Spatial Representations in the Entorhino-Hippocampal Circuit

Federico Stella

Cognitive Neuroscience

SISSA

Supervisor
Alessandro Treves

A thesis submitted for the degree of

Doctor of Philosophy

24.01.2014

Abstract

After a general introduction and a brief review of the available experimental data on spatial representations (chapter 2), this thesis is divided into two main parts. The first part, comprising the chapters from 3 to 6, is dedicated to grid cells. In chapter 3 we present and discuss the various models proposed for explaining grid cells formation. In chapter 4 and 5 we study our model of grid cells generation based on adaptation in the case of non-planar environments, namely in the case of a spherical environment and of three-dimensional space. In chapter 6 we propose a variant of the model where the alignment of the grid axes is induced through reciprocal inhibition, and we suggest that that the inhibitory connections obtained during this learning process can be used to implement a continuous attractor in mEC.

The second part, comprising chapters from 7 to 10 is instead focused on place cell representations. In chapter 7 we analyze the differences between place cells and grid cells in terms on information content, in chapter 8 we describe the properties of attractor dynamics in our model of the Ca3 network, and in the following chapter we study the effects of theta oscillations on network dynamics. Finally, in Chapter 10 we analyze to what extent the learning of a new representation, can preserve the topology and the exact metric of physical space.

Contents

Contents iii

List of Figures vii

1 General Introduction 1

2 A Glimpse on Medial Temporal Lobe Spatial Representations 5
- 2.1 Hippocampus and Place Cells . . . 6
  - 2.1.1 Remapping . . . 7
  - 2.1.2 Dorsal and Ventral Hippocampus . . . 9
  - 2.1.3 Re-play, Pre-play and other off-line Activity . . . 10
- 2.2 Entorhinal Cortex and Grid Cells . . . 11
  - 2.2.1 Entorhinal Cortex Layers and other Spatial Cells . . . 14
  - 2.2.2 Grid Cells in other Mammals . . . 15
- 2.3 Ontogeny of Spatial Representations . . . 16
- 2.4 Theta Rhythm and Phase Precession . . . 16

3 Models on the Edge of Gridness 19
- 3.1 A Computational Headache . . . 19
  - 3.1.1 Path Integrating Models of Grid Cells . . . 20
  - 3.1.2 Path Integration is not Sufficient for Stable Representations . . . 23
  - 3.1.3 Path Integration Relies on Pre-Wired Connectivity . . . 23
  - 3.1.4 An Alternative Approach . . . 24
- 3.2 A Model Based on Adaptation . . . 24
  - 3.2.1 Analytical Formulation . . . 25
  - 3.2.2 Full Model Implementation . . . 26
  - 3.2.3 Development and Environment Geometry . . . 26

4 Grids for Small Worlds 29
- 4.1 Life on a Sphere: a Mathematical Analysis . . . 29
  - 4.1.1 Solutions from Combinations of Spherical Harmonics . . . 29
  - 4.1.2 Cost Function Evaluation . . . 31
  - 4.1.3 The Possible Appearance of Other Solutions . . . 33
- 4.2 Numerical Simulations of Grid Development . . . 35
  - 4.2.1 Network Model . . . 35
  - 4.2.2 Results . . . 37
- 4.3 Conclusions . . . 38

5 Learning to Fly 41
- 5.1 Some Remarks on 3D Navigation . . . 41

5.1.1 Neural Recordings in 3D . . . 41
5.1.2 Addressing 3D Representations with the Adaptation Model . . . 42
5.2 Sphere Packing . . . 43
5.3 3D Close Packing . . . 44
5.3.1 FCC . . . 44
5.3.2 HCP . . . 46
5.3.3 The Cost Compared . . . 47
5.4 Simulations . . . 49
5.4.1 The Model . . . 49
5.4.1.1 HD Input . . . 49
5.4.1.2 Collateral Weights . . . 49
5.4.2 Simulation Results . . . 50
5.5 Conclusions . . . 56

6 Building a Continuous Attractor through Adaptation 57
6.1 Introduction . . . 57
6.2 Inhibitory Connections . . . 57
6.3 The Model . . . 58
6.4 Results . . . 59
6.4.1 Grid Alignement in mEC . . . 59
6.4.2 Self-Organization of Inhibitory Connections . . . 59
6.4.3 A Link to Continuous Attractor Models? . . . 61
6.4.4 Long Lasting Effects of Development . . . 62
6.4.5 A Problem with the Phases . . . 63
6.5 Summary . . . 63

7 Telling Apart Position and Context in mEC and Hippocampus 65
7.1 Slow Grids for Fast Place Encoding . . . 65
7.2 Beyond the Rigidity of mEC code . . . 66
7.3 The Hippocampus in our Life . . . 66
7.4 Continuous or not? Attractors in the Medial Temporal Lobe . . . 69
7.4.1 Discrete and Continuous Attractors . . . 69
7.4.2 An Ideal too Far? . . . 71
7.5 Information Measures . . . 72
7.6 Coding Precision of the Grid Cells Code . . . 73
7.7 mEC and Context Discrimination . . . 74
7.8 Information in CA3 . . . 75
7.9 The Next Sections . . . 76

8 Attractor Dynamics in Ca3 79
8.1 Dissecting the Problem . . . 79
8.2 The Model . . . 79
8.2.1 Simulations . . . 80
8.3 Results . . . 82
8.4 Energy . . . 83
8.5 Drifting Towards Crowded Locations . . . 85
8.6 Multiple Environments Case . . . 89
8.7 Discussion . . . 91

9 Theta Oscillations and Attractor Dynamics **95**
9.1 Theta Oscillations in the Rodent Hippocampus . . . 95
9.1.1 Not One Theta Oscillator but Many . . . 95
9.1.2 Influence on Firing Rates . . . 97
9.2 Flickerings in CA3 . . . 98
9.3 Our Model of Teleportation . . . 99
9.3.1 The Model . . . 99
9.3.2 Results . . . 101
9.4 Discussion . . . 103
9.4.1 Some Remarks on Information Flow in the Hippocampus . . . 107
9.4.2 Beside Theta . . . 109

10 Metric Structure of the Internal Representations of Space **111**
10.1 Introduction . . . 111
10.2 Metric Content . . . 112
10.3 Simulations . . . 112
10.4 Analysis and Results . . . 113
10.4.1 Global Metric . . . 113
10.4.2 Decoding . . . 114
10.4.3 The Spatial Confusion Matrix . . . 115
10.4.4 Metric Content . . . 116
10.4.5 Full and Reduced Matrix . . . 116
10.4.6 The Metric Induced by an Active DG . . . 118
10.4.7 Metric Resolution without DG input . . . 120
10.5 Discussion . . . 122

11 Conclusions **125**

Appendix Appendix A **129**
A-1 Drift . . . 131

Appendix Appendix B **135**
B-2 Basic Model for CA3 Simulations . . . 135
B-2.1 The Storage of New Representations . . . 136
B-2.2 Recurrent Collateral Plasticity . . . 137

List of Figures

2.1 Hippocampus. Position and anatomy . . . 6
2.2 Hippocampus. Internal circuit . . . 7
2.3 Spatial pattern of a place cell activity . . . 8
2.4 Connectivity organization of the hippocampal region in the medial temporal lobe . . . 12
2.5 Entorhinal cortex location and connections . . . 13
2.6 Spatial pattern of a grid cell activity . . . 13
2.7 Head direction and border cells . . . 15
2.8 Phase precession . . . 17

3.1 Continuous attractor model of grid cells . . . 21
3.2 interference model of grid cells . . . 22
3.3 Adaptation model of grid cells . . . 25

4.1 Analytical solutions for grids on the sphere . . . 30
4.2 Legendre projections of the Exponential kernel . . . 33
4.3 Legendre projections of the Gaussian kernel . . . 34
4.4 Numerical minimization results . . . 34
4.5 Summary of the model components . . . 36
4.6 Simulation results for different sphere size . . . 39
4.7 Examples of grids expressing intermediate configurations . . . 40

5.1 Layer alternation in FCC and HCP . . . 44
5.2 Cost function for FCC and HCP . . . 47
5.3 Analytic formulation of FCC and HCP, and their autocorrelograms . . . 48
5.4 Example of a 3D grid . . . 51
5.5 Mean field spacing . . . 52
5.6 FCC and HCP score evaluation . . . 53
5.7 Layer emergence and orientation . . . 54
5.8 Layer arrangement scores . . . 55

6.1 Examples of grid cells developed with inhibitory connectivity . . . 60
6.2 Grid score and axis orientation with inhibitory connectivity . . . 60
6.3 Distance-Phase Relation . . . 61
6.4 Grid cells in 1D . . . 62
6.5 Adjacency matrix in 1D . . . 63
6.6 Phase distribution of the developed grid cells . . . 64

7.1 Convergence of spatial and non-spatial information in the hippocampus 68
7.2 Discrete and continuous attractors . . . 70

7.3 Continuous attractor as an ideal concept . . . 72
7.4 Position and context discrimination in mEC and CA3 . . . 75
7.5 Hippocampus as a generator of multiple maps . . . 77

8.1 Effect of connectivity dilution on drift . . . 82
8.2 Effect of place fields locations disorder on drift . . . 83
8.3 Temporal dependance of drift for different network size . . . 84
8.4 Examples of two CA3 representations energy landscapes, with low and high disorder . . . 84
8.5 Temporal dynamics of drift: velocity and energy variation. . . . 85
8.6 Energy range for different network size . . . 86
8.7 Example of an energy landscape against the corresponding density of fields spatial distribution . . . 87
8.8 Energy-Fields Density relation . . . 88
8.9 Final configurations have low energies and high fields density . . . 88
8.10 Number of environment transitions . . . 90
8.11 Distribution of balance between entering and exiting trajectories among environments . . . 92
8.12 Correlation and shared units between transition locations . . . 93
8.13 Indirect transitions . . . 93

9.1 Phase relationships of theta oscillations and firing rate modulation in MTL . . . 96
9.2 Examples of simulation outcomes in the teleportation model . . . 102
9.3 Flickrs dependance on the parameters of the model . . . 103
9.4 Effect of collaterals strength on flickers frequency and duration . . . 103
9.5 Theta oscillations internal attractor dynamics . . . 104
9.6 Theta oscillations alternate recurrent and input driven dynamics . . . 105
9.7 Different forms of attractors dynamics . . . 107
9.8 Transition induced by a single oscillation event . . . 110

10.1 Global topology of a Ca3 spatial representation with MDS . . . 114
10.2 Full and reduced confusion matrix . . . 117
10.3 Information and percent correct dependance on sample size . . . 118
10.4 The Gaussian fit parameters dependance on sample size . . . 120
10.5 Metric resolution of a CA3 spatial representation . . . 121

1 Solutions of the minimization. . . . 131
2 Standard deviation of the solutions . . . 132
3 Energy dependance on the field density . . . 132
4 Analytical evaluation of drift length. . . . 134

Chapter 1

General Introduction

The medial temporal lobe, the system of networks surrounding the hippocampus, is particularly appealing to modeling for two reasons: on the one hand it presents a collection of closely related spatial representations (discovered in rodents over a 40 years span), with similar properties, expressed by groups of neurons at a distance of 1-2 synapses. On the other hand it exhibits an architecture of the connections among these networks which, if not completely straightforward, certainly is much simpler and better defined than anywhere else in the cortex, shaping a clear hierarchical structure with the hippocampus sitting at the top of multiple lines of converging inputs. This combination, the equivalent of which can be found perhaps only in the early visual areas, is ideal to develop causal models of the function of the region: ideally, one knows which inputs one should combine, and knows what information is conveyed by those inputs. Although not all the pieces of the puzzle are known, the amount of knowledge accumulated in the last decades on the hippocampus, and more recently on the neighboring areas, allows to at least try to define a division of work among these networks, to describe the mechanics underlying the generation of the various representations, each of them performing specific computations and providing different kinds of information to the neighbors.

One of the distinctions that has more vividly emerged from this attempt to define a work flow in the medial temporal lobe, is the one between place cells (O'Keefe and Dostrovsky 1971) and grid cells(Hafting, Fyhn, Molden, et al. 2005). Discovered at a distance of 30 years from one another, they are the prominent units found in the hippocampus and in medial entorhinal cortex, respectively. Grid cells sit one synapse away from hippocampal place cells both as an input of the hippocampal circuit, and as an output (Kloosterman, Haeften, and Lopes da Silva 2004; Witter and E. Moser 2006). Their interaction is reflected in the similarity of the two representations, which nevertheless exhibit a high degree of independence in response to change in the external conditions, and moreover, seem to depend on different dynamical processes.

This thesis is articulated in two main parts which reflect the contrast between grid cells and place cells. The first half is devoted to discuss extensions of the adaptation model for grid cell generation initially introduced by Kropff and Treves (2008). After a general preamble on spatial representations (Chapter 2) and on computational approaches to modeling grid cell emergence (Chapter 3), we discuss the effect a non-planar topology of the environment has on the appearance of grid cells produced by our model. In Chapter 4 (published in Stella et al. (2013a)) we address the development of grids on

a spherical surface, while in Chapter 5 we test our model in the case of the exploration of a three-dimensional environment by a flying animal (partially published in Stella et al. (2013b)). We observe that, indeed, the outcome of the self-organization process on which our model is based is affected by the metric of the external environment. The same model, with no modifications from the planar two-dimensional case, is able to produce different grid arrangements on the sphere, where the result is a different combination of spherical harmonics, depending on the size of the sphere, and in the three-dimensional volume, where the model induces a mixture of Face Centered Cubic and Hexagonal Close Packed arrangements in the fields.

We then proceed (Chapter 6, done in collaboration with Yasser Roudi and Benjamin Dunn at the Kavli Institute for Systems Neuroscience in Trondheim) to propose a variant of the model where the alignment of the grid axes is induced through reciprocal inhibition of mEC units. In this version of the model, plasticity modifications are extended also to this additional set of inhibitory connections, so that the final state is completely the product of self-organization. We argue that the inhibitory connections obtained during this learning process, might be functional to establish the necessary scaffold for the implementation of a continuous attractor network in mEC. In this way, the configuration of fields that has emerged from adaptation might be transferred to the particular organization of the recurrent connections, which in turn might be later used to reproduce the same grid pattern.

The constancy of grid cell pattern of activity across environments is in marked contrast with the high degree of variability expressed by place cell representations. The second half of the thesis is an attempt to characterize some of the aspects of this variability. In particular how the presence of flexible relations between the activity of units and of multiple, distinct maps, stored together, can influence the functionality of a network like the CA3 subfield of the hippocampus. In Chapter 7 (based on Stella, Cerasti, Si, et al. (2012)), we start by discussing how a dissociation between mEC and CA3 representations can be unveiled by means of information content analysis. The results are consistent with the view of mEC entirely dedicated to provide accurate estimation of the spatial position only, and of the hippocampus working to produce context specific representations, in which the spatial component is only one of the many contributions converging to shape the representation.

In the following two chapters, Chapter 8 and Chapter 9 (an expanded version of Stella and Treves (2011)) we study attractor driven dynamics of a CA3-like network, as a way to investigate the structure of the stored spatial representations, and their effective resemblance to a continuous attractor. In agreement with the information content argument, we show that input-free dynamics in the network is far from approximating a continuous attractor. Rather than stably representing the current position, the activity in the network "drifts" toward certain configurations, a sub-set of specific locations in the environment corresponding to the minima of an energy function (In Appendix 11 we also present an analytic approach to the same subject, developed in a collaboration with Remi Monasson at ENS in Paris). When multiple maps are stored on the network, the activity is also found to leave the original basin of attraction, transitioning to another environment representation. This part of the work is also the result of discussions I had with Remi Monasson and Sophie Rosay, at ENS in Paris, who have been

working on the same kind of analysis on their model of CA3 (presented in Rosay and Monasson 2013) and who contributed to this section. We then show how a modulation of the network activity, such as theta oscillations, can induce a switch in the dynamics, from input-driven to recurrent-driven. We use this mechanism to model the results of the "teleportation" experiment (Jezek et al. 2011), but the model offers a more general perspective on the way the network could access the information stored on the internal connections, independently of the external input, or only in a partially independent way, with just an initial external cue.

Finally, in Chapter 10 (based on Stella, Cerasti, and Treves (2013)) we analyze to what extent the learning of a new representation, a mostly random process of self-organization, can preserve the topology and the exact metric of physical space. We present a way to assess this issue quantitatively, and find that in a simple neural network model of CA3, the average topology is largely preserved, but the local metric is loose, retaining only a portion of the optimal spatial resolution.

Chapter 2

A Glimpse on Medial Temporal Lobe Spatial Representations

The study of spatial representations in the Medial Temporal Cortex has had a major influence on the neuroscientific endeavor of the last half of a century. Before the seventies, all informations relevant to understand neural computations was obtained from the very first sensory areas, and especially the early visual stream (Hubel and Wiesel 1959). Any attempt to move further down-stream, away from them, was frustrated by difficulties in interpreting the neural correlates of the observed activity and even more by the exponentially growing complexity of the connectivity structure, making it increasingly difficult to properly trace the afferent and efferent brain regions diagram (the famous Fellman and Van Essen graph is paradigmatic in this respect (Felleman and Van Essen 1991)).

The discovery of neural representations in the hippocampus brought a revolution to the field. It clearly demonstrated that neuronal coding could be studied scientifically also beyond the sensory-motor periphery, deep inside the brain, into the advanced cortices (E. Moser and M. Moser 2013) responsible for the more complex aspects of behavior.

Today, a collection of diverse neural populations, mainly located in hippocampus, entorhinal cortex and other regions of the medial temporal lobe are labelled as spatial representations. In the years more and more members have joined the group, starting with Place Cells (O'Keefe and Dostrovsky 1971), followed by Head Direction cells (Taube, Muller, and Ranck 1990a), Grid cells (Hafting, Fyhn, Molden, et al. 2005), Border/Boundary Vector cells (Solstad et al. 2008; Lever et al. 2009). Together they form an extended space circuit distributed between hippocampus and entorhinal cortex (E. Moser, Kropff, and M. Moser 2008; E. Moser and M. Moser 2013).

In this first chapter we will quickly review the experimental data concerning these representations, to provide the neurobiological basis for the later discussions of modeling studies. It is intentionally partial and incomplete, as a full account of the last 50 years of research on the subject would need entire books (Andersen et al. 2006; S. J. Mizumori 2008). We will therefore focus on some of the basic aspects characterizing the activity of these cells. We will return on some of these later in thesis.

2.1 Hippocampus and Place Cells

The hippocampus is a non-uniform structure with an articulated internal organization. It is comprised of sub-fields, the most important of which are the Dentate Gyrus (DG), CA3 and CA1, that are also the consecutive steps in the flow of signals traveling from the cortex into the hippocampus and then out again. We will mention them in the following only when a particular fact is restricted to one of them, otherwise discussing the general and common properties of the hippocampus as a whole.

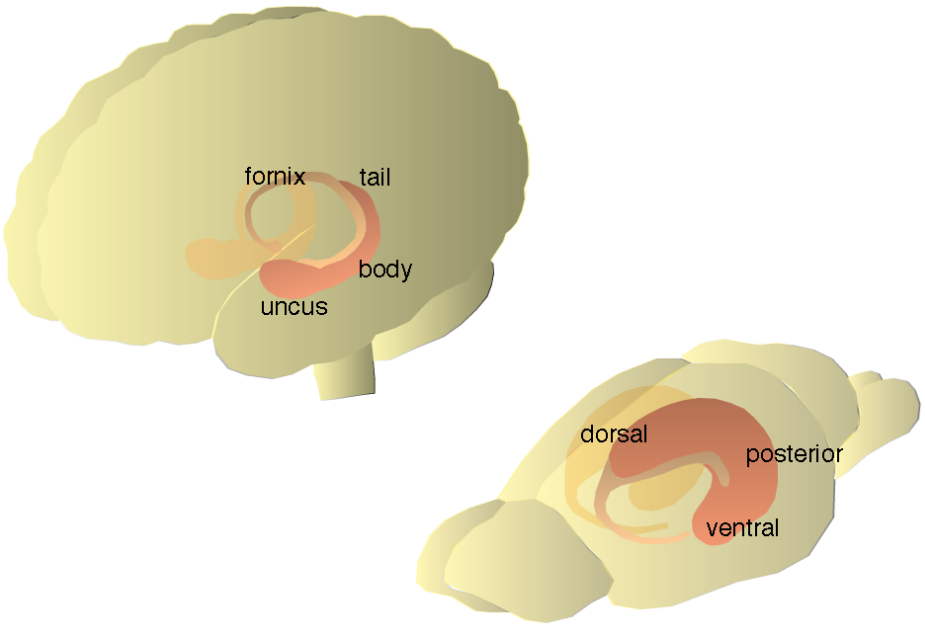


Figure 2.1: Position and shape of the hippocampus in humans (top) and in rodents (bottom). Adapted from Buzsaki, Scholarpedia page on the hippocampus.

A history of place cells can be summarized in very simple terms. O'Keefe and Dostrovsky discovered that neurons in the rat hippocampus respond to the animal's location within the environment. Each neuron is activated whenever the rat is in a particular place, the so-called place field. They published their first findings in 1971, and several details were clarified in the following years.
The activity of each neuron can be reduced visually to its place field (tacitly assumed to be unique, and neatly circular in shape, although real place fields often deviate from the stereotype). Different neurons are active in different locations, so that the entire environment is covered with these place fields; by considering the entire population, each location in the environment is associated with the activity of some of the neurons, an "ensemble" (Wilson and B. McNaughton 1993). The place representation was shown to be non-topographic in the sense that the place fields of neighboring cells (in terms of physical distance in the hippocampal tissue) appeared no more similar than place fields of more widely spaced neurons. The nature of place cell activity was immediately considered as a realization of Tolman's idea of the cognitive map (Tolman 1948; O'Keefe and Nadel 1978). Consequently, most of the following research has been aimed at defining the properties of the external environment and of the animal

behavior affecting these representations.

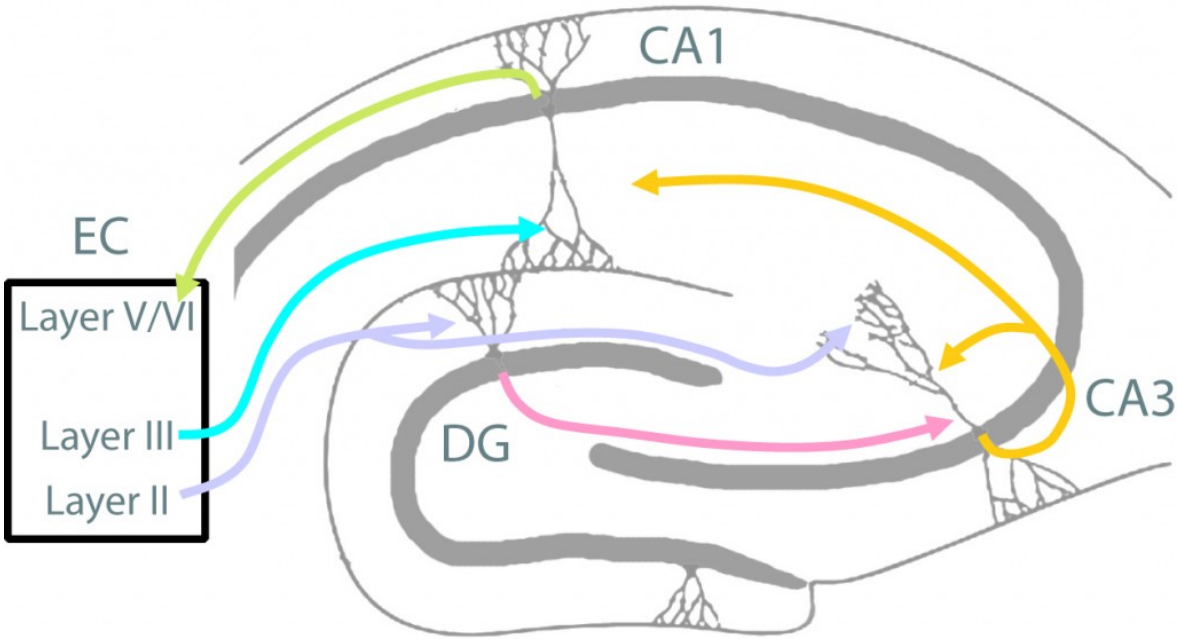


Figure 2.2: Basic scheme of the hippocampal internal circuit. The three main areas (DG, CA3, CA1) are indicated and the arrows show the main directions of signal propagation. Entorhinal Cortex (EC) provides the main input to the complex from its superficial layers (II/III) and receives most of the output to deep layers (V/VI). Adapted from Kamran Diba

2.1.1 Remapping

Global remapping Place cell representations respond to changes in the external properties of the environment in more than one way. At first it was observed that changing the experimental enclosure where recordings are performed, induced a complete rearrangement of cell activity. Most of the cells coding for the new environment were silent in the first one, and if they were active, there was no relation between the place field positions in the two. This kind of response is now called "global remapping" and it basically amounts to the generation of orthogonal, uncorrelated maps for different environments, sharing a very low degree of overlap (S. Leutgeb, J. Leutgeb, Treves, et al. 2004). This phenomenon is consistent with the idea of place cells as a cognitive map, as it is necessary for successful behavior to distinguish among enclosures, even when they are similar in shape or some other property.

Moreover this property of place cell representations was used to successfully demonstrate attractor dynamics in the CA3 sub-field of hippocampus (Guzowski, J. Knierim, and E. Moser 2004; J. Leutgeb, S. Leutgeb, Treves, et al. 2005; Wills, Lever, et al. 2005; Colgin, S. Leutgeb, et al. 2010). When two different recording boxes are morphed one into the other (for example by gradually transforming a rectangular shape into a circular one) it is observed that CA3 representation switch abruptly, between the maps associated to the two end-points of the series, at a certain stage of the morphing continuum.

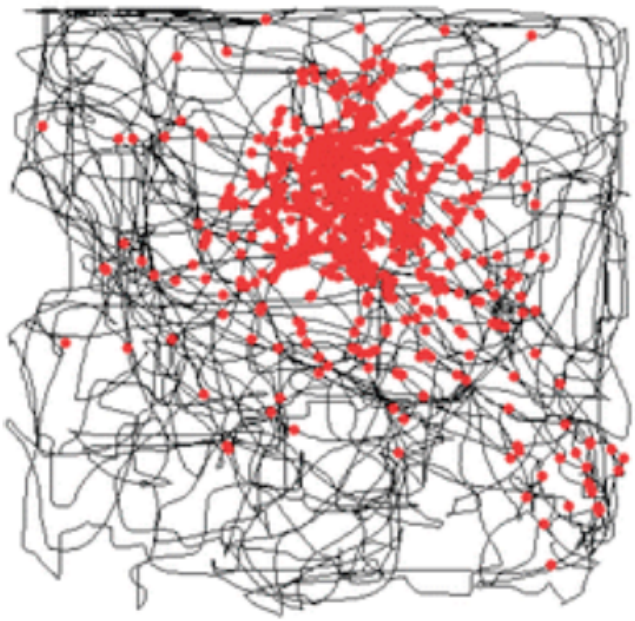


Figure 2.3: Example of the spatial distribution of spikes of an hippocampal place cell. In black the trajectory of the rat. In red, superimposed, the positions of occurrence of spikes. Adapted from (Marozzi and Jeffery 2012)

Rate Remapping The above is not the only way place cells code for modifications in the surroundings. It took thirty years, and more refined recording techniques to actually identify an alternative remapping mechanism, which was named "rate remapping" (S. Leutgeb, J. Leutgeb, Barnes, et al. 2005; S. Leutgeb, J. Leutgeb, E. Moser, et al. 2006). In this case, the same set of cells is maintained to code for the new environment, and they also exhibit the same place fields arrangement. Only the peak firing rate inside the place fields is changed independently by the various cells. In this way, while the place fields remain stable, the local code for each position, the ensemble, is modified, since the relative firing rates among the units active in the same spot are altered. This kind of remapping is observed when the modifications in the environment are minor, and in particular when the rat is kept in the same room, although in a different box. Rate remapping is thought to be a way to encode subtle perceived differences in an otherwise familiar situation and is therefore alternative to the global one that instead signals a drastic change (Ahmed and Mehta 2009).

Another line of inquiry has investigated the influence of the shape of the environment, and in particular of symmetries in the shape, on the place cell code. Classic examples are the hairpin maze (Derdikman, Whitlock, et al. 2009) and the multiple arms (or multicompartement) maze (Singer et al. 2010; Spiers et al. 2013). These set-ups are used to understand how place cells code responds to the similarities between portions of the environment and to stereotypical behavior, such as a repetitive series of left and right turns. Although there are some differences between results obtained with different paradigms, there is also evidence that most of the place cells do not discriminate

among similar parts of the maze (Derdikman, Whitlock, et al. 2009; Spiers et al. 2013), getting reset and repeating the same firing pattern in each of them.

Non-Spatial Components of Place Cells Representations While the previous examples all relate to the effects on the place cell code of the transformation of the physical properties of the environment, others experiments have tried to delineate to what extent the non-spatial elements of a task may contribute to place cell code variability. Some of the early indications in this direction came from the observation that during an alternation task on a Y-shaped maze, a different set of place cells was active on the central arm, depending on the side of the next turn (Wood et al. 2000; Ferbinteanu and Shapiro 2003; Smith and S. Mizumori 2006; Bahar, Shirvalkar, and Shapiro 2011; Ferbinteanu, Shirvalkar, and Shapiro 2011; Allen et al. 2012). A more evident example of the presence of an "episodic" component in place cells activity comes from a limited number of studies linking the properties of the representation to the positions of the goal (generally the foraging or reward location). The studies agree on a tendency of place cells to exhibit a higher density of place fields around and near goal locations (Hollup et al. 2001; Tabuchi, Mulder, and Wiener 2003; I. Lee, Griffin, et al. 2006; Hok et al. 2007; O'Neill, Senior, Allen, et al. 2008).

2.1.2 Dorsal and Ventral Hippocampus

The hippocampal formation is also characterized by a functional organization along its dorso-ventral axis. The dorsal hippocapmus is connected to dorsal entorhinal cortex as ventral hippocampus connects to ventral Entorhinal Cortex. Moreover, anatomical studies indicates that the input and output connections of the dorsal hippocampus and ventral hippocampus are distinct (Strien, Cappaert, and Witter 2009).
Place cells are found along almost the entire axis but place fields dimensions exhibit a continuous gradient moving along it, with dorsal place cells forming small, very localized fields, while progressively more ventral cells have broader and broader fields, covering an increasing portion of the environment (Jung, Wiener, and B. McNaughton 1994).

This differentiation has led to hypothesize that place cells belonging to different portions of hippocampus might be used to simultaneously represent the environment at different spatial scales. Not necessarily contradicting the idea, it is also known (E. Moser, M. Moser, and Andersen 1993; M. Moser et al. 1995; Bannerman et al. 2003) that performing a standard Morris water maze task, rats with a lesioned dorsal hippocampus are severely impaired, while the effect is not observed in rats with equivalent ventral lesions, that show no decrease in performance. That spatial navigation functions might be concentrated in the dorsal domain of the hippocampus is the conclusion to which many studies on stress, emotions and affect came, since they found an inverse correlation in these non-spatial functions, with the ventral, more than the dorsal hippocampus, deeply involved in their regulation (Fanselow and Dong 2010; Komorowski et al. 2013).

The real existence of two completely distinct, functionally separated structures is still a matter of debate and is quite implausible (Lubenov and Siapas 2009; Patel, Fujisawa, et al. 2012), still, the observation of a similar pattern of dissociation in human studies (Poppenk et al. 2013) hints at a complex and rich interaction between various parts of

the region.

2.1.3 Re-play, Pre-play and other off-line Activity

Reactivation and Sharp Wave Ripples One of the signatures of place cell coding is that it represents the actual position of the animal in space. Place cells get activated when the rat traverses the position associated to its firing field. Also when lights are extinguished place fields remain stable, keeping track of the position of the animal in space. As a matter of fact this description is only partial. Recording from the hippocampus during slow wave sleep, Wilson and B. McNaughton (1994), in 1994 found that rather than being silent, place cells were reactivated in sequences. Cells that fired together when the animal occupied particular locations in the environment exhibited an increased tendency to fire together during subsequent sleep. The sequential activation of these cells can be viewed as a neural representation of the animal's trajectory. These sequences are repeated or replayed during sharp wave ripple complexes (SWRs), network events observed in the hippocampal local field potential (O'Keefe and Nadel 1978; Buzsaki, Leung, and Vanderwolf 1983). Initially reported during sleep (Wilson and B. McNaughton 1994; Skaggs and B. McNaughton 1996; Nadasdy et al. 1999; Sutherland and B. McNaughton 2000; A. Lee and Wilson 2002), these events has been then reported during rest periods between trials (Foster and Wilson 2006; Jackson, A. Johnson, and A. Redish 2006; O'Neill, Senior, and Csicsvari 2006; Csicsvari, O'Neill, et al. 2007; Diba and Buzsaki 2007; A. Johnson, Jackson, and A. D. Redish 2008; Davidson, Kloosterman, and Wilson 2009; Karlsson and Frank 2009; Gupta et al. 2010), and more recently during the trial itself (Dupret, Joseph, Barty, et al. 2010; Pfeiffer and Foster 2013).

During the awake rest state, representations of previous experiences are reactivated(Jensen and Lisman 2000; Jackson, A. Johnson, and A. Redish 2006; O'Neill, Senior, and Csicsvari 2006), and are replayed both in the order that they were experienced (forward replay; Diba and Buzsaki (2007); A. Johnson, Jackson, and A. D. Redish (2008); Davidson, Kloosterman, and Wilson (2009); Karlsson and Frank (2009)) and in the reverse order that they were experienced (backward replay; Foster and Wilson (2006); Csicsvari, O'Neill, et al. (2007); Diba and Buzsaki (2007); Davidson, Kloosterman, and Wilson (2009)). In 2010, Gupta et al. (Gupta et al. 2010) also showed that the sequences replayed do not necessarily correspond to a trajectory actually experienced in the rat. Using a multiple choice T-maze, they were able to observe the activation of sequences of units coding for shortcuts and alternative routes, that were never taken by the rat.

Replay of behavioral sequences in the hippocampus provides a possible mechanism for several critical functions in which the hippocampal formation plays a role: consolidation of experiences into long-term memory (Marr 1971b; Buzsaki 1989; Sutherland and B. McNaughton 2000), incorporation or consolidation of information into cognitive schemas or cognitive maps (O'Keefe and Nadel 1978; Tse et al. 2007), and learning and planning future experience (Foster and Wilson 2006; Diba and Buzsaki 2007; Karlsson and Frank 2009; Lansink et al. 2009). To understand the role of SWRs and replay in hippocampal function, recent studies have blocked CA3 output (the site of SWR generation; Ylinen et al. (1995)) (Nakashiba, Buhl, et al. 2009) and have directly disrupted SWRs after learning, during sleep (Girardeau et al. 2009; Ego-Stengel and

Wilson 2010), showing that these manipulations impair performance on hippocampal-dependent behavioral tasks. Although these studies show that SWRs are important for correct task performance, they cannot specify the nature of learning taking place during these events.

Future Paths Apart from speculations on the functional role of reactivation, until very recently, there was very little that allowed to establish a direct link between the sweeps of reactivated trajectories and behavioral planning. The relation was thought to be at most deferred, with replays inducing the building of a flexible cognitive map. Instead Pfeiffer and Foster (2013) were able to detect the activation of place cells coding for paths ahead of the animal, while the rat was performing a foraging task. The study showed how these paths were then likely to be the ones chosen by the rat in the movements that followed the reactivation event. Moreover, the end point of the activated sequence was often a meaningful location of the environment, such as the foraging spot. The sequence thus traced a direct link between the actual position of the rat and a goal location at which the rat then headed. The results need to be confirmed in other experiments, but surely they appear as a clear evidence of a mechanism allowing the hippocampus to actively probe its internal representations of space in order to provide information on spatial locations beyond the animal current position.

Preplays Another form of predictive activity in the hippocampus is the one described in a number of studies by Dragoi and Tonegawa (Dragoi and Tonegawa 2011; Dragoi and Tonegawa 2013). Again they recorded off-line activity in the hippocampus during SWR events, but this time they did it in the period preceding the experiment. What they found is that sequential activation of cells is observed anyway, but most remarkably there is a very high degree of overlap between these sequences and the cells then used to represent the environment. These early activations look like "preplays": even though the rat has no experience of the environment yet, it is already possible to observe the groups of cells that will be recruited in the following session to build the spatial representation. The results are challenging for the traditional view of hippocampal representations as random constructs, and might hint at some sort of pre-wired structure present in the circuit, or to some aspects of learning not yet fully understood (Mizuseki and Buzsaki 2013).

2.2 Entorhinal Cortex and Grid Cells

Since the discovery of place cells, the apparent uniqueness of these units and the elegant spatial code expressed in the subfields of the hippocampus led many investigators to presume that place signals were mostly a product of computations performed entirely inside hippocampal circuitry. But just as soon as part of the scientific community was getting comfortable with the idea that the hippocampus was central to spatial navigation, Fyhn et al. in 2004 made a discovery that messed things up and shed a new light on the flow of information in the medial temporal lobe (Fyhn, Molden, et al. 2004).

There was already some clue indenting the idea of the hippocampus as the original generator of the position code. Spatial firing was found to persist in CA1 after removal of the intrahippocampal connections arising from the dentate gyrus (B. McNaughton, Barnes, et al. 1989) and also from CA3 (Brun, Otnass, et al. 2002). These experiments

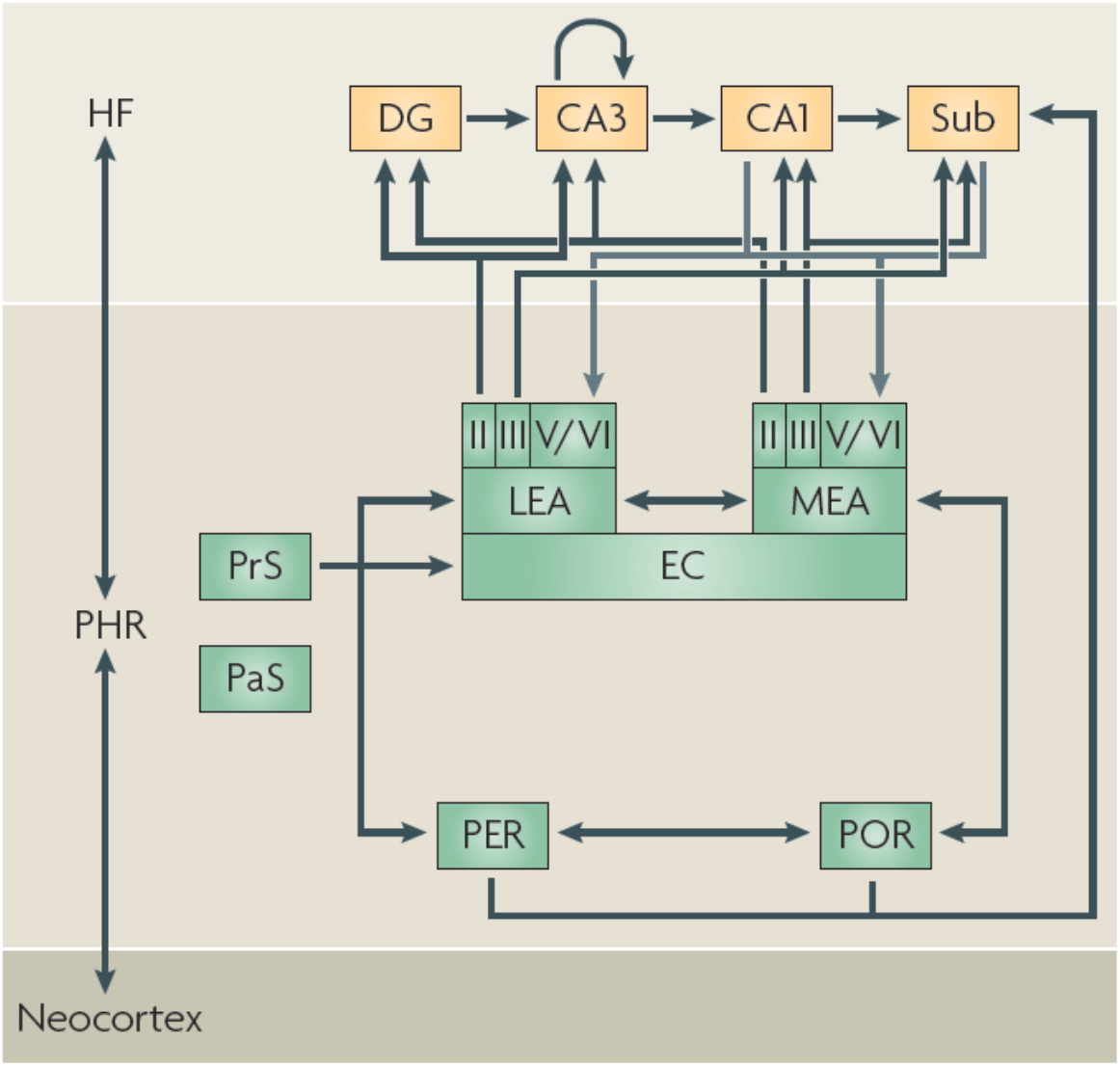


Figure 2.4: Schematic representation of the standard view of the parahippocampal-hippocampal circuitry. According to this standard view, neocortical projections are aimed at the parahippocampal region (PHR), which in turn provides the main source of input to the hippocampal formation (HF). Adapted from (Strien, Cappaert, and Witter 2009)

suggested that some position information must be present already in the entorhinal cortex, to be conveyed to CA1 through the direct perforant-path projections.

Grid Cells What Fhyn et al. found, looking at the activity of neurons in the medial part of the entorhinal cortex, was that these cells responded to the position in space of the rat in a way very similar to place cells; but instead of having only one, or very few randomly displaced place fields, these cells appeared to have many place fields, arranged throughout the environment in a surprisingly regular and precise fashion (Hafting, Fyhn, Molden, et al. 2005). Their firing fields occupy the vertex of a triangular pattern that covers the entire environment. Because of that, and because of

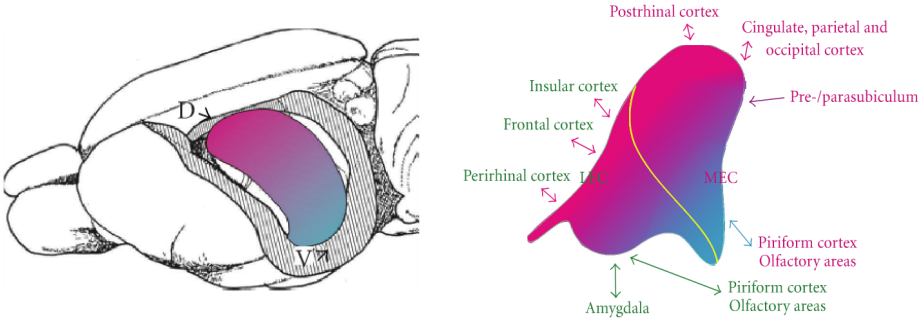


Figure 2.5: Left: Entorhinal Cortex location in the rodents brain. Right: Connectivity differentiation between the dorsal (purple) and ventral (blue) portions of Lateral (LEC) and Medial (MEC) Entorhinal Cortex. Adapted from (Canto, Wouterlood, and Witter 2008).

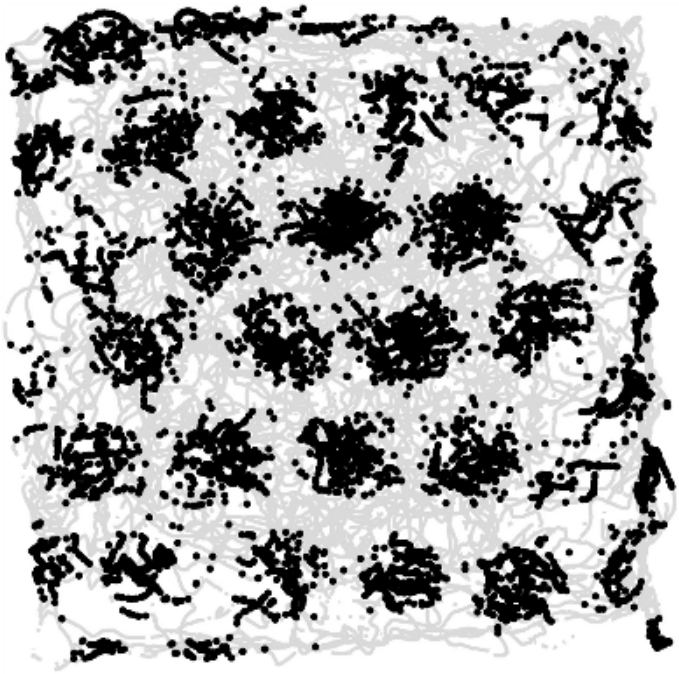


Figure 2.6: An example of the typical triangular arrangement of fields expressed by an entorhinal grid cell. Adapted from (Marozzi and Jeffery 2012)

their resemblance to place cells, they have been named grid cells. Grid cells can differ from each other in their grid spacing and grid phase, while there is growing evidence for an essentially common grid orientation throughout medial entorhinal cortex (Stensola et al. 2012). Cells very close to each other in the tissue have their grids shifted by a seemingly random phase, so that there is no topographic organization in the representation of space.

Dorso-Ventral Gradient The only sign of a topographical organization can be found in a gradual increase in the dimensions of the fields and in the distance between them, going from the more dorsal to the more ventral portion of the medial entorhinal cortex, symmetrically reflecting the analogous hippocampal place field size increase (Brun, Solstad, et al. 2008). Indeed, recent studies, involving extensive recordings of many cells from the same animal, unveiled a modular organization of the grids along the axis. Sampling from a vast portion of the dorso-ventral extension of the region Stensola et al. (2012) were able to identify four or five of these modules. Each module was characterized by a common grid spacing and a common orientation. Therefore the increase in spacing along the axis appears to happen by discrete jumps. What is more, the extension of the jumps is very regular: the grid scale increases, on average, by a factor 1.4 from one module to the next. Interestingly, the change in grid scale is accompanied by a parallel progressive variation of the physiological properties of the cells (Giocomo, Zilli, et al. 2007; Giocomo, Hussaini, et al. 2011; Yoshida, Giocomo, et al. 2011; Yoshida, Jochems, and M. Hasselmo 2013).

When the rat is moved to a new environment, grid cells retain their firing structure, the size of the fields and the distance between fields is maintained, while the grid is shifted and rotated, changing the grid phase and orientation. Most remarkably, the shift and the rotation are common to all nearby cells (Fyhn, Hafting, Treves, et al. 2007) belonging to the same module.

2.2.1 Entorhinal Cortex Layers and other Spatial Cells

Much like the hippocampus, medial entorhinal cortex (mEC) is not an homogeneous structure, and like the rest of mammalian cortices is organized in layers. Each of the five layers is characterized by the presence of different neural types and by a different pattern of afferent and efferent connectivity (Canto, Wouterlood, and Witter 2008). Grid cells studies soon started to probe the different layers to see to what extent grid cells could be found. They found that grid cells actually intermingle with other cell types in the deeper layers of mEC (Sargolini et al. 2006). While grid cells are the dominant cell type in layer II, a large fraction of the intermediate and deeper layers populations is comprised of head direction cells. Interestingly, starting from layer III, most grid cells are also directionally tuned. They show a compositional selectivity, firing only when the animal passes through the grid vertices with its head in a certain direction.

Head Direction Cells The discovery of abundant head direction tuned cells in the medial entorhinal cortex did not constitute a first observation of this kind of cell. They were first reported by James Ranck in 1983 in the rat dorsal presubiculum, and then described in Taube, Muller, and Ranck (1990a); Taube, Muller, and Ranck (1990b). They have subsequently been observed in other species including mice and monkeys. The activity of these cells is highly correlated with the direction of the animal's head and are independent of eye movements. They are controlled by external landmarks and self-motion cues and they are active irrespective of the animal ongoing-behavior. Since their discovery, head direction cells have been found in several other brain areas including thalamus, lateral mammillary nuclei, retrosplenial cortex, striatum, and entorhinal cortex.

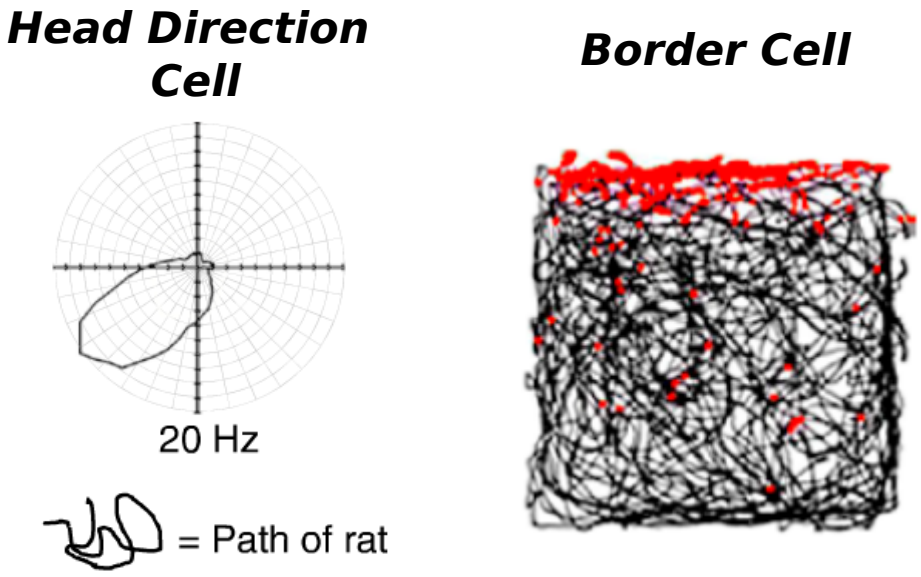


Figure 2.7: Examples of the firing patterns of a Head Direction Cell (left) and of a Border Cell (right). Adapted from (Marozzi and Jeffery 2012)

Border Cells But this was not all, as in 2008 a new kind of spatially selective cell was reported: border cells (Savelli, Yoganarasimha, and J. Knierim 2008; Solstad et al. 2008). These are cells that tie their activity to the proximity of one or more borders of the environment, like a wall or an edge. The firing field of the unit follows modifications of the position of the border, and new fields are added when new borders are introduced in the environment.

Grid cells, head direction cells, and border cells were found to coexist not only in the medial entorhinal cortex, but also in the adjacent presubiculum and parasubiculum (Boccara et al. 2010), demonstrating the extent of the distribution of spatial maps in the entire region.

2.2.2 Grid Cells in other Mammals

Bats Most of the data on grid cells has been obtained using rodents, but recently reports of grid activity in other mammals have begun to accumulate. First of all grid cell activity has been extensively reported in crawling bats (Yartsev, Witter, and Ulanovsky 2011). Notably, the critical features of rodent grid cells are the same in bats: hexagonal firing, coherence of spacing and orientation, the lack of topographic organization of phases and the dorso-ventral gradient. Also the same mixture of grid, head direction and border cells is found in bat mEC. It is important to note the phylogenetic distance between rats and bats.

Monkeys Reports of grid cells also come from a study on monkeys using visual stimuli (Killian, Jutras, and Buffalo 2012). In this case mEC units responses depended not on the position of the monkey but on where the monkey fixated on a visual image. This vision/navigation duality between monkeys and rodents is not new, as visual equivalent of place cells, "spatial view cells", were reported almost 20 years ago (E. Rolls and O'Mara 1995; E. Rolls, Robertson, and Georges-Francois 1997; Kornblith et al. 2013).

Humans Very recently, the generality of grid phenomenon has been reinforced by direct neural recordings in human epileptic patients (Jacobs et al. 2013), which led to reports of grid-like activity in humans, in their Entorhinal Cortex during a virtual navigation task. Although the hexagonal pattern was not as clear as in rodents data, it is indicative of the potential relevance of grid phenomenon also outside the rodent and bat domain.

2.3 Ontogeny of Spatial Representations

The problem of defining the origin of spatial representations has led to perform a series of recording in newborn rodents to gather data on the neural activity in the developing hippocampus and entorhinal cortex (Langston et al. 2010; Wills, Cacucci, et al. 2010; Wills, Barry, and Cacucci 2012). Of all the cells previously mentioned, head direction cells are the first to come online. Head direction cells in pre- and parasubiculum show adultlike properties at the time rat pups leave the nest for the first time, often only hours after eye-opening.

In contrast, both place cells and grid cells take some time after birth to show good spatial selectivity and especially to stabilize their firing fields. Even though a rudimental map of space can be identified after a couple of weeks after rat birth, a period of about 4 weeks of development is required for it to approach its adult-like properties. The stabilization happens at different speed for the two representations, with grid cells being the last to fully develop. While many place cells show consistent receptive fields across trials at younger ages, most grid cells acquire their spatial periodicity gradually, with the mature pattern appearing at the end of the 4 week period. It is important to note that this 4 weeks interval coincides with the emergence of adult-like connectivity within the entorhinal microcircuit.

2.4 Theta Rhythm and Phase Precession

The rodent hippocampal formation shows a prominent and pervasive modulation of its activity in the 8-12 Hz range of frequencies: the theta rhythm. According to the simplest model of theta generation, the medial septum in the basal forebrain functions as a pacemaker, enforcing a global rhythm into which hippocampal and entorhinal cortex (EC) networks are entrained. In general this rhythmic activity appears, at least in rodents, during rodent's active exploration and REM sleep.

An interesting phenomenon associated with theta is the so called phase precession (O'Keefe and Recce 1993). As the rat runs across the field of a place or of a grid cell, the spiking of the unit occurs at a progressively earlier phases in the theta cycle. It has been argued that such a pattern of activity can express a coding mechanism, alternative to the one based on the rate count, allowing for greater accuracy in the estimation of the animal position.

While all layers in mEC are theta modulated, phase precession is not common to all of them. It is particularly strong in layer II units, but it appears very sparsely in the neighboring layer III (Hafting, Fyhn, Bonnevie, et al. 2008). Similarly, also hippocampal sub-fields present a different prevalence of phase precessing units, with those in

CA1 being more strongly affected (Mizuseki, Sirota, et al. 2009).

It is worth noting that now grid cells have been recorded in species in which theta oscillations occur only intermittently, such as bats (Yartsev, Witter, and Ulanovsky 2011; Heys et al. 2013) and possibly in monkeys (Killian, Jutras, and Buffalo 2012). It increasingly appears that theta oscillations (and therefore phase precession) is the only element of a large mosaic of structural elements, connectivity patterns and neural activity that is not universal across species. Without diminishing the importance of these phenomenon it is possible that many considerations on the functional relevance of theta oscillations and phase coding for spatial cognition might be applicable to rodents and might not be directly and automatically applied to other animal models.

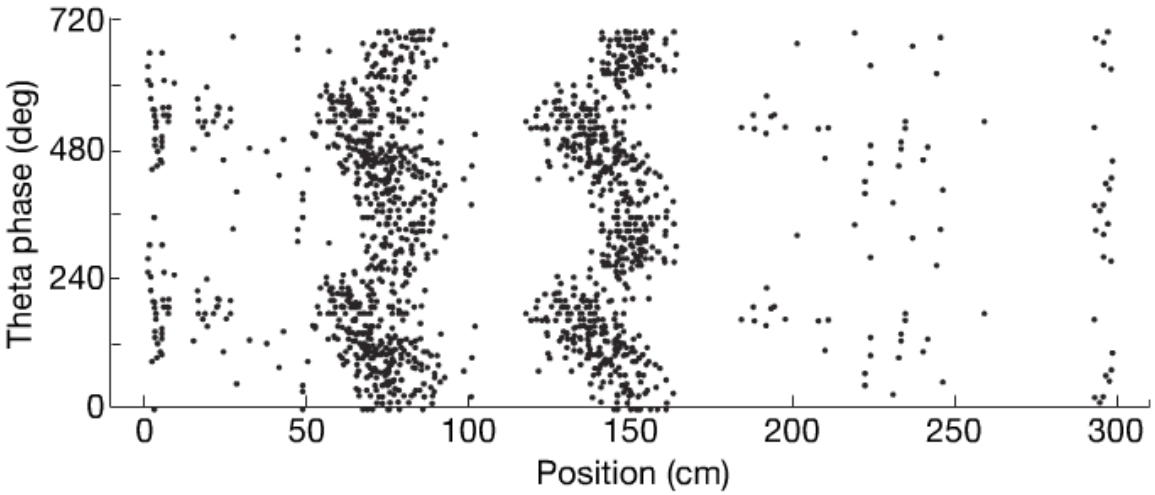


Figure 2.8: Example of a phase precessing cell in Layer II of mEC. Adapted from (Hafting, Fyhn, Bonnevie, et al. 2008)

Chapter 3

Models on the Edge of Gridness

The computational description of grid cells is by no means an isolated field of theoretical neuroscience. The emergence and differentiation of tuning curves among a population of neurons has been an intriguing computational problem for many years. This issue is of great relevance and encompasses phenomena common to a large number of brain regions, ranging from primary sensory areas to higher level ones. From a certain point of view, the study of neural dynamics itself is the study of the origin of neural selectivity and its properties.

In the framework of spatial representations, the discovery of grid cells (Fyhn, Molden, et al. 2004; Hafting, Fyhn, Molden, et al. 2005) happened some 30 years after place cells (O'Keefe and Dostrovsky 1971) and 20 years after head direction cells (Taube, Muller, and Ranck 1990a; Taube, Muller, and Ranck 1990b). The scientific community had already time to think about spatial representations that have a lot in common with grid cells, and in fact most of the tools applied to grid cells case were already available as they were developed before their discovery. Head direction cells and place cells, each with their own specificity, provided challenging examples of continuous spatial representations in one and two dimensions, respectively. Indeed, path integration, continuous attractor networks and oscillations interference are all ideas that although applied to grid cells, have been originally proposed to describe other phenomena. Examples are models originally thought for place cells (Touretzky and A. Redish 1996; Samsonovich and B. McNaughton 1997; B. McNaughton, Battaglia, et al. 2006) and later extended to the new grid cells.

3.1 A Computational Headache

Grid cells are, according to a basic description, cells showing selective responses for certain portions of the environment the animal is in. But this is true also for CA3 place cells, that also have the animal position as a behavioral correlate. One might add that grid cells always show preference for multiple, distinct locations in the environment, but again, one could say the same for CA1 place cells, which are very commonly exhibiting multiple firing fields (Fenton et al. 2008). Of course grids owe their name to the precise position of their multiple fields and are inherently characterized by their strikingly regular appearance. Most of the solutions devised by modelers for the emergence of neural selectivity across the brain are dramatically challenged by the need of combining selectivity and periodicity. Addressing grid cells required the elaboration of new solutions.

The question causal models of mEC units activity have to answer is:

> *How does a Grid cell know when to fire?*

or more precisely:

> *How does a Grid cell know how to form equilateral triangles with its firing fields?*

3.1.1 Path Integrating Models of Grid Cells

In the last ten years, a certain number of proposals have been presented to answer the question (Giocomo, M. Moser, and E. Moser 2011; Zilli 2012), and although all these models differ from each other, they can be effectively grouped in a small number of basic prototypes. The classification can in fact be based on the handful of principles that drive grid generation. Most of the models assume that grid pattern is a reflection of path integration[1], with grid cells either performing it or just receiving the outcome of computations performed somewhere else. Path integration is not a expressed exclusively by grid cells, and in fact it is widely used to describe other forms of neural activity (eg. head direction cells; K. Zhang (1996); Eliasmith (2005); Song and X. Wang (2005); Qu, Yi, and X. Wang (2009)). Therefore these models need some additional ingredient that allows the generation of hexagonal patterns from it. The ingredient used distinguishes the path integrating models into two main sub-classes:

- Continuous Attractor Neual Networks (CANN)
- Interference Based Models

CANN These models rely on the collective activity of the units in the network to produce the grid pattern (Conklin and Eliasmith 2005; Fuhs and Touretzky 2006; B. McNaughton, Battaglia, et al. 2006; Guanella, Kiper, and Verschure 2007; Burak and Fiete 2009; Mhatre, Gorchetchnikov, and Grossberg 2012; Navratilova et al. 2012; Bonnevie et al. 2013; Couey et al. 2013). They can be schematized as a two step process. First, a notion of distance between the units in the network has to be defined. To do so one can think of taking all the units and arranging them on a virtual layer, so that one can then say that unit i and unit j are neighbors, or that unit k is at a distance d from unit i. Connections among these units are then introduced based on this distance. Here is where differences between various implementations emerge. Specific models use a mixture of excitatory and inhibitory connections, or only inhibitory, continuously valued or all-or-none connections and so on, but in all of them the value of the interaction is just a function of the distance between units and this function, the connectivity profile, is uniform and translationally invariant across the virtual layer. The connectivity is chosen so that when the network receives an external perturbation, it settles in a stable state where unit activity shows an hexagonal grid organization (on the virtual layer).

What is then necessary, the second step of the grid formation process, is to transfer the grid from this virtual space to the real space where the animal movements take

[1] Path integration is the process by which a system continuously keeps track of the movements made to provide an estimate of the current position of the agent.

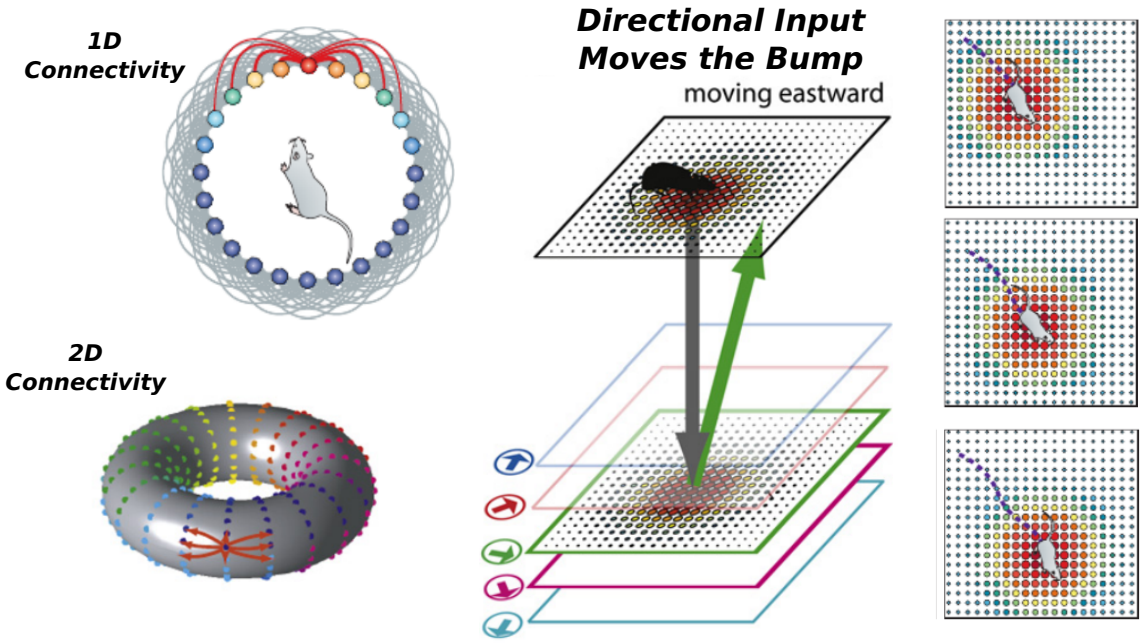


Figure 3.1: Schematic illustration of the building blocks for a continuous attractor grid model. On the left, neighboring units are connected to form stable bumps of activity and, on the right, a layer of conjunctive cells moves network activity on the grid cell layer in the appropriate direction based on rat movement. Adapted from B. McNaughton, Battaglia, et al. (2006)

place. This is generally achieved by assigning a directional selectivity to the units and introducing a combined directional asymmetry in the connections of each of them. By then plugging into the system an input representing the direction and speed of the animal movements, one can literally push the activity profile around the virtual layer, activating different group of cells in time. If one takes one of these units and projects on the animal trajectory the times at which the unit was active, one sees that they actually reproduce an hexagonal grid in the physical environment.

As already said, there are minor differences in the way these two steps are implemented in each of the CANN models present in the literature, but all of them are based on the possibility of moving around, with some external input linked to the rat behavior, an intrinsically stable activity pattern generated by the network. The core of the model is the way units are connected with each other, as the symmetries of the connectivity (radial and translational) enables the activity pattern to emerge and stabilize. Again, the connectivity has to be thought of in an abstract space of distances. The possibility that these distances may correspond to actual physical distances in the cortical tissue has been made implausible by the available evidence (Fyhn, Hafting, Witter, et al. 2008).

Interference Models The other group of models rooting grids into path integration is instead based on a single cell mechanism(Blair, Welday, and K. Zhang 2007; Burgess, Barry, and O'Keefe 2007; M. Hasselmo, Giocomo, and Zilli 2007; Burgess

2008; M. Hasselmo 2008; M. Hasselmo and Brandon 2008; Welday et al. 2011). It doesn't need interaction among units in the mEC network to produce grids, but is instead relying on oscillations at the single neuron level. In the one dimensional case the rationale of the model is simple: assuming that a cell receives two inputs in the form of oscillating signals, then if the two oscillations have a slightly different period, for example one is the omni-pervading theta oscillation, and the other has a slightly higher frequency, then their interference yields in a periodical series of depolarization intervals in which the cell is very likely to fire. If one further assumes that frequency of the second oscillator is linearly modulated by the velocity of the rat (in the direction of the one dimensional movement), then one obtains that the positions of the depolarization fields are independent of the speed of of the animal.
In two dimension the matter is more complicated, as the interference between theta and a single oscillator, whose frequency is modulated by speed only along a particular direction, would produce high activity bands orthogonal to the preferred direction of the velocity tuned oscillator. To actually obtain grids one has to introduce multiple above-theta oscillators all projecting to the same unit, and to consider their combined effect on the cell polarization. In particular, the presence of three in-phase oscillators with the same frequency and with preferred directions differing by $\pi/3$ does the job of producing hexagonal grid patterns.

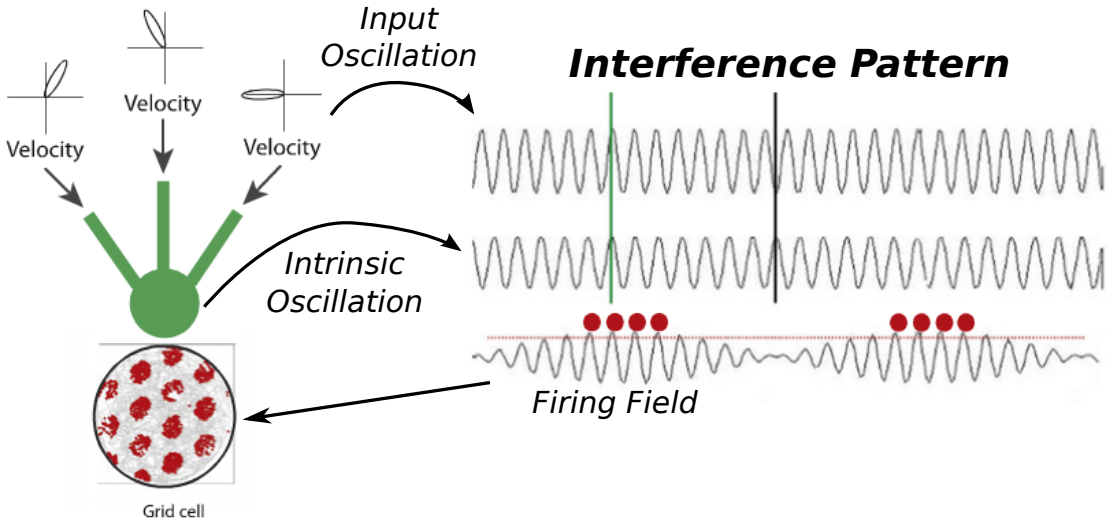


Figure 3.2: Schematic illustration of the basic interference model (Burgess et al. 2007, Giocomo et al. 2007). Speed-modulated head direction input influences the frequency of oscillations in three different dendrites. The three head direction inputs have preferred firing directions offset by 60 degrees relative to each other, as indicated by polar plots above each dendrite. An interference pattern is generated when the dendritic oscillations interact with the baseline soma oscillation. Adapted from Giocomo, M. Moser, and E. Moser (2011).

3.1.2 Path Integration is not Sufficient for Stable Representations

Even though both these classes of models rely on path integration as the main source of positional information to the grids, this is not sufficient to make it all work. Indeed a spatial representation purely based on path integration has two main problems. First, the management the error in the tracking of the position based on proprioceptive informations alone, inevitably accumulates and therefore generates an ever increasing discrepancy between the real position of the animal and the represented one. Second, simple path integration does not lead to the same arrangement of fields when visiting the same environment at later times. The two hypothesis described, are affected by the random trajectory of the rat when generating the grid since the phase and orientation of the grid are induced anew on every new exploration session.

Both these models need an anchoring mechanism to stabilize grids in familiar environments and to update the encoded position of the rat when external landmarks are present, thus resetting the path integrator annulling the error. Generally this anchoring is thought to be provided by some external and potentially independent system, often identified with the hippocampus and with place cells (Bonnevie et al. 2013). This parallel system is supposed to rely on direct sensory information and on memory processes. Place cell fields would be associated to corresponding grid fields, so that reactivation of the former would bring the latter into the correct position.

3.1.3 Path Integration Relies on Pre-Wired Connectivity

Both of these groups of models produce hexagonal grid patterns in an open arena, and they make use of certain properties of the mEC region such as theta oscillations, internal connectivity and a variety of velocity tuned signals in the neighboring cortices and in mEC itself. They also generate a number of predictions on various aspects of network behavior, from input currents on single neurons to global activity states. These predictions can be verified experimentally or, sometimes, rejected (Domnisoru, Kinkhabwala, and Tank 2013; Heys et al. 2013; Schmidt-Hieber and Hausser 2013; Yoon et al. 2013). While there is no final word yet on the validity of either of these models, they suffer also from a more intrinsic problem. Indeed, both if you think grids are the product of a continuous attractor network implemented in some connections of mEC, or that they reflect the existence of directionally tuned oscillators somewhere upstream, you are assuming the existence of some pre-wired organization of the connections already in place, that deliver grid cells.

Even if one of these models were to be proven valid, there would remain the necessity of providing the means to put all the scaffolding these models require to work is in its precise place. The challenge of explaining the appearance of order (in the form of grid cells) in an otherwise messy place such as the brain, in these models, is just recast as that of setting up a set of connections or velocity tuning specifically designed to produce hexagonal grids in an open environment. The problem seems to have just moved some steps away, also, there is the assumption that the outcome of the system has to necessarily be hexagonal grids whatever the conditions the animal finds itself in. The fact that until now, only hexagonal grids have been experimentally observed in mEC, is surely a good reason to start with them, but the possibility that this might be the effect of the environment experiments use, cannot be neglected. The data currently

available still leaves open a number of questions and issues on the universality of the hexagonal form of the grids.

3.1.4 An Alternative Approach

These considerations suggest that an alternative approach to the problem might be welcome. If we leave aside the assumptions that grid cells rely directly on path integration for expressing their characteristic pattern of activity, we might turn to another class of models that aims to generate grid patterns, those based on self-organization (Kropff and Treves 2008; Grossberg and Pilly 2012; Si, Kropff, and Treves 2012; Pilly and Grossberg 2013). In this case the spotlight is on how learning processes can shape the activity of mEC cells to generate, without any kind of supervision or instruction, the regular pattern. Generally, these models start from some sort of spatially modulated activity that can be more or less structured (random place fields or stripe cells) which works as an input to the network representing mEC. How this spatially informative input is generated, whether through path integration or some other mechanism, is not described by these models, which instead focus on how this non-grid input can be combined and processed by mEC units to yield grid-like activity.

The model on which we are going to focus, initially described in Kropff and Treves (2008) and subsequently expanded in Si, Kropff, and Treves (2012), stems from the idea that place cells and grid cells representations may be formed on different time scales. The former appear to rapidly encode changes in the external environment through the fast recruitment of pools of hippocampal cells (S. Leutgeb, J. Leutgeb, Treves, et al. 2004). Such assemblies would be organized in multiple quasi-continuous attractors (Cerasti and Treves 2010) which might be rapidly acquired during exploration. These kind of representations is also characterized by the high degree of flexibility (S. Leutgeb, J. Leutgeb, Barnes, et al. 2005) and of representational capacity (J. Leutgeb, S. Leutgeb, M. Moser, et al. 2007) necessary to keep track of the extreme variability of the external world. mEC grid cell representations might instead be established more slowly, through a gradual process that shapes unit responses and, possibly, mutual connections, to generate a prototypical representation, rigid and invariant, that can then be applied on the go to any new environment the animal encounters.

3.2 A Model Based on Adaptation

This scenario is described in a model that shows how the grids may emerge spontaneously from disorder under the influence of firing rate adaptation, a very general property of cortical neurons. Such temporal modulation induces a self-organization process that, over time, sculpts the activity of single units to produce a triangular spatial pattern Kropff and Treves 2008.

The gradual process can be observed in simulations of a model network representing mEC, that receives spatially tuned inputs from the hippocampus or from parahippocampal cortices Kropff and Treves 2008. The combination of adaptation in mEC units and plasticity on the feed-forward connectivity is enough to generate increasingly regular grid patterns, over time scales corresponding to a couple of weeks of rodent development.

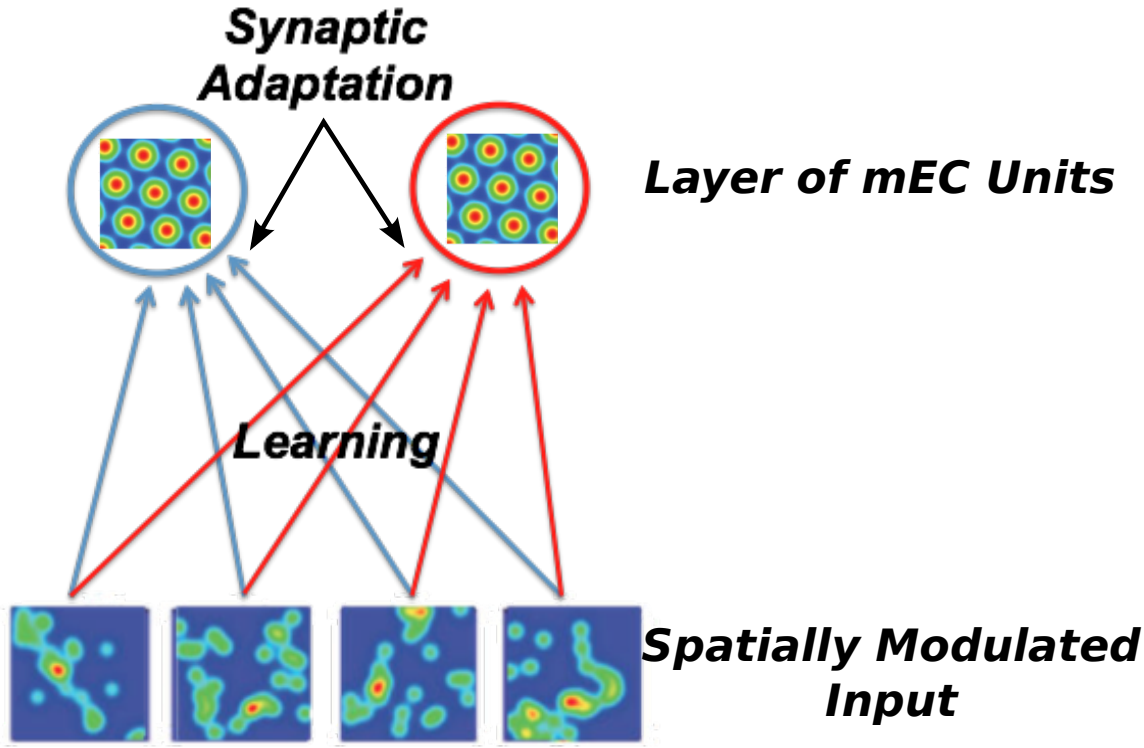


Figure 3.3: Schematic illustration of the adaptation model. An input layer provides mEC with spatially modulated input. Under the influence of synaptic adaptation, units in mEC self-organize, modifying their activity and the afferent connections strength. The resulting configuration of the activity for each unit is an hexagonal grid pattern. Adapted from Kropff and Treves (2008).

3.2.1 Analytical Formulation

The same process, at the single-unit level, can be described in analytical terms as an unsupervised optimization process, if one neglects the collateral interactions that are presumed to align the grids (Si, Kropff, and Treves 2012). The simplified version of the model which can be analyzed mathematically is very abstract, and does not specify most of the parameters necessary to the simulations. Nevertheless, it indicates which are the asymptotic states that should be reached by the system after having evolved for a certain time.

The asymptotic states are defined in terms of a variational principle amounting to the minimization of a cost function of the form:

$$H = H_K + H_A = \int d\chi [\bigtriangledown \Psi(\chi)]^2 + \gamma \int d\chi \int dt' \Psi(\chi(t)) K(t - t') \Psi(\chi(t')) \qquad (3.1)$$

where χ is the spatial coordinate and Ψ represents the firing rate of the neuron across the environment.
The functional is defined based on the hypothesis that the activity of the units reflects only two simple constraints:

1. The minimization of the variability of the maps across space, i.e a preference for

smooth maps. Such smoothness is expected to stem from the smoothness of the spatial inputs and of the neuronal transfer function. This constraint is expressed in the first term of the functional, the *kinetic* one.

2. The penalization of maps that require a neuron to fire for prolonged periods of time, despite the presence of neuronal fatigue. The second term of the functional, the adaptation term, represents this constraint.

The parameter γ parametrizes the relative importance of the two constraints.

In the second term, taking into account the averaging effect of a long run over many trajectories, we can substitute the time-dependent kernel $K(\Delta t)$ with an effective space-dependent one, $K(\chi' - \chi)$:

$$H = H_K + H_A = \frac{1}{A}\int_S d\chi[\bigtriangledown\Psi(\chi)]^2 + \gamma\frac{1}{A}\int_S d\chi\Psi(\chi)\int_S d\chi'\Psi(\chi')K(|\chi' - \chi|) \quad (3.2)$$

where we have also made explicit the normalization by the area A of the environment S.

In Kropff and Treves 2008 it is shown that the grid pattern exhibited by the mEC neurons minimizes this cost function, as it is the optimal compromise solution (among periodic and fully symmetric solutions) between the two opposite constraints, for a number of reasonable choices of the kernel K.

3.2.2 Full Model Implementation

The model can also be expanded in a full dynamical description of a neural network representing mEC units. For this, numerical simulations are required to virtually reproduce the navigation of an agent (mimicking a real lab rat) through an environment. During exploration, the position of the rat is linked to the activation of some neural populations in the network, whose evolution is controlled by a set of equations, taking into account adaptation and plasticity effects. These simulations can then be run for some time to reproduce hours or days of virtual exploration and of associated neural activity. In Kropff and Treves (2008) it is shown how during this time, units in the mEC layer of the model develop an hexagonal pattern of firing fields, in agreement with the analytical approach and with experiments.

It has to be reaffirmed that this model is completely self-organized: the feed-forward connections between the hippocampal layer and the mEC one are initialized to random values at the start of the simulations and it is the learning occurring during the rat exploration that generate the right associations between the two set of units by strengthening some of the connections and suppressing others. The model pays a price in terms of speed, as it is unable to express good and stable grids in an interval of seconds or minutes. It thus requires an extensive coverage of the environment to reach the desired outcome.

3.2.3 Development and Environment Geometry

The long time it takes for the grids to be generated, in the simulations, suggests that the process is not repeated multiple times in different environments, and it unfolds

instead over a significant portion of the rat life. As we already pointed out in the previous chapter 2.3, grid cells take some time to evolve in the newborn rat. Their appearance actually seems to come after that of place cells and to require a longer gestation. It is therefore reasonable to conjecture that adaptation might play its role over some days following the eye-opening. According to the most recent experimental data Langston et al. 2010; Wills, Cacucci, et al. 2010, the critical period for such slow process may extend between the 15^{th} and the 35^{th} postnatal day.

Since the phenomena described by this model are presumed to extend over a long developmental time, it is expected that the topological properties of the environment in which the animal develops affect the shape and the appearance of the grid maps (Barry, Hayman, et al. 2007; Savelli, Yoganarasimha, and J. Knierim 2008; Solstad et al. 2008; Derdikman and E. Moser 2010). Until now, models of grid cells have dealt only with planar, two-dimensional topologies. In the following two chapters we are going to show how to extend our single-cell adaptation model (Kropff and Treves 2008) to include the third dimension in the environment in order to study two non-planar topologies:

- the surface of a sphere (Stella et al. 2013a)
- the fully three-dimensional space. (Stella et al. 2013b)

The two cases will offer a test bench for the ability of the model to generate grids adjusted to the specific environment. They will also provide predictions on the grid structure that should be experimentally observed in similar conditions. New experiments that make use of virtual reality on a revolving sphere (Harvey et al. 2009), or in which rats are reared inside a spherical cage, will likely be a direct test for our model on the sphere, while experiments with flying bats will provide evidence on the feasibility of a genuine 3-dimensional representation of space in terms of grid units.

Chapter 4

Grids for Small Worlds

4.1 Life on a Sphere: a Mathematical Analysis

Here we discuss a formulation of the adaptation model on the sphere. This formulation is both experimentally relevant and mathematically treatable. While on the 2D plane the solutions to the minimization problem were formulated in terms of Fourier components, a similar approach can be applied on a sphere in terms of spherical harmonics (Courant and Hilbert 1962).

4.1.1 Solutions from Combinations of Spherical Harmonics

Using the following set of coordinates to parametrize the unitary sphere ($R = 1$)

$$\begin{aligned} x &= \sin(\theta)\cos(\varphi) \\ y &= \sin(\theta)\sin(\varphi) \\ z &= \cos(\theta) \end{aligned}$$

we define a function $\Psi(\theta, \varphi)$ whose values on the sphere represent the firing rate of a unit in the different locations.
We normalize this function such that:

$$\Psi(\theta, \varphi) \geqslant 0 \tag{4.1}$$

$$\int \Psi(\theta, \varphi) = 1 \tag{4.2}$$

corresponding to allowing only positive firing rates and to fixing the mean rate.
We start by considering the linear combinations of spherical harmonics that generate an ordered and symmetrical grid-like arrangement of firing fields. Among these, we then look for potential solutions to the minimization problem.
The number of fields obtained in the solution depends on the angular momentum (l) of the harmonics used in the linear combination. Solutions for l up to 6 are the following (Figure 4.1). The solution with a single field is obtained with $l = 1$ spherical harmonics:

$$\Psi_1(\theta, \varphi) = d[cY_0^0 + Y_1^0] = \frac{1}{4\pi} * (1 + \cos(\theta)) \tag{4.3}$$

$l = 2$ harmonics lead to a 2-field map:

$$\Psi_2(\theta, \varphi) = d[cY_0^0 + Y_2^0] = \frac{3}{4\pi}\cos^2(\theta) \tag{4.4}$$

$l = 3$ to 4 fields:

$$\begin{aligned}\Psi_3(\theta,\varphi) &= d\left[cY_0^0 + [aY_3^0 + b(Y_3^3 - Y_3^{-3})]\right] = \\ &= \frac{2\sqrt{2}}{\pi(16\sqrt{2}+3\pi)}\left[2 + \cos(\theta)(\frac{5}{2}\cos(2\theta) - \frac{1}{2}) + \sqrt{2}\sin^3(\theta)(2\cos(2\varphi)+1)\right]\end{aligned} \tag{4.5}$$

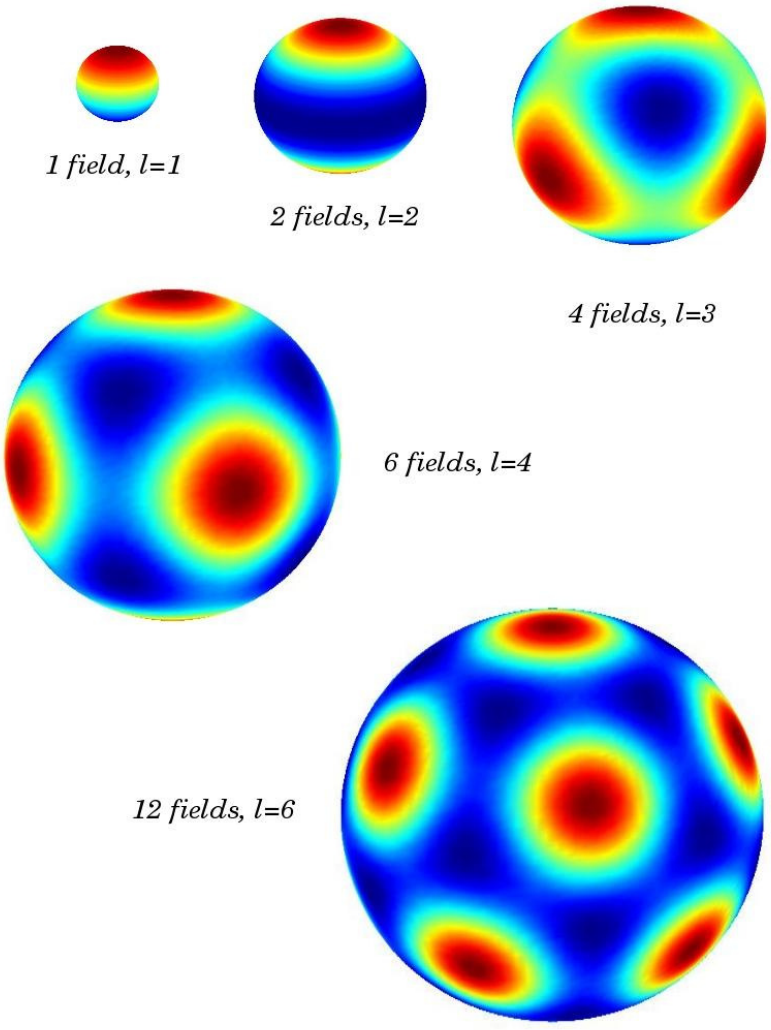


Figure 4.1: The analytical solutions considered. Larger spheres lead to more fields, given an approximately constant spacing.

$l = 4$ to 6 fields:

$$\begin{aligned}\Psi_4(\theta,\varphi) = &\quad d\left[cY_0^0 + [aY_4^0 + b(Y_4^4 + Y_4^{-4})]\right] \\ = &\quad \tfrac{1}{8\pi}[2 + 3(\cos^4(\theta) + \sin^4(\theta)\cos^4(\varphi) + \sin^4(\theta)\sin^4(\varphi)) - \\ &\quad -9(\cos^2(\theta)\sin^2(\theta) + \sin^4(\theta)\cos^2(\varphi)\sin^2(\varphi))]\end{aligned} \tag{4.6}$$

$l = 6$ to 12 fields, i.e., a soccer ball:

$$\begin{aligned}\Psi_6(\theta,\varphi) = & \quad d[cY_0^0 + [aY_6^0 + \tfrac{b}{2}(Y_6^{-5} - Y_6^5)]] \\ = & \quad d[c\tfrac{1}{2}\sqrt{\tfrac{1}{\pi}}+ \\ & +[\sqrt{\tfrac{143}{137\pi}}\tfrac{1}{32} * (231\cos^6(\theta) - 315\cos^4(\theta) + 105\cos^2(\theta) - 5)+ \\ & +\tfrac{143}{137\pi}\tfrac{21}{6} * \cos(5\varphi)\sin^5(\theta)\cos(\theta)]] \end{aligned} \tag{4.7}$$

where d and c are used to satisfy conditions (4.1) and (4.2) while a and b values are chosen in each case so that the solutions are fully symmetric.

4.1.2 Cost Function Evaluation

We now want to calculate the value of the 'energy', i.e., the cost function associated with these solutions. Consider our cost function H

$$H = H_K + H_A = -\frac{1}{A}\int_S d\chi[\bigtriangledown\Psi(\chi)]^2 + \gamma\frac{1}{A}\int_S d\chi\Psi(\chi)\int_S d\chi'\Psi(\chi')K(|\chi' - \chi|). \tag{4.8}$$

To evaluate this function for every l, we expand our solutions on the spherical harmonics:

$$\Psi_{l^\star}(r) = \sum_{l=0,l^\star}\sum_{m=0,m^\star} a_{lm}Y_l^m(r). \tag{4.9}$$

Then, using the spherical harmonics defining property

$$R^2 \bigtriangledown^2 Y_l^m(r) = -l(l+1)Y_l^m \tag{4.10}$$

we can write the kinetic term L_1 for a given $l^\star$:

$$H_K = \frac{1}{4\pi}\frac{1}{R^2}l^\star(l^\star+1)(a_{l^\star,0}^2 + 2a_{l^\star,m^\star}^2) \tag{4.11}$$

To calculate the adaptation term H_A we project the adaptation kernel on a basis of Lagrange polynomials. Suppose we consider an exponential kernel

$$K(\vec{x}\cdot\vec{y}) = \frac{1}{v\tau_L}e^{-\frac{q}{v\tau_L}} - \rho * \frac{1}{v\tau_S}e^{-\frac{q}{v\tau_S}} \tag{4.12}$$

which expresses the fact that adaptation effects become significant over time-differences τ_S and decay away after time differences τ_L. It means that, after a neuron emitted a spike in a certain position, the spatial region of strong adaptation is comprised in a ring between radii $v\tau_S$ and $v\tau_L$ around that position.
In our case q is the distance between two vectors $\vec{x}, \vec{y}$ having coordinates (θ,φ) and (θ',φ') on the sphere. We write the kernel in terms of the angle between these two vectors:

$$K(\vec{x}\cdot\vec{y}) = \frac{1}{v\tau_L}e^{-\frac{1}{v\tau_L}*R*\arccos(\vec{x}\cdot\vec{y})} - \rho * \frac{1}{v\tau_S}e^{-\frac{1}{v\tau_S}*R*\arccos(\vec{x}\cdot\vec{y})} \tag{4.13}$$

The kernel is thus a function on the [-1,1] domain and can therefore be projected on a set of Legendre Polynomials which are complete on that range (Abramowitz and Stegun 1965).

$$K(q) = \sum_l k_l P_l(q) \tag{4.14}$$

H_A is thus:

$$H_A = \frac{1}{4\pi} \sum_{l=0,l^\star} \sum_{m=0,m^\star} a_{lm} \sum_{l'=0,l^\star} \sum_{m'=0,m^\star} a_{l'm'} \int d\Omega Y_l^m(r) \int d\Omega' Y_{l'}^{m'}(r') K(\vec{x} \cdot \vec{y})$$
$$= \frac{1}{4\pi} \sum_{l=0,l^\star} \sum_{m=0,m^\star} \sum_{l'=0,l^\star} \sum_{m'=0,m^\star} a_{lm} a_{l'm'} \int d\Omega Y_l^m \int d\Omega' Y_{l'}^{m'} \sum_{l''} k_{l''} P_{l''}(\vec{x} \cdot \vec{y})$$

Using the addition theorem (Unsold 1927)

$$P_l(\vec{x} \cdot \vec{y}) = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} (-1)^m Y_l^m(\vec{x}) Y_l^{-m}(\vec{y}) \tag{4.15}$$

the previous expression becomes:

$$H_A = \frac{1}{4\pi} \sum_{l,l'=0,l^\star} \sum_{m,m'=0,m^\star} a_{lm} a_{l'm'} \sum_{l''} k_{l''} \frac{4\pi}{2l''+1} \sum_{\mu=-l''}^{l''} (-1)^\mu \tag{4.16}$$
$$\int d\Omega Y_l^m(\vec{x}) Y_{l''}^{\mu}(\vec{x}) \int d\Omega` Y_{l`}^{m'}(\vec{y}) Y_{l''}^{-\mu}(\vec{y}).$$

Given the orthonormality rules for spherical harmonics:

$$\int d\Omega Y_l^m(\vec{x}) Y_{l'}^{m'}(\vec{x})^* = \delta_{l,l'} \delta_{m,m'} \tag{4.17}$$

and that $l, l' = 0, l^\star$ and $m, m' = 0, m^\star$, the surviving combinations in the sum are

$$H_A = \frac{1}{2l^\star+1} k_{l^\star} (a_{l^\star,0}^2 + 2a_{l^\star,m^\star}^2) + k_0(a_0^2). \tag{4.18}$$

The last term is identical for all solutions, as k_0 and a_0 are always the same.

The total energy is, therefore:

$$H = (a_{l^\star,0}^2 + 2a_{l^\star,m^\star}^2) * (\frac{1}{4\pi} \frac{1}{R^2} l^\star(l^\star+1) + \gamma \frac{1}{2l^\star+1} k_{l^\star}) + \gamma k_0(a_0^2) \tag{4.19}$$

In the region of parameters where $(\frac{1}{4\pi} \frac{1}{R^2} l^\star(l^\star+1) + \gamma \frac{1}{2l^\star+1} k_{l^\star}) > 0$ any solution is worse than the constant solution with all the $a_{lm} = 0$.
Since the first term in $(\frac{1}{4\pi} \frac{1}{R^2} l^\star(l^\star+1) + \gamma \frac{1}{2l^\star+1} k_{l^\star})$ is always positive, for the sum to be negative it has to be at the very least that $k_l < 0$. As shown in Figure 4.2, the different components of the kernel projection k_l have a qualitatively similar dependence on the radius R, but the range in which they take negative values is progressively shifted towards larger radii. This result does not depend, at least qualitatively, on the particular choice of the kernel. Choosing a kernel consisting of radially symmetric Gaussians

$$K(\vec{x} \cdot \vec{y}) = \frac{1}{v\tau_L\sqrt{2\pi}} e^{-\frac{1}{2(v\tau_L)^2} * (R * \arccos(\vec{x} \cdot \vec{y}))^2} - \rho * \frac{1}{v\tau_S\sqrt{2\pi}} e^{-\frac{1}{2(v\tau_S)^2} * (R * \arccos(\vec{x} \cdot \vec{y}))^2} \tag{4.20}$$

produces the same behavior, Figure 4.3. The choice of the adaptation time scale (τ_L,τ_S) determines the spatial scale of these transitions. Slower adaptation dynamics means that all curves are shifted towards larger values of R. In Figures **??** curves are plotted with a fixed ratio $\tau_L/\tau_S = 3$ and R is expressed in units of $v\tau_L$, so that higher values of the time constants correspond to higher absolute values for R.

Conclusion This allows us to conclude that the solutions associated with increasing values of the angular momentum become energetically favored one after the other as the size of the sphere increases. The precise radius at which the switch occurs between two solutions depends also on the values of the components a_{lm}, on the particular value of γ and on the choice of the kernel, among other factors, but the point of transition from negative to positive values, of the corresponding component of the kernel, provides an upper bound for the validity of the solution. Therefore, the different solutions are not clearly separated and it is well possible that the energy minima are quite shallow. Transitions from one solution to the next may thus be gradual, and allow for intermediate spurious solutions consisting of a mix of different locally minimal configurations. These questions cannot be analytically addressed and are thus left for numerical simulations.

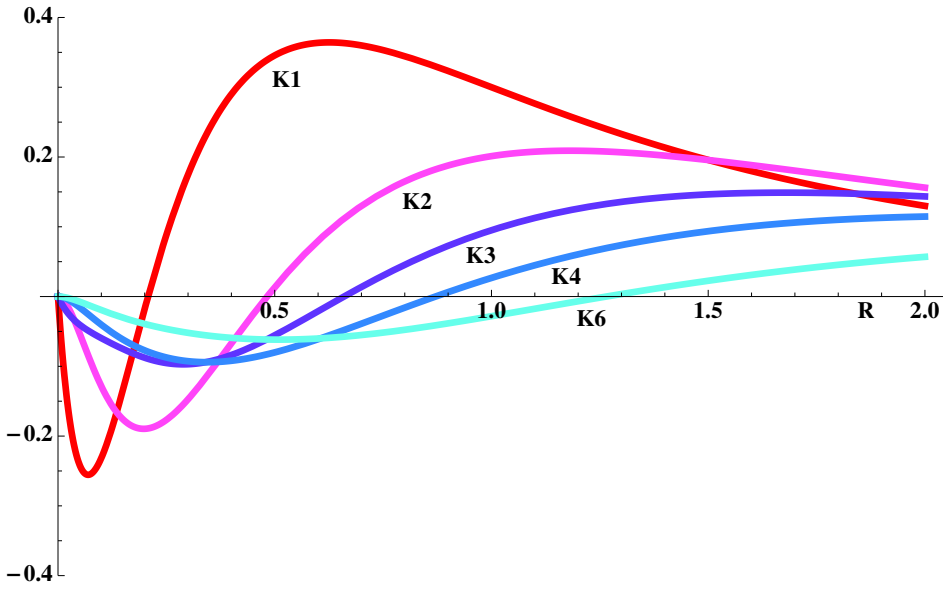


Figure 4.2: Legendre projections k_l of the Exponential kernel

4.1.3 The Possible Appearance of Other Solutions

These calculations are not conclusive about the effective optimality of the combinations of spherical harmonics we have defined. Once we set a certain radius we know that solutions including spherical harmonics with a certain angular momentum are favored. The particular combination that minimizes the cost function is determined by the values of the a_{lm} associated to that solution by the constraints in Eq. 4.2. What about other periodical but non-symmetrical solutions?

We have used a numerical approach to explore the existence of other solutions representing global minima for our functional. We have looked for any combination of the $l+1$ real spherical harmonics associated with each eigenvalue l. The optimal solution for a given l was defined as maximizing the energy function

$$H_{Num} = a_{l,0}^2 + 2 * \sum_{m=1}^{l} a_{l,m}^2 - \lambda \int d\chi \Theta(-\Psi_{a_{l,m}}(\chi)) \tag{4.21}$$

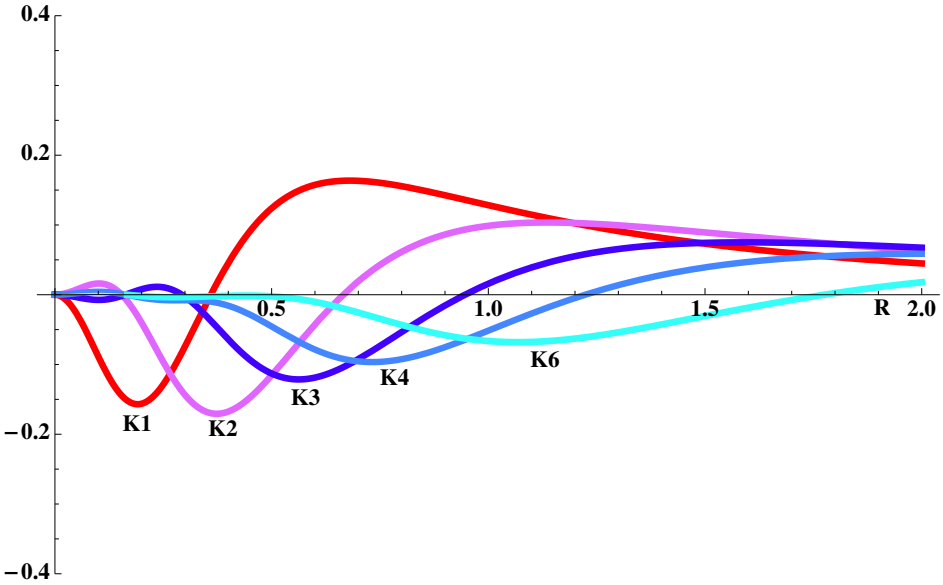
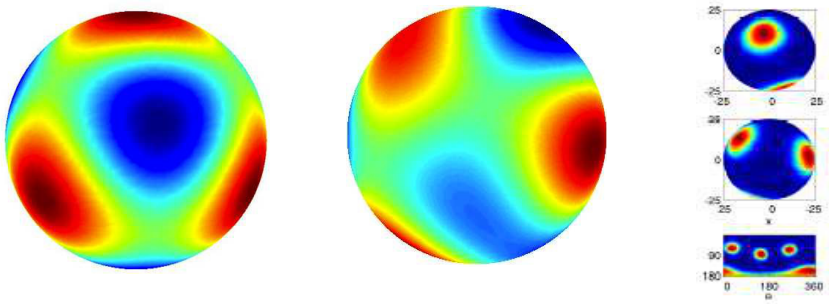


Figure 4.3: Legendre projections k_l of the Gaussian kernel

where $\Psi_{a_{l,m}}$ is the solution corresponding to a particular set of coefficients, $\lambda >> 1$, and the last term is used to enforce constraint (4.1).

With this approach, we find that for every l the solutions reached by the numerical optimization procedure are equivalent, or strongly resembling, our symmetric ansatz, as shown in the example of Fig. 4.4.

We limit this approach to the single l case for computational reasons. Not only the parameter space would significantly grow when any combination of different l is allowed, but also the energy minima become dependent on the particular choice of the adaptation kernel, making a comprehensive approach impossible. This situation is similar to the one described for the planar case in Kropff and Treves (2008), where a single wave vector length k is selected to produce the solutions.



Symmetrical solution *Numerical Optimization* *Network Simulation*

Figure 4.4: Comparison of the symmetrical solution for $l = 3$ (*left*) with the result of the numerical minimization of the energy functional (*center*) and with the asymptotical state of network simulations (*right*, see next section).

4.2 Numerical Simulations of Grid Development

The solutions we have studied in the previous section should correspond to the asymptotic behavior of a more detailed model, which explicitly expresses the effects of adaptation on neural dynamics. We now proceed to study the more detailed model with numerical simulations. If our analytical approach is robust, we expect the firing maps generated by our simulated model units to correspond to the combinations of spherical harmonics we have previously described.

4.2.1 Network Model

In our simulations, time is discretized in steps of length $\Delta t = 0.01s$. The virtual rat moves on the surface of spheres of different radii with a constant speed of $v = 0.4m/s$. For simplicity, the change in running direction between two consecutive steps of the virtual rat is sampled from a Gaussian distribution with zero mean and standard deviation $\sigma_h = 0.15$ radians. The virtual rat always runs along the great circle determined by its running direction. Our model comprises two layers, the input network represents e.g. the CA1 region of hippocampus and contains $N_{inp} = 4\pi R^2 \rho$ units, where R is the radius of the sphere and $\rho = 8000/m^2$ is the density of input units. The output network represents a population of $N_{mEC} = 100$ would-be grid units in mEC, all with the same adaptation parameters.
Each mEC unit receives afferent spatial inputs which, as already discussed in Kropff and Treves (2008), we take for simplicity to arise from regularly arranged place cells, although they could also arise from spatially modulated units in the adjacent cortices. The input to unit i at time t is then given by h_i^t:

$$h_t^i = \sum_j W_{ij}^{t-1} r_j^t. \tag{4.22}$$

The weight W_{ij}^t connects input unit j to mEC unit i. We will assume that at the time the mEC units develop their firing maps, spatially modulated or place cell-like activity is already present, either in parahippocampal cortex or in the hippocampus. The network model works in the same way with any kind of spatially modulated input, but the place-cell assumption reduces the time necessary for learning. Moreover, as recent studies have shown (Langston et al. 2010; Wills, Cacucci, et al. 2010), it is entirely plausible that place cells develop adult-like spatial code earlier than grid cells do. We thus model the place field as an exponential bump centered in the place cell preferred position $\vec{x}_{j0}$

$$r_j^t = \exp(-\frac{\|\vec{x}^t - \vec{x}_{j0}\|^2}{2\sigma_p^2}), \tag{4.23}$$

where $\vec{x}^t$ is the position at time t of the simulated rat, $\sigma_p = 0.05m$ is the width of the firing field and $\|a - b\|$ is the great-circle distance between two points a and b on a sphere. The firing rate Ψ_i^t of mEC unit i is determined by a non-linear transfer function

$$\Psi_i^t = \frac{\pi}{2} \arctan\left[g^t(\alpha_i^t - \mu^t)\right] \Theta(\alpha_i^t - \mu^t), \tag{4.24}$$

which is normalized to have maximal firing rate equal to 1 (in arbitrary units) and $\Theta(\cdot)$ is the Heaviside function.

The variable α_i^t represents the adaptation-mediated input to the unit i. It is related to the h_i^t as follow:

$$\begin{aligned} \alpha_i^t &= \alpha_i^{t-1} + b_1(h_i^{t-1} - \beta_i^{t-1} - \alpha_i^{t-1}) \\ \beta_i^t &= \beta_i^{t-1} + b_2(h_i^{t-1} - \beta_i^{t-1}), \end{aligned} \tag{4.25}$$

where β_i has slower dynamics than α_i, with $b_2 = b_1/3, b_1 = 0.1$. This adaptive threshold makes it more difficult for a neuron to fire for prolonged periods of time, and corresponds to the kernel K considered in the analytical treatment. The gain g^t and threshold μ^t are iteratively adjusted by Eqs.4.26 at every time step to fix the mean activity $a = \sum_i \Psi_i^t / N_{mEC}$ and the sparsity $s = (\sum_i \Psi_i^t)^2/(N_{mEC}\sum_i \Psi_i^{t2})$ within a 10% relative error bound from pre-specified values, $a_0 = 0.1$ and $s_0 = 0.3$ respectively. If k is indexing the iteration process:

$$\begin{aligned} \mu^{t,k+1} &= \mu^{t,k} + b_3(a^k - a_0) \\ g^{t,k+1} &= g^{t,k} + b_4 g^{t,k}(s^k - s_0). \end{aligned} \tag{4.26}$$

$b_3 = 0.01$ and $b_4 = 0.1$ are the sizes of the iteration steps. a^k and s^k are the values of mean activity and sparsity determined by $\mu^{t,k}$ and $g^{t,k}$ in the intermediate iteration steps. The iteration stops once the gain and threshold have been brought within the 10% error range, and the activity of mEC units are determined by the final values of the gain and threshold in Eq.4.24.

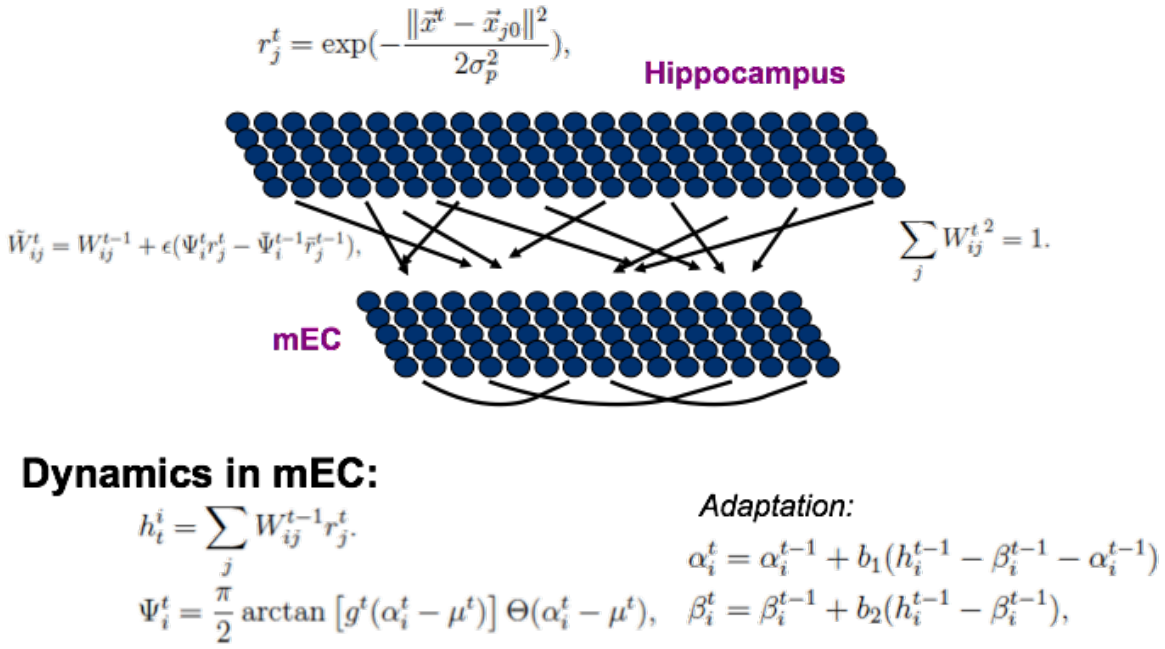


Figure 4.5: The network used for the numerical simulations in this and the following chapter.

The learning process modifies the strength of the feed-forward connections according to an Hebbian rule

$$\tilde{W}_{ij}^t = W_{ij}^{t-1} + \epsilon(\Psi_i^t r_j^t - \bar{\Psi}_i^{t-1}\bar{r}_j^{t-1}), \tag{4.27}$$

with a rate $\epsilon = 0.002$. $\bar{\Psi}_i^t$ and $\bar{r}_j^t$ are estimated mean firing rates of mEC unit i and place unit j that are adjusted at each time step of the simulation

$$\begin{aligned} \bar{\Psi}_i^t &= \bar{\Psi}_i^{t-1} + \eta(\Psi_i^t - \bar{\Psi}_i^{t-1}), \\ \bar{r}_j^t &= \bar{r}_j^{t-1} + \eta(r_j^t - \bar{r}_j^{t-1}), \end{aligned} \tag{4.28}$$

with $\eta = 0.05$ being a time averaging factor.
After each learning step, the provisional $\tilde{W}_{ij}^t$ weights are normalized into unitary norm

$$\sum_j {W_{ij}^t}^2 = 1. \tag{4.29}$$

Units that win during competitive learning manage to establish strong connections with units that provide strong inputs. As the learning proceeds the units establish fields where they receive strong inputs and at the same time are recovering from adaptation. The emergence of the grid map is the product of averaging over a period of time corresponding to some hours of run for the animal. It remains to be assessed whether the mechanism we propose might also account for the formation of novel grid representations in each new environment the animal encounters, or if, instead, it can only be applied to the developmental period.

4.2.2 Results

We have run several simulations of 3×10^7 learning time steps, i.e. 83 hours of animal time, considerably longer than needed to produce clear grids in a flat 2D environment. The fields developed by model mEC units are then expected to be in the asymptotically stable regime, and they are depicted, for different values of the radius of the sphere, in Fig.4.6. The figure presents the firing activity of a random sample of mEC units. The units develop a grid-like pattern of activation and the emerging configurations are in reasonable agreement with those predicted by the analytical model. For each value of the radius, the units typically express the same number of fields. As the radius increases, the units express an increasing number of grid fields, regularly tiling the surface. As already discussed for the analytical approach, the particular association between radius and number of fields is the result of the choice of the adaptation time scale. A different value of $b1$ (and $b2$) will change the configuration of fields, but their appearance will again be the same if the radius of the sphere is rescaled accordingly. In the case of a population of cells with heterogeneous adaptation properties, we expect the coexistence of different solutions on the same sphere.

The particular set of adaptation parameters we used would produce a spacing of about $56cm$ in the planar case. We notice that this spacing is approximately maintained on the sphere, when its radius its varied. As a consequence of this constraint, the number of fields expressed by the cells grows roughly quadratically with the radius of the sphere.

Generally the configurations correspond to those indicated as asymptotic solutions (1, 2, 4, 6, 12 fields), but in the case of the $40cm$ radius, some units exhibit a non symmetrical configuration of 8 to 10 fields, Figure 4.7, that emerges at the transition, as it were, from the 6-field to the 12-field solution. The existence and the stability of such a configuration is most probably the expression of the very shallow nature of the energy profile. As previously noted, we cannot exclude the presence of other

asymptotically stable solutions. The results of the simulations seem to point in this direction.

4.3 Conclusions

In this chapter we presented the study of grid development in the specific case of a spherical environment. Using an analytical approach based on the minimization of an energy function, together with more detailed numerical simulations of the system (though still a network of non interacting units), we study the configurations of activity that develop after a long rearing period. We find that these configurations are described by linear combinations of spherical harmonics, that express a grid-like arrangement of the fields on the surface of the sphere. By combining harmonics of increasing angular momentum one obtains solutions with an increasing number of fields (from 1 to 12) and an increasing coordination number (from 1 to 5). The radius of the sphere determines which of these is energetically favored, and thus prevails in the asymptotic configuration of the activity, as shown also by simulations.

One problem that remains open is the existence of intermediate, stable states present for certain radii, at the transition between one fully symmetrical solution and the next. Their appearance may reflect the shallow nature of the energy profile described by our model, or the inability of our approach to successfully identify all the minima of the energy function defining the asymptotic states of the system.

In the model the grid pattern is a product of a self-organization process driven by neural adaptation, with additional recurrent interactions to align the grids, which can however be regarded as a second order correction (Si, Kropff, and Treves 2012). This mechanism does not rely on any internal hard-wired structure to work. The final arrangement of the grid fields is a spontaneous effect of the protracted exploration of the environment by the animal. The model thus differs from those based on a pre-wired attractor structure (Fuhs and Touretzky 2006; Burak and Fiete 2009), which rely on a specific connectivity established *a priori* to produce the grid pattern of activity. It is not clear to us what such models would predict in the case of rodents raised in or on a sphere, or in fact in any other environment not matching the flat 2D geometry of the pre-wired connectivity (it is likewise not clear what such pre-wired attractor models would predict in the case of bats, flying in 3D). Unless attractor models can be shown to also lead to symmetric combinations of spherical harmonics, or something similar, experiments with a spherical environments might help contrast different models of grid field formation. In parallel, the observation of spherical grid fields with $l \leq 6$ would immediately lead to refute models of grid-cell formation based on the combination of striped fields at 60 or 120 degrees of each other (Burgess, Barry, and O'Keefe 2007; M. Hasselmo, Giocomo, and Zilli 2007).

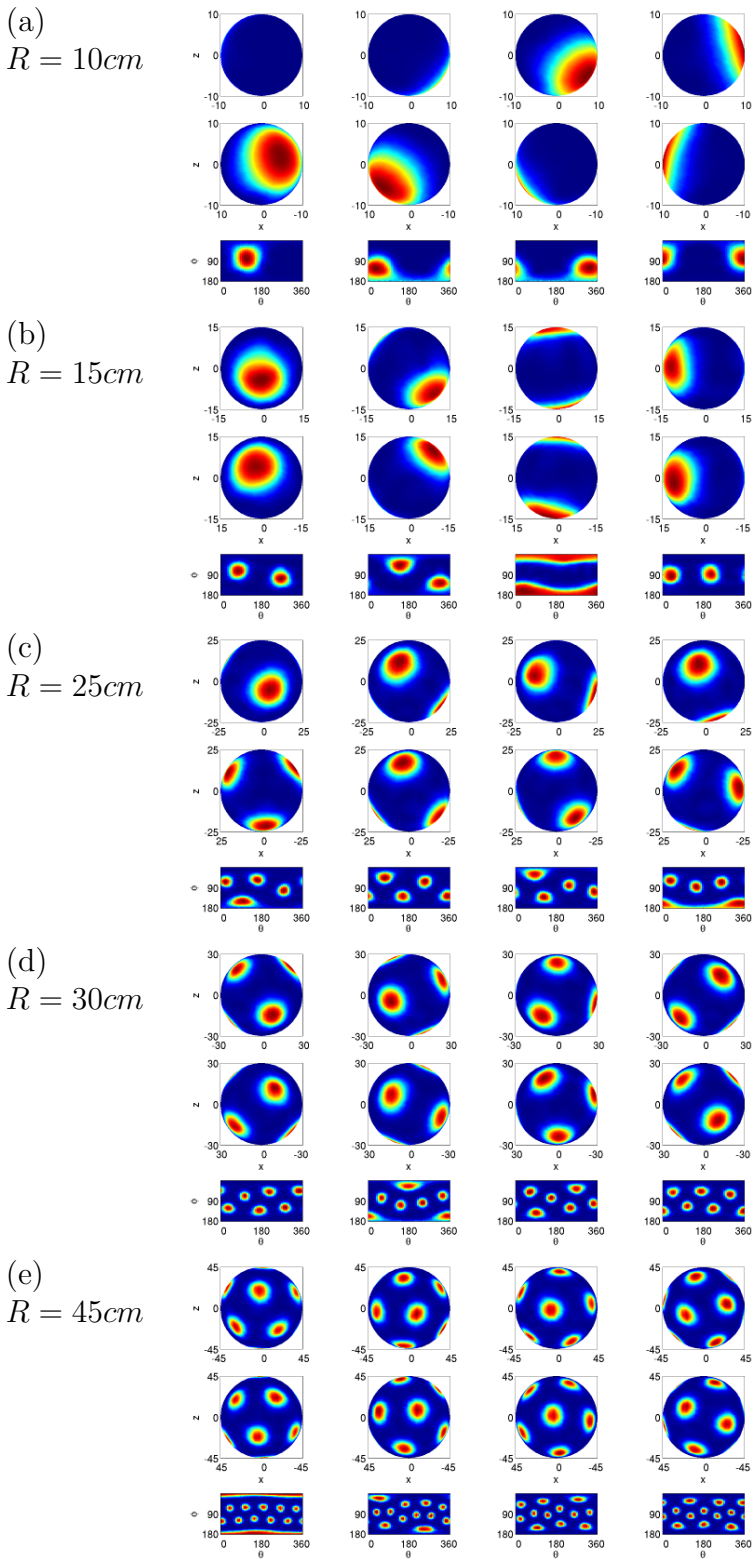


Figure 4.6: Units in an adapting network form grid-like firing maps on spheres. The number of fields increases from 1, 2, 4, 6 to 12 when the radius of the sphere R increases. In each panel, the first two rows show 4 examples of firing maps from two opposite views perpendicular to the x-z plain. The bottom rows of each panel are the Miller cylindrical projections of the corresponding maps.

$R = 40cm$

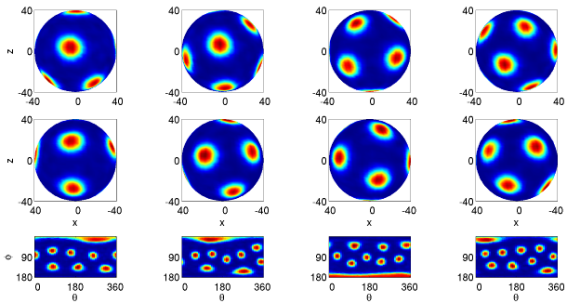


Figure 4.7: Units develop intermediate numbers of fields if the size of the sphere does not match the adaptation parameters of the network. The radius here is 40 cm, and the 4 units in the sample shown develop from 8 to 10 fields. Representation as in Fig.4.6

Chapter 5

Learning to Fly

5.1 Some Remarks on 3D Navigation

Although we naturally think of spatial navigation as a two-dimensional problem, most of the times navigation actually takes place in three dimensions. Not only our surrounding environment is far from being flat, but we should also consider the perspective of those beings that, rather than being confined to a surface in their movements, can fly and freely explore volumetric spaces (Healy, Kort, and Clayton 2005; Tsoar et al. 2011; Wu and Dickman 2012).

The study of spatial cognition in three dimensions has lagged behind the more extensive studies of two dimensional navigation, carried out on non-flying animals like rodents (Tolman 1948). In particular, if at least some description of the behavior of flying animals has accumulated over the years, until very recently there was no available data about neural activity related to spatial navigation in these animals (Jeffery et al. 2013). In fact, after 40 years of extensive research, the neural circuitry underlying planar navigation has in many cases been identified, and it now appears that the building blocks of navigation have been reasonably well identified. Although between the discovery of these structures and the construction of a causal model of spatial navigation and memory is a large gap, still neural recordings offer the most powerful tool at our disposal, to open a window into the mechanism underlying the integration of sensory inputs and the building of a cognitive map in these animals.

5.1.1 Neural Recordings in 3D

The neural populations distributed in the medial temporal lobe, like Place, Grid, Head Direction and Border Cells, whose contribution is thought to be essential for navigation, have been described and characterized mostly in conditions in which the third dimension was absent. Therefore, very little is known about how the firing properties of these units would extend to the case of a three-dimensional environment (J. Knierim, B. McNaughton, and Poe 2000; J. Knierim and B. McNaughton 2001; Calton and Taube 2005; Gibson, Butler, and Taube 2013).

3D Spatial representations in bats Recent findings in bats (Yartsev and Ulanovsky 2013) show that at least in the case of place cells, the representation of the volume of a room generated in the hippocampus by a flying bat is a straightforward generalization of its 2-D counterpart. Place cells recorded in the bat hippocampus are selective for

roughly spherical portions of the room. Different parts of the volume are evenly represented and there is no evidence for a bias in the number of place fields or in their shape along the vertical direction. Therefore in the case of bats, place cells seem to naturally extend their selectivity in the presence of additional dimensions, and to retain most of the properties observed in two-dimensional environments. Also, unpublished data about the bat subicular head directional system show a similar behavior, indicating the presence of units with mixed selectivity for all the three Euler angles (and combinations of them) that completely describe heading in 3-D. Grid cells firing during flight have not been explored yet, although they are no markedly different from rat ones when the animal is crawling (Yartsev, Witter, and Ulanovsky 2011).

What kind of periodical activity grid cells would develop in 3-D? Another set of experiments performed on rats, however, reached very different conclusions. Hayman, Verriotis, et al. (2011) studied the neural representations activated by a rat exploring a set of experimental set-up artfully designed to allow the animal to climb along the vertical direction. In these conditions they found a clear anisotropy in the geometry of the representation. Place and grid cell firing fields had normal horizontal characteristics but were elongated vertically, with grid fields forming stripes. Although a direct comparison between the results of this study and the one on bats is still not possible (due to the lack of grid cell data in the latter), nevertheless it suggests a possible profound discrepancy between the two animal models.

Making predictions on the appearance of grid cells in 3-D based on existing models is not a trivial exercise (Jeffery et al. 2013). Popular models for grid cells generation (**??**), namely those based on path-integration, either the continuous attractor variant or the interference variant, are all based on the 2D planar case and fine-tuned to this condition, in order to obtain, as the final outcome of the model network activity, the hexagonal tiling of space observed in experiments. Of course this perspective is induced by the fact that, indeed, experimental data has been obtained only in the 2D planar case and that no other example of tiling of different spatial topologies is currently available. Nevertheless, by restricting oneself to the Euclidean case and by forcing the model to only produce hexagonal grids one is limiting the perspective not only on the possible phenomenology of grid cells activity in other conditions, but perhaps also on the theoretical account of grid cells function. To actually adapt these models to cover also the 3-D case would presumptively require the introduction of ad-hoc additional features that seem ad odds with the alleged role of grid cells in spatial representations (to provide a regular and invariant reference frame to compute distances between locations in the environment), and with the generality of the phenomena that this role should imply.

5.1.2 Addressing 3D Representations with the Adaptation Model

A different outlook on the grid cells formation in 3D might come from the models based on self-organization. In these models the triangular pattern emerges spontaneously at the single unit level and is a product of firing rate adaptation. The model is based on two relevant assumptions: first, the regular tiling of space is a consequence of the isotropic exploration of that space. Second, the self-organization process takes an amount of time that makes it feasible to describe the formation of grid cells dur-

ing post-natal development Within a properly set experimental paradigm, bats indeed explore the environment quite uniformly and moreover they experience flight during development. Therefore bats behavior fall adequately into these assumptions. The same cannot be said about rats which can move through vertical space only in artificial conditions and are raised in flat cages with little or no experience of vertical movement.

The model therefore is not only aimed at predicting the spatial periodicity that would be observed in the grid cells of flying bats, but also, perhaps, at providing the base to link different behaviors to different expression of spatial representations in the brain.

5.2 Sphere Packing

What should we expect to observe in flying animals in the light of this model? As we have already shown in the two dimensional cases of the plane and the sphere, the conjunctive action of adaptation and competitive learning produces an arrangement of fields which is fundamentally a solution to the problem of packing spheres in the corresponding environment (Cohn and Kumar 2007; Petkovic 2009). Therefore the same principle should apply when an additional dimension is added. As for the two-dimensional case, the packing of spheres in a three dimensional volume (also known as *cannonball problem*) is a well studied problem(Conway 1992). The calculation of the maximal packing density, that is, the fraction of space actually occupied by the spheres, goes back to Gauss, who calculated its value: $\frac{\pi}{3\sqrt{2}} \simeq 0.74$. The lattice structure that achieves this limit is not unique. Any of these solutions is based on layers of spheres arranged in a hexagonal pattern (like the one of planar grids) stacked one upon the other (Fig 5.1). Given one of these layers, and taking it as a reference with positioning A, there are two possibles arrangements (B and C) of the next layer, obtained with a translation of A, that puts all the spheres at the same distance from their neighbors. Any sequence of A, B and C without immediate repetitions has the same, maximum, packing score. Any of these sequences can be generated from two regular prototypes:

- Face Centered Cubic (**FCC**) = ABCABCABCA
- Hexagonal Close Packed (**HCP**) = ABABABABAB

In both combinations each sphere has 12 first neighbors, and if d is the diameter of a sphere (or the distance between the centers of two neighboring spheres), then the inter-layer separation is $\frac{\sqrt{6}}{3}d$. If we consider the unit cell of 13 spheres (a central sphere + 12 neighbors) (figura), then FCC and HCP differ only for the position of 3 spheres. In fact, while in FCC neighbor spheres are arranged in 6 pairs with symmetrical positions with respect to the center, in HCP there are only 3 of these pairs, those on the central plane.

From the perspective of tiling theory, then, 3D space admits a multitude of equally optimal orderings of spheres, unlike the plane, where the hexagonal tiling is always best, and the spherical surface, where a unique solution is defined given the size of the sphere itself. Even restricting the choice to the two regular solutions, the fact that the two configurations are undistinguishable in terms of regularity, coordination number and of packing density, makes them both optimal solutions for the tiling of three dimensional space and therefore equally plausible candidates for the final outcome of our model arrangement of fields.

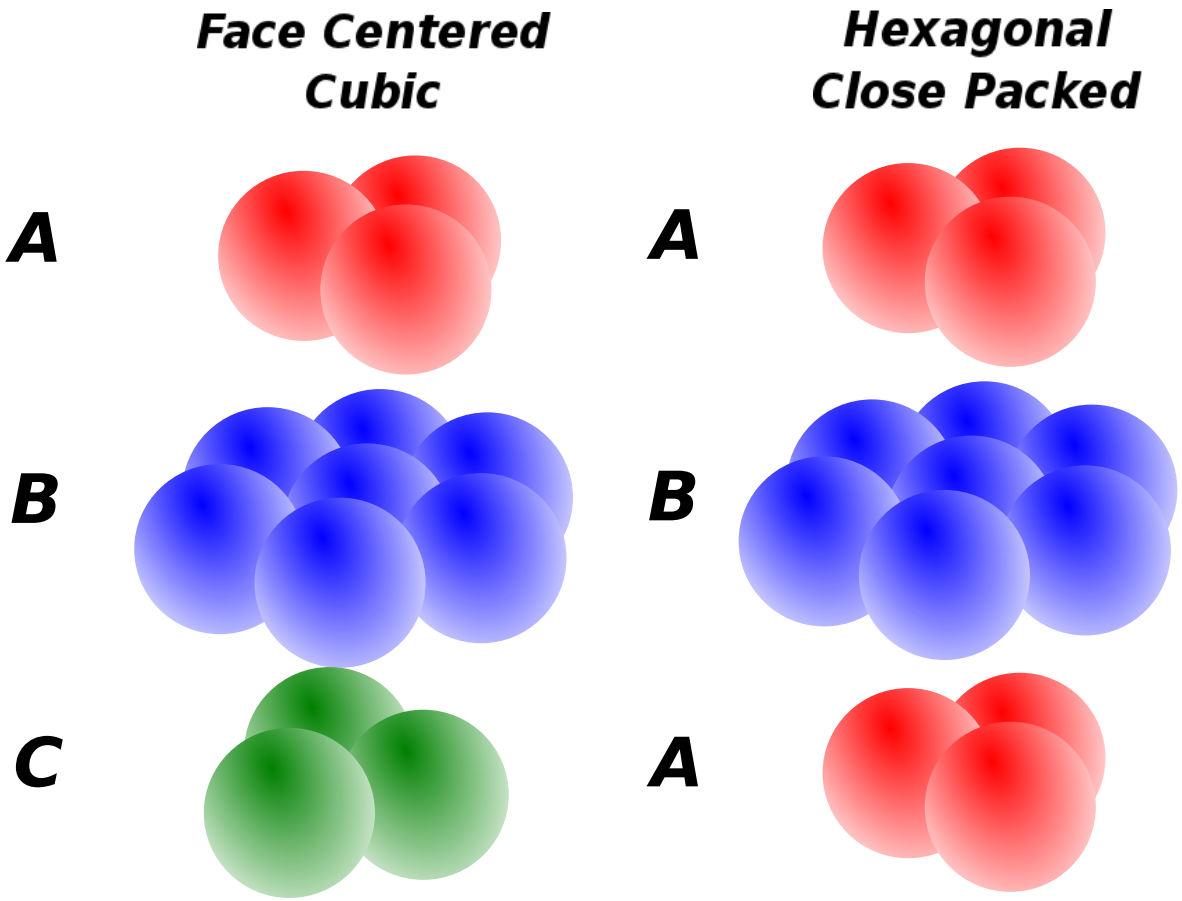


Figure 5.1: Difference of the layers sequence in FCC (ABCABC) and HCP (ABABAB)

As we have already shown, the output of the self-organization process implemented in our model network can be predicted based on a simplified analytical approach 4.1. It is well possible, although unlikely, that FCC and HCP, put on the same footing by these preliminary observations, lead to very different values for our cost function when used in the formula 3.1.

5.3 3D Close Packing

As we did in the previous chapter, we ask which is the favorite arrangement of the fields in the 3D space in terms of the minimization of the functional:

$$L = L_1 + L_2 = \frac{1}{V}\int_V dx\,[\bigtriangledown\psi(x)]^2 + \frac{\gamma}{V}\int_V dx\psi(x)\int_V dx'\psi(x')K(|x'-x|) \tag{5.1}$$

We thus define two analytical expressions for the Face Centered Cubic and the Hexagonal Close Packed configurations. We look for combinations of plane waves as they are well suited to be treated in this formulation of the problem.

5.3.1 FCC

To represent the Face Centered Cubic arrangement of fields we use the following expression:

$$\psi^{FCC}(r) = \frac{8}{9}(1 + \frac{1}{4}\sum_{i=1}^{4} cos(k_i \cdot r))^2 \tag{5.2}$$

a combination of four plane waves (Figure 5.3, top left), with the four wave vectors k_i components given by the matrix:

$$k_i = \frac{2\pi}{a}\begin{pmatrix} 0 & 0 & \sqrt{3/2} \\ 2/\sqrt{3} & 0 & -1/\sqrt{6} \\ -1/\sqrt{3} & 1 & -1/\sqrt{6} \\ -1/\sqrt{3} & -1 & -1/\sqrt{6} \end{pmatrix} \tag{5.3}$$

The directions of the wave vectors are equivalent to those of the center-to-vertex axes in a tetrahedron.
This choice gives

$$\text{Spacing} = a \tag{5.4}$$

$$\text{Normalization} < \psi^{FCC} >= 1 \tag{5.5}$$

$$|k_i|^2 = \frac{3}{2} \tag{5.6}$$

With the use of this normalization we get

$$< \psi^2 >= 9/8 \tag{5.7}$$

and consequently, after some algebra

$$< \psi^{FCC} \bigtriangledown \psi^{FCC} >= \frac{71}{162}k^2 \tag{5.8}$$

which corresponds to the first term, L_1, of the functional.
The second term can be calculated as a straightforward extension of the 2D case (Kropff and Treves 2008), by substituting to the sum over three terms the terms coming from the square in the formula.
Introducing the 2D Fourier Transform of the adaptation kernel K

$$\tilde{K}(k_i) = \int_V dq K(q) cos(k_i \cdot q), \tag{5.9}$$

we obtain:

$$L_2^{FCC} = \tilde{K}(0) + \frac{32}{81}\tilde{K}(|k|) + \frac{1}{648}\tilde{K}(2|k|) + \frac{1}{108}\tilde{K}(2/3^{1/2}|k|) + \frac{1}{108}\tilde{K}(2*2^{1/2}/3^{1/2}|k|). \tag{5.10}$$

By combining the two terms one finally obtains:

$$L^{FCC} = \gamma\tilde{K}(0) + \frac{71}{162}k^2 + \gamma\left[\frac{32}{81}\tilde{K}(|k|) + \frac{1}{648}\tilde{K}(2|k|) + \frac{1}{108}\tilde{K}(2/3^{1/2}|k|) + \frac{1}{108}\tilde{K}(2*2^{1/2}/3^{1/2}|k|)\right] \tag{5.11}$$

which can be simplified by retaining only the dominant term in the bracket:

$$L^{FCC} \simeq \gamma\tilde{K}(0) + \frac{71}{162}k^2 + \gamma\left[\frac{32}{81}\tilde{K}(|k|)\right] \tag{5.12}$$

The adaptation kernel may take various forms, but for reasons that will become clear in the next section, here we consider a kernel in a form that makes it factorable over

the spatial variables. Without loss of generality we will use a difference of radially symmetric Gaussians:

$$K(q) = K_L(q) - \rho K_S(q) = \frac{1}{(2\pi v\tau_L)^{3/2}} exp\left[-\frac{q^2}{(2v\tau_L)^2}\right] - \frac{\rho}{(2\pi v\tau_S)^{3/2}} exp\left[-\frac{q^2}{(2v\tau_S)^2}\right]. \tag{5.13}$$

The Fourier transform of this kernel is:

$$\tilde{K}(k_i) = \tilde{K_L}(k_i) - \rho\tilde{K_S}(k_i) = exp\left[-\frac{1}{2}(k_i v\tau_L)^2\right] - \rho exp\left[-\frac{1}{2}(k_i v\tau_S)^2\right]. \tag{5.14}$$

This choice allows us to modify Eq.5.11 into

$$L^{FCC} = \gamma\tilde{K_L}(0) - \gamma\rho\tilde{K_S}(0) + \frac{71}{162}k^2 + \gamma\left[\frac{32}{81}\tilde{K_L}(|k|) - \rho\frac{32}{81}\tilde{K_S}(|k|)\right] \tag{5.15}$$

5.3.2 HCP

Since the Hexagonal Close Packing does not have central symmetry, the choice of a function reproducing the arrangement of fields is less evident. We opt for:

$$\begin{aligned}\psi^{HCP}(r) = (1/2 + 1/2cos(k_z \cdot r))\left[1 + \frac{2}{3}\sum_i^3 cos(k^i_{xy}) \cdot r\right] \\ + (1/2 + 1/2cos(k_z \cdot (r + \Delta z)))\left[1 + \frac{2}{3}\sum_i^3 cos(k^i_{xy}) \cdot (r + \Delta x)\right]\end{aligned} \tag{5.16}$$

where two separate wave vectors are present: k_{xy} fixing the spacing on the hexagonal layers, and k_z used instead to regulate the distance between the layers (Figure 5.3, top right).
The components of k_{xy} are, again

$$k_i = \frac{2\pi}{a}\begin{pmatrix} 2/\sqrt{3} & 0 \\ -1/\sqrt{3} & 1 \\ -1/\sqrt{3} & -1 \end{pmatrix} \tag{5.17}$$

with $\mid k_{xy} \mid^2 = \frac{4}{3}$, while the z component is set to $|k_z| = \frac{\sqrt{3}}{2\sqrt{2}}$ and $\mid k_z \mid^2 = \frac{3}{8}$.
To obtain the correct HCP arrangement of fields Δx and Δz should be set to:

$$\Delta x = (\frac{1}{\sqrt{3}}, 0) \tag{5.18}$$

$$\Delta z = \sqrt{\frac{2}{3}} \tag{5.19}$$

Spacing and normalization are the same as for FCC.
With this choice of numbers we get

$$\begin{aligned}L^{HCP} = \tilde{K_L}(0)\tilde{K_L^z}(0) - \rho\tilde{K_S}(0)\tilde{K_S^z}(0) + \frac{1}{6}[\mid k^2_{xy} \mid + 3/2(\mid k_{xy} \mid^2 + \mid k_z \mid^2) \\ + \tilde{K_L}(k_{xy})\tilde{K_L^z}(0) - \rho\tilde{K_S}(k_{xy})\tilde{K_S^z}(0) + 3/2(\tilde{K_L}(k_{xy})\tilde{K_L^z}(k_z) - \rho\tilde{K_S}(k_{xy})\tilde{K_S^z}(k_z))]\end{aligned} \tag{5.20}$$

where the factorization of the adaptation kernel has been made explicit.
If we take $k^2_{HCP} = k^2_{xy} + k^2_z$ and $\tilde{K^z_{L,S}}(0) = 1$, then the previous expression can be rearranged in:

$$L^{HCP} = \tilde{K_L}(0)\tilde{K^z_L}(0) - \rho\tilde{K_S}(0)\tilde{K^z_S}(0) + \frac{1}{4}[\mid k^2 \mid + \tilde{K_L}(k) - \rho\tilde{K_S}(k)] + \frac{1}{6}[\mid k_{xy} \mid^2 + \tilde{K_L}(k_{xy}) - \rho\tilde{K_S}(k_{xy})] \quad (5.21)$$

where the similarity with the L^{FCC} (eq.5.15) dependence on k has been made explicit and the additional xy term is in evidence.

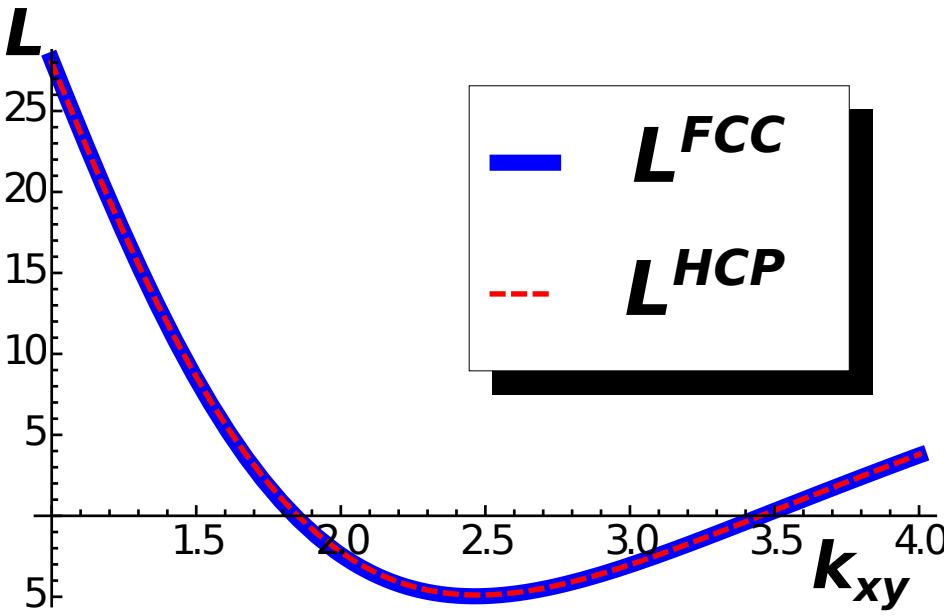


Figure 5.2: FCC and HCP arrangements are equivalent in terms of our cost function. The values for the plot have been evaluated with 5.23

5.3.3 The Cost Compared

We can compare the expressions for the cost of FCC (5.15) and HCP (5.21) by explicitly writing them as a function of k_{xy}

$$\begin{aligned} |k^{FCC}_{xy}| &= \frac{8}{9}|k^{FCC}| \\ |k^{HCP}_{xy}| &= \frac{1}{\sqrt{(1+9/32)}}|k^{HCP}| \end{aligned}$$

What we get is (omitting the $\tilde{K}(0)$ terms that are common to the two expressions and thus irrelevant for their comparison):

$$L^{FCC}(k_{xy}) = \frac{71}{162}k^2_{xy} + \gamma\left[\frac{32}{81}e^{-\frac{9}{16}k^2_{xy}\tau_L} - \frac{9}{16}\rho e^{-\frac{41}{64}k^2_{xy}\tau_S}\right] \quad (5.22)$$

$$L^{HCP}(k_{xy}) = \frac{529}{1152}k^2_{xy} + \gamma\left[\frac{1}{4}e^{-\frac{41}{64}k^2_{xy}\tau_L} - \frac{1}{4}\rho e^{-\frac{41}{64}k^2_{xy}\tau_S} + \frac{1}{6}e^{-\frac{1}{2}k^2_{xy}\tau_L} - \frac{1}{6}\rho e^{-\frac{1}{2}k^2_{xy}\tau_S}\right] \quad (5.23)$$

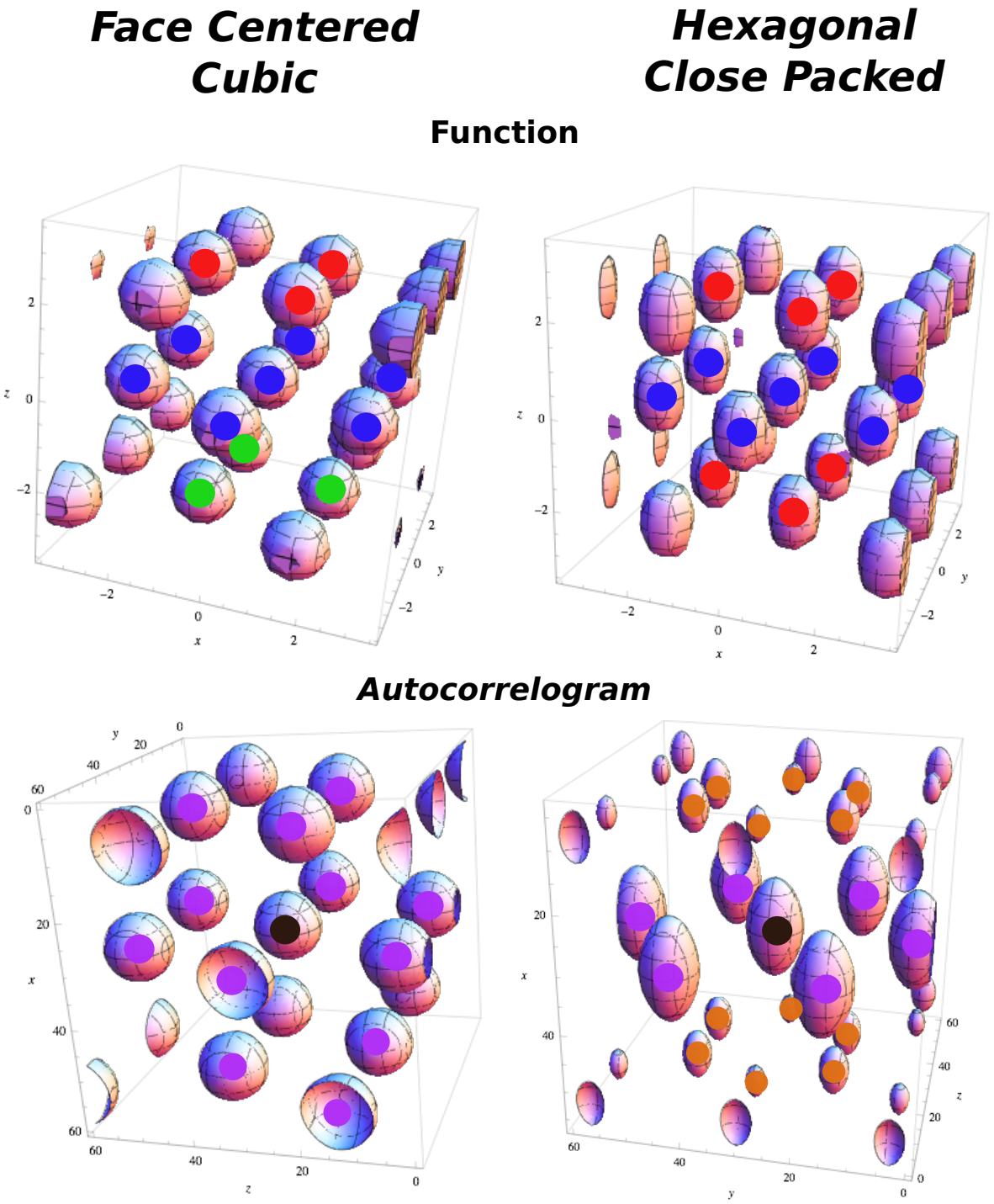


Figure 5.3: Top: Functions used in cost function calculation for FCC and HCP, color dots refer to 5.1. Bottom: Autocorrelograms of the above functions. Purple: full peaks. Orange: half peaks

In Figure 5.2 we plot these functions for $\tau_L = 1$, $\tau_S \ll \tau_L$, $\gamma = 200$, $\rho = 0.3$. They are almost perfectly coincident. And indeed, even though their (awkward) expressions look very different one can see, going back to them, that indeed: $\frac{529}{1152} \simeq \frac{71}{162}$ with a 1% margin, $\frac{1/6+1/4}{2} \simeq 32/81(3\%)$ and $\frac{41/64+1/2}{2} \simeq 9/16$ (1%), indicating that the quadratic term is nearly identical and the exponential term in L^{FCC} is roughly the average of the

two exponentials in L^{HCP}.

Although the two cost functions are not identical, they are very close across the parameter space. They have a very close minimum position, meaning they share the same optimal spacing, and they have the same value at the minimum, thus equivalently optimal.

5.4 Simulations

As done for the previous study on the sphere, we now perform numerical simulations aimed at reproducing the functioning of the more complex model with the full dynamical details.

5.4.1 The Model

The model used to perform the simulations is basically the same as the one described in the previous chapter.
The most obvious difference is the change of the environment explored by the virtual animal, which is now a cube of $1m$ side. The input to the mEC layer is still composed of units firing like place cells as in equation 4.23, with their preferred positions $\vec{x_{j0}}$ uniformly distributed in the volume. The norm $\|a - b\|$ appearing in the equation is now the Euclidean distance in three dimensions.
Two more substantial differences are represented by the introduction in the model of Head Direction information, through the assignment of preferred directions to mEC units, and the presence of recurrent connections in the mEC layer beside the feed-forward set between the two layers (Si, Kropff, and Treves 2012). Both these additions to the previous version are related to the grid alignment problem that we are going to study in the 3D case.
Therefore, the overall input to the unit j is now:

$$h_t^i = f_{\theta_i}(\omega_t)(\sum_j W_{ij}^{t-1} r_j^t + \rho \sum_k W_{ik} \Psi_k^{t-\tau}) \tag{5.24}$$

with $\rho = 0.1$ a factor setting the relative strength of feed-forward (W_{ij}^t) and collateral weights (W_{ik}), and $\tau = 25$ steps is a delay in the signal transmission.

5.4.1.1 HD Input

The multiplicative factor $f_{\theta_i}(\omega_t)$ in 5.24 is a tuning function which is maximal when the current direction of the animal movement ω_t is along the unit i assigned preferred direction (K. Zhang 1996).

$$f_\theta(\omega) = c + (1 - c)\exp[\nu(\cos(\theta - \omega) - 1)] \tag{5.25}$$

where $c = 0.2$ and $\nu = 0.8$ are parameters determining the minimum value and the width of the cell tuning curve.
Preferred head directions are randomly assigned to mEC units and they uniformly span the 4π solid angle.

5.4.1.2 Collateral Weights

The basic function of collateral weights in the model is to favor the development of fields in the post-synaptic unit with a certain phase shift relative to the pre-synaptic unit, and consequently to induce the alignment of grids, producing a common orientation in

the population. In this section we will not deal with collateral weight self-organization, only with feed-forward weight learning. For simplicity, the collateral weight are set at convenient values at the beginning of the simulations and left unchanged afterwards. We will deal with the issue of recurrent connection plasticity in the next chapter, in the scope of a somewhat different model.
Collateral weights are set in the following way: each mEC unit is temporally assigned a preferred position, an auxiliary field at coordinates (x_i, y_i, z_i). The coordinates are randomly chosen among the place field centers of the input layer. These auxiliary fields are introduced only for the sake of weights definition and will not play any role in other parts of the simulations. The collateral weight between unit i and unit k is then calculated as

$$W_{ik} = \left[f_{\theta_i}(\omega_{ik}) f_{\theta_k}(\omega_{ik}) \exp\left(-\frac{d_{ki}^2}{2\sigma_f^2} \right) - \kappa \right]^+ \tag{5.26}$$

where $[*]^+$ denotes the Heaviside step function, $\kappa = 0.05$ is an inhibition factor to favor sparse weights, $f_{\theta_i}(\omega_{ik})$ is the tuning function defined above (5.25), ω_{ik} is the direction of the line connecting the auxiliary fields of unit i and k, $\sigma_f = 20cm$ defines how broad the connectivity is, and d_{ki} is defined such that

$$d_{ki}^2 = [x_i - (x_k + l\cos(\omega_{ik}))]^2 + [y_i - (y_k + l\cos(\omega_{ik}))]^2 + [z_i - (z_k + l\cos(\omega_{ik}))]^2 \tag{5.27}$$

is the distance between the auxiliary fields with an offset $l = 10cm$. Note that l corresponds to the distance traveled by the virtual rat in a time τ, when moving at a speed $v = 0.4$m/s, as used in the simulations.
Normalization on this set of connections is performed similarly as in (4.29). The definition of the weights is such that it generates strong interactions between cells with similar preferred HD.

5.4.2 Simulation Results

What arrangement of fields emerges from our simulations? And in what time? Visualization of data in 3D is tricky, especially when trying to identify regular arrangement of fields (Fig 5.4, left). Calculating the autocorrelogram is thus critical to identify symmetries in the rat map, even more than in 2D (Fig 5.4, right). As we have seen in the previous section a FCC lattice has a central symmetry. In the same way as the autocorrelogram of a hexagonal grid is a hexagonal grid, calculating the autocorrelogram of function 5.2 would just reproduce the same configuration (Figure 5.3, bottom left) of fields, with 6 symmetric pairs of equivalent peaks surrounding the central one. Indeed the symmetries of the structure are such that one can find 4 planes passing throughout the origin which contain peaks arranged in a hexagonal way (Fig 5.6), these planes form angles of $\simeq 72°$ and are all equivalent.

The case is different for HCP, where the central symmetry is missing. In this case the autocorrelogram extracted from 5.16 does not reproduce the original form of the function. In Figure 5.3, bottom right, one can see that the autocorrelogram presents 9 pairs of peaks around the central one. But in this case these peaks are not all the same. The HCP structure is again periodical for translations along a plane, generating 6 peaks of height 1, like those of FCC (Figure 5.3, purple peaks). As the structure is translated out of this preferred plane, the ABAB arrangement of the HCP layers

is such that there are no translations that reproduce the exact same configuration of fields, in the autocorrelogram. The 6 peaks above the central one ((Figure 5.3, orange peaks) are indeed half-peaks, corresponding to an overlap of only half (6 out of 12) of the peaks of the basic unit. Therefore, although one can build 7 planes with hexagonally arranged peaks on this autocorrelogram (Fig 5.6), they are not all equivalent to those in FCC. Only one of them contains all the peaks of height 1 and forms an angle of $\simeq 72°$ with the other 6, constructed with half-peaks and forming an angle of $\simeq 56°$ between them.

We can thus try to exploit this difference in the autocorrelogram that the two configurations should theoretically produce, to distinguish in our simulations between units developing grids FCC-like or HCP-like. The difference is in the relative positions of the fields across three layers. Observing only one, or even only two layers of fields would make it impossible to distinguish between the two.

We simulate the exploration of the environment for $3 * 10^5 s$ of equivalent "bat time", a duration significantly greater of that used for the model in 2D (Si, Kropff, and Treves 2012), but necessary for the estimated longer developmental time that the higher dimensionality of the environment will presumably require.

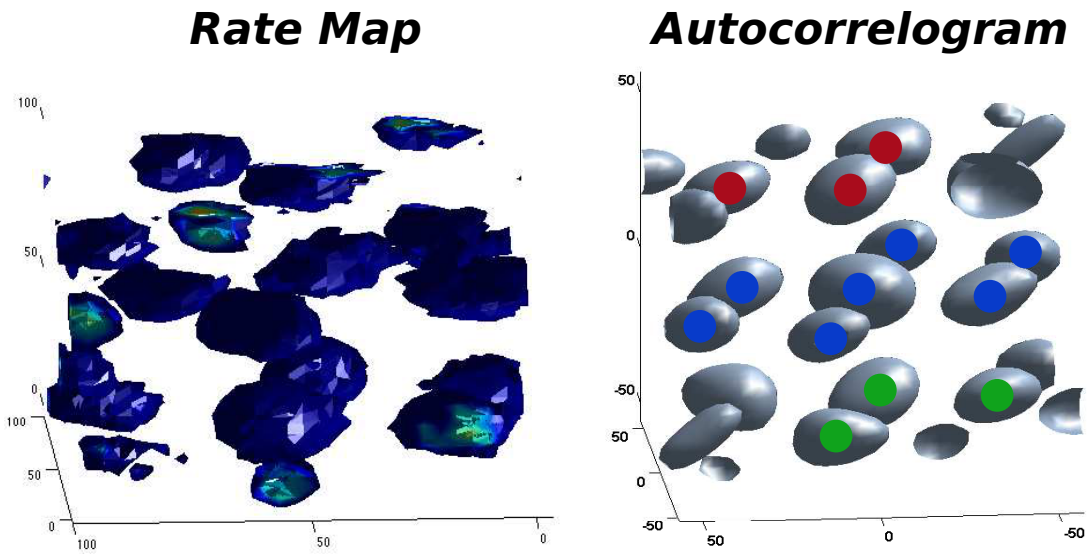


Figure 5.4: 3D firing rate map and autocorrelogram for one example cell.

First, by running a Fast Fourier Transform on the units autocorrelograms, one can see that the power spectrum of frequencies has a major peak around $\lambda = 50cm$, which indicates the presence of a narrow distribution of inter-peak distances, something espected from the adaptation process. Interestingly, a second weaker peak appears around $\lambda = 40cm$. By considering tetrahedron geometry, one quickly realizes that the ratio between the two lengths is the same found between the vertex spacing and the height of the solid. This second peak is thus induced by the presence of periodical layers of activity at an interval $\lambda_z = \sqrt{6}/3\lambda$, where λ indicates the fields spacing.

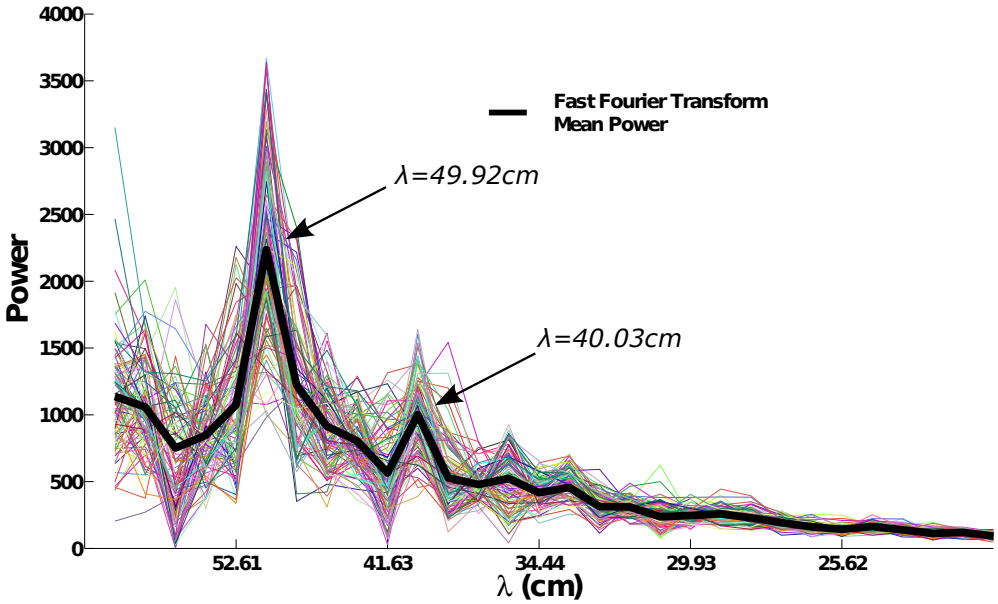


Figure 5.5: Fast Fourier Transform of the units autocorrelograms: colored lines are the power spectra of single units, the thick black line is the mean across units. The first peak at $\lambda \sim 50cm$ corresponds to the fields spacing. The second peak at $\lambda \sim 40cm$ is the effect of the interlayer distance. The two distances are in very good agreement with the ratio expected from a tetrahedron-like arrangement of the fields in space.

To study the formation of the three dimensional grids we will divide the problem in two steps. We will address separately the formation of a layered structure of hexagonal planes and the relative phases developed by these planes.

Layers formation and alignment To investigate the structure of the activity on the single layers, we take multiple slices of the autocorrelogram matrix. We take sections passing through the center, with different angles of azimuth and elevation (Figure 5.6). We then take the correlation values on each of these planes, and we compare it with a hexagonal template of equidistant peaks with $\pi/3$ periodicity (Figure 5.7, left). The plane most resembling a hexagonal pattern (the "best plane") is selected together with its similarity score, ζ_1 (we used a correlation measure). This method provides us with an equivalent of the traditional gridscore used to judge the quality of planar grid cells. By looking at the evolution in time of the scores in the population we can estimate the development of the hexagonal layers in the volume. Figure 5.7 shows how the units in the network actually reach high overlaps with the hexagonal arrangement (> 0.8 correlation) in about 40.000 seconds. This time is not dramatically larger than the one necessary for an equivalent model to generate grids in two dimensions (Si, Kropff, and Treves 2012). The effects of collateral interactions should be seen in the way these best planes align in space. We looked for the dispersion of their orientations, calculating the angles formed by the best planes of different cells (β), and the angle between the hexagonal grid axis of cells sharing a similar best plane (ω) (Figure 5.7, center and right). We indeed see that collateral interaction tends to align in time the

principal layer of different cells, defining a preferred orientation for the global structure of the grid. The emergence of the common alignment of the layers is much slower than the formation of the layers themselves. It takes about 100.000 seconds of exploration time to significantly reduce the dispersion of the angles (note that β is slightly faster, maybe an effect of the preference of the system to first define the layers and then arrange shifts within them). Although this time is larger than the one observed in the planar equivalent (about double), it is not dramatically higher.

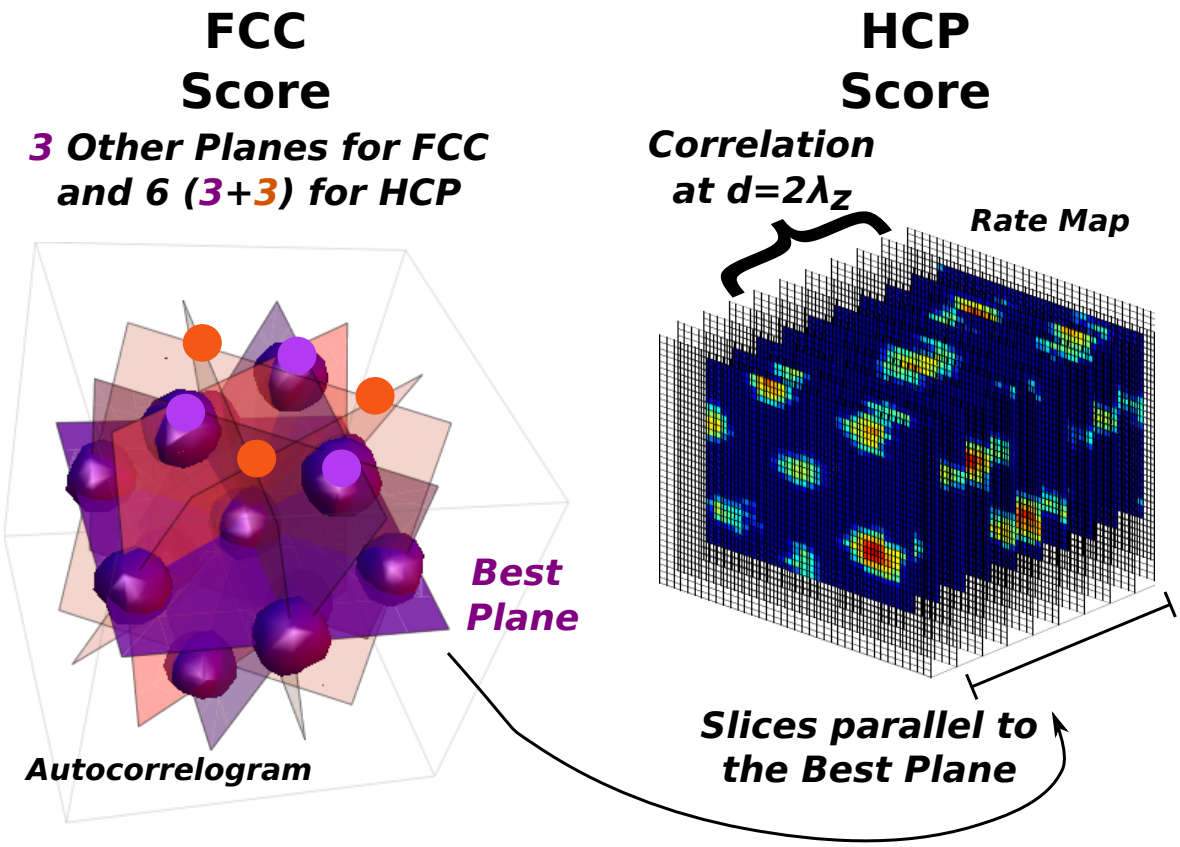


Figure 5.6: Illustration of the calculation procedure for our FCC and HCP Scores.

Mutual arrangement of layers sequences The previous measures do not tell us anything about the relationships between neighboring layers. How can we distinguish FCC and HCP in our autocorrelograms? We devised the following measures to quantify the degree of similarity of them to the two prototypes (Figure 5.6). We use the fact that once the best plane has been identified, the FCC has 3 more planes with hexagonal symmetry, at $\sim 72°$ from the best plane and between one another. HCP instead has 6 of them, again at $\sim 72°$ from the best plane, but at $\sim 56°$ between them. We then take our slice scores, take all the slices at an angle of $\sim 72°$ from the best plane and sum the scores of the best triplet of slices with $\sim 72°$ of separation (ζ_{2-4}). We then exclude them and take a second triplet of slices again with a $\sim 72°$ distance from one another and a distance of $\sim 56°$ from the first triplet (ζ_{5-7}). These two numbers tell us about the number of different planes with hexagonal symmetry that can be built from our autocorrelograms. Both scores run from -3 to 3, as they are the sum of three correlations. We expect ζ_{2-4} to be high for both FCC and HCP arrangements, and

its value should be considered as an indicator of the general quality of the grid. ζ_{5-7} instead should be high only for those grids presenting an HCP type of arrangement, but again its value might be affected by the quality of the grid. We thus define a score for the degree of FCC similarity as:

$$\chi_{FCC} = (\zeta_{2-4} - \zeta_{5-7})/\zeta_{2-4} \tag{5.28}$$

that should be close to 1 in the presence of FCC, and to 0 in the HCP case.

On the other hand, HCP is characterized by the repetition of the same fields positions every two layers, while FCC has a periodicity of three layers. Then another way to characterize the grids is to look for similarities between layers. Since the best plane we calculated indicates the direction of stacking of layers in the HCP (along its normal vector), we can go back to the firing rate map, take slices along this direction (which is, slices with the same best plane orientation) and calculate the correlation between planes separated by a two-layer distance ($2 * \lambda_z$).

$$\chi_{HCP} = \rho_{auto}(2 * \lambda_z) \tag{5.29}$$

Specularly to the previous score, this one should be close to 1 when HCP is expressed, and to 0 when FCC is.

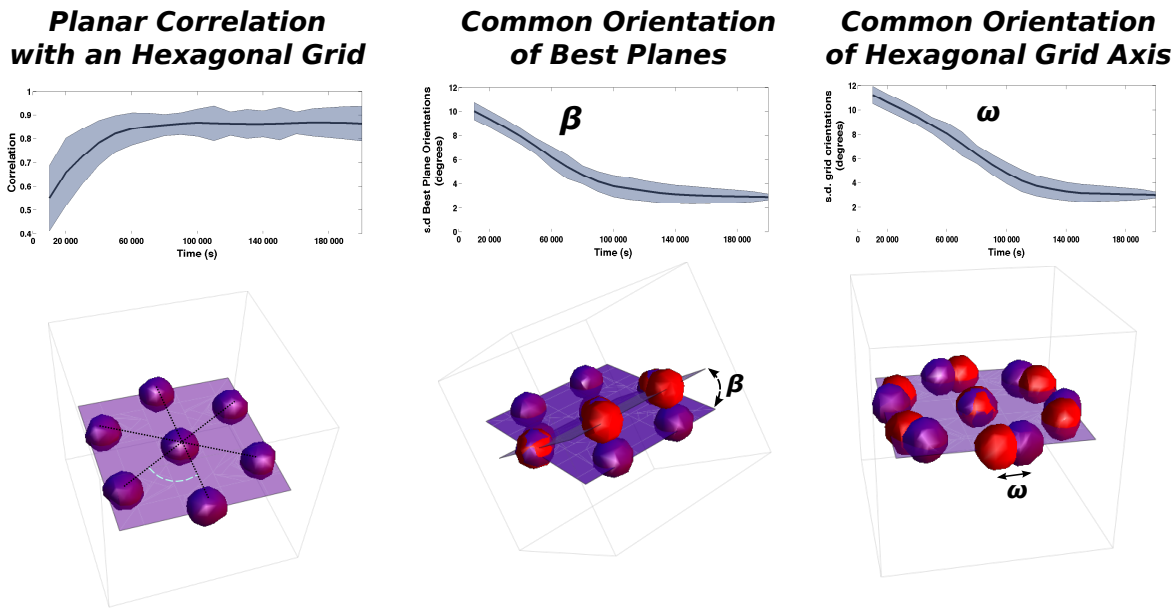


Figure 5.7: Emergence and alignment of the layers. Left: Correlation between the autocorrelogram best plane and a hexagonal grid. Center: standard deviation of best planes orientation. Right: standard deviation of grid orientations on the plane.

By looking at the combination of the two scores across the population we can assess the distribution of the two prototype packings (or of mixtures of them) (5.8). If we look separately at ζ_{2_4} and ζ_{5-7}, we see that they reach their asymptotic value (~ 2.6 and ~ 1.5, respectively) in a time course similar to the one needed to generate the single hexagonal planes (**??**, left), and that they do not show any further variation (5.8, a).

The same happens for the two scores we defined, they have a quick initial transient to an asymptotic value, intermediate between 0 and 1 (~ 0.6 for χ_{FCC}, not shown, ~ 0.5 for χ_{HCP}) that is then almost constant for the rest of the simulation. The value of the scores indicates the presence of both arrangements in the system. If we look at the scores for each unit at a given time (5.8, b) we find that these are not clustered in two groups, each one homogeneously expressing an HCP or FCC arrangement, but instead they cover an entire continuum of score distribution between the two extremes, expressing all intermediate arrangements.

Therefore, it appears that the system does not manage to arrange the sequence of layers in a common way. The units of the population express different offset between the fields on the different layers, corresponding to different combinations of FCC and HCP. While the emergence of the hexagonal symmetry on the single layers is a process unfolding in a similar time as in 2D, the positioning of the layers with respect to one another appears more complex, accomplished on a much longer time scale, or maybe not accomplished at all. Further, longer, simulations should address this point.

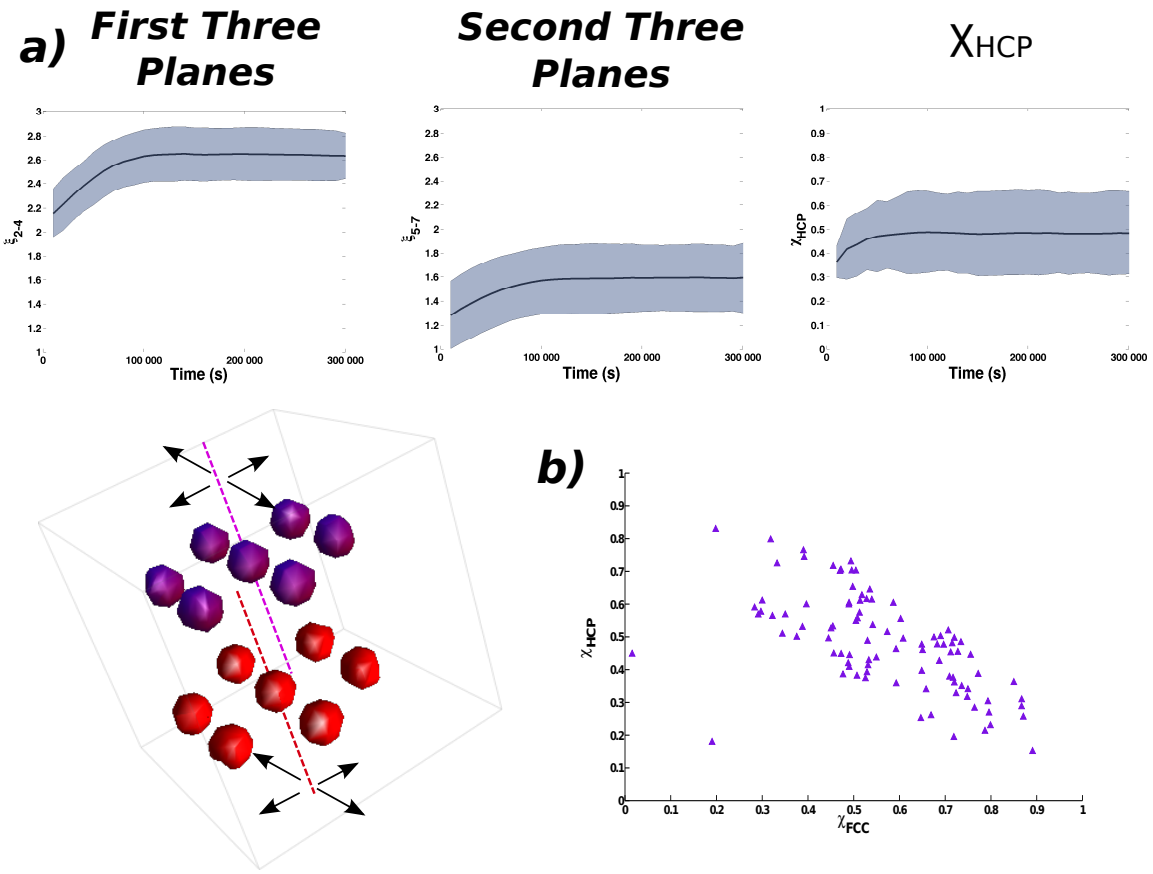


Figure 5.8: Multiple layers arrangement scores. a) Temporal evolution of ζ_{2-4}(left), ζ_{5-7}(center) and χ_{HCP}(right). b)Example of the continuous distribution of scores across the population.

5.5 Conclusions

Our model shows how grids in three dimensions may be produced starting from the very same principles of auto-organization, without increasing costs in higher dimensions. At least three time frames can be recognized. The shorter one, the one necessary to organize the fields on a hexagonal pattern is almost identical to the time needed to obtain the same hexagonal arrangement on the plane (Si, Kropff, and Treves 2012). Also the common orientation between the grids is developed in a limited, although longer time. Again, this interval is not significantly longer than its equivalent on the plane. A significant difference with the planar case is the observed lack of convergence towards a common phase arrangement of the different layers, either coincident with FCC, or HCP.

The difficulties of the system to find a common arrangement might lie in the increased complexity brought by the higher dimensionality, but it might also be that the presence of multiple, "energetically" equivalent arrangements, namely FCC and HCP, to chose from, in 3D, is the origin of this lack of coordination. In two dimensions the solution of the cost function minimization is unique, in three there are two similar but not identical solutions, and perhaps a entire set of intermediate arrangements.

Chapter 6

Building a Continuous Attractor through Adaptation

6.1 Introduction

In the previous sections we have seen how our model based on adaptation is able to produce different grid structures depending on the environment to which the animal is exposed during development, and without the need of any change in the network properties to adjust to different situations.

For the model to provide a more complete picture of the development proper grid cells, nevertheless, there are at least two more questions that need to be addressed. These two issues are of critical importance as they concern any implementation of the present model and affect its explanatory power. They are:

- Through which kind of process do recurrent connections self-organize from scratch, to induce a common alignment in grids?
- What mechanism makes the map, generated during development, available for use in novel environments, experienced by a rat later in its life?

As we will see, these two issues have features in common, and from an answer to the first, apparently more technical, question we might also learn something about the second.

6.2 Inhibitory Connections

The adaptation model described in Kropff and Treves (2008) has at its core a single-unit process, that would induce grid formation also in a neuron in complete isolation from the rest of mEC population. Interactions among units have to be introduced when it is necessary to make sure that they express grids in a coordinated way, by differentiating their phases and aligning their axes. As we have seen in the three-dimensional case, the latter effect can be obtained through the presence of recurrent connections in mEC layer. In the model described above these connections were excitatory and their strength depended on the HD preference of the connected units. This choice was motivated by the abundance of head direction information present in the deep layers of the Entorhinal Cortex (Brandon, Bogaard, Libby, et al. 2011; Canto, Koganezawa, et al. 2012), not only in the form of head direction cell, but, especially in layer V, of

conjunctive cells (grid x head direction) (Sargolini et al. 2006).

Following this line, the question of the self-organization of these internal connections has been addressed in a study (Si and Treves 2013) where directional inputs and the layer structure of mEC play major roles. In the model, pure grid formation in superficial layers of mEC is obtained via the mediation of a population of conjunctive cells in the deeper layers, that self-organize their activity and their recurrent connections.

In the following chapter, we will address the same issue from another perspective, which is alternative to the one adopted in (Si and Treves 2013). Again, we will investigate a mechanism to induce common orientation among mEC units while they are forming grids, but with different ingredients than in Si and Treves (2013). Basically, the main difference is the use of inhibitory connectivity to wire mEC units.

Inhibitory Connections in mEC Layer II The choice is motivated by recent findings (Couey et al. 2013) indicating how stellate cells (generally identified with grid cells) in layer II of the adult mEC are interconnected solely through inhibitory interneurons. The study found that any sign of a direct excitatory interaction between stellate cells disappears after a few post natal days and even before P10 it is an extremely rare event. The result is limited to layer II, and there are no indications as to whether it may extend or not to the deeper layers. Given the profound differences in the neural populations at different layers (Sargolini et al. 2006) there is no reason to assume that. The possibility that superficial layers get direct hippocampal inputs still has to be addressed experimentally. The model does not constraint the source of the spatial input to superficial mEC. It might be the hippocampus directly (Bonnevie et al. 2013), or it might be hippocampal input mediated by the deeper layers, or some other spatially modulated signal present in abundance in the parahippoacampal regions (Boccara et al. 2010). It is also true that from a modeling perspective it is interesting to explore different routes to the same objective and to study the implications and the consequences of particular network designs, especially when the experimental indications are so clear-cut.

6.3 The Model

To study the effects of inhibitory collaterals on the development of grids, we have used the usual Kropff and Treves model in the planar environment (Kropff and Treves 2008). The number of units in the input layer is $N_{inp} = 400$ and an equal number of units N_{mEC} is present in the mEC layer. The environment is a square box of 1.5m side. The details of the simulations are as in the standard case, the total input on unit i being

$$h_t^i = \sum_j \overset{+}{W}{}_{ij}^{t-1} \; r_j^t + \sum_k \overset{-}{W}{}_{ik}^{t-1} \; \Psi_k^{t-1} \tag{6.1}$$

where $\overset{+}{W}$ are the usual feed-forward connections between units of the input layer and the mEC layer of the model, whereas the second term represents the effect on unit i of the rest of the mEC population. $\overset{-}{W}$ are negative collateral weights implementing the inhibitory interaction between grid cells, mediated by a population of interneurons which are not explicitly included in the model.

Both sets of connections are now affected by plasticity, and in both cases the delay in the transmission is assumed to be short, of the order of a simulation time-step.
Following the model for inhibitory interaction proposed in Couey et al. (2013), we use binary inhibitory weights:

$$\bar{W}_{ik}^{t} = 0, -I. \tag{6.2}$$

where $-I = -0.02$. While the learning rule for the feed-forward connections is unmodified from other versions of the model, the update procedure for the weights of the collaterals requires some explanation.
Inhibitory connections values are not updated continuously during simulations like excitatory ones. Instead plasticity effects are taken to become effective only at the end of entire exploration sessions lasting roughly 1hr of virtual time. After the end of one session and before the start of the next one, inhibitory weights are checked and, if necessary, updated. Being binary variables, changing a connection means taking its value from 0 to $-I$, or the opposite. The update is made on the basis of the value of a hidden variable, which is instead update at every time step during the session, much like excitatory weights.

$$\chi_{ij}^{t} = \chi_{ij}^{t-1} + \epsilon_I(\Psi_i^t\Psi_j^t - \bar{\Psi}_i^{t-1}\bar{\Psi}_j^{t-1}) \tag{6.3}$$

The rate is $\epsilon_I = 0.0001$, $\bar{\Psi}$ is defined as in 4.28. The variable is allowed to assume negative values, and its role is to keep track of the mean overlap between different units, for the duration of the simulation.
At the end of a session the check is made and the weights are updated according to:

$$\bar{W}_ik = \begin{cases} -I, & \text{if}\chi_{ij}^t \leqslant 0 \\ 0, & \text{if}\chi_{ij}^t > 0 \end{cases} \tag{6.4}$$

Therefore, inhibitory connections are established between units with anti-correlated activity, while units with similar spatial selectivity do not interact.
At the beginning of the simulations the inhibitory set of connections is completely random. Each unit is assigned with a certain number of non-zero collateral connections.

6.4 Results

6.4.1 Grid Alignement in mEC

We first asked whether mEC units can form grid maps coherently aligned while at the same time inhibitory connectivity develops. We studied the evolution of the activity in the mEC layer over time looking for the appearance of grid-like patterns, and for the emergence of a common alignment (Figure 6.1). As we show in Figure 6.2, the two processes unfold more or less at the same time, but while the units increase their grid score gradually, the emergence of a common orientation is an abrupt event that occurs between $3 * 10^5 s$ and $6 * 10^5 s$, depending on the particular parameters of the simulation. Once this transition happens, the standard deviation of the grid axis drops to almost 0, while the mean grid score progresses to values well above 1 (Figure 6.2).

6.4.2 Self-Organization of Inhibitory Connections

In the model, the common orientation is the product of collateral self-organization that modifies their values according to a Hebbian-like learning rule. As this connectivity is

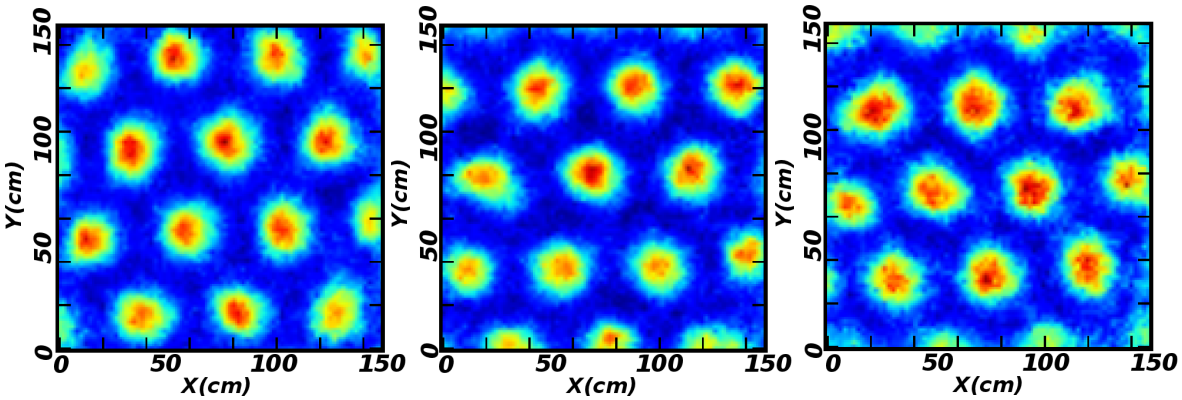


Figure 6.1: Three examples of grid obtained with the model.

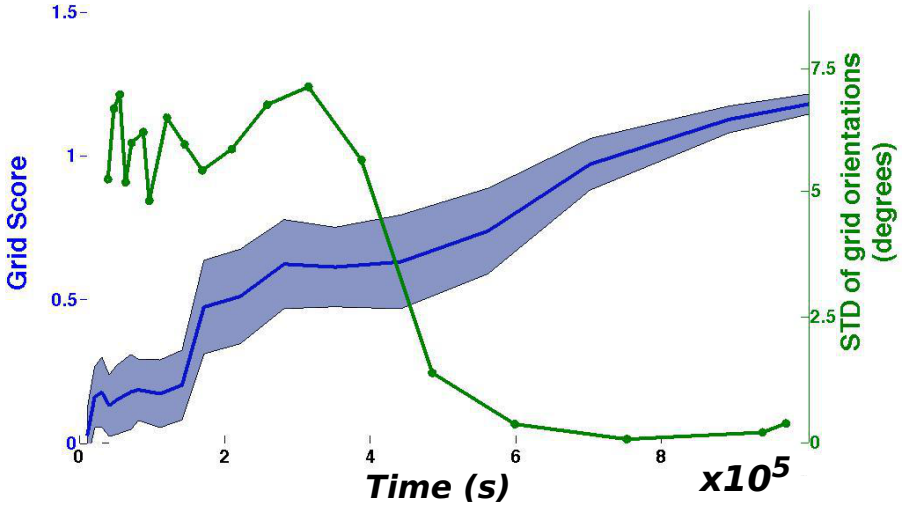


Figure 6.2: Emergence of aligned grid cells. Blue (left axis): mean grid score in time (shade = standard deviation); Green (right axis): standard deviation of grid axes direction distribution.

shaped by plasticity, it induces a coherent alignment of the grid axes, that, in turn is reflected on further changes in the connectivity due to changing correlations between units.

What is the final outcome of this recursive process? As at the end of the simulations grids are finally aligned, and have the same spacing, the correlation between units should depend only on their spatial phases. Therefore also the existence of an inhibitory interaction between two units can be attributed to the phase relation between two units. A unit should have connections with those that units cover different portions of the environment and should lack connections with the rest, those that are neighbors in terms of phase distance.

Indeed, if we plot the probability of having a connection as a function of the relative phase distance across the mEC population (figure), we observe that this relation indeed exists. The connection probability exhibits a clear transition from values close to zero,

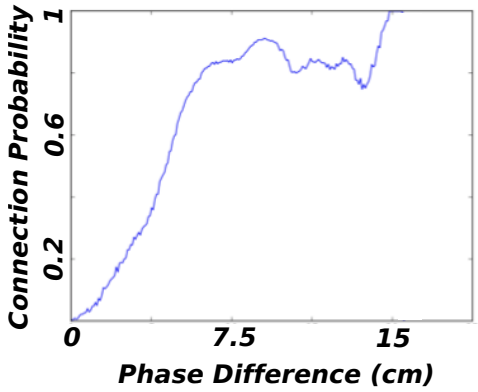


Figure 6.3: Probability of connection as a function of phase difference between units.

for small distances, to almost full connectivity after a critical phase difference.

6.4.3 A Link to Continuous Attractor Models?

The presence of this kind of connectivity profile at the end of the learning process, suggests a possible development of this model. In fact its properties closely resemble those of the connections used to implement a continuous attractor, in the models for grid cells based on path integration (Guanella, Kiper, and Verschure 2007; Couey et al. 2013). Of course, the model in its present form can not perform path integration. For that, it would be necessary to introduce velocity selectivity in mEC units and HD signals to the inputs of the system (like the one described in Couey et al. (2013)). Nevertheless, the mechanism just described might represent a valid candidate for setting up the particular connectivity pattern necessary for any of these models to work.

It might be possible that the initial formation of grid cells, sculpting the mEC internal connections over an extended developmental period, is functional to continuous attractor dynamics in mEC, supported by those internal connectivity. In turn, after the end of this critical period, mEC might produce grid patterns also independently from the hippocampus, relying on integrating directional signals from other sources and thus allowing for fast grid expression in novel environments, as reported in various experiments (Barry, Ginzberg, et al. 2012).

In this perspective, the role of the hippocampus might seem a transient one, helping the mEC to prepare the scene, and then fading away; but instead the interaction between these two structures is likely to continue after the developmental period (as suggested by many experiments and more directly by the recent study by Bonnevie et al. (2013). Possibly, after this phase, the mEC network is able to work in two modalities. It can produce grid cells with common orientation through two alternative mechanisms, with attractor dynamics performing path integration on incoming velocity stimuli (as described in Couey et al. (2013)), or integrating spatially tuned inputs coming for example from the hippocampus, or another source in the neighboring regions (as described in Kropff and Treves (2008)). It is also possible that the system associates the two inputs when learning plasticity acts on the connections projecting from the hippocampus. In particular, the two dynamical regimes are potentially relevant in different time frames, with path integration dominating the initial expression of grid

cells in a new environment, while hippocampal input may take the lead, and become necessary for stable grid expression, after learning has taken place (Bonnevie et al. 2013).

6.4.4 Long Lasting Effects of Development

Another interesting consequence of the present model, is the relevant role that has the environment, in which the animal develops, on later expression of grids in other environments.
The internal connectivity supporting grid alignment develops together with the first grids and its structure reflects the phase relationships established among the units during this process. But it is clear that the phase relationships are affected by the nature of the environment the animal can move in. Considering the example of a rat that spends its developmental period constrained to move along a one dimensional environment, will make the point clear.

Naturally there is nothing to align in one dimension, but the model can be used anyway to produce grids with periodical firing fields and to illustrate the differences with the two-dimensional case. We use the very same model described above, this time having the rat running continuously on a linear track of 1m length.

The emergence of fields at regular intervals along the track is very fast in this case (one session is enough), reflecting the much reduced complexity of the process due the reduced dimensionality. Different units share the same field spacing and phases are distributed uniformly (Figure 6.4). If we now take a look at the way inhibitory

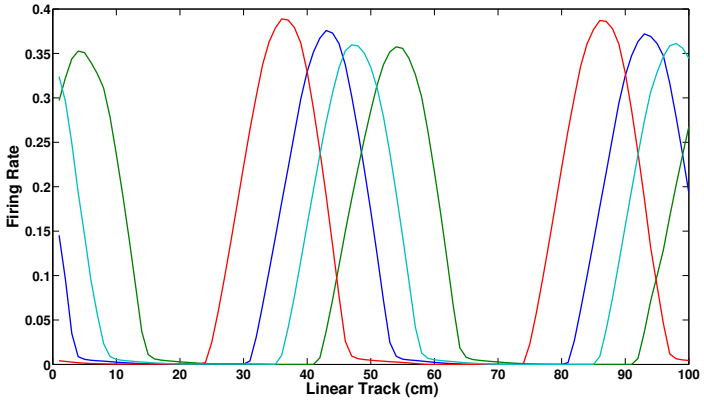


Figure 6.4: Example of grid cells developed in 1D by the model

connections are organized, Figure 6.5, we can rearrange the connectivity matrix rows and columns to unmask the regular pattern generated during learning. As in the 2D case, the emergence of preferred phases in the activity of mEC units can be used to define a notion of distance among the units and to induce the formation of a wiring based on that distance.

By rearranging the units in the matrix, we are uncovering the phase distance between them. We are rearranging the labeling of the units in the network so that units with

similar phases are close to each other. As in the previous case, the resulting connectivity could be used to sustain a bump of activity in the network, and it can provide the basis to perform path integration, moving around the bump on the continuous attractor. The only difference is that indeed, the dimensionality of the attractor is 1 in this case, a ring attractor. Phases are evenly distributed on a line and therefore the connectivity is reflecting the existence of a right and left neighborhood only. It is evident the impossibility of using a network with this sort of wiring to integrate movements in 2 or more dimensions. The same argument applies to a two-dimensional continuous attractor, where there is no way for the bump of activity to move along the z direction and track vertical displacements. Of course such a network could be used in a lower dimensionality, just by projection.

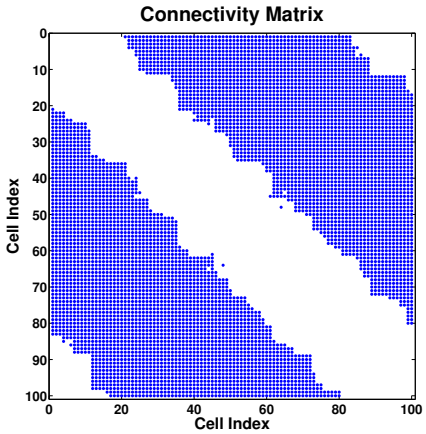


Figure 6.5: Reordered adjacency matrix for a 1D development process.

6.4.5 A Problem with the Phases

Coming back to the two dimensional case we have to mention that the model is still not perfectly functioning. In fact, while the gridness score and the common orientation develop correctly in time, the grids exhibit an inhomogeneous distribution of phases at the end of the learning phase. To a closer look (Figure 6.6) it is evident how the phases tend to cluster in 6-7 groups with a spacing of about 10cm. A comparison with a similar model ((Si and Treves 2013)), that indeed presents shorter learning times and evenly distributed phases, may suggest that a solution might come from borrowing some of its components, namely by adding a further mEC layer or by introducing HD selectivity.

6.5 Summary

The structures that underlie spatial representations are common across mammals and to a certain degree also to birds and reptiles (Striedter 2005). At the same time these animals show different spatial behaviors and actually inhabit completely different environments. They crawl, they fly, they climb, they like open fields or obscure holes. Reasonably this heterogeneity of behaviors is reflected in a differential necessity of representing the various properties of space in which they navigate (Jeffery et al. 2013). The model that we have described in the previous chapters offers a possible, although

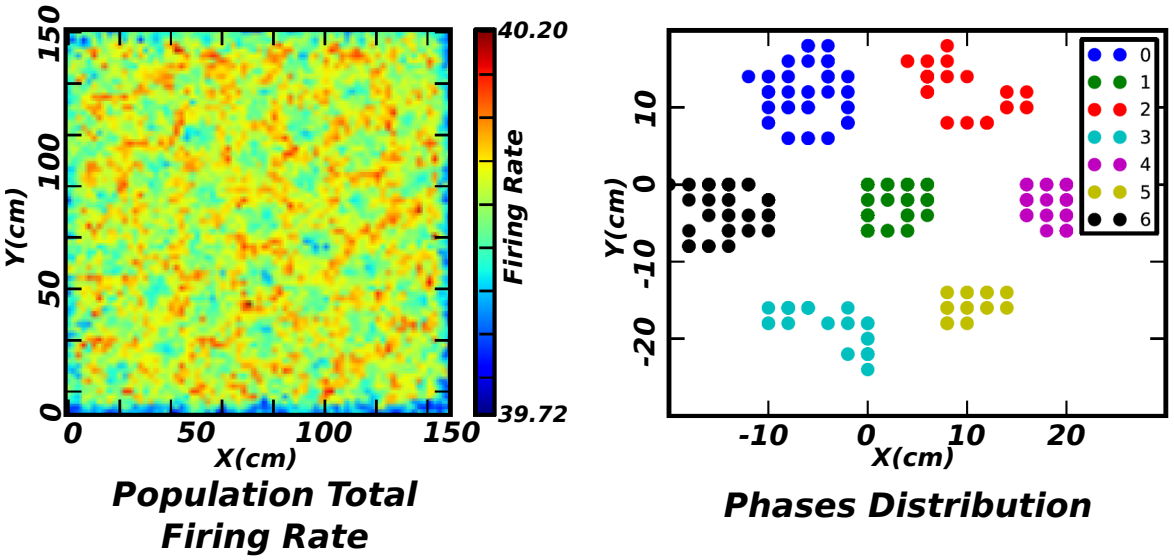


Figure 6.6: Phase distribution in the population. Left: aggregated firing rate of all units across space. Right: clustering of units according to their fields center. Note that the spatial scale is reduced because of grid periodicity.

still tentative, mechanism to link these two aspects.

In the model, the initial exploration of space during a critical developmental period (Langston et al. 2010; Wills, Cacucci, et al. 2010) determines what kind of grid the units in mEC will express. To keep it simple, moving on a plane would result in hexagonal, planar grids, while flying in space will generate three dimensional arrangements of fields. One can speculate that if this early generation of a regular firing activity in space is transferred to the pattern of the internal connectivity, that is maximally plastic in the same period of time(Langston et al. 2010) and then stabilizes (Couey et al. 2013), then early life experiences will have a long-term effect on the ability of mEC to produce grids later in life. In particular, the fixation of certain phase relationships among units in mEC will then affect the possibility of apply the grid to environments with an higher number of dimensions, or a different metric structure.

Chapter 7

Telling Apart Position and Context in mEC and Hippocampus

7.1 Slow Grids for Fast Place Encoding

In the previous sections we have described a neural process enabling the generation of a regular grid-like pattern of activity in a layer of mEC units.

Two aspects of the model are of particular relevance for understanding the role of a grid cell representation in the larger scheme of the hippocampal-entorhinal circuit. According to a series of theoretical studies on grid coding efficiency (Fiete, Burak, and Brookings 2008; Sreenivasan and Fiete 2011; Mathis, Herz, and Stemmler 2012a; Mathis, Herz, and Stemmler 2012b), the periodicity of grid representations is "optimal" for estimating spatial position and therefore superior to the accuracy attainable by a set of place cells. Our model is on the same line as these studies, providing a causal mechanism to obtain such regular spatial periodicity from any kind of spatially modulated input, thus increasing the efficiency of the input code; but it also pushes the concept a little bit further. It provides a way to generate the optimal tessellation for the environment (Conway 1992; Cohn and Kumar 2007; Petkovic 2009). It generalizes the properties highlighted by those studies as optimal for flat 2D representation, to any kind of spatial metric, breaking free from the constraint of a hexagonal, planar grid. The outcome of the self-organization process is still unique, once the environment is given. Which structure the initially random representation will express, is defined by the external conditions.

Another meaningful aspect of the model we described is its gradual self-organization. Rather than being a flaw, the slow pace at which grids emerge defines a relevant property of grid representations, in contrast with the fast acquisition of place cells maps (Nakazawa et al. 2003; T. T. McHugh et al. 2007; Nakashiba, Young, et al. 2008; T. McHugh and Tonegawa 2009). What emerges from the slow adaptation process is a once-for-all structure that does not need any further modification to function. This is true also for other models of grid cell formation. Those based on continuous attractor networks in fact, rely on some pre-wired structure that should also be fixed. The absence of plastic changes in the connectivity of the region is a requirement for the success of that approach. Whether the connectivity is the result of genetic instructions or of experience, this does not change the fact that grid cells are supposed to emerge once and then remain untouched.

7.2 Beyond the Rigidity of mEC code

The medial entorhinal cortex appears to contain a sort of universal map of allocentric space. It is a rigid map, in which the relations between units remain fixed, almost like sheets of millimeter paper stuck to each other (Fyhn, Hafting, Treves, et al. 2007). The firing of a grid cell covers the entire environment, and it does so in a strikingly periodical fashion. The way in which these cells represent space, compared to place cells, appears to be much more metric in nature. The grid code seems to be related to the basic 2D geometry of the environment, and to carry a large amount of spatial information, larger than place cells coding. Before the discovery of grid cells, spatial information could have been thought to be encoded in its final representation within the hippocampus. Now it is clear that spatial information is available in a clean and precise form before accessing the hippocampus.

What is the hippocampus doing with it remains to be discussed. It appears that the map generated in medial entorhinal cortex is applied and used irrespective of context (Fyhn, Hafting, Treves, et al. 2007). Other parts of the entorhinal cortex, like its more lateral portion, receive context-specific signals (Lu et al. 2013). The hippocampus can thus operate on the universal map and on the context-specific signals, merge them and create context-specific metric representations of space (Hayman and Jeffery 2008; Pennartz et al. 2009; S. Mizumori 2013), which are stored in memory. If this is correct, then, the grid cells representation should enable an evaluation of the position of the animal in the environment with great precision (Fiete, Burak, and Brookings 2008; Sreenivasan and Fiete 2011; Mathis, Herz, and Stemmler 2012a; Mathis, Herz, and Stemmler 2012b), while it should be almost impossible to determine which environment the animal is in, just by looking at the activity of grid cells. In the case of the place cells representation, instead, a possible decrease in the amount of positional information should be traded off with a good capability to distinguish among different environments.

The same argument should be valid also with other sorts of non-spatial elements, beside context only. The hippocampus seats at the top of a hierarchical system of converging sensory-motor inputs (Lavenex and D. Amaral 2000). It receives extensive inputs from cortical (Naber et al. 1997).and sub-cortical (D. Amaral and Cowan 1980) regions This inputs contain information that has already been duly processed and therefore represents all sort of variables, including cognitive and emotional state, action-value estimates (Pennartz et al. 2009), and so on (Eichenbaum and Lipton 2008; Meer et al. 2010). In this perspective the grid cell input is just one of many contributions that interact in the hippocampal hub Lu et al. 2013; S. Zhang et al. 2013. And indeed, hippocampal maps might reflect the combination of a multiplicity of spatial and non-spatial factors, bound together to form highly processed, richly informative wraps.

7.3 The Hippocampus in our Life

In 1957, Brenda Milner reported the profound effect on memory of bilateral medial temporal lobe resection, carried out to relieve epilepsy in patient H.M (Scoville and Milner 1957). After the surgery, the patient exhibited profound forgetfulness but in the absence of any general intellectual loss or perceptual disorders. He could not form new memories (anterograde amnesia) and also could not access some memories acquired

before his surgery (retrograde amnesia). His impairment extended to both verbal and non-verbal material and it involved information acquired through all sensory modalities.

Which structures within H.M.s lesion are really important for memory was understood only gradually, during the 1980s, following more case studies with more specific lesions (Zola-Morgan, L. R. Squire, and D. G. Amaral 1986; L. Squire and Zola-Morgan 1991), and the successful development of an animal model of human amnesia in the non-human primate (Mishkin 1982; Zola-Morgan, L. R. Squire, and Mishkin 1982). Cumulative behavioral work with the animal model, together with neuropsychological studies, thus identified the critical anatomical component of the medial temporal lobe memory system in the hippocampus itself; and pointed at the additional roles of the adjacent entorhinal, perirhinal, and parahippocampal cortices, that make up the parahippocampal region (Suzuki and D. G. Amaral 1994; Eichenbaum, Yonelinas, and Ranganath 2007).

The salient feature of memory dysfunction following hippocampal lesions is a specific inability to remember events. At least patients with circumscribed hippocampal damage, however, typically have a normal vocabulary and their general knowledge of facts remains intact (Kensinger, Ullman, and Corkin 2001; Schmolck et al. 2002). The type of memory that involves personally experienced events was termed episodic memory by Tulving (1995), who contrasted it with other aspects of declarative memory, such as knowledge of facts independently of (memory for) the context in which they were learned (semantic memory) or context-independent recognition of stimuli based on a feeling of familiarity. Thus, the hippocampus appears to be necessary for framing events within their context in the initial formation of the corresponding memories.

Central to most of the theories focusing on the role of the hippocampus in retrograde amnesia is the notion expressed by Ribot more than 100 years ago, that the loss of memories from the past could be temporarily graded, i.e. more severe for more recently acquired memories (Ribot 1881). This phenomenon has influenced a number of accounts, which have ascribed to the hippocampus a more or less extended but nevertheless time-limited role in memory. Marr was the first to cast into a mathematical network model the idea that the hippocampus acts as a temporary memory store for the storage of new information (Marr 1971a), while the neo-cortex, which he treated in a separate network model, acts as a permanent memory store (Marr 1970).

In his model, later extended and clarified by many other investigators (B. L. McNaughton and R. G. M. Morris 1987; E. T. Rolls 1989) a sparse hippocampal representation of an event is rapidly encoded through the modification of recurrent connections in the CA3 area of the hippocampus. This representation affords subsequent retrieval by an incomplete cue through pattern completion. Marrs model expresses the hypothesis that the hippocampus is required for the initial encoding of multi-modal information that is represented in cortical areas: it provides a convergence zone to mediate these associations. Then, over a certain period of time, these memories are taken to become gradually independent of the hippocampus and to reside in neocortical storage sites. This is part of the so-called memory consolidation processes (Nadel and Moscovitch 1997). The time-scale of transfer, or perhaps more accurately the time for neocortical

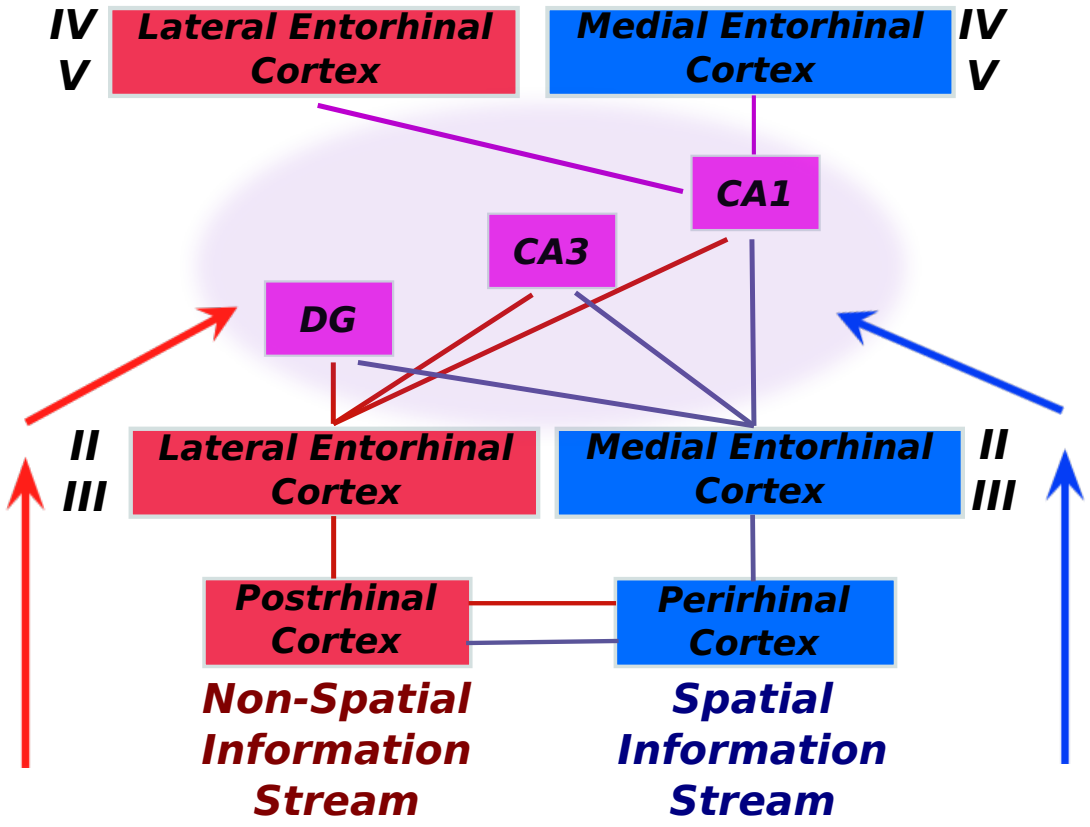


Figure 7.1: Sketch of the two streams of information converging in the hippocampus.

memory traces to organize and consolidate, is suggested by the majority of studies of retrograde amnesia in humans, that do find sparing of older memories, to be of the order of years or even decades. The slowness of this process poses a challenge to the simpler-minded view of this phenomenon, as if it were simply a case of pulling a memory item out of a display window, and putting it in permanent storage in a drawer in the back of the shop. Such a slow transfer, in fact, would have taken almost the entire lifetime of our prehistoric ancestors. If neocortical consolidation occurred so slowly, one wonders how useful it could have been for them. Moreover, there is also some conflicting evidence of hippocampally lesioned patients showing no gradient in their amnesia (Sanders and Warrington 1971; Cipolotti et al. 2001).

The link between the storage of a memory item in the hippocampus and the notion of "cell assembly", has been directly tested by a number of studies that used optogenetic tools to interact with hippocampal neurons during the formation of memories. In these studies the possibility was demonstrated of artificially inducing the formation of a new, spurious memory (Garner et al. 2012) or, even, of modifing an existing one (Ramirez et al. 2013). This approach has highlighted the fact that, by targeting a limited pool of identified units, one can affect the memory-based behavior of the animal, thus providing experimental evidence for the otherwise purely theoretical notion of cell assembly. Another confirm of the direct link between medial temporal lobe neurons and episodic memory, this time in humans, came from electrophysiological recordings in epileptic patients (Quiroga et al. 2005; Viskontas, Quiroga, and Fried 2009; Cerf et al. 2010).

7.4 Continuous or not? Attractors in the Medial Temporal Lobe

The involvement of the hippocampus in memory has drawn attention toward attractor dynamics(J. Knierim and K. Zhang 2012), i.e. the spontaneous tendency of an interacting system to approach one of a set of special configurations, the attractors (Amit 1995). In the hippocampus, particularly if implemented by the rich recurrent structure unique (within the mammalian hippocampus) to the CA3 region, it could be the key network mechanism underlying its involvement in memory (E. T. Rolls 1989), consistent with observations in CA1 (Wills, Lever, et al. 2005) and recent ones in CA3 (Jezek et al. 2011).

The notion of attractor was borrowed by neuroscience from the dynamical system vocabulary. In the simplest case of discrete fixed-point attractors, it identifies specific configurations which the system tends toward, if it is left to evolve. At least ideally, these states are stable, in the sense that once the system has reached them, in the absence of further perturbations, it will remain there for an indefinite time. To each of these attractors is associated a basin of attraction. This means that if the system is initialized in a configuration inside one of these basins, then it will, given a certain time, approach and eventually reach the corresponding attractor.

Donald Hebb, in his inspiring book (Hebb 1949), developed the similar notion of a cell assembly, as a group of cells whose coactivity represents something in the world, and provides a reference point toward which neighboring configurations of activity tend to evolve. The same idea found a precise formalization with the Hopfield model, in 1982 (Hopfield 1982). The notion of an attractor was slowly recognized as relevant to understanding memory functions (Daelli and Treves 2010). In particular, since one of its most relevant properties is pattern completion, to associative, content-addressable memories, in which the presentation of portion of the original pattern of activation leads to the retrieval of the entire representation.

7.4.1 Discrete and Continuous Attractors

The classical attractor formulation, however, was based on a finite number of attractors, neatly separated one from the other. This is the case of discrete attractors, which correspond to single memory items stored in the network. Other states, intermediate among these stable points, are not attractors, and the system may only go through them on its way to a final stable configuration. Applied to the brain, the notion of a stable point is of course an idealization, but one may still be able to observe convergence toward putative attracting states within psychological time scales of 100-200 ms (Akrami et al. 2009a).

In the case of a continuous attractor, instead, there is a multiplicity of quasi-stable attractors making up a continuous manifold. These attractors are quasi-stable in the sense that in the absence of perturbations the system can remain stable in one of them forever, but it takes an arbitrarily small perturbation to make the system move to a new attractor arbitrarily close to the first one, and perhaps continue gliding on the manifold. Each point on the attractor manifold is an equilibrium point but this equi-

librium is precarious, because other attractors are not separate and no barrier keeps them at bay. As a consequence, at least in the ideal notion of a continuous attractor, the system can be initialized in any of the points of the attractor manifold and be stable on it, or it can move effortlessly along the manifold. The manifold is typically low dimensional (Yoon et al. 2013), and of course each point on the manifold is attractive with respect to the many other dimensions of the (practically infinitely dimensional) configuration space.

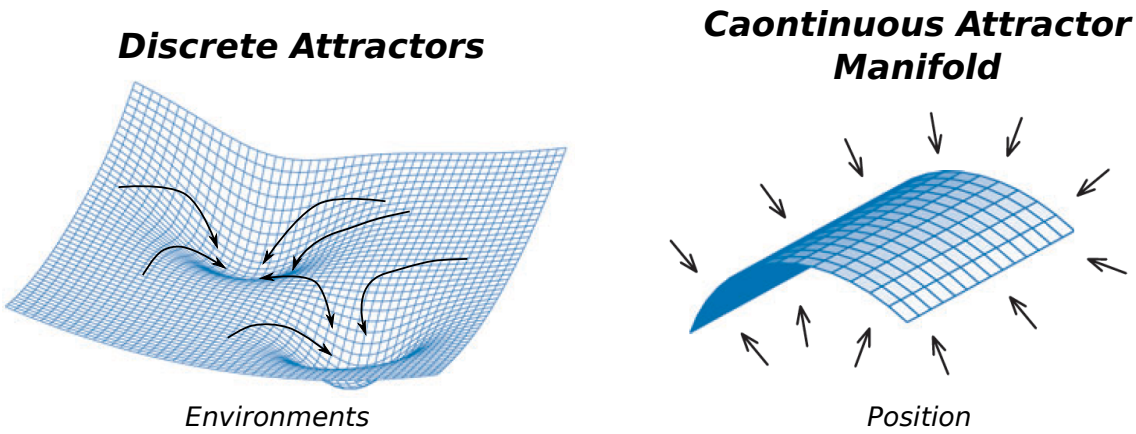


Figure 7.2: Pictorial illustration of the concepts of discrete attractors (2 in the landscape on the left) and of a single continuous attractor (right). The latter can be thought as exerting an attractive force towards the surface, but irrespective of the particular position on the plane manifold. Adapted from (J. Knierim and K. Zhang 2012)

The continuous attractor notion has been used in neuroscience to model the ability of certain brain structures to track continuously changing stimuli smoothly. Apart from models of sensory tuning, as in orientation selectivity (Ben-Yishai, Bar-Or, and Sompolinsky 1995) and in the head direction system (Xie, Hahnloser, and Seung 2002), which do not, in principle, require the storage of information in long term memory, this notion has been applied to the description of components of a spatial navigation system, including place cells (Tsodyks 1999; Romani and Tsodyks 2010) and grid cells (B. McNaughton, Battaglia, et al. 2006). In this case the quasi-stable attractors of the manifold are used to represent the different locations the animal can visit while it moves around in the environment. To each of these locations is associated an attractor, part of a continuous manifold isomorphic to the real physical space. Inasmuch as these animals are able to remember where they are when temporarily deprived of sensory (and proprioceptive) inputs, the stability of the continuum of fixed points is used for memory purposes, and not just the error correction or pattern completion capability implicit in the convergence to the attractor.

In 1997, Samsonovich and McNaughton showed how, with their multiple-chart model (Samsonovich and B. McNaughton 1997), one could conceive of fixed points organized in multiple two-dimensional continuous manifolds, each of which maps the animal position in a distinct environment. The attractor dynamics of a system with these characteristics thus follow two separate directions. First, the sensory cues of the envi-

ronment lead to the choice of one of the multiple charts stored in the network. Then, the position of the animal in the environment lead to the corresponding activation pattern of the network within that chart. Whatever the choice of the chart, the network is able to represent the position for that particular environment, making the two processes almost independent. While the switching between different charts is an abrupt phenomenon (Jezek et al. 2011; Stella and Treves 2011)(see Chapter 9), the dynamics internal to the single chart is a smooth process, in which the quasi-stability of the attractors allows small signals to gently displace the bump of activity representing animal position (Romani and Tsodyks 2010).

7.4.2 An Ideal too Far?

The notion of a really continuous attractor is of course an idealized limit case. To fully express the continuity of the attractors and their quasi-stability, one would need a network of infinite size. When endowed with a finite number of units, a network can at best approximate the conditions described above (Roudi and Treves 2008; Cerasti and Treves 2013). At most, network dynamics tend toward one of a large collection of discrete attractors, which retain their point- like nature but with little separation between each other, as they are densely distributed on the (quasi-continuous) attractor manifold. When the system is driven by its internal dynamics, then, not all positions in space can be represented with absolute precision. Rather, the discretized attractor manifold defines families of positions represented, effectively, by the same pattern of activity. The resolution of a chart can be higher or lower depending on the number of units comprising the system. When the external input is introduced, and it is dominant over the internal connections, the continuity is retrieved (Kali and Dayan 2000) even if (a) the roughness of the attractor manifold is still reflected in a dishomogeneity of the correlations among population vectors associated to different positions in space; and (b) the memory capability of the attracting manifold are not really utilized - the network merely represents, or recodes, what it currently receives. This lack of precision in representing spatial continuity with a recurrently connected neural network was first noted by Tsodyks and Sejnowski (1995b) in a one-dimensional model and then the result was extended to the two-dimensional case by Brunel and Trullier (1998) and, in a slightly different context, by Roudi and Treves (2008).

Therefore, a continuous attractor is an idealization in at least two distinct ways. First, stable attractor configurations are reached only asymptotically, and what is relevant over plausible time scales is transient dynamics toward such states, not the states themselves; second, such stable states can be conceptualized as forming a continuum, but in fact in any physical implementation they are a bunch of discrete attracting points. Moreover, one should consider that even when close to the ideal case, e.g. with very large networks, a problem with the accuracy of the representation arises again if the decoding of the position can only rely on a limited number of units. Sampling only a limited portion of the total population vector produces anyway distortions in the charting of space.

These considerations sound like a serious limit to the actual capability of a network with the characteristics of CA3 to accurately represent the spatial position of an animal in the environment. Furthermore, this very same spatial position has already been

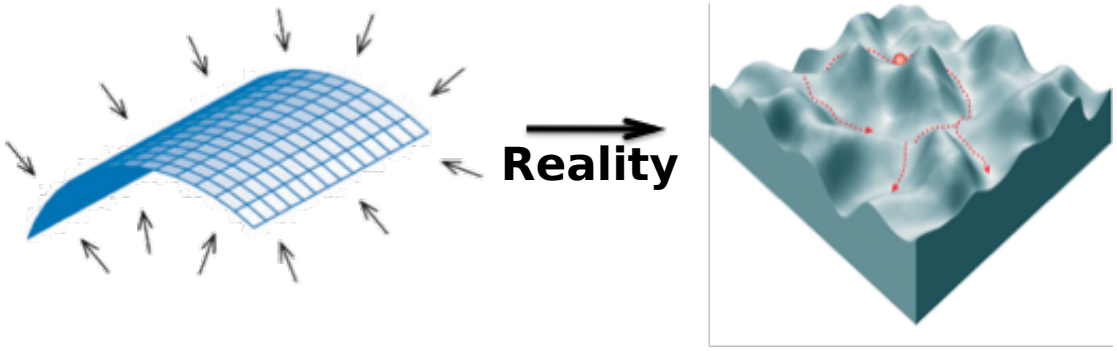


Figure 7.3: Actual implementations of a continuous attractor network can achieve only some approximation of that ideal concept. What they might produce is a collection of multiple quasi-stable points distributed over space. Adapted from Macarthur, Ma'ayan, and Lemischka (2009)

computed somewhere upstream of the hippocampus, and it is represented, as discussed above, in the remarkably beautiful and precise grid representation of medial entorhinal cortex. It seems reasonable to conceive of CA3 as specialized for something else than encoding again spatial relations, e.g. for utilizing them to construct complex memories (E. Rolls 1996; Eichenbaum, Dudchenko, et al. 1999; Cerasti and Treves 2013). If spatial computation per se is not the prerogative of hippocampus, it may well be that in the construction of memory representations with a spatial component a certain loss of spatial precision is acceptable, if compensated by the capability to distinguish the current environment from other similar-looking ones, that would lead to confusion between different memory items.

7.5 Information Measures

As we have already mentioned, there is significant evidence supporting the idea of a dissociation between the mEC and the hippocampus, in how well they represent contextual and spatial information. Nevertheless, spatial representations (in terms of place cells) are found also in the hippocampus. How do they differ? In a neural context, a quantitative approach, to address this sort of issues, is provided by information theory.

Why information theory? We can analyze brain processing by understanding what is encoded by the neuronal firing at each stage of the brain, and determine how what is encoded changes from stage to stage (E. Rolls and Treves 2011). Each neuron responds differently to a set of stimuli, and it is this that allows different stimuli to be represented. We can only address the richness of the representation therefore by understanding the differences in the responses of different neurons, and the impact that this has on the amount of information that is encoded. These issues can only be adequately and directly addressed at the level of the activity of single neurons and of populations of single neurons, and understanding at this neuronal level is essential for understanding brain computation (E. Rolls and Treves 2011). Information measures, provide the only correct way to relate representations by populations of individual real neurons to their counterparts in a given model.

In the previous sections we have summarily looked at the experimental data available for the hippocampus, and outlined the dichotomy between those who consider the hippocampus as the storage site of episodic memories and those who consider it important for spatial navigation. A challenge for current research on the hippocampus is to fill this gap and to see whether these two different perspectives may fit together (Buzsaki and E. Moser 2013; Eichenbaum 2013). To meet this challenge, a clear quantitative measure of what is encoded in a certain neural representation is essential. What hippocampal neurons are doing is less of an issue than how much they are doing it.

Information theory provides the means for quantifying how much neurons communicate to other neurons. The fundamental quantity to measure the amount of information contained in a given activity relative to some external correlate, or to the activity of another population, is the mutual information:

$$< I(s,r) >= \sum_{s,r} P(s,r) \log_2 \frac{P(s,r)}{P(s)P(r)} \tag{7.1}$$

where $P(s,r)$ denotes the joint probability of the pair of events s and r. For example, r may be the response of a population of neurons to stimulus s.

7.6 Coding Precision of the Grid Cells Code

To quantitatively investigate the coding capacity of the grid cells representation, we apply information measures. We base our model of grid cells representation on the one developed by Kropff and Treves (2008) and Si, Kropff, and Treves (2012), and described above, although alternative models for the generation of the grid pattern would yield very similar results for its coding capacity. We thus run the model to produce a system of grid units, aligned with each other and phase-shifted like the real ones. We use a population code comprised of 256 cells with a given spacing and orientation, and random phases, to compute the spatial information that can be extracted from the observation of the activity of this ensemble of virtual grid cells, during the simulation of exploratory behavior in a typical recording box.

First, we construct templates of the activity of the network in different regions of the environment, restricting ourselves to an 8x8-bin square spanning the basic periodic cell of the grid representation (for a more comprehensive methodological description, see Chapter 10). Then, to calculate the spatial information conveyed by the code, we sample subsets of units of various sizes and through a decoding algorithm we compare the decoded position, obtained by comparing the activity of the units in the sample with the activity of the same units in the different templates, with the position actually occupied by the virtual rat. We therefore construct a confusion matrix, representing the probability of a location in the environment to be decoded as belonging to any of the spatial bins in which we have divided the basic periodic unit of the representation. Then we can compute the mutual information between the real and the decoded position, contained in the values of the confusion matrix.

As expected, the spatial code expressed by our model mEC network turns out to be excellent (Fig. 7.4, left). The amount of retrievable spatial information quickly saturates close to the maximum value of 6 bits (having used 8x8 = 64 different spatial

bins, the maximum amount of information is log2(64) = 6). The information is very high even for small samples of units, and it rises above 5 bits already for 16 cells. We fit the information values for different sample sizes to exponentially saturating curves, in order to get the values of the two most relevant parameters that describe their shape: the initial slope I_{slope} (i.e. the average information conveyed by the activity of individual units) and the total amount of information $I_{\max}$ (i.e. the asymptotic saturation value). The function we used for the fit is the following

$$I = I_{\max}(1 - e^{-N\frac{I_{\text{slope}}}{I_{\max}}}) \tag{7.2}$$

where N is the number of cells in the sample. From the simulated mEC data in Fig. ?, we obtain $I_{\max} \sim 5.8$ bits and $I_{\text{slope}} \sim 1.2$ bits.

Our simulations support the idea, already expressed in other studies (Fiete, Burak, and Brookings 2008; Sreenivasan and Fiete 2011; Mathis, Herz, and Stemmler 2012a; Mathis, Herz, and Stemmler 2012b), that the grid cell code is particularly accurate in localizing the animal in space.

7.7 mEC and Context Discrimination

The rigid arrangement of the grid fields of multiple units is very well suited to precisely code for the position of the animal in a single environment, but this very same structure obstructs discrimination between two or more different environments. Intuitively, the multiplicity of grid fields, and their regular disposition, imply that (after a common rotation and a shift) the representation will preserve a high degree of self similarity. Evidence supporting this intuition comes from the experiment of Fyhn, Hafting, Treves, et al. (2007). They recorded simultaneously from mEC and CA3, and calculated for both regions the mean mutual information between recording session and distribution of population vectors. In mEC, the mutual information was not larger when comparing sessions in different boxes than in the same box, indicating that grid cells that are coactive in one context remain coactive in others, too, and the coactivity patterns are unable to discriminate between boxes. Therefore coactivity patterns are largely overlapping, or covering roughly the same manifold in the abstract space spanned by mEC population activity. The same study shows that the situation is very different in CA3, where the degree of similarity between the activity patterns in different recording boxes is much below the similarity between different sessions in the same box.

We then looked to our model of grid cells to have a statistically more reliable measure of the context sensitivity of mEC activity. To replicate the remapping (i.e. shifting and rotating) mechanism of the grid cells, we took a grid cell representation and we coherently rotated and shifted all individual grids, as observed to occur in the real data (Fyhn, Hafting, Treves, et al. 2007) and in the model of grid formation we use (Si, Kropff, and Treves 2012). In this way we produced another map, representing another environment of the same shape and dimension. We then generated again templates, as well as random virtual rat trajectories in both environments. To determine the information contained in the units activity, about which environment the (virtual) animal is in, we ran an identical analysis as described in the previous spatial case. However, this time it was not the position that was decoded, rather the comparison between the activity during the test phase with the templates was used to determine to which of

the two environments the most likely template belonged. We thus extracted contextual information, quantifying how well the two environments can be discriminated on the basis of the observation of network activity. The results support the notion of the grid cells code as nearly context-independent (Fig 7.4 right). The contextual information is very small, almost zero, and it has a weak dependence on the sample size. The initial slope of the curve that best fits the data points is $I_{\text{slope}} \sim 0.001$ bits.

It seems that the scenario of two separate streams of inputs, one for spatial information, the other for non-spatial/contextual information (Hunsaker et al. 2007), is consistent with our analysis of a simulated mEC code. If these two streams were to converge in the hippocampus (Kesner 2007), then we would expect to see something different in the way spatial and contextual information are expressed there.

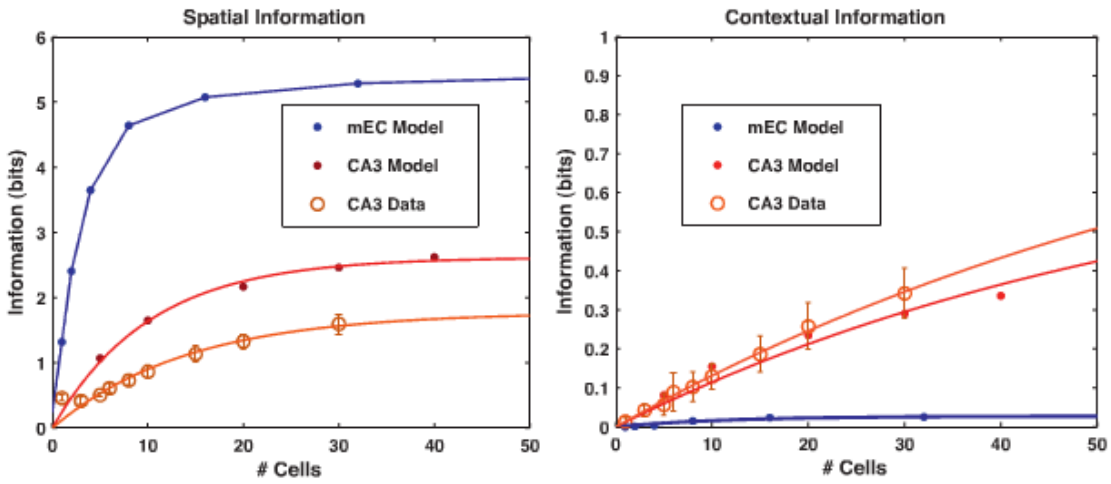


Figure 7.4: Comparison of spatial (left) and contextual (right) information content of the grid and place cells representations, for cell samples of increasing size. Simulation results for mEC (blue dots) and CA3 (red dots) are shown together with experimental data from CA3 recordings (orange circles). CA3 data points are fitted by the corresponding curves (in the case of mEC, just connecting the dots).

7.8 Information in CA3

As done already above for mEC, we apply the tools of information theory to quantitatively assess the nature of the CA3 representations. Again, we formulate the problem in terms of spatial and contextual information. The latter cannot arrive to the hippocampus through the projections from mEC, as it is not expressed there. So any non-vanishing contextual information should be ascribed to the action of some other converging input that is integrated in the hippocampal representation (J. J. Knierim, I. Lee, and Hargreaves 2006; Deshmukh and J. Knierim 2011; Yoganarasimha, Rao, and J. Knierim 2011; Deshmukh, J. Johnson, and J. Knierim 2012). This is of course fundamental to any notion of the hippocampus as the generator of episodic memories, or at least to its role in providing a spatio-temporal context to such memories .

We have applied information measures both to experimental data and to computer

simulations of the activity of the CA3 network. The experimental data was obtained by Karel Jezek in Trondheim by recording the activity of CA3 cells in rats free-foraging in two different environments, as baseline data for the teleportation experiment (Jezek et al. 2011). The training had been arranged such that the two charts established for the two environments involved global remapping, and not rate remapping, from one to the other (S. Leutgeb, J. Leutgeb, Barnes, et al. 2005). The recordings were made in CA3. (Note that analogous recordings from mEC do not typically include the number of units and the coverage statistics required to utilize information measures, even though they appear consistent with the simulation results reported in Fig 7.4).

The simulations were performed based on a model of the DG-CA3 network (Cerasti and Treves 2013) describing the flow of information, conveyed through the mossy fiber, during the storage of new spatial representations in CA3. The original model (Cerasti and Treves 2010) was extended to describe reverberations along recurrent connections and the storage of multiple environments.

Fig 7.4 shows that, as expected, both the simulations and the experiment, even if with some difference between them, show a drastic change with respect to the mEC case. CA3 expresses a significant amount of contextual information, at the population level (Fig 7.4 right; recordings: $I_{\max} \sim 1.02$ bits; $I_{\text{slope}} \sim 0.01$ bits; simulations: $I_{\max} \sim 0.76$ bits; $I_{\text{slope}} \sim 0.01$ bits). This major improvement at representing context can be seen to be not only quantitative but also qualitative, as it marks the passage from an almost complete inability of mEC, reflected by vanishing information, to a discriminatory capacity which grows monotonically with the size of the sampled population. This achievement does not entail a total inability by CA3 to distinguish locations within the chosen environment. What we observe is a partial loss of spatial information with respect to the strong performance by mEC (Fig 7.4; recordings: $I_{\max} \sim 1.78$ bits; $I_{\text{slope}} \sim 0.1$ bits; simulations: $I_{\max} \sim 2.62$ bits; $I_{\text{slope}} \sim 0.2$ bits). The CA3 code is less precise and it generates more confounds about different positions in the same environment. But this is possibly a necessary trade-off to efficiently represent differences between distinct environments (Fig **??**).

7.9 The Next Sections

Attractor dynamics is not an exclusive feature of CA3. Even in the restricted domain of spatial representations, attractors are not a prerogative of place cells representations. Low-dimensional dynamics, a typical signature of attractors, have been very recently described in the mEC (Yoon et al. 2013), demonstrating the variety of phenomena that can be ascribed to this network mechanism. But the fact is that the same unifying principle can lead to very different outcomes when applied in its slightly differing variants, outcomes as different as the single, rigid, grid cell representation and the multitude of flexible, place cell representations.
Convergent agreement between experiments (J. Leutgeb, S. Leutgeb, M. Moser, et al. 2007; Bakker et al. 2008; Lacy et al. 2011; Burghardt et al. 2012; A. Morris et al. 2012; Azab, S. Stark, and C. Stark 2013; Deng, Mayford, and Gage 2013) and theoretical work (Treves and E. Rolls 1992; Si and Treves 2009; Almeida, Idiart, and Lisman 2010; Renno-Costa, Lisman, and Verschure 2010; Monaco and Abbott 2011) indicates in the DG the origin of the highly differentiated hippocampal spatial representations. But

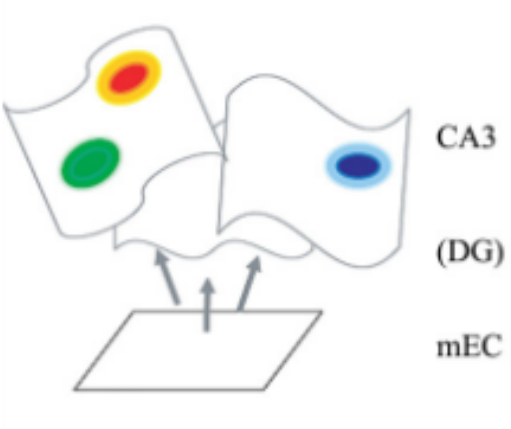


Figure 7.5: Pictorial representation of DG as a spatial random number generator/pattern separator. By passing through DG, grid maps that span the same continuous manifold in mEC are recoded as fully decorrelated CA3 charts, which can be stored as attractors in memory.

once these representations are in place, CA3 is not expected to simply relay forward the external input, but to actively retrieve the stored information. In the next sections we will dwell on the retrieval of spatial maps from the network, and on the way in which the external input and the recurrent collaterals interact during this process.

Chapter 8

Attractor Dynamics in Ca3

8.1 Dissecting the Problem

In this chapter we extend the analysis of a similar model, presented in Cerasti and Treves (2013). We address the same kind of problem, that is, the nature of the attractor landscape effectively stored in a recurrent network like CA3 (Papp, Witter, and Treves 2007). But instead of focusing on the learning process and on the DG role during the storage of the maps, as in the paper, we take a different slant and discuss how the different elements of the representation determine the properties of the neural dynamics.

To do so, we limit the complexity of the model initially used in Cerasti and Treves (2013), and reduce it to a simplified toy model of CA3 spatial representations, in which we are able to manipulate each element of the network, such as the connectivity and the positions of the place fields. To understand the degree of approximation to a continuous attractor that can be achieved by networks with the characteristics of those found in CA3, we start from an ideal case and we then perturb it, to model some of the constraints of a real system. Indeed, while often models of continuous attractor are based on full connectivity and a regular arrangement of fields (but see, (Renart, Song, and X.-J. Wang 2003; Itskov, Hansel, and Tsodyks 2011)), neither of the two is found in the hippocampus: although CA3 is known for the high degree of internal connectivity this is far from being full, and, as well known from experiments (Wilson and B. McNaughton 1993), the place fields arrangement shows no regularity. On the contrary, their disposition looks almost random, apart from some relationship observed with goal locations (Hollup et al. 2001; Hok et al. 2007; Dupret, Joseph, Barty, et al. 2010).

8.2 The Model

The model we use is mostly a simplified version of the one described in Cerasti and Treves (2013), scaled down to retain only the elements pertaining to memory retrieval and recurrently driven dynamics. It simulates a representation of a 1x1m square environment. It is based on threshold-linear units. The firing rate of N units, η_i, is calculated at discretized intervals in time, with the following equation:

$$\eta_i(t) = g(t) \left[\text{Ext}(\vec{x_0}) + \sum_j W_{ij}\eta_j(t-1) - T(t) \right]^+ \tag{8.1}$$

where $\mathrm{Ext}(\vec{x})$ is a an external activation current used to initialize the system (see below), $\vec{x}$ is the position in the environment and $[\star]^+$ is the Heaviside function. g and T are two parameters adjusted at every time step to regulate the average activity of the system. In particular g fixes the mean firing rate $\sum_i \eta_i/N = 0.1$, while the threshold T is used to fix the sparsity level of the network $a = (\sum_i \eta_i)^2/\sum_i \eta_i^2$.

The connectivity is pre-wired, defined at the start of each simulation. We imagine that plastic modifications of the connections have already taken place during some previous learning session, used to deposit a map in the network. The weights thus depend now only on the distance between the fields of the units, according to an exponential function

$$W_{ij} = \mathrm{C}exp(-\frac{(\vec{x_{i0}} - \vec{x_{j0}})^2}{2s^2}) \tag{8.2}$$

where $\vec{x}_{i0}$ is the position in the environment assigned to the unit place field of unit i, and $s = 5cm$.

8.2.1 Simulations

Generation of the Maps We start from an ideal and perfectly regular map: each unit is assigned a single place field, all the place fields are arranged on a square lattice of size NxN (the size of the network) and the network is fully connected. This produces a quasi-continuous attractor state. We expect this symmetric configuration of the network to accurately represent the environment. Errors should only be due to finite-size effects and be quickly negligible when increasing the size of the network. We then start to introduce disorder to break the translational invariance of the system and to study the induced deviations.

This disorder is of two types:

- The random deletion of an increasing number of connections
- The random displacement of the place field centers from the vertices of the starting lattice

For the first type we simulate conditions in which the 20%,40%,60% and 80% of the connections have been eliminated from the entire network. The deletion is always symmetrical, affecting both the connection from unit i to unit j, and the reciprocal one, from j to i. For the second type we reassign the field center of each of the unit, drawing the new coordinates from a Gaussian distribution centered in the old location of the field. Different conditions correspond to increasing the width σ of the distribution, from $\sigma = 1$cm to $\sigma = 100$cm. Since we are interested in structural deviations from the ideal attractor network, we will neglect the effect of fast noise on the functioning of the system. The dynamics are completely deterministic in all the simulations and are only determined by the weight matrix, without any influence of stochastic components of the dynamics.

Once the map is defined, we have the configuration of unit activity assigned to each point in space. We construct a 100x100 binning of space, associating to each bin the activity of the units expected from the assigned place field arrangement. We will use these "templates" to reconstruct the network dynamics during the simulations. Note

that for the templates, we use the same constraints on mean firing rate and sparsity used during the simulations.

Dynamics We then select 100 positions of the environment and we start 100 separate trials by initializing the system in the expected configuration for each of these positions. This initialization, working effectively as an external input cueing the retrieval of the stored representation, is kept active for a limited number of time steps in each trial, namely 8 time steps, while its strength $\text{Ext}(\vec{x_0})$ (see equation 8.1 is gradually decreased by 10% after each iteration. It thus becomes ineffective in an even shorter time, after which the update of the units is entirely determined by their own activity reverberating through the recurrent connectivity.

The total length of a trial is 40 time steps. Since this is a toy model with very simplified neural equations, there is no direct correspondence between time in the simulations and real neural time evolution. Still, the duration of the time bin should be sufficient for the propagation of a signal from one neuron to its neighbors. One can think of it as being roughly of the order of 10ms, with a trial corresponding to more or less 400ms of simulated neural dynamics. During this time the activity of the network evolves with no external influence. In the ideal case of the perfect continuous attractor, one expects that after having initialized the system in a given location one could come back after 40 time steps and find the activity configuration almost unchanged, still representing the initial position. But since this might not be the case in our simulations, we can track the evolution of the system by comparing its configuration at each time step with those initially collected as "templates". By choosing one of these templates according to some distance measure (in our case the minimal Euclidean distance between the two population vectors) one can thus decode the activity and reconstruct the represented position, in time. Of course, the real activity of the network will not be completely identical to any of these templates. One is projecting the position of the system in a very highly dimensional activity space, on a two-dimensional sheet corresponding to the stored map. One of course assumes that the distance of the real configuration from this plane is not large, and as we will see later in this chapter, this condition actually holds, at least when only a limited number of maps are stored in the network.

Analysis As already done in the paper by Cerasti and Treves (2013), we can describe the distribution in space of the 100 positions represented by the system at the end of the trials, by means of a combination of three measures (Figure 8.1 8.2 8.3). First, we can count how many different, distinct final positions are actually being represented(Figure 8.1 8.2 8.3, left). Of course, 100 is the maximum and the optimal condition, in which to each initial seed corresponds an unique final configuration. A number below 100 means there is convergence of multiple trajectories towards the same final position.

Then we can ask what is the degree of clustering of the trajectories (Figure 8.1 8.2 8.3, center). Clustering can be calculated using the score:

$$Clu = \frac{1}{S(S-1)} \sum_{k \neq l}^{S(S-1)} e^{-(\vec{x}_k - \vec{x}_l)^2} \tag{8.3}$$

where S is the number of initial seeds, 100 in our case, and $\vec{x}_k$ is the decoded position

in space. A low clustering score means that the points are evenly distributed on the environment.

Finally, and perhaps most importantly, we can address directly the issue of stability of the representation. That is, we can calculate how much is the distance between the position represented by the network after relaxation and the one initially assigned (Figure 8.1 8.2 8.3, right). We can thus evaluate what is the distance "travelled" by the system in the absence of an external input, the so called "drift" (Cerasti and Treves 2013). Again, a low score is what is expected from a continuous attractor, while the presence of only a limited set of stable configurations is reflected in a growing degree of displacement of the activity.

8.3 Results

By varying independently the connectivity dilution, the disorder of the fields and the size of the network, we can appreciate the effects on the three measures of each one of these factors. As expected, networks of any size behave very similarly, and very well, when no noise is present, showing no or very little drift. Not surprisingly, small networks (blue lines) are strongly affected when noise is introduced, even in small quantities. What is apparent, however, is that in the case of connectivity dilution (Fig 8.1), the increase in the size of the network strongly reduces these effects. Indeed, big networks (red lines) show no clustering and full score in the final positions, even when only a small fraction of the connections are still in place. Even though some drift is still present ($<$ 5cm) (Fig8.1, right) it seems that it has a linear dependance on the total number of connections only, irrespective of the dilution rate, and that it can thus be rapidly reabsorbed by slightly bigger networks.

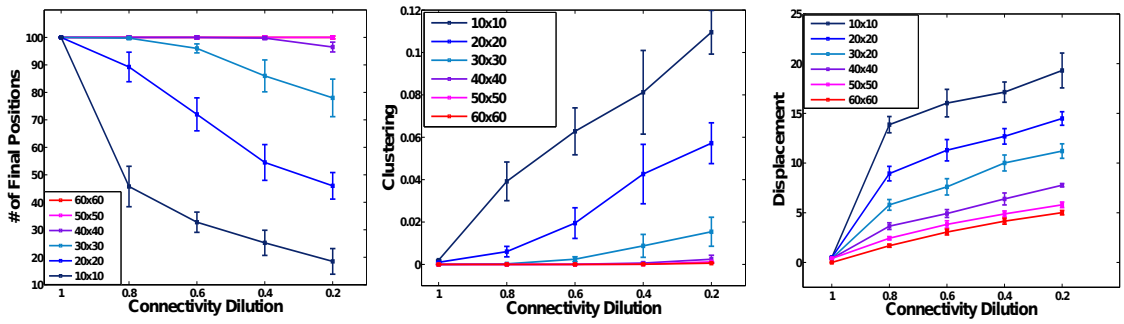


Figure 8.1: Final positions, clustering and displacement with diluted connectivity for different network sizes. On the x axis, the percentage of surviving connections.

On the contrary, the effect of place fields disorder is much more dramatic. Not only randomly arranged fields ($\sigma = 10$ Fig 8.2) produce non zero clustering and a restriction on the number of final positions even in large networks, but the degree of displacement is much larger ($>$ 10 cm, Fig 8.2,right) and does not appear to vanish at a significant rate with the increase in the number of units (Fig 8.3).

Above a certain number of units ($\sim 4000-5000$ units) the effects of size gets very small for what matters drift, suggesting that the complete removal of drift might require a

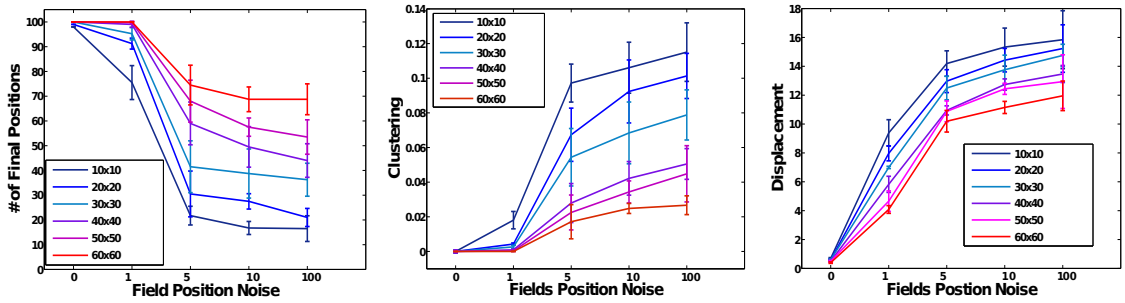


Figure 8.2: Final positions, clustering and displacement with disordered place field centers for different network size. On the x axis the standard deviation of the field centers displacement, in cm.

number of neurons above to that available in a real rodent hippocampus (D. Amaral, Ishizuka, and Claiborne 1990), or, even, in a human hippocampus. What size seems to affect, instead, is the speed at which drift takes place, as shown in Fig8.3, right, representing the amount of displacement after different time steps. Small networks drift further earlier, to then saturate at some maximum drift amount, that does not depend significantly on the size of the network. The modulation of speed caused by different network size, has been also discussed in (Itskov, Hansel, and Tsodyks 2011), in a different context, namely in the case of a one dimensional bump of activity subject to random, stochastic noise.

In Figure 8.5, left, we explicitly plot the speed of the drift as a function of the elapsed time from the beginning of the trial. Different curves correspond to different combinations of size and noise. On the one hand, the plot shows how over the span of the 40 time steps, the drift has almost completely run out. Speed approaches zero very fast in all the conditions. Actually, most of the distance is travelled in the first half of the trial, 20 time steps, after which minor adjustments take place. On the other hand, while all the curves share a similar general behavior, there is a marked difference in the first few time steps among networks of different size, as the smaller ones express much higher speeds, at least in a limited time window. The feeling is that smaller networks are "rolling down a steeper hill", a figurative expression that could be justified when we introduce energy to describe the dynamics.

8.4 Energy

The aim of the simplified model with symmetrical interactions is to be able to define an energy for the states of the system (Treves 1990) and to use it to try to formulate a synthetic description of the dynamics of the system. Given the equations of motion chosen for our simulations 8.1, we define the energy as:

$$E(\vec{x}) = -\frac{1}{2}\sum_{ij}\frac{W_{ij}}{C}\eta_i(\vec{x})\eta_j(\vec{x}) + \sum_i \eta_i(\vec{x})T(\vec{x}) + \frac{\sum_i \eta_i^2(\vec{x})}{2g(\vec{x})} \tag{8.4}$$

where C is the average number of connections per unit. The energy in this form is an extensive quantity in the number of units in the network. It is thus sufficient to

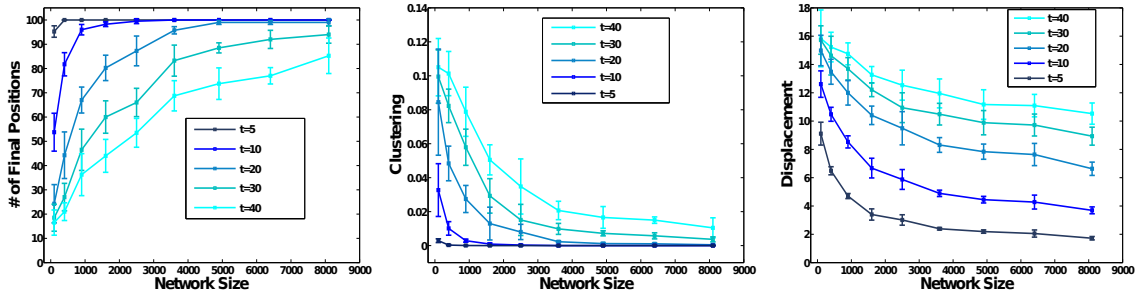


Figure 8.3: Final positions, clustering and displacement with disordered place field centers for different extensions of time. On the x axis the total size of the network. Different curves correspond to different time steps, from dark blue to green.

normalize it by N to directly compare results for networks of different size.

We can then calculate the energy value for each of the 100x100 bins of a map, and obtain an energy landscape of the environment. By superimposing on this landscape the trajectories taken by the activity of the network during drift (Fig8.4), one can see that the system appears to travel along directions leading it to local or global minima of the energy.

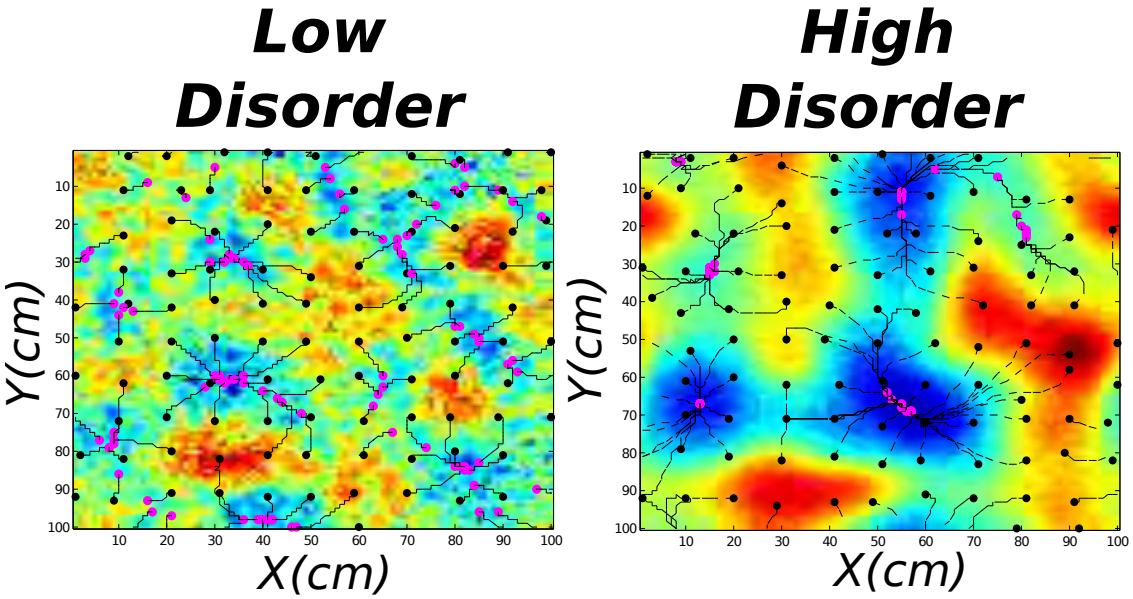


Figure 8.4: Energy Landscapes with low and high disorder. The dynamics of the simulations are overlaid. Black dots: starting points, purple dots: ending points, black traces: trajectories.

Figure 8.5, right, shows the variation in time of the energy associated to the decoded positions during the evolution of the system with respect to the energy with which it was initialized. The negative values of the curves indicate that the dynamics of the system (even though it is a projection of the real dynamics on a two-dimensional man-

ifold) aims at minimizing equation 8.4. The time evolution is similar to the plot on the right, with most of the drop in energy taking place in the first half of the time interval. In this case however, the different conditions do not converge to the same final value. Smaller networks exhibit a larger reduction of the energy, while for progressively larger networks the steepness of the curve is reduced.

This difference might be attributed to certain properties of the energy landscape depending on the number of units. In Figure 8.6 we then plot the values of the maximum and the minimum values of the energy as a function of network size. In the absence of noise the energies are all very similar, and the difference between its maximum and minimum value in each network is almost zero, exhibiting the (quasi-)flatness of a (quasi-)continuous attractor energy landscape. As we increase the noise the gap between the extremes in the energy values widens. The difference is more relevant for smaller networks, which have also on average higher values of the energy. The random distribution of a fewer number of elements generates larger deviations from the mean, a fact that can explain the larger extension of energy values in small networks of randomly arranged units.

We have also compared the energy values of the stored representations to the ones that would be associated to a novel, random spatial representation that is presented to the network, with no modification of the synaptic strengths. To do this we calculated the same energy for a number of versions of the same network where the positions of the fields is shuffled, keeping constant the underlying connectivity structure. The decoupling of the fields positions and of the connections produces higher values of the energy. Interestingly, even in the presence of high levels of disorder, the two distributions never overlap. The activity patterns associated to familiar locations are always energetically favored with respect to novel patterns of activation, even those sitting on the top of the hills of the energy landscape.

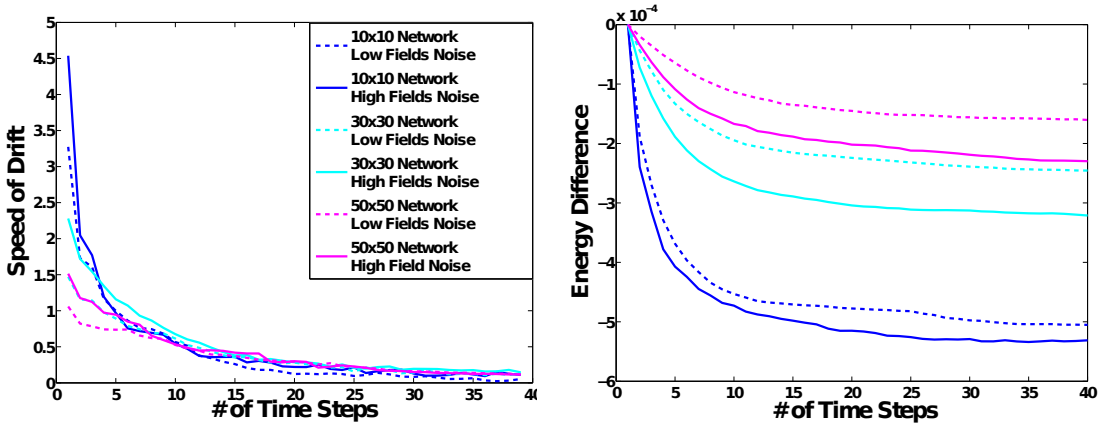


Figure 8.5: Time dependence of speed and energy during drifting.

8.5 Drifting Towards Crowded Locations

In the previous section we have seen how the role of place fields disorder is dominant in determining the behavior of the system during recurrent driven dynamics. Can we

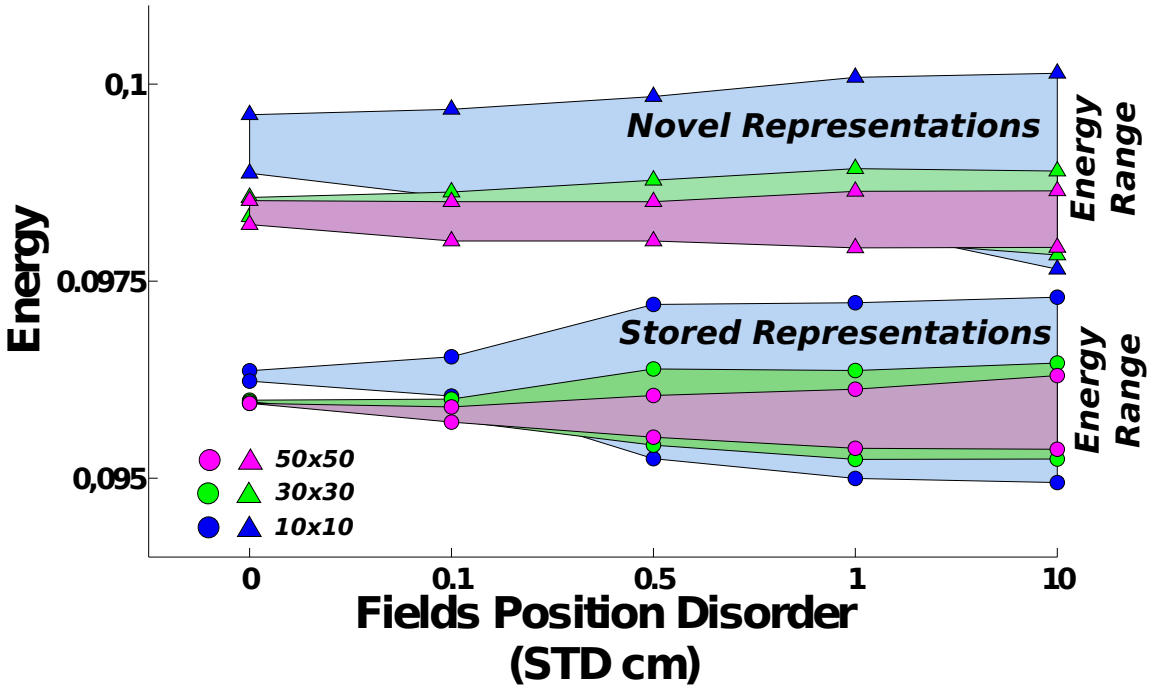


Figure 8.6: Dots: Maximum and minimum value of the energy of stored representations for different degrees of disorder. Colors: network size. Triangles: same but for non-stored representation

characterize this dependence more precisely? By moving the place fields away from the regular lattice, one is inevitably introducing some non-homogeneity in their distribution. While on the lattice the density of fields is constant, a random arrangement results in non-uniform values across space. The degree of variability of these values is a signature of the break in the symmetry of the regular lattice.

By comparing the energy landscape with the distribution of place fields density in space (for example like in Figure 8.7), one can notice the strong correlation between the position of the throughs of the energy and the of peaks of the density of fields (in the figure the color code of the energy is inverted, to emphasize the similarities). Of course the correlation can be quantified by comparing the relative values of density and energy across space. If we plot the results for one case (30x30 network, $\sigma = 10$ cm Fig 8.8), the presence of a strong anti-correlation is evident, and indeed the correlation score is -0.77 (the correlation scores for the different conditions are all very similar).

Locations happening to be associated to a higher number of fields are thus most probably corresponding to the stable configurations of the network. It is well possible therefore, that over-represented locations are also preferentially selected as the ending points of the drift dynamics. To check for this possibility we compared the values of the energy and of the fields density of the drift final locations, to the mean energy and density values respectively (Fig 8.9). We find that indeed, trajectories are attracted by positions of the environment that offer a combination of low energy values, and high number of place field centers in the surrounding.

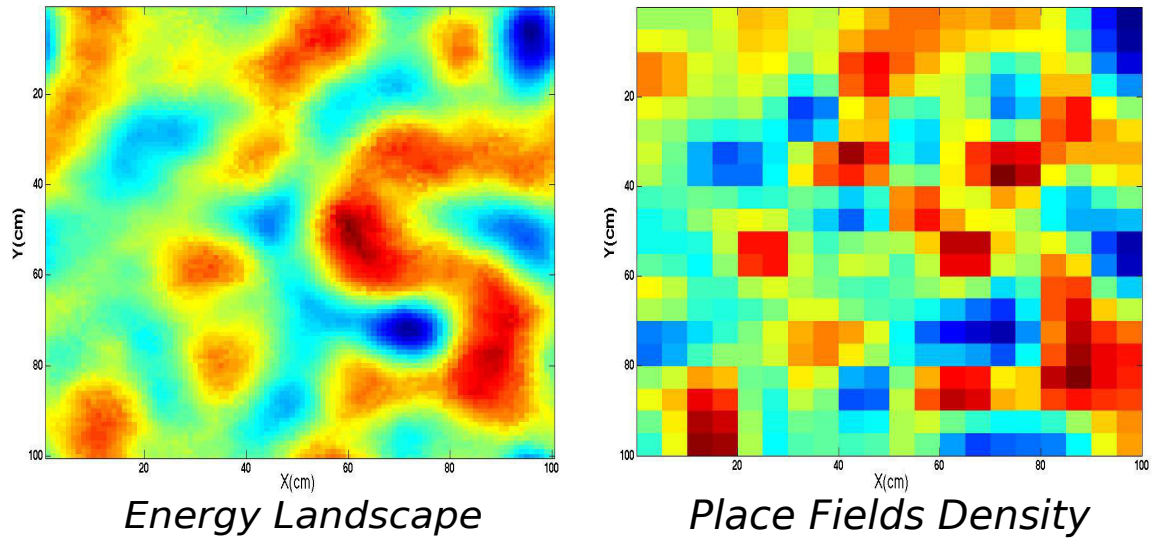


Figure 8.7: Comparison of an energy landscape with the corresponding spatial distribution of fields density. To visualize the similarities, the color code for energy has been inverted.

In summary, the arrangement of fields appears to be the dominant factor in determining the energy landscape of the manifold and thus of the dynamics of the system. Naturally, it might be related to the particular implementation of our model, which indeed is very far from the real biological system it is trying to describe. In Appendix A, we provide an alternative, more theoretical approach to the problem of the stability of the activity in the presence of disordered place fields, which leads to results consistent with those presented in this chapter.

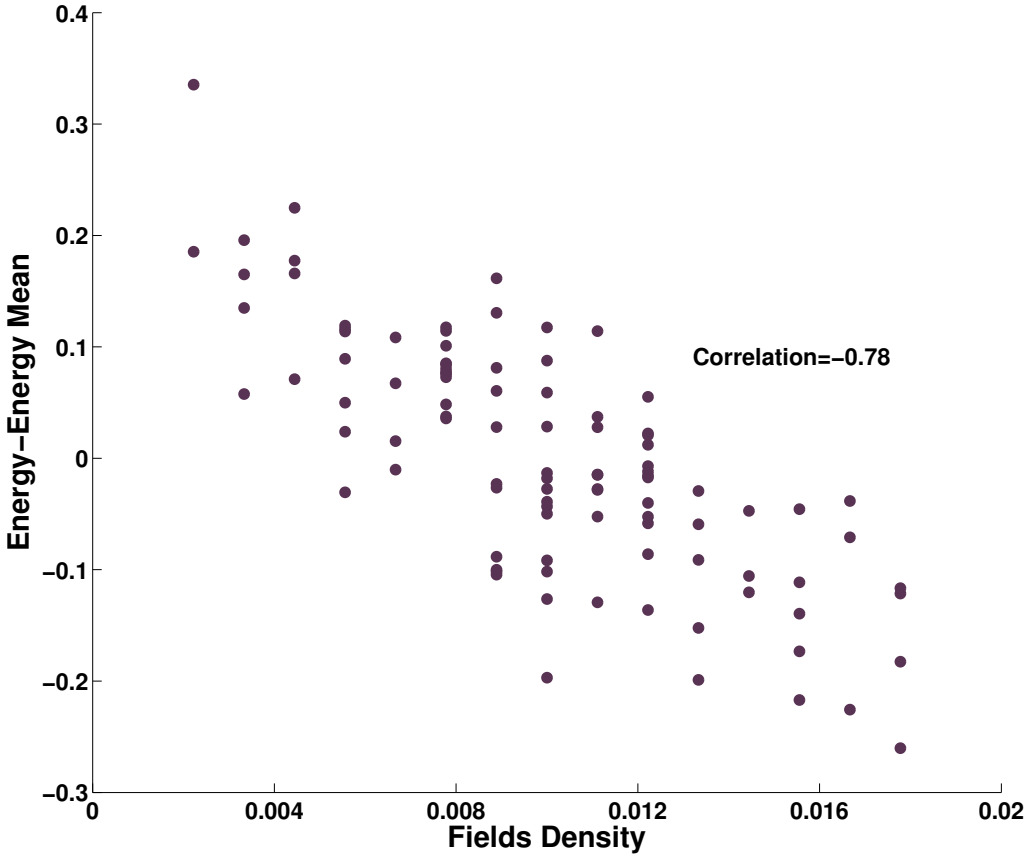


Figure 8.8: Relationship between the energy value of a location and its fields density

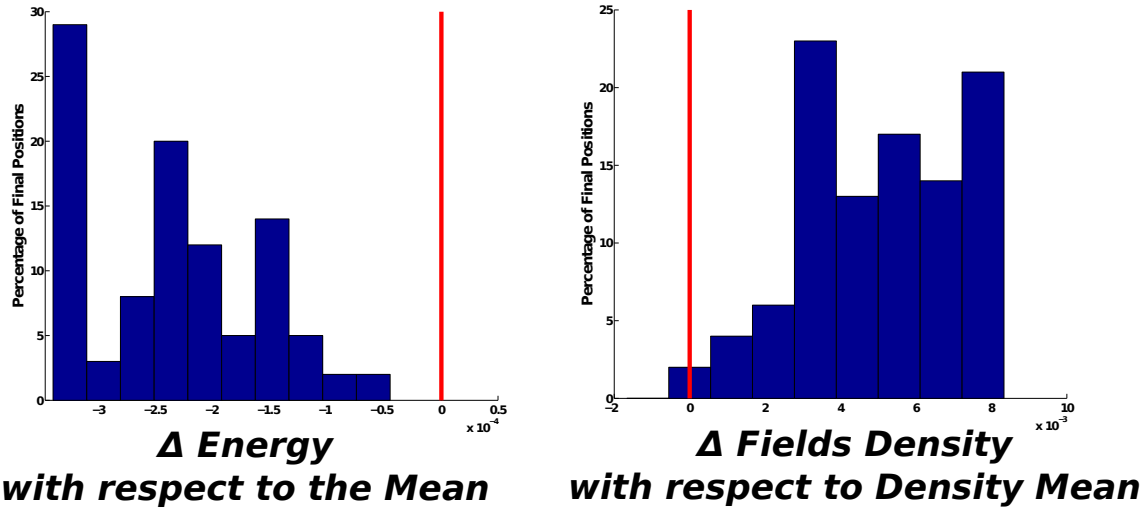


Figure 8.9: Distribution of the energies and fields density of the final location of drift trajectories. Values on the x-axis are relative to the mean (the red bar).

8.6 Multiple Environments Case

According to a popular view of spatial memories in the hippocampus (Samsonovich and B. McNaughton 1997), the representations would be organized as a collection of distinct manifolds, or charts. Each of these manifolds is associated to a different environment, and they would act as almost completely separate entities, with no or little interaction between different charts.

In the previous section, we argued that the surface of one chart mapping a single environment, might not be as smooth as expected by the classic continuous attractor theory (J. Knierim and K. Zhang 2012). Attractor dynamics seems instead to be happening on a mostly uneven landscape, leading the system to settle on a restricted number of stable final positions. The analysis was performed assuming that only one environment was stored in the network, but what about the presence of a set of similar charts, all stored on the same group of units (Battaglia and Treves 1998; Monasson and Rosay 2013; Rosay and Monasson 2013)? We have extended the model to include the effects on the synaptic matrix of a multiplicity of representations. The basics of the simulations are the same with the following differences. We have used a network of 1500 units. Each of these units has a probability p of being assigned to a particular map and to get a place field in the environment, and for most of the simulations $p = 1/3$. We store M representations, repeating the pooling procedure for each of them. The connection between two units is the average of equation 8.2 over the environments in which both units are active. The number of stored environments is varied between 2 and 50. The same analysis, applied to a different model of CA3, is present in Rosay and Monasson 2013.

For each of the maps we run 100 trials, as in the previous section. The decoding procedure is the same, only, this time, the templates include the representations of all the stored maps. Therefore we can track the drift of the system comparing its activity with that of all the stored manifolds, selecting the most similar template from any of the charts. Transitions are detected whenever the decoded template belongs to a map different from the one we initialized the system in. Do we observe such transitions, can spontaneously evolving activity move from one map to the next, or is the wandering of the system restricted to the first chart, with no possibility of leaving it without an external drive?

In Figure 8.10 we show that by increasing the number of memorized maps we can induce transitions between them. The increase in the number of trajectories leaving the environment in which they were initiated (light purple), starts above a certain threshold (10 environments) and then builds up linearly with the storage load. This increase is paralleled by a loss of coherence of the moving activity bump. In fact, if we calculate the distance of the activity pattern from the closest environment, we see that it progressively increases, meaning that the retrieved activity is less and less similar to the stored templates. The activity bump gets more unstable and in fact, when the number of maps gets larger, above 25, most of the configurations in which the dynamics settles are a mixture of states, not corresponding to any of the initially imposed configurations. In the plot we separated those trajectories that finished having a high coherence with only one of the maps (dark purple). These might be considered successful transitions and they appear to be expressed by the system only in a limited range

of stored maps. Between 10 and 25 maps they indeed comprise almost the entirety of the transitions.

The choice of the pooling probability p is of course affecting the storage capacity of the network (Battaglia and Treves 1998). A lower p leads to less overlap across the populations coding for different environments and thus to larger storage capacities (Fig 8.10, dashed lines). By reducing p to 1/4 we observe a shift of the curves towards higher values of M.

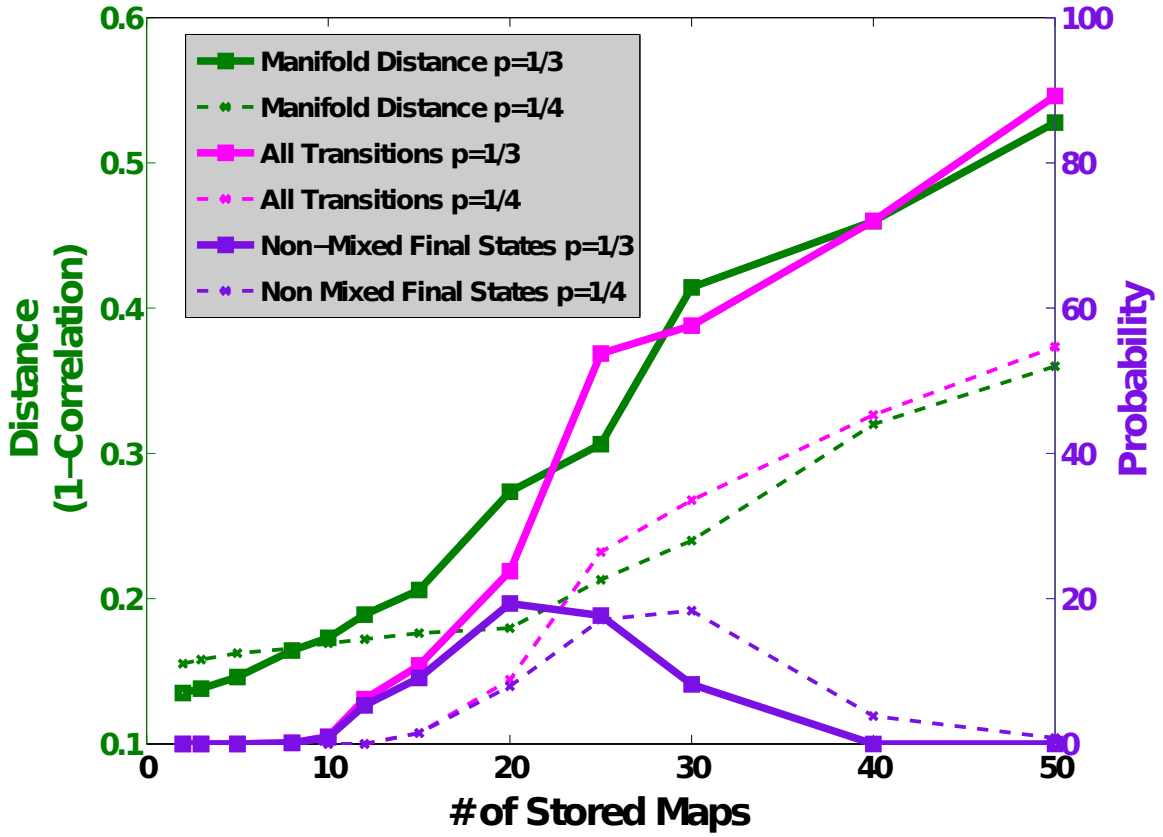


Figure 8.10: Solid lines, probability of having a field in a chart p=1/3. Dashed lines: $p = 1/4$. (Left Axis) Green: average distance from the closest environment manifold. (Right Axis) Light purple: Number of trajectories leaving the initial map. Dark purple: Number of trajectories leaving the first environment and ending in another one ("successful transitions").

How are the transitions points distributed (( )rosaycross2013)? If we look at a single environment (Fig. 8.11,a and b) we see that both the locations in which trajectories left for another map, and the locations where trajectories coming from outside appeared, are clustered in space, expressing some preferential distribution of these events. Moreover the distributions of the entry and the exit locations are not coincident, with some of the points in the environment that swallow the traveling bump, and some others they eject it, much like the dark and white holes of theoretical physics.

But the distribution of outs and ins is uneven also at the charts group level (Fig 8.11, c). Here we see that the various charts fall into a continuous distribution of the ratio between number of incoming trajectories and outgoing ones, with extreme cases of environments being only final destinations or, on the contrary being only abandoned rather than found. The results suggests the presence of an uneven probability of finding the system representing one of the environments and of a second-order organization of the energy landscape, with global minima at the level of the collection of charts.

What determines the existence of a transition location? If we look at the correlation between the last decoded template in the old environment and the first one in the new one, during a transition, we see that the values are higher than the global mean over all the configurations pairs coming from different environments. But these values are very low indeed (¡0.1 correlation), as a consequence of the random way in which the representations were generated. The slightly higher correlations are mostly due to the sharing of a relatively high number of units in the two population vectors (Fig, 8.12, right), which does not in itself correspond to a large overlap between the configurations.

The above analysis suggests that the transitions are not happening with a direct passage between one point of environment A to another of environment B. Instead the activity configuration goes through a mixed state, having a non-zero overlap with more than one of the maps. Only after transitioning through this mixture of states, a new environment is selected, and the system "re-descends" on it. To compare this transition type with one induced by the similarities of the representations of two different environments, we artificially generated an overlap between two maps, by imposing a common patch of place fields on the two which acts as a direct gateway to the other map. In Figure 8.13 we plot two illustrative examples for the two different situations. In the direct case the system never actually leaves the manifold, smoothly slipping into the new chart. This situation requires a very high degree of similarity between the representations of two locations in two different environments, and it is very unlikely to be obtained when the maps are randomly generated, even with a limited number of cells like in our simulations (of course, one can think of mechanisms for non-random generation of patterns in the setting up of CA3 representations (Dragoi and Tonegawa 2013; Mizuseki and Buzsaki 2013)).The kind of transitions one observes instead is a trajectory passing through an "uncharted" territory in the portion of the configurations space lying between the manifolds represented by the stored connections. The attractive nature of these manifolds appears from the eventual re-emergence of the system in another environment, even though the increase of final mixed states indicates a weakening of this attractive force with the increase of the number of memories (Amit 1992).

8.7 Discussion

The scenario presented by our simulations is pointing to a rich collection of dynamical phenomena characterizing the attractor-driven evolution of our system. We have shown how a reasonable implementation of CA3-like charts in a model network leads to a performance very distant from that of an ideal continuous attractor network. The activity bump in our model is extremely unstable and systematically drifts away from the initially represented position, as soon as the external input is weakened its mag-

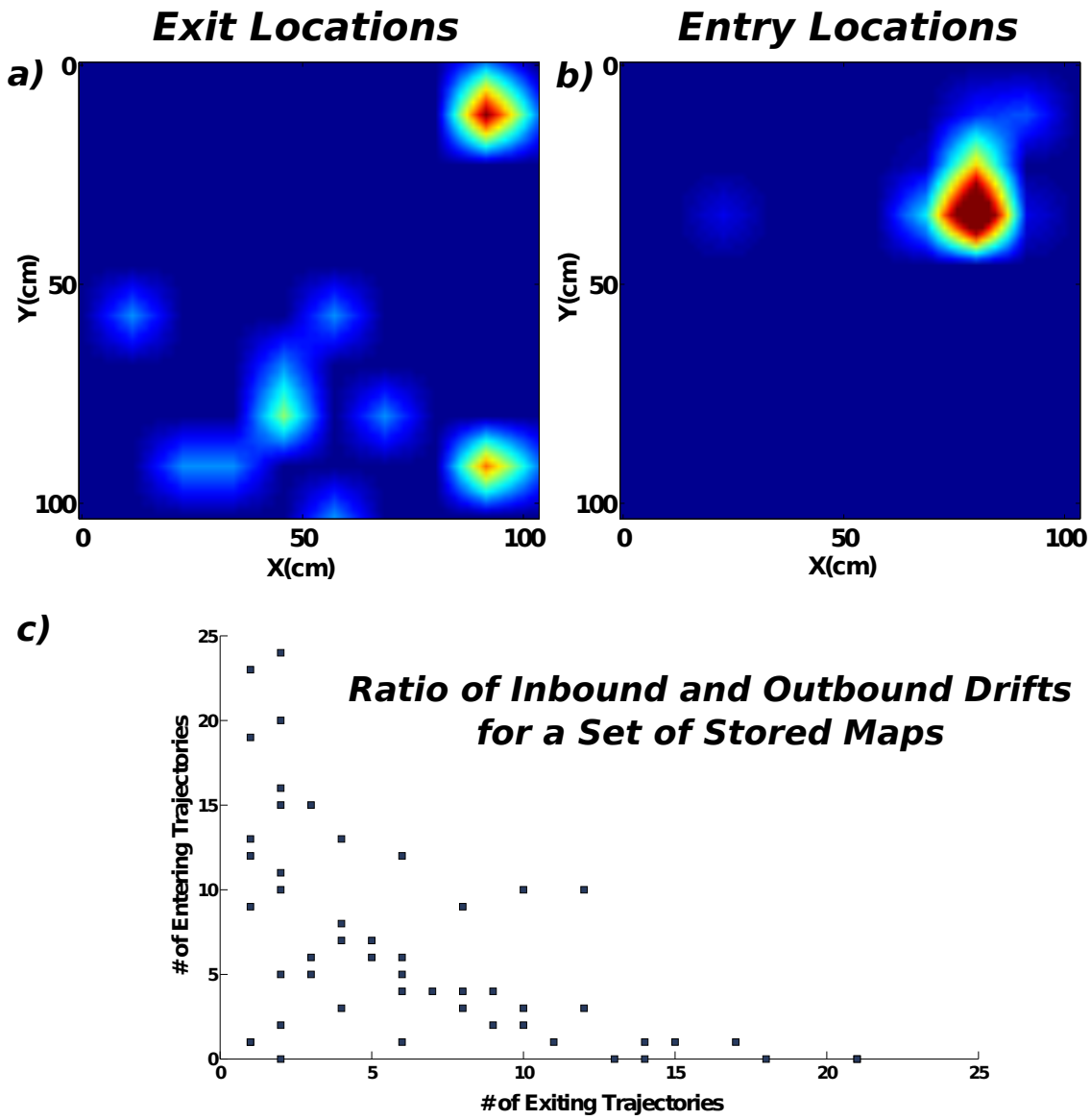


Figure 8.11: Spatial distribution of transition locations (as in Rosay and Monasson 2013). a) Example of the distribution of exit locations for one map. b) Distribution of the entry locations for the same environment. c) Distribution of entry and exit probability on a set of maps stored on the same network.

nitude. We observe both the displacement of the bump around a single environment, and its transition from one environment to another when more than one chart is present.

A fact of potentially functional relevance is the observed coincidence between the minima of the energy (that provides a synthetic description of the properties of the attractor landscape) and the over-represented locations. Given that place fields density is not always homogeneous and might be linked, in certain situations, to relevant spots in the environment, such coincidence might express a default mechanism by which the recurrent network allocates importance to some of the locations in the stored representation. Moreover, considering the tendency of the drift to trace a trajectory, from the rat position (indicated by the external input, when is present) to one of these locations,

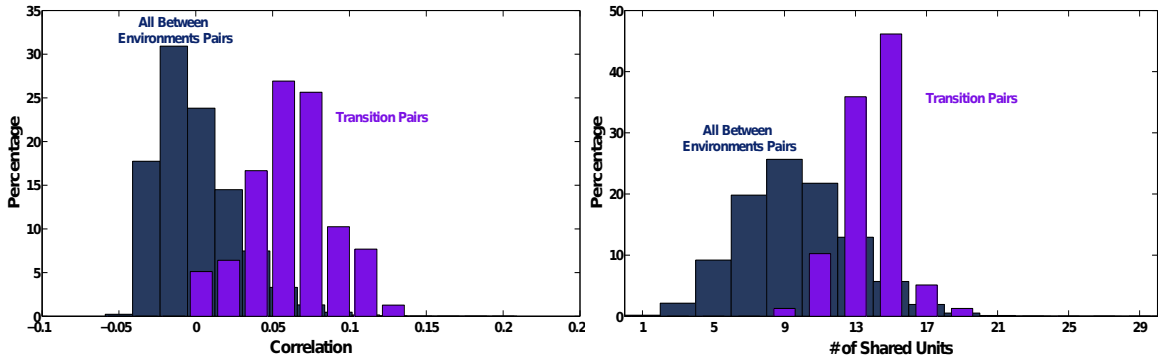


Figure 8.12: Left. Dark blue: distribution of the correlation values between pairs of representations corresponding to points of distinct environments. Purple: correlation values distribution on the subset of pairs at the two ends of a transition. Right: same but with the number of units common to the two population vectors.

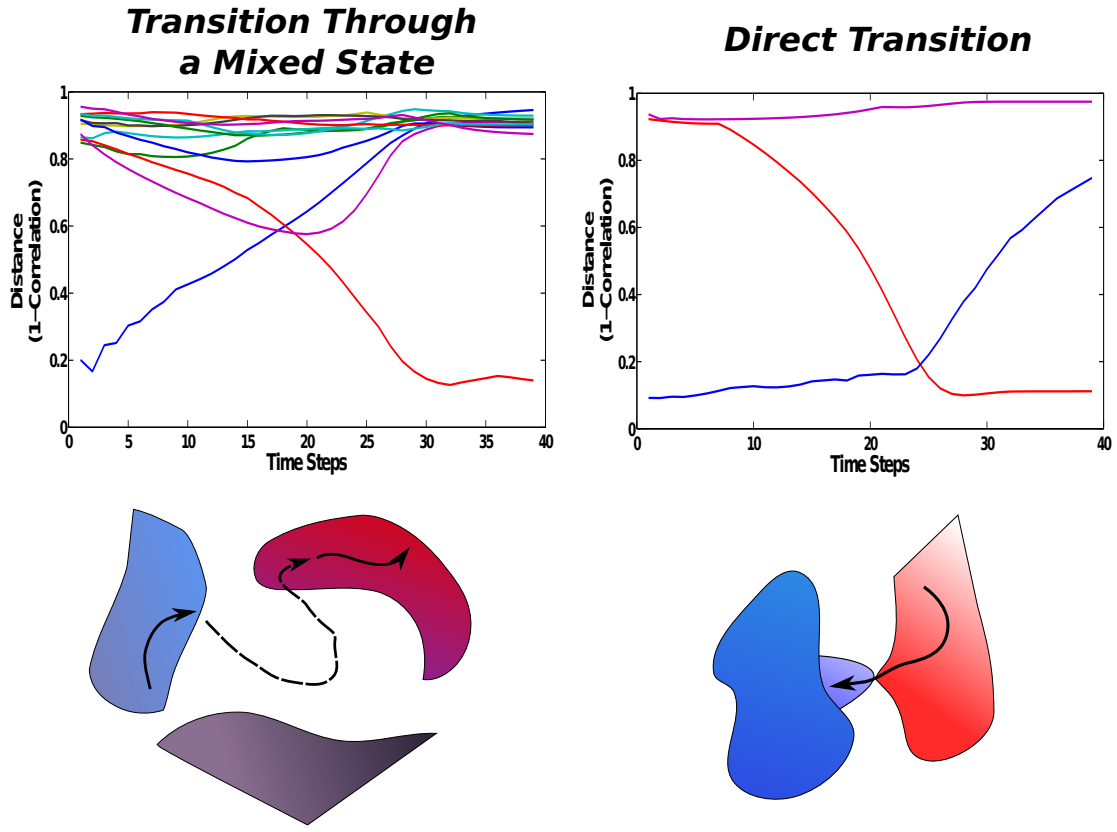


Figure 8.13: Left: example of a transition going through an intermediate mixed state. Different colored lines correspond to different environments stored in the network. Right: direct transition obtained when using highly correlated maps.

the dynamics we describe might eventually provide the basis for some sort of mental exploration mechanism.

If this sort of dynamics has any relevance for the function of the hippocampus and to guide behavior, then how does the system switch from a regime dominated by the

external input, to one in which the expression of the stored representation is allowed? To some extent, the dichotomy between recurrent driven dynamics and input driven dynamics is equivalent to the known issue of separating storage and retrieval in a recurrent network like CA3 (Treves and E. Rolls 1992; M. E. Hasselmo, Bodelón, and Wyble 2002); with the difference that in this case the contrast is not between so-called pattern separation and pattern completion, but revolves about the ability of the system to elaborate the incoming inputs based on the information stored in the connections. In the next section we will discuss a possible solution to this contrast involving oscillations, and in particular the rhythmic theta modulation of hippocampal activity. If the network is effectively switching between operational regimes in coincidence with particular physiological states, then it would be very important to better define the temporal dynamics of the model in terms of real time. Knowing how fast the drift is would allow to understand how much of the trajectories described before could fit into a theta cycle, or a sharp wave ripple, and therefore to understand what fraction of the dynamics (if only the initial transient, or the full relaxation) one can expect to observe in the real system. Of course, it would also allow for comparison with experimental evidence of sequential activation (Pfeiffer and Foster 2013).

Chapter 9

Theta Oscillations and Attractor Dynamics

9.1 Theta Oscillations in the Rodent Hippocampus

The hippocampal theta rhythm is a most prominent clocking mechanism in the forebrain Buzsaki 2006. The rhythm, in the 510 Hz band in the rat, can be readily detected as a macroscopic local field potential (LFP) in the dorsal hippocampus during exploratory behavior and REM sleep (Vanderwolf 1969; Buzsaki 2006). According to the simplest model of theta generation, the medial septum in the basal forebrain functions as a pacemaker, enforcing a global rhythm into which hippocampal and entorhinal cortex (EC) networks are entrained. However, evidence obtained with numerous experimental manipulations indicates that several rhythm-generating mechanisms and theta current dipoles are independently at work Montgomery, Sirota, and Buzsaki 2008; Mizuseki, Sirota, et al. 2009; Montgomery, Betancur, and Buzsaki 2009. The idea of a coherent wave of activity turning hippocampal neurons periodically on and off, as a single entity, has faded with the discovery of a much more complex pattern of activation of the various hippocampal subpopulations.

9.1.1 Not One Theta Oscillator but Many

The hippocampus is not passively responding to a single theta generator. Several studies (Buzsaki 2002; Montgomery, Betancur, and Buzsaki 2009) have shown that each lamina of the hippocampal formation acts as a distinct oscillator, with its own theta rhythm. Although these multiple oscillators differ in phase and amplitude, they generally present the same frequency (Mizuseki, Sirota, et al. 2009). The uniformity in the oscillation frequency across different layers nevertheless points at the influence of an external common pacemaker, generally identified in the medial septal area. Indeed, the inactivation of this area, or the interruption of its projection to the hippocampus, has a dramatic effect on theta dynamics in the region (Rawlins, Feldon, and Gray 1979; S. Mizumori et al. 1990; Yoder and Pang 2005)[20]. The LFP theta rhythm, along with rhythmic hippocampal unit activity, is eliminated, together with any coherence among different regions of the hippocampal formation. In normal conditions, sharing the same frequency but at different phases results in the appearance of particular coordination profiles among these regions (Figure 9.1). Recorded LFPs show increasing and decreasing intensity with characteristic inter-layer intervals, which remain more or less constant across time. More precisely, the two most prominent theta dipoles,

the one in the stratum lacunosum moleculare of CA1 and the other in the stratum moleculare of the DG, alternate in phase, while the current-source density in other layers shows intermediate phase relations. The theta modulation recorded in entorhinal cortex, which has been demonstrated to be at least partly independent (Hafting, Fyhn, Bonnevie, et al. 2008; Deshmukh, Yoganarasimha, et al. 2010; Quilichini, Sirota, and Buzsaki 2010), shows an inversion in phase between layer I and deeper layers. Layers II and III, where most of the projections to the hippocampus originate, thus oscillate in phase, and this phase is the same as the one typical of the dipole located in the stratum lacunosum moleculare of CA1.

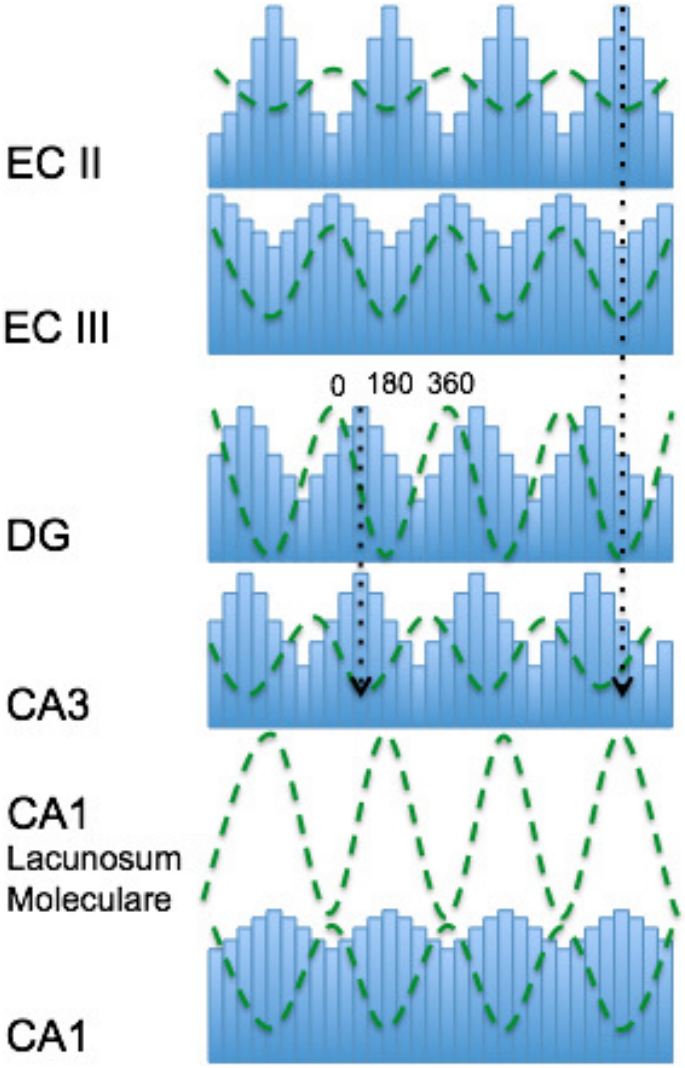


Figure 9.1: Illustration of the different phase relationships existing between theta modulation in different regions of the MTL. Dashed lines indicates LFP amplitudes at different locations, while the bars represents the induced firing rate modulation at the same locations. Redrawn from Mizuseki, Sirota, et al. (2009)

When examining the amplitude, frequency, phase, and coherence relationships among these dipoles during various aspects of behavior, however, it has been found that the power, the coherence, and the phase of theta oscillations exhibit layer-specific changes

that depend on behavioral task demands [19, 24]. As a consequence, the relations among the timing of theta oscillations in the various hippocampal layers can be modified by the particular cognitive state of the animal. Because the different layers are associated with particular physiological structures, such as distal and proximal dendrites and cell bodies, these modifications can influence the way spiking activity in the various layers is generated. This implies that oscillations can support specific processes, necessary only in certain conditions. A recent study [25] has complicated this situation even further, by indicating that theta oscillations are not synchronized within stratum oriens of CA1 but are traveling waves that propagate predominantly along the septotemporal hippocampal axis. This study is in accordance with the finding that theta dynamics are not homogeneous along the same axis: theta power and the fraction of theta-rhythmic neurons are reduced in the ventral portion of CA3 compared to the dorsal portion (Jung, Wiener, and B. McNaughton 1994; Royer et al. 2010). Again, this inhomogeneity has been hypothesized to enable specific computational functions (Fanselow and Dong 2010; Geisler et al. 2010). This body of experimental evidence indicates that theta rhythm is not a monolithic clock signal: hippocampal theta oscillations interact with a heterogeneous consortium of transmembrane currents, reflecting layer-specific processing that can be modulated by extrahippocampal inputs or differential modes of operation within the hippocampus.

9.1.2 Influence on Firing Rates

In a network of synchronized neurons, the oscillatory phase determines the degree of excitability of the neurons. In which way do hippocampal populations discharge with respect to the theta phase? We expect the maximal activity of neurons to be coincident with what is conventionally denoted as the trough of theta, when the membrane is depolarized. And because of the variety and heterogeneity of theta oscillations, we expect that neurons belonging to different subregions should fire at different times, in accordance with the local dipole. This is found to be only partially true (Mizuseki, Sirota, et al. 2009). Indeed, the activity of all the neurons of the hippocampus is modulated, to different degrees, by the theta rhythm. The strength of this modulation is not defined by the amplitude of the corresponding theta oscillation, however. CA3 neurons show the strongest modulation, while CA1 modulation is significantly less pronounced, even if the strongest theta dipole is in the CA1 region. The preferred phase does not necessarily correspond to the theta trough, either. So, for example, the neurons of EC layers II and III fire at opposite phases of the common reference oscillation, layer II neurons at the trough, and layer III neurons at the peak (Figure 9.1). This results in the fact that while the peak of population activity in an upstream region may match the timing of dendritic excitation of downstream target neurons (current sinks) in the hippocampus, the discharge of the respective target populations could be offset of even half a theta cycle. Also interneurons show the effects of the theta modulation (Freund and Buzsaki 1996; Klausberger and Somogyi 2008). As pyramidal neurons do, also the different classes of interneurons have phase preferences. Different types have different preferred phases (Halasy et al. 1996; Miles et al. 1996; Klausberger, Magill, et al. 2003; Klausberger, Marton, et al. 2004; Somogyi and Klausberger 2005; Czurkó et al. 2011). Interestingly, it appears that dendritic targeting interneurons tend to group together and to have the same discharge phase, different from the one preferred by somatic targeting interneurons, similarly grouped on their own (Miles

et al. 1996). The local theta does not seem to constrain the discharge to occur at its trough. Each population has a preferred phase of its local theta oscillation. And if we take a global look at the distribution of spikes, taking as a reference a single theta oscillator, we see that different regions discharge with specific times with respect to others, generating a complex sequence of activation connecting the various parts of the circuit (Montgomery, Sirota, and Buzsaki 2008; Mizuseki, Sirota, et al. 2009; Montgomery, Betancur, and Buzsaki 2009). This pattern does not need to be the same at all times. Also the preferred phase of certain populations shifts according to the task (Manns et al. 2007).

9.2 Flickerings in CA3

New insight on the dynamics of spatial representations in CA3 has recently come from work by Jezek et al. (2011). The aim of this study was to investigate the time course of map replacement, by determining the evolution of neuronal activity in response to an abrupt transition between two environments, resulting from a sudden change of the external cues used to define each spatial context. The recordings were performed in CA3.

Rats were trained in two different text boxes, designated A and B, which differed only in the type of illumination. The rats spent several days exploring these environments. The paradigm was intended to induce global remapping of the activity in CA3, by generating two uncorrelated representations for the two environments. The test phase took place after such uncorrelated representations had stabilized in the training phase. The rat was started in one of the two familiar environments, then after some time the rat was teleported to the other environment: the lights were abruptly switched to those of the other environment, taking the rat instantaneously, as it were, from one place to another.

A number of analyses (Jezek et al. 2011) indicate that even during these extremely rapid transitions between reference frames, the activity in the hippocampus is determined by attractor dynamics. The representation expressed by CA3 at any time is almost always one corresponding to one of the environmentsfew intermediate representations are observed, corresponding to mixtures of the two representations for box A and B, with elements from them both. The competition between the representations, stimulated by the fast transition, is revealed by the appearance, just after the teleportation, of rapid switches from one representation to the other, a phenomenon labeled by the authors as flickering. One peculiarity of such flickering is its relation to theta oscillations. The re-expression of one of the two representations is generally completed within the time span defined by a theta cycle. The theta rhythm seems to regulate the timing of the dynamics also by marking the alternation between the two representations.

These results seem to indicate a major role of the theta oscillations in regulating attractor dynamics in the CA3 network. The propagation of activity through recurrent collaterals appears to be controlled by the periodical modulation of the physiological variables in the network. The raison detre of this phenomenon might be to allow a trial-and-error-like procedure. Repeated convergence to attractor states might offer the rodent the opportunity to self-correct, a possibility that can be very useful under

conditions where input cues are ambiguous, weak, and confusing. In the case of this experiment, the conflict is between the external cues, characterizing the final box, and cues coming from path integration which, because of the very fast transition, still refer to the initial box. The possibility to continuously integrate clues about location and to formulate distinct hypotheses within a limited time may be important in order to properly adjust behavior to the external world.

9.3 Our Model of Teleportation

To test the feasibility of an involvement of the theta rhythm in the dynamics of memory retrieval and to check that the phenomenon observed in the teleportation experiment can be ascribed to these oscillations, we have tried to simulate the results with a simple model representing the CA3 region of the hippocampus. The goal of these simulations is to see whether the flickering events may arise from a modulation of the activity in the presence of partial and ambiguous external inputs. We thus compare the activity in the network in the presence of theta oscillations and without them, when all other conditions are the same. Even if the neural model is intended to represent the CA3 region of the hippocampus and its dynamics, the structure of the network and its operating parameters do not exactly reproduce those in the real brain. T o focus on the effects of theta modulation on network performance and to reduce any other spurious effect coming from uncontrolled features of the model, not relevant for the present issue, we have tried to reduce the complexity of the simulations as much as possible. For the sake of clarity the model then represents a system only similar to that present in the brain, retaining-it is hoped-the characteristics important for the problem at hand.

9.3.1 The Model

We first construct the representations of the two environments. They are two geometrically identical square boxes of size S. The network is comprised of N units. In the standard condition, the two environments are represented by two separate populations of units, half active in environment A, the other half in environment B. Each unit has a single place field in the assigned environment. Its firing rate around the place field center is defined by a Gaussian function. Place field centers are regularly spaced, arranged on a grid covering the whole environment. The two representations are therefore identical and differ only in the subset of units which are active. The Gaussian bumps representing the place fields are defined as all having the same effective size and height. The network is fully connected, and the strength of the connections is prewired. The structure of the connections is functional to the establishment of a quasi-continuous attractor surface, expressed by synaptic weights that follow an exponential decreasing function of the distance between place field centers. The weight of the connection between cell i and j is then written as

$$W_{\text{rec}}^{ij} = C * \exp(-\frac{d_{ij}}{2 * s^2}) \tag{9.1}$$

if the two cells have place fields in the same environment, while it is set to zero if they belong to different environments. The total strength of the weights of each unit is then normalized as

$$\sum_j W_{ij} = |W_{TOT}| \tag{9.2}$$

The model is a firing rate model and the neurons are simple threshold linear units. The activity is thus governed by the following equation:

$$\eta_i(\vec{x};t) = g\left[\mathrm{Ext}(\vec{x}) + \sum_j W_{\mathrm{rec}}^{ij}\eta_j(\vec{x};t-1) - T\right]^+ . \tag{9.3}$$

For each cell, the activating current is comprised of two terms: an external input and the recurrent collateral contribution. The gain g and the threshold term T are modified to regulate the average activity $\mu = \sum_i \eta_i/N$ and the sparsity $a = (\sum_i \eta_i)^2/\sum_i \eta_i^2$ of the network. The value of the normalization $|W_{TOT}|$ 9.2, should be compared with the strength of the external input. Their ratio is the relevant quantity for the model. In the following we will set the maximum value of $Ext(\vec{x})$ to 1.

Generating the Templates The simulation consists of two phases. First, we generate a trajectory for the virtual rat. The simulation is continuous in space but discrete in time. In each time step, intended to correspond to roughly 15 ms, the virtual rat moves by 0.5 cm in a direction similar to the direction of the previous time step, with a small amount of noise. At each time step we determine the input activity associated with the position occupied by the rat. To that effect, movements are considered to occur along the same trajectory in both environments, and so inputs are updated for both environments. The activity of each unit is determined according to (9.3), with the external input set according to the location in the current environment; while the term representing the activation induced by the contribution of the recurrent collaterals depends on the activity of the network calculated at the last step. This procedure emulates the activity that is elicited in CA3 during navigation in each environment, providing the templates of the spatial representation of the two environments. To study the effects of teleportation on the dynamics of CA3, such templates are then compared with the state the network takes when, instead of a full cue about one environment, it receives conflicting inputs about both.

Test Phase The second part of the simulation reproduces the teleportation experiment. The same path is followed as for building the templates, so that the rat runs through the same trajectory. We then divide the path into three sections. In the first part, we feed the network with a full cue about environment A. Then, we teleport the rat to environment B. The external cue is suddenly changed. To simulate the ambiguity and the conflict between external cues and internal path integration, we modify the input to represent this confusion. While before we fed the network with a full cue, we now reduce it to 20%, in that only a subset of the cells now receives external input. Moreover, the cue does not represent a single environment: a fraction refers to environment A and the rest to environment B, which means that both units coding for A and for B receive external input. Time steps are grouped in sets of eight, which then represent eight phases within individual theta cycles, when we introduce theta modulation. During each set/cycle the cue is kept constant, in the sense that, while the input changes as the animal moves and changes location, the units that receive external input remain the same. At the beginning of each cycle these cells are selected with a certain probability, representing the proportion of environment A and B in the external cue, which changes at each cycle. While the number of neurons receiving an external input is kept constant, the amount of activated neurons in each group has a

25% variability.

This state of confusion is held for a certain time, 10 sets or theta cycles in our simulations. At the end of this phase, we turn the external input to a full representation of environment B. The network is then fed a full cue corresponding to the second environment, simulating a rat that has finally understood which box it is in. All the simulations are first run without any periodic modulation of the activity in the network. We then introduce theta oscillations by periodically modulating the mean activity and the sparsity of the population, with a period of eight time steps (Scarpetta, Zhaoping, and Hertz 2002). We also modulate the strength of the external cue in a similar way, by varying both the number of units receiving external input and the intensity of the stimulation. We vary the amplitude of this modulation and the relative phase Δ between the oscillations of the network and of the external input.

$$\mu_{CA3}(t) = \mu_0 + A * \cos(\pi/2 + 2\pi/L * t) \tag{9.4}$$
$$\mu_{Ext}(t) = \mu_0 + A * \cos(\pi/2 + 2\pi/L * (t + \Delta)) \tag{9.5}$$
$$a_{CA3}(t) = a_0 + A * \cos(\pi/2 + 2\pi/L * t) \tag{9.6}$$
$$a_{Ext}(t) = a_0 + A * \cos(\pi/2 + 2\pi/L * (t + \Delta)) \tag{9.7}$$
$$\tag{9.8}$$

where μ_0 and a_0 are both equal to 0.1 and L is the number of time steps of the simulation corresponding to the length of a theta cycle, here $L = 8$.

9.3.2 Results

In Figure 9.2 we plot the correlation between the activity generated during this test phase and the activity obtained during normal exploration, in the same location. Four representative examples are illustrated. As expected, the network with no modulation (Figure 9.2, top row) generally fails to respond to changes in the external environment. After the first switch, the action of the recurrent collaterals is to maintain the network in the attractor relative to environment A, even if the external input has changed. The external cue can produce an effective change in the dynamics of the network only when very strong. At the end of the simulation, with the full cue of environment B, the system actually follows it and retrieves the other representation. In general, however, in this range of parameters, the dynamics of the network do not respond rapidly to changes in the external environment. The system cannot find an intermediate dynamic range between being completely externally driven and the settling in one attractor. In Figure 9.3, left, we show how varying the strength of the collaterals relatively to the external input leads to the switch between two regimes. When the collaterals are weak, we actually observe alternation between maps (green line), but the ability of the system to retrieve the representation is very limited (blue line). As the strength of the collaterals becomes sufficient to initiate the retrieval, the transitions are quickly lost.

Flickering events appear, instead, when we introduce strong theta oscillations and in antiphase with the modulation of the input (Figure 9.2, second row). Such rapid transitions between the retrieval of one attractor and the retrieval of the other, in the time span of a theta cycle, are similar to those observed in the experiment of Jezek et al. (2011) and appear with the first switch between the environments. The simulations

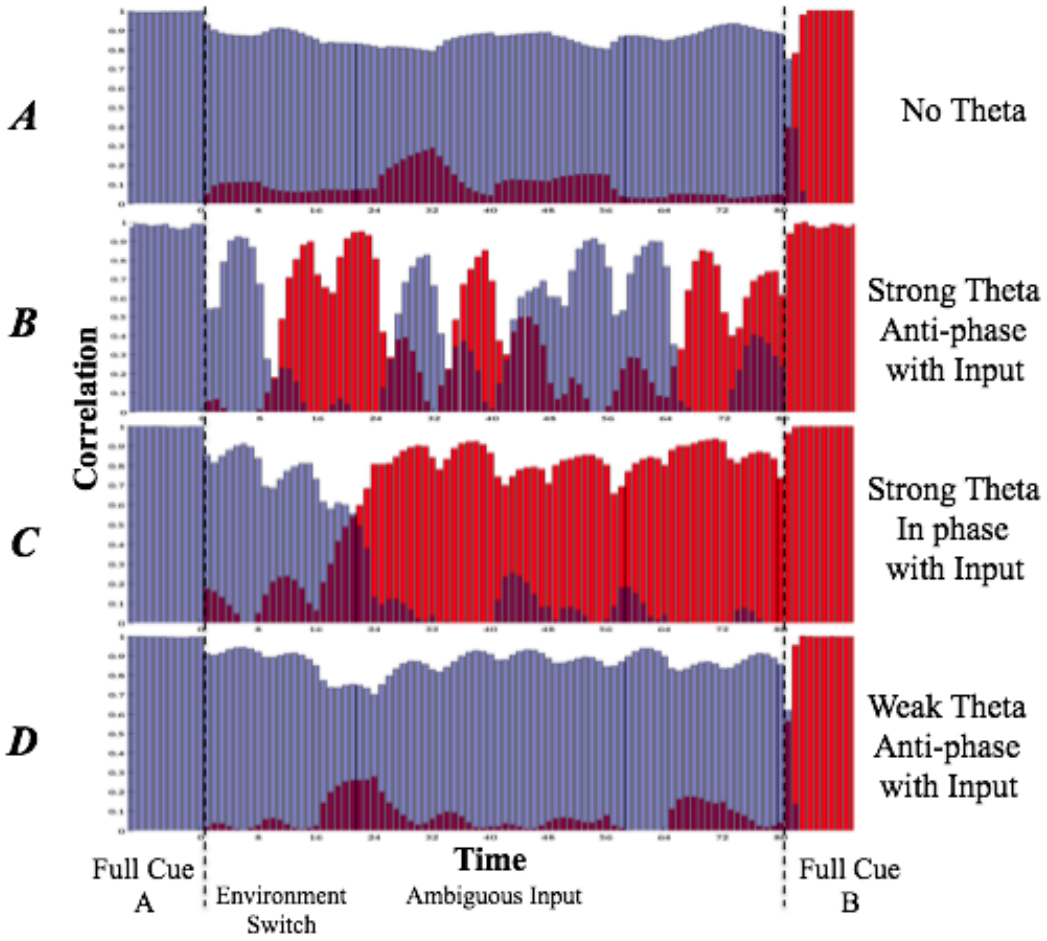


Figure 9.2: Examples from the results of our simulation of the teleportation experiment in different parameter ranges. Bars represent correlation with the two different representations at each time step (in the simulations a theta cycle is comprised of 8 time steps). Different plots are relative to different conditions: strong is 75% variation in firing rate and sparsity between the peak and the through and weak is 20% variation; in phase means $\Delta = 0°$ and out of phase means $\Delta = 180°$.

show the effect of modulating the activity in enhancing the flexibility of the network and its ability to respond rapidly and accurately to changes in the input. Note that, while reacting to the changing external cue, the network does not simply reproduce the ambiguity of the input. In each theta cycle, with a few exceptions, the system retrieves one of the two representations and discards the other (Figure 9.5). At every theta cycle this retrieval appears to be as good as the one obtained in the case with no oscillations. Figure 9.5 shows how the retrieval of the representation is happening in the second half of the cycle, while in the first half there is a preponderance of mixed states with components from both maps. Also this segregation finds confirmation in the data from Jezek et al. (2011). The strength of the collaterals still affects attractor dynamics in the presence of theta oscillations (Fig 9.4). Stronger collaterals actually induce a longer permanence in the same attractor and a decrease in the flicker probability. The system therefore, is not completely memory-less, even though weaker

collaterals significantly reduce the effects of the retrieval history.

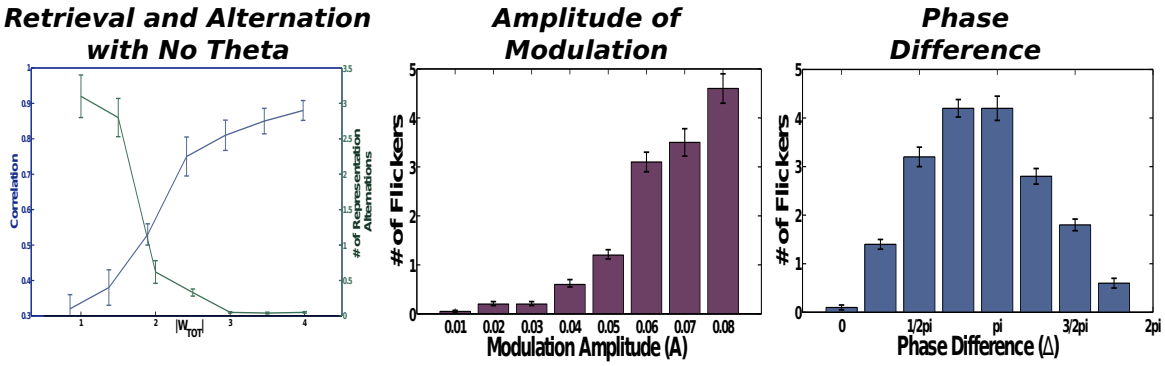


Figure 9.3: Left: Collateral strength effect on the retrieval in the absence of theta modulation. Center: larger amplitude leads to more flicker events. Right: Flickers appear preferentially when the external input modulation phase has an offset close to π.

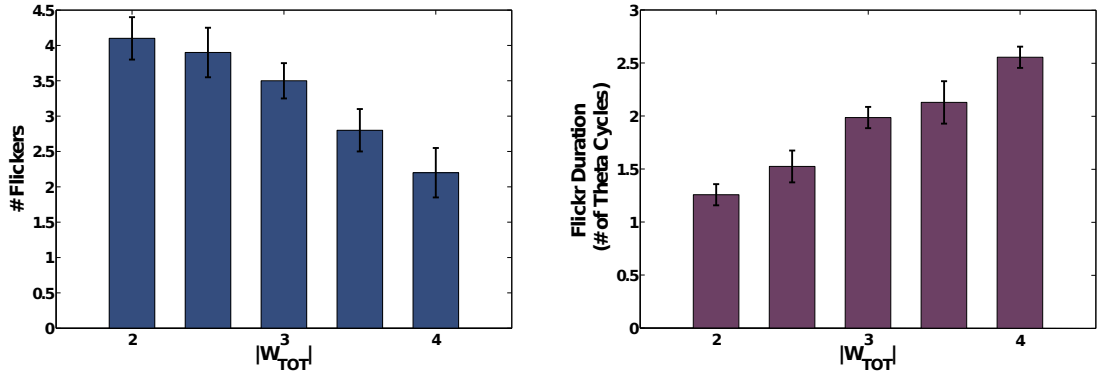


Figure 9.4: Flicker appearance is affected by collateral strength. Stronger internal connectivity increases memory effects in the dynamics: the same representation persists across more theta cycles, and flicker probability decreases.

The theta efficacy in inducing flickerings has a strong dependency on the parameters A and Δ, the amplitude of the degree of activity modulation and the relative phase of the external and internal oscillations (Figure **??**). Varying the two parameters produces a different number of switching events. As already evident from the examples in Figure 9.2, flickerings appear only with a combination of strong amplitudes and half-cycle offsets. This effect is quantified in the two plots showing the increase of events with the increase of the amplitude (Fig **??**) and the preferred offset around π (Fig 9.3, right). Perfectly synchronized oscillations produce very few switches.

9.4 Discussion

Our model suggests that modulating the activity of the CA3 region of the hippocampus may have significant effects on memory processes, but only if the modulation is strong.

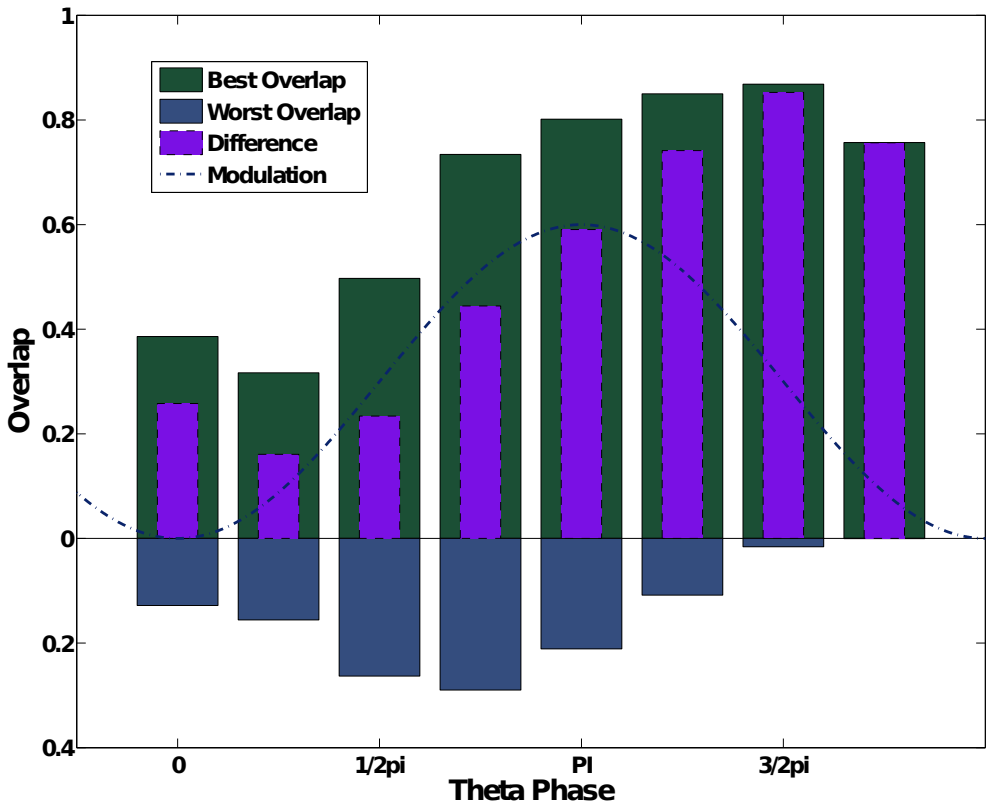


Figure 9.5: Theta oscillations entrains the attractor dynamics of the system. Full retrieval of the representation is confined to the second half of the cycle (larger activity), while on the first half of the cycle mixed states predominate.

In the model, flickering events as described by Jezek et al. (2011) in their teleportation experiment appear, but only in the presence of strong theta modulation. This is consistent with an active role of theta activity in stimulating fast transitions between attractor states (Harris et al. 2003).

The Conflict between Attractor Dynamics and External Input During memory retrieval, the activity generated by an external stimulus is propagated through the collaterals connecting the units of an associative memory network. Since memories are stored by modifying the strength of these connections, this propagation leads to an activation of the network which corresponds to one of the stored memories-what we call an attractor. Which one of the multiple memories stored in the network will be retrieved depends on the characteristics of the initial external input. Such associative memory networks function optimally when the external input serves only as a cue and neuronal activity is driven maximally by the recurrent connections. The internal dynamics of the network should be relatively unconstrained by the external input, to allow for convergence to the attractor state, even in the presence of an incomplete or distorted cue. The drawback of this operating regime is that the network loses in flexibility. The network is likely to remain in an attractor once it has reached one, and it

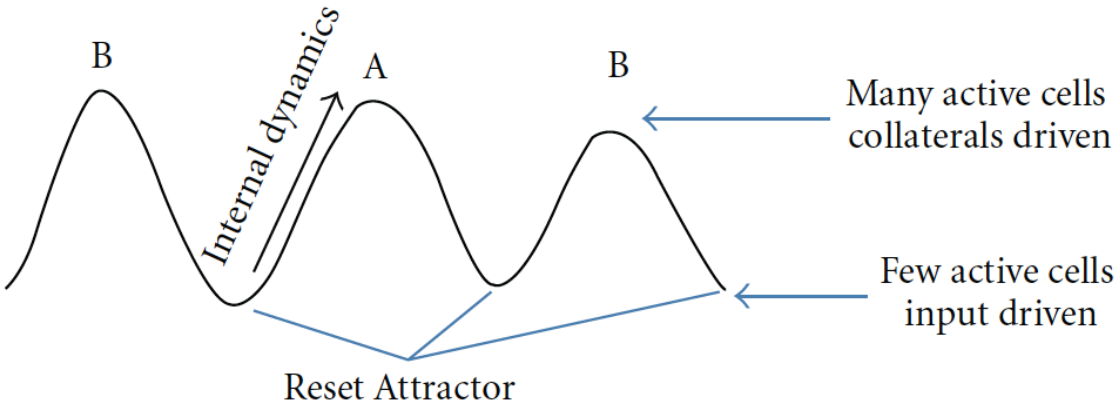


Figure 9.6: Illustration of the proposed mechanism of alternation of externally- and recurrent-driven dynamics by means of regular modulation of the system

needs a very strong shove to leave it. It is not reactive to external inputs unless this input is strong and significantly different from its current state. In our case, in which the external input is a mixture of two different representations, the action of the recurrent collaterals is to let the system choose between one of them, generally the stronger of the two, and it drives the activity of the neurons towards the corresponding attractor. But as the external input changes, if neither representation clearly prevails, it is very difficult for the network to modify its state, at least, that is, when no modulation is present.

Theta Oscillations as a Regulator of External Input Dominance The effect of introducing a modulation of the activity and of regulating accordingly the firing rates of the neurons is to control the relative strength of the inputs each neuron receives, the external cue, and the recurrent reverberation. As the number of active units and the level of global activation changes, even if the strength of the connections remains unvaried, the driving force of the system is shifted from external to internal inputs and vice versa. When only few cells are active, those which receive a strong external input have a major advantage in the competition and therefore are more likely to pass the inhibition threshold and to fire. They in turn activate new units as inhibition gets weaker. As new units are activated through the current injected in the collaterals, the activity in the network evolves, driven by the internal dynamics generated by the initial pool of active cells. If cells representing both maps are activated, then as the threshold is lowered there is a rush to recruit new units, in a competition between the two representations. When the inhibition reaches its minimum, one of the two eventually takes over, and the system falls in one of the two attractors. What happens next is crucial for the appearance of flickers. Once the inhibition rises again, the active population starts to shrink. In the beginning the collaterals continue to determine the active units, but as the inhibition approaches its maximum again, the contribution of the external input becomes critical. In a regime of very strong competition, the activity of the system is strongly related to the input that it receives at that time. As before, this small core of units is then responsible for the activation in the next theta period, while the state the system was in is forgotten. The system goes through a sort of bottleneck that allows resetting the state of the units and forcing the system

to discard the information conveyed by the recurrent collaterals. At each onset of the cycle the internal dynamics of the system are thus almost independent of the past activity. The system is potentially able to select a new attractor at each theta cycle. The attractor dynamics observed in the teleportation experiment look quite different from both point-like attractor and continuous attractors dynamics, and seem to reflect the presence of periodic modulation (Figure 9.7).

In the non-modulated condition, either the system is driven by external input, with little room for attractor dynamics, which merely serves to keep it within the correct valley (Samsonovich and B. McNaughton 1997; Zugaro et al. 2003; A. Johnson and A. Redish 2007; Akrami et al. 2009b), or, if inputs subside, it is capable of memory retrieval, but rigid and unable to rapidly adapt to further external changes (Akrami et al. 2009b). Introducing theta modulation allows implementing both these regimes in the same system, by segregating them to different theta phases. The network is input driven in the presence of strong inhibition and then progressively switches to be driven by recurrent connections as the theta period progresses. In this perspective, the addition of theta oscillations, while not essential, provides a new twist in an already consolidated schema. Flickers could be the result of successive descents in different attractor basins, while the internal attractor dynamics, taking place along the continuous attractor representing each environment, may remain unaffected by the modulation (Figure 9.7).

Strong Modulation . . . Our simulations indicate that to obtain this particular effect the amplitude of the theta oscillation is crucial. The difference between the number of active units and the average firing rate at the peak and at the trough of the theta cannot be too small. Of course this conclusion is limited by the particular assumptions and simplifications we made for our model. The way we implemented theta effects on the system is simple and most straightforward. We do not explicitly model the inhibitory neurons present in the hippocampus, nor we take into account complex interactions between them and the pyramidal neurons. The effects of inhibition are modeled as a homogeneous regulation of the gain and of the threshold applied to the firing rate of all the neurons. Actually, as we have already said, there is growing evidence on the complex dynamics linking the excitatory and the inhibitory populations in the hippocampus (Buzsaki and Chrobak 1995; Hangya et al. 2010; Dupret, Joseph, and Csicsvari 2013). The different phase locking exhibited by distinct inhibitory neuron types and the fact that this differentiation roughly corresponds to the classification in dendritic or somatic interneurons suggest the possibility that such articulation is functional to an inhibitory mechanism not described in detail by our model (Csicsvari, Hirase, Czurko, et al. 1999; Maurer et al. 2006). Moreover, the particular feedback inhibition microcircuitry present in the hippocampus, which regulates local activity and is not detailed in our model, may be another source of complex effects not expressed in our simulations (Acsady et al. 1998; Banks, White, and Pearce 2000; Pouille and Scanziani 2001; Remondes and Schuman 2002; Freund 2003; Lawrence and McBain 2003; Pouille and Scanziani 2004; Ang, Carlson, and Coulter 2005; Szabadics and Soltesz 2009; Torborg et al. 2010). Input-specific suppression and other nonlinearities, which may arise from the incorporation of this finer structure in a model, may extend our results also to the case of gentler oscillations.

. . . And out of Phase Another indication that emerges from our model is the role of the timing of the activity in different elements of the hippocampal circuit. As the simulations show, the relative phase of the oscillations of the incoming input is important to maximize the effects of the theta modulation on the dynamics of our recurrent network. It appears that the hypothesized mechanism functions optimally when the two oscillations are out of phase. This makes sense, as such a phase relation of the two oscillators enhances the separation between the two regimes of input dominance and recurrent dominance, which is at the base of our model. A concurrent rise and fall of the activity in the network and in the strength of the input actually cancels the benefits obtained with the introduction of the modulation. The input then generates interference with the internal activity when the dynamics of the system should be driven by the latter, and it is weak when its turn comes to lead. On the contrary, an external input which is in antiphase with the modulation of the network can act effectively during periods of low activity and then gently bow out when the network retrieves the attractor. In conclusion, our data better reproduce experimental observations when the recurrent network is driven by an external input which oscillates in magnitude with the same period but a different phase.

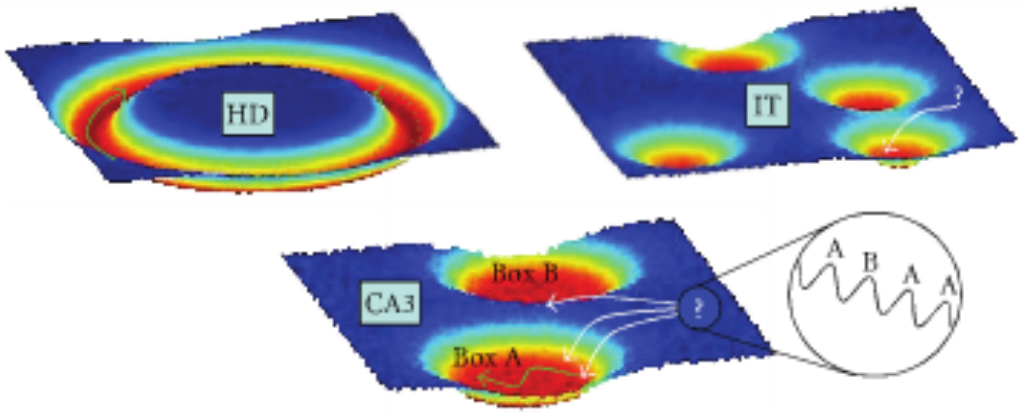


Figure 9.7: Different types of (idealized) attractors suggest distinct interpretations of neural dynamics observed in experiments. In the 1D ring attractor putatively underlying the head direction system (Zugaro et al. 2003) (top left), changing head direction implies moving, whether continuously (green arrow) or in jumps (dashed arrow), along the attractor. Convergent neural activity in IT (Akrami et al. 2009b) (top right) and rapid flickers in CA3 (Jezek et al. 2011) (bottom) can be interpreted as rolling down into the attractor basin (white arrows). In CA3, spatial representation activity can also move along the flat attractor valley (A. Johnson and A. Redish 2007) (green arrow).

9.4.1 Some Remarks on Information Flow in the Hippocampus

This leads to some comments about the operation of the real hippocampus. As reviewed above, the phase relations of the activity of the different regions of the hippocampus and of the neighboring afferent regions of the entorhinal cortex are rather articulated. Consider the CA3 region. The two external inputs this region receives have specific

and different theta phase modulation of the firing rate (Mizuseki, Sirota, et al. 2009). The neurons in the DG, which projects to CA3 through the mossy fibers, fire almost in conjunction with cells in CA3, and moreover the pattern of activation is very similar: the activity in the two regions spreads and then shrinks in almost perfect synchrony (Montgomery, Betancur, and Buzsaki 2009). On the other hand, activity in the EC layer II, which projects to CA3 through the perforant path, appears to be maximal at a completely different theta phase (Buzsaki 2002; Dragoi and Buzsaki 2006). Between the two bumps of activity, the one in EC and the one in CA3, there is approximately a quarter of a theta cycle (Mizuseki, Sirota, et al. 2009). This configuration seems to replicate the one described by our model. If we want to infer from the simulations, we should conclude that during the teleportation experiment, just after the switch of environments has occurred, the recorded activity in CA3 is mainly due to stimuli coming from the entorhinal cortex, in essence bypassing the DG. This conclusion suggests a revision of the prevailing view of information flow in the hippocampus in normal conditions. The normal path of activation, involving in the order layer II of entorhinal cortex, DG, CA3 and finally CA1, may indeed be the normal one, but during encoding, or when inputs are strong. During retrieval or when inputs are weak, there is some evidence that CA3 may operate in different and alternative regimes.

(a) First, behavioral experiments on spatial learning in rodents indicate that DG inactivation and DG lesions greatly impair the acquisition of new memories while sparing the retention of already stored ones (Lassalle, Bataille, and Halley 2000; I. Lee and Kesner 2004; B. McNaughton, Battaglia, et al. 2006). These studies indicate that the combined action of perforant path and recurrent collaterals is sufficient to perform retrieval of stored memories. So, while minor during robust DG activation, the contribution of the perforant path emerges when left alone. Synaptic plasticity acting on the perforant path allows associating the activity of the entorhinal cortex with the representation imposed by dentate activity, thus enabling future pattern retrieval.

(b) This alternative route, which forfeits the contribution of the DG, appears to contribute to another phenomenon observed in the hippocampus, the so-called sharp wave ripples (SWRs) (Buzsaki 1986). Indeed, this physiological pattern of activity originates directly in CA3, triggered by synchronized activation of pyramidal cells, and does not involve at all the dentate gyrus (Buzsaki 1986; Csicsvari, Hirase, Mamiya, et al. 2000). It is plausible to suppose that these sudden bursts of activity may be initiated by a weak input arriving from the entorhinal cortex(Sirota and Buzsaki 2005; Isomura et al. 2006). The activation of even a small subset of CA3 neurons could initiate a cascade of excitation across the recurrent collaterals, leading to reverberating activity that eventually spreads to downstream regions like CA1 (Cheng and Frank 2008; Karlsson and Frank 2009) (possibly with a more active and independent role of the latter in processing this signal than initially thought, as recent findings show that the phenomenology of the CA1 ripples is too complex to be a mere reproduction of CA3 activity during these episodes (Nakashiba, Buhl, et al. 2009; Sullivan et al. 2011)).

It is possible that inhibitory dynamics similar to that underlying SWR generation Csicsvari, Hirase, Mamiya, et al. 2000; Sullivan et al. 2011; Patel, Schomburg, et al. 2013 may operate in the rat hippocampus during the interval following the teleportation. The confusing and misleading context may limit the activity in the EC to a

fraction of that normally observed. In this case, the input originating from this region may be insufficient to properly activate the corresponding spatial representation in the DG. Still it can be enough to generate a cascade of activity in CA3. The presence of recurrent connections may act as an amplifier of an otherwise weak signal. The concomitance of maximal DG and CA3 activation is in accordance with the generic idea that phase synchronization is functional to optimize neural communication and synaptic plasticity (Buehlmann and Deco 2010). Having the neurons in the two regions active at the same time can be used to maximize the information transfer between the two regions during the storage of new representations, even if we cannot exclude that this scenario may be partially modified by specific effects arising from the particular structure of the feedforward inhibition of DG on CA3 (Lawrence and McBain 2003). On the other hand, the prevailing phase difference observed in the firing of the excitatory populations in EC and CA3 may result in an effect not captured by the generic idea. This particular timing allows the dynamics of CA3 activity to select a new attractor at each theta cycle. Of course if the external cue, even if partial, is correlated with a single stored memory, the system will simply retrieve the corresponding attractor. This mechanism becomes interesting when the cue is ambiguous and is comprised of a mixture of different representations. What happens then is that the choice of the attractor may amplify a bias or even a statistical fluctuation of the composite cue. If we observe the choices the system makes over a certain period, the number of times it chooses a given memory could be related to the fraction of the cue representing that particular memory. Moreover, because the choices are distributed in time, this mechanism may allow integrating information arriving at the hippocampus at different moments. Faced with a changing and unstable input, multiple choices provide statistical information about the environment and allow the system to correct an occasional wrong choice. If the network proceeds three times towards the attractor corresponding to A and only once towards B, it can reasonably concluded that the box is A, even if the cue was ambiguous. The freedom left to the system, from the external input, and the resetting of the state of the system at every cycle, both obtained with the introduction of the theta oscillation, are the elements generating flickering after teleportation, in our model.

9.4.2 Beside Theta

Although the model we presented was devised thinking to the results in Jezek et al. (2011), and most of its features are set to reproduce the flickering phenomenon, nevertheless the attractor dynamics modulation we describe may be considered as a general principle of the interaction between oscillations and local neural activity. Theta oscillations are known to influence the temporal expression of cell assemblies in the hippocampus also in other conditions, different from those of the teleportation experiment (Kelemen and Fenton 2010; Gupta et al. 2012; Brandon, Bogaard, Schultheiss, et al. 2013). Theta may act as a regulator of the relative strength of external inputs and internal computations beyond the extreme conditions of the experiment, maybe in less pronounced, milder ways. The periodical oscillation entraining the frequency of the selection of the attractor does not necessarily have to be theta, moreover. One can think of building a very similar model from a temporal rescaling of the one we present, substituting theta oscillations with, say, gamma oscillations (Csicsvari, Jamieson, et al. 2003; Colgin and E. Moser 2010; Buzsaki and X. Wang 2012) another prominent feature of the hippocampal LFP. The kind of input-recurrent switch we discuss, can be

obtained also at an higher frequency, perhaps nested in the theta cycle (Lisman and Buzsaki 2008; Colgin, Denninger, et al. 2009; Lisman and Jensen 2013).

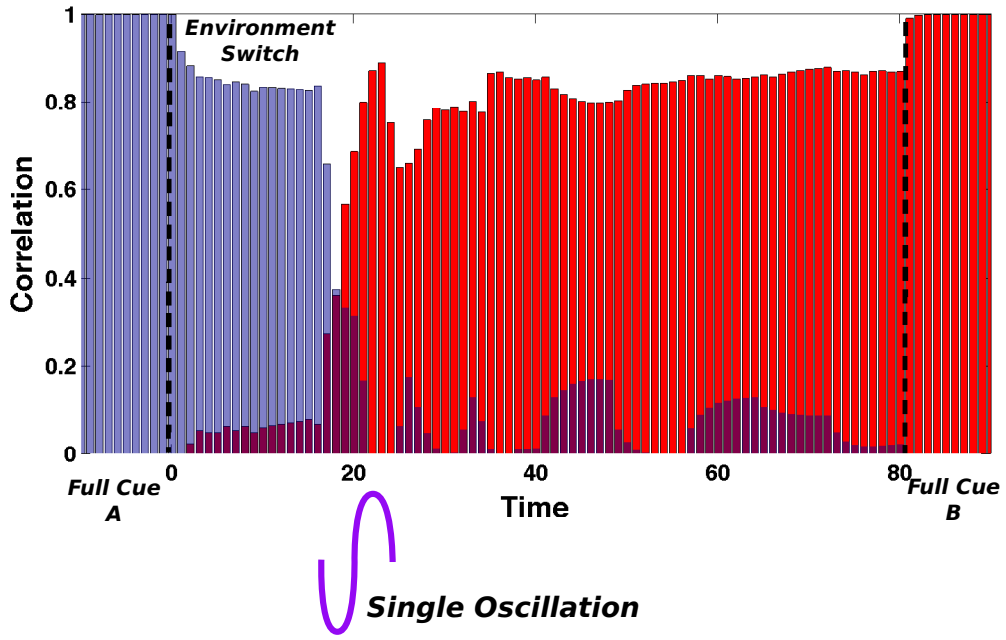


Figure 9.8: Theta modulation induces a regular and periodical alternation of the representations, but the basic mechanism can be obtained when only one oscillation is introduced. Here is an example of the same model run with no theta, except for one cycle around t=20. The single perturbation is sufficient to induce a one-time switch of the representation.

Another important aspect to point out is the periodical fashion of the attractor dynamics. The phase relationships necessary for flickering we discussed in the previous section, are necessary constraints when both the input and the network are theta modulated. In Figure 9.8 we show a different implementation of the model in a system with no theta modulation, apart from a single perturbation taking place after the teleportation, when ambiguous input is present. Initially, how we already said, the absence of oscillations leaves the system in the same attractor, unperturbed by the incoming mixed input. The figure shows how then a single oscillatory event is sufficient to take the system out of the old representation and restart attractor dynamics, in this case leading to the other representation. Regular oscillations are necessary to reproduce the flickering effect, a repetitive alternation between two attractors, but a single abrupt variation in the activity of an un-modulated network (of the kind of a SWR event) can regulate attractor dynamics as well, in a more restricted time domain (Girardeau et al. 2009).

Chapter 10

Metric Structure of the Internal Representations of Space

10.1 Introduction

What is specific about physical space, and makes it different from other correlates of neuronal activity, is its intrinsic topographical structure. A set of spatial stimuli are naturally endowed with a canonical topology and with a continuous metric, defined by the relative position of locations in the environment. These stimuli thus span a multi-dimensional manifold, which, in the typical experimental situation of a recording box, is two-dimensional and Euclidean. How spatial is the internal representation generated in CA3? How much of the external metric is preserved inside the brain? (Samsonovich and B. McNaughton 1997; Stringer et al. 2002; B. McNaughton, Battaglia, et al. 2006).

One may pretend to ignore the real-world metric, and study the metric of the virtual manifold established by the patterns of neuronal activity with which physical space has been associated (Muller and Stead 1996; Muller, Stead, and Pach 1996). How? In CA3 the movements of an animal traversing an environment elicit, on repeated trials, a distribution of responses which one can use to define distances between pairs of locations (Brown et al. 1998; Deneve, Latham, and Pouget 1999; Averbeck, Latham, and Pouget 2006), in terms e.g., of the mean overlaps in the corresponding distributions of population vectors, and one can analyze the overall structure of such pairwise distances in geometric terms. Ideally a faithful mapping of space should produce isometric representations, i.e., whose relationships mirror the relationships induced in real space by the Euclidean metric (Curto and Itskov 2008). Such an ideal mapping is, however, unfeasible with any finite neuronal population, even more so with a random self-organization process (Tsodyks and Sejnowski 1995a; Hamaguchi, Hatchett, and Okada 2006; Papp, Witter, and Treves 2007; Roudi and Treves 2008). But how to assess the degree of deviation from isometricity?

Spatial representations in CA3 depend, of course, not just on the physical structure of external space but also on how it is perceived by the animal, and on the effective dimensionality of the representation, as spanned by animal behavior (Hayman, Verriotis, et al. 2011; Ulanovsky and Moss 2011). Distant locations might be seen as similar or confused altogether, irrelevant dimensions might be ignored, e.g., on a linear track, the relative distance between locations might be distorted, not all the locations might be assigned a representation in the population, etc. In a word, the Euclidean nature

of external space may be altered arbitrarily. To start with, it is useful, however, to remove such arbitrariness and consider a model situation in which there is nothing but the Euclidean metric of physical space to be represented, through self-organization. This is what we set out to do in this study.

10.2 Metric Content

In Treves (1997), and later papers, e.g., Treves, Georges-Francois, and Panzeri (1998); Ciaramelli, Rosapia, and Treves (2006); Rosapia et al. (2007), a metric content index was introduced in order to characterize the amount of perceived metric in the representation of a discrete set of stimuli, such as faces. It was shown empirically that such metric content index is almost invariant as one varies the sample of cells used to assess the representation. Thus, it approaches the role of an objective or intrinsic measure, insensitive to the procedure used to extract it (e.g., how many and which cells are recorded in a particular experiment). Can a similar descriptor be applied to the representation of real space?

Although physical space is low dimensional, each spatial variable can span during everyday behavior a small or large interval along any of the dimensions and, crucially, can do so continuously. To actually define a set of distinct locations, for data analysis, one has in practice to discretize space in a finite number of bins and to assign to each of them a reference population vector, resulting from averaging over the activity expressed when in that bin (the actual procedure to perform this average can vary). One would want these bins to be as small as possible, to retain part of the continuity compromised by the binning. On the other hand, any refinement of the bin resolution leads unavoidably to a low sampling problem. This curse of dimensionality (Golomb et al. 1997; Panzeri et al. 2007)(where the dimensionality referred to is not of space itself, but of multiple locations in space) limits the current feasibility of the metric content analysis of a set of real, experimental data. The length of a recording session necessary to properly sample the distribution of population vectors tends to be prohibitive. One may, however, turn to computer simulations, which can be as long as needed, to produce the data necessary for the analysis.

10.3 Simulations

The full details of the model used here, which is the same we used for the estimation of information in CA3 in Chapter 7, are reported in Appendix A-1.

The mathematical model described in the Appendix, was simulated with a network of 45000 DG units and 1500 CA3 units. A virtual rat explores a continuous two dimensional space, intended to represent a 1sqm square environment but realized as a torus, with periodic boundary conditions. In each time step (of 125ms) the virtual rat moves 2.5cm in a direction similar to the direction of the previous time step, with a small amount of noise.

Before and during the learning session, all recurrent connections weights take the same value J_0^{RC}; after the learning phase, they take the values resulting from the sum of all modifications occurred during the session, and described by Eq.B-4, with learning rate γ. The trajectory of the virtual rat during the learning session is a random path, extended over a time long enough for it to effectively visit repeatedly all possible lo-

cations in space: 10000 time steps to cover the entire environment. This is taken to correspond to about 20 min of exploration in real time. Such synaptic modifications start to have an effect on the CA3 firing rate only at the end of the learning session, when the RC weights are updated to their new values.

After the learning phase for the CA3 network, we run a training phase for our decoding algorithm, during which we build the so called activity templates (E. Rolls and Treves 2011): the environment is discretized in a grid of $8x8$ locations and to each of these bins we associate a reference population vector. This reference vector for a bin is obtained averaging the activity of the network in all the time steps that the rat spends in that bin during a simulation of 100,000 time steps.

Subsequently a test phase is run. This time the rat is left wandering on a random trajectory through the environment and at each time step the activity generated in the network is compared to the templates previously defined. The dynamics is analyzed when the input coming from the DG units is either on, to characterize externally driven representation, or turned off, to characterize instead memory driven attractors (Wills, Lever, et al. 2005; Colgin, S. Leutgeb, et al. 2010). The noise level we use is kept very low ($\delta = 0.002$), as we are more interested to probe the microstructure of the spatial representation rather than to test its robustness. To have a roughly even coverage of the surface of the environment and to produce a sufficient number of visits to each one of the locations in which it is divided, simulations are run for 400,000 time steps (nearly 35hr of virtual rat time).

For simulations without DG direct input, aimed at describing attractor properties, in each step of the virtual rat trajectory, activity is allowed to reverberate for 15 time steps; with a full DG input during the first one, an input reduced to 1/3 during the second, and to 0 for the remaining 13 time steps. The final configuration attained by the system after this interval is then observed and the rat is moved to the next position and the procedure is repeated.

10.4 Analysis and Results

10.4.1 Global Metric

What does the global structure of the CA3 representation generated in our simulations look like? One may start addressing this question by using the most comprehensive measure of activity in the network. In our simulations, templates are generated using the entirety of the cells and averaging their activity over the spatial extent of the relative bin. They are the fingerprints of the representation of the environment contained in CA3. Templates are vectors in a high-dimensional space, namely the number of dimensions corresponds to the number of units in the network, 1500 in the simulations we use for the analysis. To visualize the configuration of these vectors and their arrangement, one should use some procedure to reduce their dimensionality and to produce a readable picture. One can use the correlation between the vectors associated to different positions to construct a similarity matrix containing the relative distance of all the vector pairs. Multidimensional scaling (MDS) uses this similarity matrix to assign positions to the templates in a Euclidean space of a specified dimension (in our case, 3-dimensional), so as to best preserve the ordering of the distances in the distance matrix. One can then directly compare the configuration obtained from

the algorithm to the topology of the external environment, which, in the case of our simulations, that make use of periodic boundary conditions, is a two-dimensional torus.

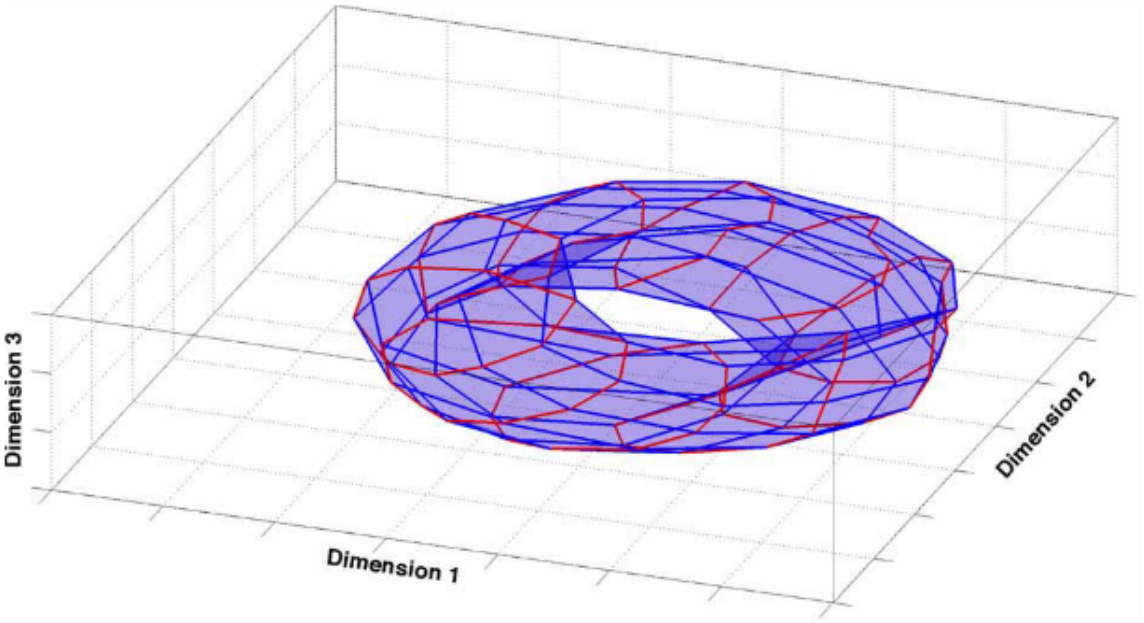


Figure 10.1: The figure shows the 3-dimensional coniguration of template vectors based on pairwise correlations obtained by multidimensional scaling (Sammon criterion (Sammon Jr 1969)), applied to the spatial representation produced by the model CA3 network. The configuration closely reproduces the original similarity matrix (Stress = 0.04). Red and blue lines correspond to orthogonal directions in the external environment.

In fact, also configurations produced by a metric MDS (Sammon algorithm (Sammon Jr 1969), Stress=0.04) tend to have a torus-like topology (Figure **??**). The global properties of the model CA3 representation thus faithfully reproduce those of external space, at least in terms of average templates. This assessment, however, is only qualitative and it misses finer details that may hide at different levels of resolution. The actual responses of the network while the animal is traversing the environment are averaged away in building the templates. Moreover a measure based on the activity of the whole population does not provide indications as to the distribution of information sampled, in practice, from the few neurons which can be recorded simultaneously, leaving doubts as to whether such a global measure has any experimental relevance. We are then left with the problem of constructing a synthetic description of the representation that comprises information on its local properties, as gauged by observing a few units at a time.

10.4.2 Decoding

To quantitatively assess the features of a neural representation of any set of stimuli we can rely on some standard procedures (Quian Quiroga and Panzeri 2009). Decoding the spike trains emitted by a population of neurons, when one (s) of a set of stimuli is

presented, means applying an algorithm that estimates, given the current spike train and those previously recorded in response to each stimulus, the likelihood for each (s) of the possible stimuli to be the current one. The stimulus for which the likelihood is maximal is the stimulus predicted on the basis of the chosen decoding algorithm. One repeats this all the times s is the current stimulus, to generate a table $P(s\prime|s)$. The decoded stimulus is not necessarily the correct one, and a first measure of the nature of the representation is just the fraction of correct hits in the table, $\sum_s P(s|s)P(s)$, where $P(s|s)$ are just the diagonal elements of the confusion matrix. A more complex, yet more complete, measure is given by the mutual information I (7.1) which also reflects the distribution of errors, and thus provides further insight on the way the stimuli are encoded (E. Rolls and Treves 2011). These two quantities, percent correct and mutual information, depend on the pool of neurons used to perform the decoding, and most crucially on the size of this pool. More cells obviously allow for better decoding.

The two quantities are not completely independent, as there are mutual constraints between them. At the same time, one does not completely define the other: given a certain percent correct there is a possible range of information values that depends on the way errors are distributed among incorrect locations. At one extreme, when there is no overlap between the representations of different stimuli, we expect errors to be distributed at chance: the distance between any pair of stimuli is maximal and effectively the same, and no non-trivial metric can be defined. Conversely, any non-uniform overlap influences the way errors are produced. The more the distribution of errors deviates from being flat, the more the representation contains overlapping and interfering elements.

10.4.3 The Spatial Confusion Matrix

During the test phase of our simulations, at each time step, the firing vector of a set of CA3 units is compared to all the templates recorded at each position in the 8x8 grid, for the same sample, in a test trial (these are the template vectors). The comparison is made by calculating the Euclidean distance between the current vector and each template, and the position of the closest template is taken to be the decoded position at that time step, for that sample. This procedure applies to our spatial analysis what has been termed maximum likelihood Euclidean distance decoding (where the distance between population vectors should not be confused with the distance between locations in the environment). The frequency of each pair of decoded and real positions are compiled in a so-called confusion matrix. Should decoding work perfectly, in the sense of always detecting the correct position in space of the virtual rat, the confusion matrix would be the identity matrix. The confusion matrix is thus L^2xL^2 (64x64 for our simulations) and its dimension grows very fast when increasing the number of bins, requiring a prohibitively longer number of time steps to properly fill all its entries.

The confusion matrix gives us the value of the percent correct and of the mutual information. The size of the sample is then varied to describe the dependence of these quantities on the number of cells in the pool. We used samples of up to 256 units, a number which can be compared with the total population in our CA3 network, 1500 units.

10.4.4 Metric Content

For a given percent correct (f_{cor}) there is in general, in a non-spatial paradigm, a certain range of possible amounts of information contained in the confusion matrix. Ideally the information should be comprised between a minimum value

$$I_{min} = \log_2 S + f_{\text{cor}} log_2 f_{\text{cor}} + (1 - f_{\text{cor}}) log_2 (1 - f_{\text{cor}}) - (1 - f_{\text{cor}}) log_2 (S - 1) \quad (10.1)$$

(where S is the number of elements in the stimulus set) corresponding to an even distribution of errors among all the possible stimuli, while the maximum is attained when all the errors are concentrated on a single incorrect stimulus

$$I_{max_{bias}} = \log_2 S + f_{\text{cor}} log_2 f_{\text{cor}} + (1 - f_{\text{cor}}) log_2 (1 - f_{\text{cor}}) \quad (10.2)$$

However, this maximum corresponds to a systematic misclassification of the current location by the network. It might therefore be reasonable to assume that our system is an unbiased classifier, which implies that incorrect stimuli can at most be chosen as frequently as the correct one, and reformulate the previous maximum in the following terms

$$I_{max} = \log_2 S + \log_2 f_{\text{cor}} \quad (10.3)$$

The metric content index can then be defined, in such a non-spatial paradigm, as

$$\lambda = \frac{I - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}} \quad (10.4)$$

This is the measure considered in our previous studies. With this choice, though, we wouldneglect the intrinsic topological structure of spatial information. In fact, the high correlation existing between the representations of neighboring locations constrains the distribution of errors. We need to distinguish between errors originating from random correlations in the representation of distant, unrelated locations, and errors emerging form the continuity of the representation. The latter can be considered as structural to the representation and they are expected to be present around the correct location even in the case of optimal decoding, but with limited spatial resolution.

10.4.5 Full and Reduced Matrix

The confusion matrix conveys information which is location-specific: we know for each of the locations how it was decoded during the test phase. The appearance of a typical example of these location-specific matrices shows how a decoding approach reveals characteristics of the representation otherwise overseen (Figure 10.2A). Far from being close to the real position, the decoded positions appear to be distributed in multiple locations over the environment, often far away from the correct spot. The pattern of distribution of decoding probability depends on the choice of the sampled neurons, and on the size of the sample, and it seems to lack any regular principle of organization.

If any regularity in the distribution of errors across positions exists, we can try to reveal it by extracting the translational invariant component of this distribution. By deriving a position-averaged version of the matrix (Figure 10.2B), we construct a simplified matrix $Q(x - x0)$, which averages over all decoding events with the same vector displacement between actual ($x0$) and decoded (x) positions. $Q(x - x0)$ is easily constructed on the torus we have used in all simulations, and it is only LxL, much smaller

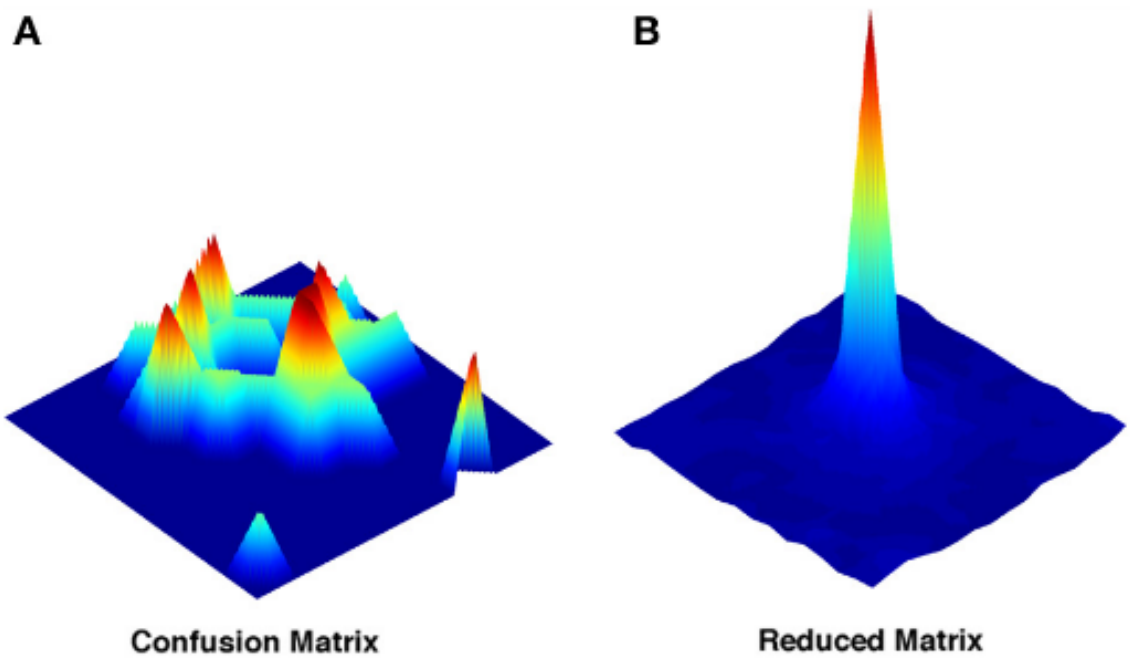


Figure 10.2: Representative examples of the full and of the reduced confusion matrix. (A) Decoding probability distribution for a single location extracted from the complete confusion matrix. (B) Decoding probability distribution obtained from the reduced matrix. Both are obtained with a sample size of N = 4. Color coding just reflects the relative heights of the points in the probability distributions.

than the complete confusion matrix.

The two procedures, given that the simplified matrix is obtained just by averaging the full confusion matrix after a row translation, might be expected to yield similar measures, but this is not the case (Cerasti and Treves 2010). In fact the amount of information that can be extracted from the reduced matrix is significantly inferior to that of the full matrix, even if the difference decreases as the sample of neurons gets larger (Figure 10.3A). The discrepancy between the two measures reflects the presence of a distribution of errors which is not translational invariant. The distribution found in the reduced matrix instead represents the performance of the system when the effects of specific firing configurations are averaged away. Being quasi-randomly distributed, the errors of the complete matrix average out and produce a smooth, radially decreasing distribution around the central, correct position in the reduced matrix (Figure 10.2B). We are thus able to extract the average error dependence on the distance between two points, regardless of their specific position. The reduced matrix is the expression of the overlaps induced by the external metric of space together with the continuity of the internal place field representation.

All the quantities of interest, the information contained in the full and reduced matrix and the percent correct, have a dependence on the sample size which can be fit rather

precisely by a sigmoid function. The function we use to fit the information is

$$I_{\text{fit}} = \frac{I_{\text{sat}}}{1 + (n_0/n)^b} \tag{10.5}$$

where n is the number of units in the decoding sample and I_{sat}, n and b are the fit parameters. An analogous form is used for the percent correct

$$f_{\text{fit}} = f_{\min} + \frac{f_{\max} - f_{\min}}{1 + (n_0/n)^b} \tag{10.6}$$

The change in the convexity of the data is particularly evident when using a log scale for the number of unit in the sample (Figure 10.3A, B).

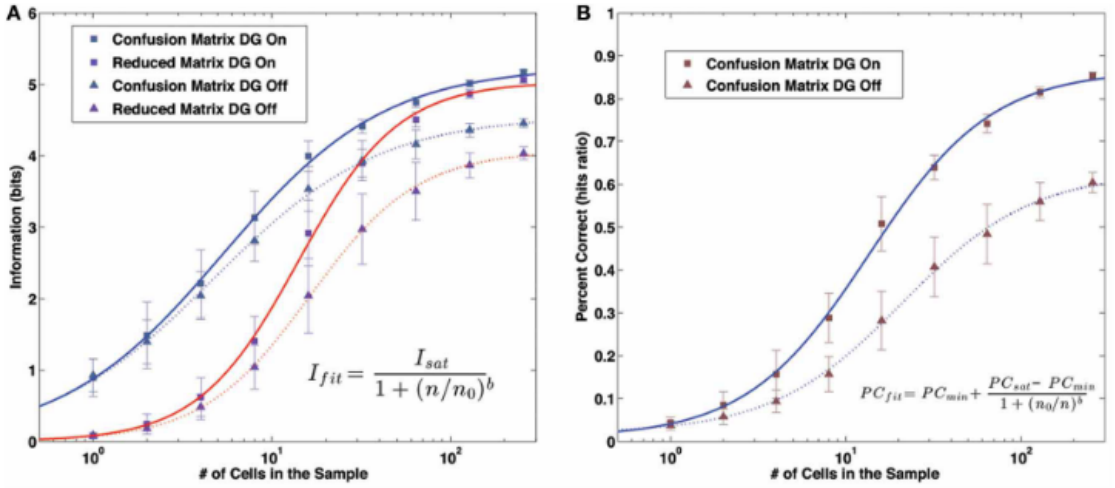


Figure 10.3: Information and Percent Correct dependence on sample size. (A) Information values. Markers: model data. Lines: fit curves. Gray markers/blue lines: information content of the confusion matrix. Purple markers/red lines: information content of the reduced matrix. Square markers/solid lines: results with the DG input active. Triangular markers/dashed lines: results without the DG input. (B) Percent Correct values. Markers: model data. Lines: fit curves, as in (A).

10.4.6 The Metric Induced by an Active DG

We can use the information contained in the reduced matrix to separate the effects of the external metric on the CA3 representation from the other sources of correlation. For this aim, the condition with the active DG is better suited, as we consider the original map imposed by the external input, before the modifications induced by the storage, which will be described in the following section.

If we fit the distribution of values in the reduced matrix with a Gaussian we can extract the parameters describing the distance dependence of the errors, namely the height of the central peak (pc), the width of the distribution (w) and the total volume below the distribution (a). They correspond, respectively, to the number of correct hits, the spread of errors around the central location and to the proportion of decoding steps associated with a single location. These parameters depend on the sample size (Figure

10.4): a larger sample corresponds to a higher peak of the Gaussian (corresponding to an increase in the percent correct), to a lower standard deviation around the mean and to a higher total volume. It should be noted, however, that the observed width w bundles together the distribution of outright errors and the spread of correct but spatially imprecise responses.

A non-zero standard deviation σ, which can then be regarded as a component of the width w, expresses the difference between decoding a set of spatial stimuli and decoding a non-spatial one. We can use this measure of the structural confusion between locations of the environment to reformulate our measure of metric content for the case of spatial information. We can argue that given a certain percent correct, the minimal information would be obtained when the decoding distribution corresponds to a Gaussian of width σ and total volume a centered on the correct location, plus the remaining $(1-a)$ evenly distributed on all the spatial bins. Analogously the maximal information attainable would correspond to the situation in which $1/a$ Gaussians of the same shape sit on the same number of different locations (one of them, of course, should be placed on the correct one). Indeed this is a first maximum, and it corresponds to the maximal unbiased information defined above for the non-spatial case.

For the spatial case, however, these minimum and maximum information values largely reflect simply the definition of the reduced and full confusion matrix. They tell us about the procedure used in the analysis more than about the representations being analyzed. We can however go further, as we can also extract the limit towards the discrete case for our model, attained when the spatial resolution is optimal, i.e. the spatial code is precise, spatially exact. By sending the standard deviation σ to zero, in the limit, and by replacing the Gaussian distribution with a single peak of height a located in the central spot, we retrieve the situation in which, in the absence of errors due to fluctuations in the topology, only the retrieval (i.e., identification) of a certain location is taken into account. This upper maximum does not differ from the previous one in terms of the number of different locations erroneously decoded as the correct one, but it modifies the way in which individual decoded locations are distributed around them. Of course all the conditions between these two extremes can be obtained, using an intermediate value of the standard deviation.
We can then define a spatial descriptor of the metric content applicable to spatial representations, as

$$\chi = 1 - \frac{\sigma}{w} \tag{10.7}$$

Note that $\chi = 0$ implies that the entire width of the Gaussian in the reduced confusion matrix is due to the poor spatial resolution, as decoding has a standard deviation $\sigma = w$; $\chi = 1$ instead implies that decoding is spatially exact, and the apparent width w emerges entirely from averaging the errors in the full confusion matrix. We can call χ the metric resolution index.

We can see where the data we generated with our simulations sit in relation these three reference curves (Figure 10.5A). For simplicity and clarity, instead of the original data we will use the fit curves we previously calculated. For each sample size, the parameters of the fit define the two maxima (lowest and highest dark blue dashed curves for $\chi = 0$ and $\chi = 1$, respectively) and the minimum (dark red dashed line) for the information, and we can compare then with the actual value of the simulated network,

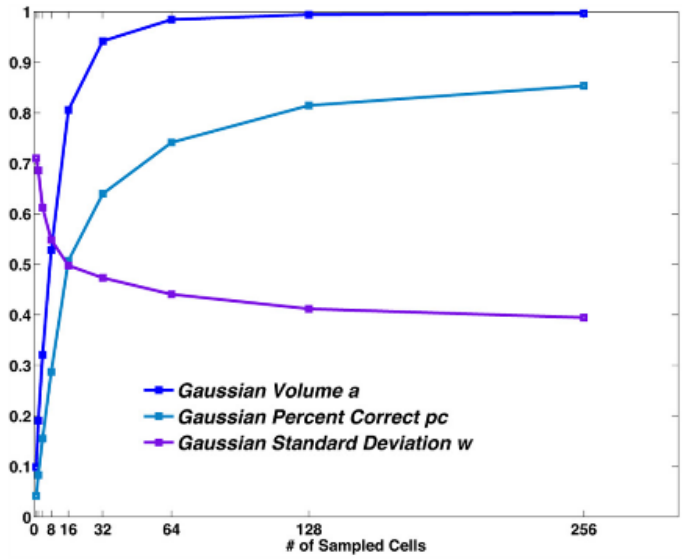


Figure 10.4: The Gaussian distribution parameters vs. sample size. The values obtained from the fit of the central bump of the reduced matrix, are plotted for different sample sizes. Pc and a are probabilities. w is measured in spatial bin units (each bin is 12.5 cm in size).

both the value extracted from the complete confusion matrix (blue solid line), and the value coming from the reduced one (Figure 10.5A, red solid line). Both series of values appear to sit on a curve of constant metric resolution. The metric resolution expressed by the complete confusion matrix is just above the one defined by our first maximum curve, in fact it sits roughly at $\chi = 0.1$. This corresponds to having a matrix in which there are essentially no randomly distributed errors, while those associated with specific locations are spread out locally, with resolution $\sigma = 0.9w$.

As expected, the form of the reduced matrix is quite different. It belongs to a complete different category: here the errors are almost entirely randomly distributed, with the exception of those giving rise to the central bump.

This measure of metric resolution, thus, extends the previous, non spatial notion of metric content, and captures the qualitatively different nature of the information expressed by the two different types of matrices, full and reduced. It allows us to shed light on the way spatial information is encoded in a CA3-like network by quantifying the presence and the relative importance of the different sources of confusion present in the system. Moreover it indicates that the nature of the confusion is an invariant of the spatial code expressed by the network and it does not depend on the size of the sample of cells used for the decoding.

10.4.7 Metric Resolution without DG input

One may ask how the procedure previously described applies to the other condition, in which the DG input is removed and the representation starts to drift away from the starting position. As already shown by (Cerasti and Treves 2013) the removal of the external drive causes a major drop in the information about the position of the rat

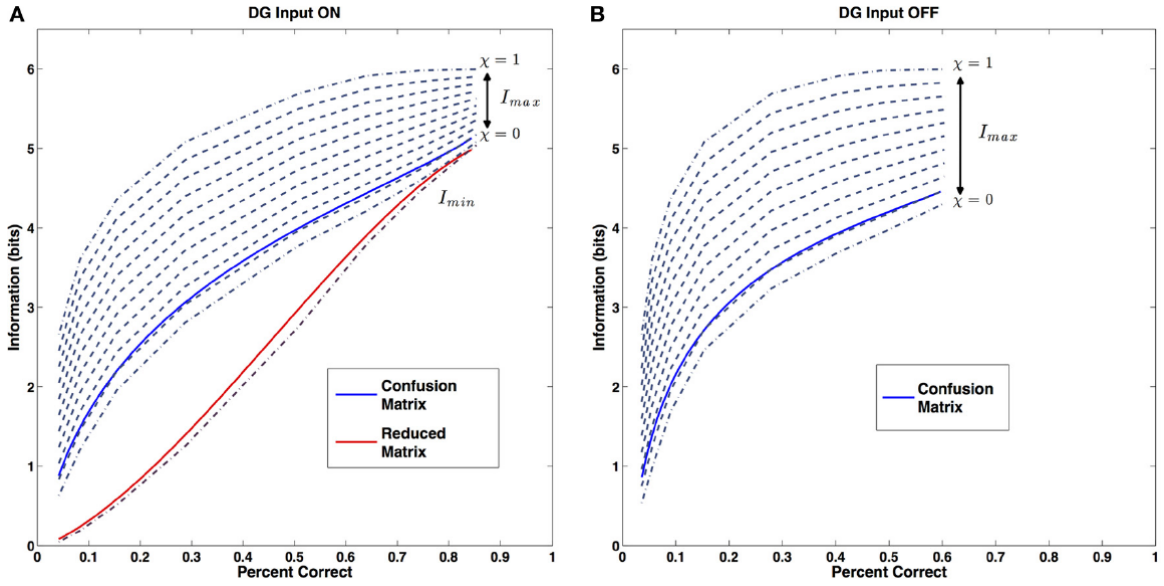


Figure 10.5: Metric Resolution. (A) Metric resolution of the CA3 representation generated with the external DG input. Blue solid line: information content of the confusion matrix (close to $\chi \sim 0.1$). Red solid line: information content of the reduced matrix. Dark red dashed line: theoretical minimum for the information. Dark blue dashed lines: theoretical maxima of the information for progressively decreasing values of the error dispersion. (B) Metric resolution of the CA3 representation based on the internal collateral activity only (also close to $\chi \sim 0.1$). Color code, same as (A). Only the metric resolution of the full confusion matrix is shown.

retrievable from the system (also shown in the figure). This drop comes with a parallel decrease in the percent correct obtained by the system (Figure 3B).

The main effect of drift on our analysis is to introduce a systematic bias in the decoding of some of the positions in the environment. Since we discretize space in bins of side L, every time drift crosses bin boundaries, the decoded position will belong to a neighboring bin. Not all the points in the same bin are decoded as belonging to another bin, however, some of them will be assigned correctly, while those who are not may move in different directions to different neighboring bins.

This phenomenon introduces a series of non-negligible effects in our measure. It undermines one of the fundamental assumptions of the approach, the unbiased nature of the decoding. In many cases, in those locations where the drift is strong enough, the maximum of the decoding matrix will be associated to an incorrect bin. Moreover it invalidates the use of the reduced matrix to extract information about the distribution of errors around the correct location. Being limited in space, the bump obtained in the reduced matrix is a combination of metric errors and of neighboring bins reached by the drift. This makes the parameters of the bump very unreliable and of little use.

It turns out, however, that the effects of drift can be effectively accommodated in the analysis. To estimate the metric content of the representation generated without the contribution of the external input we follow the same procedure used in the DG On

condition with a single modification. We calculate the minimal and maximal information levels as in the previous case, using the same parameters extracted from the previous fits and therefore still coming from the reduced matrix of the DG On case. The only adjustment we introduce is to impose that each of the 1/a bumps present in the confusion matrix is transformed to a combination of two adjacent Gaussian bumps, with the same width of the original, but half of the height. In this way the amount a of hits associated with the location is preserved but the drift away from (some of) the locations comprised in the bin is taken into account. This solution is an average of the effects present at different locations in the environment, as in some of them no drift is taking place, while in others the phenomenon is present with different strengths and a variable number of directions.

The result proves that this simple procedure captures very well the effects of drift on the properties of the representation and justifies the 2-bump choice. As shown in Figure 10.5B, the data sits on a curve of constant metric resolution, which happens to coincide with the one expressed in the previous condition.

The calculation of the metric resolution offers insights on the characteristics of this representation. In fact it shows how, factoring out the drift phenomenon, the representation expressed by the recurrent collaterals alone appears to have the very same metric properties as the one obtained with the active contribution of the DG. For each sample size, the distribution of errors around the central location is indeed the same, as are the number of locations erroneously chosen by the decoding algorithm. As in the previous case, no portion of the errors is allocated randomly. This consistency is noteworthy as we applied a method adapted, in the previous study, to a completely different set of data, where both the information and the percent correct values are drastically different. The metric content analysis shows a remarkable persistence of the relationship between these two quantities, even when the external input is removed

10.5 Discussion

The ability of the hippocampus to operate as a cognitive map and to successfully guide spatial behavior (R. Morris et al. 1982) lies in the amount of information about the external environment that can be stored and retrieved from its representations (Battaglia and Treves 1998; Káli and Dayan 2000). In particular one would expect that the global topography expressed by these representations, the set of relationships existing among the patterns of activation associated with the different points of the environment, should reflect the one which the animal actually experiences (Curto and Itskov 2008; Dabaghian et al. 2012). If we imagine an internal observer trying to decode the position of the rat, and the structure of the environment in which it is moving, only relying on the sequence of units that get activated in the hippocampus, similarities between the population vectors would be the most probable criteria to infer the geometrical location of points traversed at different times, or to predict the existence of traversable routes (Gupta et al. 2010; Dragoi and Tonegawa 2011).

In fact we observe that the topology realized by the averaged activity of our network actually reproduces the torus in which the virtual rat of our simulations is moving. From such a coarse-scale perspective, a CA3 representation is a true reconstruction of

the metric the rat is experiencing. Within minor deviations, the correlation between two average population vectors is a good estimate of the distance of the two locations over which their individual population vectors were recorded.

But does such a general approach capture all the variability expressed by the network during its normal operation? We actually find that at a finer scale, the structure of a CA3 spatial representation is richer than it appears at the coarser scale. The difference we observe, between the information content of the full confusion matrix and its reduced version, is telling us about other, non-metric sources of correlation between different parts of the representation. This difference appears when decoding is based on limited samples of network units, but the effects persist up to substantial sizes of the sample, and one should also consider the experimental impracticality of a decoding procedure based on the whole population.

To quantify the properties of the finer aspects of the spatial representation we propose the use of an index, the metric resolution, that like the metric content used earlier with non-spatial stimuli, combines the measures of information and percent correct, which are obtained from a confusion matrix. The index allows for an assessment of the metric in the internal representation of space. The measure of metric resolution is applied to our model of a CA3 network in two different regimes, when driven by an external input coming from DG, and when this external input is absent and the active representation is solely an expression of the activity reverberating in the collateral connections. In both cases we find that the metric resolution of the representation is roughly constant, at a value $\chi \sim 0.1$, when the size of the sample is varied, and that the same combination of metric and non-metric structure is present in the confusion matrix generated from these different samples. The metric resolution measure tells us that the self-organizing process has generated an internal representation of space, in the model, which is consistent with external space but has limited metricity, i.e. limited spatial resolution, close to the minimum amount possible (which would yield $\chi = 0.0$). We acknowledge the possible dependence of the results on the size of the network that represents the real CA3. Although a network of 1500 units is significantly smaller than the actual rodent hippocampus, in some respects it can produce representations which faithfully reproduce the characteristics of the ones experimentally observed (Cerasti and Treves 2010; Cerasti and Treves 2013). Moreover the difference in size is less extreme if one regards the real CA3 network as comprised of partially independent sub-modules. The fact that metric content, in our model system, is independent of sample size and of the mode of operation (whether or not externally driven) is an indication of the robustness of this measure.

The errors in decoding, that do not depend on the similarity in the representation due to close distance, might be seen as hindering the proper operation of the hippocampal memory system, by introducing ambiguities in the reconstruction of the external space. But another possibility is that these holes in the consistency of the representation offer the hippocampus a handle to introduce non-spatial elements of information in the representation itself (Wood et al. 2000; S. Leutgeb, J. Leutgeb, Barnes, et al. 2005; Singer et al. 2010; Eichenbaum, Sauvage, et al. 2011). Thus the metric content analysis reinforces the suggestion (Cerasti and Treves 2013) that the CA3 network, with its effectively random drive from the Dentate Gyrus, might be best adapted to serve a

memory function, rather than to produce faithful representations of space.

Chapter 11

Conclusions

Present day proposals on the functioning of the hippocampus can be seen as the result of the convergence of two different line of thoughts. On one side the perspective of the hippocampus as a part of the medial temporal lobe circuit for spatial representations, mostly based on neurophysiological studies and single-cell recordings in rodents (O'Keefe and Dostrovsky 1971), on the other side the classical neuropsychological idea about the role of the hippocampus in the formation of declarative memories (Scoville and Milner 1957), certainly more respectful of its central position in the brain, but at the same time often unrelated to the concrete cellular and network mechanisms underlying cognitive functions in this part of the brain.

The emerging synthesis is moving away from considering a mere involvement of the hippocampus in spatial navigation, which it would subserve by providing just a where-am-I signal. Instead hippocampal place cell representations are seen more and more as the substrate with which rodent brains organize more complex and assorted kind of memories, comprised of both spatial and non-spatial elements, more resembling the notion of episodic memory. The position of the hippocampus at the converging focus of multiple streams, and the identification of the neural representations expressed in the regions part of the different streams, are pushing toward a characterization of the different networks comprised in the medial temporal lobe on the basis of the relationships occurring between them and on the way specific representations can be obtained from the processing of afferent inputs.

Of particular importance for the development of this conceptual framework has been the discovery of grid cells in the medial portion of entorhinal cortex (Hafting, Fyhn, Molden, et al. 2005). They proved the existence of not only spatial information, but of highly processed spatial information, present in the form of beautifully regular hexagonal pattern of activation, just at the gates to the hippocampus. The discovery accelerated the effort to differentiate the function of the hippocampus from that of the neighboring regions comprising the medial temporal lobe, insisting on their higher specialization (e.g. mEC in expressing spatial information, LEC in conveying context discriminant signals, but poorly spatially informative) with respect to the hippocampus, which is thought to "put together" all the pieces.

This thesis work investigates some of the aspects of this proposal. Place cell and grid cell representations are discussed in the frame of two different models, focusing on the peculiarities of these two phenomena and trying to relate them to their functional role.

In the case of grid cells we address the point of how such a pattern of activity can emerge from raw spatial inputs. In our model based on adaptation (Kropff and Treves 2008) the grid pattern is the result of a gradual self-organization process that, over the course of an extended time window, arranges the grid fields to form a regular tiling of the environment. The model is devoid of any pre-wired element, and the formation of the grid pattern is entirely based on the animal active exploration of space. This fact is reflected on the dependance of the particular grid pattern generated on the metric structure of the external environment. We show how, without the need of changes or additional ingredients, the same network that generates hexagonal grids on a flat surface, produces grids with different coordination number (from 1 to 5) on the surface of a sphere (Chapter 4), and a combination of Face Centered Cubic and Hexagonal Close Packed arrangement of fields in the three-dimensional space (Chapter 5). Thus, we argue that the hexagonality of grids observed in the experiments is a consequence of the particular topology of the environment used, and not an intrinsic property of grid cell activity. Our model presents the formation of grids as a spontaneous tendency of the system (guided by adaptation) to attain a regular tessellation of the environment presented to the animal during a critical developmental period.

In Chapter 6 we discuss the effect of introducing inhibitory interactions among units in the mEC layer of the model. Simulating the network with this kind of internal connectivity, that is initially random and then self-organizes together with the rest of the system, leads to the emergence of regular grids with common orientation, even though the model still fails to produce an even distribution of the phases, which tend to be clustered in space. Interestingly, however, the resulting inhibitory connectivity shows the typical pattern used to wire continuous attractor networks, with a neat dependence of the probability of a connection on the phase difference between two units.(Zilli 2012; Couey et al. 2013) This implementation of the model, then, could represent a possible link between the initial generation of the grid pattern through self-organization, and the implementation of a continuous attractor in mEC, of the kind proposed by many models of grid cells. Some of the assumptions of the model, like the once-for-all definition of the connectivity in mEC, in a critical developmental window after which no plastic change in the synaptic weights is possible, may have interesting consequences on the use of the network to produce grid patterns after this critical period. Namely, grids would reflect the fields organization originally developed by adaptation, induced in turn by the metric of the explored environment.

The reproducibility of grids over different environments is in sharp contrast with the seemingly random arrangement of place fields. While the focus of the grid model was on the possibility of generating a unique, extremely regular chart from any spatially modulated input, even the more disordered one (Kropff and Treves 2008), in the case of CA3, we address an almost opposite issue, that is the effects of disorder and multiple charts on the attractor dynamics of the region and on its ability to faithfully represent spatial location.

In Chapters 8 and 10 we present a series of results pointing towards a fundamental inadequacy of the CA3 network to provide an accurate spatial representation of the environment. Indeed, the internal metric of space found in our model of CA3 presents

distortions of the local topology and introduces non-metric correlations between different parts of the representation. Moreover, the recurrent dynamics of the network in the absence of a constant external input, is far from that of continuous attractor. The activity configuration "drifts" away from the initial position, towards a limited number of stable locations in the environment. It also transit from one chart to another, in the presence of multiple stored maps. The results are at odds with a popular view of the hippocampal attractor landscape as composed by a collection of flat manifolds, representing locations in an environment, well separated from each other and clearly distinguished (the different environments) (Samsonovich and B. McNaughton 1997). In our model the resulting "energy" landscape inside each chart is not flat, and moreover different charts (though randomly generated and largely uncorrelated) are connected by sorts of wormholes which enables the bump of activity to move from one to another.

This situation is reflected in the spatial information provided by such a network (Chapter 7). When compared to that conveyed by a grid map of the environment, the latter clearly comes out as the superior winner. But the situation is reversed when taking into account the ability of the two networks to discriminate between different environments. Here the high variability in the CA3 maps is an advantage to clearly distinguish different contexts.

The nature of hippocampal representations seems to promote a trade-off between the ability to convey precise and accurate information about specific elements of the external world (in this case, spatial position), and the ability to distinguish and recognize unique combinations of these elements (in our example, the context). But if the internal dynamics of CA3 are functional to process information, the network needs a way to address the content of its memories, independently from the drive of an external input. In Chapter 9 we present a possible solution to the conflict between the need to collect inputs from the afferent regions, and the elaboration of this input through recurrent dynamics. It involves the presence of the theta rhythm regularly modulating the activity of the area. In our model, the two functional regimes are segregated to different phases of the oscillation, with the external input dominating the half with higher inhibition, and attractor dynamics taking over during the other part of the cycle. This sort of effect allows the system to periodically retrieve an attractor, not necessarily the same one, on the basis of the changing external input. The model reproduces the results of the teleportation experiment (Jezek et al. 2011), but it might also offer a more general hint at the interaction between oscillations of any sort (like gamma, or sharp wave ripples) and recurrent dynamics.

Taken together these results, contribute to the characterize the complementarity between grid cells and place cells. On one side the focus is on generating a prototypical arrangement of fields and in permanently fixing the phase relations among units. On the other, the accent is on the properties that derive to the representation from the flexible generation of multiple and heterogeneous cell assemblies. Both processes rely on plastic changes of set of connections, and both of them are described in our models as the outcome of a self-organization process, but in the case of mEC the learning happens only once, gradually, on an extended time scale; in the case of the hippocampus the units contributing to a representation are randomly recruited from a vaster pool, and the relationship between their activity pattern is highly dependent on chance, or

possibly on active mechanisms like the accumulation of place fields at the goal location. We argue that the difference in the way the two representations are formed is reflected in the different functional role of the two regions, with the circuitry of mEC cortex specifically designed to provide ready-to-use, accurate metric relations (not necessarily of locations in space), and the hippocampal representations expressing relationships between different elements of the external world, whose combinations it is necessary to rapidly encode in memory.

Appendix A

In this Appendix we present an analytical approach to the problem of dynamics in the presence of disordered place fields centers. To do so we consider a minimal formulation of the problem, N place cells are modeled by interacting binary units σ_i (Monasson and Rosay 2013)

$$\sigma_i = 0, 1 \tag{A-1}$$

corresponding to, respectively, silent and active state.
We suppose that after learning each cell is randomly assigned a preferential firing location, its place field. The centers of the place fields are randomly arranged on the environment, which, for simplicity, is taken to be a one-dimensional segment of length L. The place fields are represented by a Gaussian bump:

$$\phi(x) = \frac{1}{(2\pi w^2)^{d/2}} \exp(\frac{-x^2}{2w^2}) \tag{A-2}$$

where x is the position in the environment.
Interactions between units depend on the extent of the overlaps between their place fields:

$$J_{ij} = \int_V dx \phi(x - x_i)\phi(x - x_j) \tag{A-3}$$

We assume that the total neural activity is fixed by the action of a global inhibition. We introduce the parameter f to denote the fraction of active cells:

$$\sum_{i=1}^{N} \sigma_i = fN \tag{A-4}$$

The couplings J_{ij} and the constraints on the activity define the activity of the network $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$ which is assumed to be drawn from the probability distribution:

$$P_J(\boldsymbol{\sigma}) = \frac{1}{Z} \exp(-\beta H) \tag{A-5}$$

where the energy of the configuration is

$$\mathcal{H} = -\sum_{i<j} J_{ij}\sigma_i\sigma_j \tag{A-6}$$

and Z is the partition function.
If we introduce replicas

$$\bar{Z}^n = \int_V \prod_i dx_i \sum_{\sigma_i^a}^{*} e^{\beta \sum_{i<j} J_{ij} \sum_{a=1}^{n} \sigma_i^a \sigma_j^a} \tag{A-7}$$

(where $\sum^*$ stands for the sum over the configurations constrained by Equation A-4), the conjugated variable $y(x)$ (which represents the density of activity in space) and we consider the set of scalings:

$$\rho = \frac{N}{L}, N, V \to \infty \tag{A-8}$$
$$\phi(x) = \hat{\phi}(\hat{x})/L \tag{A-9}$$
$$x = \hat{x}L \tag{A-10}$$
$$y^a(x) = \hat{y}^a(\hat{x}) \tag{A-11}$$

after some calculations we get to a free energy of the form:

$$\bar{Z}^n = \int \prod_a D\hat{y}^a(\hat{x}) e^{-N\mathcal{F}[\{\hat{y}^a(\hat{x})\}]} \tag{A-12}$$
$$\mathcal{F} = \frac{1}{2\beta\rho} \sum_a \int d\hat{x} \hat{y}^a(\hat{x})^2 - \log \left[\int d\hat{r} \sum_{\sigma^a}^{*} e^{\sum_a \int d\hat{x} \hat{\phi}(\hat{x}-\hat{r}) \hat{y}^a(\hat{x})} \right] \tag{A-13}$$

By assuming replica symmetry:

$$\hat{y}^a(\hat{x}) = \hat{y}(\hat{x}), \tag{A-14}$$

we finally obtain:

$$\mathcal{F}/n = \frac{1}{2\beta\rho} \int d\hat{x} \hat{y}(\hat{x})^2 - \int d\hat{r} \log \left[1 + e^{\mu + \int d\hat{x} \hat{\phi}(\hat{x}-\hat{r}) \hat{y}(\hat{x})} \right] \tag{A-15}$$

with μ used to enforce the constraint:

$$\int d\hat{x} \hat{y}(\hat{x}) = f\rho \tag{A-16}$$

The minimization equation with respect to y reads:

$$\frac{1}{\beta\rho} \hat{y}(\hat{x'}) = \int d\hat{r} \frac{e^{\mu + \int d\hat{x} \hat{\phi}(\hat{x}-\hat{r}) \hat{y}(\hat{x})} \hat{\phi}(\hat{x'} - \hat{r})}{1 + e^{\mu + \int d\hat{x} \hat{\phi}(\hat{x}-\hat{r}) \hat{y}(\hat{x})}} \tag{A-17}$$

the one for μ is instead:

$$f = \int d\hat{r} \frac{e^{\mu + \int d\hat{x}\hat{\phi}(\hat{x}-\hat{r})\hat{y}(\hat{x})}}{1 + e^{\mu + \int d\hat{x}\hat{\phi}(\hat{x}-\hat{r})\hat{y}(\hat{x})}} \tag{A-18}$$

We numerically solve the system of equations A-17 A-18 varying the value of ρ (Figure 1). By increasing the density of fields the solution switch from flat to localized and then get increasingly peaked. In Figure 2 we plot the standard deviation of the minimization as a function of ρ (with $f = 0.1$ and $w = 0.05$).
We can then calculate the energy $\mathcal{F}$ of the solution (Figure 3) and compare it to that of a uniform distribution in space of the activity density:

$$\hat{y}(\hat{x}) = f\rho \tag{A-19}$$

The minimal energy initially coincides with that of the uniform distribution (and indeed the solution to A-17 is almost flat for small values of ρ), but above a certain density (the same of the transition to localized solutions in Figure 2) it departs from it becoming energetically favorable. Moreover an higher field density correspond to lower energies.

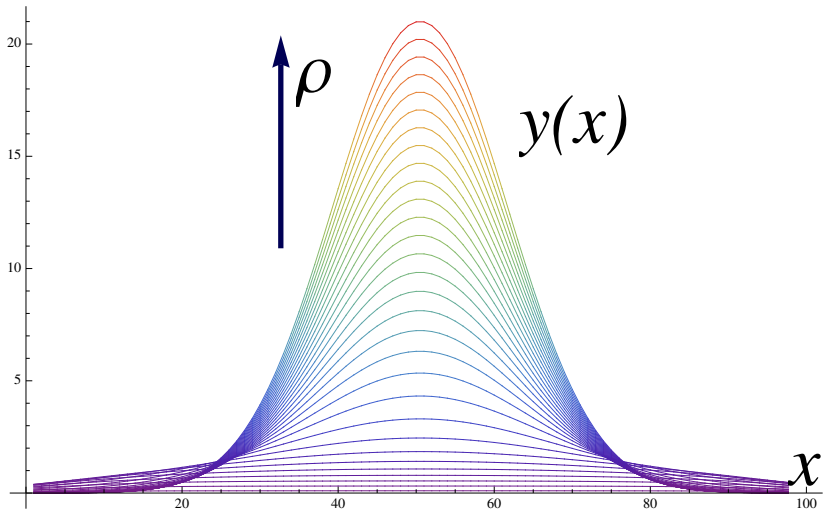


Figure 1: Plot of the spatial shape of the minimization solution for different values (colored lines) of the fields density ρ, 1D

The results have a dependency on the parameters w and f but are qualitatively the same.

A-1 Drift

To quantify the amount of drift produced by our activity probability distribution, we introduce an external field in the hamiltonian, representing an initial cue, initializing the bump of activity around the point $x = 0$. The external field h will be then sent to 0, to estimate the displacement.

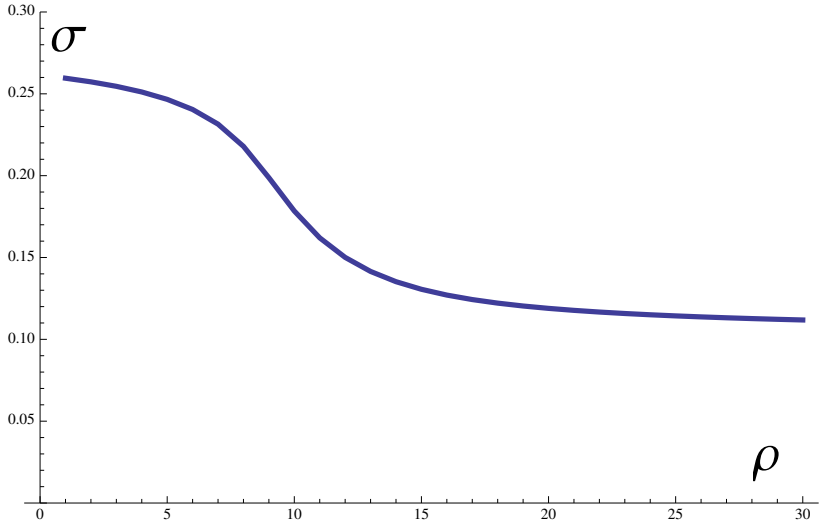


Figure 2: Standard deviation of the minimal energy solution as a function of the fields density ρ, 1D

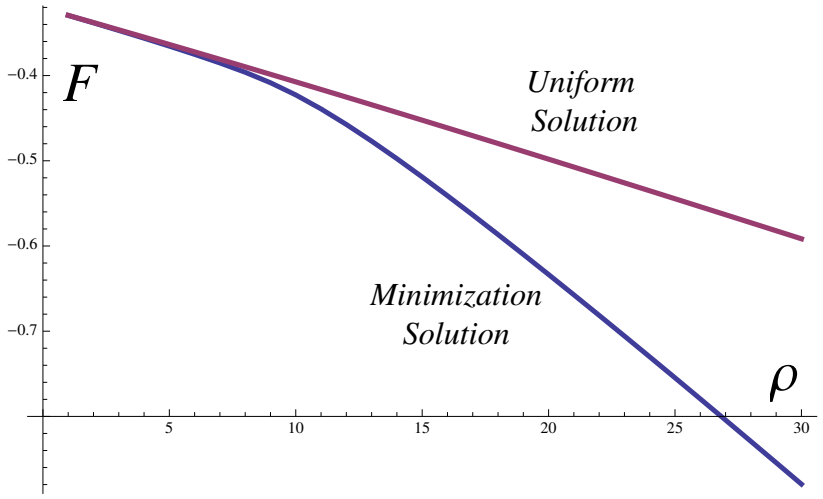


Figure 3: Energy of the solution as a function of the fields density ρ, 1D

The hamiltonian is then:

$$\mathcal{H} = -\frac{1}{2}\sum_{ij} J_{ij}\sigma_i\sigma_j - h\sum_i \sigma_i x_i \tag{A-20}$$

which gives a Partition Function:

$$\overline{Z^n} = \int \prod_a Dy^a e^{-\frac{N}{2\beta\rho}\sum_a \int dx y^a(x)^2 + N log[\int dr \prod_a (1+e^{dx\phi(x-r)y^a(x)+h\beta r})]} \tag{A-21}$$

Considering Replica Symmetry we get:

$$\overline{Z^n} = \int Dy e^{-\frac{nN}{2\beta\rho}\int dx y(x)^2 + nN log[\int dr (1+e^{dx\phi(x-r)y(x)+h\beta r})]} \tag{A-22}$$

The drift can be estimated in two ways:

$$D^2 = \frac{1}{(fN)^2}\overline{<(\sum_i \sigma_i x_i)^2>} = \frac{1}{(fN)^2}\overline{\sum_{ij} <\sigma_i\sigma_j> x_i x_j} \tag{A-23}$$

$$D^{\dagger 2} = \frac{1}{(fN)^2}\overline{(\sum_i <\sigma_i> x_i)^2} = \frac{1}{(fN)^2}\overline{\sum_{ij} <\sigma_i><\sigma_j> x_i x_j} \tag{A-24}$$

where $<\cdot>$ stands for the mean over the probability distribution of σ, and $\overline{(\cdot)}$ is instead the quenched average over the arrangement of the fields.The two definitions of displacement differ in the way the average over the "thermal" noise is performed and we expect that $D > D^\dagger$.
If we perform the calculations we find:

$$\begin{aligned}\frac{1}{f^2N^2}D^2 = \overline{\frac{1}{f^2N^2}(\sum_i <\sigma_i> x_i)^2}|_{h=0} = \frac{1}{f^2N^2}\overline{\frac{1}{Z}\triangle_h Z} = \\ = \frac{1}{f^2}\int dr r^2 \frac{e^{\mu+\int dx y(x)\phi(x-r)}}{(1+e^{\mu+\int dx y(x)\phi(x-r)})}\end{aligned} \tag{A-25}$$

and

$$\begin{aligned}\frac{1}{f^2N^2}D^{\dagger 2} = \overline{\frac{1}{f^2N^2}(\sum_i <\sigma_i> x_i)^2}|_{h=0} = \frac{1}{f^2N^2}\overline{(\frac{1}{Z}\frac{\partial Z}{\partial h})^2} = \\ = \frac{1}{f^2}\int dr r^2 \frac{e^{2(\mu+\int dx y(x)\phi(x-r))}}{(1+e^{\mu+\int dx y(x)\phi(x-r)})^2}\end{aligned} \tag{A-26}$$

Results are depicted in Figure 4 (with $f = 0.1$ and $w = 0.05$). Consistently with their definition, $D > D^\dagger$. Both measures show a decrease with the increase of ρ, but neither of the two go to zero.

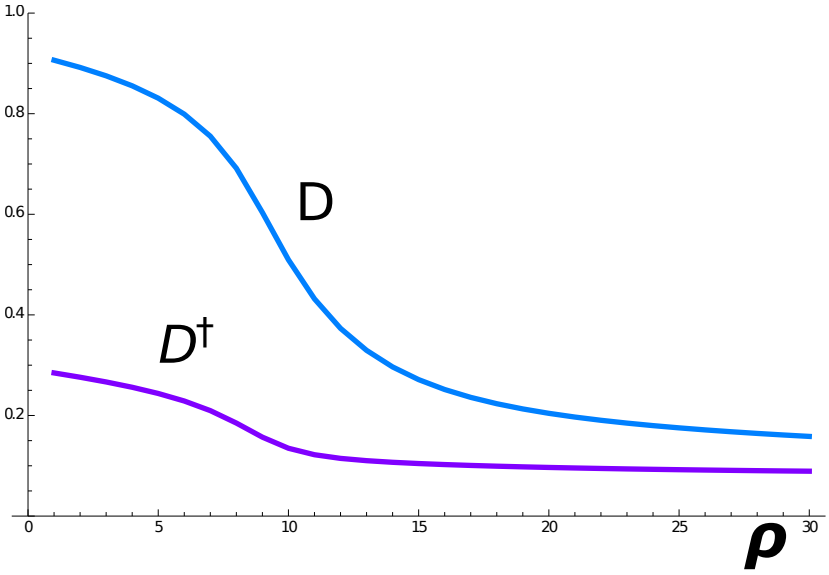


Figure 4: Analytical evaluation of drift length.

Appendix B

Here we describe the model used for the simulations in Chapters 6 and 9.

B-2 Basic Model for CA3 Simulations

The model we consider is, as in Cerasti and Treves 2013, an extended version of the one used in a previous study (Cerasti and Treves, 2010), where the firing rate of a CA3 pyramidal cell, η_i, was determined, as the one informative component, by the firing rates $\{\beta\}$ of DG granule cells, which feed into it through mossy fiber (MF) connections. The model used for the neuron was a simple threshold-linear unit (Treves, 1990), so that the firing of the unit results from an activating current (which includes several non-informative components) and is compared to a threshold:

$$\eta_i(\vec{x}) = g\left[\sum_j c_{ij}^{MF} J_{ij}^{MF} \beta_i(\vec{x}) + \tilde{\delta}_i - \tilde{T}_i\right]^+ \tag{B-1}$$

where g is a gain factor, while $[]^+$ equals the sum inside the brackets if positive in value, and zero if negative. The effect of the current threshold for activating a cell, along with the effect of inhibition, and other non-informative components, are summarized into a single subtractive term, with a mean value across CA3 cells expressed as $\tilde{T}$, and a deviation from the mean for each particular cell $i$as $\tilde{\delta}_i$, which acts as a sort of noise; threshold and inhibition, in fact, while influencing the mean activity of the network, are supposed to have a minor influence on the coding properties of the system. In the earlier reduced model, however, $\tilde{T}$ and $\tilde{\delta}$ also included the effect of other cells in CA3, through RC connections, and that of the perforant path, both regarded as unspecific inputs - this based on the assumption that information is driven into a new CA3 representation solely by MF inputs.

Since we are interested in the ability of the RC system (Amaral et al., 1990;van Strien et al., 2009) to express spatial representations, we separate out the RC contribution, and redefine $\tilde{T}$ and $\tilde{\delta}$ into T and δ - which sum the remaining unspecific inputs, including the perforant path, not analyzed here:

$$\eta_i(\vec{x}) = g\left[\sum_j c_{ij}^{MF} J_{ij}^{MF} \beta_i(\vec{x}) + \sum_k c_{ik}^{RC} J_{ij}^{RC} \eta_k(\vec{x}) + \delta_i - T_i\right]^+ \tag{B-2}$$

Connections between cells are indicated by the fixed binary matrices $\left\{c^{MF}\right\}$, $\left\{c^{RC}\right\}$,

whose non-zero elements (which take value 1) represent the existence of anatomical synapses between two cells. The synaptic efficacies are instead indicated by the matrices of weights $\{J^{MF}\}$, $\{J^{RC}\}$, whose elements are allowed to take positive values. The notation is chosen to minimize differences with our previous analysis of other components of the hippocampal system (e.g. (Treves, 1990; Kropff and Treves, 2008)).

The perforant path inputs from Entorhinal Cortex are not explicitly included in the model, in line with the hypothesis that the they relay the cue that initiates the retrieval of a previously stored representation, and have no role in the storage of a new representation and in defining the properties of the attractors in CA3. This perspective has been theoretically described in (Treves and Rolls) and has found experimental support (Lassalle et al.;Lee and Kesner). The firing rates of the various populations are all assumed to depend on the spatial position $\vec{x}$ of the animal; and the time scale considered for evaluating the firing rate is of order the theta period, about a hundred msec, so the finer temporal dynamics over shorter time scales is neglected. To be precise, in the simulations, we take a time step to correspond to 125ms of real time, or a theta period, during which the simulated rat moves 2.5 cm, thus at a speed of 20 cm/s. This is taken to be an average over a virtual exploratory session, familiarizing with a new environment.

B-2.1 The Storage of New Representations

The important novel ingredient that was introduced by (Cerasti and Treves, 2010), and that makes the difference from previous models of self-organizing recurrent networks, is a realistic description of the patterns of firing in the inputs, i.e. in the dentate gyrus. As the virtual rat explores the new environment, the activity β_j of DG unit j is determined by the position $\vec{x}$ of the animal, according to the expression:

$$\beta_j(\vec{x}) = \sum_{k=0}^{Q_j} \beta_0 e^{-\frac{(\vec{x}-\vec{x_{jk}})^2}{2\sigma_f^2}} \tag{B-3}$$

The firing rate of the granule cells is then a combination of Q_j Gaussian functions, resulting in bumps, or fields in the firing map of the environment, centered at random points . The environment is taken to have size A, and the fields are defined as all having the same effective size $\pi(\sigma_f)^2$ and height β_0. Q_j, which indicates the multiplicity of fields of DG cell j, is drawn from a Poisson distribution with mean value q, which roughly fits the data reported by (Leutgeb et al.). According to the same experimental data, we assume that only a randomly selected fraction $p_{DG} \ll 1$ (here set at $p_{DG} = 0.033$) of the granule cells are active in a given environment. Hence population activity is sparse, but the firing map of individual active granule units need not be sparse (it would only be sparse if $q\pi(\sigma_f)^2/A \ll 1$, which we do not assume to be always the case).

The activity of DG units determines the probability distribution for the firing rate of any given CA3 pyramidal unit, once the connectivity level between the two layer has been fixed: $\{C^{MF}\} = 0, 1$, with $P(C_{ij}^{MF} = 1) = \frac{C^{MF}}{N_{DG}} \equiv c^{MF}$. In agreement with experimental data, we set $C^{MF} = 50$, a value in the range of the ones providing an optimal information transmission from DG to CA3 (Cerasti and Treves, 2010). The MF

synaptic weights are set to be uniform in value, $J_{ij}^{MF} = J$, and similarly $J_{ik}^{RC} = J_0^{RC}$ initially. Subsequently, during the learning phase, RC weights are modified according to the simulated learning process and under the influence of the input coming from the MF connections. Following the simplified hypothesis that the mossy fibers carry all the information to be stored without contributing anything to the retrieval process, which is left to the recurrent collateral, MF weights are kept fixed to their initial values J; note that we have found, in our earlier study, that MF connections appear to be inadequate, even when associatively plastic, to support retrieval of spatial representation (Cerasti and Treves, 2010).

The connectivity among CA3 cells is given by the matrix $\left\{C^{RC}\right\} = 0, 1$ with $P(C_{ij}^{RC} = 1) = \frac{C^{RC}}{N_{CA3}} \equiv c^{RC}$, where $C_{RC} = 900$ in most simulations. The activity of the network is regulated by the constraint we impose on its mean and on its sparsity a_{CA3}, i.e. the fraction of the CA3 units firing significantly at each position, which is an important parameter affecting memory retrieval ((Treves, 1990); more precisely, $a_{CA3} = \langle \eta_i \rangle^2 / \langle \eta_i^2 \rangle$). Here we set the sparsity of each representations as $a_{CA3} = 0.1$, in broad agreement with experimental data (Papp et al., 2007), and at each time step we regulate the threshold T accordingly, to fulfill such requirement, while keeping the mean activity $\langle \eta_i \rangle = 0.1$ by adjusting the gain g.

B-2.2 Recurrent Collateral Plasticity

During the learning phase, the activity of CA3 is driven by DG inputs, and RC connections contribute through weights uniformly set to their initial value J_0. While the virtual rat explores the environment, RC weights are allowed to change according to an associative Hebbian learning rule, such that the total change in the synaptic weights is given as a sum of independent terms

$$\Delta J_{ik}^{RC} = \gamma \eta_i(t) \left(\eta_j(t) - \Gamma_j(t)\right) \tag{B-4}$$

where $\Delta J_{ij}(t)$ indicates the variation of the connection weight between cells i and j occurring at a given time step t, η_i and η_j are the postsynaptic and presynaptic firing rate, while γ is the learning rate. This associative learning rule includes the contribution of a trace, Γ, of the recent past activity of the presynaptic cell, defined as

$$\Gamma_j(t) = \frac{1}{\tau} \sum \sum_{t_s=1}^{\tau} \eta_j(t - t_s) \tag{B-5}$$

where τ is taken equal to 14 time steps (1750ms). RC weights are forced to be non-negative, so they are reset to zero each time they become negative. Moreover, the total of the synaptic weights afferent to a single postsynaptic CA3 cell is normalized at the end of the learning process, so that $\sum_k J_{ik}^{RC} = 1$ per each CA3 cell. In words, the synaptic plasticity on recurrent connections allows the system to store the information about the current environment conveyed by mossy fiber inputs; such information is expressed in the form of place-like patterns of activity in CA3 units, and the Hebb-like learning rule strengthens the connections between units that show overlapping fields.

References

Abramowitz, M. and I. Stegun, eds. (1965). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Dover Publications (cit. on p. 31).

Acsady, L., A. Kamondi, A. Sik, T. Freund, and G. Buzsaki (1998). "GABAergic cells are the major postsynaptic targets of mossy fibers in the rat hippocampus." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 18.9, pp. 3386–3403 (cit. on p. 106).

Ahmed, O. and M. Mehta (2009). "The hippocampal rate code: anatomy, physiology and theory." In: *Trends in neurosciences* 32.6, pp. 329–338. DOI: 10.1016/j.tins.2009.01.009. URL: http://dx.doi.org/10.1016/j.tins.2009.01.009 (cit. on p. 8).

Akrami, A., Y. Liu, A. Treves, and B. Jagadeesh (2009a). "Converging neuronal activity in inferior temporal cortex during the classification of morphed stimuli". In: *Cerebral Cortex* 19 (4), p. 760 (cit. on p. 69).

Akrami, A., Y. Liu, A. Treves, and B. Jagadeesh (2009b). "Converging neuronal activity in inferior temporal cortex during the classification of morphed stimuli." In: *Cerebral cortex (New York, N.Y. : 1991)* 19.4, pp. 760–776. DOI: 10.1093/cercor/bhn125. URL: http://dx.doi.org/10.1093/cercor/bhn125 (cit. on pp. 106, 107).

Allen, K., J. Rawlins, D. Bannerman, and J. Csicsvari (2012). "Hippocampal place cells can encode multiple trial-dependent features through rate remapping." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 32.42, pp. 14752–14766. DOI: 10.1523/JNEUROSCI.6175-11.2012. URL: http://dx.doi.org/10.1523/JNEUROSCI.6175-11.2012 (cit. on p. 9).

Almeida, L. de, M. Idiart, and J. Lisman (2010). "The single place fields of CA3 cells: A two-stage transformation from grid cells". In: *Hippocampus* (cit. on p. 76).

Amaral, D. and W. Cowan (1980). "Subcortical afferents to the hippocampal formation in the monkey." In: *The Journal of comparative neurology* 189.4, pp. 573–591 (cit. on p. 66).

Amaral, D., N. Ishizuka, and B. Claiborne (1990). "Neurons, numbers and the hippocampal network." In: *Progress in brain research* 83, pp. 1–11 (cit. on p. 83).

Amit, D. J. (1992). "Modeling brain function: The world of attractor neural networks". In: (cit. on p. 91).

Amit, D. J. (1995). "The Hebbian paradigm reintegrated: Local reverberations as internal representations". In: *Behavioral and Brain Sciences* 18 (cit. on p. 69).

Andersen, P., R. Morris, D. Amaral, T. Bliss, and J. O'Keefe (2006). "The hippocampus book". In: (cit. on p. 5).

Ang, C., G. Carlson, and D. Coulter (2005). "Hippocampal CA1 circuitry dynamically gates direct cortical inputs preferentially at theta frequencies." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 25.42, pp. 9567–9580. DOI: 10.1523/JNEUROSCI.2992-05.2005. URL: http://dx.doi.org/10.1523/JNEUROSCI.2992-05.2005 (cit. on p. 106).

Averbeck, B., P. Latham, and A. Pouget (2006). "Neural correlations, population coding and computation." In: *Nature reviews. Neuroscience* 7.5, pp. 358–366. DOI: 10.1038/nrn1888. URL: http://dx.doi.org/10.1038/nrn1888 (cit. on p. 111).

Azab, M., S. Stark, and C. Stark (2013). "Contributions of human hippocampal subfields to spatial and temporal pattern separation." In: *Hippocampus*. DOI: 10.1002/hipo.22223. URL: http://dx.doi.org/10.1002/hipo.22223 (cit. on p. 76).

Bahar, A., P. Shirvalkar, and M. Shapiro (2011). "Memory-guided learning: CA1 and CA3 neuronal ensembles differentially encode the commonalities and differences between situations." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31.34, pp. 12270–12281. DOI: 10.1523/JNEUROSCI.1671-11.2011. URL: http://dx.doi.org/10.1523/JNEUROSCI.1671-11.2011 (cit. on p. 9).

Bakker, A., C. Kirwan, M. Miller, and C. Stark (2008). "Pattern separation in the human hippocampal CA3 and dentate gyrus". In: *Science* 319 (5870), p. 1640 (cit. on p. 76).

Banks, M., J. White, and R. Pearce (2000). "Interactions between distinct GABA(A) circuits in hippocampus." In: *Neuron* 25.2, pp. 449–457 (cit. on p. 106).

Bannerman, D., M. Grubb, R. Deacon, B. Yee, J. Feldon, and J. Rawlins (2003). "Ventral hippocampal lesions affect anxiety but not spatial learning." In: *Behavioural brain research* 139.1-2, pp. 197–213. DOI: 10.1016/S0166-4328(02)00268-1. URL: http://dx.doi.org/10.1016/S0166-4328(02)00268-1 (cit. on p. 9).

Barry, C., L. Ginzberg, J. O'Keefe, and N. Burgess (2012). "Grid cell firing patterns signal environmental novelty by expansion." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.43, pp. 17687–17692. DOI: 10.1073/pnas.1209918109. URL: http://dx.doi.org/10.1073/pnas.1209918109 (cit. on p. 61).

Barry, C., R. Hayman, N. Burgess, and K. Jeffery (2007). "Experience-dependent rescaling of entorhinal grids." In: *Nature neuroscience* 10.6, pp. 682–684. DOI: 10.1038/nn1905. URL: http://dx.doi.org/10.1038/nn1905 (cit. on p. 27).

Battaglia, F. and A. Treves (1998). "Attractor neural networks storing multiple space representations: a model for hippocampal place fields". In: *Physical Review E* 58.6 (cit. on pp. 89, 90, 122).

Ben-Yishai, R., R. Bar-Or, and H. Sompolinsky (1995). "Theory of orientation tuning in visual cortex". In: *Proceedings of the National Academy of Sciences* 92 (9), p. 3844 (cit. on p. 70).

Blair, H., A. Welday, and K. Zhang (2007). "Scale-invariant memory representations emerge from moire interference between grid fields that produce theta oscillations: a computational model." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27.12, pp. 3211–3229. DOI: 10.1523/JNEUROSCI.4724-06.2007. URL: http://dx.doi.org/10.1523/JNEUROSCI.4724-06.2007 (cit. on p. 21).

Boccara, C., F. Sargolini, V. Thoresen, T. Solstad, M. Witter, E. Moser, and M. Moser (2010). "Grid cells in pre- and parasubiculum." In: *Nature neuroscience* 13.8, pp. 987–994. DOI: 10.1038/nn.2602. URL: http://dx.doi.org/10.1038/nn.2602 (cit. on pp. 15, 58).

Bonnevie, T., B. Dunn, M. Fyhn, T. Hafting, D. Derdikman, J. Kubie, Y. Roudi, E. Moser, and M. Moser (2013). "Grid cells require excitatory drive from the hippocampus." In: *Nature neuroscience* 16.3, pp. 309–317. DOI: 10.1038/nn.3311. URL: http://dx.doi.org/10.1038/nn.3311 (cit. on pp. 20, 23, 58, 61, 62).

Brandon, M., A. Bogaard, C. Libby, M. Connerney, K. Gupta, and M. Hasselmo (2011). "Reduction of theta rhythm dissociates grid cell spatial periodicity from directional tuning." In: *Science (New York, N.Y.)* 332.6029, pp. 595–599. DOI: 10.1126/science.1201652. URL: http://dx.doi.org/10.1126/science.1201652 (cit. on p. 57).

Brandon, M., A. Bogaard, N. Schultheiss, and M. Hasselmo (2013). "Segregation of cortical head direction cell assemblies on alternating theta cycles." In: *Nature neuroscience* 16.6, pp. 739–748. DOI: 10.1038/nn.3383. URL: http://dx.doi.org/10.1038/nn.3383 (cit. on p. 109).

Brown, E., L. Frank, D. Tang, M. Quirk, and M. Wilson (1998). "A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 18.18, pp. 7411–7425 (cit. on p. 111).

Brun, V., M. Otnass, S. Molden, H. Steffenach, M. Witter, M. Moser, and E. Moser (2002). "Place cells and place recognition maintained by direct entorhinal-hippocampal circuitry." In: *Science (New York, N.Y.)* 296.5576, pp. 2243–2246. DOI: 10.1126/science.1071089. URL: http://dx.doi.org/10.1126/science.1071089 (cit. on p. 11).

Brun, V., T. Solstad, K. Kjelstrup, M. Fyhn, M. Witter, E. Moser, and M. Moser (2008). "Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex." In: *Hippocampus* 18.12, pp. 1200–1212. DOI: 10.1002/hipo.20504. URL: http://dx.doi.org/10.1002/hipo.20504 (cit. on p. 14).

Brunel, N. and O. Trullier (1998). "Plasticity of directional place fields in a model of rodent CA3". In: *Hippocampus* 8 (6), pp. 651–65 (cit. on p. 71).

Buehlmann, A. and G. Deco (2010). "Optimal information transfer in the cortex through synchronization." In: *PLoS computational biology* 6.9. DOI: 10.1371/journal.pcbi.1000934. URL: http://dx.doi.org/10.1371/journal.pcbi.1000934 (cit. on p. 109).

Burak, Y. and I. Fiete (2009). "Accurate path integration in continuous attractor network models of grid cells." In: *PLoS computational biology* 5.2. DOI: 10.1371/journal.pcbi.1000291. URL: http://dx.doi.org/10.1371/journal.pcbi.1000291 (cit. on pp. 20, 38).

Burgess, N. (2008). "Grid cells and theta as oscillatory interference: theory and predictions." In: *Hippocampus* 18.12, pp. 1157–1174. DOI: 10.1002/hipo.20518. URL: http://dx.doi.org/10.1002/hipo.20518 (cit. on p. 21).

Burgess, N., C. Barry, and J. O'Keefe (2007). "An oscillatory interference model of grid cell firing." In: *Hippocampus* 17.9, pp. 801–812. DOI: 10.1002/hipo.20327. URL: http://dx.doi.org/10.1002/hipo.20327 (cit. on pp. 21, 38).

Burghardt, N., E. Park, R. Hen, and A. Fenton (2012). "Adult-born hippocampal neurons promote cognitive flexibility in mice." In: *Hippocampus* 22.9, pp. 1795–1808. DOI: 10.1002/hipo.22013. URL: http://dx.doi.org/10.1002/hipo.22013 (cit. on p. 76).

Buzsaki, G. (1986). "Hippocampal sharp waves: their origin and significance." In: *Brain research* 398.2, pp. 242–252 (cit. on p. 108).

— (1989). "Two-stage model of memory trace formation: a role for noisy brain states." In: *Neuroscience* 31.3, pp. 551–570. DOI: 10.1016/0306-4522(89)90423-5. URL: http://dx.doi.org/10.1016/0306-4522(89)90423-5 (cit. on p. 10).

Buzsaki, G. and J. Chrobak (1995). "Temporal structure in spatially organized neuronal ensembles: a role for interneuronal networks." In: *Current opinion in neurobiology* 5.4, pp. 504–510. DOI: 10.1016/0959-4388(95)80012-3. URL: http://dx.doi.org/10.1016/0959-4388(95)80012-3 (cit. on p. 106).

Buzsaki, G., L. Leung, and C. Vanderwolf (1983). "Cellular bases of hippocampal EEG in the behaving rat." In: *Brain research* 287.2, pp. 139–171 (cit. on p. 10).

Buzsaki, G. (2002). "Theta oscillations in the hippocampus." In: *Neuron* 33.3, pp. 325–340. DOI: 10.1016/S0896-6273(02)00586-X. URL: http://dx.doi.org/10.1016/S0896-6273(02)00586-X (cit. on pp. 95, 108).

— (2006). *Rhythms of the Brain*. Oxford University Press (cit. on p. 95).

Buzsaki, G. and E. Moser (2013). "Memory, navigation and theta rhythm in the hippocampal-entorhinal system." In: *Nature neuroscience* 16.2, pp. 130–138. DOI: 10.1038/nn.3304. URL: http://dx.doi.org/10.1038/nn.3304 (cit. on p. 73).

Buzsaki, G. and X. Wang (2012). "Mechanisms of gamma oscillations." In: *Annual review of neuroscience* 35, pp. 203–225. DOI: 10.1146/annurev-neuro-062111-150444. URL: http://dx.doi.org/10.1146/annurev-neuro-062111-150444 (cit. on p. 109).

Calton, J. and J. Taube (2005). "Degradation of head direction cell activity during inverted locomotion." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 25.9, pp. 2420–2428. DOI: 10.1523/JNEUROSCI.3511-04.2005. URL: http://dx.doi.org/10.1523/JNEUROSCI.3511-04.2005 (cit. on p. 41).

Canto, C., N. Koganezawa, P. Beed, E. Moser, and M. Witter (2012). "All layers of medial entorhinal cortex receive presubicular and parasubicular inputs." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 32.49, pp. 17620–17631. DOI: 10.1523/JNEUROSCI.3526-12.2012. URL: http://dx.doi.org/10.1523/JNEUROSCI.3526-12.2012 (cit. on p. 57).

Canto, C., F. Wouterlood, and M. Witter (2008). "What does the anatomical organization of the entorhinal cortex tell us?" In: *Neural plasticity* 2008. DOI: 10.1155/2008/381243. URL: http://dx.doi.org/10.1155/2008/381243 (cit. on pp. 13, 14).

Cerasti, E. and A. Treves (2010). "How informative are spatial CA3 representations established by the dentate gyrus?" In: *PLoS computational biology* 6.4. DOI: 10.1371/journal.pcbi.1000759. URL: http://dx.doi.org/10.1371/journal.pcbi.1000759 (cit. on pp. 24, 76, 117, 123).

— (2013). "The spatial representations acquired in CA3 by self-organizing recurrent connections." In: *Frontiers in cellular neuroscience* 7. DOI: 10.3389/fncel.2013.00112. URL: http://dx.doi.org/10.3389/fncel.2013.00112 (cit. on pp. 71, 72, 76, 79, 81, 82, 120, 123).

Cerf, M., N. Thiruvengadam, F. Mormann, A. Kraskov, R. Quiroga, C. Koch, and I. Fried (2010). "On-line, voluntary control of human temporal lobe neurons." In: *Nature* 467.7319, pp. 1104–1108. DOI: 10.1038/nature09510. URL: http://dx.doi.org/10.1038/nature09510 (cit. on p. 68).

Cheng, S. and L. Frank (2008). "New experiences enhance coordinated neural activity in the hippocampus." In: *Neuron* 57.2, pp. 303–313. DOI: 10.1016/j.neuron.2007.11.035. URL: http://dx.doi.org/10.1016/j.neuron.2007.11.035 (cit. on p. 108).

Ciaramelli, E., L. Rosapia, and A. Treves (2006). "Dissociating episodic from semantic access mode by mutual information measures: evidence from aging and Alzheimer's disease." In: *Journal of physiology, Paris* 100.1-3, pp. 142–153. DOI: 10.1016/j.jphysparis.2006.09.008. URL: http://dx.doi.org/10.1016/j.jphysparis.2006.09.008 (cit. on p. 112).

Cipolotti, L., T. Shallice, D. Chan, N. Fox, R. Scahill, G. Harrison, J. Stevens, and P. Rudge (2001). "Long-term retrograde amnesia... the crucial role of the hippocampus". In: *Neuropsychologia* 39 (2), pp. 151–172 (cit. on p. 68).

Cohn, H. and A. Kumar (2007). "Universally optimal distribution of points on spheres". In: *Journal of the American Mathematical Society* 20. DOI: 10.1090/S0894-0347-06-00546-7. URL: http://dx.doi.org/10.1090/S0894-0347-06-00546-7 (cit. on pp. 43, 65).

Colgin, L., T. Denninger, M. Fyhn, T. Hafting, T. Bonnevie, O. Jensen, M. Moser, and E. Moser (2009). "Frequency of gamma oscillations routes flow of information in the hippocampus." In: *Nature* 462.7271, pp. 353–357. DOI: 10.1038/nature08573. URL: http://dx.doi.org/10.1038/nature08573 (cit. on p. 110).

Colgin, L., S. Leutgeb, K. Jezek, J. Leutgeb, E. Moser, M. Bruce, and M. Moser (2010). "Attractor-map versus autoassociation based attractor dynamics in the hippocampal network." In: *Journal of neurophysiology* 104.1, pp. 35–50. DOI: 10.1152/jn.00202.2010. URL: http://dx.doi.org/10.1152/jn.00202.2010 (cit. on pp. 7, 113).

Colgin, L. and E. Moser (2010). "Gamma oscillations in the hippocampus." In: *Physiology (Bethesda, Md.)* 25.5, pp. 319–329. DOI: 10.1152/physiol.00021.2010. URL: http://dx.doi.org/10.1152/physiol.00021.2010 (cit. on p. 109).

Conklin, J. and C. Eliasmith (2005). "A controlled attractor network model of path integration in the rat." In: *Journal of computational neuroscience* 18.2, pp. 183–203. DOI: 10.1007/s10827-005-6558-z. URL: http://dx.doi.org/10.1007/s10827-005-6558-z (cit. on p. 20).

Conway, J. (1992). *Sphere Packings, Lattices and Groups.* Ed. by M. Berger and B. Eckmann. Springer (cit. on pp. 43, 65).

Couey, J., A. Witoelar, S. Zhang, K. Zheng, J. Ye, B. Dunn, R. Czajkowski, M. Moser, E. Moser, Y. Roudi, and M. Witter (2013). "Recurrent inhibitory circuitry as a

mechanism for grid formation." In: *Nature neuroscience* 16.3, pp. 318–324. DOI: 10.1038/nn.3310. URL: http://dx.doi.org/10.1038/nn.3310 (cit. on pp. 20, 58, 59, 61, 64, 126).

Courant, R. and D. Hilbert (1962). *Methods of mathematical Physics*. Wiley-Interscience (cit. on p. 29).

Csicsvari, J., H. Hirase, A. Czurko, A. Mamiya, and G. Buzsaki (1999). "Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving Rat." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 19.1, pp. 274–287 (cit. on p. 106).

Csicsvari, J., H. Hirase, A. Mamiya, and G. Buzsaki (2000). "Ensemble patterns of hippocampal CA3-CA1 neurons during sharp wave-associated population events." In: *Neuron* 28.2, pp. 585–594. DOI: 10.1016/S0896-6273(00)00135-5. URL: http://dx.doi.org/10.1016/S0896-6273(00)00135-5 (cit. on p. 108).

Csicsvari, J., B. Jamieson, K. Wise, and G. Buzsaki (2003). "Mechanisms of gamma oscillations in the hippocampus of the behaving rat." In: *Neuron* 37.2, pp. 311–322. DOI: 10.1016/S0896-6273(02)01169-8. URL: http://dx.doi.org/10.1016/S0896-6273(02)01169-8 (cit. on p. 109).

Csicsvari, J., J. O'Neill, K. Allen, and T. Senior (2007). "Place-selective firing contributes to the reverse-order reactivation of CA1 pyramidal cells during sharp waves in open-field exploration." In: *The European journal of neuroscience* 26.3, pp. 704–716. DOI: 10.1111/j.1460-9568.2007.05684.x. URL: http://dx.doi.org/10.1111/j.1460-9568.2007.05684.x (cit. on p. 10).

Curto, C. and V. Itskov (2008). "Cell groups reveal structure of stimulus space." In: *PLoS computational biology* 4.10. DOI: 10.1371/journal.pcbi.1000205. URL: http://dx.doi.org/10.1371/journal.pcbi.1000205 (cit. on pp. 111, 122).

Czurkó, A., J. Huxter, Y. Li, B. Hangya, and R. Muller (2011). "Theta phase classification of interneurons in the hippocampal formation of freely moving rats." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31.8, pp. 2938–2947. DOI: 10.1523/JNEUROSCI.5037-10.2011. URL: http://dx.doi.org/10.1523/JNEUROSCI.5037-10.2011 (cit. on p. 97).

Dabaghian, Y., F. Mémoli, L. Frank, and G. Carlsson (2012). "A topological paradigm for hippocampal spatial map formation using persistent homology." In: *PLoS computational biology* 8.8. DOI: 10.1371/journal.pcbi.1002581. URL: http://dx.doi.org/10.1371/journal.pcbi.1002581 (cit. on p. 122).

Daelli, V. and A. Treves (2010). "Neural attractor dynamics in object recognition". In: *Experimental Brain Research* 203 (2), pp. 241–248 (cit. on p. 69).

Davidson, T., F. Kloosterman, and M. Wilson (2009). "Hippocampal replay of extended experience." In: *Neuron* 63.4, pp. 497–507. DOI: 10.1016/j.neuron.2009.07.027. URL: http://dx.doi.org/10.1016/j.neuron.2009.07.027 (cit. on p. 10).

Deneve, S., P. Latham, and A. Pouget (1999). "Reading population codes: a neural implementation of ideal observers." In: *Nature neuroscience* 2.8, pp. 740–745. DOI: 10.1038/11205. URL: http://dx.doi.org/10.1038/11205 (cit. on p. 111).

Deng, W., M. Mayford, and F. Gage (2013). "Selection of distinct populations of dentate granule cells in response to inputs as a mechanism for pattern separation in mice." In: *eLife* 2. DOI: 10.7554/eLife.00312. URL: http://dx.doi.org/10.7554/eLife.00312 (cit. on p. 76).

Derdikman, D. and E. Moser (2010). "A manifold of spatial maps in the brain." In: *Trends in cognitive sciences* 14.12, pp. 561–569. DOI: 10.1016/j.tics.2010.09.004. URL: http://dx.doi.org/10.1016/j.tics.2010.09.004 (cit. on p. 27).

Derdikman, D., J. Whitlock, A. Tsao, M. Fyhn, T. Hafting, M. Moser, and E. Moser (2009). "Fragmentation of grid cell maps in a multicompartment environment." In: *Nature neuroscience* 12.10, pp. 1325–1332. DOI: 10.1038/nn.2396. URL: http://dx.doi.org/10.1038/nn.2396 (cit. on pp. 8, 9).

Deshmukh, S., J. Johnson, and J. Knierim (2012). "Perirhinal cortex represents nonspatial, but not spatial, information in rats foraging in the presence of objects: comparison with lateral entorhinal cortex." In: *Hippocampus* 22.10, pp. 2045–2058.

DOI: 10.1002/hipo.22046. URL: http://dx.doi.org/10.1002/hipo.22046 (cit. on p. 75).

Deshmukh, S. and J. Knierim (2011). "Representation of non-spatial and spatial information in the lateral entorhinal cortex." In: *Frontiers in behavioral neuroscience* 5. DOI: 10.3389/fnbeh.2011.00069. URL: http://dx.doi.org/10.3389/fnbeh.2011.00069 (cit. on p. 75).

Deshmukh, S., D. Yoganarasimha, H. Voicu, and J. Knierim (2010). "Theta modulation in the medial and the lateral entorhinal cortices." In: *Journal of neurophysiology* 104.2, pp. 994–991006. DOI: 10.1152/jn.01141.2009. URL: http://dx.doi.org/10.1152/jn.01141.2009 (cit. on p. 96).

Diba, K. and G. Buzsaki (2007). "Forward and reverse hippocampal place-cell sequences during ripples." In: *Nature neuroscience* 10.10, pp. 1241–1242. DOI: 10.1038/nn1961. URL: http://dx.doi.org/10.1038/nn1961 (cit. on p. 10).

Domnisoru, C., A. Kinkhabwala, and D. Tank (2013). "Membrane potential dynamics of grid cells." In: *Nature* 495.7440, pp. 199–204. DOI: 10.1038/nature11973. URL: http://dx.doi.org/10.1038/nature11973 (cit. on p. 23).

Dragoi, G. and G. Buzsaki (2006). "Temporal encoding of place sequences by hippocampal cell assemblies." In: *Neuron* 50.1, pp. 145–157. DOI: 10.1016/j.neuron.2006.02.023. URL: http://dx.doi.org/10.1016/j.neuron.2006.02.023 (cit. on p. 108).

Dragoi, G. and S. Tonegawa (2011). "Preplay of future place cell sequences by hippocampal cellular assemblies." In: *Nature* 469.7330, pp. 397–401. DOI: 10.1038/nature09633. URL: http://dx.doi.org/10.1038/nature09633 (cit. on pp. 11, 122).

— (2013). "Distinct preplay of multiple novel spatial experiences in the rat." In: *Proceedings of the National Academy of Sciences of the United States of America* 110.22, pp. 9100–9105. DOI: 10.1073/pnas.1306031110. URL: http://dx.doi.org/10.1073/pnas.1306031110 (cit. on pp. 11, 91).

Dupret, D., O. Joseph, P. Barty, and J. Csicsvari (2010). "The reorganization and reactivation of hippocampal maps predict spatial memory performance." In: *Nature neuroscience* 13.8, pp. 995–991002. DOI: 10.1038/nn.2599. URL: http://dx.doi.org/10.1038/nn.2599 (cit. on pp. 10, 79).

Dupret, D., O. Joseph, and J. Csicsvari (2013). "Dynamic Reconfiguration of Hippocampal Interneuron Circuits during Spatial Learning." In: *Neuron*. DOI: 10.1016/j.neuron.2013.01.033. URL: http://dx.doi.org/10.1016/j.neuron.2013.01.033 (cit. on p. 106).

Ego-Stengel, V. and M. Wilson (2010). "Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat." In: *Hippocampus* 20.1, pp. 1–10. DOI: 10.1002/hipo.20707. URL: http://dx.doi.org/10.1002/hipo.20707 (cit. on p. 10).

Eichenbaum, H., P. Dudchenko, E. Wood, M. Shapiro, and H. Tanila (1999). "The Hippocampus, Memory, Review and Place Cells: Is It Spatial Memory or a Memory Space?" In: *Neuron* 23, pp. 209–226 (cit. on p. 72).

Eichenbaum, H., A. R. Yonelinas, and C. Ranganath (2007). "The medial temporal lobe and recognition memory". In: *Annual Review of Neuroscience* 30, p. 123 (cit. on p. 67).

Eichenbaum, H. (2013). "Memory on time." In: *Trends in cognitive sciences*. DOI: 10.1016/j.tics.2012.12.007. URL: http://dx.doi.org/10.1016/j.tics.2012.12.007 (cit. on p. 73).

Eichenbaum, H. and P. Lipton (2008). "Towards a functional organization of the medial temporal lobe memory system: role of the parahippocampal and medial entorhinal cortical areas." In: *Hippocampus* 18.12, pp. 1314–1324 (cit. on p. 66).

Eichenbaum, H., M. Sauvage, N. Fortin, R. Komorowski, and P. Lipton (2011). "Towards a functional organization of episodic memory in the medial temporal lobe." In: *Neuroscience and biobehavioral reviews*. DOI: 10.1016/j.neubiorev.2011.07.

006. URL: http://dx.doi.org/10.1016/j.neubiorev.2011.07.006 (cit. on p. 123).

Eliasmith, C. (2005). "A unified approach to building and controlling spiking attractor networks." In: *Neural computation* 17.6, pp. 1276–1314. DOI: 10.1162/0899766053630332. URL: http://dx.doi.org/10.1162/0899766053630332 (cit. on p. 20).

Fanselow, M. and H. Dong (2010). "Are the dorsal and ventral hippocampus functionally distinct structures?" In: *Neuron* 65.1, pp. 7–19. DOI: 10.1016/j.neuron.2009.11.031. URL: http://dx.doi.org/10.1016/j.neuron.2009.11.031 (cit. on pp. 9, 97).

Felleman, D. and D. Van Essen (1991). "Distributed hierarchical processing in the primate cerebral cortex." In: *Cerebral cortex (New York, N.Y. : 1991)* 1.1, pp. 1–47. DOI: 10.1093/cercor/1.1.1. URL: http://dx.doi.org/10.1093/cercor/1.1.1 (cit. on p. 5).

Fenton, A., H. Kao, S. Neymotin, A. Olypher, Y. Vayntrub, W. Lytton, and N. Ludvig (2008). "Unmasking the CA1 ensemble place code by exposures to small and large environments: more place cells and multiple, irregularly arranged, and expanded place fields in the larger space." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28.44, pp. 11250–11262. DOI: 10.1523/JNEUROSCI.2862-08.2008. URL: http://dx.doi.org/10.1523/JNEUROSCI.2862-08.2008 (cit. on p. 19).

Ferbinteanu, J. and M. Shapiro (2003). "Prospective and retrospective memory coding in the hippocampus." In: *Neuron* 40.6, pp. 1227–1239. DOI: 10.1016/S0896-6273(03)00752-9. URL: http://dx.doi.org/10.1016/S0896-6273(03)00752-9 (cit. on p. 9).

Ferbinteanu, J., P. Shirvalkar, and M. Shapiro (2011). "Memory modulates journey-dependent coding in the rat hippocampus." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31.25, pp. 9135–9146. DOI: 10.1523/JNEUROSCI.1241-11.2011. URL: http://dx.doi.org/10.1523/JNEUROSCI.1241-11.2011 (cit. on p. 9).

Fiete, I., Y. Burak, and T. Brookings (2008). "What grid cells convey about rat location." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28.27, pp. 6858–6871. DOI: 10.1523/JNEUROSCI.5684-07.2008. URL: http://dx.doi.org/10.1523/JNEUROSCI.5684-07.2008 (cit. on pp. 65, 66, 74).

Foster, D. and M. Wilson (2006). "Reverse replay of behavioural sequences in hippocampal place cells during the awake state." In: *Nature* 440.7084, pp. 680–683. DOI: 10.1038/nature04587. URL: http://dx.doi.org/10.1038/nature04587 (cit. on p. 10).

Freund, T. and G. Buzsaki (1996). "Interneurons of the hippocampus." In: *Hippocampus* 6.4, pp. 347–470. DOI: 10.1002/(SICI)1098-1063(1996)6:4<347::AID-HIPO1>3.0.CO;2-I. URL: http://dx.doi.org/10.1002/(SICI)1098-1063(1996)6:4%3C347::AID-HIPO1%3E3.0.CO;2-I (cit. on p. 97).

Freund, T. (2003). "Interneuron Diversity series: Rhythm and mood in perisomatic inhibition." In: *Trends in neurosciences* 26.9, pp. 489–495. DOI: 10.1016/S0166-2236(03)00227-3. URL: http://dx.doi.org/10.1016/S0166-2236(03)00227-3 (cit. on p. 106).

Fuhs, M. and D. Touretzky (2006). "A spin glass model of path integration in rat medial entorhinal cortex." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26.16, pp. 4266–4276. DOI: 10.1523/JNEUROSCI.4353-05.2006. URL: http://dx.doi.org/10.1523/JNEUROSCI.4353-05.2006 (cit. on pp. 20, 38).

Fyhn, M., T. Hafting, A. Treves, M. Moser, and E. Moser (2007). "Hippocampal remapping and grid realignment in entorhinal cortex." In: *Nature* 446.7132, pp. 190–194. DOI: 10.1038/nature05601. URL: http://dx.doi.org/10.1038/nature05601 (cit. on pp. 14, 66, 74).

Fyhn, M., T. Hafting, M. Witter, E. Moser, and M. Moser (2008). "Grid cells in mice." In: *Hippocampus* 18.12, pp. 1230–1238. DOI: 10.1002/hipo.20472. URL: http://dx.doi.org/10.1002/hipo.20472 (cit. on p. 21).

Fyhn, M., S. Molden, M. Witter, E. Moser, and M. Moser (2004). "Spatial representation in the entorhinal cortex." In: *Science (New York, N.Y.)* 305.5688, pp. 1258–1264. DOI: 10.1126/science.1099901. URL: http://dx.doi.org/10.1126/science.1099901 (cit. on pp. 11, 19).

Garner, A., D. Rowland, S. Hwang, K. Baumgaertel, B. Roth, C. Kentros, and M. Mayford (2012). "Generation of a synthetic memory trace." In: *Science (New York, N.Y.)* 335.6075, pp. 1513–1516. DOI: 10.1126/science.1214985. URL: http://dx.doi.org/10.1126/science.1214985 (cit. on p. 68).

Geisler, C., K. Diba, E. Pastalkova, K. Mizuseki, S. Royer, and G. Buzsaki (2010). "Temporal delays among place cells determine the frequency of population theta oscillations in the hippocampus." In: *Proceedings of the National Academy of Sciences of the United States of America* 107.17, pp. 7957–7962. DOI: 10.1073/pnas.0912478107. URL: http://dx.doi.org/10.1073/pnas.0912478107 (cit. on p. 97).

Gibson, B., W. Butler, and J. Taube (2013). "The head-direction signal is critical for navigation requiring a cognitive map but not for learning a spatial habit." In: *Current biology : CB* 23.16, pp. 1536–1540. DOI: 10.1016/j.cub.2013.06.030. URL: http://dx.doi.org/10.1016/j.cub.2013.06.030 (cit. on p. 41).

Giocomo, L., S. Hussaini, F. Zheng, E. Kandel, M.-B. Moser, and E. Moser (2011). "Grid cells use HCN1 channels for spatial scaling." In: *Cell* 147.5, pp. 1159–1170. DOI: 10.1016/j.cell.2011.08.051. URL: http://dx.doi.org/10.1016/j.cell.2011.08.051 (cit. on p. 14).

Giocomo, L., M. Moser, and E. Moser (2011). "Computational models of grid cells." In: *Neuron* 71.4, pp. 589–603. DOI: 10.1016/j.neuron.2011.07.023. URL: http://dx.doi.org/10.1016/j.neuron.2011.07.023 (cit. on pp. 20, 22).

Giocomo, L., E. Zilli, E. Fransén, and M. Hasselmo (2007). "Temporal frequency of subthreshold oscillations scales with entorhinal grid cell field spacing." In: *Science (New York, N.Y.)* 315.5819, pp. 1719–1722. DOI: 10.1126/science.1139207. URL: http://dx.doi.org/10.1126/science.1139207 (cit. on p. 14).

Girardeau, G., K. Benchenane, S. Wiener, G. Buzsaki, and M. Zugaro (2009). "Selective suppression of hippocampal ripples impairs spatial memory." In: *Nature neuroscience* 12.10, pp. 1222–1223. DOI: 10.1038/nn.2384. URL: http://dx.doi.org/10.1038/nn.2384 (cit. on pp. 10, 110).

Golomb, D., J. Hertz, S. Panzeri, A. Treves, and B. Richmond (1997). "How well can we estimate the information carried in neuronal responses from limited samples?" In: *Neural computation* 9.3, pp. 649–665. DOI: 10.1162/neco.1997.9.3.649. URL: http://dx.doi.org/10.1162/neco.1997.9.3.649 (cit. on p. 112).

Grossberg, S. and P. Pilly (2012). "How entorhinal grid cells may learn multiple spatial scales from a dorsoventral gradient of cell response rates in a self-organizing map." In: *PLoS computational biology* 8.10. DOI: 10.1371/journal.pcbi.1002648. URL: http://dx.doi.org/10.1371/journal.pcbi.1002648 (cit. on p. 24).

Guanella, A., D. Kiper, and P. Verschure (2007). "A model of grid cells based on a twisted torus topology." In: *International journal of neural systems* 17.4, pp. 231–240. DOI: 10.1142/S0129065707001093. URL: http://dx.doi.org/10.1142/S0129065707001093 (cit. on pp. 20, 61).

Gupta, A., M. A. van der Meer, D. Touretzky, and A. Redish (2010). "Hippocampal replay is not a simple function of experience." In: *Neuron* 65.5, pp. 695–705. DOI: 10.1016/j.neuron.2010.01.034. URL: http://dx.doi.org/10.1016/j.neuron.2010.01.034 (cit. on pp. 10, 122).

— (2012). "Segmentation of spatial experience by hippocampal theta sequences." In: *Nature neuroscience* 15.7, pp. 1032–1039. DOI: 10.1038/nn.3138. URL: http://dx.doi.org/10.1038/nn.3138 (cit. on p. 109).

Guzowski, J., J. Knierim, and E. Moser (2004). "Ensemble dynamics of hippocampal regions CA3 and CA1." In: *Neuron* 44.4, pp. 581–584. DOI: 10.1016/j.neuron.

2004.11.003. URL: http://dx.doi.org/10.1016/j.neuron.2004.11.003 (cit. on p. 7).

Hafting, T., M. Fyhn, T. Bonnevie, M. Moser, and E. Moser (2008). "Hippocampus-independent phase precession in entorhinal grid cells." In: *Nature* 453.7199, pp. 1248–1252. DOI: 10.1038/nature06957. URL: http://dx.doi.org/10.1038/nature06957 (cit. on pp. 16, 17, 96).

Hafting, T., M. Fyhn, S. Molden, M. Moser, and E. Moser (2005). "Microstructure of a spatial map in the entorhinal cortex." In: *Nature* 436.7052, pp. 801–806. DOI: 10.1038/nature03721. URL: http://dx.doi.org/10.1038/nature03721 (cit. on pp. 1, 5, 12, 19, 125).

Halasy, K., E. Buhl, Z. Lorinczi, G. Tamas, and P. Somogyi (1996). "Synaptic target selectivity and input of GABAergic basket and bistratified interneurons in the CA1 area of the rat hippocampus." In: *Hippocampus* 6.3, pp. 306–329. DOI: 10.1002/(SICI)1098-1063(1996)6:3<306::AID-HIPO8>3.0.CO;2-K. URL: http://dx.doi.org/10.1002/(SICI)1098-1063(1996)6:3<306::AID-HIPO8>3.0.CO;2-K (cit. on p. 97).

Hamaguchi, K., J. Hatchett, and M. Okada (2006). "Analytic solution of neural network with disordered lateral inhibition." In: *Physical review. E, Statistical, nonlinear, and soft matter physics* 73.5 Pt 1 (cit. on p. 111).

Hangya, B., Y. Li, R. Muller, and A. Czurko (2010). "Complementary spatial firing in place cell-interneuron pairs." In: *The Journal of physiology* 588.Pt 21, pp. 4165–4175. DOI: 10.1113/jphysiol.2010.194274. URL: http://dx.doi.org/10.1113/jphysiol.2010.194274 (cit. on p. 106).

Harris, K., J. Csicsvari, H. Hirase, G. Dragoi, and G. Buzsaki (2003). "Organization of cell assemblies in the hippocampus." In: *Nature* 424.6948, pp. 552–556. DOI: 10.1038/nature01834. URL: http://dx.doi.org/10.1038/nature01834 (cit. on p. 104).

Harvey, C., F. Collman, D. Dombeck, and D. Tank (2009). "Intracellular dynamics of hippocampal place cells during virtual navigation." In: *Nature* 461.7266, pp. 941–946. DOI: 10.1038/nature08499. URL: http://dx.doi.org/10.1038/nature08499 (cit. on p. 27).

Hasselmo, M. E., C. Bodelón, and B. P. Wyble (2002). "A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning". In: *Neural Computation* 14 (4), pp. 793–817. ISSN: 0899-7667 (cit. on p. 94).

Hasselmo, M. (2008). "Grid cell mechanisms and function: contributions of entorhinal persistent spiking and phase resetting." In: *Hippocampus* 18.12, pp. 1213–1229. DOI: 10.1002/hipo.20512. URL: http://dx.doi.org/10.1002/hipo.20512 (cit. on p. 22).

Hasselmo, M. and M. Brandon (2008). "Linking cellular mechanisms to behavior: entorhinal persistent spiking and membrane potential oscillations may underlie path integration, grid cell firing, and episodic memory." In: *Neural plasticity* 2008. DOI: 10.1155/2008/658323. URL: http://dx.doi.org/10.1155/2008/658323 (cit. on p. 22).

Hasselmo, M., L. Giocomo, and E. Zilli (2007). "Grid cell firing may arise from interference of theta frequency membrane potential oscillations in single neurons." In: *Hippocampus* 17.12, pp. 1252–1271. DOI: 10.1002/hipo.20374. URL: http://dx.doi.org/10.1002/hipo.20374 (cit. on pp. 21, 38).

Hayman, R. and K. Jeffery (2008). "How heterogeneous place cell responding arises from homogeneous grids–a contextual gating hypothesis." In: *Hippocampus* 18.12, pp. 1301–1313. DOI: 10.1002/hipo.20513. URL: http://dx.doi.org/10.1002/hipo.20513 (cit. on p. 66).

Hayman, R., M. Verriotis, A. Jovalekic, A. Fenton, and K. Jeffery (2011). "Anisotropic encoding of three-dimensional space by place cells and grid cells." In: *Nature neuroscience* 14.9, pp. 1182–1188. DOI: 10.1038/nn.2892. URL: http://dx.doi.org/10.1038/nn.2892 (cit. on pp. 42, 111).

Healy, S., S. de Kort, and N. Clayton (2005). "The hippocampus, spatial memory and food hoarding: a puzzle revisited." In: *Trends in ecology & evolution* 20.1, pp. 17–22. DOI: 10.1016/j.tree.2004.10.006. URL: http://dx.doi.org/10.1016/j.tree.2004.10.006 (cit. on p. 41).

Hebb, D. (1949). *The organisation of behaviour*. New York: Wiley (cit. on p. 69).

Heys, J., M. Katrina, C. Moss, and M. Hasselmo (2013). "Bat and rat neurons differ in theta-frequency resonance despite similar coding of space." In: *Science (New York, N.Y.)* 340.6130, pp. 363–367. DOI: 10.1126/science.1233831. URL: http://dx.doi.org/10.1126/science.1233831 (cit. on pp. 17, 23).

Hok, V., P.-P. Lenck-Santini, S. Roux, E. Save, R. Muller, and B. Poucet (2007). "Goal-related activity in hippocampal place cells." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27.3, pp. 472–482. DOI: 10.1523/JNEUROSCI.2864-06.2007. URL: http://dx.doi.org/10.1523/JNEUROSCI.2864-06.2007 (cit. on pp. 9, 79).

Hollup, S. A., S. Molden, J. G. Donnett, M.-B. Moser, and E. I. Moser (2001). "Accumulation of hippocampal place fields at the goal location in an annular watermaze task". In: *The Journal of Neuroscience* 21.5, pp. 1635–1644 (cit. on pp. 9, 79).

Hopfield, J. (1982). "Neural networks and physical systems with emergent collective computational abilities". In: *Proceedings of the national academy of sciences* 79 (8), p. 2554 (cit. on p. 69).

Hubel, D. and T. Wiesel (1959). "Receptive fields of single neurones in the cat's striate cortex." In: *The Journal of physiology* 148, pp. 574–591 (cit. on p. 5).

Hunsaker, M., G. Mooy, J. Swift, and R. Kesner (2007). "Dissociations of the medial and lateral perforant path projections into dorsal DG, CA3, and CA1 for spatial and nonspatial (visual object) information processing." In: *Behavioral neuroscience* 121.4, pp. 742–750. DOI: 10.1037/0735-7044.121.4.742. URL: http://dx.doi.org/10.1037/0735-7044.121.4.742 (cit. on p. 75).

Isomura, Y., A. Sirota, S. Ozen, S. Montgomery, K. Mizuseki, D. Henze, and G. Buzsaki (2006). "Integration and segregation of activity in entorhinal-hippocampal subregions by neocortical slow oscillations." In: *Neuron* 52.5, pp. 871–882. DOI: 10.1016/j.neuron.2006.10.023. URL: http://dx.doi.org/10.1016/j.neuron.2006.10.023 (cit. on p. 108).

Itskov, V., D. Hansel, and M. Tsodyks (2011). "Short-Term Facilitation may Stabilize Parametric Working Memory Trace." In: *Frontiers in computational neuroscience* 5 (cit. on pp. 79, 83).

Jackson, J., A. Johnson, and A. Redish (2006). "Hippocampal sharp waves and reactivation during awake states depend on repeated sequential experience." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26.48, pp. 12415–12426. DOI: 10.1523/JNEUROSCI.4118-06.2006. URL: http://dx.doi.org/10.1523/JNEUROSCI.4118-06.2006 (cit. on p. 10).

Jacobs, J., C. Weidemann, J. Miller, A. Solway, J. Burke, X. Wei, N. Suthana, M. Sperling, A. Sharan, I. Fried, and M. Kahana (2013). "Direct recordings of grid-like neuronal activity in human spatial navigation." In: *Nature neuroscience* 16.9, pp. 1188–1190. DOI: 10.1038/nn.3466. URL: http://dx.doi.org/10.1038/nn.3466 (cit. on p. 16).

Jeffery, K., A. Jovalekic, M. Verriotis, and R. Hayman (2013). "Navigating in a three-dimensional world." In: *The Behavioral and brain sciences* 36.5, pp. 523–543. DOI: 10.1017/S0140525X12002476. URL: http://dx.doi.org/10.1017/S0140525X12002476 (cit. on pp. 41, 42, 63).

Jensen, O. and J. Lisman (2000). "Position reconstruction from an ensemble of hippocampal place cells: contribution of theta phase coding." In: *Journal of neurophysiology* 83.5, pp. 2602–2609 (cit. on p. 10).

Jezek, K., E. Henriksen, A. Treves, E. Moser, and M. Moser (2011). "Theta-paced flickering between place-cell maps in the hippocampus." In: *Nature* 478.7368, pp. 246–249. DOI: 10.1038/nature10439. URL: http://dx.doi.org/10.1038/nature10439 (cit. on pp. 3, 69, 71, 76, 98, 101, 102, 104, 107, 109, 127).

Johnson, A., J. Jackson, and A. D. Redish (2008). "Measuring distributed properties of neural representations beyond the decoding of local variables: implications for cognition". In: *Mechanisms of information processing in the brain: encoding of information in neural populations and networks*, pp. 95–9119 (cit. on p. 10).

Johnson, A. and A. Redish (2007). "Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27.45, pp. 12176–12189. DOI: 10.1523/jneurosci.3761-07.2007. URL: http://dx.doi.org/10.1523/jneurosci.3761-07.2007 (cit. on pp. 106, 107).

Jung, M., S. Wiener, and B. McNaughton (1994). "Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 14.12, pp. 7347–7356 (cit. on pp. 9, 97).

Kali, S. and P. Dayan (2000). "The involvement of recurrent connections in area CA3 in establishing the properties of place fields: a model". In: *The Journal of Neuroscience* 20 (19), p. 7463 (cit. on p. 71).

Káli, S. and P. Dayan (2000). "The involvement of recurrent connections in area CA3 in establishing the properties of place fields: a model." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 20.19, pp. 7463–7477 (cit. on p. 122).

Karlsson, M. and L. Frank (2009). "Awake replay of remote experiences in the hippocampus." In: *Nature neuroscience* 12.7, pp. 913–918. DOI: 10.1038/nn.2344. URL: http://dx.doi.org/10.1038/nn.2344 (cit. on pp. 10, 108).

Kelemen, E. and A. Fenton (2010). "Dynamic grouping of hippocampal neural activity during cognitive control of two spatial frames." In: *PLoS biology* 8.6. DOI: 10.1371/journal.pbio.1000403. URL: http://dx.doi.org/10.1371/journal.pbio.1000403 (cit. on p. 109).

Kensinger, E. A., M. T. Ullman, and S. Corkin (2001). "Bilateral medial temporal lobe damage does not affect lexical or grammatical processing: Evidence from amnesic patient HM". In: *Hippocampus* 11 (4), pp. 347–360 (cit. on p. 67).

Kesner, R. (2007). "A behavioral analysis of dentate gyrus function." In: *Progress in brain research* 163, pp. 567–576. DOI: 10.1016/S0079-6123(07)63030-1. URL: http://dx.doi.org/10.1016/S0079-6123(07)63030-1 (cit. on p. 75).

Killian, N., M. Jutras, and E. Buffalo (2012). "A map of visual space in the primate entorhinal cortex." In: *Nature*. DOI: 10.1038/nature11587. URL: http://dx.doi.org/10.1038/nature11587 (cit. on pp. 15, 17).

Klausberger, T., P. Magill, L. Marton, J. Roberts, P. Cobden, G. Buzsaki, and P. Somogyi (2003). "Brain-state- and cell-type-specific firing of hippocampal interneurons in vivo." In: *Nature* 421.6925, pp. 844–848. DOI: 10.1038/nature01374. URL: http://dx.doi.org/10.1038/nature01374 (cit. on p. 97).

Klausberger, T., A. Marton Laszloand Baude, J. Roberts, P. Magill, and P. Somogyi (2004). "Spike timing of dendrite-targeting bistratified cells during hippocampal network oscillations in vivo." In: *Nature neuroscience* 7.1, pp. 41–47. DOI: 10.1038/nn1159. URL: http://dx.doi.org/10.1038/nn1159 (cit. on p. 97).

Klausberger, T. and P. Somogyi (2008). "Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations." In: *Science (New York, N.Y.)* 321.5885, pp. 53–57. DOI: 10.1126/science.1149381. URL: http://dx.doi.org/10.1126/science.1149381 (cit. on p. 97).

Kloosterman, F., T. van Haeften, and F. Lopes da Silva (2004). "Two reentrant pathways in the hippocampal-entorhinal system." In: *Hippocampus* 14.8, pp. 1026–1039 (cit. on p. 1).

Knierim, J. and B. McNaughton (2001). "Hippocampal place-cell firing during movement in three-dimensional space." In: *Journal of neurophysiology* 85.1, pp. 105–116 (cit. on p. 41).

Knierim, J., B. McNaughton, and G. Poe (2000). "Three-dimensional spatial selectivity of hippocampal neurons during space flight." In: *Nature neuroscience* 3.3, pp. 209–

210. DOI: 10.1038/72910. URL: http://dx.doi.org/10.1038/72910 (cit. on p. 41).

Knierim, J. J., I. Lee, and E. L. Hargreaves (Sept. 1, 2006). "Hippocampal place cells: Parallel input streams, subregional processing, and implications for episodic memory". en. In: *Hippocampus* 16 (9), pp. 755–764. ISSN: 1098-1063. (Visited on 10/10/2011) (cit. on p. 75).

Knierim, J. and K. Zhang (2012). "Attractor Dynamics of Spatially Correlated Neural Activity in the Limbic System." In: *Annual review of neuroscience*. DOI: 10.1146/annurev-neuro-062111-150351. URL: http://dx.doi.org/10.1146/annurev-neuro-062111-150351 (cit. on pp. 69, 70, 89).

Komorowski, R., C. Garcia, A. Wilson, S. Hattori, M. Howard, and H. Eichenbaum (2013). "Ventral hippocampal neurons are shaped by experience to represent behaviorally relevant contexts." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33.18, pp. 8079–8087. DOI: 10.1523/JNEUROSCI.5458-12.2013. URL: http://dx.doi.org/10.1523/JNEUROSCI.5458-12.2013 (cit. on p. 9).

Kornblith, S., X. Cheng, S. Ohayon, and D. Tsao (2013). "A network for scene processing in the macaque temporal lobe." In: *Neuron* 79.4, pp. 766–781. DOI: 10.1016/j.neuron.2013.06.015. URL: http://dx.doi.org/10.1016/j.neuron.2013.06.015 (cit. on p. 15).

Kropff, E. and A. Treves (2008). "The emergence of grid cells: Intelligent design or just adaptation?" In: *Hippocampus* 18.12, pp. 1256–1269. DOI: 10.1002/hipo.20520. URL: http://dx.doi.org/10.1002/hipo.20520 (cit. on pp. 1, 24–27, 34, 35, 45, 57, 58, 61, 73, 126).

Lacy, J., M. Yassa, S. Stark, L. Muftuler, and C. Stark (2011). "Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity". In: *Learning & Memory* 18 (1), p. 15 (cit. on p. 76).

Langston, R., J. Ainge, J. Couey, C. Canto, T. Bjerknes, M. Witter, E. Moser, and M. Moser (2010). "Development of the spatial representation system in the rat." In: *Science (New York, N.Y.)* 328.5985, pp. 1576–1580. DOI: 10.1126/science.1188210. URL: http://dx.doi.org/10.1126/science.1188210 (cit. on pp. 16, 27, 35, 64).

Lansink, C., P. Goltstein, J. Lankelma, B. McNaughton, and C. Pennartz (2009). "Hippocampus leads ventral striatum in replay of place-reward information." In: *PLoS biology* 7.8. DOI: 10.1371/journal.pbio.1000173. URL: http://dx.doi.org/10.1371/journal.pbio.1000173 (cit. on p. 10).

Lassalle, J., T. Bataille, and H. Halley (2000). "Reversible inactivation of the hippocampal mossy fiber synapses in mice impairs spatial learning, but neither consolidation nor memory retrieval, in the Morris navigation task." In: *Neurobiology of learning and memory* 73.3, pp. 243–257. DOI: 10.1006/nlme.1999.3931. URL: http://dx.doi.org/10.1006/nlme.1999.3931 (cit. on p. 108).

Lavenex, P. and D. Amaral (2000). "Hippocampal-neocortical interaction: a hierarchy of associativity." In: *Hippocampus* 10.4, pp. 420–430 (cit. on p. 66).

Lawrence, J. and C. McBain (2003). "Interneuron diversity series: containing the detonation–feedforward inhibition in the CA3 hippocampus." In: *Trends in neurosciences* 26.11, pp. 631–640. DOI: 10.1016/j.tins.2003.09.007. URL: http://dx.doi.org/10.1016/j.tins.2003.09.007 (cit. on pp. 106, 109).

Lee, A. and M. Wilson (2002). "Memory of sequential experience in the hippocampus during slow wave sleep." In: *Neuron* 36.6, pp. 1183–1194. DOI: 10.1016/S0896-6273(02)01096-6. URL: http://dx.doi.org/10.1016/S0896-6273(02)01096-6 (cit. on p. 10).

Lee, I., A. Griffin, E. Zilli, H. Eichenbaum, and M. Hasselmo (2006). "Gradual translocation of spatial correlates of neuronal firing in the hippocampus toward prospective reward locations." In: *Neuron* 51.5, pp. 639–650. DOI: 10.1016/j.neuron.2006.

06.033. URL: http://dx.doi.org/10.1016/j.neuron.2006.06.033 (cit. on p. 9).

Lee, I. and R. Kesner (2004). "Encoding versus retrieval of spatial memory: double dissociation between the dentate gyrus and the perforant path inputs into CA3 in the dorsal hippocampus." In: *Hippocampus* 14.1, pp. 66–76. DOI: 10.1002/hipo.10167. URL: http://dx.doi.org/10.1002/hipo.10167 (cit. on p. 108).

Leutgeb, J., S. Leutgeb, M. Moser, and E. Moser (2007). "Pattern separation in the dentate gyrus and CA3 of the hippocampus." In: *Science (New York, N.Y.)* 315.5814, pp. 961–966. DOI: 10.1126/science.1135801. URL: http://dx.doi.org/10.1126/science.1135801 (cit. on pp. 24, 76).

Leutgeb, J., S. Leutgeb, A. Treves, R. Meyer, C. Barnes, M. Bruce, M. Moser, and E. Moser (2005). "Progressive transformation of hippocampal neuronal representations in "morphed" environments." In: *Neuron* 48.2, pp. 345–358. DOI: 10.1016/j.neuron.2005.09.007. URL: http://dx.doi.org/10.1016/j.neuron.2005.09.007 (cit. on p. 7).

Leutgeb, S., J. Leutgeb, C. Barnes, E. Moser, M. Bruce, and M. Moser (2005). "Independent codes for spatial and episodic memory in hippocampal neuronal ensembles." In: *Science (New York, N.Y.)* 309.5734, pp. 619–623. DOI: 10.1126/science.1114037. URL: http://dx.doi.org/10.1126/science.1114037 (cit. on pp. 8, 24, 76, 123).

Leutgeb, S., J. Leutgeb, E. Moser, and M. Moser (2006). "Fast rate coding in hippocampal CA3 cell ensembles." In: *Hippocampus* 16.9, pp. 765–774. DOI: 10.1002/hipo.20201. URL: http://dx.doi.org/10.1002/hipo.20201 (cit. on p. 8).

Leutgeb, S., J. Leutgeb, A. Treves, M. Moser, and E. Moser (2004). "Distinct ensemble codes in hippocampal areas CA3 and CA1." In: *Science (New York, N.Y.)* 305.5688, pp. 1295–1298. DOI: 10.1126/science.1100265. URL: http://dx.doi.org/10.1126/science.1100265 (cit. on pp. 7, 24).

Lever, C., S. Burton, A. Jeewajee, J. O'Keefe, and N. Burgess (2009). "Boundary vector cells in the subiculum of the hippocampal formation." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29.31, pp. 9771–9777. DOI: 10.1523/JNEUROSCI.1319-09.2009. URL: http://dx.doi.org/10.1523/JNEUROSCI.1319-09.2009 (cit. on p. 5).

Lisman, J. and G. Buzsaki (2008). "A neural coding scheme formed by the combined function of gamma and theta oscillations." In: *Schizophrenia bulletin* 34.5, pp. 974–980. DOI: 10.1093/schbul/sbn060. URL: http://dx.doi.org/10.1093/schbul/sbn060 (cit. on p. 110).

Lisman, J. and O. Jensen (2013). "The theta-gamma neural code." In: *Neuron* 77.6, pp. 1002–1016. DOI: 10.1016/j.neuron.2013.03.007. URL: http://dx.doi.org/10.1016/j.neuron.2013.03.007 (cit. on p. 110).

Lu, L., J. Leutgeb, A. Tsao, E. Henriksen, S. Leutgeb, C. Barnes, M. Witter, M. Moser, and E. Moser (2013). "Impaired hippocampal rate coding after lesions of the lateral entorhinal cortex." In: *Nature neuroscience* 16.8, pp. 1085–1093. DOI: 10.1038/nn.3462. URL: http://dx.doi.org/10.1038/nn.3462 (cit. on p. 66).

Lubenov, E. and A. Siapas (2009). "Hippocampal theta oscillations are travelling waves." In: *Nature* 459.7246, pp. 534–539. DOI: 10.1038/nature08010. URL: http://dx.doi.org/10.1038/nature08010 (cit. on p. 9).

Macarthur, B., A. Ma'ayan, and I. Lemischka (2009). "Systems biology of stem cell fate and cellular reprogramming." In: *Nature reviews. Molecular cell biology* 10.10, pp. 672–681. DOI: 10.1038/nrm2766. URL: http://dx.doi.org/10.1038/nrm2766 (cit. on p. 72).

Manns, J., E. Zilli, K. Ong, M. Hasselmo, and H. Eichenbaum (2007). "Hippocampal CA1 spiking during encoding and retrieval: relation to theta phase." In: *Neurobiology of learning and memory* 87.1, pp. 9–20. DOI: 10.1016/j.nlm.2006.05.007. URL: http://dx.doi.org/10.1016/j.nlm.2006.05.007 (cit. on p. 98).

Marozzi, E. and K. Jeffery (2012). "Place, space and memory cells." In: *Current biology : CB* 22.22, R939–R942. DOI: 10.1016/j.cub.2012.10.022. URL: http://dx.doi.org/10.1016/j.cub.2012.10.022 (cit. on pp. 8, 13, 15).

Marr, D. (1970). "A theory for cerebral neocortex". In: *Proceedings of the Royal Society of London. Series B, Biological Sciences*, pp. 161–234 (cit. on p. 67).

— (July 1, 1971a). "Simple memory: a theory for archicortex". In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 262 (841), pp. 23–81. ISSN: 0080-4622. (Visited on 07/08/2011) (cit. on p. 67).

Marr, D. (1971b). "Simple memory: a theory for archicortex". In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, pp. 23–81 (cit. on p. 10).

Mathis, A., A. Herz, and M. Stemmler (2012a). "Optimal population codes for space: grid cells outperform place cells." In: *Neural computation* 24.9, pp. 2280–2317. DOI: 10.1162/NECO_a_00319. URL: http://dx.doi.org/10.1162/NECO_a_00319 (cit. on pp. 65, 66, 74).

— (2012b). "Resolution of nested neuronal representations can be exponential in the number of neurons." In: *Physical review letters* 109.1. DOI: 10.1103/PhysRevLett.109.018103. URL: http://dx.doi.org/10.1103/PhysRevLett.109.018103 (cit. on pp. 65, 66, 74).

Maurer, A., S. Cowen, S. Burke, C. Barnes, and M. Bruce (2006). "Phase precession in hippocampal interneurons showing strong functional coupling to individual pyramidal cells." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26.52, pp. 13485–13492. DOI: 10.1523/JNEUROSCI.2882-06.2006. URL: http://dx.doi.org/10.1523/JNEUROSCI.2882-06.2006 (cit. on p. 106).

McHugh, T. T., M. Jones, J. Quinn, N. Balthasar, R. Coppari, J. Elmquist, B. Lowell, M. Fanselow, M. Wilson, and S. Tonegawa (2007). "Dentate gyrus NMDA receptors mediate rapid pattern separation in the hippocampal network." In: *Science (New York, N.Y.)* 317.5834, pp. 94–99. DOI: 10.1126/science.1140263. URL: http://dx.doi.org/10.1126/science.1140263 (cit. on p. 65).

McHugh, T. and S. Tonegawa (2009). "CA3 NMDA receptors are required for the rapid formation of a salient contextual representation." In: *Hippocampus* 19.12, pp. 1153–1158. DOI: 10.1002/hipo.20684. URL: http://dx.doi.org/10.1002/hipo.20684 (cit. on p. 65).

McNaughton, B. L. and R. G. M. Morris (1987). "Hippocampal synaptic enhancement and information storage within a distributed memory system". In: *Trends in Neurosciences* 10 (10), pp. 408–415. ISSN: 0166-2236 (cit. on p. 67).

McNaughton, B., C. Barnes, J. Meltzer, and R. Sutherland (1989). "Hippocampal granule cells are necessary for normal spatial learning but not for spatially-selective pyramidal cell discharge." In: *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale* 76.3, pp. 485–496. DOI: 10.1007/BF00248904. URL: http://dx.doi.org/10.1007/BF00248904 (cit. on p. 11).

McNaughton, B., F. Battaglia, O. Jensen, E. Moser, and M. Moser (2006). "Path integration and the neural basis of the 'cognitive map'." In: *Nature reviews. Neuroscience* 7.8, pp. 663–678. DOI: 10.1038/nrn1932. URL: http://dx.doi.org/10.1038/nrn1932 (cit. on pp. 19–21, 70, 108, 111).

Meer, M. A. van der, A. Johnson, S. Neil, and A. Redish (2010). "Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task." In: *Neuron* 67.1, pp. 25–32. DOI: 10.1016/j.neuron.2010.06.023. URL: http://dx.doi.org/10.1016/j.neuron.2010.06.023 (cit. on p. 66).

Mhatre, H., A. Gorchetchnikov, and S. Grossberg (2012). "Grid cell hexagonal patterns formed by fast self-organized learning within entorhinal cortex." In: *Hippocampus* 22.2, pp. 320–334. DOI: 10.1002/hipo.20901. URL: http://dx.doi.org/10.1002/hipo.20901 (cit. on p. 20).

Miles, R., K. Toth, A. Gulyas, N. Hajos, and T. Freund (1996). "Differences between somatic and dendritic inhibition in the hippocampus." In: *Neuron* 16.4, pp. 815–823 (cit. on p. 97).

Mishkin, M. (1982). "A memory system in the monkey". In: *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 298 (1089), p. 85 (cit. on p. 67).

Mizumori, S. J. (2008). *Hippocampal place fields: Relevance to learning and memory*. Oxford University Press, USA (cit. on p. 5).

Mizumori, S., G. Perez, M. Alvarado, C. Barnes, and B. McNaughton (1990). "Reversible inactivation of the medial septum differentially affects two forms of learning in rats." In: *Brain research* 528.1, pp. 12–20. DOI: 10.1016/0006-8993(90)90188-H. URL: http://dx.doi.org/10.1016/0006-8993(90)90188-H (cit. on p. 95).

Mizumori, S. (2013). "Context Prediction Analysis and Episodic Memory." In: *Frontiers in behavioral neuroscience* 7 (cit. on p. 66).

Mizuseki, K. and G. Buzsaki (2013). "Preconfigured, skewed distribution of firing rates in the hippocampus and entorhinal cortex." In: *Cell reports* 4.5, pp. 1010–1021. DOI: 10.1016/j.celrep.2013.07.039. URL: http://dx.doi.org/10.1016/j.celrep.2013.07.039 (cit. on pp. 11, 91).

Mizuseki, K., A. Sirota, E. Pastalkova, and G. Buzsaki (2009). "Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop." In: *Neuron* 64.2, pp. 267–280. DOI: 10.1016/j.neuron.2009.08.037. URL: http://dx.doi.org/10.1016/j.neuron.2009.08.037 (cit. on pp. 17, 95–98, 108).

Monaco, J. and L. Abbott (2011). "Modular Realignment of Entorhinal Grid Cell Activity as a Basis for Hippocampal Remapping". In: *The Journal of Neuroscience* 31 (25), p. 9414 (cit. on p. 76).

Monasson, R. and S. Rosay (2013). "Crosstalk and transitions between multiple spatial maps in an attractor neural network model of the hippocampus: Phase diagram". In: *Physical Review E* 87. DOI: 10.1103/PhysRevE.87.062813. URL: http://dx.doi.org/10.1103/PhysRevE.87.062813 (cit. on pp. 89, 129).

Montgomery, S., M. Betancur, and G. Buzsaki (2009). "Behavior-dependent coordination of multiple theta dipoles in the hippocampus." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29.5, pp. 1381–1394. DOI: 10.1523/JNEUROSCI.4339-08.2009. URL: http://dx.doi.org/10.1523/JNEUROSCI.4339-08.2009 (cit. on pp. 95, 98, 108).

Montgomery, S., A. Sirota, and G. Buzsaki (2008). "Theta and gamma coordination of hippocampal networks during waking and rapid eye movement sleep." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28.26, pp. 6731–6741. DOI: 10.1523/JNEUROSCI.1227-08.2008. URL: http://dx.doi.org/10.1523/JNEUROSCI.1227-08.2008 (cit. on pp. 95, 98).

Morris, A., J. Churchwell, R. Kesner, and P. Gilbert (2012). "Selective lesions of the dentate gyrus produce disruptions in place learning for adjacent spatial locations." In: *Neurobiology of learning and memory* 97.3, pp. 326–331. DOI: 10.1016/j.nlm.2012.02.005. URL: http://dx.doi.org/10.1016/j.nlm.2012.02.005 (cit. on p. 76).

Morris, R., P. Garrud, J. Rawlins, and J. O'Keefe (1982). "Place navigation impaired in rats with hippocampal lesions." In: *Nature* 297.5868, pp. 681–683 (cit. on p. 122).

Moser, E., M. Moser, and P. Andersen (1993). "Spatial learning impairment parallels the magnitude of dorsal hippocampal lesions, but is hardly present following ventral lesions." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 13.9, pp. 3916–3925 (cit. on p. 9).

Moser, E., E. Kropff, and M. Moser (2008). "Place cells, grid cells, and the brain's spatial representation system." In: *Annual review of neuroscience* 31, pp. 69–89. DOI: 10.1146/annurev.neuro.31.061307.090723. URL: http://dx.doi.org/10.1146/annurev.neuro.31.061307.090723 (cit. on p. 5).

Moser, E. and M. Moser (2013). "Grid cells and neural coding in high-end cortices." In: *Neuron* 80.3, pp. 765–774. DOI: 10.1016/j.conb.2013.08.012. URL: http://dx.doi.org/10.1016/j.conb.2013.08.012 (cit. on p. 5).

Moser, M., E. Moser, E. Forrest, P. Andersen, and R. Morris (1995). "Spatial learning with a minislab in the dorsal hippocampus." In: *Proceedings of the National Academy of Sciences of the United States of America* 92.21, pp. 9697–9701. DOI: 10.1073/pnas.92.21.9697. URL: http://dx.doi.org/10.1073/pnas.92.21.9697 (cit. on p. 9).

Muller, R. and M. Stead (1996). "Hippocampal place cells connected by Hebbian synapses can solve spatial problems." In: *Hippocampus* 6.6, pp. 709–719. DOI: 10.1002/(SICI)1098-1063(1996)6:6<709::AID-HIPO13>3.0.CO;2-4. URL: http://dx.doi.org/10.1002/(SICI)1098-1063(1996)6:6%3C709::AID-HIPO13%3E3.0.CO;2-4 (cit. on p. 111).

Muller, R., M. Stead, and J. Pach (1996). "The hippocampus as a cognitive graph." In: *The Journal of general physiology* 107.6, pp. 663–694. DOI: 10.1085/jgp.107.6.663. URL: http://dx.doi.org/10.1085/jgp.107.6.663 (cit. on p. 111).

Naber, P., C. M, J. B, and M. Witter (1997). "Parallel input to the hippocampal memory system through peri- and postrhinal cortices." In: *Neuroreport* 8.11, pp. 2617–2621 (cit. on p. 66).

Nadasdy, Z., H. Hirase, A. Czurko, J. Csicsvari, and G. Buzsaki (1999). "Replay and time compression of recurring spike sequences in the hippocampus." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 19.21, pp. 9497–9507 (cit. on p. 10).

Nadel, L. and M. Moscovitch (1997). "Memory consolidation, retrograde amnesia and the hippocampal complex". In: *Current Opinion in Neurobiology* 7 (2), pp. 217–227 (cit. on p. 67).

Nakashiba, T., D. Buhl, T. McHugh, and S. Tonegawa (2009). "Hippocampal CA3 output is crucial for ripple-associated reactivation and consolidation of memory." In: *Neuron* 62.6, pp. 781–787. DOI: 10.1016/j.neuron.2009.05.013. URL: http://dx.doi.org/10.1016/j.neuron.2009.05.013 (cit. on pp. 10, 108).

Nakashiba, T., J. Young, T. McHugh, D. Buhl, and S. Tonegawa (2008). "Transgenic inhibition of synaptic transmission reveals role of CA3 output in hippocampal learning." In: *Science (New York, N.Y.)* 319.5867, pp. 1260–1264. DOI: 10.1126/science.1151120. URL: http://dx.doi.org/10.1126/science.1151120 (cit. on p. 65).

Nakazawa, K., L. Sun, M. Quirk, L. Rondi-Reig, M. Wilson, and S. Tonegawa (2003). "Hippocampal CA3 NMDA receptors are crucial for memory acquisition of one-time experience." In: *Neuron* 38.2, pp. 305–315 (cit. on p. 65).

Navratilova, Z., L. Giocomo, J.-M. Fellous, M. Hasselmo, and B. McNaughton (2012). "Phase precession and variable spatial scaling in a periodic attractor map model of medial entorhinal grid cells with realistic after-spike dynamics." In: *Hippocampus* 22.4, pp. 772–789. DOI: 10.1002/hipo.20939. URL: http://dx.doi.org/10.1002/hipo.20939 (cit. on p. 20).

O'Keefe, J. and J. Dostrovsky (1971). "The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat." In: *Brain research* 34.1, pp. 171–175 (cit. on pp. 1, 5, 19, 125).

O'Keefe, J. and M. Recce (1993). "Phase relationship between hippocampal place units and the EEG theta rhythm." In: *Hippocampus* 3.3, pp. 317–330. DOI: 10.1002/hipo.450030307. URL: http://dx.doi.org/10.1002/hipo.450030307 (cit. on p. 16).

O'Keefe, J. and L. Nadel (1978). "The hippocampus as a cognitive map". In: 3 (cit. on pp. 6, 10).

O'Neill, J., T. Senior, K. Allen, J. Huxter, and J. Csicsvari (2008). "Reactivation of experience-dependent cell assembly patterns in the hippocampus." In: *Nature neuroscience* 11.2, pp. 209–215. DOI: 10.1038/nn2037. URL: http://dx.doi.org/10.1038/nn2037 (cit. on p. 9).

O'Neill, J., T. Senior, and J. Csicsvari (2006). "Place-selective firing of CA1 pyramidal cells during sharp wave/ripple network patterns in exploratory behavior." In:

Neuron 49.1, pp. 143–155. DOI: 10.1016/j.neuron.2005.10.037. URL: http://dx.doi.org/10.1016/j.neuron.2005.10.037 (cit. on p. 10).

Panzeri, S., R. Senatore, M. Montemurro, and R. Petersen (2007). "Correcting for the sampling bias problem in spike train information measures." In: *Journal of neurophysiology* 98.3, pp. 1064–1072. DOI: 10.1152/jn.00559.2007. URL: http://dx.doi.org/10.1152/jn.00559.2007 (cit. on p. 112).

Papp, G., M. Witter, and A. Treves (2007). "The CA3 network as a memory store for spatial representations." In: *Learning & memory (Cold Spring Harbor, N.Y.)* 14.11, pp. 732–744. DOI: 10.1101/lm.687407. URL: http://dx.doi.org/10.1101/lm.687407 (cit. on pp. 79, 111).

Patel, J., S. Fujisawa, A. Berenyi, S. Royer, and G. Buzsaki (2012). "Traveling theta waves along the entire septotemporal axis of the hippocampus." In: *Neuron* 75.3, pp. 410–417. DOI: 10.1016/j.neuron.2012.07.015. URL: http://dx.doi.org/10.1016/j.neuron.2012.07.015 (cit. on p. 9).

Patel, J., E. Schomburg, A. Berenyi, S. Fujisawa, and G. Buzsaki (2013). "Local Generation and Propagation of Ripples along the Septotemporal Axis of the Hippocampus." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33.43, pp. 17029–17041. DOI: 10.1523/JNEUROSCI.2036-13.2013. URL: http://dx.doi.org/10.1523/JNEUROSCI.2036-13.2013 (cit. on p. 108).

Pennartz, C., J. Berke, A. Graybiel, R. Ito, C. Lansink, M. van der Meer, A. Redish, K. Smith, and P. Voorn (2009). "Corticostriatal Interactions during Learning, Memory Processing, and Decision Making." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29.41, pp. 12831–12838 (cit. on p. 66).

Petkovic, M. (2009). *Famous puzzles of great mathematicians*. Ameri- American Mathematical Society (cit. on pp. 43, 65).

Pfeiffer, B. and D. Foster (2013). "Hippocampal place-cell sequences depict future paths to remembered goals." In: *Nature* 497.7447, pp. 74–79. DOI: 10.1038/nature12112. URL: http://dx.doi.org/10.1038/nature12112 (cit. on pp. 10, 11, 94).

Pilly, P. and S. Grossberg (2013). "Spiking neurons in a hierarchical self-organizing map model can learn to develop spatial and temporal properties of entorhinal grid cells and hippocampal place cells." In: *PloS one* 8.4. DOI: 10.1371/journal.pone.0060599. URL: http://dx.doi.org/10.1371/journal.pone.0060599 (cit. on p. 24).

Poppenk, J., H. Evensmoen, M. Moscovitch, and L. Nadel (2013). "Long-axis specialization of the human hippocampus." In: *Trends in cognitive sciences* 17.5, pp. 230–240. DOI: 10.1016/j.tics.2013.03.005. URL: http://dx.doi.org/10.1016/j.tics.2013.03.005 (cit. on p. 9).

Pouille, F. and M. Scanziani (2001). "Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition." In: *Science (New York, N.Y.)* 293.5532, pp. 1159–1163. DOI: 10.1126/science.1060342. URL: http://dx.doi.org/10.1126/science.1060342 (cit. on p. 106).

Pouille, F. and M. Scanziani (2004). "Routing of spike series by dynamic circuits in the hippocampus." In: *Nature* 429.6993, pp. 717–723. DOI: 10.1038/nature02615. URL: http://dx.doi.org/10.1038/nature02615 (cit. on p. 106).

Qu, H., Z. Yi, and X. Wang (2009). "A Winner-Take-All Neural Networks of N Linear Threshold Neurons without Self-Excitatory Connections". In: *Neural Processing Letters* 29. DOI: 10.1007/s11063-009-9100-x. URL: http://dx.doi.org/10.1007/s11063-009-9100-x (cit. on p. 20).

Quian Quiroga, R. and S. Panzeri (2009). "Extracting information from neuronal populations: information theory and decoding approaches." In: *Nature reviews. Neuroscience* 10.3, pp. 173–185. DOI: 10.1038/nrn2578. URL: http://dx.doi.org/10.1038/nrn2578 (cit. on p. 114).

Quilichini, P., A. Sirota, and G. Buzsaki (2010). "Intrinsic circuit organization and theta-gamma oscillation dynamics in the entorhinal cortex of the rat." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30.33, pp. 11128–11142. DOI: 10.1523/JNEUROSCI.1327-10.2010. URL: http://dx.doi.org/10.1523/JNEUROSCI.1327-10.2010 (cit. on p. 96).

Quiroga, R., L. Reddy, G. Kreiman, C. Koch, and I. Fried (2005). "Invariant visual representation by single neurons in the human brain." In: *Nature* 435.7045, pp. 1102–1107. DOI: 10.1038/nature03687. URL: http://dx.doi.org/10.1038/nature03687 (cit. on p. 68).

Ramirez, S., X. Liu, P. Lin, J. Suh, M. Pignatelli, R. Redondo, T. Ryan, and S. Tonegawa (2013). "Creating a false memory in the hippocampus." In: *Science (New York, N.Y.)* 341.6144, pp. 387–391. DOI: 10.1126/science.1239073. URL: http://dx.doi.org/10.1126/science.1239073 (cit. on p. 68).

Rawlins, J., J. Feldon, and J. Gray (1979). "Septo-hippocampal connections and the hippocampal theta rhythm." In: *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale* 37.1, pp. 49–63. DOI: 10.1007/BF01474253. URL: http://dx.doi.org/10.1007/BF01474253 (cit. on p. 95).

Remondes, M. and E. Schuman (2002). "Direct cortical input modulates plasticity and spiking in CA1 pyramidal neurons." In: *Nature* 416.6882, pp. 736–740. DOI: 10.1038/416736a. URL: http://dx.doi.org/10.1038/416736a (cit. on p. 106).

Renart, A., P. Song, and X.-J. Wang (2003). "Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks." In: *Neuron* 38.3, pp. 473–485. DOI: 10.1016/S0896-6273(03)00255-1. URL: http://dx.doi.org/10.1016/S0896-6273(03)00255-1 (cit. on p. 79).

Renno-Costa, C., J. Lisman, and P. Verschure (2010). "The mechanism of rate remapping in the dentate gyrus". In: *Neuron* 68 (6), pp. 1051–1058 (cit. on p. 76).

Ribot, T. (1881). "Les maladies de la memoire [Diseases of memory]". In: *New York: Appleton-Century-Crofts* (cit. on p. 67).

Rolls, E. T. (1989). "Functions of neuronal networks in the hippocampus and cerebral cortex in memory". In: *Models of brain function*, pp. 15–33 (cit. on pp. 67, 69).

Rolls, E. and S. O'Mara (1995). "View-responsive neurons in the primate hippocampal complex." In: *Hippocampus* 5.5, pp. 409–424. DOI: 10.1002/hipo.450050504. URL: http://dx.doi.org/10.1002/hipo.450050504 (cit. on p. 15).

Rolls, E., R. Robertson, and P. Georges-Francois (1997). "Spatial view cells in the primate hippocampus." In: *The European journal of neuroscience* 9.8, pp. 1789–1794. DOI: 10.1111/j.1460-9568.1997.tb01538.x. URL: http://dx.doi.org/10.1111/j.1460-9568.1997.tb01538.x (cit. on p. 15).

Rolls, E. and A. Treves (2011). "The neuronal encoding of information in the brain." In: *Progress in neurobiology* 95.3, pp. 448–490. DOI: 10.1016/j.pneurobio.2011.08.002. URL: http://dx.doi.org/10.1016/j.pneurobio.2011.08.002 (cit. on pp. 72, 113, 115).

Rolls, E. (1996). "A theory of hippocampal function in memory". In: *Hippocampus* 6 (6), pp. 601–620 (cit. on p. 72).

Romani, S. and M. Tsodyks (2010). "Continuous Attractors with Morphed/Correlated Maps". In: *PLoS Computational Biology* 6 (8), e1000869 (cit. on pp. 70, 71).

Rosapia, L., E. Ciaramelli, C. Piccini, and A. Treves (2007). "Differential impact of brain damage on the access mode to memory representations: an information theoretic approach." In: *The European journal of neuroscience* 26.10, pp. 2702–2712. DOI: 10.1111/j.1460-9568.2007.05881.x. URL: http://dx.doi.org/10.1111/j.1460-9568.2007.05881.x (cit. on p. 112).

Rosay, S. and R. Monasson (2013). "Cross-talk and transitions between multiple environments in an attractor neural network model of the hippocampus". In: *BMC Neuroscience* 14 (cit. on pp. 3, 89, 92).

Roudi, Y. and A. Treves (2008). "Representing where along with what information in a model of a cortical patch." In: *PLoS computational biology* 4.3. DOI: 10.1371/journal.pcbi.1000012. URL: http://dx.doi.org/10.1371/journal.pcbi.1000012 (cit. on pp. 71, 111).

Royer, S., A. Sirota, J. Patel, and G. Buzsaki (2010). "Distinct representations and theta dynamics in dorsal and ventral hippocampus." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30.5, pp. 1777–1787.

DOI: 10.1523/JNEUROSCI.4681-09.2010. URL: http://dx.doi.org/10.1523/JNEUROSCI.4681-09.2010 (cit. on p. 97).

Sammon Jr, J. W. (1969). "A nonlinear mapping for data structure analysis". In: *Computers, IEEE Transactions on* 100.5, pp. 401–409 (cit. on p. 114).

Samsonovich, A. and B. McNaughton (1997). "Path integration and cognitive mapping in a continuous attractor neural network model." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 17.15, pp. 5900–5920 (cit. on pp. 19, 70, 89, 106, 111, 127).

Sanders, H. I. and E. K. Warrington (1971). "Memory for remote events in amnesic patients". In: *Brain* 94 (4), p. 661 (cit. on p. 68).

Sargolini, F., M. Fyhn, T. Hafting, M. Bruce, M. Witter, M. Moser, and E. Moser (2006). "Conjunctive representation of position, direction, and velocity in entorhinal cortex." In: *Science (New York, N.Y.)* 312.5774, pp. 758–762. DOI: 10.1126/science.1125572. URL: http://dx.doi.org/10.1126/science.1125572 (cit. on pp. 14, 58).

Savelli, F., D. Yoganarasimha, and J. Knierim (2008). "Influence of boundary removal on the spatial representations of the medial entorhinal cortex." In: *Hippocampus* 18.12, pp. 1270–1282. DOI: 10.1002/hipo.20511. URL: http://dx.doi.org/10.1002/hipo.20511 (cit. on pp. 15, 27).

Scarpetta, S., L. Zhaoping, and J. Hertz (2002). "Hebbian imprinting and retrieval in oscillatory neural networks." In: *Neural computation* 14.10, pp. 2371–2396. DOI: 10.1162/08997660260293265. URL: http://dx.doi.org/10.1162/08997660260293265 (cit. on p. 101).

Schmidt-Hieber, C. and M. Hausser (2013). "Cellular mechanisms of spatial navigation in the medial entorhinal cortex." In: *Nature neuroscience* 16.3, pp. 325–331. DOI: 10.1038/nn.3340. URL: http://dx.doi.org/10.1038/nn.3340 (cit. on p. 23).

Schmolck, H., E. A. Kensinger, S. Corkin, and L. R. Squire (2002). "Semantic knowledge in patient HM and other patients with bilateral medial and lateral temporal lobe lesions". In: *Hippocampus* 12 (4), pp. 520–533 (cit. on p. 67).

Scoville, W. B. and B. Milner (1957). "Loss of recent memory after bilateral hippocampal lesions". In: *Journal of Neurology, Neurosurgery & Psychiatry* 20 (1), p. 11 (cit. on pp. 66, 125).

Si, B. and A. Treves (2009). "The role of competitive learning in the generation of DG fields from EC inputs". In: *Cognitive Neurodynamics* 3 (2), pp. 177–187 (cit. on p. 76).

Si, B., E. Kropff, and A. Treves (2012). "Grid alignment in entorhinal cortex." In: *Biological cybernetics* 106.8-9, pp. 483–506. DOI: 10.1007/s00422-012-0513-7. URL: http://dx.doi.org/10.1007/s00422-012-0513-7 (cit. on pp. 24, 25, 38, 49, 51, 52, 56, 73, 74).

Si, B. and A. Treves (2013). "A model for the differentiation between grid and conjunctive units in medial entorhinal cortex." In: *Hippocampus*. DOI: 10.1002/hipo.22194. URL: http://dx.doi.org/10.1002/hipo.22194 (cit. on pp. 58, 63).

Singer, A., M. Karlsson, A. Nathe, M. Carr, and L. Frank (2010). "Experience-dependent development of coordinated hippocampal spatial activity representing the similarity of related locations." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30.35, pp. 11586–11604. DOI: 10.1523/JNEUROSCI.0926-10.2010. URL: http://dx.doi.org/10.1523/JNEUROSCI.0926-10.2010 (cit. on pp. 8, 123).

Sirota, A. and G. Buzsaki (2005). "Interaction between neocortical and hippocampal networks via slow oscillations." In: *Thalamus & related systems* 3.4, pp. 245–259. DOI: 10.1017/S1472928807000258. URL: http://dx.doi.org/10.1017/S1472928807000258 (cit. on p. 108).

Skaggs, W. and B. McNaughton (1996). "Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience." In: *Science (New York, N.Y.)* 271.5257, pp. 1870–1873. DOI: 10.1126/science.271.5257.1870. URL: http://dx.doi.org/10.1126/science.271.5257.1870 (cit. on p. 10).

Smith, D. and S. Mizumori (2006). "Learning-related development of context-specific neuronal responses to places and events: the hippocampal role in context processing." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26.12, pp. 3154–3163. DOI: 10.1523/JNEUROSCI.3234-05.2006. URL: http://dx.doi.org/10.1523/JNEUROSCI.3234-05.2006 (cit. on p. 9).

Solstad, T., C. Boccara, E. Kropff, M. Moser, and E. Moser (2008). "Representation of geometric borders in the entorhinal cortex." In: *Science (New York, N.Y.)* 322.5909, pp. 1865–1868. DOI: 10.1126/science.1166466. URL: http://dx.doi.org/10.1126/science.1166466 (cit. on pp. 5, 15, 27).

Somogyi, P. and T. Klausberger (2005). "Defined types of cortical interneurone structure space and spike timing in the hippocampus." In: *The Journal of physiology* 562.Pt 1, pp. 9–26. DOI: 10.1113/jphysiol.2004.078915. URL: http://dx.doi.org/10.1113/jphysiol.2004.078915 (cit. on p. 97).

Song, P. and X. Wang (2005). "Angular path integration by moving hill of activity: a spiking neuron model without recurrent excitation of the head-direction system." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 25.4, pp. 1002–1014. DOI: 10.1523/JNEUROSCI.4172-04.2005. URL: http://dx.doi.org/10.1523/JNEUROSCI.4172-04.2005 (cit. on p. 20).

Spiers, H., R. Hayman, A. Jovalekic, E. Marozzi, and K. Jeffery (2013). "Place Field Repetition and Purely Local Remapping in a Multicompartment Environment." In: *Cerebral cortex (New York, N.Y. : 1991)*. DOI: 10.1093/cercor/bht198. URL: http://dx.doi.org/10.1093/cercor/bht198 (cit. on pp. 8, 9).

Squire, L. and S. Zola-Morgan (1991). "The medial temporal lobe memory system". In: *Science* 253 (5026), p. 1380 (cit. on p. 67).

Sreenivasan, S. and I. Fiete (2011). "Grid cells generate an analog error-correcting code for singularly precise neural computation." In: *Nature neuroscience* 14.10, pp. 1330–1337. DOI: 10.1038/nn.2901. URL: http://dx.doi.org/10.1038/nn.2901 (cit. on pp. 65, 66, 74).

Stella, F., E. Cerasti, B. Si, K. Jezek, and A. Treves (2012). "Self-organization of multiple spatial and context memories in the hippocampus." In: *Neuroscience and biobehavioral reviews* 36.7, pp. 1609–1625. DOI: 10.1016/j.neubiorev.2011.12.002. URL: http://dx.doi.org/10.1016/j.neubiorev.2011.12.002 (cit. on p. 2).

Stella, F., E. Cerasti, and A. Treves (2013). "Unveiling the metric structure of internal representations of space." In: *Frontiers in neural circuits* 7. DOI: 10.3389/fncir.2013.00081. URL: http://dx.doi.org/10.3389/fncir.2013.00081 (cit. on p. 3).

Stella, F., B. Si, E. Kropff, and A. Treves (2013a). "Grid cells on the ball". In: *Journal of Statistical Mechanics: Theory and Experiment* 2013. DOI: 10.1088/1742-5468/2013/03/P03013. URL: http://dx.doi.org/10.1088/1742-5468/2013/03/P03013 (cit. on pp. 1, 27).

— (2013b). "Grid maps for spaceflight, anyone? They are for free!" In: *The Behavioral and brain sciences* 36.5, pp. 566–567. DOI: 10.1017/S0140525X13000575. URL: http://dx.doi.org/10.1017/S0140525X13000575 (cit. on pp. 2, 27).

Stella, F. and A. Treves (2011). "Associative memory storage and retrieval: involvement of theta oscillations in hippocampal information processing." In: *Neural plasticity* 2011. DOI: 10.1155/2011/683961. URL: http://dx.doi.org/10.1155/2011/683961 (cit. on pp. 2, 71).

Stensola, H., T. Stensola, T. Solstad, K. Frøland, M. Moser, and E. Moser (2012). "The entorhinal grid map is discretized." In: *Nature* 492.7427, pp. 72–78. DOI: 10.1038/nature11649. URL: http://dx.doi.org/10.1038/nature11649 (cit. on pp. 13, 14).

Striedter, G. (2005). *Principles of brain evolution.* Sinauer Associates (cit. on p. 63).

Strien, N. van, N. Cappaert, and M. Witter (2009). "The anatomy of memory: an interactive overview of the parahippocampal-hippocampal network." In: *Nature reviews.*

Neuroscience 10.4, pp. 272–282. DOI: 10.1038/nrn2614. URL: http://dx.doi.org/10.1038/nrn2614 (cit. on pp. 9, 12).

Stringer, S., T. Trappenberg, E. Rolls, and I. de Araujo (2002). "Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells." In: *Network (Bristol, England)* 13.2, pp. 217–242 (cit. on p. 111).

Sullivan, D., J. Csicsvari, K. Mizuseki, S. Montgomery, K. Diba, and G. Buzsaki (2011). "Relationships between hippocampal sharp waves, ripples, and fast gamma oscillation: influence of dentate and entorhinal cortical activity." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31.23, pp. 8605–8616. DOI: 10.1523/JNEUROSCI.0294-11.2011. URL: http://dx.doi.org/10.1523/JNEUROSCI.0294-11.2011 (cit. on p. 108).

Sutherland, G. and B. McNaughton (2000). "Memory trace reactivation in hippocampal and neocortical neuronal ensembles." In: *Current opinion in neurobiology* 10.2, pp. 180–186. DOI: 10.1016/S0959-4388(00)00079-9. URL: http://dx.doi.org/10.1016/S0959-4388(00)00079-9 (cit. on p. 10).

Suzuki, W. A. and D. G. Amaral (1994). "Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices". In: *The Journal of neuroscience* 14 (3), p. 1856 (cit. on p. 67).

Szabadics, J. and I. Soltesz (2009). "Functional specificity of mossy fiber innervation of GABAergic cells in the hippocampus." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29.13, pp. 4239–4251. DOI: 10.1523/JNEUROSCI.5390-08.2009. URL: http://dx.doi.org/10.1523/JNEUROSCI.5390-08.2009 (cit. on p. 106).

Tabuchi, E., A. Mulder, and S. Wiener (2003). "Reward value invariant place responses and reward site associated activity in hippocampal neurons of behaving rats." In: *Hippocampus* 13.1, pp. 117–132. DOI: 10.1002/hipo.10056. URL: http://dx.doi.org/10.1002/hipo.10056 (cit. on p. 9).

Taube, J., R. Muller, and J. Ranck (1990a). "Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 10.2, pp. 420–435 (cit. on pp. 5, 14, 19).

— (1990b). "Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 10.2, pp. 436–447 (cit. on pp. 14, 19).

Tolman, E. (1948). "Cognitive maps in rats and men." In: *Psychological review* 55.4, pp. 189–208 (cit. on pp. 6, 41).

Torborg, C., T. Nakashiba, S. Tonegawa, and C. McBain (2010). "Control of CA3 output by feedforward inhibition despite developmental changes in the excitation-inhibition balance." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30.46, pp. 15628–15637. DOI: 10.1523/JNEUROSCI.3099-10.2010. URL: http://dx.doi.org/10.1523/JNEUROSCI.3099-10.2010 (cit. on p. 106).

Touretzky, D. and A. Redish (1996). "Theory of rodent navigation based on interacting representations of space." In: *Hippocampus* 6.3, pp. 247–270. DOI: 10.1002/(SICI)1098-1063(1996)6:3<247::AID-HIPO4>3.0.CO;2-K. URL: http://dx.doi.org/10.1002/(SICI)1098-1063(1996)6:3<247::AID-HIPO4>3.0.CO;2-K (cit. on p. 19).

Treves, A. (1997). "On the perceptual structure of face space." In: *Bio Systems* 40.1-2, pp. 189–196. DOI: 10.1016/0303-2647(96)01645-0. URL: http://dx.doi.org/10.1016/0303-2647(96)01645-0 (cit. on p. 112).

Treves, A., P. Georges-Francois, and S. Panzeri (1998). "The metric content of spatial views as represented in the primate hippocampus". In: *Neural Circuits and Networks* (cit. on p. 112).

Treves, A. and E. Rolls (1992). "Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network." In: *Hippocampus* 2.2,

pp. 189–199. DOI: 10.1002/hipo.450020209. URL: http://dx.doi.org/10.1002/hipo.450020209 (cit. on pp. 76, 94).

Treves, A. (1990). "Graded-response neurons and information encodings in autoassociative memories." In: *Physical review. A* 42.4, pp. 2418–2430 (cit. on p. 83).

Tse, D., R. Langston, M. Kakeyama, I. Bethus, P. Spooner, E. Wood, M. Witter, and R. Morris (2007). "Schemas and memory consolidation." In: *Science (New York, N.Y.)* 316.5821, pp. 76–82. DOI: 10.1126/science.1135935. URL: http://dx.doi.org/10.1126/science.1135935 (cit. on p. 10).

Tsoar, A., R. Nathan, Y. Bartan, A. Vyssotski, G. Dell'Omo, and N. Ulanovsky (2011). "Large-scale navigational map in a mammal." In: *Proceedings of the National Academy of Sciences of the United States of America* 108.37, E718–E724. DOI: 10.1073/pnas.1107365108. URL: http://dx.doi.org/10.1073/pnas.1107365108 (cit. on p. 41).

Tsodyks, M. and T. Sejnowski (1995a). "Associative memory and hippocampal place cells". In: *International journal of neural systems* 6, pp. 81–86 (cit. on p. 111).

Tsodyks, M. (1999). "Attractor neural network models of spatial maps in hippocampus". en. In: *Hippocampus* 9 (4), pp. 481–489. ISSN: 1098-1063. (Visited on 10/09/2011) (cit. on p. 70).

Tsodyks, M. and T. Sejnowski (1995b). "Rapid state switching in balanced cortical network models". In: *Network: Computation in Neural Systems* 6 (2), pp. 111–124 (cit. on p. 71).

Tulving, E. (1995). "Organization of memory: Quo vadis". In: *The cognitive neurosciences*, pp. 839–847 (cit. on p. 67).

Ulanovsky, N. and C. Moss (2011). "Dynamics of hippocampal spatial representation in echolocating bats." In: *Hippocampus* 21.2, pp. 150–161. DOI: 10.1002/hipo.20731. URL: http://dx.doi.org/10.1002/hipo.20731 (cit. on p. 111).

Unsold, A. (1927). "Beitrage zur Quantenmechanik der Atome". In: *Annalen der Physik* 387. DOI: 10.1002/andp.19273870304. URL: http://dx.doi.org/10.1002/andp.19273870304 (cit. on p. 32).

Vanderwolf, C. (1969). "Hippocampal electrical activity and voluntary movement in the rat". In: *Electroencephalography and Clinical Neurophysiology* 26. DOI: 10.1016/0013-4694(69)90092-3. URL: http://dx.doi.org/10.1016/0013-4694(69)90092-3 (cit. on p. 95).

Viskontas, I., R. Quiroga, and I. Fried (2009). "Human medial temporal lobe neurons respond preferentially to personally relevant images." In: *Proceedings of the National Academy of Sciences of the United States of America* 106.50, pp. 21329–21334. DOI: 10.1073/pnas.0902319106. URL: http://dx.doi.org/10.1073/pnas.0902319106 (cit. on p. 68).

Welday, A., I. Shlifer, M. Bloom, K. Zhang, and H. Blair (2011). "Cosine directional tuning of theta cell burst frequencies: evidence for spatial coding by oscillatory interference." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31.45, pp. 16157–16176. DOI: 10.1523/JNEUROSCI.0712-11.2011. URL: http://dx.doi.org/10.1523/JNEUROSCI.0712-11.2011 (cit. on p. 22).

Wills, T., C. Barry, and F. Cacucci (2012). "The abrupt development of adult-like grid cell firing in the medial entorhinal cortex." In: *Frontiers in neural circuits* 6. DOI: 10.3389/fncir.2012.00021. URL: http://dx.doi.org/10.3389/fncir.2012.00021 (cit. on p. 16).

Wills, T., F. Cacucci, N. Burgess, and J. O'Keefe (2010). "Development of the hippocampal cognitive map in preweanling rats." In: *Science (New York, N.Y.)* 328.5985, pp. 1573–1576. DOI: 10.1126/science.1188224. URL: http://dx.doi.org/10.1126/science.1188224 (cit. on pp. 16, 27, 35, 64).

Wills, T., C. Lever, F. Cacucci, N. Burgess, and J. O'Keefe (2005). "Attractor dynamics in the hippocampal representation of the local environment." In: *Science (New York, N.Y.)* 308.5723, pp. 873–876. DOI: 10.1126/science.1108905. URL: http://dx.doi.org/10.1126/science.1108905 (cit. on pp. 7, 69, 113).

Wilson, M. and B. McNaughton (1993). "Dynamics of the hippocampal ensemble code for space." In: *Science (New York, N.Y.)* 261.5124, pp. 1055–1058 (cit. on pp. 6, 79).

— (1994). "Reactivation of hippocampal ensemble memories during sleep." In: *Science (New York, N.Y.)* 265.5172, pp. 676–679. DOI: 10.1126/science.8036517. URL: http://dx.doi.org/10.1126/science.8036517 (cit. on p. 10).

Witter, M. and E. Moser (2006). "Spatial representation and the architecture of the entorhinal cortex." In: *Trends in neurosciences* 29.12, pp. 671–678 (cit. on p. 1).

Wood, E., P. Dudchenko, R. Robitsek, and H. Eichenbaum (2000). "Hippocampal neurons encode information about different types of memory episodes occurring in the same location." In: *Neuron* 27.3, pp. 623–633. DOI: 10.1016/S0896-6273(00)00071-4. URL: http://dx.doi.org/10.1016/S0896-6273(00)00071-4 (cit. on pp. 9, 123).

Wu, L. and J. Dickman (2012). "Neural Correlates of a Magnetic Sense." In: *Science (New York, N.Y.)* DOI: 10.1126/science.1216567. URL: http://dx.doi.org/10.1126/science.1216567 (cit. on p. 41).

Xie, X., R. Hahnloser, and H. Seung (2002). "Double-ring network model of the head-direction system". In: *Physical Review E* 66 (4), p. 041902 (cit. on p. 70).

Yartsev, M. and N. Ulanovsky (2013). "Representation of three-dimensional space in the hippocampus of flying bats." In: *Science (New York, N.Y.)* 340.6130, pp. 367–372. DOI: 10.1126/science.1235338. URL: http://dx.doi.org/10.1126/science.1235338 (cit. on p. 41).

Yartsev, M., M. Witter, and N. Ulanovsky (2011). "Grid cells without theta oscillations in the entorhinal cortex of bats." In: *Nature* 479.7371, pp. 103–107. DOI: 10.1038/nature10583. URL: http://dx.doi.org/10.1038/nature10583 (cit. on pp. 15, 17, 42).

Ylinen, A., A. Bragin, Z. Nadasdy, G. Jando, I. Szabo, A. Sik, and G. Buzsaki (1995). "Sharp wave-associated high-frequency oscillation (200 Hz) in the intact hippocampus: network and intracellular mechanisms." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 15.1 Pt 1, pp. 30–46 (cit. on p. 10).

Yoder, R. and K. Pang (2005). "Involvement of GABAergic and cholinergic medial septal neurons in hippocampal theta rhythm." In: *Hippocampus* 15.3, pp. 381–392. DOI: 10.1002/hipo.20062. URL: http://dx.doi.org/10.1002/hipo.20062 (cit. on p. 95).

Yoganarasimha, D., G. Rao, and J. Knierim (2011). "Lateral entorhinal neurons are not spatially selective in cue-rich environments." In: *Hippocampus* 21.12, pp. 1363–1374. DOI: 10.1002/hipo.20839. URL: http://dx.doi.org/10.1002/hipo.20839 (cit. on p. 75).

Yoon, K., M. Buice, C. Barry, R. Hayman, N. Burgess, and I. Fiete (2013). "Specific evidence of low-dimensional continuous attractor dynamics in grid cells." In: *Nature neuroscience*. DOI: 10.1038/nn.3450. URL: http://dx.doi.org/10.1038/nn.3450 (cit. on pp. 23, 70, 76).

Yoshida, M., L. Giocomo, I. Boardman, and M. Hasselmo (2011). "Frequency of subthreshold oscillations at different membrane potential voltages in neurons at different anatomical positions on the dorsoventral axis in the rat medial entorhinal cortex." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31.35, pp. 12683–12694. DOI: 10.1523/JNEUROSCI.1654-11.2011. URL: http://dx.doi.org/10.1523/JNEUROSCI.1654-11.2011 (cit. on p. 14).

Yoshida, M., A. Jochems, and M. Hasselmo (2013). "Comparison of properties of medial entorhinal cortex layer II neurons in two anatomical dimensions with and without cholinergic activation." In: *PloS one* 8.9. DOI: 10.1371/journal.pone.0073904. URL: http://dx.doi.org/10.1371/journal.pone.0073904 (cit. on p. 14).

Zhang, K. (1996). "Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 16.6, pp. 2112–2126 (cit. on pp. 20, 49).

Zhang, S., J. Ye, C. Miao, A. Tsao, I. Cerniauskas, D. Ledergerber, M. Moser, and E. Moser (2013). "Optogenetic dissection of entorhinal-hippocampal functional connectivity." In: *Science (New York, N.Y.)* 340.6128. DOI: 10.1126/science.1232627. URL: http://dx.doi.org/10.1126/science.1232627 (cit. on p. 66).

Zilli, E. (2012). "Models of grid cell spatial firing published 2005-2011." In: *Frontiers in neural circuits* 6. DOI: 10.3389/fncir.2012.00016. URL: http://dx.doi.org/10.3389/fncir.2012.00016 (cit. on pp. 20, 126).

Zola-Morgan, S., L. R. Squire, and D. G. Amaral (Oct. 1986). "Human amnesia and the medial temporal region: enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus". In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 6 (10). PMID: 3760943, pp. 2950–2967. ISSN: 0270-6474. (Visited on 10/08/2011) (cit. on p. 67).

Zola-Morgan, S., L. R. Squire, and M. Mishkin (1982). "The neuroanatomy of amnesia: amygdala-hippocampus versus temporal stem". In: *Science* 218 (4579), p. 1337 (cit. on p. 67).

Zugaro, M., A. Arleo, A. Berthoz, and S. Wiener (2003). "Rapid spatial reorientation and head direction cells." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 23.8, pp. 3478–3482 (cit. on pp. 106, 107).

Acknowledgements

First and foremost I wish to thank my supervisor, Alessandro, for letting me join the group and for disclosing me the joys of the hippocampal formation, for having showed me where good science stands and most importantly for having encouraged me to undertake interesting science. I also thank him for his great hospitality and generosity, for letting me discover Israel and for cultivating with stories and anecdotes a shared interest in obscure, long-gone tribes populating remote corners of the earth in unclear times.

During these peripatetic years I had the pleasure of being welcomed by a number of (excellent) labs situated at very different latitudes. In Trondheim, I owe a huge debt to Yasser for his trust and his friendship and to Ben, for taking on his shoulders this collaboration. In Paris, I would like to thank Remi and Sophie for the hospitality they reserved me and for regularly taking me back to physics. Finally, I would like to express all my appreciation to the people at Weizmann, Nachum, Liora, Inbar, Alon, Maya, Gily for the great time I always have in Israel, and in particular to Arseny, for all the talking we had in rigorously warm climates; I would also like to include their bats, for having no theta.

I also thank Mathew, for having allowed me to take part to the life of his lab, to all the TPLL for the great science and the great food. My deepest gratitude goes to Natalia, for her dedication to the cause, for teaching me the correct way to shout at a rat in argentinian, and for having always believed in Number 33.

The good memories I have of these years in Trieste will always be tied to the wonderful group of people I found when I joined LIMBO. The pleasure of endless discussions, the intellectual richness and freedom, the mutual support and the diffused foolishness are some of the many things I could enjoy by being the "plus Fede" of these extraordinary girls. My wish to the the new and future members of the lab is to be able to create a similar aggregate of humanity and strong spirits.

It is difficult to express my affection for Valentina and Eleonora, especially as to start with one will have the other complaining for the order. But surely, in the last two years I've been missing Ele's bewilderment and sincerity. I thank her for her loyalty, understanding and most of all for being the partner in crime I could not substitute afterwards. I thank Vale for her inexhaustible source of topics and controversies, only comparable to

the inexhaustible number of fictitious aches she has. I also thank her for her support in difficult times and for her precious company that I learnt to enjoy through the folds of her wandering attention. I thank Athena for her passion, for her support of my cynicism and political un-correctness, for knowing what I was talking about; I thank Erika for the shared complaining, for fostering the polemic and for the good advises I haven't followed.

My time in Trieste would not have been the same without Claudia and her friendship. I thank her for her wisdom, generously handed-out, for her clear vision and for introducing me to an odd and funny piece of triestinian life. I then wish to thank Giulio, for being the closest thing to a male friend I had in many years and for being in Trieste every time I am not; and Olga, whose human qualities, kindness and intelligence, are only overshadow by her culinary abilities.

The list of all the people from the sector, from the rest of SISSA and from the real world, that helped me survive during the long afternoons in SISSA or offered some relief outside SISSA, would be too long. I want to thank you all for all the coffees, the aperitivi, the movies, the concerts, the dinners, the tennis, the climbings, the drinks, the early morning returns and so on, that you shared with me and that made my time in Trieste so fine.

I also renew my gratitude to my friends from previous lives, for tolerating my long absences and for not grounding our friendship on email replies.

The quantity of stress and of bad mood I have often afflicted my colleagues with, is nothing compared to the amount my parents had to sustain. I thank them for their support and affect, they have been fundamental throughout these years.

To my knowledge, I'm the first member of SISSA Neuroscience to survive the ill-famed "curse". Most of the merit of this success should go to Giorgia, to her infinite patience, strength, complicity and glad acceptance of being an odd couple.