

Decision making and population decoding with strongly inhibitory neural field models

THOMAS TRAPPENBERG

Faculty of Computer Science, Dalhousie University
6050 University Ave, Halifax, Nova Scotia, Canada, B3H 1W5

tt@cs.dal.ca

Abstract

This chapter discusses dynamic neural field models which are the mathematical quantification of cortical and subcortical maps. Several examples from different brain areas show the universality of these models that captures a principle information processing principle in the brain. The examples connect behavioural findings with physiological measurements in the brain. Such models are therefore important to close the gap.

1 Introduction

Understanding how the brain works is a scientific quest on many levels. This certainly includes areas such as the biophysics of ion channels or the developmental properties of the nervous system. But ultimately we would like to understand how specific brain mechanisms measured with neurophysiological methods give rise to cognitive mechanisms that can be measured on a behavioural level. While there has been some progress in recent years, the gap between neurophysiology and behaviour is often large. The aim of this book is to demonstrate how computational modelling approaches can help in bridging this gap.

In this contribution I discuss an example of a fairly general model of processing in neural tissues based on center-surround lateral interactions. I refer to this model as a dynamic neural field (DNF) or center-surround neural field (CDNF) models. It is called a field model because it describes a continuous (feature) space and incorporates continuous time dynamics. Describing brain functions with continuous models in space and time is not only mathematically useful and elegant, but it stresses the continuous character of cognitive processes. In contrast to many neural network models, it does not depend on the number of nodes and displays some characteristics that can be generalized to many brain functions.

I give several examples of how DNFs can elucidate neural response properties. For example, I review how tuning curves result from lateral interactions in center-surround neural tissue. While classical tuning curves describe only the maximal response of neurons, I will show in addition that experimental results for the temporal evolution of neural responses are consistent with such general models. Furthermore, important for this workshop is not only that single cell data can be explained by such models, but that these models, at the same time, explain behavioural findings. I will therefore give some examples of how the same model can explain physiological single cell recording data as well as psychophysical functions.

The examples of cell recording data in this chapter are taken from a variety of neocortical areas, including primary visual cortex, temporal cortex and parietal cortex, and a subcortical area

related to eye movements, the superior colliculus. This in itself should highlight some universality with respect to the information processing principle discussed in this section. Besides tuning curves I discuss perceptual choice tasks and experimental data from the recent literature on decision making. I also briefly discuss some issues in population decoding. Some of these data and mechanisms have been explained previously with similar or related models. However, the reason for reiterating them here is to highlight the explanation of these data within the same neural field framework and within the same parametrization for all of the various application areas. Indeed, it is good news that various researchers start to build converging consensus within computational principles rather than building highly specialized models of specific physiological or behavioural findings.

There are also some new parts to the story presented here. In particular, I concentrate to some extent on a regime in these networks with relatively strong inhibition, which is not commonly the focus in the application of such models. Furthermore, I mention new data from eye movement studies that demonstrate how surprising nonlinearities in behavioural studies can be explained with the neural field model of the superior colliculus.

2 Dynamic neural field model and cortical feature representation

2.1 The basic computational model

Center-surround neural field models have been introduced and discussed at least since the early 1970s (Wilson & Cowan, 1973; Grossberg, 1973). In the following I discuss a neural field model of the formulation introduced by Amari (Amari, 1977) where the dynamics of an internal state variable $u(x)$ at ‘location’ x is give by

$$\tau \frac{\partial u(x, t)}{\partial t} = -u(x, t) + \int_y w(|x - y|) r(y, t) dy + I^{\text{ext}}(x, t). \quad (1)$$

In this formula, $I^{\text{ext}}(x, t)$ represents an external stimulus as received by the neural field, and $r(x)$ is an external state variable. This rate is related to the internal state variable $u(x)$ through a gain function $r = g(u)$. I use a sigmoidal gain function $g(u) = 1/(1 + \exp(-\beta u))$ for the demonstrations in this chapter. Other gain functions are sometimes biologically more appropriate but will not change the principle findings in the examples presented here. Most of the discussions in this chapter are done with 1-dimensional neural fields, but all of the discussions can easily be generalized to higher dimensions where the 1-dimensional feature space is replaced by a higher dimensional feature map. An example of a 2-dimensional model is discussed in the last section of this chapter when discussing eye movements and a model of the superior colliculus. Also, we always use period boundary conditions to minimize boundary effects. Thus the feature space is formally a ring in one dimensions, where we frequently use a feature value of $(0, 2\pi]$, or more generally a torus in higher dimensions.

Equation (1) defines the dynamic evolution of the neural field, and it is important to note that this formulation is continuous in time and space. To evaluate such models on a computer one needs to *discretize* both space and time. For example, one can define a small but finite time step Δt and spacial resolution Δx and write the differential equations as a difference or iterative equation for each *node* with index i as

$$u_i(t + \Delta t) = (1 - \frac{\Delta t}{\tau}) u_i(t) + \frac{\Delta t}{\tau} (\sum_j w_{ij} r_j(t) \Delta x + I_i^{\text{ext}}(t)). \quad (2)$$

This equation reduces to the widely used equation for recurrent attractor networks in the stationary limit or with time steps $\Delta t = \tau$. It is important to stress that the discretization is only a necessary

step for computational purposes, but that the processes described by this model should approach continuity of time and (feature) space which is appropriate to describe cognitive processes. Thus, we have to make sure that numerical results do not change by altering the time step or including more nodes. Also, the interpretation of each node is not that of a simple neuron, rather, it stands for a collection (population) of neurons that receive similar input in a certain task (Wilson & Cowan, 1972; Gerstner, 2000), or that of a cortical minicolumn (Johansson & Lansner, 2007). However, we will see below some examples where such a population average describes quite well the trial-averaged activity of neurons.

The lateral interaction kernel $w(x, y)$, or its discrete version w_{ij} , is chosen to be a shifted Gaussian

$$w_{ij}^{\text{ex}} = A_w \left(\frac{1}{\sqrt{4\pi\sigma_r}} e^{-((i-j)*\Delta x)^2/4\sigma_r^2} - C \right). \quad (3)$$

Other popular choices are weight profiles in form of a Mexican-hat, but these functions result in similar behaviour when the extent of inhibition and excitation are matched in the periodic feature maps discussed here. Many studies assume this interaction structure from the start, and it is possible that such structures are genetically coded, in particular if the feature space is topographic. However, such weights can also be learned from activity-dependent Hebbian learning on Gaussian patterns with width σ_r (Stringer, Trappenberg, Rolls, & Araujo, 2002; Trappenberg, 2002; Wu & Trappenberg, 2007). Such learning might be, for example, important in the formation of place fields in the hippocampus (Stringer et al., 2002).

Amari (Amari, 1977) discussed such neural field models in detail and showed that there are several solutions to the dynamic equations. In particular, there is a regime of intermediary inhibition values in which an area of activity will persist even in the absence of continuous external input. Such an active area is sometimes called a bubble, a bump, or an activity packet, and this persistent activity has been implicated with physiological working memory (Wang, 1999, 2001). There are many variations of the basic model with different shapes of weights, various gain functions or slightly different formulations of the dynamic equation. However, as long as there is some balance between local excitation and long-distance inhibition the principle findings are very similar. For further discussions of this point, see (Hansel & Sompolsky, 2005).

Neural field models of lateral-inhibition type are in common use on various levels of description and application domains. For example, Wilson and Cowan (Wilson & Cowan, 1973) introduced such a model from course-graining neuronal sheets of spiking neurons with excitatory and inhibitory neuronal pools. In this sense the model can be viewed as physiologically based. On the other hand, Gregor Schöner (Schöner, 2007) has argued for many years that such models are a useful descriptions of behavioural phenomena, in particular since they describe the dynamics of many behavioural findings. In this chapter I want to describe specifically some examples where the model simultaneously describes physiological data of trial-averaged firing rates of single neurons and behavioural data such as reaction times and detection accuracies.

As a final comment on the neurophysiological plausibility of such models, the origins of long-range lateral inhibition may be subject to debate. Anatomically it seems that inhibitory neurons have a more local projection area than lateral excitatory connections (McDonald & Burkhalter, 1993; Kisvrdy & Eysel, 1993; Hirsch & Gilbert, 1991). However, mechanistically there are several possible ways to accomplish extended inhibition. For example, Wilson and Cowan cited experimental evidence (see page 62 in (Wilson & Cowan, 1973)) that excitatory to inhibitory connections extend further than excitatory to excitatory connections. Or, Taylor and Alavi (Taylor & Alavi, 1993) argued that long-range inhibition could be mediated by cortico-thalamic interactions thorough the nucleus reticularis of the thalamus. In general, one could think about the long-range global inhibition as an inhibitory background field, often described as a inhibitory pool (Deco & Lee, 2002). Also, effective long-range inhibition can be seen experimentally (Nelson & Frost,

1978; Grinvald, Lieke, Frostig, & Hildesheim, 1994).

When thinking about the cortical implementation of this model one is foremost bound to think about this model as a representation of a cortical organization of a hypercolumn. Indeed, the discussions in the next section will illustrate this in some detail. Processing in hypercolumns are often thought of as a winner-take-all process (Hubel & Wiesel, 1977), and the neural field models discussed here implement such a mechanism through lateral interactions. In this chapter we will also discuss competition between object representations beyond the typical representation in hypercolumns of early sensory areas. There are two ways of thinking about this. First, functional structures of hypercolumns in higher cortical areas will of course represent much more abstract concepts and global object features compared to hypercolumns in early sensory cortex. Secondly, competitive interactions can also be interpreted, in a more abstract way, as competition between hypercolumns or even whole objects as represented by a series of features represented by hypercolumns (Johansson & Lansner, 2007).

2.2 Competitive feature representations in cortical tissue

Hubel and Wiesel (Hubel & Wiesel, 1962) found that some neurons in the primary visual cortex of anesthetized cats respond strongly to moving bars in their receptive field with a specific orientation and that the same neurons respond less when the orientation of the moving bar deviates from this preferred direction. An example of such a tuning curve from experiments by Henry et al. (Henry, Dreher, & Bishop, 1974) is shown by solid circles and an interpolated Gaussian curve in Figure 2.2. In this example the preferred direction of this cell is plotted as zero degrees, and the rate diminishes with a bell-shaped curve with increasing distance from the preferred direction.

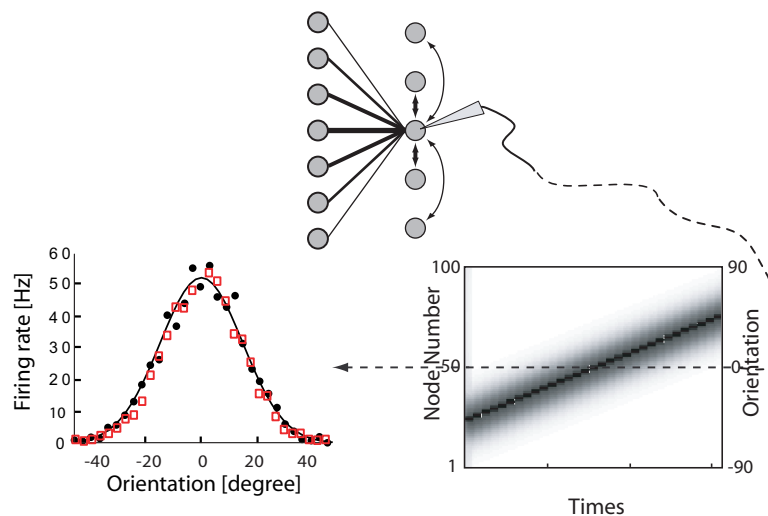


Figure 1: Experimental data (solid circles) from Henry et al. (1974) plot the firing rate of a neuron in cat primary visual cortex as a function of the orientation of moving line segments. Open squares correspond to the corresponding measurements in a neural field model. The evolution of the activity in the neural field is shown in the plot on the right.

Such tuning curves can be modeled in different ways. For example, we could assume that the neuron receives input from cells that only fire to specific orientations, and that the input weight of these connections is a bell-shaped function. However, since the orientation selectivity only arises in V1, we follow here the common idea that local arrangements of V1 cells provide a focal orientation selectivity as suggested by Hubel and Wiesel (see also (Nagano & Kurata, 1980)) but that the broadness of the tuning curves are mediated by lateral interactions (Bullier, 2006;

Ben-Yishai, Bar-Or, & Sompolinsky, 1995). This general idea is captured by the neural field model as illustrated in Figure 2.2. Each node in the model receives input from a certain area of the visual field (its receptive field) and is most responsive to a specific orientation (tuning curve). In the shown simulated experiment I probed the response of the neural field by activating externally a very small region, essentially covering only one node in the simulations. The neurons for these specific stimuli are very active, and, most importantly, the activity is also ‘bleeding out’ to neighboring nodes. The maximal response for different orientations of input for a neuron with preferred orientation of zero degrees is plotted with open squares on top of the experimental tuning curve, demonstrating how the neural field model can reproduce the experimental findings.

In our interpretation of tuning curves we see that there is a wide response of neurons even when presenting a ‘pure’ image. This is important as such states are likely to represent probabilities of specific events (Zemel, Dayan, & Pouget, 1998). Cortical neurons are also very noisy which was combated with many trials in the experiments and by neglecting noise in the simulations. Decoding the information from a population becomes crucial for reliable information processing in the brain. I return to this point later in this chapter. The tuning curves discussed here only show the maximal response during each stimulus presentation and I will now turn to some temporal aspects that can illuminate other experimental findings.

3 Neural field models of perceptual choice and decision making

3.1 Temporal aspects of neural field responses

In the following experiments we have a closer look at the temporal developments in the neural field. While doing this, we will generalize the application of this model to other feature representations in cortex. Higher cortical areas represent increasingly complex features beyond orientation selectivity of V1 neurons, but we can still think of these representations in terms of neural fields. The only difference is that the represented feature value (orientation in V1) should be replaced by another feature value to which the other cortical area responds.

In this section I discuss cell recording data from Monkey’s inferior temporal (IT) area associated with object recognition. Figure 2 shows data from recordings in this area by Chelazzi et al. (Chelazzi, Miller, Duncan, & Desimone, 1993). In their experiment they showed different objects to monkeys and selected objects to which the recorded IT cell responded strongly (*good* objects) and weakly (*bad* objects). In the illustration of Experiment 1, the good object is represented with a square. The average firing rate of an IT cell to this good stimulus is shown in Figure 2C as solid line. The period when the cue stimulus is presented is indicated by the left grey bar in this figure. The neuron responds strongly to this stimulus, but the trial-averaged rate of this neuron decays basically to the background level after the cue is removed.

The ‘bad’ object is illustrated with a triangle in Figure 2, and the trial-averaged response of the same neuron recorded during Experiment 2 is shown as dashed line in Figure 2C. Instead of increasing the rate during the cue presentation, this neuron seems to respond with a firing rate below the usual background rate.

Several aspects of this response are captured by the neural field model shown in Figures 2C and D. The solid line represents the activity of a node within the response bubble of the neural field such as a node recorded at position O_1 in Figure 2D. The activity declines after the external stimulus is removed. This was achieved by using a relatively strong inhibition constant of $c = 0.06$ in the simulations. A slightly smaller inhibition constant of $c = 0.05$ would result in a sustained activity bubble as used in working memory models. The activity of the dashed line corresponds to the activity of a node outside the activity bubble. There is a diminished activity through lateral inhibition during the stimulus presentations similar to the experimental data.

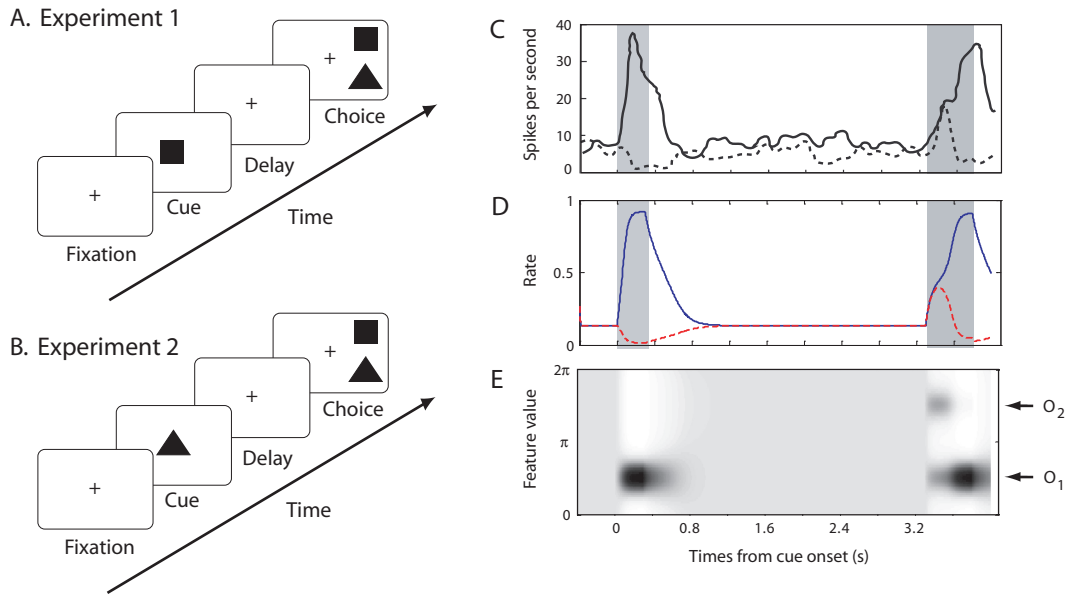


Figure 2: Experimental paradigm and data from Chelazzi et al. (1993) and their interpretation in the neural field model. Monkeys are given a single visual cue and, following a delay, the monkey is shown two stimuli and must moves its eyes to the remembered stimulus. The average response of an IT neuron is shown in panel (C) where the solid line corresponds to a ‘good’ object and the dashed line to a ‘bad’ object. The corresponding response in the neural field model are shown in panels (D) and (E).

The details of the program and the chosen parameters are detailed in the Appendix. It is worth noting that all the experiments in the different application areas are basically run with the same model and comparable parameters.

3.2 Competitive aspects of neural field responses

The second half of the experiments by Chelazzi et al. is even more interesting. After a delay the monkey is shown the two objects, the ‘good’ and the ‘bad’ object, at the same time (right grey area in the panels on the right). The monkey is instructed to saccade to the cued objects. Interestingly, the initial response of the IT neuron to the display is an increase in rate regardless of which object was cued. However, after around 200ms the response in trials with good and bad object cues diverge. The activity increase further in trials where the cue is a good object. In contrast, further activity is suppressed in trials where the bad object was the cue.

The interpretation of these data with the neural field model is straight forward. Presenting two objects to the neural field means that two different regions are stimulated simultaneously. We take the (top-down) task instruction into account by making one of the inputs slightly stronger, only 1% in the shown simulations. Thus the requirement to memorize the cue is not explicitly modeled in these effects, rather, only the effect of this memory is taken into account. Indeed, it is possible and common to use a similar model to store this short term memory which feeds into this IT map. This module could, for example, simulate working memory in prefrontal cortex or working memory loops between prefrontal and other cortical areas. The interesting point here is that we take only a small bias in the input to the neural field into account. At first, this small difference in input strength does not make a difference in the neural field response. However, after enough activity is generated in the neural field, the two active centers begin to interact though lateral inhibition, and the slight advantage of one center leads to a winning active area.

I am not the first to offer this interpretation of the these data. Usher and McClelland (Usher & McClelland, 2001) have recently argued that these data can be understood and described with competitive integrators that correspond to the neural field models discussed here. Indeed, their model represents a discrete implementation of the neural field model with strong inhibition. Usher and McClelland also discuss nicely how the neural field model unifies two historical models of decision making, accumulator models and diffusion models. With accumulator models, evidence for stimuli accumulates over time and a decision is reached when evidence for a particular stimulus exceeds a decision threshold. With diffusion models, it is not the absolute evidence that is accumulated, but the relative difference between the evidence for each stimulus. The neural field model is thus a dynamic implementation of an intermediate model where there is some coupling between different areas of the neural field.

3.3 Decision making, psychophysics and uncertainty

My discussion of perceptual choice has centered on electrophysiological data, but with crude behavioural resolution. To provide a cognitive perspective, behavioural measurements with higher resolution are required. Those are provided by motion discrimination experiments. In a typical motion discrimination task, a display of dots is presented on a computer screen and some fraction of the dots are repeatedly displaced. The dots appear to move in the direction of displacement, typically to the left or to the right. Monkeys are trained to move their eyes in the direction of perceived movement, thus indicating their choice. Experimentalists manipulate the *coherence* of movement by controlling the percentage of displaced dots, allowing them to finely control the certainty of evidence. Recording response time and accuracy provides psychophysical measurements in response to the controlled parameter (coherence in this case).

Psychophysical measurements from Roitman and Shadlen (Roitman & Shadlen, 2002) are shown as dots in Figures 3A and B. The figures show the *psychometric* and *chronometric* functions respectively, depicting accuracy and reaction time as a function of the certainty of evidence. At low coherence levels, the monkey makes a lot of errors and has long reaction times. With increasing coherence, the monkey's accuracy increases and its reaction times are lower. Results from corresponding neural recordings in monkey lateral intraparietal area (LIP) are shown in Figure 3C. The data are similar to output from simulations, shown in Figure D. With increased coherence, the activity of neurons with a preferred response to the direction of motion increases (solid lines). The response of neurons with other preferred directions increases at first, before rapidly decreasing (dashed lines). Three different values of the coherence parameter are depicted in the figure. Stronger input (greater certainty) more quickly 'defeats' the weaker stimulus (dashed lines) leading to a quicker decision, as shown by the solid lines. Reaction times are equated with the time required for activity to reach a threshold level, and decrease with increased coherence (Figure 3B). If the evidence has not reached threshold by 800ms, the monkey responds according to the accumulated evidence, shown in Figure 3A. In the figure, lines show model output, where accuracy is poor.

The accuracy curves from the model are proportional to the evidence accumulated in LIP. In the simplest case, the evidence from the neural population can be read out by calculating the normalized maximum response for each perceptual choice, where the normalization provides a probabilistic interpretation of the activity in LIP. This calculation is shown by the solid line in Figure 3A. The curve describes the experimental data, providing a mechanistic description of LIP activity in Roitman and Shadlen's task.

Despite its ability to explain the data, the solid curve in Figure 3A does not run through the dots. The monkey is perfectly accurate at even moderate coherence levels, but the solid line approaches this limit smoothly. Thus, a decision based on LIP activity cannot be as simple as suggested by the probabilistic interpretation. In the figure, the dotted line shows a linear transfor-

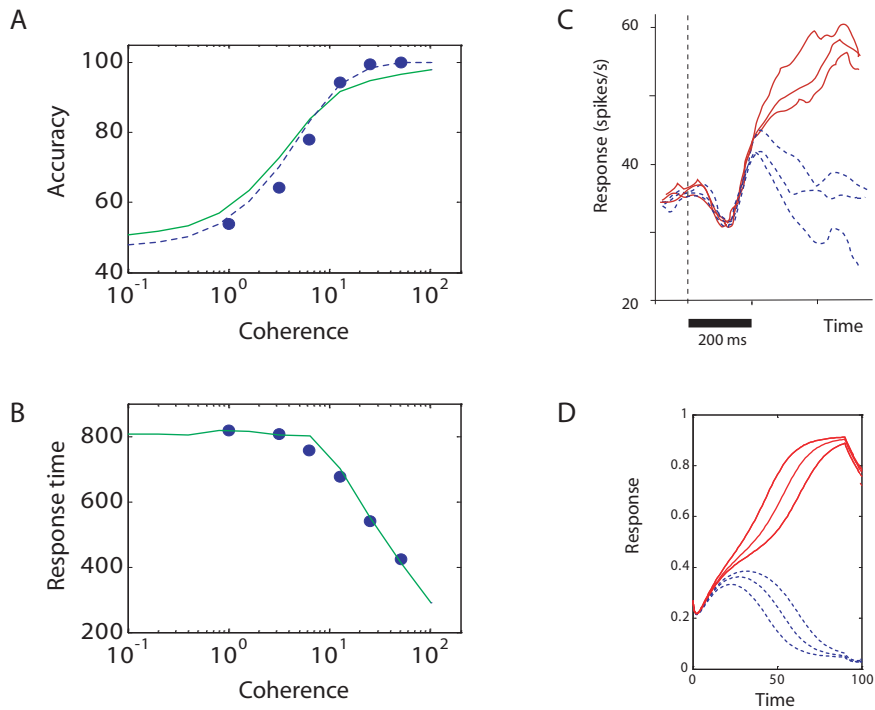


Figure 3: Data from Roitman and Shadlen (2002) is explained by the neural field model. (A) The *psychometric* function plotting accuracy as a function of certainty of evidence. Experimental data are shown as dots. The solid lines shows normalized output from the model. A linear transformation weighted in favor of high-rate activity is shown by the dashed line. (B) The *chronometric* function plotting response time as a function of certainty of evidence. Again, dots and lines correspond to data and output from the model respectively. (C) Neural response over time for different levels of certainty of evidence. Solid and dashed lines show the spike rate of neurons responsive to the chosen and unchosen stimuli respectively. (D) Output from the model in response to simulation of the task in C.

mation of the output of the model, where high-rate activity is given a greater weight than low-rate activity. This curve fits the data more accurately, but does not tell us anything new. In this regard, it is sometimes easy to fit curves to experimental data with enough parameters, but a simpler model may be more scientifically useful, serving to identify information processing principles in the brain.

4 Population decoding with strong inhibition

As pointed out at the beginning of this chapter, the models discussed here describe instantaneous rates of populations of neurons. Single neurons in the cortex have typically low firing rates and their firing is also very noisy. In the previous examples I averaged over such neuronal populations and I also compared the model behaviour to a trial-based average of corresponding experiments. However, the question arises how a neural sheet receiving very noisy input can achieve reliable feature estimation from noisy population activity.

A possible answer to noise population decoding comes from the recognition that center-surround neural field models are ideal to solve this noisy decoding task. The theory of population decoding with neural fields was developed recently in some detail (Pouget, Zhang, Deneve, & Latham, 1998; Deneve, Pouget, & Latham, 1999; Deneve, Latham, & Pouget, 1999; Pouget, Dayan, & Zemel, 2000; Deneve, Latham, & Pouget, 2001; Wu, Amari, & Nakahara, 2002). It is common in these applications of neural field models to use a normalizing gain function that represent divisive inhibition so that the resulting activity packet has a Gaussian shape. It can then be shown (Pouget et al., 1998; Wu et al., 2002) that the activity pattern is equivalent to a maximum likelihood (ML) estimate of Gaussian tuning curves.

Population decoding in the above mentioned papers is usually done by applying input as initial states of the recurrent network. The decoded signal is then usually considered to be the asymptotic state of sustained activity packets (bubbles). Here I summarize my recent study (Trappenberg, 2007) where I investigated the population decoding with DNFs with several additions and simplifications that add to their biological realism and investigate which regime of the DNF models is best suited for population decoding. I thereby investigate these networks with sustained input over a period of time, either with fixed noisy input or with temporal changing input. All the simulations in this section are done without divisive inhibition that is commonly used in the related literature, showing that these form of global operation is not necessary for the population decoding. The major question I ask in this section is if there is an optimal value of global inhibition which controls the amount of competition in the network.

4.1 Population decoding with recurrent networks

I consider again the generic network model described above. As noted previously, this model exhibits several regimes characterized by different possible asymptotic states (Amari, 1977). If the inhibition C is low compared to excitation in the network, then the excitation will spread through the network resulting in runaway activity. In contrast, if inhibition is dominating, then any activity in the field will decay without external reinforcement. In an intermediate regime it is possible to have activity packets where localized activity is stable. I call this mode of the model the memory regime. The later two regimes are depicted in Figure 4 in the context of population decoding. In these experiments I supplied a static noisy input to the field over 20 timesteps. This input was chosen as a Gaussian around the middle node ($x_0 = 50$) with additive white noise, η , of strength $n_\eta = 0.5$,

$$I_{\text{ext}} = I_0 + A(e^{-((i-j)*\Delta x)^2/2\sigma^2} + n_\eta\eta), \quad (4)$$

where I_0 is a background field and η is a normal distributed random number. When an inhibition constant of $c = 0.05$ is used, the field developed into a clean bubble around the middle node after the input was removed at time $t = 20$, demonstrating perfect decoding. The sustained localized activity in the neural field without external input demonstrates the above mentioned memory regime.

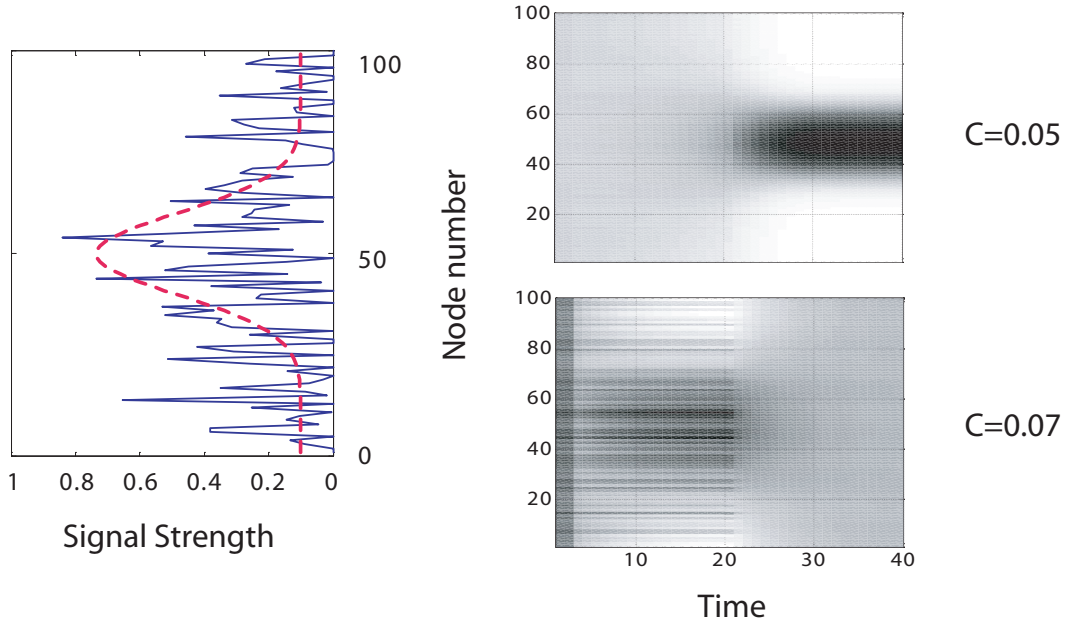


Figure 4: Noisy population decoding with weak and strong inhibition in neural fields. The noisy input (solid line in the left graph), which is derived from the dashed line in the left graph, is applied until $t = 20$ to the neural fields. The time evolution of the neural field is shown in the right graphs, with moderate inhibition constant, $C = 0.05$, in the upper graph and larger inhibition constant, $C = 0.07$, in the lower graph. The activity of the neural field is thereby depicted on a grey scale with different ranges in the two figures. Therefore, the noisy input until $t = 20$ in the upper graph is not well visible since the activity over time is dominated by the sustained bubble at later times.

Traces of the noisy input are not apparent in the $c = 0.05$ case, and it seems one has to wait until the bubble forms to perform accurate decoding. This is different when running the same simulation with larger inhibition. Figure 4 shows the case for $c = 0.07$. Traces of the noisy input are now also visible during the time the external input is supplied, which is partly enhanced by the fact that a smaller range of values is depicted by the gray scale in the figure. The inhibition is now too large to sustain an activity packet after input is removed. However, the increased competition facilitates a cleaning of the signal even during the time when the signal is applied so that some form of population decoding is supported. While this might be less accurate than in the previous case, an advantage would certainly be that the decoding can be achieved much earlier. This assertion is investigated in the next section.

4.2 Quality of decoding with varying inhibition over time

To assess the quality of decoding with time and different inhibition constants I ran decoding experiments over 100 trials in each condition. While I have used signals with static noise in Figure 4, I report now on the results when changing the noise after each unit of time, simulating ongoing fluctuations in the input signal over time. This represents a more plausible implementation of decoding conditions in the brain, although I found similar results in the static noise case.

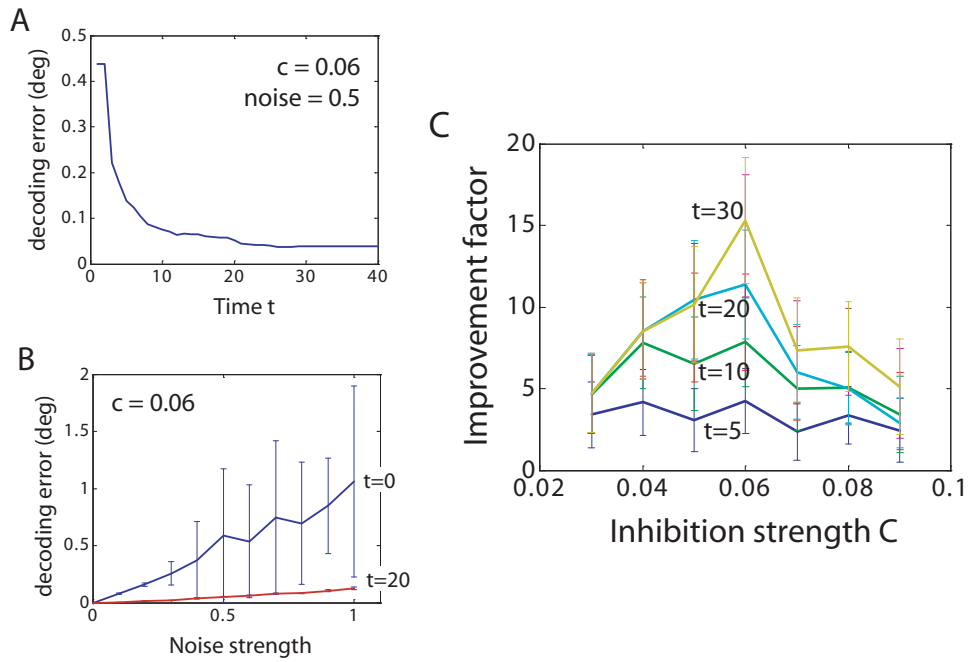


Figure 5: Temporal aspects and decoding quality with different levels of inhibition. (A) Time course of the decoding error (in radiant degrees) with an inhibition constant of $C = 0.06$. The decoding error is decreasing even while the external input is supplied until $t = 20$. The noise level in the shown example is $n_\eta = 0.5$. (B) The decoding error as a function of noise strength, n_η , when decoding is performed at $t = 0$, which corresponds to the original signal before modified by the recurrent network, and at $t = 20$ when the recurrent network was able to process the signal. (c) The improvement factor is calculated as the average decoding error at a specific time $t > 0$ compared to the average decoding error of the original signal, $t = 0$.

To assess the decoding quality I use a center of mass scheme to determine the prediction of the feature value encoded in the neural field. This was done for the original input signal and at each time step during the dynamic decoding in the recurrent network. An example trace of the decoding error is shown in Figure 5A. For the shown strength of inhibition, the decoding error continuously decreases even while external input is supplied. The decoding error only improves slightly after the external input is removed at $t = 20$.

As can be expected, the decoding error increases with increasing noise level as shown in Figure 5B, but the decoding error at $t = 20$ is always much smaller than the center of mass decoding of the original signal. Finally, a major question motivating this study was to determine which network regime, in terms of inhibition strength, would be most suitable for decoding in such networks. The results of these studies are summarized in Figure 5C which shows the decoding improvement at different time steps for different inhibition parameters. Early on there is little dependence of the results on the strength of inhibition, but later there is some advantage for inhibition values around 0.06. Interestingly, this is close to the transition region between the domain of decaying input and the domain of sustained activity.

5 The Interaction of bottom up and top down manipulations on saccade reaction times in monkeys

The last application of the neural field model that I want to mention here is recent work with Robert Marino and Doug Munoz (Marino, Trappenberg, & Munoz, 2006). Robert studied eye movements with monkeys in a gap paradigm where he manipulated two principle factors, the number of possible locations where a target could appear in each trial, and the intensity of a target. The resulting median saccade reaction times (SRT) is shown in Figure 5D. Reaction times are fairly long with faint targets, but SRTs become faster with increasing intensity. However, the surprising effect Rob discovered is that the rate of decrease slowed down and reaction times became even slightly longer for very strong intensities. Another surprising non-monotonicity also appeared when analyzing the reaction times to the number of possible target locations shown in Figure 5F. While reaction times first increase with increasing number of possible target locations, saccade onsets are again faster for 8 possible target locations.

These results are stunning in itself, but are also not that obvious to reconcile with the model of the intermediate layer of the superior colliculus (SC for short in the following) that we devised some time ago (Trappenberg, Dorris, Klein, & Munoz, 2001). The SC is thought to be a topographically organized motor map that drives fast eye-movements, so called saccades. We found that a large variety of effects of saccadic eye movements can be explained when considering that this motor map integrates a variety of endogenous and exogenous signals with a competitive mechanisms described by the DNF models considered in this chapter. Indeed, we even found direct evidence of the center-surround interaction structure (Trappenberg et al., 2001; Dorris, Olivier, & Munoz, 2007).

When applying this model to the new experimental conditions with increased levels of external input to the neural field simulating increased target intensities, then the buildup of SC activity become faster with increasing intensity and there is no reason that reaction times become slower again. In the case of increasing number of possible target locations we expect an increasing number of input representing the expectation of a target at different locations which increases the inhibition within the neural field. This in turn should increase in reaction times with an increasing number of target locations. However, the decrease in reaction time with 8 possible target locations is stunning.

The puzzle can be resolved by considering area effects. If intensities will not only increase the strength of input but will mainly alter the area of the stimulated region within the SC as illustrated

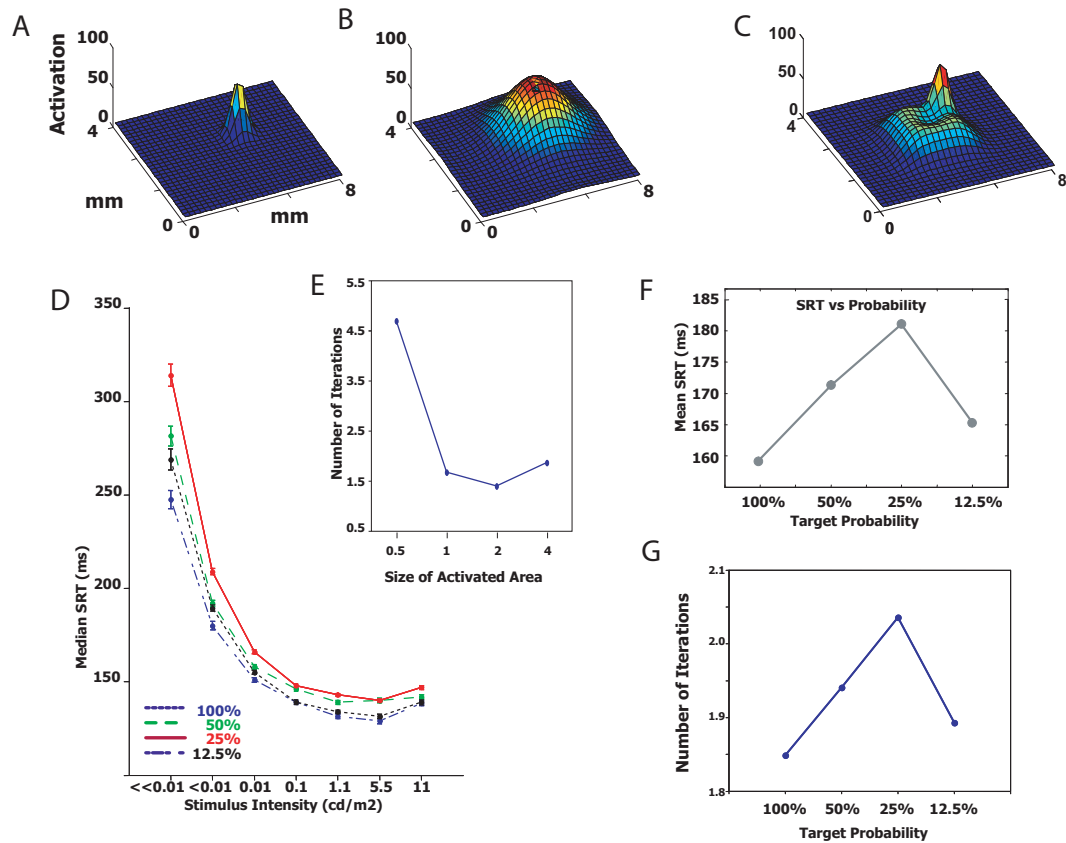


Figure 6: Data of Behavioural experiments and 2d DNF simulations of saccadic eye movements. The upper row shows the 2d DNF neural sheet activity with a (A) narrow and (B) wider input. (C) Neural sheet activity with 4 endogenous and 1 exogenous inputs. (D) Behavioural monkey data and (E) DNF simulation data with different target intensities. Larger target intensity was thereby simulated with wider input in the DNF simulations. (F) Monkey data and (G) DNF simulation data of experiments with a varying number of possible target locations.

in Figure 5A and B, then we get a reaction time curve shown in Figure 5E that resembles the nonlinearity of the experimental results. The reason for the slowdown in reaction times with the largest intensities comes thereby from the fact that the activated area extends to the inhibitory surround in the neural field. As long as the input is small, increasing input will add to the buildup of SC activity and thus will increase SRTs. However, when the activated area becomes comparable with the extend of the center-surround scale, then SRTs will become longer again.

A corresponding explanation can be given to the effect of altering the number of possible target locations. At first there is an increase in areas that are stimulated by endogenous input (top-down input that represents expectation) as shown in Figure 5C, but a further input will not alter much inhibitory input from remote areas but will help the winning location to build up faster. Thus, the model predicts an increase in the area of activation with increased intensity. Robert is currently running experiments to investigate this prediction.

6 Conclusion

Measuring behavioural responses in specific circumstances is a major tool to study cognition. Results of such studies are valuable in itself. For example, knowing the specific parameters of a speed-accuracy tradeoff can help to design appropriate systems for human-computer interactions. On the other side of neuroscience, neurophysiological studies are a major tool in understanding brain mechanisms. In particular, single cell recordings in behaving monkeys and functional brain imaging in behaving humans have started to map brain activities in specific tasks. Such studies are in itself valuable. For example, such studies can illuminate effects of neurodegenerative diseases. But to understand how the brain thinks we have ultimately to close the gap between neurophysiology and behaviour, and computational modelling approaches are a central tool in this domain.

In this chapter I have given some examples where models are able to explain the principal trends in cell recordings and psychophysical curves. Making contact with both levels is important. This allows us to embed behavioural findings into physiological mechanisms on the one hand, and on the other hand this allows us to specify the consequences of neuronal activity on behaviour. It also provides us with a functional explanation of behavioural findings. For example, cell recordings in the perceptual choice task have shown that there is first a common buildup of activity before they split, and on a behavioural level that there is an accuracy-speed tradeoff in responses to such perceptual choices. The model quantifies the hypothesis that the perceptual choice is facilitated by competitive mechanisms in the neural representation of the stimuli, and that these mechanisms will directly influence behaviour. Using computational specific implementation of such models will even allow us to specify these relations quantitatively.

The model studied here is fairly simple. It consists of a neural sheet with center-surround lateral interactions. The continuous formulation of this model is useful as this allows mathematical elegant treatments and also to make it somewhat universal. While the model is simple, I find it interesting to see that it can be applied to so many different areas in neuroscience and cognitive science. I believe that this demonstrates that this model captures principal brain mechanisms which are important for brain-style information processing. The fact that this model can also explain so many effects measured in behavioural studies and neurophysiological studies shows that the experimentally measured effects are understood. Thus, the simplicity of the model is not a drawback; indeed it should be a major attraction of this model. This fact also highlights that, while the experimental effects look on a first glance exciting, it might be worthwhile to investigate new questions where the behaviour can not be explained with such simple models and might therefore highlight some novel information processing mechanisms in the brain. Models should indeed guide us more in discovering such novel mechanisms by making concrete predictions that can be

verified experimentally.

APPENDIX: The basic program

The following is a complete listing of the program to simulate the experiment by Chelazzi et al. (Chelazzi et al., 1993). The other experiments use basically the same program.

The following listing is the main program with the specific experimental settings. After defining some constants, Hebbian learning is used to set up the center-surround connectivity matrix that drives the experiments. Training on Gaussian training patterns leads to a Gaussian connectivity matrix.

After training we define some input patterns and then call a numerical differential equation solver to integrate with neural field over specific times with the various inputs and plot the results at the end of the program. The four short functions called in the main program are listed further below.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% 1-d Continuous Attractor Neural Network with Hebbian learning
% two gaussian signal: decision network
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clear; close all;
nn = 100; dx=2*pi/nn; % number of nodes and resolution in deg

%weight matrices
sig = 2*pi/20;
w_sym=hebb(nn, sig, dx);
w_inh=0.07;%use 0.04, 7, 6, 3; 3*(sqrt(2*pi)*sig)^2/nn;
w=500*(w_sym-w_inh);

%inputs
perc=0.01; Is=11;
Ia=(1+0.5*perc)*Is;
Ib=(1-0.5*perc)*Is;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Experiment %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

param=0;
%%% no external input
u0 = zeros(nn,1)-10;
I_ext=zeros(nn,1);
tspan=[0, 40];
[t,u]=ode45('rnn_ode_u', tspan, u0, [], nn, dx, w, I_ext);
r=f1(u);

%%% external cue
u0 = u(size(t,1), :);
```

```

I_ext=zeros(nn,1);
loc1=pi/2;%+pi/16;
loc2=3*pi/2;%-pi/16;
I_ext=I_ext+in_signal_pbc(loc1,Is,sqrt(2)*sig,nn,dx);
tspan=[40 70];
[t2,u]=ode45('rnn_ode_u',tspan,u0,[],nn,dx,w,I_ext);
r=[r;f1(u)];
t=[t;t2];

%% no external input
u0 = u(size(t2,1),:);
I_ext=zeros(nn,1);
param=0;
tspan=[70,370];
[t2,u]=ode45('rnn_ode_u',tspan,u0,[],nn,dx,w,I_ext);
r=[r;f1(u)];
t=[t;t2];

%% both external inputs
u0 = u(size(t2,1),:);
I_ext=zeros(nn,1);
loc1=pi/2;%+pi/16;
loc2=3*pi/2;%-pi/16;
I_ext=I_ext+in_signal_pbc(loc1,Ia,sqrt(2)*sig,nn,dx);
I_ext=I_ext+in_signal_pbc(loc2,Ib,sqrt(2)*sig,nn,dx);

tspan=[370 420];
[t2,u]=ode45('rnn_ode_u',tspan,u0,[],nn,dx,w,I_ext);
r=[r;f1(u)];
t=[t;t2];

%% no external input
u0 = u(size(t2,1),:);
I_ext=zeros(nn,1);
param=0;
tspan=[420,440];
[t2,u]=ode45('rnn_ode_u',tspan,u0,[],nn,dx,w,I_ext);
r=[r;f1(u)];
t=[t;t2];

%% plot results

surf(t',1:nn,r','linestyle','none'); view(0,90);
figure
plot(t,r(:,round(loc1/dx))); hold on;
plot(t,r(:,round(loc2/dx)),'r')

```

The numerical ODE solver calls the file `rnn_ode_u.m` that contains the differential equation:

```
function udot=rnn(t,u,flag,nn,dx,w,I_ext)
```



```

% odefile for recurrent network
tau_inv = 1./2;      % inverse of membrane time constant
r=f1(u);
sum=w*r*dx;
udot=tau_inv*(-u+sum+I_ext);
return

```

The function `f1.m` simply contains the sigmoidal gain function which can easily be replaced by other gain functions:

```

function f1=rnn(u)
% gain function: logistic
beta =.1;
alpha=.0;%-100*(u>0);
f1=1./(1+exp(-beta.*(u-alpha)));
return

```

The function `in_signal_pbc.m` produces a gaussian signal with periodic boundary conditions:

```

function y = in_signal_pbc(loc, ampl, sig, nn, dx)
y=zeros(nn,1);
for i=1:nn;
    di=min(abs(i*dx-loc),2*pi-abs(i*dx-loc));
    y(i)=ampl/(sqrt(2*pi)*sig)*exp(-di^2/(2*sig^2));
end
return

```

Finally, we also included Hebbian learning to choose the weight matrix:

```

function w = hebb(nn, sig, dx)
% self organization of symmetric cann interactions
lrate=1;      % learning rate
w=zeros(nn);
%%%%%% learning session
for epoch=1:1
    for loc=1:nn;
        r=in_signal_pbc(loc*dx,1,sig,nn,dx);
        dw=lrate.*r*r';
        w=w+dw;
    end
end
w=w/nn;
return

```

References

- Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27, 77–87.
- Ben-Yishai, R., Bar-Or, R. L., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 92(9), 3844–3848.
- Bullier, J. (2006). 23 problems in system neuroscience. In J. L. van Hemmen & T. Sejnowski (Eds.), (pp. 103–132). Oxford University Press.
- Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, 363, 345–347.
- Deco, G., & Lee, T. (2002). A unified model of spatial and object attention based on inter-cortical biased competition. *Neurocomputing*(44-46), 775–781.
- Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: a neural implementation of the ideal observer. *Nature Neuroscience*, 2, 740–745.
- Deneve, S., Latham, P. E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, 4, 826–831.
- Deneve, S., Pouget, A., & Latham, P. (1999). Divisive normalization, line attractor networks and ideal observers. In D. C. Michael Kearns Sara Solla (Ed.), *Advances in neural information processing systems* (Vol. 11). The MIT Press.
- Dorris, M. C., Olivier, E., & Munoz, D. P. (2007). Competitive integration of visual and preparatory signals in the superior colliculus during saccadic programming. *Journal Of Neuroscience*.
- Gerstner, W. (2000). Population dynamics of spiking neurons: Fast transients, asynchronous states, and locking. *Neural Computation*, 12(1), 43-89.
- Grinvald, A., Lieke, E., Frostig, R., & Hildesheim, R. (1994). Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *Journal of Neuroscience*, 14, 2545–2568.
- Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52, 217–257.
- Hansel, D., & Sompolinsky, H. (2005). Continuous attractor neural networks. In L. N. de Castro & F. J. V. Zuben (Eds.), *Recent developments in biologically inspired computing* (pp. 398–425). IDEA Group Publishing.
- Henry, G. H., Dreher, B., & Bishop, P. O. (1974). Orientation specificity of cells in cat striate cortex. *Journal of Neurophysiology*, 37, 1394–1409.
- Hirsch, J., & Gilbert, C. (1991). Synaptic physiology of horizontal connections in the cat's visual cortex. *Journal of Neuroscience*, 11, 1800–1809.
- Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, London*, 160, 106–154.
- Hubel, D., & Wiesel, T. (1977). Ferrier lecture: Functional architecture of macaque monkey visual cortex. *Proc. Roy. Soc. Lond. B. Biol. Sci.*, 198, 1–59.
- Johansson, C., & Lansner, A. (2007). Towards cortex sized artificial neural systems. *Neural Networks*, 20, 48–61.
- Kisvrdy, Z. F., & Eysel, U. T. (1993). Functional and structural topography of horizontal inhibitory connections in cat visual cortex. *European Journal of Neuroscience*, 5(12), 1558–1572.
- Marino, R. A., Trappenberg, T. P., & Munoz, D. P. (2006). Modeling the bottom-up and top-down effects on saccadic reaction times in monkeys using artificial neural network models

- of the superior colliculus. In *28th international symposium montral: Computational neuroscience*.
- McDonald, C. T., & Burkhalter, A. (1993). Organization of long-range inhibitory connections within rat visual cortex. *Journal of Neuroscience*, *13*(2), 766–761.
- Nagano, T., & Kurata, K. (1980). A model of the complex cell based on recent neurophysiological findings. *Biological Cybernetics*, *33*(2), 103–105.
- Nelson, J., & Frost, B. (1978). Orientation-selective inhibition from beyond the classical visual receptive field. *Brain Research*(139), 357–365.
- Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience*, *1*, 125–132.
- Pouget, A., Zhang, K., Deneve, S., & Latham, P. (1998). Statistically efficient estimation using population coding. *Neural Computation*, *10*(2), 373–401.
- Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *The Journal of Neuroscience*, *22*(21), 9475–9489.
- Schöner, G. (2007). Dynamical systems approaches to cognition. In R. Sun (Ed.), *Cambridge handbook of computational cognitive modeling*. Cambridge University Press.
- Stringer, S., Trappenberg, T., Rolls, E., & Araujo, I. (2002). Self-organising continuous attractor networks and path integration: One-dimensional models of head direction cells. *Network: Computation in Neural Systems*, *13*, 217–242.
- Taylor, J., & Alavi, F. (1993). A global competitive network for attention. *Neural Network World*, *5*, 477–502.
- Trappenberg, T. (2002). *Fundamentals of computational neuroscience*. Oxford University Press.
- Trappenberg, T. (2007). Dynamics of population decoding with strong inhibition. In *International conference on cognitive neurodynamics iccn'07*.
- Trappenberg, T., Dorris, M., Klein, R., & Munoz, D. (2001). A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience*, *13*, 256–271.
- Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, *108*, 550–592.
- Wang, X.-J. (1999). Synaptic basis of cortical persistent activity: the importance of nmda receptors to working memory. *J. Neurosci.*, *19*, 9587–9603.
- Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neuroscience*, *24*, 455–46.
- Wilson, H., & Cowan, J. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.*, *12*, 1–24.
- Wilson, H., & Cowan, J. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, *13*, 55–80.
- Wu, S., Amari, S. ichi, & Nakahara, H. (2002). Population coding and decoding in a neural field: a computational study. *Neural Computation*, *14*, 999–1026.
- Wu, S., & Trappenberg, T. (2007). Learning in sparse attractor networks with inhibition. In *International conference on cognitive neurodynamics iccn'07*.
- Zemel, R., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population code. *Neural Computation*, *10*, 403–430.