this is normally not the case. In the following we therefore employ a model with variable firing rates to illustrate that the method can also cope with more realistic spike trains. The basic assumption of the model outlined in figure 2.4 is that there is some sort of stereotypical firing rate response which is related to some fictitious *experimental task*. For this purpose we have generated some arbitrary rate profile as shown in the top of the figure[2]. The simplest configuration would then be to add a constant rate offset to each trial to introduce some rate variance (model type I). The reduction in FF observed in cortical data [Rickert et al., 2009; Churchland et al., 2010] would then simply be due to a relative increase of the mean firing rate with respect to the (constant) rate variance. This scenario is depicted in the left-hand side ($l.h.s.$) of the figure. The magnitude of the rate variance has been chosen to roughly replicate the values of FF found in our data set (see section 2.3). Time resolved FF and CV^2 were calculated from 200 trials of simulated activity. The bottom panel shows that equation 2.17 performs well in recovering the simulated constant rate variance.

An alternative scenario would be that σ_ν^2 as well as the FF are modulated during the course of a trial (model type II). To simulate this we have generated trial wise rate offsets that are proportional to the inverse of the average rate profile in such a way that $\sigma_\nu^2(t)$ varied between $30[1/s^2]$ at trial onset and $15[1/s^2]$ at the time of maximum mean rate deflection. This particular method of constructing $\sigma_\nu^2(t)$ is again arbitrary and only serves the purpose of illustrating that the method can also cope reasonably well with modulated rate variances, as shown in the $r.h.s.$ bottom panel of figure 2.4.

It is interesting to note that although [Churchland et al., 2011] have stated that their *VarCE* is basically equivalent to the FF, the two models type I and II cannot be qualitatively distinguished from the time resolved estimates of FF and CV^2 as shown in the figure. Slight differences in the magnitude of the FF can hardly be identified in physiological data. The modulation in the estimated time resolved rate variance however allows a clear distinction between the two models.

Figure 2.4 has been generated with 200 trials of simulated activity. Although not an unrealistically high number, physiological data is often available only for smaller numbers of repetitions. In figure 2.5, the mean squared error ($m.s.e.$) between the theoretical and estimated rate variances is plotted versus the

[2] $\nu(t) = 25\Big[exp\Big(-(800-t)^2/25000\Big) + 0.5 * exp\Big(-(1200-t)^2/20000\Big)\Big][1/s]$

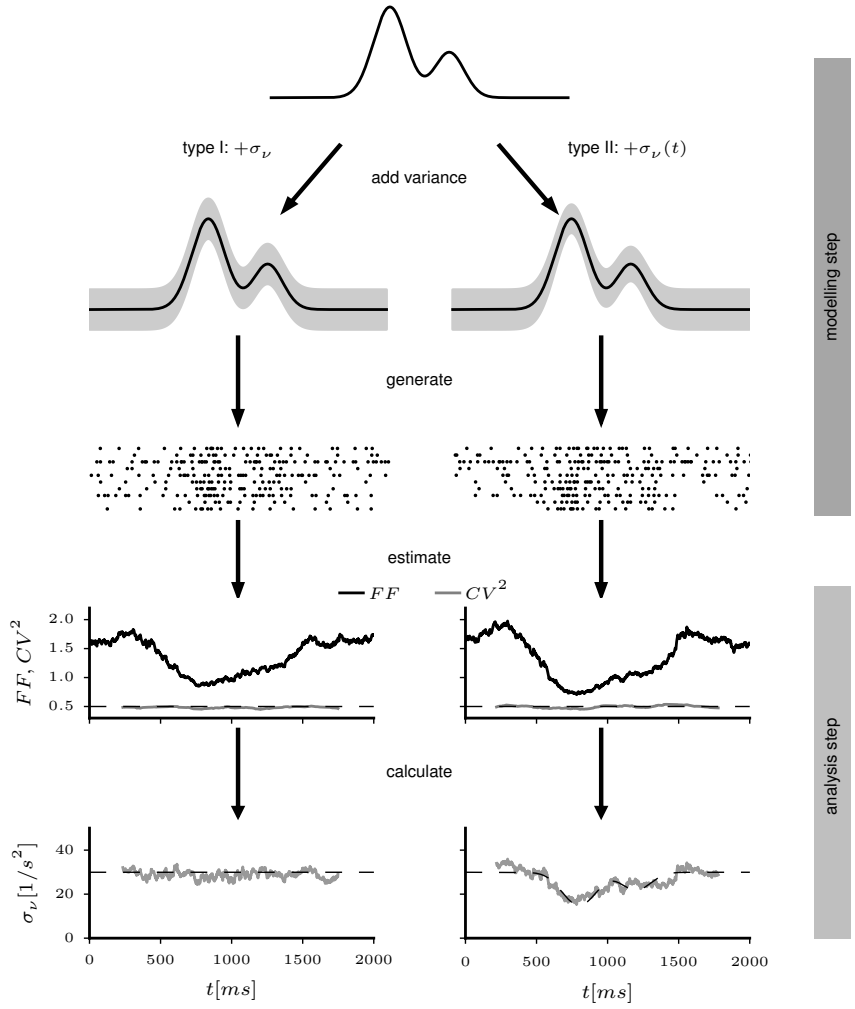

Figure 2.4: Illustration of alternative compositions of neural variability. Top panel: Stereotypical rate profile. Trial rates are generated by drawing rate offsets for each tria which are either constant ($l.h.s.$) or reduced inversely proportional to the mean rate ($r.h.s.$). 200 trials of Gamma spike trains are then generated with $\alpha = 2$. Consequently FF and CV^2 are estimated from those spike trains and finally σ_ν^2 is computed according to equation 2.17 (bottom panels). Dashed lines represent theoretically expected values.

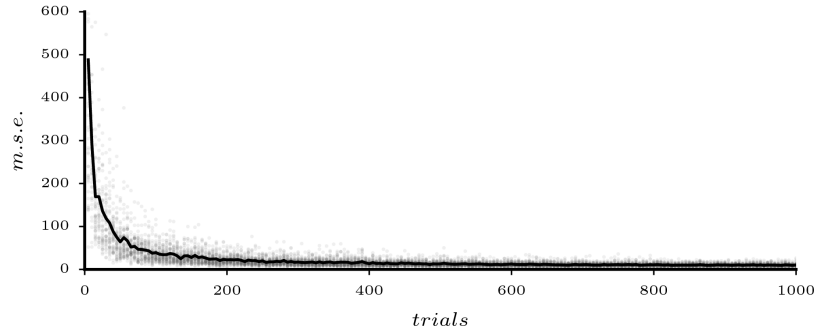

Figure 2.5: $m.s.e.$ between simulated and estimated σ_ν^2 vs the number of simulated trials for model type II in figure 2.4. Dots represent individual simulation results, black line is the average over 50 repetitions.

number of simulated trials. For small numbers of trials, the error is large but quickly decays to some asymptotic value. The average $m.s.e.$ can never go to zero because of the finite sizes of the sliding windows in which the individual statistics are computed. A trade-off exists between small estimation windows with high bias and variance of estimation and wider windows which cannot finely capture the temporal modulations of the statistics.

In the following section we will apply the methods described so far to a set of physiological data.

2.3 Dissecting Variability Components in Monkey Motor Cortex

The dataset used throughout this work has been kindly provided by Alexa Riehle's lab at Institut de Neurosciences de la Timone CNRS & Universite Aix-Marseille, France and has been previously published in [Bastian et al., 2003] and [Rickert et al., 2009].

Two Macaque monkeys were trained to perform a delayed centre-out reach task. The subjects had to press one of six buttons at the end of a delay period previous to which varying degrees of target information were given according to one of three experimental conditions. The experimental settings are explained in more detail in chapter 5. For the present purpose only data from condition 1 is used, where complete target information was given as a preparatory signal (PS), $500ms$ after trial start (TS) which was initiated by the animals. After a delay of one second, the animals were prompted to perform the movement at the response signal (RS). Correct execution was rewarded with

fruit juice. Extracellular recordings were taken from the primary and pre-motor areas. Only spike times resulting from online spike-sorting of successful trials were available to us. We applied some selection criteria to exclude artefacts which left us with 76 units for monkey 1 and 66 for monkey 2 (see appendix A for more details).

We computed the variability statistics described in the previous sections for each unit and direction separately. Figure 2.6 shows the results for a sample unit for monkey 1. At PS, the direction to reach for is indicated. The unit responds with a slight bump in the firing rate (figure 2.6a). The FF drops from a high spontaneous value to around unity, while the CV^2 stays low and relatively unmodulated throughout the trial. Since our calculation of the CV^2 requires some relatively strong assumptions we have included estimates of the L_V and CV_2 for comparison. We confirm the result of Ponce-Alvarez et al. [2010] that all three quantities are relatively constant while it roughly holds that $L_V < CV^2 < CV_2$.

The estimation of σ_ν^2 is based on the renewal process assumption which states that the CV^2 is constant. Having established that this is more or less the case in the data, we use the CV^2 as calculated on the whole duration of the trial (including rate-unwarping and bias-correction) in the calculation of the rate variance. This has the benefits of alleviating the under-estimation of the trial wise CV^2 seen in figure 2.3b as well as enabling us to calculate σ_ν^2 for the whole duration where the FF is available.

The rate variance clearly reduces during the preparatory period (PS to RS, panel e), suggesting that the simply type I model presented earlier is not sufficient to explain the dynamics. The rate peak after RS in panel a indicates that this particular unit is *most involved* in the actual movement execution. This is in accordance with the fact that the FF (panel d) as well as the rate variance (panel f) are further reduced when the trials are aligned with respect to movement onset.

Figure 2.7 shows the same analyses but averaged over all 142 neurons which passed our selection criteria. The results are qualitatively similar to the sample unit in figure 2.6. Averaging over many measurements naturally yields smoother estimates. Note that since the time-unwarping method of estimating the CV^2 results in traces of varying lengths, we show only the range where a minimum of 100 values were available for averaging. For the grand average, the FF is higher and the CV^2 lower than for the single unit shown. Consequently the average rate variance is also higher. Having many measurements available allowed us to perform a statistical test on the modulation of σ_ν^2. As reference values we defined those estimated in a the latest window fully contained in

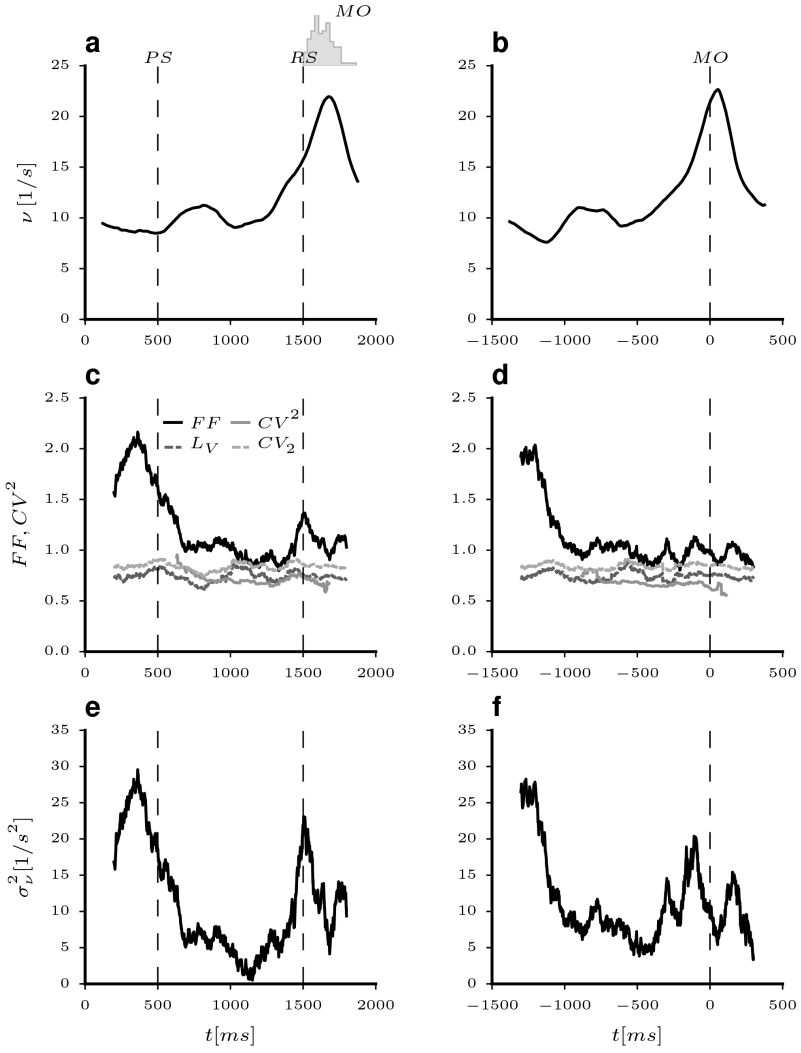

Figure 2.6: Variability components for a sample unit of monkey 1. $l.h.s.$: data aligned to TS. The histogram at the top shows the distribution of times to movement onset (MO). $r.h.s.$: data aligned to MO **a-b**) Firing rate estimates with a triangular kernel with $\sigma = 50ms$. **c-d**) FF, L_V and CV_2 estimated in $400ms$ windows, CV^2 estimated in rate-unwarped window of $10ot$. **e-f**) time resolved rate variance estimated using equation 2.17.

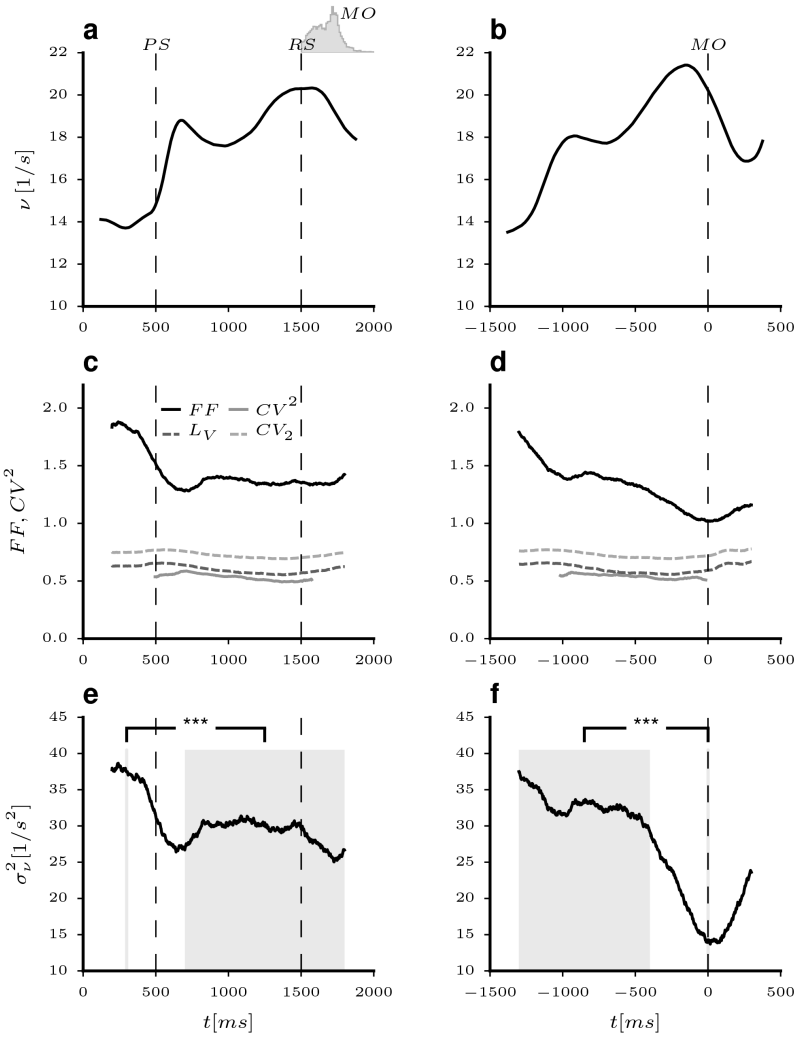

Figure 2.7: Same as figure 2.6 but averaged over all 142 units for both monkeys. Significances in σ^2_ν are from a Wilcoxon signed rank test between **e**) the last window full contained before PS and all subsequent windows not overlapping with PS and **f**) the window centred on MO and all previous windows not overlapping the window centred on MO. *** indicates $p < 0.001$ for the two-sided p-value of the test.

the interval before PS (centered around the light grey line in figure 2.7 e). A Wilcoxon signed rank test was applied between these values and those from and all subsequent time windows which did not overlap with the reference interval (figure 2.7 e, shaded area). The reduction in rate variance was highly significant for the whole range tested ($p < 10^{-3}$). The significance was even higher when the test was applied comparing the window centred around MO and those contained in the preparatory period ($p < 10^{-14}$, figure 2.7 f).

The significant reduction in rate variance at cue presentation present in the grand average suggests that the simple assumption of ongoing activity which varies so slowly that it can be modelled by a constant rate offset per trial [Nawrot, 2010] is not sufficient to explain the variability reduction. The rate variance is modulated on time-scales matching those of the behavioural task. One possible explanation could be the alignment to external experimental events. Before PS, in the spontaneous phase, the monkeys are not doing anything specific. This means that the trial-to-trial variance is computed over arbitrary network conditions and therefore expected to be high. Movement onset is not an experimentally dictated event but one determined by behaviour. It is therefore no surprise that MO-aligned spike trains yield the lowest rate variance at the time of movement initiation. The intermediate variability level in the preparatory period between initial stimulus presentation and movement onset could then be interpreted as an *intermediate misalignment* due to different reaction times over trials. This variability may be somewhat reduced if the trial wise times are linearly *stretched* so that PS and MO coincide for all trials [Riehle and Renoult, 2006].

This view is however still based on a stereotypical rate modulation superimposed onto some sort of background modulation of the network states - both of which remain unexplained by such a *model*. In recent years, some interesting observations have been made in the field of attractor dynamics in neural networks [Deco and Hugues, 2012; Litwin-Kumar and Doiron, 2012] which provide possible explanations for both the origin of rate variance and its task related reduction. The average firing rates modulations can then be interpreted as the result of probabilistic winner-less competition [Lagzi and Rotter, 2015] between competing attractors rather than as a fixed *task-related* rate profile. We will explore some aspects of those mechanisms in the following chapters.

Chapter 3

Balanced Networks of Binary Neurons

In chapter 2 we have seen that the activity of cortical neurons is variable on several time scales. The observation that highly reliable firing patterns can be induced with current injections in vitro if the statistics of the injected currents are matched to those of synaptic input in cortex [Mainen and Sejnowski, 1995] raises the question how such seemingly chaotic firing can be induced in networks of deterministic units. All sensory input is intrinsically stochastic because all stimulus modalities are either thermodynamic or quantum mechanical in nature [Faisal et al., 2008]. This sensory noise can however only account for a small part of variability in the brain as neural responses become more variable with increasing distance from sensory input stages [Kara et al., 2000].

In cortex, individual neurons receive large numbers of excitatory and inhibitory synaptic inputs from the surrounding network. It has been shown that a condition exists in which these neurons fire in a chaotic manner at low firing rates. This condition was termed the *Balanced State* and occurs if excitation and inhibition to each cell cancel each other on average so that spike emission is triggered by fluctuations in the input rather than its mean [van Vreeswijk and Sompolinsky, 1996, 1998; Brunel, 2000]. Using networks of binary units, van Vreeswijk and Sompolinsky [1996] showed that this dynamic equilibrium occurs without much fine tuning of the parameters if a few conditions are met. Brunel [2000] characterised different types of firing activity in networks of leaky integrate-and-fire neurons (LIFs) as a function of the strength of an external drive to the network and the relative strength of excitation and inhibition.

Depending on the parameters, balanced networks can produce asynchronous firing with CVs of unity and above [Brunel, 2000]. If synaptic strength is in-

creased above some critical value, bursty spiking with even higher variability can be achieved [Ostojic, 2014]. Unstructured balanced networks can however not capture the firing rate variance observed in spontaneous cortical recordings (see chapter 2). Although it is possible to introduce variable firing rates through the external drive, the experimentally observed suppression of this variance during stimulus presentation would then also have to come from outside the network.

More recently it has been shown in a number of studies [Deco and Hugues, 2012; Litwin-Kumar and Doiron, 2012; Doiron and Litwin-Kumar, 2014; Mazzucato et al., 2015] that competition between attractors can induce rate variance in balanced networks. Attractors were introduced into the networks by defining sub-assemblies or *clusters* in the excitatory populations of the networks and by increasing the synaptic efficacies between units inside clusters relative to those to the remaining units. If the ratio of intra-cluster weights to inter-cluster weights is low, no change in the dynamics occurs. If the ratio is too high, the attractors become too *deep*, resulting in winner-take-all dynamics where one cluster has a high firing rate and suppresses the activity in the other assemblies [Lagzi and Rotter, 2015]. In an intermediate range, the variance in the in the population firing rates causes the networks to *switch* between states where different clusters have higher firing rates. This results in a scenario where individual units exhibit multi-stability in their firing rates and as a results introduce variance in firing rates that increases the trial-to-trial variability. In addition, selective stimulation of subsets of clusters causes certain attractors to become more stable, which in turn quenches the switching-dynamics and thereby the FF.

One problem with this family of models is that the active clusters tend to have firing rates close to saturation, where firing becomes very regular. Also, the range of synaptic strength ratios in which state-switching can occur is quite narrow.

In the present chapter we will analyse the dynamics of cluster competition and investigate possible improvements to the model. We will first recapitulate the conditions of the balanced state using a simplified binary neuron model for which and extensive mean field theory has been described [van Vreeswijk and Sompolinsky, 1998; Renart et al., 2010]. We will then use the mean field approach to analyse the attractors of clustered networks and show that the introduction of inhibitory clustering can moderate the firing rates of the active clusters.

We first consider networks of binary units which capture some of the features of spiking neurons while being easier to analyse. The networks consist of N_E excitatory and N_I inhibitory units with asynchronous updates. Unless stated

otherwise, we model a total of $N = 5000$ units with $N_E = 4N_I$. In each simulation step t, one unit is randomly chosen and its state $\sigma \in 0, 1$ is updated according to the rule

$$\sigma_i(t+1) = \Theta\left(\sum_{j=1}^{N} J_{ij}\sigma_j(t) - \theta_i + J_{iX}m_X\right), \tag{3.1}$$

where Θ is the Heaviside step function, Jij is the synaptic weight between pre-synaptic unit j and post-synaptic unit i, θ_i is the threshold and m_X is the rate of some external drive to the unit which is modelled as a constant (rather than a spike source) weighted by J_{iX}. If a unit is updated to the *up*-state ($\sigma = 1$) it remains in that state until the next update. Hence the network has an integration time scale τ equal to the average time between updates. Since time is not explicitly modelled, we assign a value of $\tau = 10ms$ similar to neuronal membrane time constants for illustration purposes only. Note that τ is proportional to N.

The connection strengths $J^{ij}_{\alpha\beta}$ from unit j in population β to unit i in population α ($\alpha, \beta \in E, I$) are $J_{\alpha\beta}$ with probability $p_{\alpha\beta}$ and zero otherwise. Uniform connection probabilities between 1 and 20 % are commonly used in the literature (e.g. [Brunel, 2000; Renart et al., 2010; Ostojic, 2014; Kriener et al., 2014; Litwin-Kumar and Doiron, 2014]). To allow a comparison to other clustered network studies we here adopt the approach taken by Litwin-Kumar and Doiron [2012] and Mazzucato et al. [2015] and set the connection probability of excitatory to excitatory units to $p_{EE} = 0.2$ and all those involving the inhibitory population to $p_{EI} = p_{IE} = p_{II} = 0.5$.

Conditions for the values of the remaining model parameters arise from an analysis of the balanced state using a mean field description of the network dynamics.

3.1 Mean Field Description and the Balanced State

The conditions for the balanced state and its stability arising from mean field considerations are reproduced from van Vreeswijk and Sompolinsky [1998] and Renart et al. [2010]. We do however guide the reader through some of the derivations as the references do not show them for the specific types of networks used here.

Since the balanced state requires that spiking is fluctuation driven, the mean inputs to each unit need to cancel while the variance has to be on the order of the thresholds. Hence the synaptic strengths are scaled with network

size as $J_{\alpha\beta} = j_{\alpha\beta}/\sqrt{N}$ where $j_{\alpha\beta}$ is a constant. Also, the number of input spikes required to reach the threshold needs to be small. We adopt the scaling used in [van Vreeswijk and Sompolinsky, 1998] so that $\sqrt{K}$ excitatory spikes arriving during one time constant suffice to elicit a postsynaptic one, where $K_{\alpha\beta} = p_{\alpha\beta}N_\beta$ is the average number of connections a unit in population α receives from population β. This condition implies:

$$\sqrt{p_{\alpha E}N_E}J_{\alpha E} = \theta_\alpha \tag{3.2}$$

To achieve balance between excitation and inhibition, we need the excitatory and inhibitory inputs to each population to cancel. If excitatory and inhibitory population rates are equal, this means:

$$0 = N_E p_{EE} J_{EE} + \frac{1}{g} N_I p_{EI} J_{EI} \tag{3.3}$$

$$0 = N_E p_{IE} J_{IE} + N_I p_{II} J_{II} \tag{3.4}$$

Here we have introduced a factor g to control the relative strength of excitation and inhibition to accommodate external inputs. Combining equations 3.2 and 3.3, we can now compute the excitatory weights as:

$$j_{EE} = \frac{\theta_E}{\sqrt{p_{EE}n_E}} \tag{3.5}$$

$$j_{EI} = -g j_{EE} \frac{p_{EE}n_E}{p_{EI}n_I} \tag{3.6}$$

where $n_E = N/N_E$. Similarly, for the inhibitory population:

$$j_{IE} = \frac{\theta_I}{\sqrt{p_{IE}n_E}} \tag{3.7}$$

$$j_{II} = -j_{IE} \frac{p_{IE}n_E}{p_{II}n_I} \tag{3.8}$$

For large N, the central limit theorem allows the treatment of the synaptic input to each unit as a Gaussian random variable. The dynamics of the population rates $m_\alpha(t)$ in networks of asynchronously updated binary units can be described as [van Vreeswijk and Sompolinsky, 1998; Renart et al., 2010]:

$$\tau_\alpha \frac{d}{dt} m_\alpha(t) = -m_\alpha(t) + \mathrm{H}\left(-\frac{\mu_\alpha(t)}{\sqrt{s_\alpha^2(t)}}\right) \tag{3.9}$$

Here, the population *activity*-rate is defined as the average of the instantaneous states $\sigma \in [0, 1]$ in population α, $m_\alpha(t) = \langle \sigma_\alpha(t) \rangle$ and H is the complementary

error function

$$\mathrm{H}(z) = \frac{1}{\sqrt{2\pi}} \int_z^\infty dx\, \mathrm{e}^{-\frac{x^2}{2}} \tag{3.10}$$

Note that the m_α is not equivalent to the firing rate ν_α because spikes are only counted when units update their state from 0 to 1. The population time constant τ_α corresponds to the average time between successive updates of the states. μ_α and s_α^2 are the mean and variance of the input to population α. The population average input is:

$$\mu_\alpha(t) = \sum_\beta \bar{J}_{\alpha\beta} m_\beta(t) + J_{\alpha X} m_X - \theta_\alpha \tag{3.11}$$

where $\bar{J}_{\alpha\beta} = j_{\alpha\beta} p_{\alpha\beta} n_\beta \sqrt{N}$ is the average weight from population β to α. Neglecting correlations between fluctuations in the input currents to the populations the variance of the input is determined by the variance in the weights $\bar{J}^{(2)}_{\alpha\beta}$ [van Vreeswijk and Sompolinsky, 1998; Renart et al., 2010].

$$s_\alpha^2(t) = \sum_\beta \bar{J}^{(2)}_{\alpha\beta} m_\beta(t) \tag{3.12}$$

For constant weights, $\bar{J}^{(2)}_{\alpha\beta}$ is determined by the stochasticity in the connectivity so $\bar{J}^{(2)}_{\alpha\beta} = p_{\alpha\beta}\big(1 - p_{\alpha\beta}\big) j^2_{\alpha\beta} n_\beta$.

From equation 3.9 it can be seen that the steady state rates of the populations are:

$$m_\alpha = \mathrm{H}\left(-\frac{\mu_\alpha}{\sqrt{s_\alpha^2}} \right) \tag{3.13}$$

However, in the balanced state both excitatory and inhibitory inputs to each unit are much larger than the threshold. To obtain finite firing rates, excitation and inhibition therefore need to balance [van Vreeswijk and Sompolinsky, 1998]:

$$\bar{J}_{\alpha E} m_E + \bar{J}_{\alpha I} m_I + J_{\alpha X} m_X = 0 \tag{3.14}$$

We can hence deduct the population rates in the balanced state without solving equation 3.13:

$$m_E = \frac{\big(J_{EX}\bar{J}_{II} - J_{IX}\bar{J}_{EI}\big)}{\bar{J}_{EI}\bar{J}_{IE} - \bar{J}_{EE}\bar{J}_{II}} m_X \tag{3.15}$$

$$m_I = \frac{\big(J_{EX}\bar{J}_{IE} - J_{IX}\bar{J}_{EE}\big)}{\bar{J}_{EE}\bar{J}_{II} - \bar{J}_{EI}\bar{J}_{IE}} m_X \tag{3.16}$$

Using the definition of $\bar{J}_{\alpha\beta}$ and the expressions for the weights in equations 3.5 through 3.8, we can express the balanced rates in terms of the network

parameters and the strength of the external input:

$$m_E = \frac{m_x}{\sqrt{N_E}(g-1)}\left(\frac{J_{EX}}{\theta_E\sqrt{p_{EE}}} - g\frac{J_{IX}}{\theta_I\sqrt{p_{IE}}}\right) \tag{3.17}$$

$$m_I = \frac{m_x}{\sqrt{N_E}(g-1)}\left(\frac{J_{EX}}{\theta_E\sqrt{p_{EE}}} - \frac{J_{IX}}{\theta_I\sqrt{p_{IE}}}\right) \tag{3.18}$$

For m_E and m_I to be positive and finite and assuming that $\theta_E = \theta_I$ we thus require either

$$g < 1, \frac{J_{EX}}{J_{IX}} < g\frac{\sqrt{p_{EE}}}{\sqrt{p_{IE}}} \tag{3.19}$$

or

$$g > 1, \frac{J_{EX}}{J_{IX}} > g\frac{\sqrt{p_{EE}}}{\sqrt{p_{IE}}} \tag{3.20}$$

If either equation 3.19 or 3.20 is satisfied, a fixed point with finite rates is ensured.

To asses the stability of fixed points we need to compute the partial derivatives of equation 3.9 with respect to m_β,

$$\frac{\partial}{\partial m_\beta}\left(\frac{dm_\alpha}{dt}\right) = -\frac{1}{\tau_\alpha}\left(\frac{\partial m_\alpha}{\partial m_\beta} + \mathrm{H}'\left(-\frac{\mu_\alpha}{s_\alpha}\right)\frac{\bar{J}_{\alpha\beta}s_\alpha - \frac{1}{2}\mu_\alpha\bar{J}^{(2)}_{\alpha\beta}s_\alpha^{-1}}{s_\alpha^2}\right) \tag{3.21}$$

where

$$\mathrm{H}'(x) = -\frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \tag{3.22}$$

The stability of a fixed point depends on the ratio of population time constants [van Vreeswijk and Sompolinsky, 1998]. If we denote the partial derivatives in equation 3.21 evaluated at a fixed point $\vec{m}_0$ as $\mathrm{f}_{\alpha\beta}\big|_{\vec{m}_0}$ we can write the stability matrix S as:

$$S = \begin{bmatrix} \mathrm{f}_{EE}|_{\vec{m}_0} & \mathrm{f}_{EI}|_{\vec{m}_0} \\ \mathrm{f}_{IE}|_{\vec{m}_0} & \mathrm{f}_{II}|_{\vec{m}_0} \end{bmatrix} \tag{3.23}$$

.

This matrix has eigenvalues $\lambda_{1,2} = \frac{1}{2}\left(\mathrm{T}_S \pm \sqrt{\mathrm{T}_S^2 - 4\,\delta_S}\right)$. Here, T_S and δ_S are the trace and determinant of S. If we define $\mathrm{f}'_{\alpha\beta} = \mathrm{f}_{\alpha\beta}\big|_{\vec{m}_0}\tau_\alpha$, and set $\tau_E = 1$, we can make the dependence of the stability on the inhibitory time constant explicit and write:

$$\mathrm{T}_S = \mathrm{f}'_{EE} + \mathrm{f}'_{II}/\tau_I \tag{3.24}$$

$$\delta_S = \frac{1}{\tau_I}\left(\mathrm{f}'_{EE}\,\mathrm{f}'_{II} - \mathrm{f}'_{EI}\,\mathrm{f}'_{IE}\right) \tag{3.25}$$

The eigenvalues are hence negative and real if the following conditions are fulfilled:

$$T_S < 0 \tag{3.26}$$

$$|T_S| > \sqrt{T_S^2 - 4\delta} \tag{3.27}$$

$$T_S^2 \geq 4\delta \tag{3.28}$$

Condition 3.26 is satisfied if $\tau_I < -\frac{\mathrm{f}'_{II}}{\mathrm{f}'_{EE}}$. Equation 3.27 is satisfied as long as $g > 1$. Inserting 3.24 and 3.25 into 3.28 results in a quadratic equation in τ_I. The System consequently has three bifurcations at the critical time constant ratios:

$$r_1 = A - \sqrt{A^2 - B^2} \tag{3.29}$$

$$r_2 = -B \tag{3.30}$$

$$r_3 = A + \sqrt{A^2 - B^2} \tag{3.31}$$

with

$$A = \frac{\mathrm{f}'_{EE}\,\mathrm{f}'_{II} - 2\,\mathrm{f}'_{EI}\,\mathrm{f}'_{IE}}{(\mathrm{f}'_{EE})^2} \tag{3.32}$$

$$B = \frac{\mathrm{f}'_{II}}{\mathrm{f}'_{EE}} \tag{3.33}$$

The values of those ratios depend on the specific parameters and their effect on the fixed point is illustrated in figure 3.1a.

For $\tau_I/\tau_E < r_1$, both eigenvalues are real and negative and the fixed point is a stable node, i.e. all trajectories in its vicinity converge directly towards it along the eigenvector corresponding to the largest eigenvalue. Above r_1 the eigenvalues become complex and the activity rates show damped oscillations towards the fixed points. When $\tau_I/\tau_E > r_2$, the fixed point becomes unstable and the firing rates escape towards an oscillatory limit cycle with large amplitude (see figure 3.2b). In simulations, different time constants for the populations are achieved by scaling the probability $P_{U\alpha}$ that a unit from population α is updated so that $P_{UE}/P_{UI} = \tau_I/\tau_E$. The update rule results in exponentially distributed intervals between updates (see figure 3.1 b and c). A consequence of this fact is that interval statistics of spikes generated from this model are difficult to interpret, as the update process is already Poissonian. An alternative possibility would be to update every unit sequentially and then shuffle the update list before the next round. This results in update distributions as shown in a lighter shade in figure 3.1 b and c. In this case however, the mean field description is no

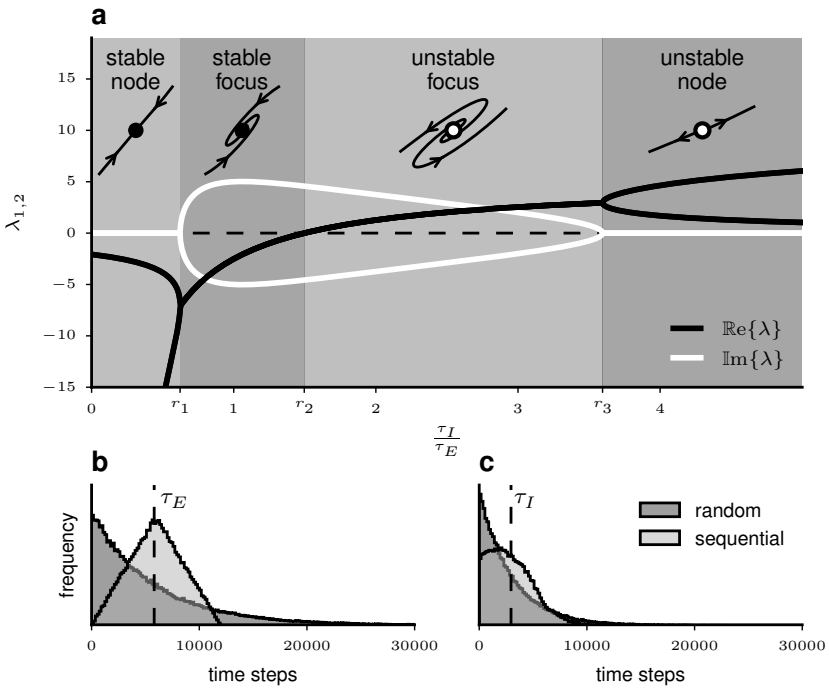

Figure 3.1: **a**) Dependence of eigenvalues of the stability matrix at the fixed point on the ratio of inhibitory and excitatory time constants. **b**) Distribution of inter-update-intervals for the excitatory units of a network with $N_U = 1$ for random and sequential updates. **c**) same as **b** for inhibitory population. Remaining parameters are given in table 3.1.

longer accurate (see figure 3.2 d).

The mean field theory description portrayed above shows good agreement with network simulation for moderate values of N. Figure 3.2 shows some characteristic examples for the parameters given in table 3.1. The activity rates for an excitatory and an inhibitory population are plotted against each other and arrows represent the derivatives at sample points in phase space. In panel a, the stability condition in equation 3.20 is met and $\tau_I/\tau_E < r_1$. The fixed point is hence a stable node and the simulated network rates behave as predicted by mean field theory and follow the flow field directly to the fixed point. In panel b, $\tau_I/\tau_E > r_2$ and the rates cycle through large amplitude oscillations. The match between mean field theory and simulation is also good when conditions 3.19 and 3.20 are violated (panel c). When the units are updated sequentially, the mean field description breaks down. Although the predicted steady state rates are eventually reached, the transient trajectory of the rates differs from the theoretical prediction.

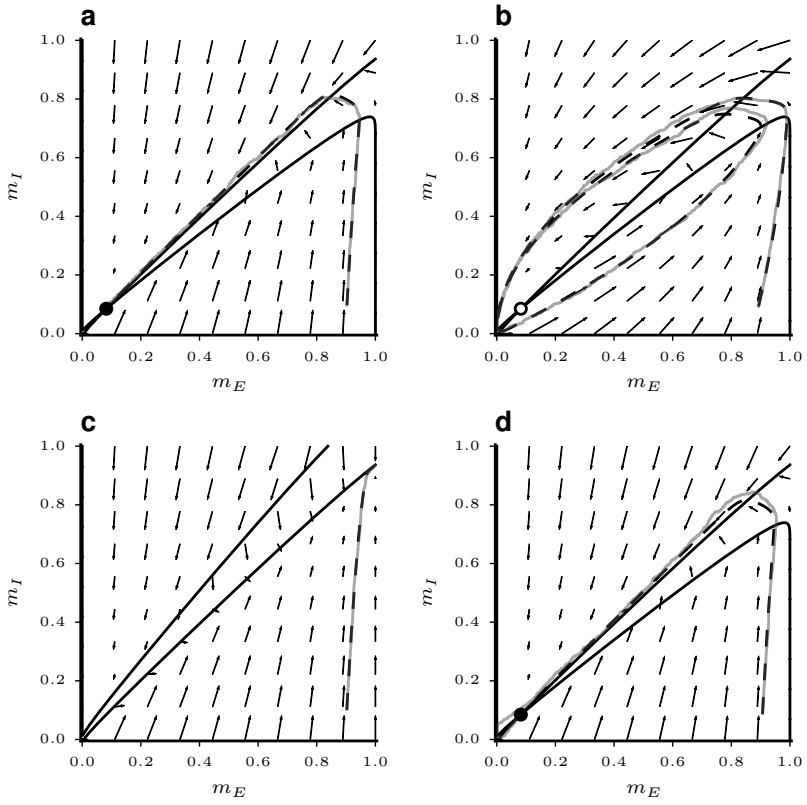

Figure 3.2: Comparison of mean field description and simulations for networks with $N_E = 4000$ illustrating various parameter settings. Dashed trajectories are mean field theory, grey traces are network simulations. Arrows represent derivatives ofequation 3.9. Nullclines are drawn as solid black lines. Filled (empty) circles indicate stable (unstable) fixed points. **a**) $g = 1.2$, $\tau_I/\tau_E = 0.5$ **b**) $g = 1.2$, $\tau_I/\tau_E = 2$,**c**) $g = 0.8$ **d**) sequential updates

Unless stated otherwise, we use the parameters summarised in table 3.1 throughout.

3.2 Clusters in the Excitatory Population

Having established the conditions for the balanced state, we turn to introducing clustered connectivity in the excitatory population. Amit and Brunel [1997] modelled working memory and persistent activity in attractor networks but did not consider variability dynamics. Their model consisted of an unstructured background population and a number of attractor assemblies which were formed by increasing the weights between units belonging to the same assembly by a factor J_+ while across-cluster weights were decreased by a factor J_- to

Parameter	Value
N	$4000(E), 1000(I)$
θ	1
τ_I	$0.5\tau_E$
p_{EE}	0.2
p_{EI}, p_{IE}, p_{II}	0.5
g	1.2
J_{EX}	$\sqrt{p_{EE}N_E}$
J_{IX}	$0.8\sqrt{p_{EE}N_E}$
m_X	0.03

Table 3.1: Summary of parameters used in the binary network simulations

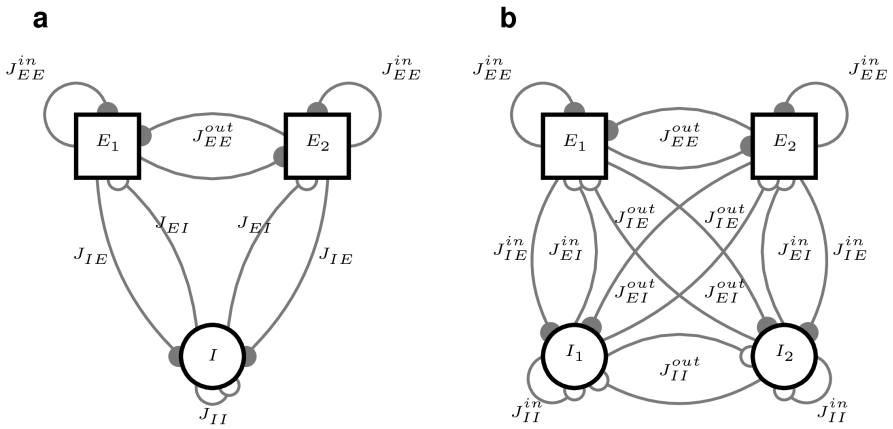

Figure 3.3: Network architecture for $Q = 2$ showing all connection types. **a**) EE-clustering: Two excitatory assemblies and a single unstructured inhibitory population. **b**) EI-clustering: Each excitatory cluster has an associated inhibitory population.

maintain overall balance. A similar approach was taken by Mazzucato et al. [2015] and Deco and Hugues [2012] although the latter did not explicitly model the background or inhibitory populations. [Litwin-Kumar and Doiron, 2012] on the other hand increased the synaptic strength as well as the connection probabilities within clusters.

Since a background population is not necessary for the clustering effects (see figure 3.5) and a simultaneous adjustment of synaptic strengths and connection probabilities adds unnecessary confusion we choose in the present work to divide the excitatory population into Q equally sized clusters with uniform connection probabilities. Connections between units in the same cluster are multiplied by a factor $J_+ > 1$ and to maintain a balance of weights, connections

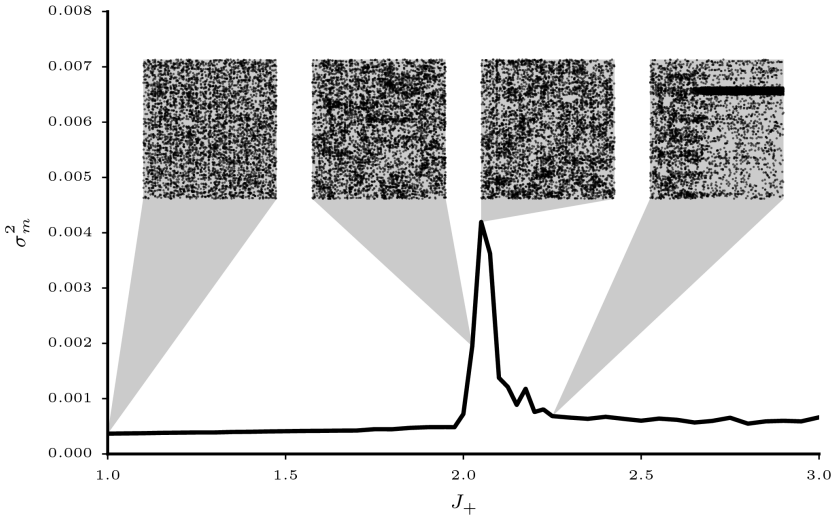

Figure 3.4: Average variance of the instantaneous mean cluster activity rates σ_m^2 of 20 trials of $1000ms$ duration for different values of the excitatory cluster strength J_+. The line shows the average over 20 network realizations. Insets show sample trials of spiking activity.

between units belonging to different clusters are multiplied by a factor

$$J_- = \frac{Q - J_+}{Q - 1} \tag{3.34}$$

Consequently $J_+ = 1$ leads to homogeneous connectivity while at $J_+ = Q$ the populations are completely decoupled. A schematic depiction of a network with $Q = 2$ excitatory clusters and all occuring connections between populations is given in figure 3.3 a.

Due to the random update process, the FF of spike trains from binary networks has no meaningful interpretation. When the activity rates saturate, i.e. when units fire each time they are updated, the FF is unity while for lower rates $FF > 1$. To quantify the effect of clustering on the network dynamics we therefore calculate the instantaneous variance in the mean cluster activity rates and average this quantity over time and clusters (σ_m^2). Figure 3.4 shows this variance for different values of the cluster strength J_+ for networks with $Q = 20$ clusters. At low values of J_+, the dynamics are not influenced by the clustered connectivity and σ_m^2 does not change much compared to the unstructured case at $J_+ = 1$. Around $J_+ \sim 2$, there is a sharp increase in rate variance. This is due to random activations of individual clusters in a winnerless competition regime, as can be seen in the more structured looking raster plots in figure 3.4.

After a sharp peak σ_m^2 quickly drops again as clustering becomes so strong that clusters tend to remain active for increasingly long times, effectively producing winner-take-all dynamics (right-most raster plot). Note that the location of the peak with respect to J_+ depends on the number of clusters as well as on the size of the network [Litwin-Kumar and Doiron, 2012]

To gain a better understanding of the underlying mechanisms, we employ the mean field approach to examine the stationary rate points of the clustered network. For this purpose, for each of the Q clusters as well as for the I-population an activity rate equation (eq. 3.9) is formed. Similar to the method described in [Mazzucato et al., 2015], we then solve the resulting system of equations for the stationary states and check for stability as described in section 3.1. Numerically solving such a multi-dimensional system of coupled differential equations requires that the initial guess for the solution is close to a fixed point. To sample the space of possible rate configurations we therefore initialised the rates randomly between 0 and 1 and then integrated the system for a number of time steps before finding the exact fixed points using the Nelder-Mead simplex algorithm implemented in `scipy` [Jones et al., 01]. This process was repeated many times to ensure that most of the existing fixed points will be found. Note that the method can only find stable solutions of the system.

The resulting stable fixed points for different network structures are shown in figure 3.5. The solid lines with high rates indicate the rates of the clusters in the active states while the lower solid lines represent the rates of the remaining populations. The numbers indicate how many populations occupy a state simultaneously. The activities of the inhibitory populations are drawn as dashed lines and the dotted line represents the homogeneous state where all populations fire at the same rate. To justify our choice of clustering method, the figure includes the attractor landscapes for networks with 10% of the units not belonging to any cluster (i.e. background population) as in [Mazzucato et al., 2015] (panel a), and for clustering of the connection probabilities as in [Litwin-Kumar and Doiron, 2012] (panel b). The clustering parameter $R_{EE} = p_{in}/p_{out}$ quantifies the ratio between connection probabilities within clusters to those across clusters. The results are not qualitatively different from networks without background population and cluster independent connection probabilities (panel c). As the cluster strength increases, more stable states occur with increasing numbers of simultaneously active clusters, resulting in multistability of the rates. At some critical value, the base-state in which all populations share the same low firing rate disappears. The common property of all three cases is that the *up*-states, i. e. the rates of the active clusters, show high activity rates and quickly approach the saturation value as the cluster parameter increases.

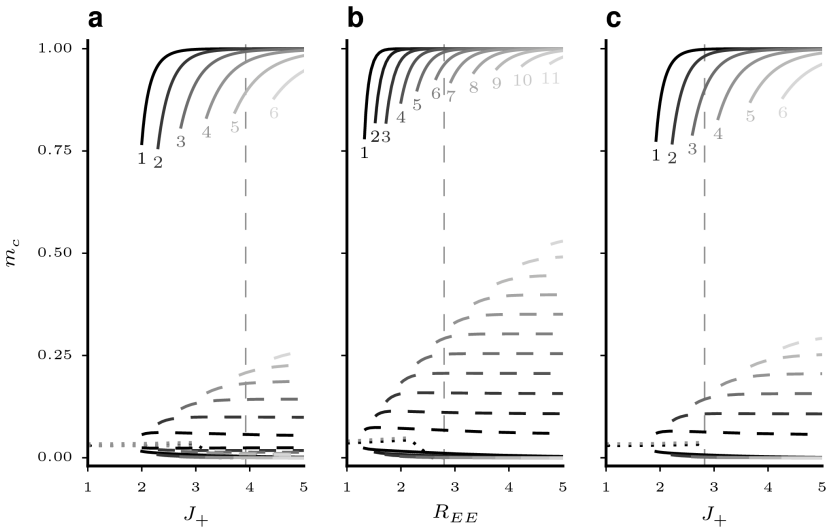

Figure 3.5: Sampled stable fixed points of the mean field equations for networks with 20 excitatory clusters versus the clustering parameter. **a**) Clustering by weight increase with 10% of the E-units as an unstructured background population. **b**) Clustering by increase in connection probability and synaptic strength. **c**) Clustering by weight increase of the entire E-population. Vertical dashed lines correspond to the clustering strength where the homogeneous state is no longer stable.

It is evident in figure 3.5 that the state with the highest activity rate is always that with a single active cluster. That means that this rate forms an upper bound for the active cluster rates. We therefore carry on our analysis of the cluster dynamics by solving only for those cases. This is achieved by constraining $Q-1$ of the cluster populations to have equal rates. Also, since the above described method of random sampling of the activity space is costly and can yield only stable fixed points we employ a more systematic procedure for analysing the dynamics.

For single population models, the fixed points of neural activity can be found graphically by plotting the neurons' gain function and the firing rates against the synaptic input and finding the intersections of the two lines [Gerstner et al., 2014]. When the input to the gain function depends not only on one population rate, i.e. when there are several coupled differential equations, the approach is no longer feasible. Mascaro and Amit [1999] describe an effective response function (EFR) approach for multi-population models which puts one or more populations in focus while still incorporating the full dynamics of the remaining populations.

For a network model with P populations, the individual population rates can

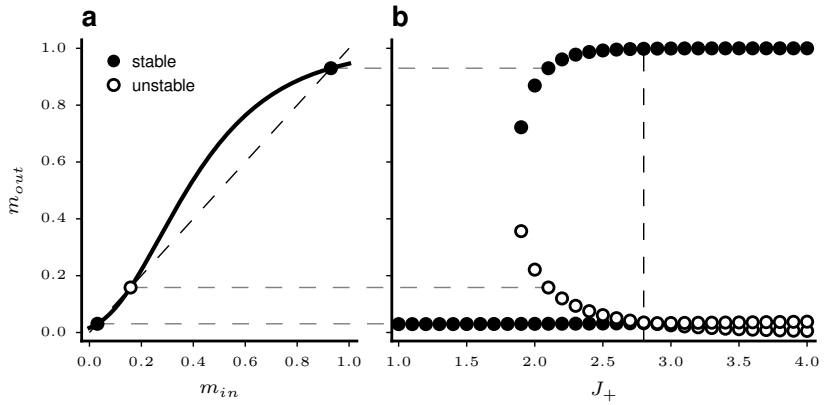

Figure 3.6: EFRs for states with a single active cluster in networks with excitatory clustering. Filled/empty circles represent stable/unstable fixed points.**a**) Full EFR for $J_+ = 2.1$ **b**) Fixed points of EFRs versus cluster strength J_+.

be expressed as functions of all the population rates in the network.

$$\begin{aligned} m_1 &= \Phi_1(m_1, m_2 \ldots, m_P) \\ m_2 &= \Phi_2(m_1, m_2 \ldots, m_P) \\ &\vdots \\ m_P &= \Phi_P(m_1, m_2 \ldots, m_P) \end{aligned}$$

For the present case Φ takes the form of equation 3.13. The EFR-approach works by treating the rate of a focus population as a parameter. I.e. we fix $m_1 = \bar{m}_1$ and solve the $P-1$ equations for the remaining rates.

$$\begin{aligned} m_2 &= \Phi_1(\bar{m}_1, m_2 \ldots, m_P) \\ &\vdots \\ m_P &= \Phi_P(\bar{m}_1, m_2 \ldots, m_P) \end{aligned}$$

The solution $m'(\bar{m}_1)$ to those equations will drive the rate of the focus population to a value m_{1out} given by

$$m_{1out} = \Phi_1\big(\bar{m}_1, m'(\bar{m}_1)\big) = \Phi_{\text{eff}}(\bar{m}_1) \tag{3.35}$$

Mascaro and Amit [1999] called the resulting input-output relation for the focus population the EFR. When $m_{1out} = \bar{m}_1$, i.e. when the EFR intersects the diagonal, m_{1out} is a fixed point of the system. If the slope of the EFR at the

intersection is larger than unity, the fixed point is unstable. For slopes smaller than unity, the fixed point is stable for m_{1out} *given* $m'(\bar{m}_1)$. Since this is a one dimensional representation of potentially multi dimensional systems and since the stability of fixed points depends additionally on the ratios of population time constants which are not captured by the EFR, those points are not generally globally stable for the whole system [Mascaro and Amit, 1999]. We therefore assess the stability of fixed points by examining the eigenvalues of the stability matrix.

Figure 3.6 illustrates the EFR for $Q = 20$ clusters and $q = 1$ unconstrained cluster (corresponding to the line labelled 1 in figure 3.5 c). Panel a shows the full EFR for $J_+ = 2.1$ where equation 3.35 has been evaluated on a dense grid for the whole range of activity rates. This representation reveals an additional unstable fixed point between the low rate attractor where all excitatory populations have the same firing rate and the *up*-state of the focus population. In panel b, the EFR has been calculated for different cluster strengths and only the fixed points of the system are plotted versus J_+. At $J_+ = 1$, the EFR is simply a flat line. I.e. an increase in m_{in} has no significant effect on m_{out}. As J_+ increases, the self amplification of the focus population causes m_{out} to increase with m_{in} until the EFR touches the diagonal. At this point ($J_+ \sim 1.8$) a stable *up*-state with an intermediate unstable fixed point arises.

To switch between *up*-states of different populations, the variance in the population activities has to be sufficient to cross the unstable fixed point. It can be seen in panel b that this becomes increasingly less likely as the separation between fixed points widens with increasing J_+. The peak of σ_m^2 in figure 3.4 occurs in the narrow range where stable *up*-states exist at rates below the saturation limit. A potential solution to the problem of high firing rates could be to simply increase the strength of inhibition. Changing the relative inhibition parameter g has however not yielded qualitatively different attractor structures. Since inhibition is global, an increase in firing rate of a single population leads only to a small rise inhibitory activity, so that the self excitation of the active cluster is not balanced and the rate saturates.

We therefore propose, that to reduce the firing rates of active clusters it is necessary that inhibition is also cluster specific so that a rate increase in an excitatory cluster is *balanced* by a corresponding inhibitory population. This idea will be explored in the following section.

3.3 Clustering E and I Populations

We have seen in the previous section that excitatory clusters in networks with global inhibition lead to rate saturation in active clusters which impedes state switching because the attractors of different cluster states are very far apart. We will now show how this problem can be overcome by introducing structure in the inhibitory connections as well. Litwin-Kumar and Doiron [2012] briefly describe how clustering the inhibitory units leads to stimulation-induced variability reduction in the inhibitory units. For this purpose, they segmented the inhibitory population into an equal number of clusters and assigned one inhibitory cluster to each excitatory one. The variability modulation in the inhibitory units was then induced by increasing the connection probabilities (p_{IE}) from an excitatory cluster to its inhibitory counterpart with respect to the remaining I-units. Although Litwin-Kumar and Doiron [2012] did not address the effect of this alteration on the firing rates of the active clusters, it is clear that it is not sufficient for our present purpose.

We require that an E-population selectively excites its corresponding I population which in turn selectively inhibits the E units. It is therefore necessary to close the loop and cluster both the EI and IE synapses. Like the E-population, the inhibitory units are equally divided into Q clusters, resulting in a total of $2Q$ populations. We rename the clustering factor for the excitatory population as J_{E+}. For simplicity, all connections involving the inhibitory population are lumped into a single cluster parameter J_{I+}. Balance is again maintained by rescaling across-cluster connections according to equation 3.34 so that the average row sum remains constant in each quadrant of the connectivity matrix. An overview over the possible connections is given in figure 3.3 b for $Q = 2$, where within/across cluster connections are denoted by the superscript in/out. So we have $J^{in}_{EE} = J_{E+}J_{EE}, J^{out}_{EE} = J_{E-}J_{EE}$ and $J^{in}_{\alpha\beta} = J_{I+}J_{\alpha\beta}, J^{out}_{\alpha\beta} = J_{I-}J_{\alpha\beta}$ for $\alpha\beta \in (EI, IE, II)$.

Since we have seen before that the highest *up*-state rates (m_{up}) are always reached by a single active cluster, we again constrain the population equations so the $Q - 1$ excitatory as well as their corresponding inhibitory populations have the same rate, resulting in a total of four distinct equations to solve. To examine the effect of inhibitory clustering, we start by keeping J_{E+} fixed at 2.9 (the first point in figure 3.6 where the homogeneous state is unstable) and then increase J_{I+}. The resulting fixed points are shown in figure 3.7 a.

For low values of J_{I+}, m_{up} is again at the saturation limit. As the inhibitory cluster strength increases however, the active cluster rates decrease and the stable homogeneous state reappears. When J_{I+} increases further a bifurcation

occurs and the active cluster state no longer exists. This can be understood intuitively. An increase in J_{I+} strengthens the coupling between the active cluster and its corresponding inhibitory population. When the coupling becomes too strong this selective inhibition prevents the focus population from attaining higher firing rates. It can therefore be concluded that low firing rates in active clusters can be obtained if the inhibitory cluster strength is present but smaller than the excitatory parameter.

Having established that the inhibitory clustering needs to be weaker than that of the E population, we introduce a proportionality factor R_J, so that

$$J_{I+} = 1 + R_J(J_{E+} - 1) \tag{3.36}$$

I.e. when $R_J = 0$ the inhibitory connections are un-clustered and for $R_J = 1$ we have $J_{E+} = J_{I+}$. Having defined a relationship between the excitatory and inhibitory cluster parameters, we can now examine the fixed point landscape for different values of R_J. This is shown in figure 3.7 b through f. Here, J_{E+} is varied over a wide range. It can be seen in the sequence of plots that increasing R_J has three effects. Firstly, it moves the appearance of *up*-states to higher values of J_{E+}. Secondly, it causes the *up*-states to become unstable when both R_J and J_{E+} are high. Finally, an increase in R_J leads to a gradual decrease in the maximum rates reached by the active clusters. For R_J close to one, the regime where stable *up*-states exists becomes increasingly narrow and when $J_{E+} = J_{I+}$ the active cluster states vanish as before (figure 3.7 f).

We have hence shown that increasing excitatory and inhibitory cluster strength proportionally can yield the desired effect of preventing the active cluster rates from saturating and consequently reducing the gap between *up* and *down* states which should in turn facilitate spontaneous switching between active clusters.

Figure 3.8 illustrates that this is indeed the case. Panel a shows sampled fixed points of a network with 20 clusters and $R_J = 3/4$ as before for the case without inhibitory clustering in figure 3.5 c. This time we have increased J_{E+} all the way to Q, at which point the inter-cluster connections of the excitatory populations vanish. It can be seen that over the whole range of cluster strengths, the maximum rates do not exceed 0.7. Also, even when the excitatory clusters are fully decoupled, there are still multiple different configurations of active clusters. The dashed lines below the *up*-states represent the inhibitory counterparts to the active clusters. This selective increase in inhibition is what prevents the active states from saturating.

Panel b shows the corresponding rate variance plot with sample raster plots. Compared to the equivalent plot for excitatory clustering only in figure 3.4,

the peak of σ_m^2 has shifted to higher clustering strengths while its amplitude and width have at the same time increased significantly. The useful range for computation is likely to lie in the rising branch of the σ_m^2 curve but for illustration purposes, the whole range is shown. At $J_{E+} = 4$, the point where the homogeneous state becomes unstable, the activity cycles through the clusters with moderate firing rates as desired. For higher cluster strengths, the activities of the *down*-state clusters become increasingly suppressed and the active clusters remain in the *up*-states for increasingly long times. The sample raster plots show however, that switching between states still occurs even when the excitatory clusters are fully decoupled at $J_{E+} = 20$.

In figure 3.9 two cases with and without inhibitory clustering are compared in more detail. Since inhibitory clustering shifts the onset of cluster dynamics to higher values of J_{E+}, we compare the dynamics at the point where the homogeneous state has just become unstable. The left panels of the figure show the case where $J_{E+} = 2.9$ and $R_J = 0$. On the $r.h.s.$ R_J was $3/4$ and $J_{E+} = 4$.

Since the homogeneous state is unstable for the parameters chosen the activity quickly moves from the random initial state to an active cluster in both cases. In the raster plot in panel a it is evident that the network remains in that state for the remainder of the simulation period for EE-only clustering. The firing rate plot in panel c confirms that the active cluster immediately goes into rate saturation. Since this increases the rate of the inhibitory population, all other E-populations experience a reduction in rate, which further widens the gap between the high and low cluster states. In the EI-clustered network on the other hand, the activity cycles happily between different attractors. It can be seen that the inhibitory clusters closely follow the rate excursions of their excitatory counter parts. Although all cluster have equal sizes and the weight of the same type (i.e. within or across populations) are all identical the switching between active clusters seems to occur at random.

The firing rates for the EI-clustered network also seem to follow the predictions obtain from the mean field model (panel d). To illustrate this we have plotted the distributions of instantaneous maximum cluster activity rates for 100 repeated network simulations in the bottom of figure 3.9. The dashed lines and dots represent the stable *up*-states predicted by the mean-field model. For the E-only cluster model the theory predicts three different stable configurations (one, two and three active clusters respectively). However, in 100 separate simulation runs almost exclusively the state with a single saturated active cluster was reached (panel e). For the EI-case the theory yielded two stable configurations. The maximum rates obtained from network simulations had a wider distribution

but the shape coincides with the theoretical fixed points. The configuration with two active clusters at lower rate seems to occur more frequently than the higher rates of individual clusters. Maximum rates were higher than the stable points predicted by the model. Note however that the probability of finding activity rates higher than ~ 0.7 was zero, i.e. rate saturation occurred in none of our simulations.

As mentioned above, the update mechanism used for the binary neuron simulations does not permit us to analyse the variability dynamics of these networks. We will however show in the following section that the rate dynamics resulting from the introduction of EI-clustered connectivity transfer well to networks of more realistic LIF units where variability analyses become feasible.

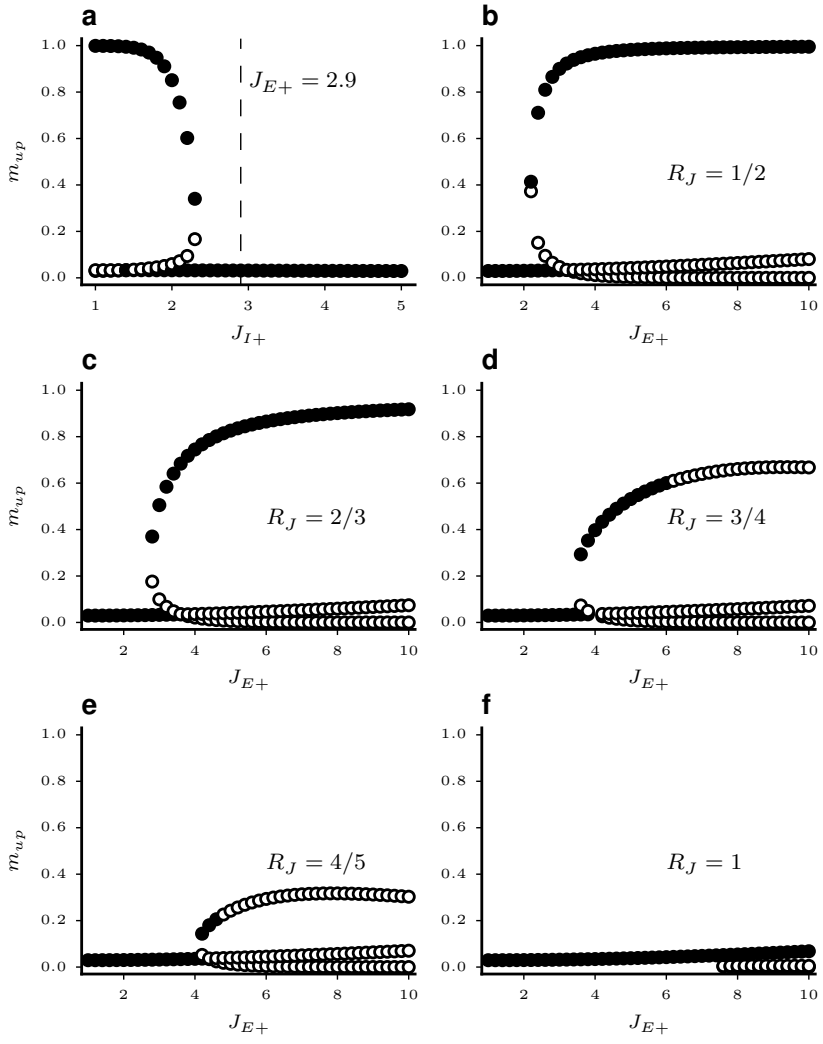

Figure 3.7: EFR fixed points for a single active population. Filled/empty circles represent stable/unstable fixed points. **a**) m_{up} vs J_{I+} with J_{E+} held constant at the value where the homogeneous fixed point becomes unstable if $J_{I+} = 1$. **b**-**f**) Proportional increase of J_{E+} and J_{I+} for different strength ratios R_J.

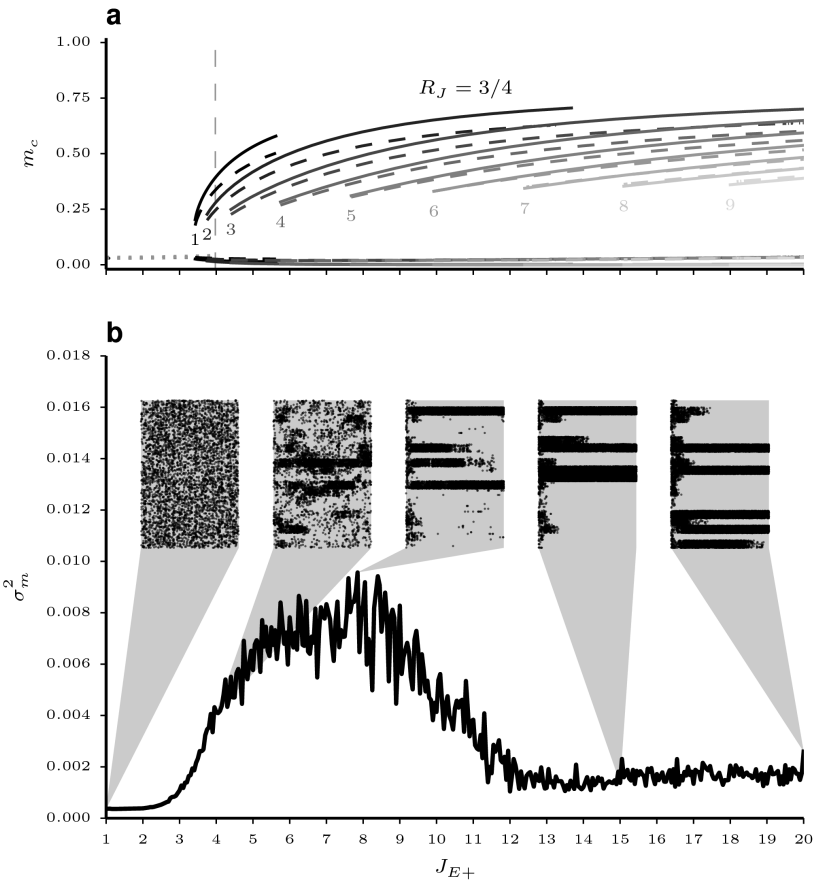

Figure 3.8: Network dynamics with $Q = 20$ and $R_J = 3/4$ with excitatory cluster strength varied from $J_{E+} = 1$ to full excitatory decoupling at $J_{E+} = Q$. **a**) Stable rate fixed points of the unconstrained mean field equations of the system. **b**) Instantaneous variance in mean cluster activity rates averaged over 20 network realisations.

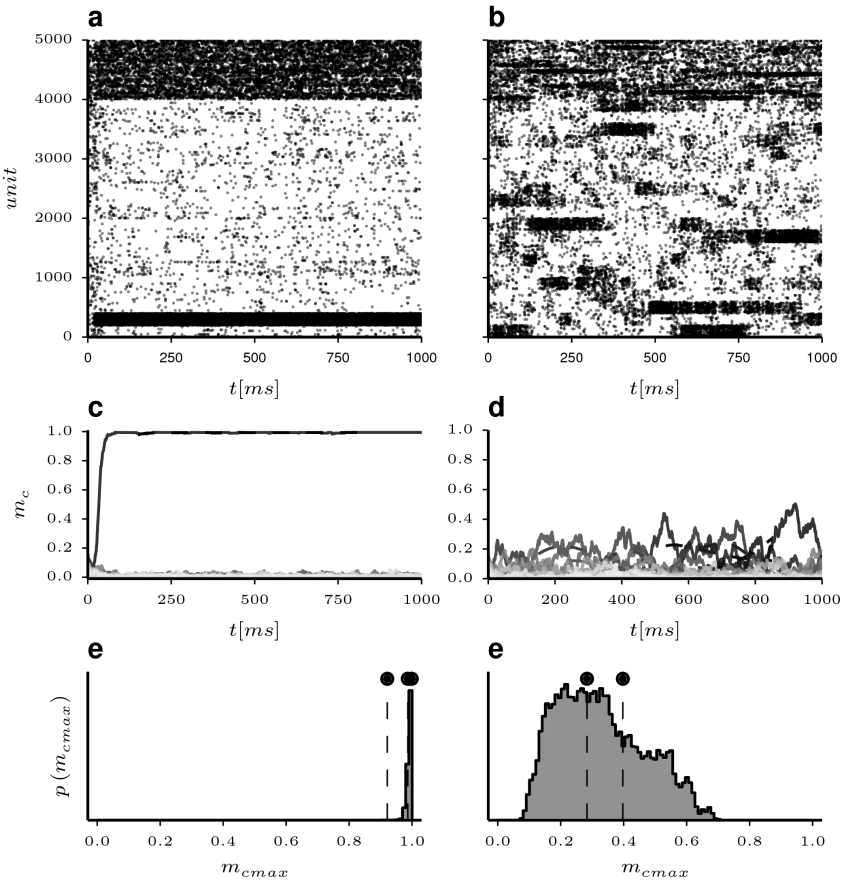

Figure 3.9: Comparison of cluster dynamics with $J_{E+} = 2.9, R_J = 0$ (left) and $J_{E+} = 4, R_J = 3/4$, (right). **a**) and **b**) Sample raster plots of E and I spiking activity. **c**) and **d**) Mean cluster activity rates m_c for the trial shown above. Dashed lines show cluster rates smoothed with a Gaussian kernel with $\sigma = 75ms$. **e**) and **f**) Distributions of instantaneous maximum cluster activity rates m_{cmax} for 100 random network realisations. Fixed points from mean field theory are indicated as dashed lines.

Chapter 4

Cluster Dynamics in Spiking Network Models

In the previous chapter we have seen how inhibitory as well as excitatory clustering reduces the distance between the *up* and *down* states for binary networks. This had the effect of facilitating cluster cycling. We now investigate whether this can be transfered to spiking networks. Although it is possible to treat attractor dynamics in networks of leaky integrate-and-fire neurons (LIFs) using a mean-field description [Amit and Brunel, 1997; Mazzucato et al., 2015] we limit ourselves here to reproducing the findings from chapter 3 in simulations. All spiking neuron simulations in this work were performed using the `NEST`-simulator [Gewaltig and Diesmann, 2007].

We model linear units with exponential synaptic currents where the sub-threshold evolution of the membrane potential V is described by the differential equation

$$\frac{dV}{dt} = \frac{-(V - E_L)}{\tau_m} + \frac{I_{syn} + I_x}{C_m} \tag{4.1}$$

In the absence of input, the membrane potential decays exponentially to the resting potential E_L with time constant τ_m. The synaptic input is represented by the current I_{syn} and I_x is an externally injected current. C_m is the membrane capacitance. If the potential reaches the threshold V_{th} a spike is emitted and V is clamped to a reset voltage V_r for an absolute refractory period τ_r. The synaptic current to a neuron i evolves according to the equation

$$\tau_{syn} \frac{dI^i_{syn}}{dt} = -I^i_{syn} + \sum_j J_{ij} \sum_k \delta\Big(t - t^j_k\Big) \tag{4.2}$$

where t_k^j is the time of the arrival of the k^{th} spike from presynaptic neuron j and δ is the Dirac delta function.

To allow comparison to other clustering results, we use parameters similar to those given in [Litwin-Kumar and Doiron, 2012; Mazzucato et al., 2015] (see table 4.1). The synaptic strengths are however, as is commonly the case, just given without explanation. For clarity, we will briefly explain how the parameters used here were derived in the following.

4.1 Calibration of the Balanced State

We follow the same approach as for the binary networks by requiring that $\sqrt{K}$ excitatory action potentials arriving within a short time suffice to drive the membrane potential form E_L to V_{th} and hence elicit a spike. For that purpose we need to compute the deflection in the membrane potential caused by a presynaptic spike.

According to equation 4.2, a spike arriving at $t = 0$ leads to a postsynaptic current of the form

$$I_{psc}(t) = Je^{-t/\tau_{syn}}\,\Theta(t) \tag{4.3}$$

where J and Θ are again the synaptic efficacy and step function respectively. Inserting this into equation 4.1 and integrating with $V = 0$ at $t = 0$ the postsynaptic potential is obtained:

$$PSP(t) = J\frac{\tau_m\tau_{syn}}{\tau_m - \tau_{syn}}\left(e^{-t/\tau_m} - e^{-t/\tau_{syn}}\right)\Theta(t) \tag{4.4}$$

The maximal deflection of the PSP, PSP_{max} occurs at $t = \frac{log\frac{\tau_{syn}}{\tau_m}}{\left(1/\tau_m - 1/\tau_{syn}\right)}$.

Note that the PSP amplitude depends on the the synaptic as well as the membrane time constants and is therefore different for each synapse type (PSP^{EE}_{max}, PSP^{EI}_{max},...). The scale-free weights are then constructed in the same way as for the binary networks (equations 3.3 - 3.8) but weighted by the

respective PSP amplitudes:

$$j_{EE} = \frac{V_{th} - E_L}{\sqrt{p_{EE} n_E}} \frac{1}{PSP^{EE}_{max}} \tag{4.5}$$

$$j_{EI} = -g j_{EE} \frac{p_{EE} n_E}{p_{EI} n_I} \frac{PSP^{EE}_{max}}{PSP^{EI}_{max}} \tag{4.6}$$

$$j_{IE} = \frac{V_{th} - E_L}{\sqrt{p_{IE} n_E}} \frac{1}{PSP^{IE}_{max}} \tag{4.7}$$

$$j_{II} = -j_{IE} \frac{p_{IE} n_E}{p_{II} n_I} \frac{PSP^{IE}_{max}}{PSP^{II}_{max}} \tag{4.8}$$

where g is again the relative strength of inhibition. The final weights $J_{\alpha\beta}$ are again obtained by dividing by $\sqrt{N}$.

Since we consider variability dynamics the external inputs are modelled as constant currents to ensure all variability arises deterministically inside the network rather than stemming from externally generated Poisson input. In analogy to the *threshold rate* of [Brunel, 2000] the external current I_x is expressed in terms of the current required to reach the threshold in the absence of synaptic input:

$$I_{th} = \frac{V_{th} - E_L}{\tau_m} C_m \tag{4.9}$$

A complex interplay exists between the E and I firing rates and the magnitude of the external currents to the populations. The tuning of the injected currents required to obtain the desired firing rates for 3 and 5 spikes per second for the E and I populations respectively was therefore achieved by modelling single units with Poissonian inputs mimicking the network input at the target firing rates. The external inputs could then be increased until the modelled units fired on average at the required rates.

Before introducing structured connectivity we now ensured that the network configuration was operating in the asynchronous-irregular (AI) regime. Irregularity was measured using the CV^2 as described in chapter 2. Synchrony of measures such as the instantaneous firing rate or the membrane potential in neural networks can be quantified according to [Golomb and Hansel, 2000] as:

$$\chi = \sqrt{\frac{\sigma^2_{pop}}{\left\langle \sigma^2_i \right\rangle}} \tag{4.10}$$

here σ^2_{pop} is the variance of the the population average and $\left\langle \sigma^2_i \right\rangle$ is the average over the individual units' variances. The measure gives unity for totally synchronised activity and for asynchronous activity in networks of size N, one

Parameter	Unit	Value
N	-	$4000(E), 1000(I)$
E_L	mV	0
V_{th}	mV	20
V_R	mV	0
C_m	pF	1
τ_m	ms	$20(E), 10(I)$
τ_{syn}	ms	$3(E), 2(I)$
τ_r	ms	5
p_{EE}	-	0.2
p_{EI}, p_{IE}, p_{II}	-	0.5
g	-	1.2
J_{EE}	pA	0.33
J_{EI}	pA	-0.89
J_{IE}	pA	0.25
J_{II}	pA	-1.34
I_x	pA	$2.13I_{th}(E), 1.24I_{th}(I)$

Table 4.1: Summary of parameters used in the spiking network simulations

expects $\chi \sim \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$. Since recording all membrane potentials in simulations is computationally expensive, we computed χ on spike counts measured in bins of $20ms$.

It can be seen in figure 4.1 that the networks show the usual characteristics of the balanced state. When excitation dominates, synchronous-regular firing near the saturation rate $1/\tau_r$ is observed. The AI state occurs when g is sufficiently large for inhibition to dominate (figure 4.1a). As in section 3, we chose $g = 1.2$, where $\chi = 0.02 \sim 1/\sqrt{N}$ and $CV^2 = 0.73$ (panel b). The raster plot shows no discernible structure (panel c) and the average firing rate is low and constant over time (panel d). The synaptic currents from excitatory and inhibitory inputs and the external current Ix cancel so that the net input fluctuates about zero (panel e). Hence the membrane potentials fluctuate at low values and only occasionally reach the threshold to produce a spike (panel f). The parameters used for all simulations in this chapter are summarised in table 4.1.

It is worth noting that the synaptic delay had to be set equal to the simulation time step ($0.1ms$) in order to avoid synchronous oscillations. This is not biologically realistic and usually synaptic delays between 0.5 and $2ms$ are used in network simulations (e.g. [Kriener et al., 2014; Ostojic, 2014; Pernice et al., 2011; Morrison et al., 2007]). Synaptic delays are known to play a role in synchronous oscillations [Brunel, 2000; Bose and Kunec, 2001]. We did not conclusively investigate the origin of those oscillations. It is however likely that it is due to the dense connectivity employed in our model which we chose to

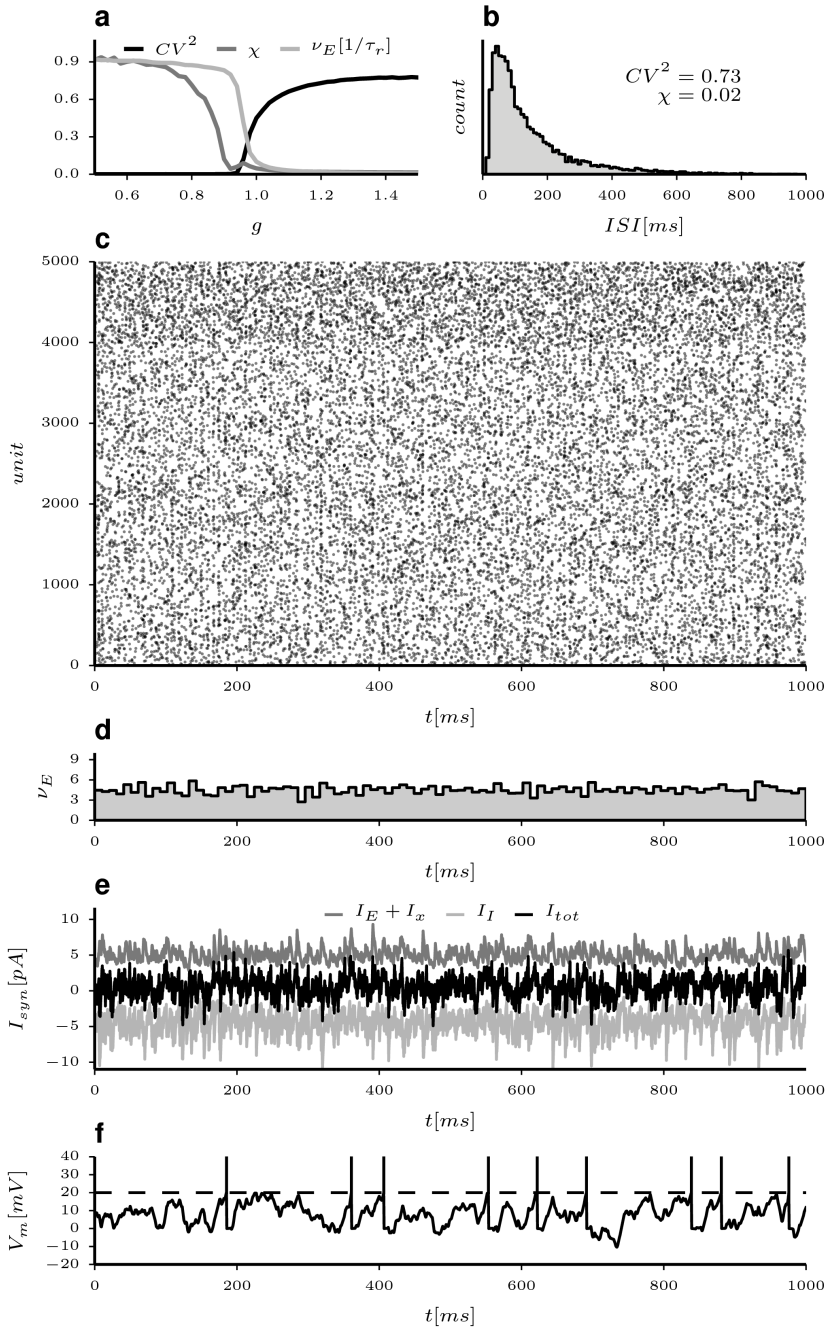

Figure 4.1: Spiking network in the balanced state. Parameters as in table 4.1. **a**) CV^2, χ and normalised excitatory firing rate versus relative inhibitory strength g. **b**) Pooled ISI distribution for the E population. **c**) Raster plot of one second of spiking activity. **d**) E population rate histogram computed in $10ms$ bins. **e**) Synaptic currents of a randomly selected E unit. **f**) Membrane potential for same unit as in **e**. Vertical bars above the threshold (dashed line) represent action potentials

enable comparisons to [Litwin-Kumar and Doiron, 2012] and [Mazzucato et al., 2015] who presumably did the same as no mention is made of a synaptic delay in either reference.

Hence our unstructured networks operate in the AI regime with the parameters given. The CV^2s are slightly lower than those usually reported for the balanced state. This might be connected to the dense connectivity issue mentioned above. The irregularity can be increased by increasing the overall scale of the synaptic weights [Lerchner et al., 2006]. We have confirmed this in our model (results not shown) but decided to stick with the weights as given by equations 4.5 through 4.8 as the resulting CV^2s are consistent with those found in the data shown in chapter 2.

We will now turn to investigating the effect of clustered connectivity on these spiking networks.

4.2 Variability Statistics during Winnerless Competition

Clustered connectivity matrices were constructed in the same way as before. As for the binary units, we compare networks with EE-only clustering to those where the inhibitory population is also clustered with a relative strength R_J as defined by equation 3.36. Since the LIF units do not involve random updates, we can now compute the FF as for the physiological data before.

Figure 4.2 shows the FF in relation to the excitatory clustering strength J_{E+} for $R_J = 0$ (panel a) and $R_J = 3/4$ (panel b) for $Q = 20$ clusters. The statistics were computed over 20 trials of $400ms$ duration and averaged over 50 random network realisations. Trials were cut from continuous simulations. In the case without inhibitory clusters, a similar effect can be observed as for the binary networks (compare figure 3.4). For low values of J_{E+}, the FF remains unchanged on a level similar to that of the unstructured CV^2. At $J_{E+} \sim 2$ an increase in FF occurs. As for the binary case, the peak is quite narrow. It can be seen in the corresponding inset in figure 4.2 a that the activity for this setting looks more structured to begin with but can suddenly switch to a state with a single active cluster. This is characteristic for the two different types of activity occurring in different realisations at this point. I.e. winnerless competition does occur but it is rather fragile. As J_{E+} is increased further states where one cluster *wins* become dominant and the FF decreases below that of the unstructured networks. The reason is that the highly active clusters fire near the saturation rate where firing is very regular. This also regularises the input to the remaining

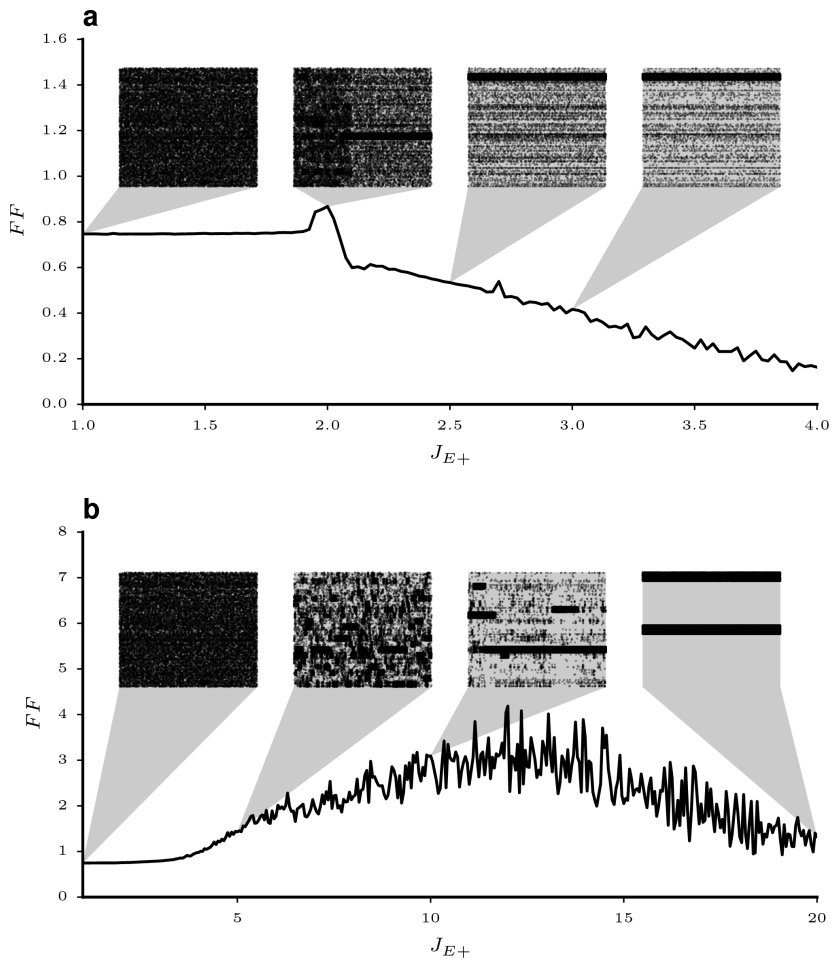

Figure 4.2: FF versus J_{E+} for networks with $Q = 20$. Other parameters as in table 4.1. FFs were computed on 20 trials of 400 ms and averaged over 50 network realisations. Insets show one second of sample spiking activity of the excitatory populations. **a**) $R_J = 0$ **b**) $R_J = 3/4$

clusters which become increasingly quiet as cluster strength increases.

In the case with $R_J = 3/4$ the results of the binary networks are confirmed and the cluster switching dynamics become more moderate and are spread over a much wider range of J_{E+}. With increasing cluster strength, the durations of individual *up*-states become longer and their amplitudes increase. As long as the *up*-state durations are not much much longer than the estimation window, this increases the observed rate variance and thereby the FF. The behaviour is again shown for the whole range of cluster strength until the excitatory populations become completely decoupled at $J_{E+} = Q$. For $J_{E+} > 5$ or

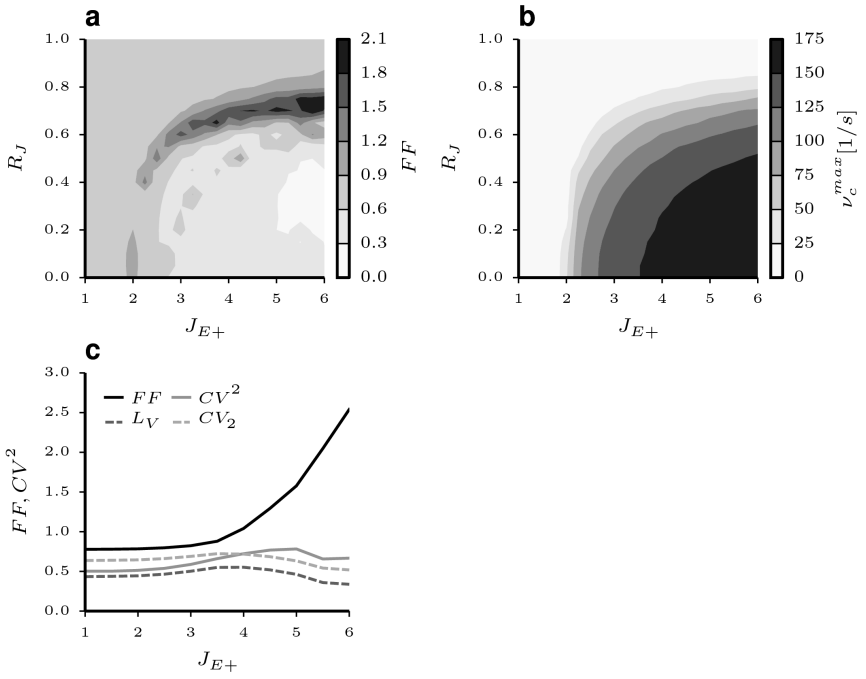

Figure 4.3: Effect of cluster parameters on variability statistics for networks with $Q = 20$. Statistics are computed from 20 trials of $400ms$ width and averaged over 20 network realisations. **a**) FF dependence on J_{E+} and R_J. **b**) Maximum average cluster firing rate vresus J_{E+} and R_J. **c**) $R_J = 3/4$. FF, CV^2 and local measures of interval variability for increasing cluster strength.

so, the average FF becomes quite noisy although it has been computed over 50 random realisations with identical parameters. For those large values the variance over repeated simulations is high as in some cases cluster cycling is quite violent while in others one or more cluster *win*, suppressing the winnerless competition dynamics. At $Q = 20$ coupling between the populations exists only through the inhibitory connections. This coupling is however still relatively strong so that most populations are quieted by a few *winners*. The whole range is again mainly shown for completeness, as the *interesting* dynamics take place at relatively low values of J_{E+}.

In chapter 2 we have used the concept of a doubly stochastic renewal process and found that it can explain the dynamics of the variability components in our data set. In this framework of interpretation the FF is expected to increase with rate variance while interval statistics should remain unaffected. An increase in cluster strength translates to more rate variance. Figure 4.3c shows the FF as well as the interval statistics CV^2, L_V and CV_2 for moderate values of J_{E+}, again for networks with $R_J = 3/4$. Before J_{E+} is sufficiently large to increase

the FF and therefore the rate variance, the relation between the measures of interval variability match those found in our data and reported in [Ponce-Alvarez et al., 2010] ($L_V < CV^2 < CV_2$, see section 2.3). This relationship does not hold for stronger clustering although the FF does seem to increase much more than the interval statistics. Note that the interval measures have been calculated in windows of $400ms$, using the bias correction for the CV^2 introduced in chapter 2, which assumes that spiking is well described by a Gamma process. We have therefore again included the local measures L_V and CV_2 which are model-free estimators. According to those measures, interval variability seems to decrease with rate variance.

So far we have only considered EI-clustered networks where the cluster strength ratio R_J was set to $3/4$. As we have seen from the mean field considerations for binary networks, increasing R_J seems to have the effect of shortening the distance between *up* and *down* states in cluster switching. In figure 4.3 a we show the FF for different combinations of R_J and J_{E+}. As expected, for small R_J there is only a small *bump* at $J_{E+} = 2$. Inhibitory clustering seems to stabilize the winnerless competition roughly for $0.4 < R_J < 0.8$. Figure 4.3 b shows the maximum cluster averaged firing rates for the same parameter space. It can be seen that the region that produces realistic values of FF is at the transition from lower to higher cluster rates. The EI-cluster model can produce a wider range of Fano factors than the model with excitatory clusters only $R_J = 0$.

Figure 4.4 illustrates the difference between pure excitatory and EI-clustering on the level of individual units. The raster plot in panel a shows the activity of five sample clusters in a network with $J_{E+} = 2$ and no inhibitory clusters. All clusters fire at low rates until at about $t = 500ms$ one cluster switches to the *up* state. The cluster averaged instantaneous firing rates, estimated with a triangular kernel with $\sigma = 50ms$, of the same run are shown in panel c. The firing rate of the active cluster increases dramatically, while the rates of the other clusters are suppressed. This is the expected behaviour as described by Litwin-Kumar and Doiron [2012]. Panel g shows the synaptic currents for a single neuron from the active population. It can be seen that while the excitatory current increases due to self excitation within the cluster, the net inhibitory input remains constant. Hence the mean current increases and spiking becomes *mean*-driven rather than *fluctuation* driven. This means that a large portion of the units in the active cluster synchronise and fire like clockworks at the saturation rate dictated by the refractory period as can be seen in panel e which depicts the membrane potential of the same unit.

The $r.h.s.$ of the figure shows the equivalent scenario for a network with

$R_J = 3/4$ and $J_{E+} = 4$, a setting where FFs around 1.2 are common. The raster plot shows that at this mild stage of winnerless competition, the clusters cycle between short burst of higher activity with occasional longer excursions to the *up* state of one or more clusters. Panel d shows that this cycling takes place at much lower firing rates. Panel h again shows the synaptic currents of a sample neuron at the transition from low to higher activity of its surrounding cluster. The effect of inhibitory clustering is that the inhibitory currents also seem to increase their magnitude during *up* states, thereby increasing the variance, but not the mean of the synaptic currents, keeping inputs balanced throughout. Hence the units remain in the fluctuation driven balanced state with irregular spike trains as is confirmed in panel f. Although synchrony arises also in this case as units in the same cluster tend to have correlated rates, the tight locking seen in panel a is not observed in the fluctuation driven case.

Having shown that the effect of inhibitory clustering on the spontaneous dynamics in spiking networks is qualitatively equivalent to our observations in binary networks we now turn to the *quenching* of variability during stimulation. In chapter 2 we have seen that the reduction in variability at stimulus onset reported in [Churchland et al., 2010] can also be found in our dataset. We now inject additional external current into units belonging to some of the clusters to mimic stimulation. It is known that this can reduce the trial-to-trial variability as measured by the FF [Litwin-Kumar and Doiron, 2012; Deco and Hugues, 2012; Mazzucato et al., 2015]. The effect of stimulation of clustered networks on the interval statistics has however to our knowledge not previously been examined.

We again use a network with $R_J = 3/4$ and $J_{E+} = 4$. Stimulation is simulated by applying a step increase in external current to two of the twenty clusters. The amplitude of this current was $0.1I_{th}$, giving an increase of less than 5% with respect to the external current during spontaneous firing. To enable a comparison to the results from the data presented in figures 2.6 and 2.7, where we required units to have mean counts of at least 15 spikes in the observation interval measuring two seconds, we analysed only units exhibiting at least 7.5 spikes per second. Stimuli were applied for one second at a time and an inter stimulus interval randomly varied between 2000 and $2500ms$. This was inserted to avoid periodic effects. A total of 200 trials were cut in windows ranging from -1000 to $2000ms$ around stimulus onsets.

A sample trial is shown in figure 4.5 a. For better visibility only 10 of the 20 clusters are shown. The stimulus was applied only to the top most couple of populations. Note that the applied stimulus is relatively weak. It increases the probability of the stimulated clusters switching the the *up*-state but does not forcefully drive them. In the trial shown, the top most cluster only hardly

becomes active. Also, the random cycling between the other clusters observed during spontaneous activity is reduced but not completely suppressed by the stimulus.

The lower panels of figure 4.5 show statistics calculated for the stimulated populations ($l.h.s.$) and the remaining clusters ($r.h.s.$) separately. The basic parameters are still those given in table 4.1 with external inputs tuned to give an average rate of 3 spikes per second. The higher average firing rates in the spontaneous state are due to the selection criterium mentioned above. The non-stimulated clusters only show a small decrease in their average firing rate during stimulus applications, while the rate of the stimulated clusters nearly doubles. Besides increasing the inhibitory effect on active clusters, inhibitory clustering also *decreases* the suppression of the activity in the remaining clusters. The suppression of trial-to-trial variability is however conserved. Both stimulated and non-stimulated populations show similar reductions in FF during the stimulus interval with a slightly stronger effect for the stimulated populations (panel e).

Interval statistics on the other hand differ between the stimulated and unstimulated populations. In the non-stimulated populations the CV^2 as well as the local measures remain fairly constant over time. The CV^2 is higher than both L_V and CV_2 which is consistent with the findings shown in figure 4.3 a. In the stimulated population the interval variability measures are slightly reduced when the stimulus is applied. Most of the interval variability was however conserved in contrast to networks with excitatory clustering only [Deco and Hugues, 2012; Litwin-Kumar and Doiron, 2012].

Because of this reduction in interval variability the rate variances shown in the bottom panels of the figure were computed using the full time resolved estimate of the CV^2 rather than the single value calculated on the whole interval as before when it was constant. The magnitude of σ_ν^2 is lower than the average observed in the data but the model captures the stimulus induced reduction.

We have shown that the introduction of inhibitory as well as excitatory clusters in balanced networks gives more control over the firing rates during winnerless competition dynamics while preserving the characteristic properties like increased rate variance and its reduction during stimulation. In the following chapter we will see that the model of winnerless competition can explain further observation on our dataset.

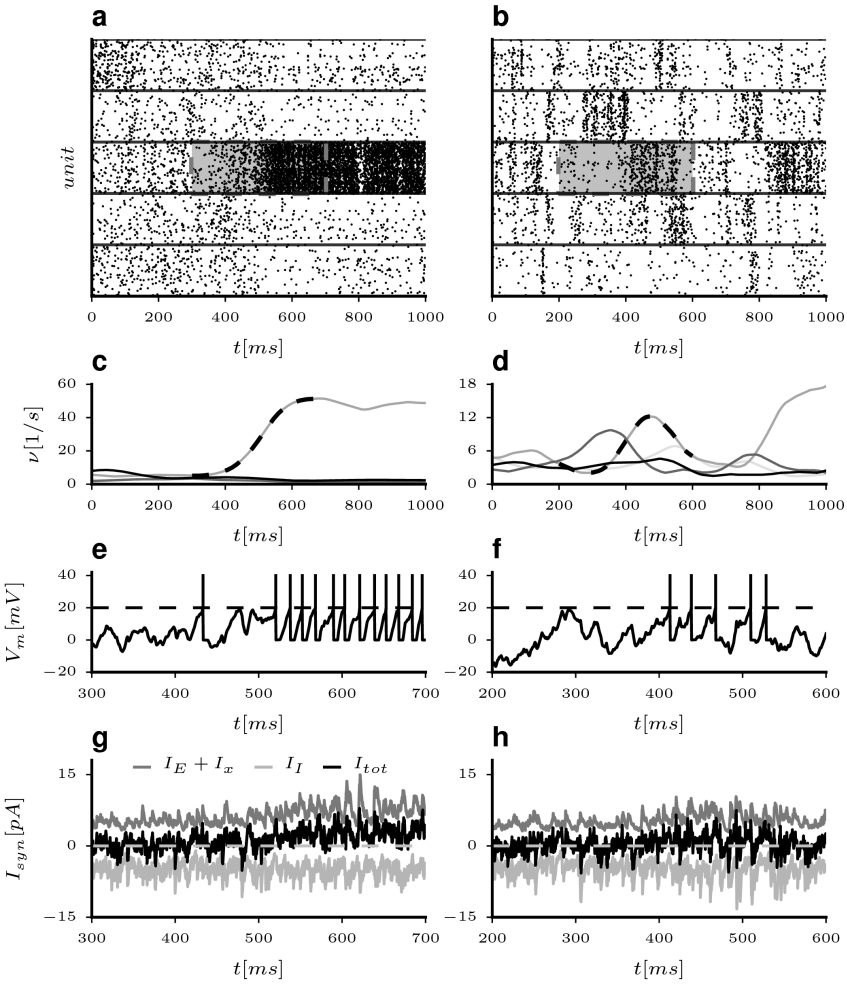

Figure 4.4: Comparison of attractor dynamics with ($r.h.s.$: $J_{E+} = 4, R_J = 3/4$) and without inhibitory clustering ($l.h.s.$: $J_{E+} = 2, R_J = 0$). **a**) and **b**) Raster plots of five excitatory clusters for $1000ms$ of spontaneous activity. Horizontal lines indicate cluster membership. **c**) and textbfd) Average firing rate of the above clusters estimated with a triangular kernel with $\sigma = 50ms$. Dashed lines correspond to the shaded region of interest in the upper panels. **e**) and textbff) Membrane potentials with superimposed spikes for randomly chosen units during the change from *low* to *up*-states for the region of interest indicated above. **g**) and textbfh) Synaptic currents for the units shown above in the same time interval.

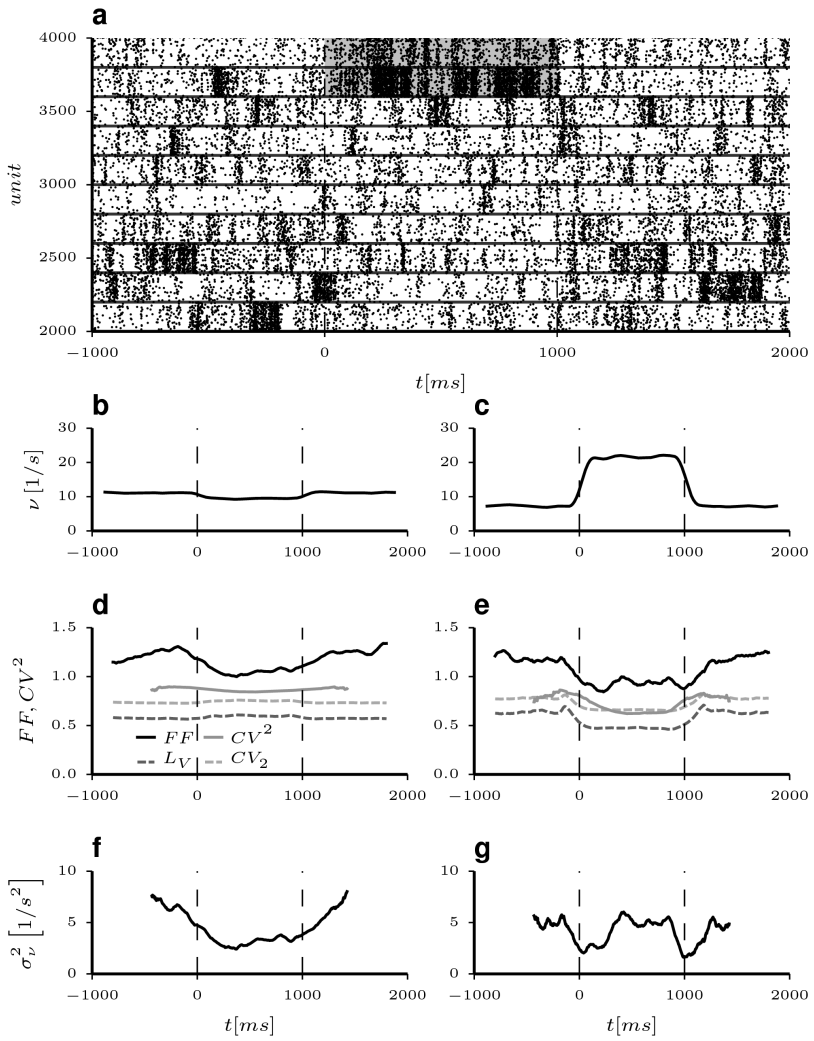

Figure 4.5: Effect of stimulation on attractor dynamics. Two of the 20 clusters have been *stimulated* with an additional constant current of $0.1I_{th}$ applied for $0 < t < 1000ms$. **a** Raster plot of a sample trial for 10 clusters. Top most populations were stimulated. **b-g**) Statistics from 200 simulation trials averaged over units in the stimulated ($r.h.s.$) and non-stimulates clusters ($l.h.s.$). Vertical dashed lines indicate the stimulation interval. **b-c**) Firing rates of the excitatory units. **d-e**) Variability statistics computed in $400ms$ windows. **f-g**) Rate variance computed using the time resolved CV^2. To allow comparison to the physiological data, only units with an average of at least 30 spikes in the analysed period were analysed.

Chapter 5

Context Dependent Variability Modulation in Monkey Motor Cortex

In chapter 2 we have seen how variability statistics can be measured and how count variance modulates during the course of experimental trials while interval statistics remain largely constant. In chapter 4 we showed that the modulation of count variance can be reproduced in balanced networks with clustered excitatory connectivity and that the preservation of interval variability requires that the excitatory units are also clustered. In the current chapter we will see that the balanced network with EI-clustering can reproduce some other effects found in our data set which have so far not been discussed.

In chapter 2 we showed the analyses for one of the three experimental conditions present in the data set. The experiment involved three conditions in which varying amounts of target information were given at the beginning of the trials. The original publication of the data set [Bastian et al., 2003] dealt mainly with the evolution of directional tuning during the preparatory phase. In [Rickert et al., 2009], a supplementary figure hinted at an aspect that will form the centre of attention of the present chapter: The magnitude of modulations in the average Fano factors depends on the amount of target information available to the monkey. Although variability has been shown to be different for different reaction times in a similar experiment [Churchland et al., 2006] and has also been found to depend on the performance history during experimental sessions [Marcos et al., 2013], such and explicit dependence on experimental conditions has to our knowledge not been discussed elsewhere.

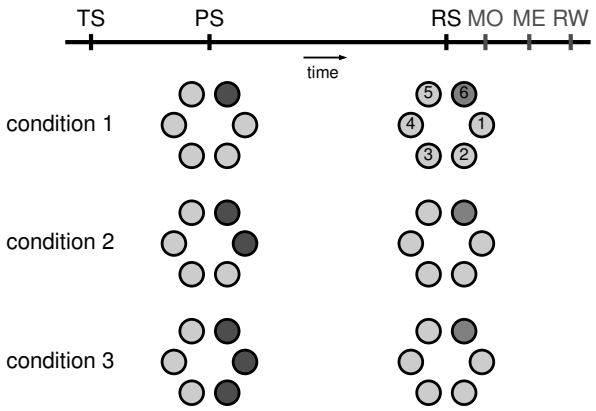

Figure 5.1: Experimental protocol of the delayed centre out reach task. $500ms$ after trial start (TS) a preparatory stimulus (PS) is presented, indicating either 1,2 or three adjacent targets, depending on the condition. The response signal (RS) follows a fixed delay of $1s$ after which the monkey is required to execute a movement to the indicated target. Movement onset (MO), movement end (ME) and the time of reward (RW) depend on behaviour and are not fixed in time.

In section 5.2 we will analyse the variability and information content of the motor cortical spike trains during the delay period in our data set for all three conditions and find that the FF decreases when more information is present. In section 5.3 we will then show how a simple attractor model with winnerless competition dynamics can reproduce those observations.

5.1 Experimental Paradigm

The experiments were conducted at Alexa Riehle's lab at Institut de Neurosciences de la Timone CNRS & Universite Aix-Marseille, France and the results were first published in [Bastian et al., 2003]. Two Macaque monkeys were trained to perform a delayed centre-out reach task with varying information about the correct movement at trial start. The subjects were seated in front of a panel featuring a hexagonal array of touch sensitive LEDs. The monkeys initiated the trials by placing their hand on a seventh LED at the centre. $500ms$ after trial start (TS), a preparatory signal (PS) was given. The PS consisted of either one, two or three of the LEDs lighting up in red depending on the condition,as illustrated in figure 5.1. After a a delay of $1000ms$ a single LED lit

up in green representing the response signal (RS) which informed the subjects where to move their hand. The times of MO and movement end (ME) were recorded and if the monkey touched the correct LED, the trial was registered as successful and a drop of juice was given as a reward. Only successful trials were available for the present study.

The conditions of one, two or three possible targets presented at PS were executed in blocks. In each block, 150 trials with randomised target directions were carried out so that each of the directions appeared on average 25 times per condition. Note that in order to obtain the same number of possible trial types in all conditions, not all possible combinations of directions for the preparatory stimulus were used in conditions 2 and 3. Since six combinations are possible for condition one, only the pairs 1-2, 3-4 and 5-6 were used in condition 2 and for condition 3, only two cases occurred (6-1-2, 3-4-5).

Extracellular recordings were taken with electrode arrays in pre-motor areas (monkey 1: M1 close to the central sulcus, monkey 2: mainly PMd). Online spike sorting resulted in up to seven simultaneously recorded single unit spike trains. On each recording day, all three conditions were measured so that the responses of individual neurons can be compared across conditions. Some of the available units were excluded from the analyses because of suspected artefacts. The interested reader may view the details of this procedure in appendix A.

Since we intend to compare variability statistics across conditions, additional precautions were taken. As in chapter 2 Fano factors are computed for each unit and direction separately and we require that units have at least 10 spikes in the two second interval after trial start and that at least ten trials were recorded per direction. To enable the comparison across conditions, we only included units and directions were those criteria were met for al conditions.

5.2 Data Analysis

The analyses for all three conditions for monkey 1 are shown in figure 5.2. Panel a shows the firing rates, again estimated using a triangular kernel of width $\sigma = 50ms$. When the preparatory stimulus is presented, the firing rates increase sharply in all three conditions. However, the peak is more pronounced for condition 1, where the complete information is given at PS. The firing rates for all conditions remain elevated throughout the preparatory period. When the response signal is given, the rate drops for condition 1 but rises sharply for conditions 2 and 3 where the information required for movement initiation is only now available.

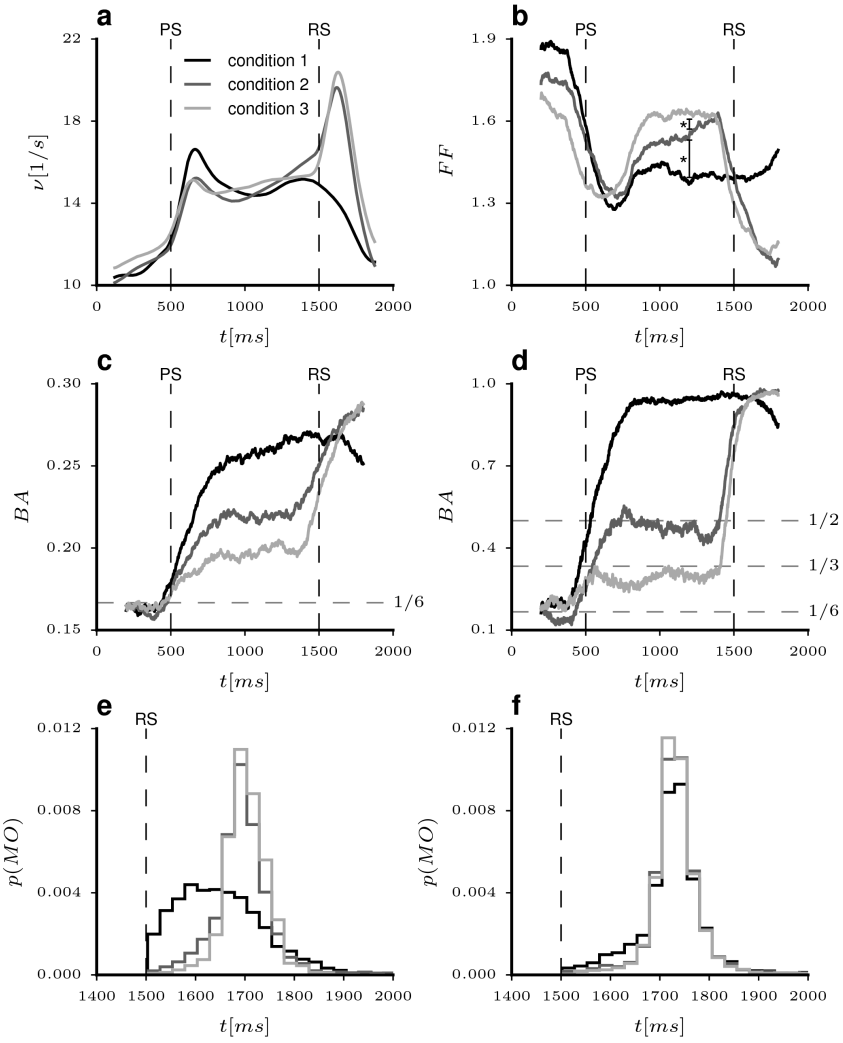

Figure 5.2: Analyses of the physiological data for all three experimental conditions. **a-e** are monkey 1, **f** is monkey 2. **a**) Firing rates estimated with a triangular kernel ($\sigma = 50ms$). **b** Fano factors computed in sliding windows of $400ms$. $*$ indicates significance from Wilcoxon signed rank test (conditions 1-2: $p = 0.007$, conditions 2-3: $p = 0.03$) **c**) Balanced accuracy of direction classification using single unit counts in windows of $400ms$. Lines show averages over 10 repetitions of 5-fold cross validation. **d**) Same as **c** but for a pseudo-population constructed from all available neurons. **e**) Distribution of reaction times (MO) for monkey 1. **f**) Distribution of reaction times for monkey 2.

Panel b shows the average FF for all units. For all conditions, the count variability is *quenched* at the presentation of the PS. The FF then recovers to different levels depending on the condition. When the complete information is given at PS, the FF remains at a low value while for condition 2 it is higher and still higher when three possible targets are presented. This difference is significant with respect to a Wilcoxon signed rank test throughout most of the preparatory period (conditions 1-2: $p = 0.007$, conditions 2-3: $p = 0.03$ at $1200ms$ after TS). After the response signal, the FF drops close to unity for conditions 2 and 3 but remains more or less constant for condition 1. This may in part be due to the fact that the movement onset times for condition 1 were more broadly distributed than for the more difficult conditions 2 and 3 where the monkey had to wait for the RS before he could form a decision. Hence for conditions 2 and 3 the trials are closer to being aligned to MO (see panel e).

To asses how much directional information is contained in the single trial counts of individual units we computed the classification performance of a logistic regression classifier using spike counts in windows of $400ms$ width. At each point in time, 10 repetitions of 5-fold shuffled cross validation were performed. We computed the balanced accuracy of the predictions (BA) to asses how well the classifiers performed. Balanced accuracy overcomes the effects of unequal class sizes in the estimation of classification performance and is simply the fraction of correctly predicted samples averaged over classes.

$$BA = \frac{1}{C} \sum_{c=1}^{C} \frac{N^c_{correct}}{N^c_{total}} \tag{5.1}$$

All classification was done using the `scikit-learn` package [Pedregosa et al., 2011].

Figure 5.2 c shows the balanced accuracy averaged over all units for the three conditions. As expected, before PS when no information is given, the classifiers perform at a chance level of $1/6$. During the preparatory period, the classification performance increases for all three conditions. Although the increase is only slight because classification is performed using a single spike count per trial, there is a clear difference between the conditions with the performance being higher the more information is presented. At RS, the three curves converge again at roughly the level of the full information condition. The fact that the accuracy increases further for conditions 2 and 3 while it starts to drop for condition 1 is again an indication that the neurons are involved in movement preparation rather than execution.

To show that the combined information of those single units is sufficient to reach good performance in the task, we have reproduced the approach of Rickert

et al. [2009] and constructed pseudo-populations of all available units. This was achieved by concatenating spike counts from randomly chosen trials for the same directions from all units and pretending that they have been simultaneously recorded. The performance was again calculated using 5-fold cross validation and averaged over 10 randomly assembled populations. The resulting balanced accuracies for the three conditions are shown in panel d. Using these pseudo populations, which is likely to be more similar to the information the monkeys have available than the single unit counts used before, the balanced accuracies approach those theoretically possible given the available information. I.e. in condition 1 where full information is available, the classification performance is close to unity. When two possible targets are given the accuracy is close to $1/2$ and in condition 3 it approaches $1/3$. At RS the performance approaches one for all conditions.

Note that panels a-d show only units recorded from monkey 1. The reason is that the two subjects followed different strategies during the experiments. While monkey 1 was keen to react as quickly as possible and often anticipated the RS in condition 1 (i.e. started moving too early,see truncated distribution in panel e) monkey 2 did not react faster in the full information condition and hence hat similar reaction times in all cases (panel f). Although the differences between conditions are also present for monkey 2, they are less pronounced.

It has been hypothesized that movements are prepared prior to their execution [Riehle, 2005]. This idea stems from the observation that movements are executed faster when information about the target of movement has been given prior to a delay period [Rosenbaum, 1980; Riehle and Requin, 1993]. A possible interpretation of the reaction time distributions is that monkey 2 simply did not start movement preparation prior to RS. This is a viable strategy because faster reaction times were not specifically rewarded. Since we concentrate on movement preparation in the following, we decided to perform the analyses in this chapter for monkey 1 only.

Movement preparation is often associated with bringing the firing rates in motor cortex to a certain *optimal subspace* [Churchland et al., 2006], or *initial condition* [Afshar et al., 2011] suitable for the movement to be executed. Alternatively, the *activation* associated with a movement is brought *close to threshold*. All these models assume that a certain combination of firing rates needs to be achieved before a movement can be executed and that this takes some time. Our findings that average firing rates are similar while movement direction can still be decoded to some degree from individual neurons and nearly optimally from populations during the preparatory period is consistent with this view. Achieving a suitable *position* in some high-dimensional firing rate

Parameter	Unit	Value
N	-	$1200(E), 300(I)$
E_L	mV	0
V_{th}	mV	20
V_R	mV	0
C_m	pF	1
τ_m	ms	$20(E), 10(I)$
τ_{syn}	ms	$3(E), 2(I)$
τ_r	ms	5
p_{EE}	-	0.2
p_{EI}, p_{IE}, p_{II}	-	0.5
g	-	1.2
J_{EE}	pA	0.60
J_{EI}	pA	-1.60
J_{IE}	pA	0.46
J_{II}	pA	-2.44
I_x	pA	$1.25I_{th}(E), 0.78I_{th}(I)$
I_{stim}	pA	0.15
Q	-	6
J_{E+}	-	3.3
R_J	-	$3/4$

Table 5.1: Parameters used in the spiking network model

space prior to movement execution could explain the reduction in FF during the preparatory period. Naturally, if the exact movement direction is known (condition 1), a more *optimal* initial condition for the movement can be prepared than when the target is ambiguous (conditions 2,3) and shorter reaction times can be achieved. An ambiguity in initial condition could also explain the higher rate variance in conditions 2 and 3.

In the following we propose a mechanistic model employing the circuitry described in chapter 3 that is based on winnerless competition and can explain the differences in count variance and information content of single units in the preparatory period as well as the difference in reaction times.

5.3 An Attractor Model of Movement Preparation

The aim of our model is to show how winnerless competition can lead to the condition dependent count variability and information content observed in our data set. We therefore simplify its construction as much as possible. The architecture is schematically shown in figure 5.3 a.

The core of the model consists of a balanced network of integrate and fire units with $Q = 6$ clusters - one for each direction. Each cluster consists of

200 excitatory units and is assigned a corresponding inhibitory cluster with 50 units. The clustering parameters J_{E+} and R_J were adjusted so that transitions between clusters occurred as we have shown in chapter 4 and the Fano factors in the spontaneous state resembled those measured in the physiological data (see figure 5.2 b). Since the network is smaller than those used in chapter 4, the strength of the external input currents had to be adjusted to obtain firing rates of approximately 3 and 5 spikes per second for the excitatory and inhibitory populations respectively. No special parameter tuning was otherwise performed. The parameters used are summarised in table 5.1.

The physiological neurons display various types of tuning curves which can also change over time [Bastian et al., 2003; Rickert et al., 2009]. We do not specifically model tuning or how it arises. We simply stimulate one cluster per direction by applying a constant current of $0.15pA$ to all its units. This means that each unit in our model is sharply selective to a single direction. The same stimulation protocol as described above for the actual experiment was applied to our model. For each condition 150 trials were executed. In each trial one of the six trial types was randomly chosen. The preparatory signal was applied to the model by direct stimulation of 1,2 or three clusters depending on the condition. Note that the amplitude of the stimulation current was identical for all conditions. After $1000ms$, i.e. at the end of the preparatory period, the PS currents were turned off except for the one corresponding to the correct target direction. The target current was applied for an additional $400ms$ to mimic the response signal in the experiment. Trials were cut from one long continuous simulation. To allow the network to relax to its spontaneous state again, inter-trial periods randomly varied between 1500 and $1700ms$ were inserted. The variance in this relaxation period was intended to avoid any effects of periodicity.

The decoder depicted in figure 5.3a is used to read out the movement direction during the $400ms$ interval where the RS is applied. Since the directional stimulation is applied to separate sub-populations of the network, *decoding* is not really necessary in our model. The decoder is merely used to demonstrate how a difference in reaction times emerges from our model.

Inspired by previous models of perceptual decision making [Gold and Shadlen, 2007; Meckenhäuser et al., 2014], the decoder has a decision variable associated with each cluster. For each direction d a leaky integrator governed by an equation of the form

$$\frac{dI_d(t)}{dt} = -\frac{I_d(t)}{\tau_I} + C_d \tag{5.2}$$

sums up the instantaneous spike counts C_d (in $1ms$ bins) of the corresponding population and *forgets* with time constant τ_I. Due to the random nature of the

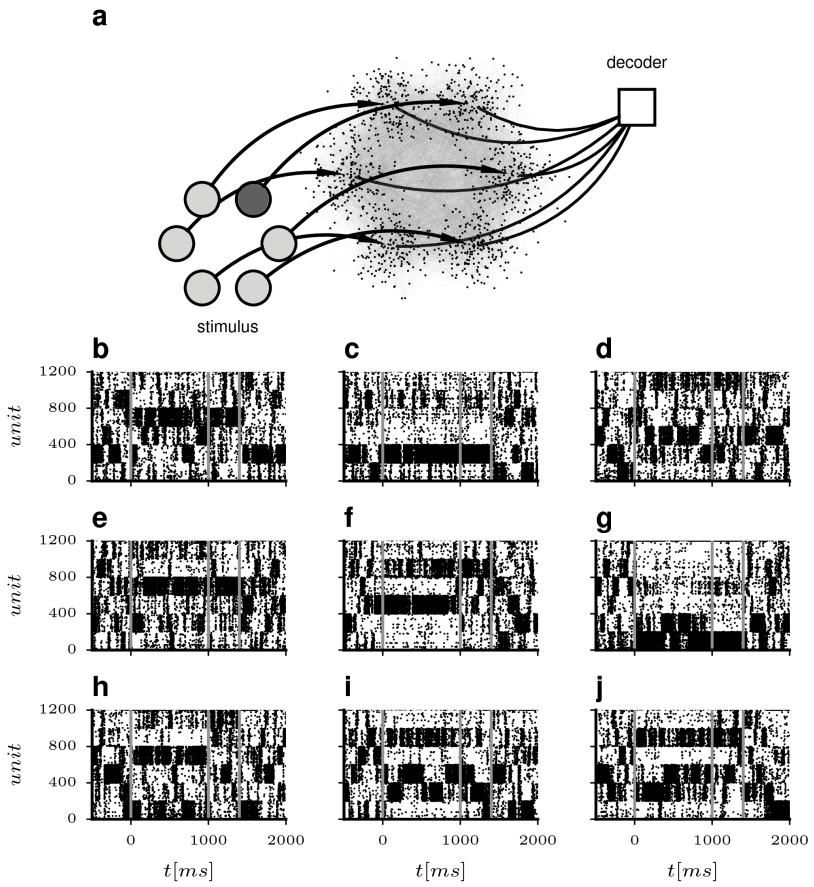

Figure 5.3: **a**) Architecture of the model. Spiking activity is modelled in a balanced network with parameters given in table 5.1. Each cluster is assigned a direction for which it receives its external input. The decoder integrates the average cluster rates and generates a decision. **b-j**) Raster plots showing sample trials of the excitatory units. Preparatory stimuli of $0.15pA$ are applied to one (**b-d**), two (**e-g**) or three (**h-j**) clusters simultaneously between 0 and $1000ms$. A response signal of the same amplitude is injected into single clusters from 1000 to $1400ms$ as indicated by the vertical lines.

balanced network, the firing rates of the individual clusters may vary considerably. To enable a decision by applying a single threshold θ to all directions, a decision variable $DV_d(t)$ is formed for all directions through normalising by the sum over all directions:

$$DV_d(t) = \frac{I_d(t)}{\sum_{j=1}^{6} I_j(t)} \tag{5.3}$$

This is similar to multi-class classification and the decision variable $DV_d(t)$ expresses the probability that direction d is the correct choice at time t. The

decoder was *armed* at RS. A decision was reached when one of the decision variables crossed the threshold in the interval where the RS was presented ($1000ms < t <= 1400ms$). Threshold crossings after that period were counted as unsuccessful trials. If a decision variable was already above threshold at the beginning of RS, the decision was counted. The threshold θ was adjusted so that the performance of the model was maximised. The time constant of integration τ_I was set to $50ms$ which represents an intermediate value between very fast *reactions* directly when the threshold is *armed* at RS and very slow integration where the threshold was not reached during the RS-interval.

Figure 5.3 b through j shows raster plots of the excitatory populations for sample trials generated from the model. Our interpretation is that a cluster in the active state means that the model is *prepared* for a movement execution in the corresponding direction. I.e. it occupies the *optimal subspace* of firing rates for a certain movement as described by Churchland et al. [2006]. It can be seen in panels b through d that the stimulation amplitude is fairly moderate. Rather than forcing a direction cluster to be active the whole time it merely increases its probability of being active. Consequently there is still considerable cycling between clusters during the PS interval. This is consistent with our observation in the data. Even for condition 1 the rate variance is not completely quenched by the stimulus presentation as can be seen by the still relatively high Fano factor in figure 5.2 b (see also figure 2.7 c,e on page 19). In condition 2, two clusters at a time are simultaneously stimulated and competition arises between them. Since two clusters now have to *share* the higher probability of being active, the number of occurring rate configurations increases and thereby the reduction in rate variance is weaker. The amount of rate variance in this model increases with the number of stimulated clusters. If all clusters are stimulated at the same time, their chances during winnerless competition are again equal and the rate variance level of the spontaneous state is recovered. Hence for condition 3 the variance is again higher than for condition 2.

The raster plots also reveal another central aspect of the model. In condition 1, the correct direction has a high change of already *being prepared*, i.e. of the corresponding cluster being in the active state at the time of the RS. In the ambiguous conditions 2 and 3 on the other hand the chance of currently being *unprepared* is higher and a switch is often required to the target direction. A correct movement can therefore in those conditions only be executed with a longer reaction time.

Figure 5.4 shows the same analyses as shown for the physiological data (fig. 5.2) computed on the spike trains generated b the model. To allow a fair comparison the same selection conditions were applied as for the experimental

data. Since the number of trials was the same for all units, only the minimum count rate criterium had an effect (10 spikes in the inspected time window). Consequently 474 of the 1200 excitatory units were included in our analysis.

The exact time course of firing rates is not captured by our simple model (panel a). As expected, our model of probabilistic sampling of prepared directions produces average rate profiles that resemble the *box*-shape of our stimulus protocol. The model does however capture the condition dependent modulation of the average FF. In the spontaneous state before the PS is turned on, the count variance is high in all conditions. During the preparatory period the Fano factors are differentially quenched for the three target conditions as in the physiological data (fig. 5.2). The differences are also highly significant with respect to a Wilcoxon signed rank test (conditions 1-2: $p = 5.7 * 10^{-9}$, conditions 2-3: $p = 4.8 * 10^{-11}$). At the RS onset, a single stimulus is applied in all conditions and the FFs converge to the same value. The Fano factors for condition 2 and 3 exhibit a second strong decrease after RS in the physiological data. This is again not captured by our model. A possible reason might be that the distributions of reaction times are much narrower for conditions 2 and 3 (see figure 5.2 e), producing more *aligned* firing rate over trials.

The probabilistic nature of switching between prepared directions does produce average decoding accuracies for single neurons which resemble those seen in the data (panel c). Although the balanced accuracy scores are slightly higher in the model, the order of conditions is reproduced. When classification is performed on the entire population of analysed units simultaneously, we recover the optimal decoding scores of 1, $1/2$ and $1/3$ for the three conditions (panel d).

Finally, the operation of the decoder is illustrated in figure 5.3. The threshold has been adjusted so that the balanced accuracy over all sessions was maximised (panel b). Figure 5.3 a shows a raster plot of the excitatory activity for a sample trial of condition 3. The PS current is applied to clusters 2,3 and 5 (counting from the bottom). The respective decision variables $DV(t)$ are superimposed on each direction. Due to competition between populations, all decision variables remain below the threshold in this case as a longer period of being in the active state is required due to the time scale of integration. When the RS is given, the *wrong* cluster is currently in the active state. It takes a short moment until the stimulus takes effect an the network activity switches to the target cluster. An even longer time is taken for the decision variable to reach the threshold. In this particular trial a *reaction time* of $\sim 200ms$ is thus produced.

Figure 5.3 c shows the distributions of reaction times produced by the model for each condition. It can be seen that as in the experimental data the average reaction times in condition 1 were much quicker than in conditions 2 and 3.

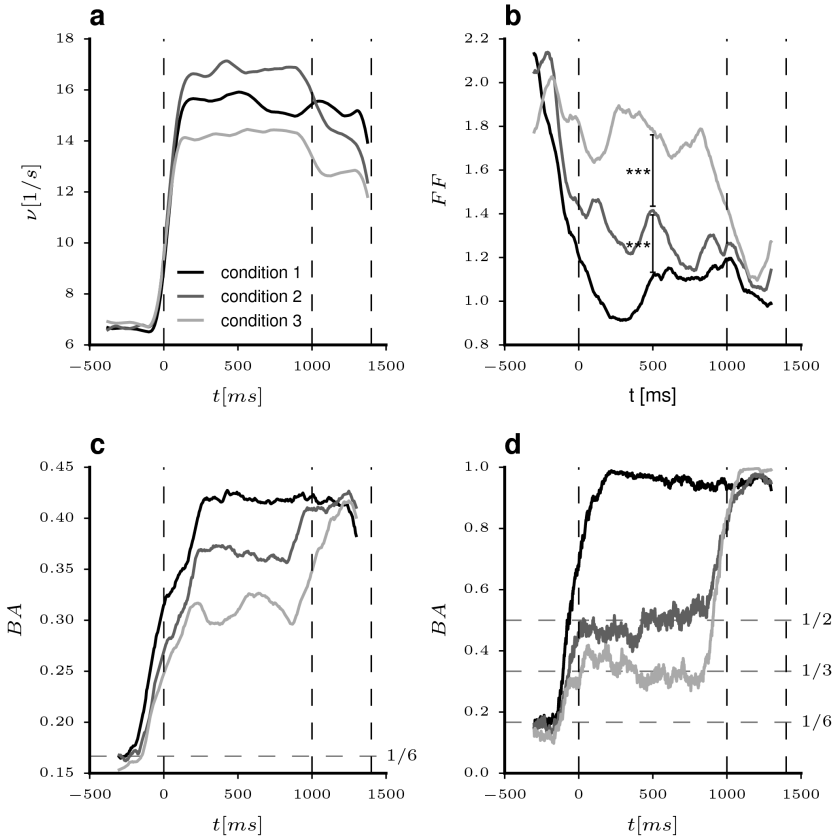

Figure 5.4: Condition-wise analyses corresponding to figure 5.2 but for the spiking network model. Vertical dashed lines represent stimulus events.**a**) Firing rates estimated with a triangular kernel ($\sigma = 50ms$). **b** Fano factors computed in sliding windows of $400ms$. $***$ indicates significance from Wilcoxon signed rank test (conditions 1-2: $p = 5.7 * 10^{-9}$, conditions 2-3: $p = 4.8 * 10^{-11}$) **c**) Balanced accuracy of direction classification using single unit counts in windows of $400ms$. Lines show averages over 10 repetitions of 5-fold cross validation. **d**) Balanced accuracy of direction classification using the counts of all units as features (10 repetitions of 5-fold cross validation).

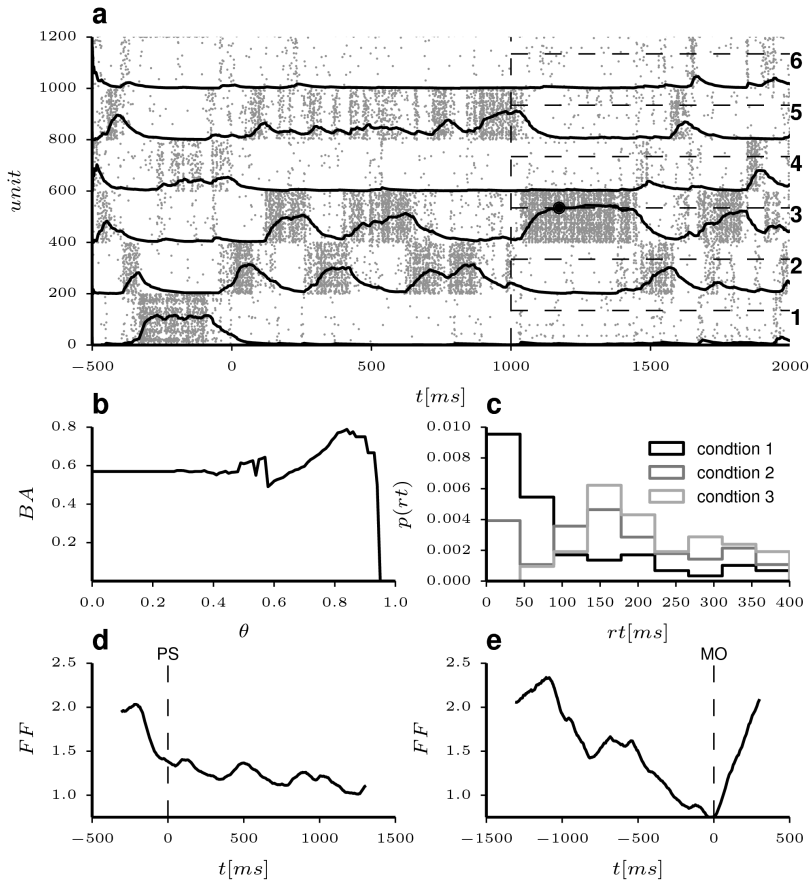

Figure 5.5: Illustration of decision model and Fano factor dynamics. **a**) Raster plot of excitatory cluster activity for a sample trial of condition 3. Solid lines represent the associated decision variable for each cluster. At $t = 1000ms$ the decoder becomes *armed* as indicated by the vertical dashed line. Horizontal dashed lines show the level of the decision threshold θ. To the right, direction numbers are indicated. During the preparatory phase, clusters 2,3 and 5 are stimulated. Target is direction 3. The black dot indicates a successful threshold crossing. **b**) Decodiung performance of the model over the possible range of θ. **c**) Histograms of *reaction times* of the model for the three conditions. The lower two panels show the average FF of the model units for condition 2 when aligned with respect to stimulus presentation (**e**) or the time of threshold crossing (i.e. the *movement onset*, panel **f**)

In contrast to the experiment, anticipated responses were not penalised in the model. If the decision variable of the correct direction was already above threshold at RS, the trial was counted as successful. In condition 1 this was frequently the case. The shape of the reaction time histogram for condition 1 in figure 5.2 e suggests that the monkey displayed a similar behaviour. The difference in response times between conditions 2 in 3 is not discernible in the model as in the data. The chance of having *prepared* for the wrong direction in the model hence explains the difference in reaction times between the full information and the ambiguous conditions. Note that the histograms in figure 5.2 look much smoother because they have been computed over many sessions with numerous trials while in our model only one session with 150 trials per condition was simulated.

Analogous to the experimental data, the reaction times can be used to align the trials of simulated data to an *intrinsic* event - the crossing of the decision threshold or *movement onset*. Figure e and f show the Fano factors of the simulated data aligned to PS and MO respectively. As in the physiological data (see figure 2.7 c,d on page 19) the Fano factor is lowest when aligned to MO.

Although our model is very simplified and captures only some aspects of the observed dynamics during movement preparation, it presents a mechanistic explanation for the models proposed by Churchland et al. [2006] and Afshar et al. [2011]. To our knowledge, it is the first model that links attractor dynamics to context dependent variability modulation in motor cortex and behavioural variability in the form of reaction times.

Chapter 6

Discussion

The current study has spanned a range of aspects of cortical variability from basic measurement techniques to a model of a complex experimental task. In chapter 2 we have discussed how to carefully quantify the variability in spike counts and interval variability which has allowed us to draw some conclusions about potential mechanisms which produce these dynamics. In chapter 3 we have employed the mean field model of balanced networks of binary neurons to show how clustering of the inhibitory as well as excitatory populations can reduce the firing rates of individual clusters and therefore facilitate winnerless competition dynamics over a wider range of parameters. In chapter 4 we have transferred this network structure to more detailed spiking neuron models and found that inhibitory clusters lead to cluster switching dynamics which remain balanced even in the active states, thus conserving interval variability contrary to previous models. Finally we employed this model to replicate some findings from an experiment where monkeys had to perform a reaching task given various degrees of previous information about the target direction. In the following we will recapitulate the main results of this work and discuss some of the problems and implications in more detail.

Relationship of Rate Variance and VarCE In chapter 2 we have seen how the description of neural activity as a doubly stochastic point process allows us to see aspects of cortical dynamics which would remain hidden if we only examined average rate profiles and Fano factors. The dissection of neural variability into rate variance and intrinsic variability is directly related to the variance of the conditional expectation ($VarCE$) of spike counts introduced by Churchland et al. [2011]. Instead of approximating the point process variance of the spike trains with the lowest FF a neuron exhibited over time we have shown

that the trial wise CV^2 represents this quantity. This relationship is however tied to a number of assumptions.

Firstly, the theoretical prediction that $FF = CV_\infty^2$ holds only for renewal processes. This does not hold strictly for neural spike trains as indicated by negative serial interval correlations [Nawrot et al., 2007]. The assumptions in calculating the CV^2 are however likely to play a greater role. The rate unwarping method for its time resolved estimation is based on the assumption that spike trains in each trial are realisations generated by some process from an identical intensity profile. One of the central results of chapter 2 is that firing rates modulate over trials which is why we introduced the trial wise estimation of the CV^2. This again assumes that the underlying rate modulations happen so slowly that they can be regarded to be constant for the width of an estimation window. Since we compute the CV^2 in operational time and require our estimation windows to be at least ten intervals, the low cortical firing rates lead to estimation windows which - if projected back to real time - in some cases correspond to the entire length of a trial. Additionally, although we have corrected for a systematic bias due to right-censoring of the interval distribution, the CV^2 is slightly under-estimated if computed per trial. Our choice of making use of the formal relationship of the CV^2 to the FF is however justified as shown by the qualitative resemblance with the local interval statistics CV_2 and L_V which do not have these problems. In our dataset all three quantities are relatively constant over time on average.

In principle it could be possible to compute the rate variance from instantaneous single trial estimates of firing rate and irregularity as described by Shimokawa and Shinomoto [2009]. Unfortunately the large number of spikes required to achieve stable results and the computational cost associated with the numerical solution make this approach infeasible.

Churchland et al. [2011] state that the $VarCE$ is qualitatively similar to the mean-matched Fano factor. Mean-matching was proposed by Churchland et al. [2010] to avoid detecting FF reductions which are only due to an increased mean rate while rate variance stays constant (the case on the $l.h.s.$ in figure 2.4, page 15). It works by computing the distribution of mean counts of a group of neurons in a range of time windows. To avoid computing the FF for different mean counts, a common count distribution is then found and in each time window units are randomly excluded until a satisfactory match with this distribution is reached. The obvious drawbacks of this technique are that it cannot be applied to single neurons and that the mean FF is potentially computed for a different combination of units in each time step. Our method of computing the rate variance from the raw FF and CV^2 overcomes these

problems.

Significant Reduction in Rate Variance The simplest model that could explain the discrepancy between rate and interval variability is that of a very slowly varying background state where rate variance is virtually constant for the duration of a trial [Nawrot, 2010]. In figure 2.4 we have simulated this case by adding a constant random offset in each trial to a fixed firing rate profile. Using the rate variance, we have made it explicit that this effect is not sufficient. The reduction in total variance as measured by the Fano factor is caused by a highly significant reduction in underlying rate variance while interval variability is largely conserved in our data (see figure 2.7).

This puts the interpretation of a somehow varying *background-state* or ongoing activity onto which a task dependent firing rate profile is superimposed into question. The observation that the variance modulates on the time scale of the experimental task suggests that there must be at least some component of ongoing activity that varies at a faster rate. The quenching of rate variability could then be achieved in two ways. Either the activity in the network is temporarily suppressed or the firing rates in individual trials become more aligned to the experimental events.

Both of these effects can be explained by winnerless competition between assemblies or clusters of cells [Deco and Hugues, 2012; Litwin-Kumar and Doiron, 2012; Lagzi and Rotter, 2015]. The rate variance caused by the switching of these clusters between states of high and low firing rates is reduced if one or more cluster receive some external drive and become more likely to remain in the active state. Inhibition then causes the remaining clusters to stay mostly in the low activity states. In an experimental set-up, the former units would likely be identified as *task* related while the latter may not even be detected by spike sorting. Also, if interpreted in this way the task dependent trial averaged firing rate could consist of combinations of *up* and *down* state periods of the assemblies that the observed units belong to. Rather than a stereotypical rate profile that is in each trial disturbed by noise, the average firing rate could then be interpreted as a correlate of the probability of certain assemblies being active at certain times during trials.

Physiological Evidence for Inhibitory Clustering Balanced networks with clustered connectivity in the excitatory population exhibit winnerless competition dynamics if the parameters are chosen appropriately [Deco and Hugues, 2012; Litwin-Kumar and Doiron, 2012; Mazzucato et al., 2015]. This usually means adjusting the clustering parameter, in our case J_+. Extensive exploration has

shown that the cluster parameter has to be quite finely tuned to achieve cluster switching at all. Switching of activity states is triggered by fluctuations in cluster firing rate. Hence the probability of switching decreases when clusters become larger because fluctuations tend to average out [Litwin-Kumar and Doiron, 2012]. Mazzucato et al. [2015] used 30 clusters of 120 units ($N_E = 4000$, 10% unstructured background populations) whereas Litwin-Kumar and Doiron [2012] used 50 clusters of 80 units ($N_E = 4000$). While fine-scale clusters have been reported to consist of tens rather than thousands of neurons [Perin et al., 2011], it is hard to believe that the cortical algorithm should already break down with the clusters of 200 units used in the present study.

Further, although excitatory clustering can produce the rate variance seen in spontaneous cortical activity and its reduction during stimulation, firing in active clusters is more regular than observed in cortex [Litwin-Kumar and Doiron, 2012] and the CV^2 is also reduced by stimulation [Deco and Hugues, 2012] which is in contrast with our findings in monkey motor cortex.

By analysing the stable fixed points of the mean-field equations for networks with excitatory clusters we have shown in chapter 3 that switching is hampered by the high rates that active clusters attain. Our proposed remedy of assigning an inhibitory population to each cluster and to increase the strength of reciprocal connections with respect to those connecting to other clusters solved both of the above problems. Since inhibition is now also selective, each excitatory population is held in check by its inhibitory counter part and hence the fixed points of the active and passive clusters move closer together. This facilitates switching even for larger clusters. In chapter 4 we have seen that this also helps to maintain interval variability because the active clusters remain in the fluctuation-driven balanced state rather than being forced into saturation. Aviel et al. [2004] have assigned *inhibitory shadow patterns* to excitatory populations to increase the capacity of an attractor based memory network. The positive effect of local balance on winnerless competition dynamics have however to our knowledge not been discussed previously.

In short, we have introduced inhibitory clustering to achieve a certain type of network dynamics. But can inhibitory clusters be justified biologically?

For excitatory neurons, *small world* structures have been reported on many spatial scales [Sporns and Zwi, 2004]. Bidirectional connections as well as clustered three-neuron patterns are much more frequent than would be expected in a random network [Song et al., 2005]. These clusters also tend to have stronger connections.

In network simulations, the units are commonly only divided into excitatory and inhibitory cells while anatomical studies have identified myriads of inhibitory

interneurons (e.g. [Markram et al., 2004; Harris and Shepherd, 2015]) and a single excitatory neuron may make connections with several types [Markram et al., 1998]. An intermediate level of detail is the distinction between Fast-Spiking (FS) and non-FS inhibitory interneurons. These are usually identified in slice preparations where intracellular recordings are taken while the surrounding tissue is stimulated (e.g. [Yoshimura and Callaway, 2005; Holmgren et al., 2003; Kätzel et al., 2011; Levy and Reyes, 2012; Dantzker and Callaway, 2000; Otsuka and Kawaguchi, 2009; Watts and Thomson, 2005])

In addition to the different spiking behaviour implied by their names these two cell types exhibit functional differences. Fast spiking cells are mainly locally connected while non-FS neurons make translaminar connections [Dantzker and Callaway, 2000; Levy and Reyes, 2012; Kätzel et al., 2011; Otsuka and Kawaguchi, 2009]. On a finer scale, most connections between excitatory and FS-interneurons are reciprocal [Holmgren et al., 2003] with inhibitory post-synaptic currents being three times larger in reciprocal connections than for unidirectional ones [Yoshimura and Callaway, 2005]. With distance from an excitatory unit in a small volume, the connection probability with local inhibitory units varies only slightly but the fraction of reciprocal connections decreases [Holmgren et al., 2003]. Reciprocally connected cell pairs also share more common input than non-connected or unidirectionally connected pairs while non-FS cells share little common input, connect to excitatory units with lower probability and reciprocal connections are rare and not stronger than unidirectional ones [Yoshimura and Callaway, 2005]. Also, interneurons of different types are frequently connected with each other [Reyes et al., 1998] and inhibition can also be exerted bisynaptically so that excitatory axons excite inhibitory cells local to other populations [Binzegger et al., 2005].

It is hence safe to say that there is a lot of structure in the inhibitory cortical connectivity. The strong reciprocal local inhibition of the FS-cells and the weaker longer range connections of the non-FS interneurons could be interpreted as an argument for the type of inhibitory clustering we have proposed. Whether inhibition is less localised than excitation as predicted by our model cannot be conclusively answered at this time. The current physiological evidence certainly does not rule out the possibility.

Time scales of Variability Dynamics A recurring theme in the context of cortical variability are the time scales on which neural activity varies. This work is centred around the hypothesis that rate variance across trials is a consequence of slow switching between clusters of neurons. In our model, we we have observed that switching tends to occur less frequently when the

clustering is stronger, thereby increasing the time that assemblies spend in the active state. Also, larger assemblies tend to lead to slower switching [Litwin-Kumar and Doiron, 2012].

A number of physiological results support the hypothesis that spontaneous activity is made up of sequences of structured activity patterns which emerge seemingly at random.

Kenet et al. [2003] used voltage sensitive dye imaging in the visual cortex of anaesthetised cats to show that spontaneous activity seems to cycle through patterns very similar to those evoked by visual stimuli and reported that transitions between patterns occurred on a time scale of around $80ms$. Berkes et al. [2011] recorded extracellular activity using multi electrode arrays in the visual cortex of developing ferrets. They found that over the course of development, the spontaneous activity became increasingly similar to evoked responses. They measured strong correlations between activity patterns for delays up to $20ms$. Ponce-Alvarez et al. [2012] fitted a hidden Markov model (HMM) to simultaneously recorded spike trains from somatosensory and motor cortex in monkeys. During the delay period in a perceptual decision task they observed switching through sequences of states with transition times on the order of $50ms$. Mazzucato et al. [2015] applied a similar analysis to spontaneous recordings from gustatory cortex in rats. Their HMM analysis yielded exponentially distributed state durations with a mean of $\sim 700ms$. Luczak et al. [2009] examined evoked and spontaneous activity in the auditory and somatosensory cortices of rats and found not only spatial patterns but temporally stereotyped profiles with durations of $\sim 100ms$.

These results point to time scales on the order of tens to hundreds of milliseconds. Teich et al. [1997] on the other hand have found that Fano factors of retinal ganglion cells in cats increase with counting window width for observation intervals of several minutes. For stationary point processes, the FF depends only weakly on window size [Nawrot, 2010]. When rate variance is added to the equation, the Fano factor will increase with counting window size until the window spans several periods of the periodicity of the underlying fluctuations. Such long time scale fluctuations do not necessarily have to originate from the cluster switching mechanisms described here. It is likely that winnerless competition dynamics in spontaneous cortical firing will happen on time scales related to the stimulus modalities or movements a particular region codes for.

As mentioned above, the time scales of switching in our model depend on the setting of the cluster parameters J_{E+} and R_J. We have adjusted these parameters so that the networks' spontaneous activities would exhibit Fano factors similar to those observed in our data set. We have not measured the

durations that attractors spend in the active states, although a HMM analysis similar to those by Ponce-Alvarez et al. [2012] and Mazzucato et al. [2015] could yield estimates. Inspection of the raster plots does however suggest that the *up* states in our model have similar time scales as those reported in the literature ($\sim 20 - 200ms$).

It would in principle be desirable to have a theoretical prediction for the time scales of cluster switching in relation to the model parameters. Lagzi and Rotter [2015] have described winnerless competition between two populations using a rate model governed by Lokte-Volterra type equations. If noise is introduced, switching can also occur in rate models. They found that the survival times of the active states could be well approximated by an exponential distribution and that the average time between switches grows faster than exponentially with cluster strength.

We have used the mean-field description of our network configuration to find the stable rate configurations. As mentioned earlier, switching between these stable attractors is a finite size effect due to chaotic fluctuations in the firing of individual units in the populations which is by definition not captured by the mean field approach. The mean field theory for networks of binary units also predicts the distribution of activity rates within populations [van Vreeswijk and Sompolinsky, 1998]. From this it is theoretically possible to compute the $2Q$-dimensional joint distribution of cluster rates. Making some assumptions about the *noise* caused by the rate fluctuations within clusters it may then be possible to make predictions about the switching dynamics. This analysis was however beyond the scope of the current study.

Cluster Dynamics as Probabilistic Inference So far we have discussed the dynamics of activity switching between clusters only from the perspective that it provides a mechanistic explanation for the high rate variance observed in cortex and its quenching by stimulation.

Traditionally, stable patterns or attractors in networks are used to model working memory (e.g. [Hopfield, 1982; Amit and Brunel, 1997; Aviel et al., 2004]). A network is *pushed* into a certain attractor by some external drive and maintains a certain configuration of firing rates which can be read out or *retrieved* at a later point. The difference between this scenario and the dynamics of winnerless competition is simply the *depth* of the attractors as characterised by the strength of clustering in the connectivity structure. This relationship lends itself to the interpretation that the stability of attractors is related to the probability of some variable encoded in their firing rate.

Mounting physiological indications exist for the hypothesis that some cor-

relate of prior probability over previously observed states is encoded in the spontaneous activity in the neocortex. Berkes et al. [2011] report that the spontaneous activity in the visual cortex of ferrets before eye-opening is unstructured and becomes more and more similar to that evoked by visual stimuli during development. They consequently interpret the spontaneous activity as the prior probability of observing certain patterns and fit a Bayesian model to their data. This observation is well matched by the results of Kenet et al. [2003] described above and Luczak et al. [2009] who found that in auditory and somatonsensory cortex of rats, spontaneous states resemble those evoked by sensory stimulation.

There is also more direct evidence that these priors are actually used during perception. Supèr et al. [2003] report that significant differences were detected in the firing rates in monkey visual cortex between trials where monkeys correctly reported the occurrence of a stimulus and those where they missed it. Similarly, Hesselmann et al. [2008] found that perceptual decisions can be predicted from ongoing activity in fMRI signals in humans 1.5 seconds prior to stimulus presentation in a face-or-vase task. The interpretation is here that the attractor that a network is currently in influences what decision is made.

We have applied the concept of winnerless competition to a model of movement preparation with varying degrees of information. Although it is not a perceptual decision task, the ambiguity between a small number of choices creates a comparable situation. The preparatory signal causes the activity in the motor cortices of the highly trained animals to be confined to a smaller subspace of possible dynamics. This subspace can be made smaller when more information is available. In condition 2 for example the activity cycles randomly between the two *ideal initial conditions* [Churchland et al., 2006; Afshar et al., 2011] for movement preparation. If the correct cluster happens to be active at the response signal, the movement can be executed with shorter reaction times.

Similar to [Supèr et al., 2003], additional evidence for or against this hypothesis could have been obtained if the trials where the animals reached for the wrong target had been available by comparing the spike counts just before movement onset. It is clear that our model only captures some aspects of the dynamics observed in the data. The firing rates in the physiological data show an overshoot effect at PS for condition 1 and a strong increase at RS for conditions 2 and 3. Also, some of the units show persistent activity in the preparatory period [Rickert et al., 2009] which could mean that they function as a working memory [Amit and Brunel, 1997]. While our model predicts different Fano factors for the three conditions which are constant for the duration of the preparatory period, the FFs in the physiological data show a sharp dip at the

onset of the stimulus which then seems to relax to a different level for each condition. This could possibly be explained by spike frequency adaptation (SFA). Farkhooi et al. [2013] have proposed an alternative (possibly complementary) model of stimulus-induced variability reduction. Their model does not include rate variance and can therefore not capture the differential suppression of count variance. It indicated however, that the introduction of SFA into our model could lead to better agreement with our data. It is likely that SFA would also influence the spontaneous dynamics of cluster switching. This effect has not been investigated in the current study.

Rickert et al. [2009] also found that some of the neurons seem to alter their preferred direction at certain times during the trial. In our model of random switching, systematic changes in tuning cannot occur. Of course in real brains, the attractors do not form randomly but are shaped by sensory input and experience or training in the case of motor experiments. Hence the sequences of activated clusters in spontaneous states or preparatory phases are likely to have meaningful structure. The strict division of units into clusters assigned to directions is of course also a strong simplification. [Litwin-Kumar and Doiron, 2012] have shown that cluster cycling and winnerless competition can also occur when units are randomly assigned to overlapping assemblies.

Outlook Attractor networks are a popular model for cortical computation and can capture working memory (e.g. [Amit and Brunel, 1997]), decision making (e.g. [Albantakis and Deco, 2011] and variability dynamics (e.g. [Litwin-Kumar and Doiron, 2012]) depending on the strength of the attractors. We hypothesize that cortical circuits can switch between these functions continuously depending on the requirements. This could be achieved through some form of neuromodulation.

At the moment, only indirect measures can be employed to asses the validity of the cluster switching model such as the HMM technique used by Mazzucato et al. [2015]. In the near future it is however likely that simultaneous recordings of large numbers of neurons in very small volumes can be made in vivo using calcium imaging or similar techniques.

Contrary to our fixed random network connectivity, structure is shaped by synaptic plasticity in vivo. A number of studies have recently been published, where clusters of excitatory units form in balanced networks through spike-timing dependent plasticity (STDP) and selective stimulation [Ocker et al., 2015; Zenke et al., 2015; Litwin-Kumar and Doiron, 2014]. The resulting connectivities exhibit high rate variability in the spontaneous state. In a related study, binary networks with synchronous updates using STDP-inspired as well as homeostatic learning

rules were shown to perform Bayesian-like inference in sequence-learning tasks [Lazar et al., 2009].

In all the above studies, some form of inhibitory plasticity was used as a homeostatic mechanism. This seems to be generally required for networks with excitatory plasticity to prevent positive feedback loops [Zenke et al., 2013]. For example [Litwin-Kumar and Doiron, 2014] use an inhibitory STDP rule for the EI-connections, i.e. the synapses that mediate inhibition from the inhibitory to the excitatory population to prevent winner-take-all dynamics. Our results predict that the EI as well as the IE connections are to some extent plastic to achieve the specificity required to obtain local balance in each assembly. It remains a subject of further study how this can be achieved in simulations and whether it can be found in biological circuits.

Bibliography

Afshar, A., Santhanam, G., Yu, B. M., Ryu, S. I., Sahani, M., and Shenoy, K. V. (2011). Single-trial neural correlates of arm movement preparation. *Neuron*, 71(3):555–64. 62, 70, 78

Albantakis, L. and Deco, G. (2011). Changes of mind in an attractor network of decision-making. *PLoS computational biology*, 7(6):e1002086. 79

Amit, D. and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, 7:237–252. 29, 43, 77, 78, 79

Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science*, 273(5283):1868. 1, 11

Aviel, Y., Horn, D., and Abeles, M. (2004). The Doubly Balanced Network of Spiking Neurons: A Memory Model with High Capacity. *Advances in Neural Information Processing Systems 16*, pages 1247–1254. 74, 77

Baker, S. N. and Gerstein, G. L. (2001). Determination of response latency and its application to normalization of cross-correlation measures. *Neural computation*, 13(6):1351–1377. 6

Barlow, R. (1989). *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*. Manchester Physics Series. Wiley. 13

Bastian, A., Schoner, G., and Riehle, A. (2003). Preshaping and continuous evolution of motor cortical representations during movement preparation. *European Journal of Neuroscience*, 18(7):2047–2058. 16, 57, 58, 64, 91

Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science (New York, N.Y.)*, 331(6013):83–7. 2, 76, 78

Binzegger, T., Douglas, R. J., and Martin, K. A. C. (2005). Cortical architecture. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3704 LNCS:15–28. 75

Bose, A. and Kunec, S. (2001). Synchrony and frequency regulation by synaptic delay in networks of self-inhibiting neurons. *Neurocomputing*, 40:505–513. 46

Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E., and Frank, L. M. (2001). The time-rescaling theorem and its application to neural spike train data analysis. *Neural computation*, 14(2):325–46. 6

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of computational neuroscience*, 8(3):183–208. 21, 23, 45, 46

Churchland, A. K., Kiani, R., Chaudhuri, R., Wang, X.-J., Pouget, A., and Shadlen, M. N. (2011). Variance as a signature of neural computations during decision making. *Neuron*, 69(4):818–31. 2, 12, 13, 14, 71, 72

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., Newsome, W. T., Clark, A. M., Hosseini, P., Scott, B. B., Bradley, D. C., Smith, M. a., Kohn, A., Movshon, J. A., Armstrong, K. M., Moore, T., Chang, S. W., Snyder, L. H., Lisberger, S. G., Priebe, N. J., Finn, I. M., Ferster, D., Ryu, S. I., Santhanam, G., Sahani, M., and Shenoy, K. V. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature neuroscience*, 13(3):369–378. 2, 11, 14, 52, 72, 94

Churchland, M. M., Yu, B. M., Ryu, S. I., Santhanam, G., and Shenoy, K. V. (2006). Neural Variability in Premotor Cortex Provides a Signature of Motor Preparation. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(14):3697–3712. 57, 62, 66, 70, 78

Cox, D. and Isham, V. (1980). *Point Processes*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis. 2, 6, 7, 11, 12

Dantzker, J. L. and Callaway, E. M. (2000). Laminar sources of synaptic input to cortical inhibitory interneurons and pyramidal neurons. *Nature neuroscience*, 3(7):701–7. 75

Dayan, P. and Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press. 6

Deco, G. and Hugues, E. (2012). Neural network mechanisms underlying stimulus driven variability reduction. *PLoS computational biology*, 8(3):e1002395. 3, 20, 22, 30, 52, 53, 73, 74

Doiron, B. and Litwin-Kumar, A. (2014). Balanced neural architecture and the idling brain. *Frontiers in computational neuroscience*, 8(May):56. 22

Faisal, a. A., Selen, L. P. J., and Wolpert, D. M. (2008). Noise in the nervous system. *Nature reviews. Neuroscience*, 9(4):292–303. 1, 21

Fano, U. (1947). Ionization yield of radiations. II. the fluctuations of the number of ions. *Physical Review*, 72(1):26–29. 7

Farkhooi, F., Froese, A., Muller, E., Menzel, R., and Nawrot, M. P. (2013). Cellular Adaptation Facilitates Sparse and Reliable Coding in Sensory Pathways. *PLoS Computational Biology*, 9(10):e1003251. 79

Gerstner, W., Kistler, W. M., Naud, R., and Paninski, L. (2014). *Neuronal dynamics: From single neurons to networks and models of cognition.* Cambridge University Press. 33

Gewaltig, M.-O. and Diesmann, M. (2007). NEST (NEural Simulation Tool). *Scholarpedia*, 2(4):1430. 43

Gold, J. I. and Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annu. Rev. Neurosci.*, 30:535–574. 64

Golomb, D. and Hansel, D. (2000). The number of synaptic inputs and the synchrony of large, sparse neuronal networks. *Neural computation*, 12(5):1095—-1139. 45

Gur, M., Beylin, a., and Snodderly, D. M. (1997). Response variability of neurons in primary visual cortex (V1) of alert monkeys. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 17(8):2914–2920. 11

Harris, K. D. and Shepherd, G. M. G. (2015). The neocortical circuit: themes and variations. *Nature Neuroscience*, 18(2):170–181. 75

Hesselmann, G., Kell, C. A., Eger, E., and Kleinschmidt, A. (2008). Spontaneous local variations in ongoing neural activity bias perceptual decisions. *Proceedings of the National Academy of Sciences of the United States of America*, 105(31):10984–9. 78

Holmgren, C., Harkany, T., Svennenfors, B., and Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *The Journal of physiology*, 551(Pt 1):139–53. 75

Holt, G. R., Softky, W. R., Koch, C., and Douglas, J. (1996). Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *Journal of Neurophysilogy*, 75(5):1806–1814. 11

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. 77

Jones, E., Oliphant, T., Peterson, P., et al. (2001–). SciPy: Open source scientific tools for Python. [Online; accessed 2016-01-27]. 32

Kara, P., Reinagel, P., and Reid, R. C. (2000). Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron*, 27(3):635–646. 1, 21

Kätzel, D., Zemelman, B. V., Buetfering, C., Wölfel, M., and Miesenböck, G. (2011). The columnar and laminar organization of inhibitory connections to neocortical excitatory cells. *Nature Neuroscience*, 14(1):100–107. 75

Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A., and Arieli, A. (2003). Spontaneously emerging cortical representations of visual attributes. *Nature*, 425(6961):954–6. 2, 76, 78

Kriener, B., Enger, H., Tetzlaff, T., Plesser, H. E., Gewaltig, M.-O., and Einevoll, G. T. (2014). Dynamics of self-sustained asynchronous-irregular activity in random networks of spiking neurons with strong synapses. *Frontiers in computational neuroscience*, 8(October):136. 23, 46

Lagzi, F. and Rotter, S. (2015). Dynamics of Competition between Subnetworks of Spiking Neuronal Networks in the Balanced State. *PloS one*, 10(9):e0138947. 20, 22, 73, 77

Lazar, A., Pipa, G., and Triesch, J. (2009). SORN: a self-organizing recurrent neural network. *Frontiers in computational neuroscience*, 3(October):23. 80

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. 48

Levy, R. B. and Reyes, A. D. (2012). Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex. *J Neurosci.*, 32(16):5609–5619. 75

Litwin-Kumar, A. and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature neuroscience*, pages 1–2. 3, 20, 22, 23, 30, 32, 36, 44, 48, 51, 52, 53, 73, 74, 76, 79

Litwin-Kumar, A. and Doiron, B. (2014). Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature communications*, 5(May):5319. 23, 79, 80

Luczak, A., Bartho, P., and Harris, K. D. (2009). Spontaneous Events Outline the Realm of Possible Sensory Responses in Neocortical Populations. *Neuron*, 62(3):413–425. 2, 76, 78

Mainen, Z. and Sejnowski, T. (1995). Reliability of spike timing in neocortical neurons. *Science*, 135(1993):1–4. 1, 2, 21

Marcos, E., Pani, P., Brunamonti, E., Deco, G., Ferraina, S., and Verschure, P. (2013). Neural variability in premotor cortex is modulated by trial history and predicts behavioral performance. *Neuron*, 78(2):249–55. 57

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nature reviews. Neuroscience*, 5(10):793–807. 75

Markram, H., Wang, Y., and Tsodyks, M. (1998). Differential signaling via the same axon of neocortical pyramidal neurons. *Proc Natl Acad Sci U S A*, 95(9):5323–8. 75

Mascaro, M. and Amit, D. (1999). Effective neural response function for collective population states. *Network: Computation in Neural . . .* , 6536(November). 33, 34, 35

Masquelier, T. (2013). Neural variability, or lack thereof. *Frontiers in computational neuroscience*, 7(February):7. 2

Mazzucato, L., Fontanini, A., and La Camera, G. (2015). Dynamics of Multistable States during Ongoing and Evoked Cortical Activity. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 35(21):8214–31. 3, 22, 23, 30, 32, 43, 44, 48, 52, 73, 74, 76, 77, 79

Meckenhäuser, G., Krämer, S., Farkhooi, F., Ronacher, B., and Nawrot, M. P. (2014). Neural representation of calling songs and their behavioral relevance in the grasshopper auditory system. *Frontiers in systems neuroscience*, 8(December):183. 64

Morrison, A., Aertsen, A., and Diesmann, M. (2007). Spike-timing-dependent plasticity in balanced random networks. *Neural computation*, 1467:1437–1467. 46

Nawrot, M. P. (2003). *Ongoing activity in cortical networks: noise, variability and context.* PhD thesis. 6

Nawrot, M. P. (2010). Analysis of Parallel Spike Trains. pages 1–22. Springer US, Boston, MA. 7, 20, 73, 76

Nawrot, M. P., Boucsein, C., Rodriguez-Molina, V., Aertsen, A., Grün, S., and Rotter, S. (2007). Serial interval statistics of spontaneous activity in cortical neurons in vivo and in vitro. *Neurocomputing*, 70(10-12):1717–1722. 72

Nawrot, M. P., Boucsein, C., Rodriguez Molina, V., Riehle, A., Aertsen, A., and Rotter, S. (2008). Measurement of variability dynamics in cortical spike trains. *Journal of neuroscience methods*, 169(2):374–90. 7, 9, 11

Ocker, G. K., Litwin-Kumar, A., and Doiron, B. (2015). Self-Organization of Microcircuits in Networks of Spiking Neurons with Plastic Synapses. *PLoS computational biology*, 11(8):e1004458. 79

Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nature neuroscience*, 17(4):594–600. 22, 23, 46

Otsuka, T. and Kawaguchi, Y. (2009). Cortical Inhibitory Cell Types Differentially Form Intralaminar and Interlaminar Subnetworks withExcitatory Neurons. *Journal of Neuroscience*, 29(34):10533–10540. 75

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830. 61

Perin, R., Berger, T. K., and Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences of the United States of America*, 108(13):5419–5424. 74

Pernice, V., Staude, B., Cardanobile, S., and Rotter, S. (2011). How structure determines correlations in neuronal networks. *PLoS computational biology*, 7(5):e1002059. 46

Ponce-Alvarez, A., Kilavik, B. E., and Riehle, A. (2010). Comparison of local measures of spike time irregularity and relating variability to firing rate in motor cortical neurons. *Journal of computational neuroscience*, 29(1-2):351–65. 11, 17, 51

Ponce-Alvarez, A., Nacher, V., Luna, R., Riehle, A., and Romo, R. (2012). Dynamics of cortical neuronal ensembles transit from decision making to storage for later report. *J Neurosci*, 32(35):11956–11969. 76, 77

Reich, D. S., Victor, J. D., and Knight, B. W. (1998). The power ratio and the interval map: spiking models and extracellular recordings. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 18(23):10090–10104. 6

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 (New York, N.Y.)*, 327(5965):587–90. 3, 22, 23, 24, 25

Renart, A. and Machens, C. K. (2014). Variability in neural activity and behavior. *Current opinion in neurobiology*, 25:211–20. 2

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. 75

Rickert, J., Riehle, A., Aertsen, A., Rotter, S., and Nawrot, M. P. (2009). Dynamic encoding of movement direction in motor cortical neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(44):13870–82. 11, 14, 16, 57, 61, 64, 78, 79, 91, 93

Riehle, A. (2005). Preparation for action: one of the key functions of the motor cortex. *Motor Cortex in Voluntary Movements: A Distributed System for Distributed Functions*, 33:213–240. 62

Riehle, A. and Renoult, L. (2006). Time is a rubberband : neuronal activity in monkey motor cortex in relation to time estimation. *Neuroscience*, 23(2004):3098–3108. 20

Riehle, A. and Requin, J. (1993). The predictive value for performance speed of preparatory changes in neuronal activity of the monkey motor und premotor cortex. *Behavioural Brain Research*, 53:35–49. 62

Rosenbaum, D. a. (1980). Human movement initiation: specification of arm, direction, and extent. *Journal of Experimental Psychology*, 109(4):444–474. 62

Scholvinck, M. L., Saleem, A. B., Benucci, A., Harris, K. D., and Carandini, M. (2015). Cortical State Determines Global Variability and Correlations in Visual Cortex. *Journal of Neuroscience*, 35(1):170–178. 11

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 : the official journal of the Society for Neuroscience*, 18(10):3870–96. 2, 12

Shimokawa, T. and Shinomoto, S. (2009). Estimating instantaneous irregularity of neuronal firing. *Neural computation*, 21(7):1931–1951. 72

Shinomoto, S., Kim, H., Shimokawa, T., Matsuno, N., Funahashi, S., Shima, K., Fujita, I., Tamura, H., Doi, T., Kawano, K., Inaba, N., Fukushima, K., Kurkin, S., Kurata, K., Taira, M., Tsutsui, K.-I., Komatsu, H., Ogawa, T., Koida, K., Tanji, J., and Toyama, K. (2009). Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLoS computational biology*, 5(7):e1000433. 11

Shinomoto, S., Shima, K., and Tanji, J. (2003). Differences in spiking patterns among cortical neurons. *Neural computation*, 15(12):2823–42. 11

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. 74

Sporns, O. and Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics*, 2(2):145–162. 74

Stein, R. B. (1967). Some models of neuronal variability. *Biophysical journal*, 7(1):37–68. 1, 6

Supèr, H., van der Togt, C., Spekreijse, H., and Lamme, V. a. F. (2003). Internal state of monkey primary visual cortex (V1) predicts figure-ground perception. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 23(8):3407–3414. 78

Teich, M. C., Heneghan, C., Lowen, S. B., Ozaki, T., and Kaplan, E. (1997). Fractal character of the neural spike train in the visual system of the cat. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 14(3):529–546. 76

van Vreeswijk, C. and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726. 21

van Vreeswijk, C. and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural computation*, 10(6):1321–71. 3, 21, 22, 23, 24, 25, 26, 77

Watts, J. and Thomson, A. M. (2005). Excitatory and inhibitory connections show selectivity in the neocortex. *The Journal of physiology*, 562(Pt 1):89–97. 75

Wiener, M. C. (2003). An adjustment to the time-rescaling method for application to short-trial spike train data. *Neural computation*, 15(11):2565–76. 9

Yoshimura, Y. and Callaway, E. M. (2005). Fine-scale specificity of cortical networks depends on inhibitory cell type and connectivity. *Nature neuroscience*, 8(11):1552–9. 75

Zenke, F., Agnes, E. J., and Gerstner, W. (2015). Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature communications*, 6:6922. 79

Zenke, F., Hennequin, G., and Gerstner, W. (2013). Synaptic plasticity in neural networks needs homeostasis with a fast rate detector. *PLoS computational biology*, 9(11):e1003330. 80

Appendix A

Selection Criteria for Physiological Data

The physiological data used in this work stems from Alexa Riehle's lab at Institut de Neurosciences de la Timone CNRS & Universite Aix-Marseille, France and has been previously published in [Bastian et al., 2003] and further analysed in [Rickert et al., 2009]. Two Macaque monkeys were trained to perform a delayed centre-out reach task with varying information about the correct movement at trial start (for a detailed description of the experimental paradigm see chapter 5). Conditions were recorded in separate blocks of around 150 trials where the trial types were chosen randomly with equal probabilities.

Extracellular recordings were taken with electrode arrays in pre-motor areas (monkey 1: M1 close to the central sulcus, monkey 2: mainly PMd). Online-spike sorting was performed resulting in up to seven simultaneously recorded units. The resulting action potentials as well as behavioural events were stored in 1 ms resulution for offline analysis. Only successful trials were available for the present study (i.e. trials where the monkey pressed the correct button after the RS). 111 and 110 neurons were available for monkeys 1 and 2 respectively.

The calculation of variability statistics sets high standards for data quality as artefacts from recording or spike sorting procedures can have dramatic impacts on the results. Also, reasonable amounts of spikes are required to get meaningful statistical measures. We therefore applied strict quality criteria for the individual neurons.

Some units showed abrupt changes in behaviour over trials in their raster plots. Though it is theoretically possible that a unit changes its behaviour over time, it is more likely that the change is due to a small movement in the recording electrode or detection problems during spike sorting. Raster plots of all trials for

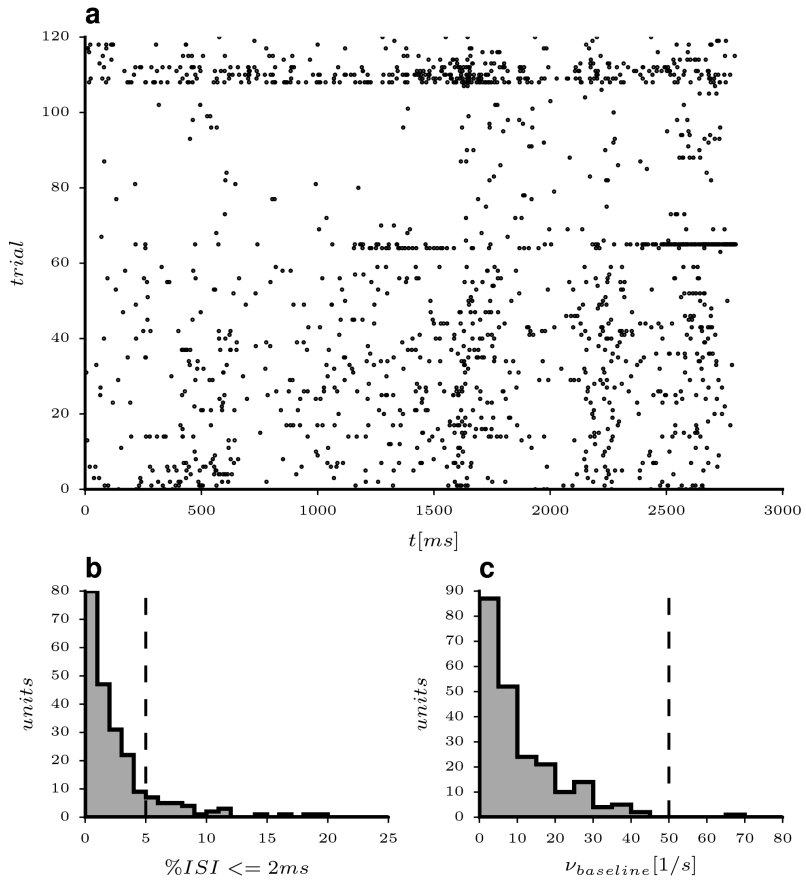

Figure A.1: Illustration of selection criteria applied to the physiological data. **a**)Example session of a unit excluded due to suspiciously abrupt changes in firing behaviour. Trials are in chronological order. **b**) Distribution of precentage of ISIs smaller or equal to two ms for all units. **c**) Disribution of firing rates in the pre-cue interval for all units. Dashed vertical lines indicate thresholds for exclusion.

each unit in chronological order were therefore visually inspected and un-natural looking units were excluded from the analyses. For an example of abruptly changing firing behaviour see figure A.1 a.

The raster plot in figure A.1 a also contains some suspicious looking bursts in which many ISIs of one or two ms occur in sequence. To exclude artefacts of this kind we counted the number of ISIs $<= 2ms$ for each unit. Although most units contained some small intervals (see figure A.1 b) the problem was particularly dramatic in some cases. We therefore decided to mistrust spike sorting when the fraction of small ISI exceed 5% of all intervals recorded from a unit and excluded those neurons from our analyses.

The distribution of spontaneous firing rates for all units (measured in the

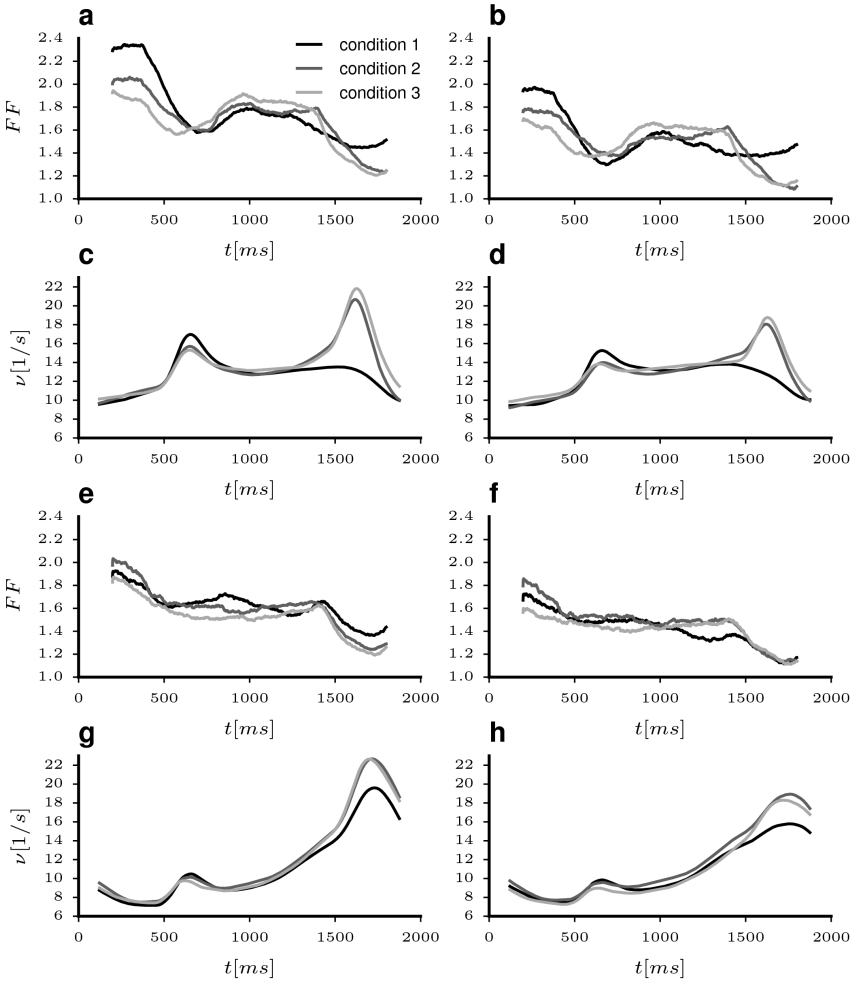

Figure A.2: Effect of neuron selection on FFs and firing rates. Left columns: all units, right column: units that passed all tests. **a**-**d**) monkey 1 **e**-**h**) monkey 2

$500ms$ interval between TS and PS) is shown in figure A.1 c. It can be seen that there are some outliers with very high spontaneous rates above 50 spikes per second. These were also excluded from the analyses. Further, we required units to be present in all three experimental conditions and that there be at least 5 trials per direction in each condition.

This selection procedure left us with 76 units for monkey 1 and 66 for monkey 2, i.e. about a third of the recorded neurons were excluded. The effect of this procedure on the average FFs was quite drastic. Figure A.2 shows the firing rates and FFs for both monkeys before ($l.h.s.$) and after ($r.h.s.$) the exclusion of suspicious units. Panel a corresponds to the figure given in the supplement of [Rickert et al., 2009]. The difference in FF before PS is no longer present in the

cleaned data (compare panels a and b). Also the maximum FFs are reduced to values closer to those reported elsewhere in the literature (e.g. [Churchland et al., 2010]). The effect of the selection on the firing rates is less pronounced.