

# Space, time and memory in the medial temporal lobe

Davide Spalla

A thesis presented for the degree of  
Doctor of Philosophy

Supervisors:  
Alessandro Treves  
Charlotte Boccara



Cognitive Neuroscience  
SISSA  
Academic year 2019-2020

# Abstract

This thesis focuses on memory and the representation of space in the medial temporal lobe, their interaction and their temporal structure.

Chapter 1 briefly introduces the topic, with emphasis on the open questions that the subsequent chapters aim to address.

Chapter 2 is dedicated to the issue of spatial memory in the medial entorhinal cortex. It investigates the possibility to store multiple independent maps in a recurrent network of grid cells, from a theoretical perspective. This work was conducted in collaboration with Remi Monasson, Alexis Dubreuil and Sophie Rosay and is published in [1].

Chapter 3 focuses on the problem of the dynamical update of the representation of space during navigation. It presents the results of the analysis of electrophysiological data, previously collected by Charlotte Boccara [2], investigating the encoding of self-movement signals (speed and angular velocity of the head) in the parahippocampal region of rats.

Chapter 4 addresses the problem of the temporal dynamics of memory retrieval, again from a computational point of view. A continuous attractor network model is presented, endowed with a mechanism that makes it able to retrieve continuous temporal sequences. The dynamical behaviour of the system is investigated with analytical calculations and numerical simulations, and the storage capacity for dynamical memories is computed.

Finally, chapter 5 discusses the meaning and the scope of the results presented, and highlights possible future directions.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Memory in the medial temporal lobe . . . . .	5
1.2	Space in the medial temporal lobe . . . . .	9
1.3	Cognitive maps and attractor neural networks . . . . .	11
1.4	Memory dynamics in MTL . . . . .	12
1.5	Open questions and thesis structure . . . . .	14
<b>2</b>	<b>Can grid cell ensembles represent multiple spaces?</b>	<b>18</b>
2.1	Two complementary network models . . . . .	22
2.1.1	Binary units . . . . .	23
2.1.2	Threshold-linear units . . . . .	23
2.2	Storage capacity . . . . .	25
2.2.1	Binary units . . . . .	26
2.2.2	Threshold-linear units . . . . .	29
2.2.3	Sparsity and noise reverberation . . . . .	29
2.3	Band solutions . . . . .	31
2.4	Discussion . . . . .	34
<b>3</b>	<b>Angular and linear speed cells in the parahippocampal circuits</b>	<b>37</b>
3.1	Angular velocity coding in the parahippocampal region . . .	39
3.2	Parahippocampal neurons upstream of the entorhinal cortex code for linear speed . . . . .	43
3.3	Conjunctive coding of primary (place, direction) and self- motion signals . . . . .	45
3.4	Self-motion coding is independent of theta modulation . . .	49
3.5	Discussion . . . . .	52

<b>4</b>	<b>Continuous attractors for dynamic memories</b>	<b>55</b>
4.1	The model . . . . .	57
4.1.1	The shift mechanism . . . . .	58
4.1.2	Storing multiple dynamic memories . . . . .	58
4.2	Dynamic retrieval . . . . .	59
4.3	Storage capacity for dynamic continuous attractors . . . . .	65
4.3.1	Analytical calculation of $\alpha_c$ in the highly diluted limit	67
4.3.2	Numerical estimation of $\alpha_c$ for a fully connected network	72
4.4	Discussion . . . . .	75
<b>5</b>	<b>Conclusions</b>	<b>77</b>
	<b>Appendices</b>	<b>81</b>
<b>A</b>	<b>Appendices to chapter 2</b>	<b>82</b>
A.1	Mean field equations: Binary Model . . . . .	82
A.2	Mean field equations: Threshold-linear Model . . . . .	83
A.3	Finite connectivity and noise reverberation . . . . .	87
A.4	Free-energy barriers in the binary model . . . . .	87
<b>B</b>	<b>Methods for Chapter 3</b>	<b>89</b>
B.1	Estimation of the behavioural correlates . . . . .	89
B.2	Firing rate calculation . . . . .	90
B.3	Speed filtering . . . . .	90
B.4	Rate maps and tuning curves . . . . .	90
B.5	Shuffling . . . . .	91
B.6	Measure used for cell type classification . . . . .	91
B.7	Estimation of the significance of overlaps between cell popu- lations . . . . .	92
B.8	Information analysis . . . . .	93
B.9	GLM analysis . . . . .	93
B.10	Tuning curve fitting . . . . .	94
<b>C</b>	<b>Supplementary figures for Chapter 3</b>	<b>95</b>
<b>D</b>	<b>Appendices to chapter 4</b>	<b>103</b>
D.1	Solution of the equation for the activity profile in the case of many maps . . . . .	103



# Chapter 1

## Introduction

The medial temporal lobe (MTL) is a system of brain structures specific to the mammalian brain. It originated from the reorganization of the medial pallium, the medial part of the upper surface of each cerebral hemisphere of therapsids, when mammals diverged from other early reptilian lineages, around three hundred million years ago [3]. The detachment of the most medial portion, and its loss of continuity with the rest of the cortex at the hippocampal sulcus, generated the hippocampus of modern-day mammals. A new population of granule cells, comprising the Dentate Gyrus, emerged between the Ammon's horn region, that retained the distinctly cortical pyramidal cells, and the surrounding neocortical regions [4]. These regions, the entorhinal, perirhinal and parahippocampal cortices, complete the structure of the medial temporal lobe.

Inputs are funneled to the hippocampus through the adjacent cortices, from virtually all neocortical association areas [6]. The Amigdala, as well as the cholinergic and other regulatory systems, provide additional input. Therefore, the medial temporal lobe acts as a hub for multimodal information coming from neocortical areas, and projects back into most of the areas from which it receives input.

Its recent appearance and its peculiar position in the cortical hierarchy suggest the involvement of MTL in advanced cognitive functions. MTL is indeed well known to play a role in the formation, consolidation and retrieval of memories, in the representation of time, space, and other contextual information and in many other cognitive functions.

In this thesis we investigate key aspects of the role of the medial temporal lobe in episodic and spatial memory, from a computational point of view. In this chapter we will briefly review the role of MTL in memory (section 1.1)

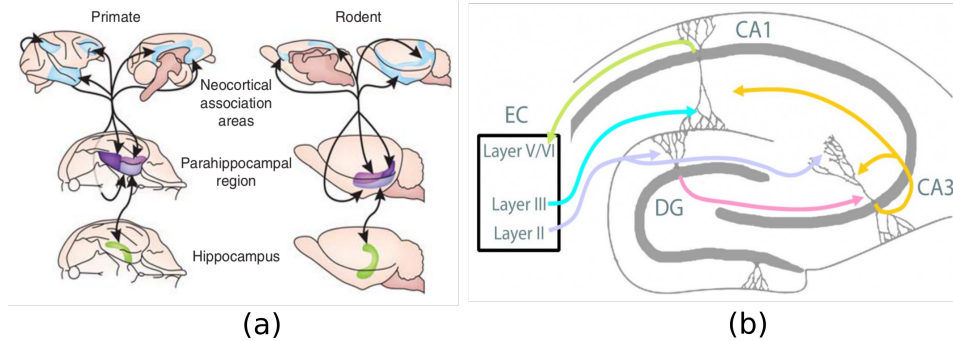


Figure 1.1: Anatomy of MTL. **(a)** The location of MTL and its connections with neocortical association areas in primates and rodents, adapted from Dickerson & Eichenbaum [5]. **(b)** Anatomy and connectivity of the hippocampus, adapted from Kamran Diba (<https://sites.lsa.umich.edu/dibablab/neural-circuits-of-the-hippocampus/>)

and space (section 1.2), some of the most influential quantitative theories of its function (section 1.3) and its dynamic nature (section 1.4). Finally, in section 1.5 we will summarize some open questions that motivate and introduce the original work presented in the rest of the thesis.

## 1.1 Memory in the medial temporal lobe

The involvement of the medial temporal lobe in memory has been known for more than a century. It was already in the last decades of the nineteenth century that the physiologists Sanger Brown and Edward Schafer observed that the ablation of MTL in monkeys resulted in memory deficits [7]. At the turn of the century, the Russian psychiatrist Vladimir Bekterev reported MTL softening in a dead patient who had shown a profound amnesia [5], and the work of neurologists and psychiatrists like Sergei Korsakoff and Alois Alzheimer showed the dramatic effects that neurodegenerative diseases involving MTL can have on cognition, and memory in particular.

In the second part of the last century, the role of the medial temporal lobe in memory became subject to extensive investigation. In a series of seminal studies on the case of their patient Henri Molaison (H.M.), the neurosurgeon William Scoville and the neuropsychologist Brenda Milner showed the striking effects that hippocampal removal had on specific components of memory [8],[9]. A first distinction was made between *declarative* memory, related to the explicit and conscious recall, and *implicit* (or *procedural*) memory, com-

prising, for example, motor learning and priming effects [10]. Subsequent neuropsychological research, as well as neuroimaging and lesion studies in animals, have shown that the hippocampus is crucial for the encoding of new declarative memories, while procedural memory is largely independent from MTL [11],[12].

More debated is the role of the hippocampus in memory retrieval, in particular of old memories. The nature and the extent of *retrograde amnesia* (RA) that follows hippocampal damage is still unclear. In some instances hippocampal-related RA is graded, with older memories being affected less than recent ones. This was, for example, the case of patient H.M., and is at the core of the theory of system consolidation [13], often referred to as the standard model of memory for its popularity. In this view, the hippocampus is crucial only for the retrieval of new memories. Over time, and through the interaction between the hippocampus and the neocortex, memories are transferred to the latter. These consolidated memories do not depend on the hippocampus for their subsequent retrieval (Fig.1.2a). On the other hand, there is evidence to suggest that an intact hippocampus is needed for the retrieval of memories rich in autobiographical and contextual information, no matter the age of the memory [14]. The so called “multiple trace theory” [15] aims to conceptualize this phenomenon. According to the theory, schematized in 1.2b, the hippocampus stores distinct aspects of each new memory, including contextual information, in largely non-overlapping cell assemblies. Neocortical memory traces of similar memories are instead largely, but not completely, overlapping. Hippocampal memories are highly episodic, rich in spatial and temporal details, and can be retrieved only with proper functionality of the hippocampus. Neocortical memories are instead semantic, and largely context free. Their formation, but not their retrieval, depends on the hippocampus.

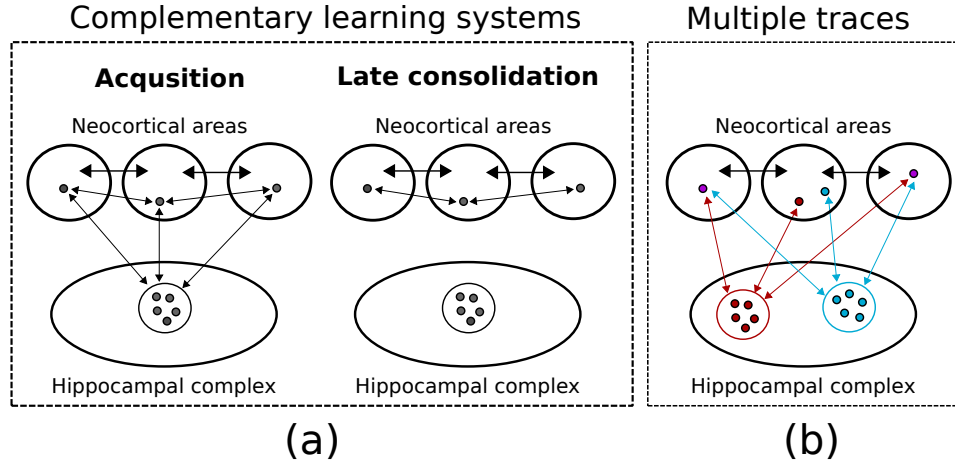


Figure 1.2: A schematic illustration of different theories on the role of hippocampus in memory, adapted from Nadel and Moscovitch [14]. **(a)** The standard system consolidation theory conjectures that the hippocampal complex is essential in the acquisition phase (left). After a consolidation period, in which information is consolidated in neocortex, the retrieval becomes independent of the hippocampus (right). **(b)** Multiple traces theory assumes that different memories are represented by largely orthogonal hippocampal assemblies, and partially overlapping neocortical populations. The hippocampus retains its role in the retrieval of memory rich in contextual information, independently of the age of the memories.

Despite their differences, these two qualitative theories agree on the crucial role that MTL has in the construction and – at least short term – storage of episodic memories. An attempt to quantify the memory capacity of the hippocampal region, and its information processing machinery dates back to David Marr [16], and was refined in the work of McNaughton and Morris [17] and Treves and Rolls [18]. The peculiar anatomical structure of the hippocampal circuit allowed a comparison of its different structures to different memory association circuits (Fig.1.3).

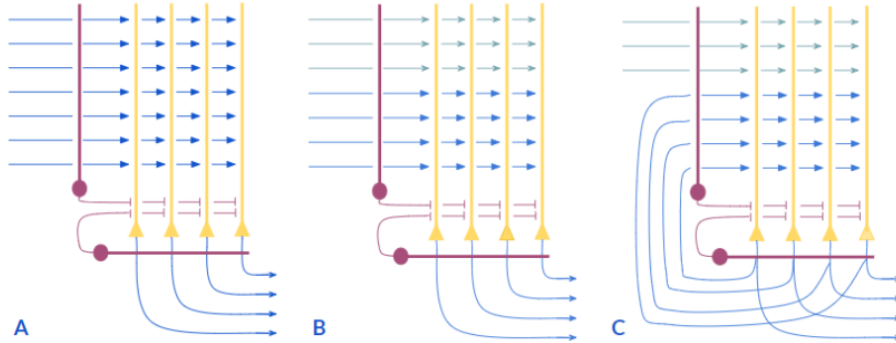


Figure 1.3: Different circuits for memory association **(a)** A simple autoassociative network, that performs pattern completion: when cued with a partial or corrupted version of a previously learned stimulus, reconstructs the full pattern. **(b)** An heteroassociator can build an association between a cue (green arrows) and a target (blue arrows) **(c)** A recurrent autoassociator can associate the input to a population to its output. These kind of network can self-sustain activity through reverberation in the recurrent connections. Figure courtesy of Oleksandra Soldatkina and Francesca Schönsberg, adapted from McNaughton et al. [17].

For example, the extensive recurrent connectivity of the CA3 region makes it suitable to work as an autoassociative memory able to retrieve stored patterns from partial cues provided by the entorhinal's perforant path inputs. The encoding of the memories is thought to be heavily dependent on the Dentate Gyrus and the Mossy Fibers that it projects onto CA3. The sparse and powerful input provided by the Dentate Gyrus would be needed in the encoding phase to separate similar inputs in non-overlapping memories.

A large class of neural network models, attractor neural networks, have been used to provide a quantification of the properties of this circuitry. We will examine some important features of these models in section 1.3, but first we will briefly review another fundamental function of the medial temporal lobe: the representation of space.

## 1.2 Space in the medial temporal lobe

Space is an important component of episodic memory, and its representation is a key role of MTL. Since the discovery of place cells by O'Keefe and Dostorovsky [19], many different functional classes of cells involved in the representation of space have been discovered in rodents (Fig 1.4). Place cells and grid cells [20] respond when the rat is in certain locations in the experimental environment (typically simple boxes, in the order of a square meter in size). Border cells [21] are active in the vicinity of boundaries or obstacles. The response of these cell classes is allocentric, i.e. is anchored to one or many places in the environment and not to the position of the animal relative to the locations they are encoding. Equally important for navigation is the representation of direction by Head Direction (HD) cells [22], that preferentially fire when the head of the animal faces a certain direction and are found in several brain areas, with particular abundance in parasubiculum and EC. These populations show a varied and complex phenomenology whose description extends beyond the purpose of this section and is a very active field of research.

The MTL is also involved in the representation of self-movement. The self-movement coding, theorized by attractor network models [23], was experimentally observed in a class of cells that respond to the speed at which the animal is moving [24]. Speed cells firing rate is proportional to the instantaneous speed of the animal, an interesting difference from the firing field coding that static representations of space and orientation adopt. In chapter 3, we report a similar phenomenology for the coding of angular head velocity in the parahippocampal region, as well as a more general form of speed coding than previously observed.

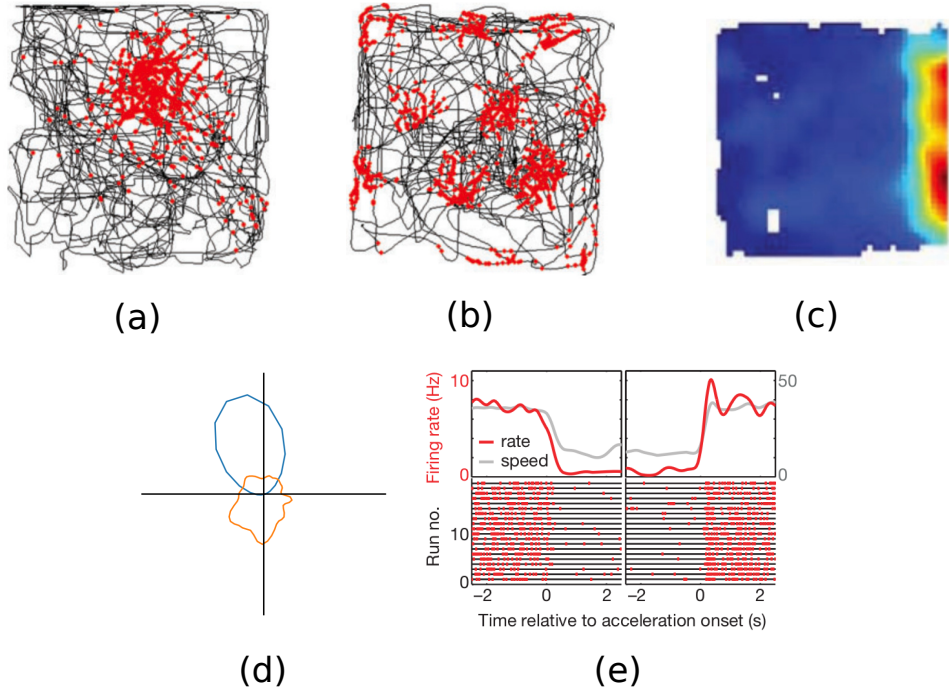


Figure 1.4: Spatially selective cells in MTL. **(a)** Place cells **(b)** Grid cells, adapted from Moser et al. [25] **(c)** Border cells, adapted from Solstad et al. [21] **(d)** Head direction cell (in blue the average firing rate, in orange the frequency of occurrence of a certain direction during the trial) **(e)** Speed cells, adapted from Kropff et al. [24].

MTL does not build a representation of space online, instantaneously during navigation, but relies heavily on spatial memory. The Morris water maze experiment [26] shows how damaging the hippocampus impairs the memorization of relevant locations. The memory property of the hippocampal place cells is apparent in the phenomenon called *remapping* [27]. In its most extreme version, global remapping, place cells acquire, change or lose their preferred firing location when the animal is in different environments. This change happens in a seemingly random fashion and independently from one cell to the other. When the animal is re-exposed to a previously seen environment, the place code is restored, showing a memory for space whose capacity is as large as experimentally probed [28].

Grid cells are believed to be very different in this respect. What is known of their response to different environments is summarized in chapter 2, where the theoretical possibility of global remapping in grid cell population is ad-

dressed.

The allocentric and memory properties of spatial coding led to the idea that MTL encodes space through *cognitive maps*. The central aspects of this theory, as well as the computational models it underlies, are the focus of the next section.

### 1.3 Cognitive maps and attractor neural networks

Cognitive maps were introduced by the psychologist Edward Tolman in 1948 [29], as a conceptual framework for the results of several experiments on the spatial behaviour of rats. Tolman hypothesized that the navigation of rodents – and humans – is based on spatial cognitive maps, i.e. global representations of the environment, rather than on simple associations. O’Keefe and Nadel [30] later interpreted the place cell system as the neural basis of cognitive maps. The cognitive map theory reconciles the memory-centric and space-centric narratives on the function of MTL, and offers a foothold for quantitative analysis.

A widely studied class of models, attractor neural networks (ANNs), are often used for the modeling of the neural mechanisms underlying cognitive maps. First introduced by Hopfield in 1982 [31], these networks can store many memories in their synaptic connectivity, and are able to perform cued retrieval. The stored memories are in fact attractive configurations for the network dynamics, and the configuration more similar to the initial – cued – configuration will be reached spontaneously. Amit, Gutfreund and Sompolinsky [32] developed the mathematical framework necessary for the study of these models and the quantification of properties such as the memory capacity or the stability of the attractive points. To describe spatial cognitive maps, ANNs have been extended to encode attractive low dimensional manifolds [33],[34],[35]. Ring attractors, with a connectivity that is short-range excitatory and long-range inhibitory have been used to describe populations of HD cells. A similar profile, in two dimensions, is at the core of the ANNs description of grid cells [36],[37].

Place cells require one step further: in order to encode several cognitive maps, and to account for global remapping, it is crucial for the system to be endowed with a reasonable storage capacity. The calculation of the storage capacity for network models of place cells was carried out by Battaglia & Treves [38] and Monasson & Rosay [39]; both analyses found a large capacity for continuous attractors, providing quantitative support to the idea that the hippocampus stores many independent spatial representations.



Evidence of neural dynamics compatible with continuous attractors has been recently found in grid cell and head direction cell populations [40],[41], whose collective activity has been shown to lie on a low dimensional subspace. Another hallmark of attractor dynamics, the abrupt switch between different stable configurations, has been investigated with a "teleportation" experiment [42]. The authors observed a switch in the neural representation of hippocampal place cells when the light conditions of the environment where suddenly changed to simulate an instantaneous change of room. Interestingly, around the "teleportation" time, CA3 place cells did not show a gradual transition between the two representations but a rapid, bi-stable flickering, as would be expected from the presence of two competing attractors.

It is worth noting that in these models cognitive maps are usually regarded as static. However, the activity of cells is in continuous motion on several temporal scales, both in active behaviour and during sleep. In the next section we focus on this dynamics, an important feature that we will incorporate in an attractor neural network model in chapter 4.

## 1.4 Memory dynamics in MTL

During both behavior and sleep, the activity of cell ensembles in the hippocampus and the surrounding areas is highly dynamic. One example is given by the phenomenon of phase precession, first reported by O'Keefe and Recce in 1993 [43]. During active exploration, the hippocampal formation of rodents shows a prominent modulation of its activity in the 8-12 Hz frequency range, called theta rhythm. As the rat runs across the receptive field of a place cell, the spiking of the cell progressively shifts towards earlier phases in the theta cycle. From the point of view of the active population, this produces a cycle-long swipe across cells with place fields around the location of the animal (see Fig. 1.5).

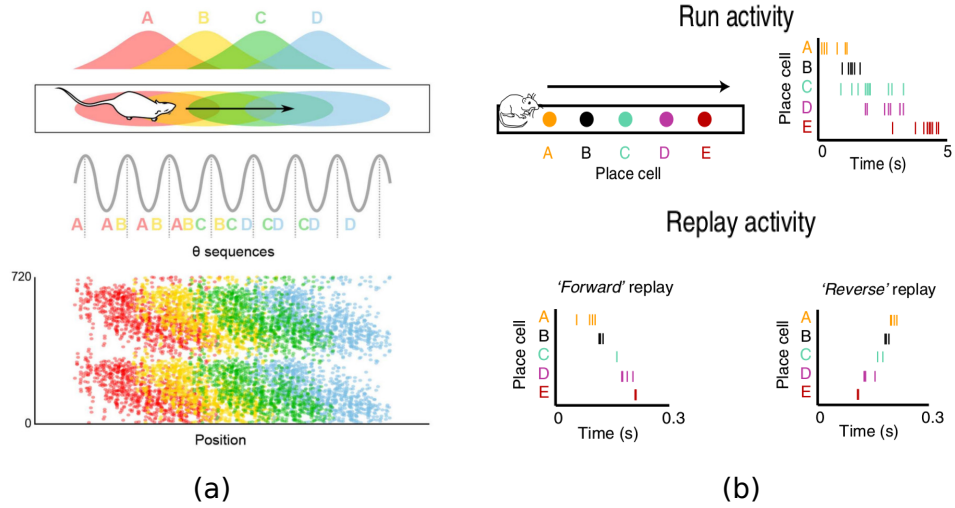


Figure 1.5: Dynamic sequences in the hippocampus **(a)** A scheme of the phase precession phenomenon, adapted from Drieu & Zugaro [44]. During a theta cycle, a sequence of place cells is activated, centered around the current location of the animal. **(b)** A scheme of hippocampal replay, adapted from Olafsdottir et al. [45]. Top row: place and time of activation of place cells during run activity. Bottom row: time compressed sequential activation of the same cells during sleep or restful wake, in the same (forward) or reverse (backward) order as in behaviour.

Sequential activation, on a shorter temporal scale, has also been observed during inactivity, in a phenomenon referred to as *replay*. In 1994, Wilson and McNaughton [46] observed reactivation of sequences of place cells in the rat hippocampus during slow wave sleep: cells with nearby firing fields showed a tendency to be reactivated together, producing a replay of recent trajectories. Neural replay takes place during sharp wave ripples, fast oscillations of the hippocampal local field potential that are particularly abundant during sleep and restful wakefulness [47],[48]. Indeed, replay has been observed not only during sleep [49],[50], but also during inter-trial rest periods [51],[52], and during still periods in navigational tasks [53],[54]. A temporally structured activation takes place also before the exposure to an environment [55], a phenomenon known as *preplay*, and a recent study showed that this dynamical feature emerges very early during development, preceding the appearance of theta rhythm [56].

Both phase precession and replay/preplay have been hypothesized to play several roles in cognition, from inference and planning to memory con-

solidation [44],[45]. Independently of their still debated function, phase precession and replay are a striking demonstration that neural activity is unwilling to sit still. Its motion is, however, disciplined: neural trajectories are constrained to lie on low dimensional manifolds in activity space, that guide sequential activation and are – at least to some degree – shaped by experience. In chapter 4 we address the problem of dynamic memory with an ANN model that goes beyond the standard equilibrium formulation, and is able to store and retrieve dynamical memories. In the next section we call the attention of the reader to some open questions that already appeared in this brief introduction, concerning the role of the medial temporal lobe in space and memory, and how and where we address them in the rest of this thesis.

## 1.5 Open questions and thesis structure

The rest of this thesis addresses several aspects of hippocampal and parahippocampal coding, with a mix of theoretical modeling and experimental investigation.

Chapter 2 tackles the problem of spatial memory in the medial entorhinal cortex (MEC). MEC grid cells are different from hippocampal place cells not only in the number of firing fields, but also in the way they remap in different environments. A seminal study on grid cell remapping [57] showed that the spatial relationship between pair of grid cells are preserved when the environment is changed. This, together with the discovery of the organization of grid cells in modules with different spatial scales [58] reinforced the idea that MEC encodes a single, low dimensional spatial map, whose principal role is position or distance encoding and path integration [23].

We summarize some recent results and theoretical arguments that challenge the “universal map” view: environment-dependent distortion, the coding of non-spatial features, different maps for different behavioural contexts and the problem of curvature.

The problem of multiple grid maps is addressed from a theoretical perspective: are there computational constraints that make it impossible for a population of grid cells to encode multiple maps? The storage capacity is found to be large enough to store tens, maybe hundreds of independent maps in a biologically realistic regime.

We then discuss the reason behind this high capacity, for which our models yield complementary insights, and the possible role of multiple maps in cognition.

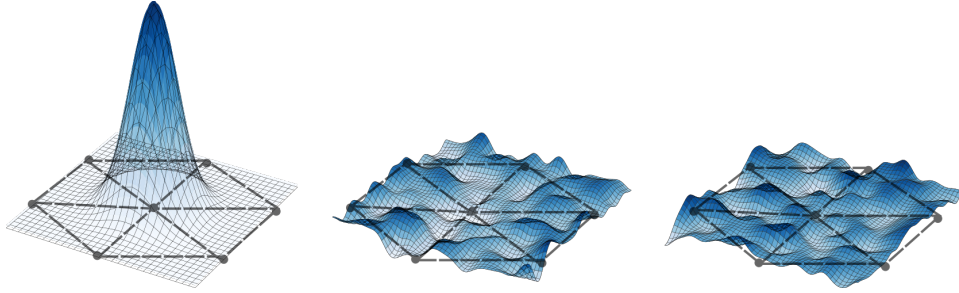


Figure 1.6: A pictorial illustration of the problem of multiple grid maps. The retrieval of one grid map (left) is represented. The activity, coherently organized in the first map, looks like random noise from the point of view of the other maps stored in the network (two in this case, on the center and right).

In chapter 3 we investigate how the representation of space built in the parahippocampal cortices can integrate self-movement, in order to update the represented position and direction.

As briefly mentioned in section 1.2, ANN models of the representation of space require the presence of self-movement signals in order to update the represented position and orientation during navigation, and in particular for performing path integration [23]. Speed cells in MEC provide this signal in the case of position, but in the case of orientation, no similar mechanism has yet been found in the parahippocampal cortices.

Analyzing electrophysiological recordings of single neurons from the medial entorhinal cortex, the presubiculum and the parasubiculum of freely moving rats, we find a population of cells whose firing activity is modulated by the instantaneous angular velocity of the head. We also show that linear speed cells are widely distributed in the parahippocampal circuit, with percentages in pre and parasubiculum similar to the one in MEC.

We then study the conjunction of this self-motion cells with grid and HD cells, and we find a rich phenomenology of conjunctive coding, compatible with the idea that different coding properties are assigned to different cell populations independently one from the other: they do not cluster and they do not segregate. This result challenges the idea of a rigid connectivity structure that many of the current models of spatial representation prescribe.

Finally, we show that the coding of speed and angular head velocity is, at the single cell level, independent from the theta rhythm. We discuss the meaning of these results in the framework of current theories and hypotheses

on the role of MTL in spatial coding.

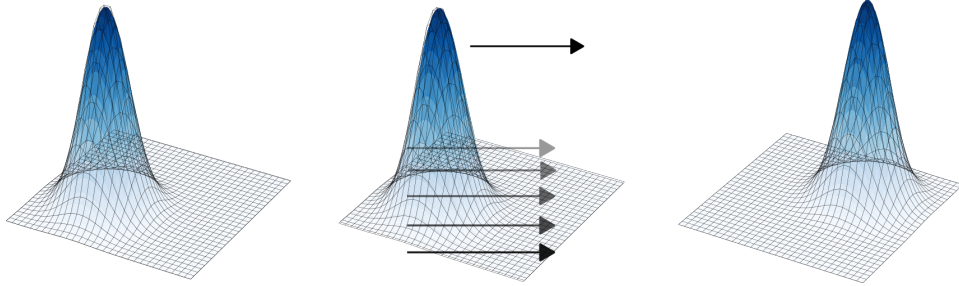


Figure 1.7: A pictorial illustration of the update of the represented position. The represented position at time  $t$  (left) shifts rightward at  $t + 1$  (right) by the action of an external speed signal (center).

In chapter 4 we address the problem of dynamical memories. The memory capacity of attractor neural networks, either discrete or continuous, is usually calculated with tools borrowed from equilibrium statistical physics, and therefore assume symmetric connections and static memories. Movement along a continuous attractor has instead been studied in the context of a single spatial map as a response to a velocity signal [33],[37],[36].

We develop an attractor network model suited to describe the dynamic retrieval of many memories, in a way reminiscent of hippocampal replay. We incorporate an asymmetric component in the synaptic connectivity, that produces a rigid shift along the retrieved manifold, and develop an analytical framework for the study of the features of this dynamical retrieval.

Analytical results and numerical simulations show that this mechanism is robust to the details of its implementation, and yields a storage capacity that is on the same order of the one for static attractors. For networks with diluted connectivity, the capacity decreases monotonically but gradually with the strength of the asymmetry. In densely connected networks, instead, moderate values of asymmetry can enhance the capacity, that shows a maximum for finite values of the asymmetry strength.

We discuss the implications of this work concerning the role of symmetry and asymmetry in synaptic connectivity, the role of temporal dynamics in memory and possible future direction in the use of out-of-equilibrium systems in the description of memory.

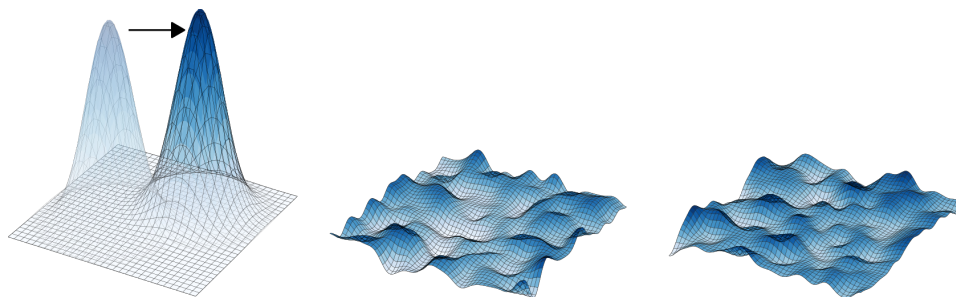
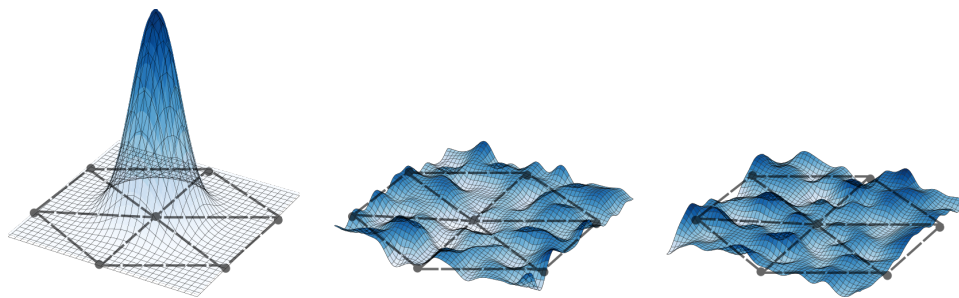


Figure 1.8: Pictorial representation of dynamic retrieval. The encoded position in the retrieved map (left) shifts spontaneously in time by the effect of asymmetric connectivity. Again, the population activity looks like noise from the point of view of two non-retrieved maps (center and right).

Finally, chapter 5 is dedicated to a general discussion of the results and models presented and their meaning in the context of the current state of neuroscientific research, and highlights open directions for future investigation.

## Chapter 2

# Can grid cell ensembles represent multiple spaces?



Grid cells appear to comprise an essential component of the cognitive representation of space in rodents [20] and in other species, e.g. bats [59]. In chapter 1 we mentioned how a study of the activity of grid cells in multiple environments [57] has shown that while the grid expressed by each neuron varies across environments in its spatial phase and orientation, between neurons the co-activity relations are largely preserved, at least for those recorded nearby in the tissue, with the same tetrode. In other words, the grids of different cells undergo a coherent rigid movement when a new environment is explored, as illustrated schematically in Fig.2.1 (a) and (b). The subsequent discovery of quasi-discrete modules [58] indicates that these relations are maintained at the local network level, presumably by recurrent collateral connections among grid cells. This finding has led to the hypothesis

that local ensembles of grid cells comprise each a single continuous attractor network, expressing a “universal”, two-dimensional map, which encodes the metric of space independently of the environmental context.

There is a crucial difference with the context-dependent spatial representations provided by hippocampal place cells, which display global remapping [60] even between very similar rooms, in particular in the CA3 field [28]: cells which were silent acquire one or more place fields, others lose theirs, and the fields that seem to have been maintained typically are in a different location (Fig.2.1B). Global remapping has motivated the conceptual model of multiple charts [34], in contrast with early and later models of continuous attractor grid cell networks, which envisage a single chart [37],[36],[61]. The dominant overall view, then, holds that the hippocampus encodes multiple, uncorrelated, context-dependent cognitive maps, while the grid system provides metric information that is independent of the environment.

Recent evidence of context-dependent distortions in the grid pattern have begun to question the view that the collective map expressed by a grid module is universal, that is, that it applies to any environment. Stensola et al. [62] have shown that, when rats explore large environments, a single grid can exhibit multiple orientations, likely due to anchoring effects to the closest wall, which in any case amount to distortions of the hexagonal pattern. These effects have been analyzed extensively in a more recent study [63]. Krupic et al. [64], [65] have shown that the grid pattern deviates from perfect hexagonality, with both global and local distortions, in response to environmental features such as the geometry of the walls. Finally, a couple of recent studies [66],[67] have shown that the presence of salient features such as goals or rewards affect the entorhinal map, changing field locations and inducing remapping in other space selective cells. These observations, moreover, refer solely to the position of the peaks of activity, i.e. the place fields of each cell, and do not take into account the fact that they vary reliably in height, independently across peaks, from one environment to the other [68]. Should we still regard grid cells as a sort of stack of millimeter paper, providing a universal metric for space?

In addition, recent studies conducted in both rodents and humans suggest that regular grids may not “measure” only physical space. Aronov and colleagues [69] find that both place cells and grid cells, in rats, are involved in the representation of a non-spatial but continuous, one-dimensional variable, such as the frequency of a sound. An fMRI study by Constantinescu et al. [70] shows an hexagonal modulation of the BOLD signal in human Entorhinal Cortex, and elsewhere, in a task that requires subjects to “navigate” the 2D space spanned by the varying leg and neck lengths of a drawing



of a bird. The representation of abstract or conceptual spaces, which could in principle be topologically and geometrically complex, would require of the grid cell system a flexibility that can hardly be reconciled with the universal grid hypothesis.

In a most interesting study [71], a subset of grid units were depolarized in transgenic mice, leading to what appears to be global remapping in the hippocampus. What is so striking is that the manipulation induces extensive changes, up and down, in the peak firing rates of the different fields of individual grid units, but not in their position. This elaborates the observation in [57], and suggests that what might be universal in the grid representation expressed by an ensemble of units, if anything, are the relative positions of the fields, whereas their peak firing rates are variable (Fig.2.1C).

On the other hand, a strict hexagonal periodicity of the field positions of individual units is only possible in flat 2D environments. The adaptation model of grid formation [72] predicts instead, on surfaces with constant positive or negative Gaussian curvature, and appropriate radius, the emergence of grids with e.g. pentagonal [73] or heptagonal [74] symmetry. In all other cases, including ecologically plausible natural environments, non-flat surfaces have varying curvature, making strictly periodic grids dubious, and rigid phase coherence most unlikely. But then, what happens to the universality of the grid in natural environments?

To address these issues we aim to answer a first fundamental question: is it at all possible to conceive of multiple, hence non-universal, ideal grid representations expressed in the same local network, when the animal is placed in *distinct*, even if flat, environments? In other words, would the storage capacity of a recurrent network of grid cells be above unity, so that multiple continuous attractors can coexist, encoded in the same synaptic efficacies? We pose this question within two alternative mathematical models, both accepting the idealized assumptions which underlie the universal map hypothesis, that is, of strict periodicity and equal peak rates, depicted in Fig.2.1D, but allowing for several uncorrelated grid representations. Under these assumptions, we analyze an ensemble of grid cells as a Continuous Attractor Neural Network, extending the frameworks developed in [38], [39] and [75] for the description of place cells. We emphasize that the storage capacity we are interested in quantifies the number of different, independent *charts* (or collective maps) that the network can store, and not the spatial resolution (which may be referred to as *information capacity*, i.e. the number of different positions that can be decoded from the ensemble activity), as considered for example in [76] and [77].

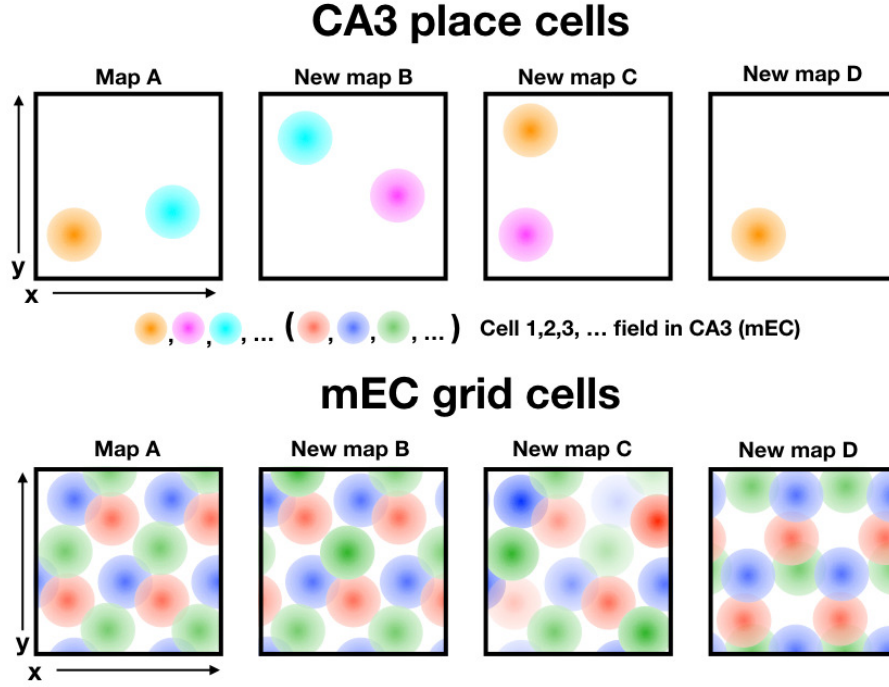


Figure 2.1: Types of change in grid cell activity in mEC (bottom) concurrent with global remapping in the CA3 field of the hippocampus (top). The universal grid map model, idealized from [57] allows only for a *coherent* translation (and possibly a rotation) into a new map B, when changing environment. Under a manipulation which does not entail changing environment, the individual fields of each unit have been observed to independently vary their peak rates, keeping their relative position ([71], new map C). We assess the hypothesis that the same network may also express other maps, such as map D, with a complete re-positioning of the grids of different units.

## 2.1 Two complementary network models

We model the grid cell population as an ensemble of units interacting through recurrent connections, whose structure defines which activity states are robust – the dynamical attractors. We assume, however, that a separate process, based *e.g.* on adaptation [72], has determined the emergence of a periodic grid, independently for each unit, during familiarization with each of  $p$  distinct environments; meanwhile, recurrent connections are shaped by a Hebbian learning process, such that neurons that happen to have nearby fields tend to fire together, strengthening their connections, while neurons with fields far apart remain weakly connected.

The connection strength  $J_{ij}$  is therefore taken to be a sum of contributions from the exploration of  $p$  environments, with each contribution, once averaged across many trajectories, a function of the relative position of the fields in that environment. Exploiting the simplifying assumption that each grid is strictly periodic, we can focus on the elementary repetitive tile, which has only one field per unit and is, in the mathematical formulation, connected by “periodic boundary conditions” to adjacent tiles. The assumption of periodic boundary conditions is motivated by the remarkable regularity of the arrangement of the fields observed in the original experiments, and by the model being meant to describe interactions within a grid module, in which all cells share the same spacing and orientation.

The contribution to the connection strength between two units  $i$  and  $j$  is then reduced to a function of their field centers  $\vec{x}_i^\pi$  and  $\vec{x}_j^\pi$  on the elementary tile in environment  $\pi$

$$J_{ij} = \sum_{\pi=1}^p K(\vec{x}_i^\pi, \vec{x}_j^\pi) \quad (2.1)$$

where we refer to  $K(\cdot)$  as the “interaction kernel”. The field peaks, or centers  $\vec{x}_i$  of  $N$  units are taken to be randomly and uniformly distributed over the elementary tile.

Our analysis focuses on two different models of neurons (binary and threshold-linear) and two types of attractor symmetry (square and hexagonal), which stem from the tile shape or the interaction kernel. Both neuron models allow, from complementary angles, a full statistical analysis, leading to otherwise inaccessible results. The storage capacity turns out to depend more on how interference reverberates through loops (expressed by the parameter  $\psi$ , see below) than on the type of units; and interference, in the densely coded and densely connected regime, affects square much more than hexagonal grids.

### 2.1.1 Binary units

The first model we consider is an extension of the model proposed by Monas-son & Rosay [39] for the modeling of place cells in CA3. Here the activity of neurons is described by binary variables, such that the pattern of activity of a network of  $N$  units is a vertex  $\{\sigma\} \in \{0, 1\}^N$ . For the binary model, the kernel  $K(i, j)$  between units  $i$  and  $j$  relative to one environment is taken to be a step function of the distance between their field centers

$$K(\cdot) = \frac{1}{N} \Theta(d_c - |\vec{x}_i - \vec{x}_j|) \quad (2.2)$$

where  $\Theta(x)=1$  for  $x > 0$  and 0 otherwise – note that the distance  $|\vec{x}_i - \vec{x}_j|$  is along the shortest path, considering the periodic boundary conditions.

The periodic structure of the attractor depends on the shape of the rhomboid unitary tile in which the field center  $\vec{x}_i$  of each unit is located. The lattice symmetry is specified by the angle  $\theta$  between its two primitive vectors.  $\theta = 60^\circ$  corresponds to the standard case of hexagonal grids, while  $\theta = 90^\circ$  describes a square grid pattern. These two cases and the resulting interaction kernel are depicted in Fig.2.2 (a) and (b).

The cut-off distance  $d_c$  sets the number of non-zero connections each unit receives from the storage of a given environment, denoted by  $wN$ :  $d_c = \sqrt{(w/\pi) \sin \theta}$ . This measure of connectivity within one environment should not be confused with the global connectivity taking into account all environments,  $C = (N-1)(1 - (1-w)^p) \sim N$  for large  $p$ .

The dynamics of the network is governed by the energy function:

$$E_J[\{\sigma\}] = - \sum_{i < j} J_{ij} \sigma_i \sigma_j \quad (2.3)$$

and constrained by the requirement that at any time a fixed fraction  $f$  of units be in the active state, i.e.  $\sum_i \sigma_i = fN$ . We call  $f$  the coding level, or sparsity of the representation. This constraint is taken to reflect some form of global inhibition. Later we shall focus only, given  $w$ , on the optimal coding level in terms of storage capacity, hence on a specific value  $f^*(w)$ , which turns out to be a monotonic function of  $w$  (see Fig.2.3). This model then allows an explicit focus on the dependence of the storage capacity on the width of the kernel and on the resulting optimal sparsity of the representation.

### 2.1.2 Threshold-linear units

We extend our analysis to firing-rate units, whose activity is described by a continuous positive value corresponding to their instantaneous firing rate.

This second model allows us to capture the graded nature of neural activity, which is salient when it represents space, which is itself continuous. The activity of the network is given by a configuration  $\{V_i\} \in (\mathbb{R}^+)^N$ , and each unit integrates the inputs it receives through a threshold-linear transfer function [78]

$$V_i = \begin{cases} g(h_i - h_0) & \text{if } h_i \geq h_0 \\ 0 & \text{if } h_i \leq h_0 \end{cases} \quad (2.4)$$

where  $g$  (the linear gain) and  $h_0$  (the activation threshold) are global parameters of the network, and the “local field”  $h_i$  is a real-valued variable summarizing the input influence on unit  $i$  from the rest of the network, which we take to come from a random but fixed set of  $C$  among the  $N - 1$  other units, as well as from external sources.

The interaction kernel  $K(\cdot)$  is given by the special sum-of-cosines form

$$K(\cdot) = \frac{1}{C} \sum_{l=1}^d (\cos[\phi^l(\vec{x}_i) - \phi^l(\vec{x}_j)] + 1) \quad (2.5)$$

which had been considered as a toy case by [38], before the discovery of grid cells. The field center of each unit on the elementary tile is expressed by a set of angles  $\phi^l(\vec{x})$ . We shall see that  $d = 2$  and  $d = 3$  are equally valid choices on the plane, as well as  $d = 1$ , which leads to “band” solutions (see below).

This model therefore allows decoupling the form of the kernel, which is extended, with interactions among units far away on the elementary tile (and the resulting coding level is correspondingly non sparse) from the connectivity, which can be made arbitrarily sparse if  $C/N \rightarrow 0$ . As a superposition of  $d$  cosine functions, the kernel can also be conveniently written as a sum of dot products. The  $+1$  term is added to enforce excitatory connections.

While not circularly symmetric like the radial kernel used in the binary model, this cosine kernel allows for the analytical study of periodic patterns that are spread out on the plane, with a large fraction of the units active at any given time. The solutions for the hexagonal kernel (Fig.2.2(d)), in particular, given by three cosine functions at a  $60^\circ$  angle from one another, have been considered as a reasonable model for experimentally observed grid cells. In the figure, the hexagonal elementary tile extends in the range  $x = \pm 1/2$  and  $y = \pm 1/\sqrt{3}$ , and the three angles span the directions  $\phi^1 = 2\pi x$ ,  $\phi^{2,3} = \pi(x \pm \sqrt{3}y)$ . The square kernel is obtained for  $d = 2$  and the two cosines at  $90^\circ$  from each other (Fig.2.2 (c)). Note that, as with the binary model,  $N$  units are concentrated on an elementary tile that in the hexagonal case is  $\sqrt{3}/2$  of the area of the square case.

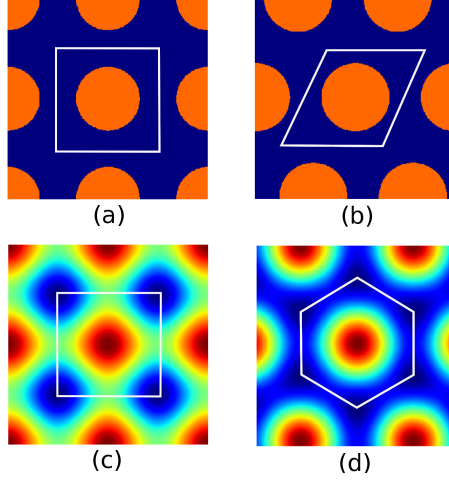


Figure 2.2: Interaction kernels for the binary (a,b) and rate (c,d) models. The white lines show the elementary tile of each lattice.

An energy function would look similar to the one in Eq. [2.3], but now expressed in terms of the continuous variables  $\{V\}$ . When  $C < N - 1$  and connections are not symmetric, the energy formalism does not apply but we can still analyze the model (see below and in appendix A.2), and again we take global inhibition, which can now also act through a modulation of the common gain  $g$ , to depend on the average activity of the network and to be such as to optimize storage capacity.

## 2.2 Storage capacity

Both models can store a single population map, as in the bottom panels of Fig.2.1A,B, and the equations for such a map admit periodic bump solutions that reproduce the shape of the kernel (as well as potentially other solutions, e.g. stripes, to be discussed later). We are interested however in their capacity to store several distinct maps, as in Fig.2.1A and D, and in the possibility to calculate such storage capacity analytically, in the mean field approximation.

The general strategy involves formulating and resolving a set of self consistent equations relating the activity of the units in the network. When the model admits an energy function, these are the saddle point equations derived from the computation of the “free energy” of the system with the

replica trick, which allows to take into account the statistics of the field centers in each environment. Without an energy function, e.g. when the connections are sparse and not symmetric, equivalent equations can be derived through the so-called Self Consistent Signal-to-Noise Analysis [79].

The solutions to these equations, that describe the activity in one map, disappear sharply at a critical value  $\alpha_c$  of the storage load  $\alpha = (p/C)$ , which measures the ratio of the number of maps to the number of connections to each unit.  $\alpha_c$  then gives the maximum number of maps that the network can store and retrieve or express, normalized by the connectivity. We have developed a novel method to assess whether below  $\alpha_c$  these solutions are indeed stable and prevail on others (Fig. 2.6 and 2.7). The details of these methods, that build on [39] and [75] for the binary model and on [78] and [38] for the rate model, can be found in appendix A.4.

We focus, in the calculation of the storage capacity, on so-called “bump states”, in which activity is localized along each of the two dimensions of the elementary tile (anywhere on the tile, given the translation invariance of the interaction kernel). Other solutions however exist, as discussed in section 2.3.

### 2.2.1 Binary units

The statistical analysis of the minima of the free energy leads to the patterns of activity  $\{\sigma\}$  that are likely to be observed given the connectivity. More precisely, we have derived self-consistent equations for the average activity  $\rho(\vec{x}) = \langle \sigma_i \rangle$  of unit  $i$  having its grid field centered in  $\vec{x}$  (in the elementary tile):

$$\rho(\vec{x}) = \int dz \frac{e^{-z^2/(2\alpha r)}}{\sqrt{2\pi\alpha r}} \Theta[\mu(\vec{x}) + z + \lambda] , \quad (2.6)$$

where

$$\mu(\vec{x}) = \int d\vec{y} K(\vec{x}, \vec{y}) \rho(\vec{y}) \quad (2.7)$$

is the signal input received by the unit through the interactions corresponding to the environment in which the bump is localized, say,  $\pi = 1$ , and  $z$  is the noisy, Gaussian input due to the interference from the other environments, say,  $\pi = 2, \dots, p$ , see Eq. (2.1). The variance  $\alpha r$  of these Gaussian inputs is, in turn, self consistently derived from the knowledge of the activity profile  $\rho$ , see appendix A.1. The uniform (inhibitory) input  $\lambda$  enforces the constraint  $\int d\vec{x} \rho(\vec{x}) = f$ . We consider here the limit case of neurons responding deterministically to their inputs, although the analysis extends naturally to stochastic noise.

We calculate, from the saddle point equations, the storage capacity  $\alpha_c(w, f)$  as the maximal value of the load  $\alpha$  for which a bump-like solution to Eq. [2.6] exists. Then, for a given value of  $w$ , we find the coding level  $f^*(w)$  that maximizes the storage capacity. Over a broad range  $0 \leq w \leq 0.5$  the optimal  $f^*$  turns out to be approximately half the value of  $w$  (see Fig.2.3).

That the optimal value for the coding level is proportional to  $w$  can be understood intuitively by considering the spatial profile of the signal  $\mu(\vec{x})$ : if too few cells are allowed to be active, the connections to the cells that are forced to be silent, within the connectivity range of the active cells, will be frustrated. On the other hand, if too many cells are active, those outside the connectivity range will contribute more to the noise than to the signal.

The optimal storage capacity is plotted in Fig.2.4, for the square and hexagonal grids as a function of  $w$ . At low  $w$  the two values are similar, but when  $w$  increases their trends diverge leading to substantially higher capacity value in the hexagonal case, of order  $10^{-2}$  for  $w \simeq 0.5$ . This value would definitely allow, in a real cortical network with order thousands (or tens of thousands) of neurons, the storage and retrieval of multiple independent grid maps.

Again considering the spatial profiles of the signal  $\mu(\vec{x})$  allows to gain intuition about this divergence. At very low  $w$ , i.e. short range interactions, what happens in other tiles can be neglected, and the two grids behave similarly. When the range is wider, the location of the fields in the immediately neighbouring tiles starts to be relevant. In the square case, there are four first neighbours, contributing to excite silent neurons in-between the fields. For an hexagonal arrangement of the fields, there are six neighbouring tiles that each contribute relatively less excitation in-between fields. Intuitively this last geometrical arrangement makes the structure more rigid and reduces the influence of the noise due to the storage of other charts.



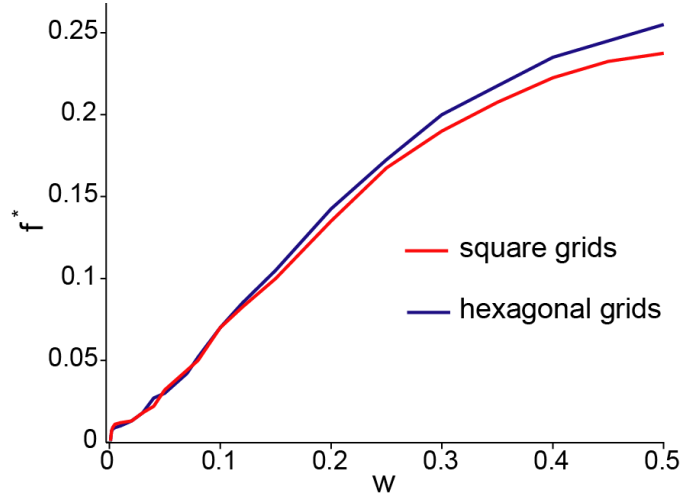


Figure 2.3: Optimal coding level for the binary model.

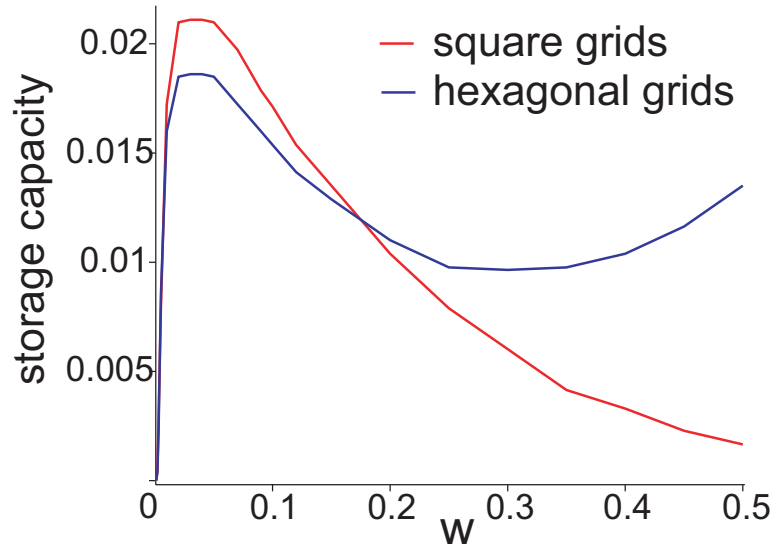


Figure 2.4: Storage capacity as a function of  $w$  for square and hexagonal grids in the binary model, given an optimal coding level  $f \simeq w/2$ .

### 2.2.2 Threshold-linear units

In this model the coding level and the connectivity range are both fixed by the shape of  $K(\cdot)$ . The mean field approach can be however extended to the case of arbitrary values of the connectivity density  $C/N$ , with the Self-Consistent Signal-to-Noise Analysis [79]. The storage capacity is given by the  $\alpha$  for which the solution to the equation

$$\bar{\mu}^2 - d \left[ 1 + \frac{C}{N} \left( \frac{(2-\psi)\psi}{(1-\psi)^2} \right) \right] \alpha r = 0 \quad (2.8)$$

disappears. In fact, the disappearance of the solution only gives an upper bound on  $\alpha_c$ , as one has to check its stability as well. The details of the derivation and the expression of the average signal  $\bar{\mu}$  and of the interference noise  $r$  are reported in appendix A.2.

We plot such critical value for square and hexagonal grids with the respective kernels, as a function of the inverse density  $N/C$ , in Fig.2.5 (full curves, blue and red). In the fully connected regime we find a result, corroborated also by computer simulations, similar to the one obtained with the binary model, with however a huge difference in capacity between square and hexagonal grids, and a value  $\sim 10^{-2}$  only for the latter. Moreover, it turns out that for the square kernel the stripe or band solutions of the next section are the global minima, and the square solutions are only marginally stable. In all cases the capacity increases as the connectivity density decreases, reaching an asymptotic value as  $N/C \rightarrow \infty$ .

The quantitative results for hexagonal grids has implications consistent with those of the binary model: it suggests that, again, a network of grid cells, for which a plausible number of synapses per neuron may be in the order of thousands, and with a connectivity, say, of order  $C/N \simeq 0.1$ , would have the capacity to encode perhaps a hundred different environments.

### 2.2.3 Sparsity and noise reverberation

The binary model shows that the difference in capacity between hexagonal and square grids results from the effective interactions among the fields in different tiles, as it emerges only with wide kernels and dense coding. When both are sparse, hexagonal and square grids are roughly equivalent. The  $w \rightarrow 0$  limit can be worked out analytically and  $\alpha_c \rightarrow 0$  in both cases, but only after having reached a maximum around  $\alpha_c \simeq 0.02$  for quite sparse codes,  $w \simeq 0.03$  and  $f \simeq 0.015$ .

Sparse coding is known to suppress noise reverberation (leading to small  $\psi$ ), but remarkably this relatively large capacity is approximately preserved

for hexagonal grids with dense coding,  $w \simeq 0.5$  and  $f \simeq 0.25$ , illustrating the efficiency with which this compact arrangement minimizes interference.

The threshold-linear model affords complementary insight, again on how the hexagonal/square capacity difference depends on the units active in each attractor reverberating their activity. Mathematically, this is expressed explicitly by the dependence of Eq.2.8 on the order parameter  $\psi$ , which quantifies the amount of reverberation through the loops in the networks. The physical meaning of  $\psi$  can be inferred from the expression derived in appendix A.2 and A.3:

$$\psi = g' \frac{T_0}{d} f. \quad (2.9)$$

The factor  $g'T_0/d$  is in fact the typical noise  $T_0/d$  amplified by the renormalized gain  $g'$  and multiplied by the average fraction of active units, the  $f$  parameter as in the binary model.  $\psi$  is then the one-step loop term in the reverberation of the noise; its effect on the capacity is illustrated by the dashed line in Fig.2.5, in which such contribution is factored out. For densely connected networks, storage capacity would massively increase and relative differences would decrease without noise reverberation.

The optimal capacity for the hexagonal kernel is then (mainly) the result of a reduced reverberation of the noise, due to the shape of the activity distribution of its attractors: the average fraction of active units ( $f \sim 0.46$ ) in the attractive state of the hexagonal kernel model is considerably lower than the same fraction in the square kernel, where it would be  $f \sim 0.79$  for the square grids, and is only somewhat reduced to  $f \sim 0.68$  for the stripes, which replace them as the stable solutions for this kernel.

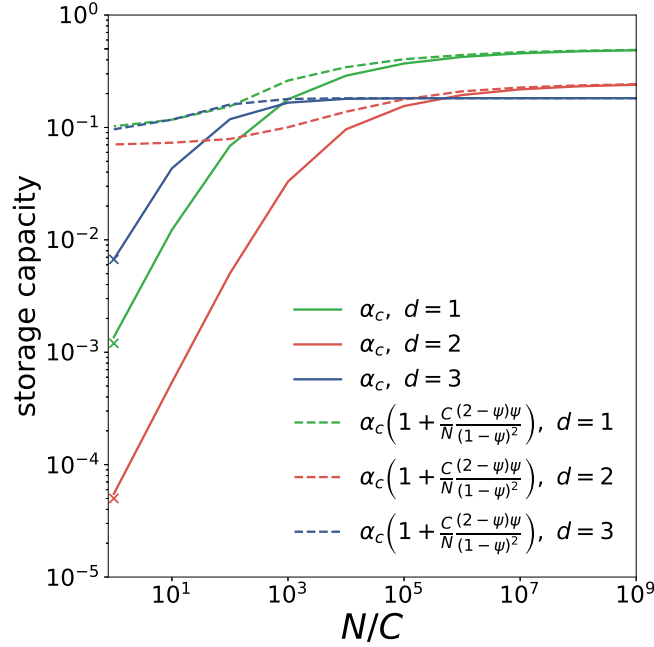


Figure 2.5: Storage capacity in the threshold-linear model as a function of the inverse connectivity density  $N/C$ , on a log-log scale. Full lines give  $\alpha_c$  for the three different interaction kernels (bands in green, square grids in red and hexagonal grids in blue). Dashed lines indicate what the capacity would be without noise reverberation. The crosses on the left show the capacity obtained with numerical simulations for a fully connected network.

## 2.3 Band solutions

In the previous analysis, we focused on “bump” states, in which activity is localized in a grid pattern. Another possibility are partially localized solutions: “band” states, where activity is localized along a single direction in the elementary tile, and extends along a stripe in the orthogonal direction.

In the binary model, these band states can be oriented along an edge of the tile (Fig.2.6(b,f)), or along the diagonal of the tile (Fig.2.6(c,g)), or in a discrete multiplicity of other orientations. Individual units “fire” along stripes of the same orientation, with relative offsets. We can study the properties of some of these band states in the  $w - f$  parameter space, to find that they are particularly favored in regions of high coding level.

Given the connectivity range set by  $w$ , bump states are the global minima of the free energy for low  $f$ , and one of the band states (which one depends on  $\theta$ ) becomes the minimum for higher  $f$ . For example, for both square and hexagonal grids, at small connectivity range  $w = 0.1$ , band states have lower free energy than the bump state for coding levels beyond 0.35, while for the larger connectivity range  $w = 0.5$ , this happens for coding levels beyond 0.4. This is intuitive, since for sufficiently large  $f$  a band state has a shorter boundary between active and quiescent units than a bump, and it is the length of the boundary that raises the free energy above its minimum value. Moreover, we can study how these different states are separated by computing the size of the free-energy barrier to cross to go from one state to another. The method to compute this barrier is sketched in Fig.2.7(c) and explained in more details in appendix D. In Fig.2.7(d) we show the size of the barriers to cross to go from a “bump” state to “band” states.

On the range of coding levels where these two kinds of states co-exist, the “bump” state is always more robust for an hexagonal grid compared to a square grid, as shown by the higher barrier size in an hexagonal grid (blue curve, from Bump to Band Edge or Band Diag. state) compared to square grid (full red curve, from Bump to Band Edge state).

A different behaviour is observed in the threshold-linear network. In this case, the rigid symmetry imposed by the 3-cosine interaction kernel makes the bump pattern a global minimum. In the 2-cosine case, instead, band state are stable solutions, corresponding to a macroscopic overlap with only one of the two cosines. We can describe bands also with a 1D interaction kernel, with a single cosine, and compare the storage capacity for band patterns with the one for square and hexagonal grids. In Fig.2.5, the green line shows the capacity for band patterns as a function of the connectivity.

For a densely connected network, it is above that for square grids, and the barrier methods indicates that these are only marginally stable to collapsing into stripes. This is in line with the reduction of the capacity from one to two dimensions shown in [38]. Interestingly, as soon as the emergence of a third cosine is allowed the capacity is instead enhanced, surpassing the 1D kernel except for very low values of connectivity density.

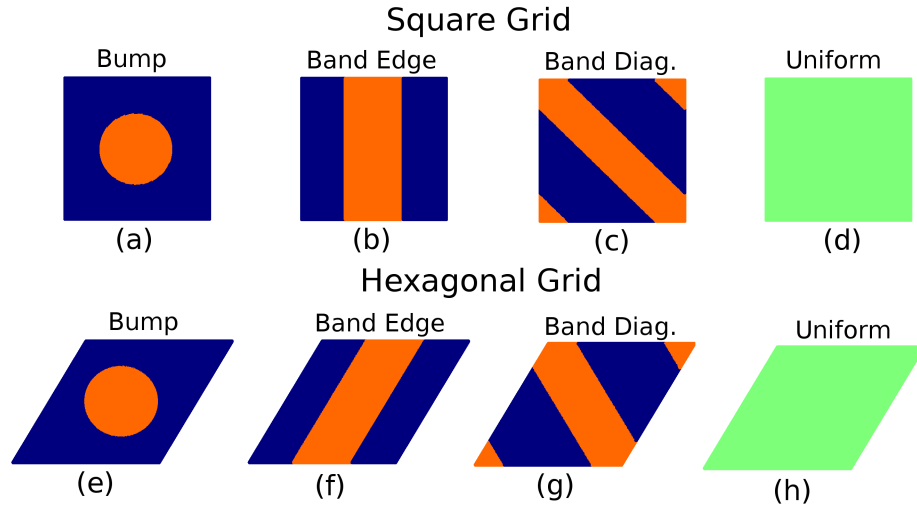


Figure 2.6: Different solutions to the saddle point equations in the binary model. Bumps (a,e) are stable at low  $f$  ( $f=0.2$  in the figure). Edge-oriented and diagonal bands are stable solutions for the  $\theta = 60^\circ$  tile at higher  $f$  (*e.g.*  $f=0.4$ , f,g), but only the former (b) are stable for  $\theta = 90^\circ$ : Uniform solutions (d,h) are always unstable below the critical capacity.

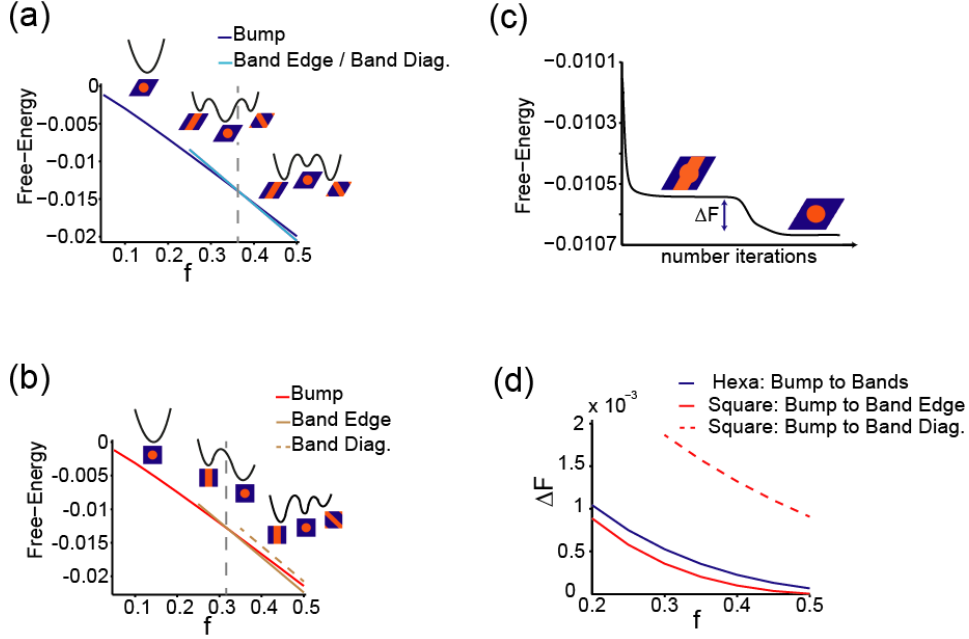


Figure 2.7: Bump and band states in the binary model. Free-energies of the bump and band states for hexagonal grids (a) and square grids (b). (c) Free-energy barriers are given by the difference in free-energies between an unstable mixed-state (band edge + bump shown here) and a metastable state (bump state shown here). (d) Size of the free-energy barriers to cross to go from the bump state to band states.  $w = 0.1$ ,  $\alpha \rightarrow 0$ .

## 2.4 Discussion

Our results indicate that, given appropriate conditions, a neural population with recurrent connectivity can effectively store and retrieve many hexagonally periodic continuous attractors. This possibility suggests that a regular grid code may not be restricted to represent only physical space; it could also express continuous abstract relations between arbitrary features, at least if they can be mapped to a two-dimensional space. This would however require a system flexible enough to store and retrieve uncorrelated grid representations. Our results show that this flexibility does not need, in principle, separate neural populations for separate representations, but can be achieved by a single local ensemble, provided it can learn effectively orthogonal representations.

Given the recent observation of non-spatial coding – a consistently tuned response to the “position” along a 1D non-spatial variable, sound frequency, during a sound manipulation task – by neurons that qualify as grid cells in a 2D spatial exploration task [69], it would be interesting to know whether a similar selectivity can be observed for a 2D non-spatial variable, as suggested by indirect observations of hexagonal modulation [70].

Several important questions are left open for future investigation. First of all, if global remapping is possible within a grid cell population, why has it not been observed experimentally? Possibly, a remapping capacity of grid cells may have been hidden by the fact that multiple mappings were only studied in simple, empty, flat environments — and then they turned out to be the same, modulo translations [57].

The hypothesis of a universal grid, that shifts without deformation across an environment and from one environment to the other, faces severe difficulties as soon as curvature is taken into consideration. In curved environments, rigid translations are not possible, and the geodesic transformations that partially substitute for them do not leave field-to-field relations unchanged, making a universal grid *a priori* impossible. Nevertheless, natural environments show a wide range of both positive and negative curvature, which does not seem to pose any problem to the navigational skills of rodents, or of other species. It is then conceivable that the apparent universality of the grid pattern comes from the experimental restriction to flat environments, which all belong to the same, rather special, class of two dimensional spaces with zero curvature, and that a richer grid behavior is required in order to code for position in more general spaces.

The emergence of grid representations in curved environments has been investigated with a model based on single cell adaptation [73][74], which illustrates the emergence of different regular patterns for distinct ranges of curvature. Estimating the storage capacity of recurrent networks expressing curved grids, however, poses some challenges. Since shifting the grid pattern along a curved surface moves individual fields by a different amount, the relationships between grid units cannot be reduced to the relationships between a single pair of their fields. Long-range translational coherence becomes impossible. Curved grids can be only partially coherent, and whether this partial coherence is sufficient to build stable attractors is an open problem [80].

A second open problem is the ability of a network encoding multiple charts to support path integration, since the noise introduced by other charts is likely to introduce discontinuities in the dynamics shifting the activity bump, impacting the accuracy of the integrator. It has recently been sug-



gested [81] that interactions between different grid modules (each encoding a single chart or coherent ensemble of maps) can enhance the robustness to noise during path integration. The possibility that this result generalizes to modules encoding multiple charts, and the analysis of the capacity deriving from interactions between modules, are beyond the scope of the present work, but deserve future investigation.

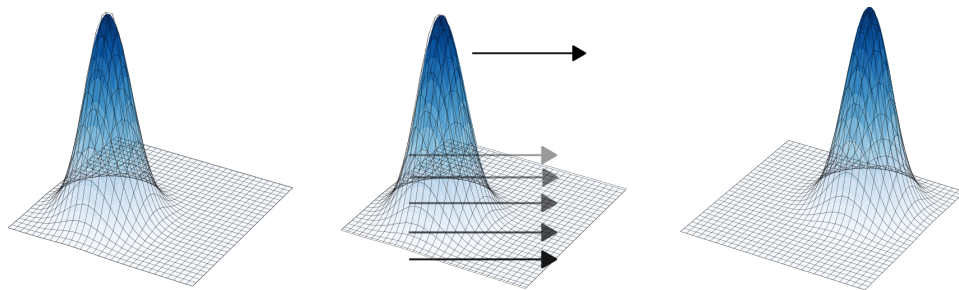
Finally, a third issue concerns the learning dynamics that sculpts the grid attractors. What is the mechanism that leads to the attractors of the recurrent network? Does a single grid dominate it, in the case of flat environments? Can *self-organization* be unleashed by the interplay between the neural populations of mEC, including non-grid units, and hippocampal place cells, aided by the dentate gyrus [82]? Including the hippocampus may be needed also to understand the distortion of the grid pattern, reported in several experimental studies [58][62][64], that by disrupting long-range order also weakens coherence.

At the system level, a finite storage capacity for the grid cell network implies the possibility that medial Entorhinal Cortex, or any other area in the brain [70] that is observed to include grid-like units, can serve context memory. This would turn upside down the widely shared notion that memory for the specific spatial features of each environment is only available downstream, in the hippocampus, and conceptually reunite medial Entorhinal Cortex with other regions of the mammalian temporal lobe, known to be dedicated to their own flavour of memory function [83].

Moreover, the possibility of multiple uncorrelated continuous attractors in flat environments, combined with the discovery of transitions between (highly correlated) states in which the grid is the same but the peak firing rate of each field is different [71], and with a new understanding of the disorder and frustration inherently associated to the grid representation of curved environment [80], puts to rest the rigid order which had appeared as the most salient character of the newly discovered grid cells. It suggests instead a sort of spin glass at intermediate temperature, i.e., that in order to code densely and efficiently for position on (many) continuous manifolds, grid cells have to be equipped with the flexibility and the ability to compromise characteristic of self-organized disordered system.

## Chapter 3

# Angular and linear speed cells in the parahippocampal circuits



As we briefly summarized in chapter 1, one of the roles of the hippocampal region seems to be the integration of multimodal information (coming from a variety of sensory and associative cortices, as well as deeper structures) to build a dynamic memory representation of an environment or an event [84] [85]. The accurate update of this representation, in response to changes in external stimuli or through the integration of self movements, is of great importance for successful navigation, as well as for other cognitive processes. In the context of spatial cognition, information integration is implemented in interconnected subareas of the hippocampal region through neurons coding for specific instantaneous navigational features such as position (place cells and grid cells) [19][20], direction (head direction cells) [86], and bound-

aries (border cells) [21]. The accurate updating of spatial representation is thought to depend crucially on self-motion signals and their integration with both positional and directional information [23],[87]. Despite their important role, where and how self-motion signals are integrated remains largely elusive.

The existence of neurons whose firing rate is modulated by running speed has been reported in the CA1 region of the hippocampus and in the medial entorhinal cortex (MEC) in conjunction with positional information or as an (apparently) self-standing code. [24][88]. In addition, speed has been reported to influence oscillatory activity recorded in the hippocampal field potential where the theta power seems correlated to locomotory activity [89],[90]. In contrast, angular velocity coding has not been established yet in principal neurons of the hippocampal region: most reports come from recordings of deeper structures (e.g. lateral mammillary nuclei, dorsal tegmental nucleus) linked to the processing of vestibular information [91],[92].

Nevertheless, many models of spatial cognition assume the existence of a continuous attractor network (CAN) in which neurons coding for instantaneous navigational correlates – such as position and direction – are connected by a population of cells conjunctively coding for position, direction and self-motion [23],[33],[36]. These conjunctive cells are hypothesised to mediate the shift of activity from position at time  $t$  to the next position at time  $t + 1$  (see Fig.3.1). They have been referred to as an "hidden layer", a term that hints at the fact that the existence of cells presenting such complex conjunction of coding is hypothetical. Support to the CAN models was recently provided by investigations of the *Drosophila melanogaster* central complex which exposed the existence of a ring of head direction cells whose activity is modulated by angular velocity [93],[94]. How much a rigid structure of this kind is common in species beyond arthropods, and how it could flexibly support high level cognitive behaviour, is however an open question.

To understand the circuit mechanism by which spatial representations can be updated in mammals, in this work we analyze the activity of 1436 principal neurons in all layers of three interconnected subareas of the parahippocampal region of rats (MEC: 396 cells; presubiculum: 605 cells; parasubiculum: 435 cells), recorded by Boccara et al. [2] in a previous study. The recordings were performed during ten minutes long free foraging sessions in open environments, and allow to study the effects of self-motion signals on the firing of individual neurons. We specifically investigate whether parahippocampal neurons could respond to both linear and angular self-motion signals. Our study reveals the existence of parahippocampal neu-

rons coding for speed and angular head velocity, as well as a rich structure of cells coding conjunctively for direction, position and self-motion. Considering that linear and angular speed are respectively the derivative in time of position and direction, we speculate that our results may uncover a more general algorithm for the updating of any type of information.

### 3.1 Angular velocity coding in the parahippocampal region

First, we ask whether parahippocampal neurons respond to an angular head velocity (AHV) signal, i.e. the derivative of head direction in time. To that end, we compute, for each cell, an angular velocity score as the Pearson product moment correlation between the instantaneous value of angular velocity and the firing rate of the cell across the recording session (appendix B, [24]). We define cells as AHV modulated when their score is greater than the 99th percentile of the shuffled distribution. This method leads us to classify a total of 246 cells as angular head velocity (AHV) cells, amounting to about one sixth of all parahippocampal cells (MEC: 16.9%; Prs: 17%; PaS: 17.2%, Fig. 3.1 and C.1). AHV modulation is uniformly distributed across all layers of each region (Fig. C.2), a part from the remarkable absence of AHV cells in MEC LII. The reason for this absence, that we find to be concomitant with the absence of head direction coding (see below), will be addressed in the discussion.

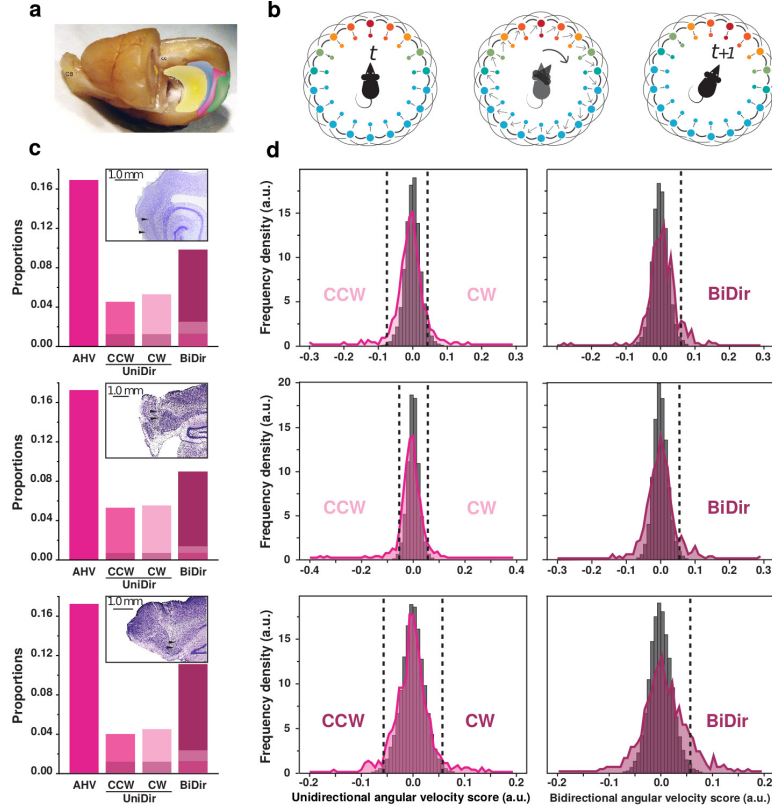


Figure 3.1: Angular velocity cell in the parahippocampal cortex  
**(a)** Whole rat brain, with partially removed left hemisphere to enable a mid-sagittal view of the right hemisphere and outlines of hippocampal formation (yellow), presubiculum (blue), parasubiculum (pink) and MEC (green). **(b)** Schematics ring attractor depicting the updating of head direction code from time  $t$  (left) to time  $t+1$  (right) following angular movement (middle). The outer layer of head direction (HD) cells is connected to a “hidden” inner layer of conjunctive HD-by-AHV cells. The colour represents neural activation from maximum (red) to minimum (blue). **(c)** Proportions of AHV cells within MEC (top), PaS (middle) and PrS (bottom). The shaded areas represent the intersection between AHV-CCW & AHV-BiDir (darker shade) or between AHV-CW & AHV-BiDir (lighter shade). Upper right corner boxes: representative Nissl stained sagittal section showing example recorded track for each area. **(d)** Distribution of unidirectional (left) and bidirectional (right) AHV scores across MEC (top), PaS (middle) and PrS (bottom) cell population comparing observed (coloured curve) and shuffled data (grey bars). Dashed lines represent 99 percentile thresholds for CCW- and CW-AHV (left) and Bidir-AHV (right).

As per our definition, AHV cells are neurons whose firing rate is positively modulated by angular velocity, meaning that these cells are more active when the animal is turning its head. About half of the AHV cells have their activity modulated solely when the animal had its head turning in one direction, either clockwise (CW) or counterclockwise (CCW). The firing rate CW (CCW) cells shows a positive (negative) correlation with the instantaneous value of angular head velocity (Fig.3.2 and 3.3a). The other half of the AHV cells are bidirectional (BiDir) and modulated by angular motion in both directions – their score was calculated by correlating their firing rate with the absolute value of the angular head velocity– (MEC: 26.8% CW, 31.0% CCW, 58.2% BiDir; Prs: 25.9% CW, 23.1% CCW, 64.4% BiDir; PaS: 32.0% CW, 30.0% CCW, 52.0% BiDir, Fig. 3.1 and C.1, note that the percentages do not sum up to 100%, since unidirectional and bidirectional populations can overlap). All layers of each region presented similar proportions of CW, CCW and bidirectional AHV cells (Fig. C.2). AHV modulation is stable in time across all regions and we observe no change in modulation intensity (AHV score) while comparing successive half-sessions (Pearson correlation  $\rho=0.52$ ,  $p < 0.001$ , Fig. 3.3b).

In order to further characterize how angular velocity is encoded in the parahippocampal region, we fit the rate-response to AHV tuning curve of modulated neurons with either a linear or a sigmoid function (see appendix B). The majority of AHV cells (68 %) are better described by a linear fit, the remaining third by a sigmoid fit (Fig.C.3). The steepness of the sigmoid is usually low (Fig. C.3): for most AHV cells the rate depends quasi-linearly on the angular head velocity value.

Given that our scoring method pre-suppose that a neuron firing rate responds quasi-linearly to AHV, we analyse all recorded cells with two additional scoring methods that do not depend on such assumption. These methods would allow us, for example, to flag cells responding to a specific speed band. The first additional score is inspired by a method commonly used to characterized spatial modulation [95]: the score of each cell is obtained by computing the information per spike that each neuron conveys about the running speed or AHV. The second additional method we use to qualify AHV modulation is a generalized linear model (GLM) approach to calculate a neuron firing profile as a function of velocity values [88]. We label cells as AHV-modulated when the fitted velocity function is a better predictor of the cell activity than a null model with constant firing profile (see appendix B and Fig. C.4). While these approaches reveal a fraction of AHV-modulated cells not captured by linear scoring method (and vice-versa), the cell populations yielded by the three methods are significantly overlapping

(Binomial test,  $p < 0.001$ , Fig. C.4). Most AHV cells solely picked up by the spatial information and GLM methods are anti-correlated with angular velocity – their activity is maximal when the animal is not turning its head. Nevertheless, given that both the fitting and the additional analyses show a dominance of linear (or quasi-linear) behaviour among the selected AHV cells, we will use the correlation method for all further analysis.

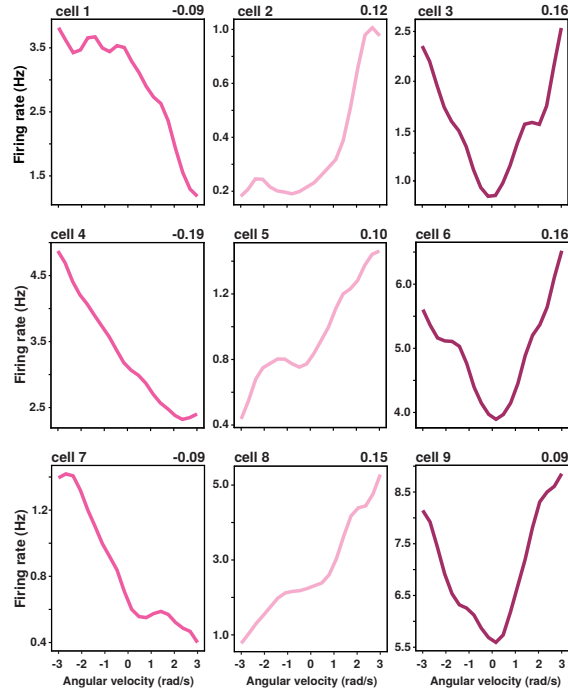


Figure 3.2: Example AHV cells in MEC (top), PaS (middle) and PrS (bottom) showing firing rate as a function of angular velocity (in rad/s), score in upper right corner. From left to right: CW-AHV (dark pink), CCW-AHV (light pink) and BiDir-AHV (purple).

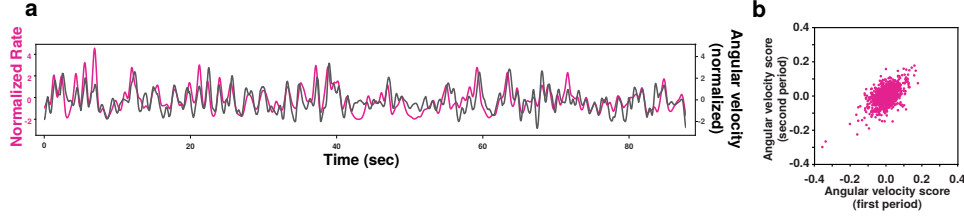


Figure 3.3: **(a)** Comparison between z-scored firing rate of an example AHV cell (pink curve) and instantaneous angular head velocity (black curve). **(b)** Correlation between the AHV score calculated in the first and second half of the recording session.

### 3.2 Parahippocampal neurons upstream of the entorhinal cortex code for linear speed

Once we established that angular velocity coding is widely spread across several parahippocampal area, we test whether linear speed coding also extends beyond the medial entorhinal cortex (MEC). To that end, we determine the speed score of each cell as a Pearson product moment correlation between the instantaneous value of rectilinear speed and the firing rate of the cells across the recording session, in a fashion similar to what we do in the case of angular head velocity and to what has been done in previous studies on speed cells [24]. With this method we classify a total of 277 speed cells. Our results show an uniform presence of speed cells in all the layers of the MEC, in similar proportions to what was previously reported [24] (MEC all: 16.7%, LII: 23.9%, LIII: 18.7%, LV: 12.7% and LVI: 13.7%; Fig 3.4, C.1 and C.2). In addition, we observe that rectilinear speed signals can be found upstream of the MEC, in about one fifth of both PrS and PaS cells (Prs: 20.6%; PaS: 19.8%, Fig. 3.4, 3.5 and C.1).



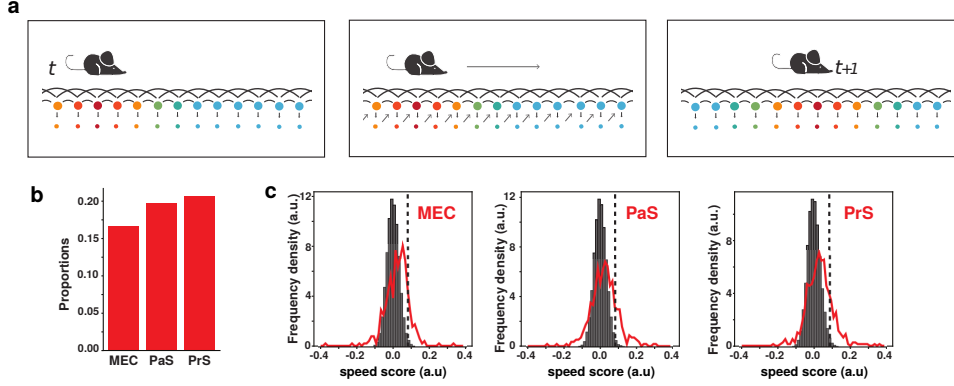


Figure 3.4: Speed cells proportions across the parahippocampal cortex. **(a)** Schematics of the update of the encoded position in a linear attractor, from position  $t$  (left) to position  $t+1$  (right) following movement (middle). The outer layer of conjunctive grid-by-HD cells is connected to a “hidden” inner layer of conjunctive grid-by-HD-by-speed cells. The colour represents neural activation from maximum (red) to minimum (blue). **(b)** Proportions of speed cells in MEC, PaS and PrS. **(c)** Distribution of speed scores in MEC, PaS and PrS. The observed (red line) and shuffle data (grey bars) are compared, dashed lines represent the 99-percentile threshold for each region.

Speed cells are uniformly distributed across all layers in each area (Fig C.2). As for AHV cells, speed cells are stable across time (Fig. 3.6).

We next explore whether all speed cells were responding similarly to increase in rectilinear motion regardless of their anatomical location. Given that speed is a derivative of position in time while AHV is a derivative of head-direction in time, we are particularly interested in modulation principles that could apply to both populations. To that end, we perform the same tuning curve fitting analyses previously performed on AHV cells. As for AHV neurons, we observe both linear and sigmoid tuning curves. Yet in contrast with the AHV population, the sigmoidal fit is slightly more predominant (56%). This result seems to be due to a saturation effect at high speeds, probably related to the low sampling in that speed band (Fig. C.3). Again, the low steepness of the observed sigmoids (Fig. C.3) allows to qualify the dependence of the firing rate on linear speed as close to linear.

Also in this case, we further analyze speed modulation with the same two methods (a spatial information score and a GLM approach) used for AHV cells. The population of speed cells we detect with these methods significantly overlaps with those obtained with our correlation scoring method

(binomial test,  $p$ -value  $< 0.001$ ; Fig. C.4).

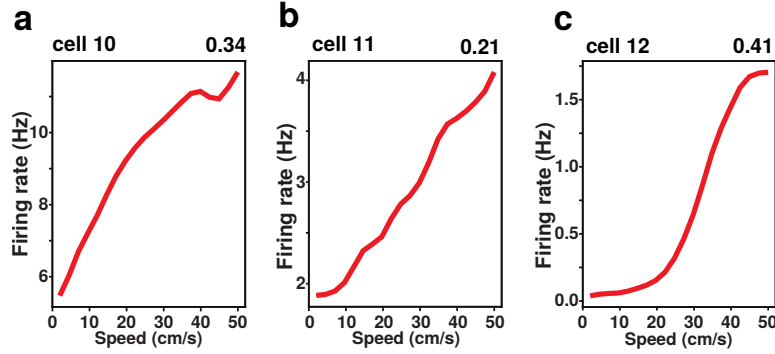


Figure 3.5: Examples of speed cells tuning curves in MEC (a), PaS (b) and PrS (c).

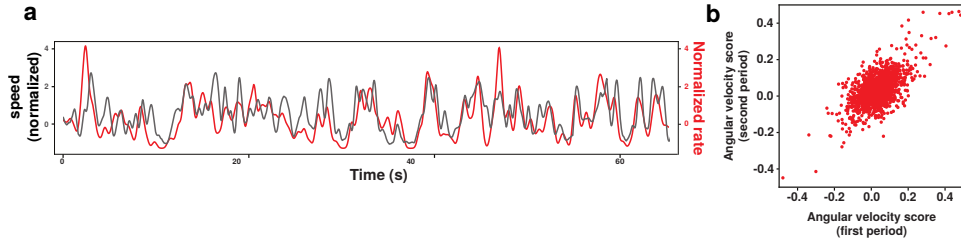


Figure 3.6: (a) Comparison between z-scored firing rate of an example speed cell (red curve) and instantaneous linear speed (black curve). (b) Correlation between the speed score calculated in the first and second half of the recording session.

### 3.3 Conjunctive coding of primary (place, direction) and self-motion signals

We proceed to investigate to which degree self-motion signals are co-existing in conjunction with other types of coding at the unit level: this interaction has an important role in many theories of spatial cognition, and in continuous attractor models in particular. We compute the grid and head direction (HD) scores of each recorded unit, and label as significantly modulated cells whose score exceed the 99th percentile of the score distribution calculated on shuffled data (appendix B). With this analysis we find that the majority

(80.4%) of AHV cells code for at least one other feature (HD: 55.2%; grid: 13.4%; grid-by-HD: 5.6%; rectilinear speed: 33.7%, Fig.3.7). These percentages are similar to the percentages observed in the general population (HD: 53.3%; grid: 17.8%; grid-by-HD: 7.7%, rectilinear speed: 19.2%).

A similar distribution of conjunctive coding is observed among speed cells (HD: 35.4%; grid: 18.7%; grid-by-HD: 3.2%; AHV: 29.9%, Fig. 3.8). We observe all possible types of conjunction of code including AHV-by-HD (the hidden layer of directional CAN models) and grid-by-HD-by-speed (the hidden layers of positional CAN models). We do not observe any significant difference in the distribution of conjunctive coding across layers of each area, a part from MEC LII, in which both HD cells and AHV cells are absent (Fig. C.2).

The scores (grid, HD, AHV and linear speed score) are independent from each other and we do not observe any significant correlation between them, with the exception of a small correlation between speed and bidirectional AHV scores (Pearson  $r = 0.29$ ,  $p\text{-value} < 0.001$ ) and a small anti-correlation between speed and HD scores (Pearson  $r = -0.14$ ,  $p\text{-value} < 0.001$ ).

It is important to observe that the observed distribution of mixed selectivity is compatible with a simple hypothesis of independent assignment of each of the coding properties in the general population: cells coding for different behavioural features neither segregate, nor cluster together. This is very different from what would be expected by models requiring a precise wiring between different sub-populations. The implications of this fact will be discussed below.

In order to grasp whether self-motion properties (i.e. speed, AHV) are encoded in a similar fashion to instantaneous properties (i.e. position, direction), we compare the firing properties of each class of neurons. We observe that cells coding for self-motion signals exhibit higher average firing rate than cells coding for position or direction (t-test,  $p\text{-value} < 0.001$ , Fig. C.6). They also show a shorter average inter-spike interval (t-test,  $p\text{-value} < 0.001$ , Fig. C.6) and a larger peak firing (defined as the fifth quintile of the rate distribution, t-test,  $p\text{-value} < 0.001$ , Fig. C.6). These differences could be explained by the fact that the monotonic firing profile used to encode motion signals is less sparse than the receptive field coding of grid and HD cells which are largely silent outside their firing field.

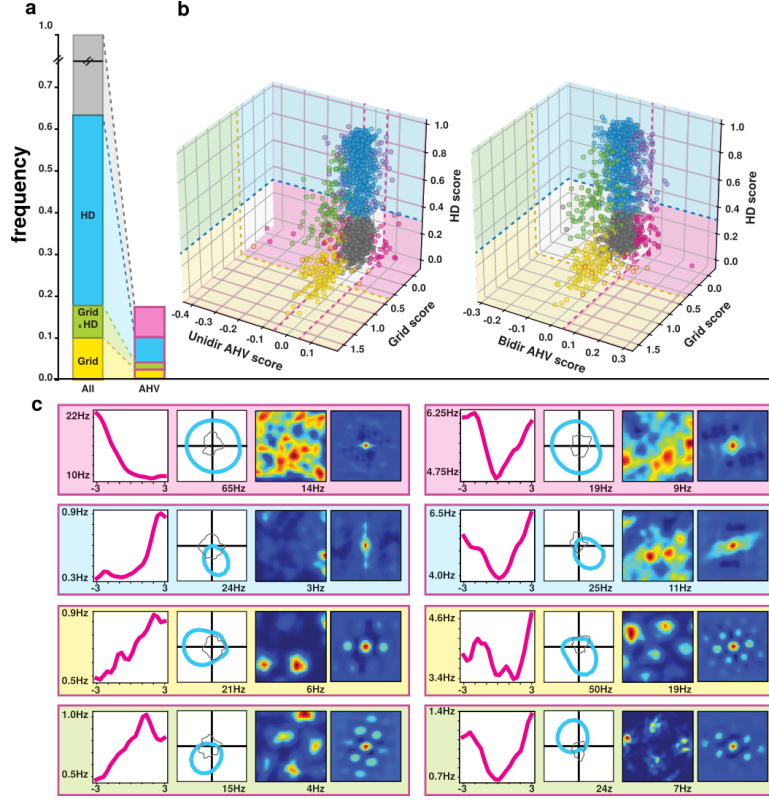


Figure 3.7: Conjunctive coding between HD, grid and AHV cells. **(a)** Intersection between angular head velocity (AHV), grid and head direction (HD) cells. The bar on the left shows the proportions of cells qualifying as grid (yellow), HD (blue) and grid x HD (green) cells in the total population. In grey are cells neither coding for HD nor grid. The bar on the right (pink contour) shows the same distribution among cells qualifying as AHV cells. **(b)** Scatter plots showing the intersection between grid, HD and AHV scores. Plot on the left: unidirectional – UniDir – AHV: i.e. CCW-AHV and CW-AHV. Plot on the right: bidirectional – BiDir – AHV. The colour code is the same as in (a). A pink contour denotes a modulation by AHV. Dotted lines represent region-averaged classification thresholds, computed to guide the visualization. **(c)** Examples of the four different kinds of AHV modulated cells, colour coded as in (a). From left to right: AHV tuning curve, HD polar plot, spatial firing rate map, spatial autocorrelogram.

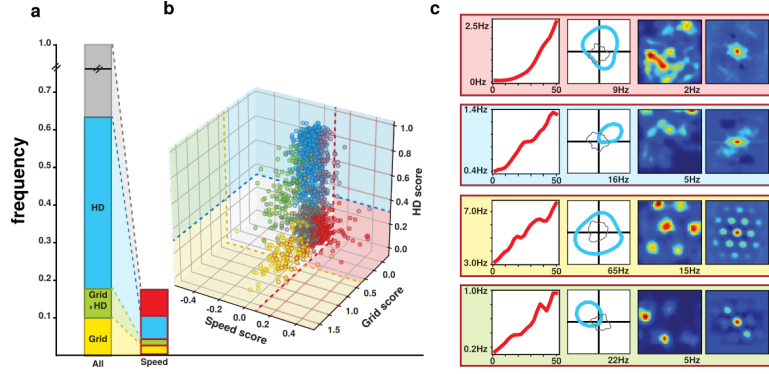


Figure 3.8: Conjunctive coding between HD, grid and speed cells. **(a)** Same as in Fig 3.7, for speed cells. The red contour bar represents percentages of principal cells within the speed cell population. **(b)** Scatter plots showing the intersection between grid, HD and AHV scores. A red contour denotes a modulation by speed. **(c)** Examples of the four different kinds of speed modulated cells, colour coded as in (b).

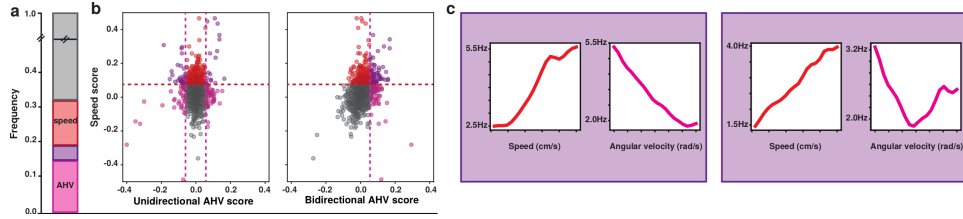


Figure 3.9: Intersection between angular head velocity and speed cells **(a)** Percentages of cells in the whole population. Pink: cells coding only for AHV; Red: cells coding only for speed; Purple: cells coding for speed and AHV ; Grey: cells neither coding for AHV nor speed. **(b)** Scatter plots of AHV and speed scores in the population. Plot on the left: unidirectional – UniDir – AHV: i.e CCW-AHV and CW-AHV. Plot on the right: bidirectional – BiDir – AHV. Colour code as in (a). Dotted lines represent region-averaged classification thresholds. **(c)** Examples of two kinds of conjunctive AHV X speed cells. Plots show the firing rate as a function of the angular velocity (pink) or speed (red). Left panel: speed x UniDir-AHV cell. Right panel: speed x BiDir-AHV cell.

### 3.4 Self-motion coding is independent of theta modulation

The theta rhythm of the local field potential has historically been strongly associated to running speed, and plays a crucial role in many theory of hippocampal function and models of spatial cognition. We therefore investigate whether AHV and speed cells have their activity modulated by theta rhythm. Following previously published work [2], we define a cell as theta-modulated when the mean spectral power around the peak in the 5-11 Hz range is at least fivefold greater than the average spectral power in the 0-125 Hz range (see appendix B). We observe that only around 40% of the AHV and the speed cells pass these criteria for theta modulation, while many of the remaining cells are not showing any modulation by theta (Fig. 3.10). The proportion of theta modulated cells within AHV/speed cells is comparable to the proportion observed in the general population (Fig. 3.12).

There is no significant correlation between AHV score and theta score, or between speed score and theta score (Fig.3.11, Pearson correlation, p-value  $> 0.05$ ). Conjunctive coding for grid or head direction does not influence the proportions of velocity cells that are theta modulated (Fig. 3.12). Theta modulation is uniformly distributed across all layers of each area except for MEC LII, that shows more theta modulation, and MEC LVI, that shows less (Fig. C.7, t-test, pval  $< 0.01$ ). The proportion of velocity cells theta modulated in each layer does not differ from what is expected based on the theta modulation observed in the general population except from MEC LVI speed cells, that showed less theta modulation than expected (Fig. C.7, t-test, pval  $< 0.05$ ). Together, these results suggest that the code for self-motion in the parahippocampal region is largely independent of theta modulation at the single cell level.

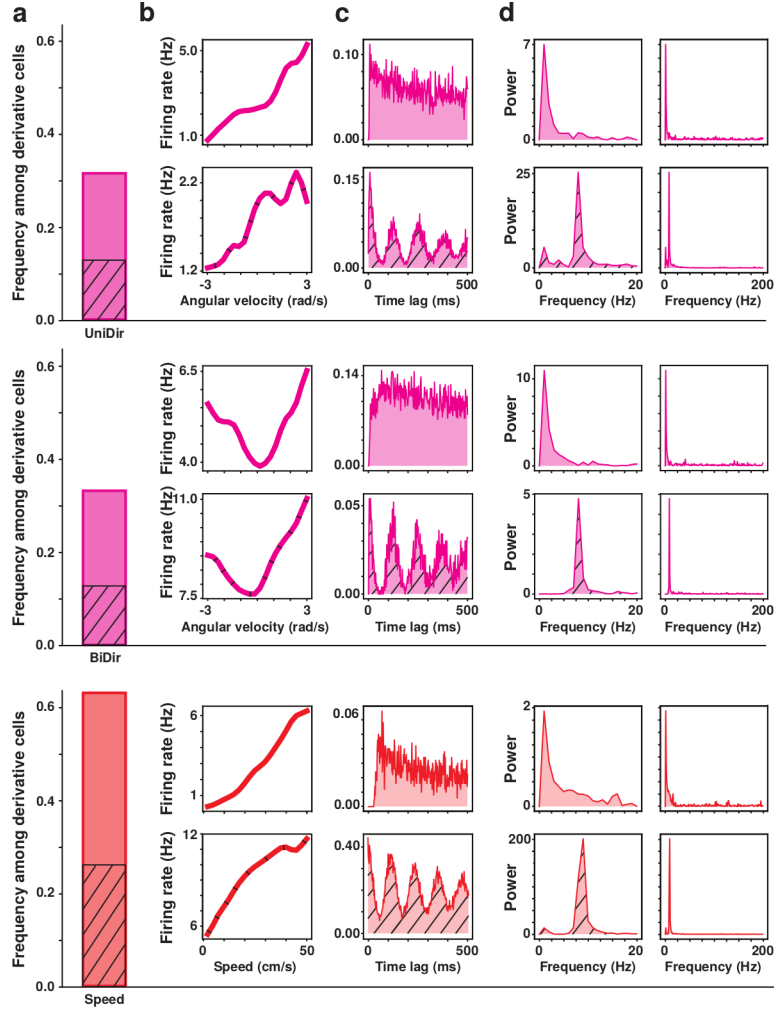


Figure 3.10: Speed and AHV coding is independent of theta **(a)** Frequency of unidirectional AHV cells (first row), bidirectional AHV cells (second row) and speed cells (third row) within the population of derivative cells (speed + all AHV). Dashed bars represent the fraction of theta modulated cells. **(b-c-d)** Examples of theta modulated (bottom row of each section) and non-modulated (top row each section) derivative cells. **(b)** Tuning curves: firing rate as a function of the angular velocity or speed. **(c)** Time autocorrelogram of the firing rate of the cell. **(d)** Power spectrum. First column: power in the range 0-20 Hz; second column: full spectrum (0-200 Hz)

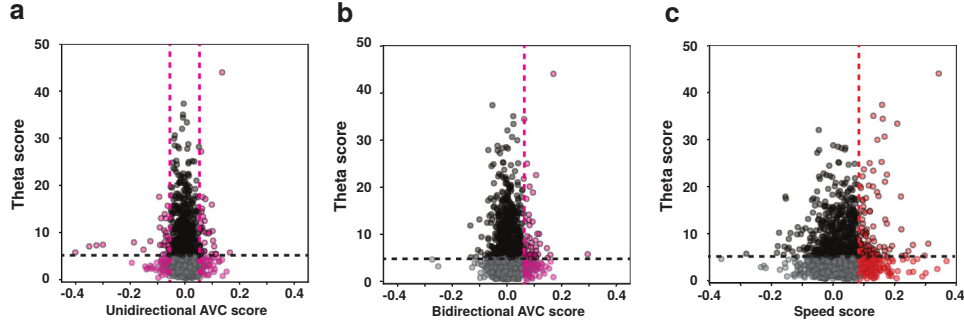


Figure 3.11: Speed/AHV scores and theta scores are uncorrelated. Scatter plots of theta scores versus unidirectional AHV (a), bidirectional AHV (b) and speed (c) scores. Cells are color coded according to their label. Black: theta modulated; red/pink: speed/AHV modulated; grey: not modulated. Dashed lines represent classification thresholds.

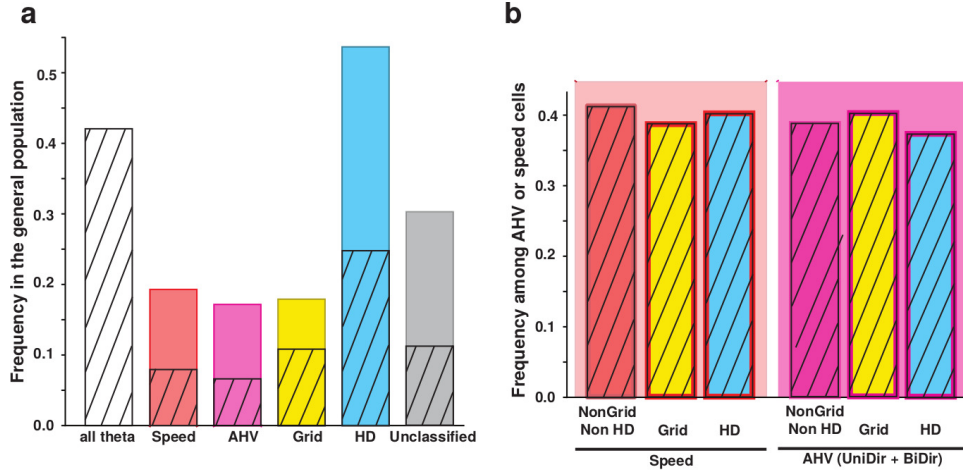


Figure 3.12: Proportion of theta modulated cells in different populations. (a) Distribution of speed cells (red), AHV cells (pink), grid cells (yellow), HD cells (blue) and unclassified cells (grey) in the whole population. Dashed bars represent the proportion of theta modulated cells. (b) Frequency of theta modulated cells among speed cells (left, red background) and AHV cells (right, pink background). Cells are divided by type: “pure speed” (red), “pure AHV” (pink), conjunctive grid x speed (yellow left), conjunctive grid x AHV (yellow right), conjunctive HD x speed (blue left) and conjunctive HD x AHV (blue right).



### 3.5 Discussion

In this study we revealed the existence of parahippocampal principal neurons whose activity is modulated by both angular and linear self-motion signals. Extensive mapping of the parahippocampal region showed that this population is spread homogeneously across all layers of several interconnected areas upstream of the hippocampal formation: the medial entorhinal cortex (MEC), the presubiculum (PrS) and the parasubiculum (PaS). We observe that some neurons modulated by self-motion signals seemed to only respond to either angular head velocity or linear speed. Yet, the majority of the recorded neurons modulated by self-motion signals are also concomitantly responding to spatial or directional information. Such integration at the unit level may be a crucial mechanism underlying the generation and the updating of the representation of position (place and grid cells) and direction (head direction cells) in the hippocampal/parahippocampal circuits.

We show that both angular and linear self-motion signals are encoded in a different manner with respect to static instantaneous navigational correlates – such as position or head direction. Head direction, grid, and place cells tend to be active only when an individual is either in a specific position or with its head in a specific direction. In contrast, only a negligible proportion of self-motion modulated neurons responded preferentially for a given speed or angular velocity. The vast majority saw their firing rate linearly (or quasi-linearly) ramp up proportionally to the correlate. This difference might be the hallmark of a general strategy for the neural coding of scalar quantities – speed and angular velocity in this case – as opposed to neural activity manifolds used to encode position and direction. In addition, self-motion modulated neurons generally showed a much higher firing activity with respect to neurons coding for instantaneous navigational correlates. Given the similarity between the responses to speed and angular head velocity signals, we hypothesize that this code for self-motion might be an just an instance of a more general sensitivity of the parahippocampal formation to the variation in time of behaviourally relevant quantities.

Before our study, angular head velocity (AHV) cells had been mainly characterized upstream of the hippocampal circuits, in the deeper structures linked to the processing of vestibular information (e.g. lateral mammillary nuclei, dorsal tegmental nucleus, thalamic nuclei and striatum) [91], [92], [96]. Some example of AHV modulation was reported in the retrosplenial cortex [97],[98] and has been linked to the accurate processing of visual inputs in the primary visual cortex [99]. In addition, angular head velocity was shown to influence the preferred orientation of some presubicular head

direction cells and to modulate the firing rate of a small fractions of fast spiking presubicular interneurons [100].

We observed cells modulated by angular head velocity in the presubiculum, the parasubiculum and the MEC. We hypothesised that such signals were not reported in previous studies either because of a restricted scope in analyses or because recordings were often clustered in the most dorsolateral part of the presubiculum, at the border of the retrosplenial cortex. Instead, the recordings analyzed here come from a much larger population, spread out across the medio-lateral and antero-posterior axis. Nevertheless, it is worth noting that these recordings do not show any topographic organisation of self-motion modulation.

Since the discovery of grid cells, many have attempted to understand how such a strikingly regular signal could be generated by individual neurons. A speed code is central to most of this theoretical work and a break-through was the characterization of speed cells and speed modulation in the MEC [24]. Interestingly, recent experimental work has demonstrated that grid cells activity is dependent on the integrity of the speed signal [101]. Likewise, the stability in head direction coding seems dependent on angular head velocity (AHV) signal [102] and vestibular inputs [92]. Therefore, self-motion signal could be similarly involved in the generation and the maintenance of both position and direction signals. It is interesting to note that in MEC LII, where there is virtually no HD signal, we find no sensitivity to AHV, arguing for the importance of a local coexistence of the two types of signals for the correct functioning of the spatial code.

The results presented here offer experimental evidence for the existence of a mechanism for the integration of speed and angular velocity signals that could subserve the update of the representation of space and orientation. The fact that we observe cells that code conjunctively for “static” and self-motion correlates could be taken, at first sight, as an experimental confirmation of the existence of the “hidden layer” of conjunctive cells hypothesized to provide the update mechanism required by continuous attractor network models. However, our results challenge theories assuming very specific architectures, that would require, for example, all speed and AHV cells to be conjunctively coding for position and head direction. Indeed, we observe that scores (i.e. grid, HD, AHV and linear speed score) show no correlation and the observed percentages of conjunctive cells are compatible with a scenario of independent assignment of coding properties. It is important that future theoretical work account for this independence. One exception to this observation is the dual absence of AHV and HD cells in MEC LII, suggesting that AHV signal is locally represented in concomitance

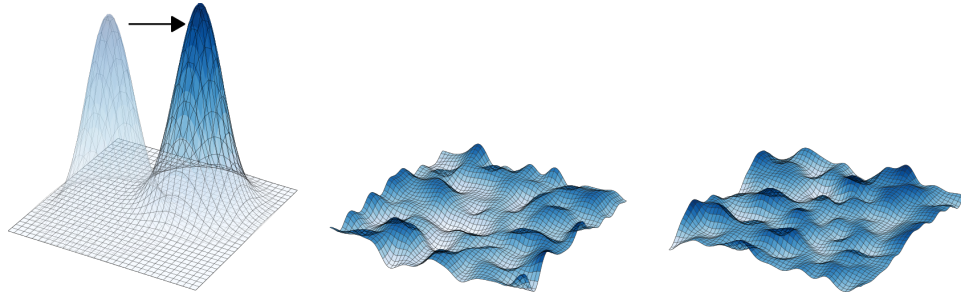
with the primary signal it serves to update. This leads us to hypothesize that the neural circuitry needed for the representation and the updating may coexist not only at the level of the brain area, but also locally within a layer. Interventional studies will be necessary to test how the conjunctive coding of static and self-motion signals is involved in the updating of the cognitive map.

Historically, running speed has been reported to show a strong correlation with the amplitude of theta oscillations recorded in the local field potential of freely behaving rodents [89],[90]. Likewise, many place and grid cells exhibit a strong modulation of their firing rate following those theta oscillations, either in a phase-locked or in a phase-precessing manner. Based on these evidence, many models have assumed that theta oscillations were inherent to the generation of the grid signal. Among them, the oscillatory interference-based models of grid cells assumed a velocity input to the grid network composed of speed and movement direction [103],[104]. Here we show that only 40% of our speed cells and AHV cells show a strong modulation by theta. This decoupling between speed/velocity signals and theta may be surprising for some. However, the role of theta in spatial coding had already been recently challenged by two new type of evidence. First, no theta oscillation had been recorded in the hippocampal local field potentials of bats who do exhibit both place and grid coding [59]. Second, the modulation of the septal oscillatory activity has no consequence on grid signal maintenance [105], therefore suggesting that non-theta septal correlates – such as attention – were involved in the grid cell signal disruption observed after septal inactivation.

In conclusion, the study presented here provides evidence of a widespread parahippocampal network involved in linear and angular speed coding that could have a crucial role in the updating of spatial cognitive maps. We hypothesize that derivative algorithms may have a generalized role in the updating of any type of information. Further studies, with either targeted inactivation of neurons or testing of non-spatial correlates will be necessary to establish whether one of the main roles of the parahippocampal region is to ensure the accurate updating of the hippocampal representation.

## Chapter 4

# Continuous attractors for dynamic memories



We have seen in chapter 1 that neural activity in the hippocampus tends to organize in temporally structured sequences. These sequences emerge early during development [56], and have been hypothesized to be crucial for memory consolidation [106] and retrieval [107], as well as for route planning [54].

The fact that sequential activity can be present before the exposure to the environment [55],[56] suggests that their dynamical nature is not specific to a role in spatial cognition, but is inherent to hippocampal operation in general. Moreover, in a recent study Stella et al. [108] showed that the retrieved positions during slow wave sleep are not always replaying experienced trajectories, but are compatible with a random walk on the low dimensional manifold that represents the previously explored environment.

This suggests that what is essential are not the sequences themselves, but the tendency to produce them: neural activity tends to move, constrained to abstract low-dimensional manifolds, which can then be recycled to represent spatial environments, and possibly non-spatial ones as well.

However, the activity is not constrained to a single subspace: replay in sleep can reflect multiple environments [109],[110], the content of awake replay reflects both the current and previous environments [107], and during behaviour fast hippocampal sequences appear to switch between possible future trajectories [111]. Further evidence comes from a recent study with human participants learning novel word pair associations [112]. The study shows that the same, pair-dependent neural sequences are played during the encoding and the retrieval phase.

A similar phenomenology – a dynamic activity on low dimensional manifolds – is present in memory schemata, cognitive frameworks that constrain and organize our mental activity [113], and have been shown to have a representation in the medial temporal lobe [114]. When we retrieve a schema, our memory follows a spontaneous flow that is guided by the boundaries of the schema.

Mechanistic models of memory usually neglect the dynamic component, treating memories as static objects, either discrete [32] or continuous [38],[39]. The production of sequences of discrete memories can be implemented with a heteroassociative component [115], usually dependent on the time integral of the instantaneous activity, that brings the network out of equilibrium and to the next step in the sequence. A similar effect can be obtained with an adaptation mechanism [116], with the difference that in this case the transitions are not imposed, but driven by the correlations between the memories in a so-called latching dynamics [117],[118]. In the case of continuous attractor networks, movement is induced by mechanisms that integrate an external velocity input and make use of asymmetric synaptic strengths [33],[37],[36]. In the simplest instantiation, the system is not a memory in the proper sense: the activity is constrained on a single attractive manifold.

Here we propose a continuous attractor network model able to store and retrieve multiple independent manifolds in a dynamical way. The model relies on a map-dependent asymmetric component in the connectivity, that produces a robust shift of the activity on the retrieved manifold. This connectivity profile is thought to be the result of a learning phase in which the mechanism of spike timing dependent plasticity (STDP) [119] produces the asymmetry. Crucially, the asymmetry is not treated here as a “pathological” feature, assumed to level out in the limit of long learning, but as a defining trait of the stored memories. The balance between two components – one

symmetric and trajectory-averaged, the other asymmetric and trajectory-dependent – is explicit in the formulation of the model, and allows to study their effects on memory storage.

In what follows we develop an analytical framework that allows to derive the dependence of important features of the dynamics, such as the shift speed and the asymmetry of the activity cluster, as a function of the relevant parameters of the model. We show with numerical simulations that the behaviour of the model is robust with respect to the details of the model, depending weakly on the shape of the interactions. Finally, we estimate the storage capacity for dynamical memories and we find it to be of the same order of the capacity for static continuous attractors, and even higher in some regimes.

## 4.1 The model

The model we consider is a continuous attractor neural network, with an additional anti-symmetric component in the connectivity strength. We consider a population of  $N$  neurons, with recurrent connectivity described by an interaction matrix  $J_{ij}$ , whose entries represent the strength of the interaction between neuron  $i$  and  $j$ . The activation function of the neurons is threshold-linear, i.e. the output  $V_i$  of neuron  $i$ , given the input  $h_i$ , is

$$V_i = [h_i]^+ = g(h_i - h_0)\theta(h_i - h_0) \quad (4.1)$$

where  $\theta$  is the Heaviside step function and the gain  $g$  and threshold  $h_0$  are global parameters of the network. The variables  $V_i$  are positive and continuous, and thought to represent the firing rates of the units.

The dynamic evolution of the network is regulated by the equations:

$$\tau \frac{\partial V_i}{\partial t} + V_i = \left[ \left( \sum_{j \neq i} J_{ij} V_j + b \left( \frac{1}{N} \sum_i V_i \right) - h_0 \right) \right]^+ \quad (4.2)$$

The first term on the right hand side represent the excitatory inputs provided to neuron  $i$  from the rest of the network through recurrent connections.  $b(\cdot)$  is a global inhibition term that, together with  $h_0$  and  $g$ , regulates the average activity and the sparsity of the network [78]. For the purpose of this work, the global inhibition term  $b$  can be reabsorbed in  $h_0$ , and will no longer be explicitly written.

A recurrent network of this kind can encode continuous maps in its connectivity matrix. In a basic model expressing static continuous attractors,

each neuron is assigned a preferential firing location  $x_i$  in the stimulus space, and the strength of the interaction between pairs of neuron is a decreasing, symmetric function of the distance between their preferred firing locations

$$J_{ij} \sim K_S(x_i - x_j). \quad (4.3)$$

This formula is assumed to come from a time-averaged Hebbian plasticity rule: neurons with nearby firing fields will fire concurrently and strengthen their connections, while far apart firing fields will produce weak interactions. The symmetry of the function  $K$ , usually called **interaction kernel**, ensures that the network reaches a static equilibrium, where the activity of the neurons represents a certain position in the map and, if not pushed, remains still.

#### 4.1.1 The shift mechanism

The assumption of symmetric interactions neglects any temporal structure in the learning phase. In the case of the learning of a spatial map, for example, the order in which recruited neurons fire along a trajectory may produce an asymmetry in the interactions as a consequence of a phenomenon called Spike Timing Dependent Plasticity [119], that requires the postsynaptic neuron to fire *after* the presynaptic one in order to strengthen the synapse. This phenomenon can be accounted for in the definition of the interaction kernel. Any asymmetric kernel can be decomposed in two contributions:

$$K(x_i - x_j) = K_S(x_i - x_j) + \gamma K_A(x_i - x_j) \quad (4.4)$$

where  $K_S$  is the usual symmetric component and  $K_A$  is an anti-symmetric function ( $K_A(x_i - x_j) = -K_A(x_j - x_i)$ ). The parameter  $\gamma$  regulates the relative strength of the two components.  $K_A$  generates a flow of activity along the direction of asymmetry: neuron  $i$  activates neuron  $j$  that, instead of reciprocating, will activate neurons downstream in the asymmetric direction. Mechanisms of this kind have been shown to produce a shift of the activity bump, without its disruption [33],[36],[37]. This effect is illustrated in Fig.4.1, and its quantitative properties are analyzed in detail in the next section.

#### 4.1.2 Storing multiple dynamic memories

A network with the connectivity structure described in Eq. 4.4 has a single dynamical attractor. In order to model the autoassociative memory properties of the hippocampus, we want the system to be able to store and retrieve

multiple manifolds, each with its own temporal structure. We construct an interaction matrix  $J_{ij}$  that is the sum of the contributions from  $p$  different, independent memories:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p K(x_i^\mu - x_j^\mu) \quad (4.5)$$

Here each  $x_i^\mu$  represent the preferred firing location of neuron  $i$  in the manifold representing memory  $\mu$ , and  $K$ , given by eq. 4.4, contains a symmetric and anti-symmetric component for each memory. The dynamic of this network, for low memory loads  $\alpha = p/N$ , evolves in two phases: a fast convergence to the retrieved manifold, and then a rigid movement along it, that replicates its temporal structure. The same activity, if projected on the other, unretrieved manifolds, appears as random noise. When the memory load  $\alpha$  is increased above a certain value  $\alpha_c$ , a phase transition occurs, and the network is not able to retrieve any memory, falling instead into a disordered state.

The value of  $\alpha_c$ , the storage capacity of the network, is estimated in section 4.3 and shown to be large enough to allow a network of biologically plausible size to store hundreds of different dynamical memories. Its dependence on the asymmetry parameter  $\gamma$  shows a non-trivial behaviour that depends crucially on the density of the connectivity on the network.

Before considering the multiple maps case, in the next section we present a quantitative study of the dynamics of the network in the case of a single map.

## 4.2 Dynamic retrieval

The presence of an asymmetry in the connection strengths prevents the system to reach a stationary equilibrium. Instead, it generates a steady flow of activity in the direction of the asymmetry. This flow is illustrated in Fig. 4.1, obtained with numerical simulation of a network encoding a one, two or three dimensional map respectively. Note that the bump of activity translates without disruption, producing a steady flow in the asymmetric direction.



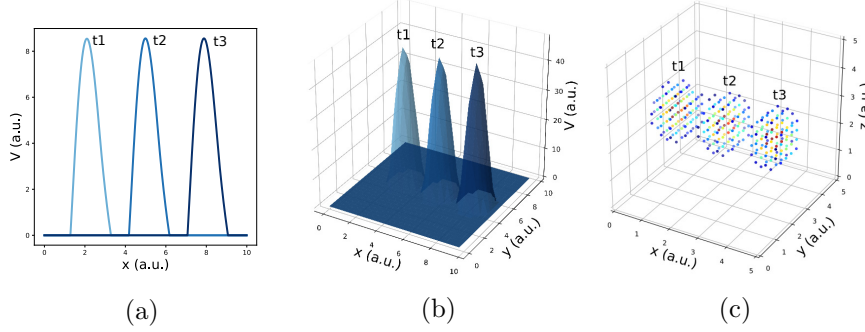


Figure 4.1: Dynamic retrieval in different dimensions. Three snapshots of the network activity at three different times (t1, t2 and t3) are shown for a system encoding a one dimensional **(a)**, two dimensional **(b)** and three dimensional manifold **(c)**. In (c), activity is color-coded (blue represents low activity, red high activity, silent neurons are not plotted for better readability). In all cases the anti-symmetric component is oriented along the x axis.

The dynamic behaviour of the system and its features can be described analytically with a generalization of the framework developed by Battaglia & Treves [38]. For this purpose, is easier to formulate the problem in the continuous limit, and describe the population activity by its profile  $V(x)$  on the attractive manifold parametrized by the coordinate  $x$ , and the dynamical evolution as a discrete step map, equivalent to Eq. 4.2.

$$V(x, t + 1) = g[h(x, t)]^+ \quad (4.6)$$

$$h(x, t) = \int_{-\infty}^{\infty} dx' K(x - x') V(x', t) - h_0 \quad (4.7)$$

The requirement of a rigid shift of population activity is then imposed by setting the activity at time  $t + 1$  to be equal at the activity at time  $t$ , but translated by an amount  $\Delta x$ , proportional to the speed of the shift. In this way we find the equation:

$$V(x + \Delta x) = g \left[ \int_{-\infty}^{\infty} dx' K(x - x') V(x') - h_0 \right]^+ \quad (4.8)$$

That we can rewrite, as:

$$V(x + \Delta x) = \begin{cases} g \int_{\Omega} dx' K(x - x') V(x') - h_0, & \text{if } x \in \Omega \\ 0 & \text{otherwise} \end{cases} \quad (4.9)$$

where  $\Omega$  is a compact domain for which there exist a solution of Eq. 4.8 that is zero at the boundary. This allows to exploit the fact that our threshold-linear system is, indeed, linear in the region in which  $V(x) > 0$ .

Equation 4.9 is valid in general, but we will focus here, to derive an analytical solution, on the one dimensional case and on an exponential kernel in the form

$$K(x - x') = e^{-|x-x'|} + \gamma \text{sign}(x - x') e^{-|x-x'|} \quad (4.10)$$

Differentiating twice Eq. 4.9, we obtain the differential equation

$$V''(x + \Delta x) = V(x + \Delta x) + 2gV(x) + 2g\gamma V'(x) + g\theta \quad (4.11)$$

This is a second order linear ODE, with constant coefficients. The presence of the shift term  $\Delta x$  inside the unknown function makes the equation non-trivial to solve. To solve the equation, we proceed in the following way: first, we look for a particular solution, that is easily found in the constant function

$$V_c = \frac{g\theta}{1 - 2g} \quad (4.12)$$

Then, we consider the associated homogeneous equation, and look for a solution in the form  $V(x) = e^{kx}$ . Where  $k$  is a solution of the characteristic equation  $C(k) = 0$ , with

$$C(k) = k^2 e^{k\Delta x} + 2g\gamma k + 2g - e^{k\Delta x}. \quad (4.13)$$

This transcendental equation has to be solved graphically in the complex domain, as shown in Fig. 4.2.

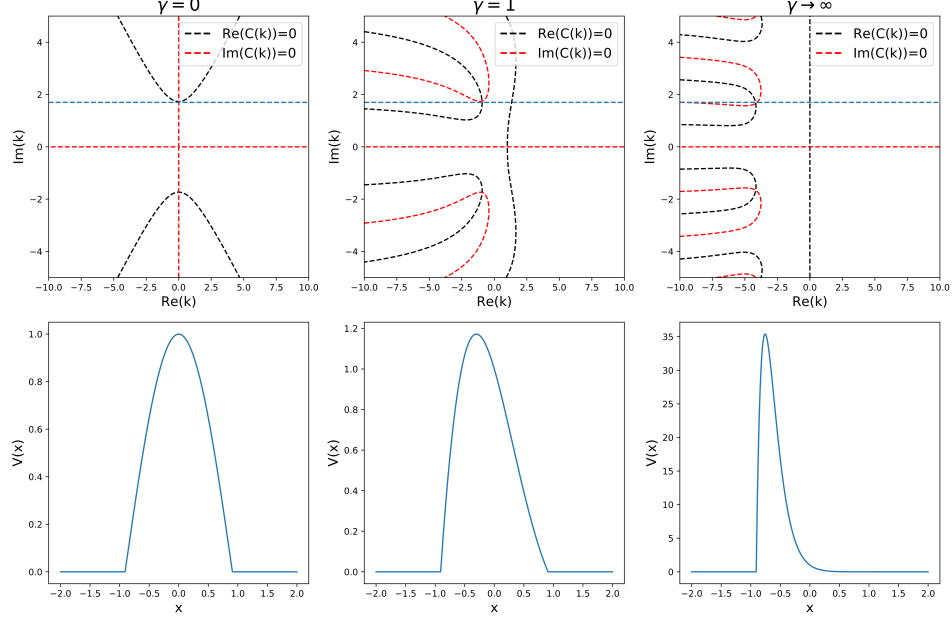


Figure 4.2: Analytic solution of equation (4.11). The top row shows the graphical procedure to find the complex zeros of the characteristic  $C(k)$  given in (4.13), for three different value of  $\gamma$ . Black and red lines show the zeros of the real and imaginary part of  $C(k)$  respectively. Their intersections are the complex solutions to  $C(k) = 0$ . The blue line represent the sparsity constraint  $k_i = k_i^S$ . The bottom row shows the corresponding solution shapes.

For each value of  $\gamma$  and  $\Delta x$ , the equation shows a pair of complex conjugate solutions

$$k_{1,2}^*(\gamma, \Delta x) = k_r(\gamma, \Delta x) \pm i k_i(\gamma, \Delta x) \quad (4.14)$$

The general solution of the equation will therefore have the form

$$V(x) = \begin{cases} C e^{k_r x} \cos(k_i x) + \frac{g\theta}{1-2g} & \text{if } -R \leq x \leq R \\ 0 & \text{if } -R > x \text{ or } x > R \end{cases} \quad (4.15)$$

In the limit case  $\gamma = 0$ ,  $\Delta x = 0$  (Fig. 4.2, first column), the solutions are pure imaginary ( $k_r = 0$ ), and we recover the solution of the symmetric case studied in [38]. From Eq.4.15 we can see that the absolute value of  $k_i$  is related to the width of the bump, and therefore to the sparsity of the

solution, by the relation

$$R = \frac{\pi}{2k_i} \quad (4.16)$$

We focus here on the case in which  $R$  is kept constant when  $\gamma$  changes, i.e. the case in which the network is constrained to operate at a certain sparsity. This constraint is enforced by requiring that the zeros of 4.13 lie in the subspace  $k_i = k_i^S$ , i.e. we have the same sparsity of the solution  $k_i^S$  of the symmetric case. This imposes a relation between  $\gamma$  and both  $\Delta x$  (related to speed of the shift) and  $k_r$  (related to the asymmetry of the shape of the solution). These relationships are shown in Fig. 4.3. The similarity between this two relationships can be understood intuitively by thinking that, for a fixed kernel shape, the larger the asymmetry of the solution, the more the bump will be translated by the evolution Eq. 4.8.

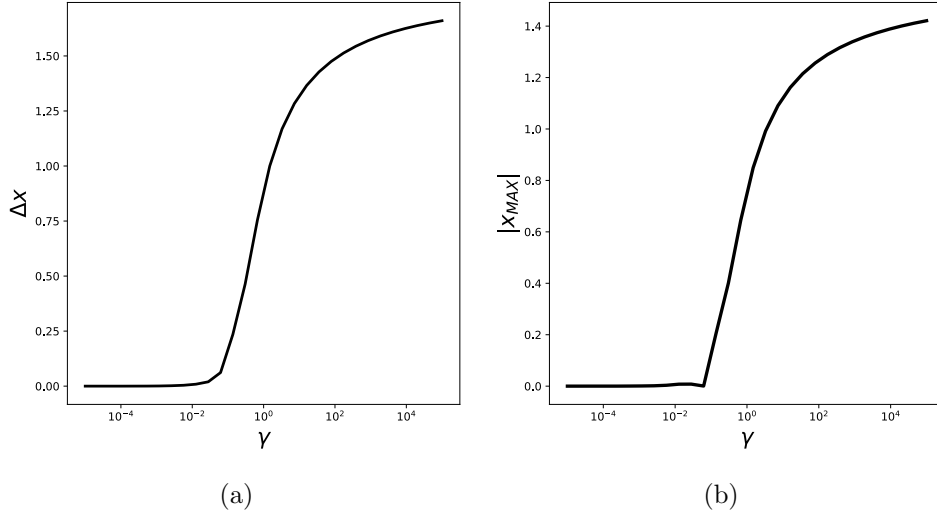


Figure 4.3: Dependence of speed and shape on  $\gamma$ . **(a)**  $\Delta x = -v\tau$  as a function of gamma. **(b)** Deviation of the position of the maximum of the bump from zero. This quantity is related to the value of the real part  $k_r$  of the solution of the characteristic equation by the relation  $x_{max} = \arctan(k_r)$ , and is one way to quantify the asymmetry in the bump shape produced by increasing  $\gamma$ . Note that the scale is logarithmic in  $\gamma$ .

The analytical results are presented here for a specific choice of the kernel, but the qualitative behaviour of the model is extremely general. In fact, numerical simulations show that a shifting bump can be obtained with a wide variety of interaction kernels, without any relationship required, for

example, between the symmetric and anti-symmetric components. Some examples are illustrated in Fig. 4.4.

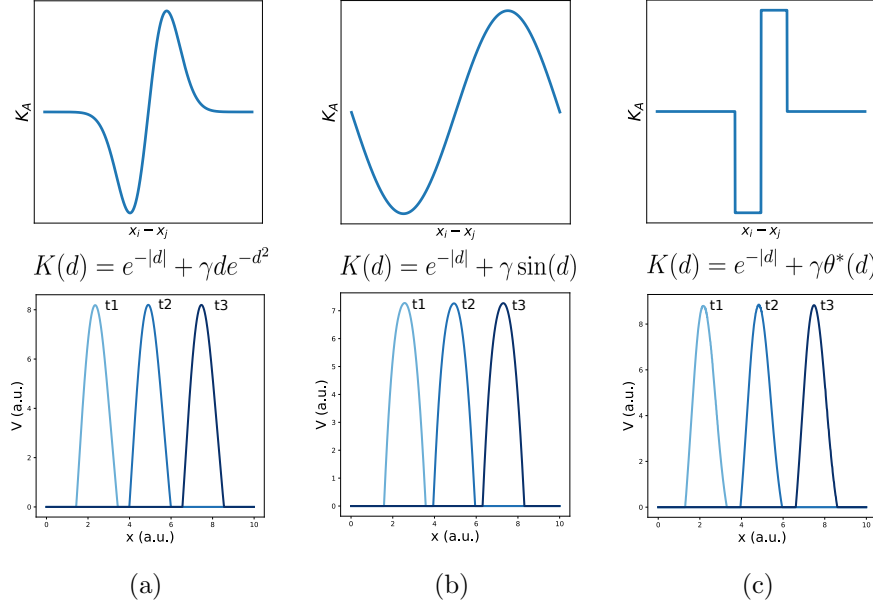


Figure 4.4: Different interaction kernels produce similar behaviour. Three examples of dynamics with the same symmetric component and three different anti-symmetric components. Top row: shape of the anti-symmetric component  $K_A$ . Bottom row: three snapshots of the retrieval dynamics for the corresponding  $K_A$ . **(a)** Gaussian derivative. **(b)** Sinusoidal. **(c)** Anti-symmetric step function ( $\theta^* = \theta(d)\theta(1-d) - \theta(-d)\theta(d-1)$ ).

Despite the robustness of the general features of the behaviour, the shape of the interaction kernel affects the details of the dynamics. Two parameters are particularly important: the relative strength  $\gamma$  between the symmetric and anti-symmetric components, and the characteristic length  $\xi$  of the anti-symmetric component. Their effect on the dynamics are shown in Fig. 4.5.

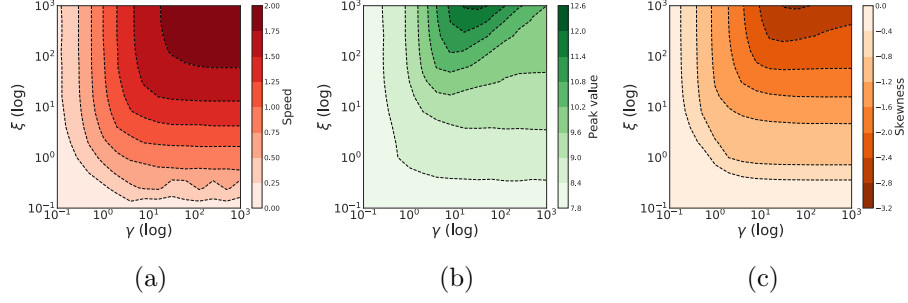


Figure 4.5: Effect of the kernel strength  $\gamma$  and its width  $\xi$ , in the case of exponential kernel  $K(d) = e^{-|d|} + \gamma \text{sign}(d)e^{-|d|/\xi}$ . **(a)** Shift speed **(e)** Maximum value of the activity bump and **(e)** Skewness of the activity bump.

Taken together, these results show that the model can implement a dynamic memory system, whose dynamics is constrained to a memorized attractive manifold and moves along it at constant speed without disruption. This behaviour depends weakly on the details of the connectivity kernel and can be implemented with a rather general type of asymmetric connections. However, in order to effectively work as a memory, the model has to be able to store and retrieve multiple different manifolds. The problem of storage capacity is addressed in the next section.

### 4.3 Storage capacity for dynamic continuous attractors

A network with the connectivity profile described in section 4.1.2 is able to store and retrieve multiple dynamic maps. The retrieval process, as in the single map case, unfolds in two phases: a fast transient in which the dynamics converges to one of the stored manifolds, and a subsequent stable shift. The second phase is illustrated in Fig. 4.6, obtained with numerical simulations.

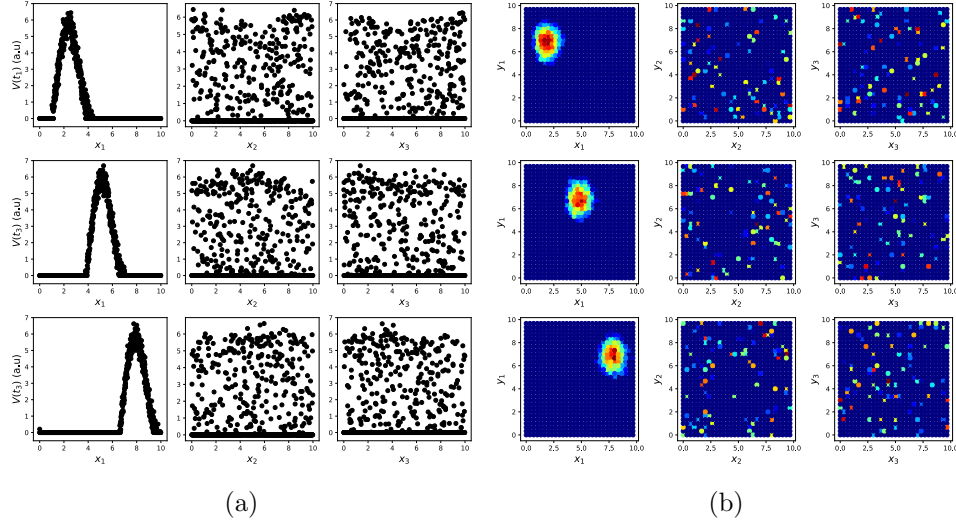


Figure 4.6: Dynamic retrieval in the presence of multiple memories in one dimension **(a)** and two dimensions **(b)**. Each row represents a snapshot of the dynamics at a point in time. The activity is projected on each of the three attractors stored in the network. In both cases, the first attractor is retrieved, and the activity organizes in a coherent bump that shifts in time. The same activity, projected onto the two non-retrieved maps looks like incoherent noise ((a) and (b), second and third columns).

The number of maps that can be stored and retrieved in this way is typically a fraction of the size of the network. Its magnitude, the storage capacity of the system, is crucial to determine if it can effectively operate as a memory.

To estimate the storage capacity for dynamic continuous attractors, we proceed along two complementary paths. For a fully connected network, where the analytical tools developed for equilibrium systems are not applicable, we take advantage of the fact that numerical simulations can be effective for the estimation of the capacity, since the number of connections per neuron  $C$  (the relevant parameters in the definition of the storage capacity  $\alpha_c = p/C$ ) scales as the number of neurons. For a highly diluted system, on the other hand, the number of neurons is much larger than  $C$ , making the simulation of the system very difficult in practice. We then resort to an analytical formulation based on a signal to noise analysis [38], that exploits the vanishing correlation between inputs of different neurons in a highly diluted network, and does not require symmetry in the connectivity.

In both cases we study the dependence of the capacity on two important parameters: the map sparsity, i.e. the ratio between the width of the connectivity kernel (fixed to one without loss of generality) and the size  $L$  of the stored manifolds, and the asymmetry strength  $\gamma$ . We present the analytic solution for the diluted case in the next section, showing that a simple approximation yields a remarkably accurate estimation of the capacity and allows to decouple the effects of map sparsity and asymmetry. The capacity is found to be monotonically but gently decreasing with both sparsity and asymmetry.

In section 4.3.2 we present the numerical results for the fully connected network. In this case, the presence of asymmetry can enhance the storage capacity, that is found to be maximal for finite values of  $\gamma$  and of map sparsity.

#### 4.3.1 Analytical calculation of $\alpha_c$ in the highly diluted limit

We consider here the highly diluted limit, the case in which the number of connections per neuron  $C$  is much smaller than the total number of neurons  $N$  ( $C/N \rightarrow 0$ ), and a number of maps  $\alpha = p/C$  is stored. This scaling makes the system extremely hard to simulate, but allows us to exploit its tree-like structure and the vanishing correlations between inputs to different neurons, and to study the network with a signal-to-noise approach that does not require its connectivity to be symmetric.

This approach, illustrated in details in [38], involves writing the local field  $h_i$  as the sum of two contributions: a signal term, due to the retrieved – “condensed” – map, and a noise term consisting of the sum of the contributions of the other, “uncondensed” maps. In the diluted regime these contributions are independent and can be summarized by a Gaussian term  $\rho z$ , where  $z$  is a random variable with zero mean and unit variance. In the continuous limit, assuming that map  $\mu = 1$  is retrieved we can write:

$$h(x^1) = g \int_L dx^{1'} K(x^1 - x^{1'}) V(x^{1'}) + \rho z \quad (4.17)$$

The noise will have variance:

$$\rho^2 = \alpha y L^2 \langle \langle K^2(x - x') \rangle \rangle \quad (4.18)$$

Where  $L$  is the size of the map,  $\langle \langle K^2(x - x') \rangle \rangle$  is the spatial variance of the kernel and

$$y = \frac{1}{N} \sum_i V_i^2 \quad (4.19)$$



is the average square activity.

We can write the fixed point equation for the average activity profile  $m^1(x)$ , incorporating the dynamic shift with an argument similar to the one made for the single map case:

$$m^1(x + \Delta x) = g \int^+ Dz (h(x) - h_0) \quad (4.20)$$

Where  $Dz = (e^{-z^2/2}/\sqrt{2\pi})dz$  and  $\int^+ f(x)dx = \int f(x)\theta(x)dx$ . The average square activity  $y$ , entering the noise term, reads

$$y = \frac{g^2}{L} \int dx \int^+ Dz (h(x) - h_0)^2 \quad (4.21)$$

Introducing the rescaled variables

$$w = \frac{-h_0}{\rho} \quad (4.22)$$

$$v(x) = \frac{m^1(x)}{\rho} \quad (4.23)$$

And the functions

$$\mathcal{N}(x) = x\Phi(x) + \sigma(x) \quad (4.24)$$

$$\mathcal{M}(x) = (1 + x^2)\Phi(x) + x\sigma(x) \quad (4.25)$$

where  $\Phi(x)$  and  $\sigma(x)$  are the Gaussian cumulative and the Gaussian probability mass function respectively, we can rewrite the fixed-point equation as

$$v(x + \Delta x) = g\mathcal{N} \left( \int dx' K(x - x')v(x') + w \right) \quad (4.26)$$

$$y = \rho^2 g^2 \int \frac{dx}{L} \mathcal{M} \left( \int dx' K(x - x')v(x') + w \right) \quad (4.27)$$

Substituting Eq.4.27 in the expression for the noise variance 4.18 we obtain

$$\frac{1}{\alpha} = g^2 L \langle \langle K^2 \rangle \rangle \int dx \mathcal{M} \left( \int dx' K(x - x')v(x') + w \right) \quad (4.28)$$

If we are able to solve Eq. 4.26 for the rescaled activity profile  $v(x)$ , we can use Eq. 4.28 to calculate  $\alpha$ . We can then maximize  $\alpha$  with respect to

$g$  and  $w$ : this yields the maximal value  $\alpha_c$  for which retrieval solutions can be found.

These equations are valid in general, but here we focus on the one dimensional case and the exponential kernel of Eq. 4.10. In this case we have

$$\langle\langle K^2(x-x')\rangle\rangle = (1 + \gamma^2)\langle\langle K_S^2(x-x')\rangle\rangle. \quad (4.29)$$

where  $K_S(x-x') = e^{-|x-x'|}$  is the symmetric component of the kernel. Eq. 4.26 can be transformed into a non-linear, delayed differential equation, that we can solve numerically. This solution procedure is illustrated in appendix D.1. Plugging the obtained form of  $v(x)$  into Eq. 4.28 we can calculate the capacity. The dependence of the capacity on  $\gamma$  is shown, for  $L = 60$  in Fig. 4.7.

We can see from the full dots in the figure that the contribution of the integral in Eq.4.28 is remarkably constant in  $\gamma$ . This is due to the fact that the distortions of the bump shape induced by the presence of the asymmetry have a negligible effect on the average square activity  $y$ , whose value is dominated by the dependence on  $\gamma$  of the spatial variance of the kernel (Eq.4.18).

This allows us to approximate the value of the integral in Eq. 4.28 with its value in the  $\gamma = 0$  case. We can then calculate the capacity as a function of  $\gamma$  and  $L$  by solving the symmetric case for different  $L$ s, and then incorporating the dependence on  $\gamma$  given by the kernel variance:

$$\alpha_c(L, \gamma) \sim \alpha_c(L, 0)/(1 + \gamma^2) \quad (4.30)$$

The result is shown in Fig 4.8.

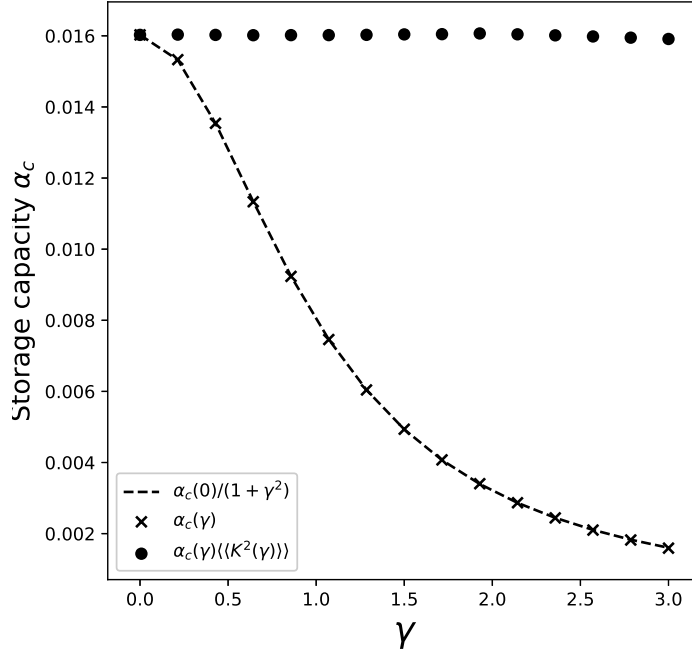


Figure 4.7: Dependence of the storage capacity on  $\gamma$ , for  $L = 60$ . The crosses show the full solution of Eq. 4.26 and 4.28. The dashed line is obtained by taking the value of the capacity  $\alpha(0)$  obtained with full solution at  $\gamma = 0$ , and multiplying it by the scaling of the kernel variance  $(1 + \gamma^2)$ . Full dots show the value of capacity obtained with the full solution and the contribution of the kernel variance factored out.

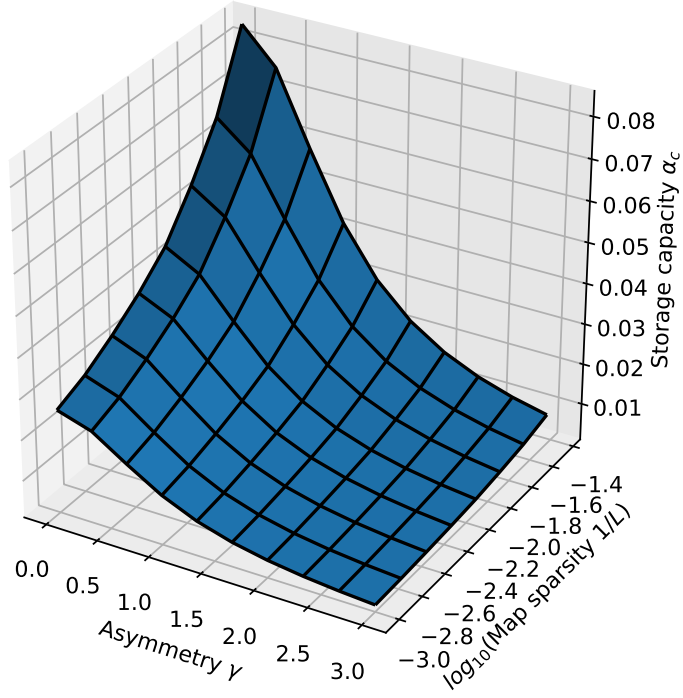


Figure 4.8: Dependence of the storage capacity on  $\gamma$  and  $1/L$  (represented as  $\log_{10}(1/L)$ ).

With this approximate decoupling we see that, for sparse maps and small values of the asymmetry, the capacity scales as

$$\alpha_c \sim -\frac{1}{\ln(1/L)(1+\gamma^2)} \quad (4.31)$$

The scaling with  $1/L$  is the same found by Battaglia & Treves [38] in the analysis of the symmetric case, as expected: for  $\gamma = 0$  the two models are equivalent. The presence of asymmetry decreases the capacity, but does not have a catastrophic effect: the decrease is continuous and scales with

a power of  $\gamma$ . There is therefore a large range of values of asymmetry and map sparsity in which a large number of dynamic map can be stored and retrieved. We will see in the next section how this picture changes in a fully connected network, in which the asymmetry can enhance the capacity.

### 4.3.2 Numerical estimation of $\alpha_c$ for a fully connected network

To estimate the storage capacity for a fully connected network, we proceed with numerical simulations. For a network of fixed size  $N$ , and for given  $\gamma$ ,  $L$  and number of maps  $p$ , we run a number of dynamics  $D$ , letting the network evolve from a random initial configuration. We consider a simulation to have performed a successful retrieval if the global overlap

$$m^\mu = \frac{1}{N^2} \sum_{i \neq j} V_i V_j K_S(x_i^\mu - x_j^\mu) \quad (4.32)$$

that quantifies the coherence of the activity with map  $\mu$ , is large for one map  $\mu^*$  (at least 95% of the overlap value obtained in the case of a single map) and low in all others maps  $\mu \neq \mu^*$ . We then define the retrieval probability as  $p_r = D_r/D$ , where  $D_r$  is the number of observed retrievals.

We repeat the process varying the storage load, i.e. the number of stored manifolds  $p$ . As  $p$  is increased, the system reaches a transition point, at which the retrieval probability rapidly goes to zero. This transition is illustrated, for various values of  $\gamma$ , in Fig. 4.9.

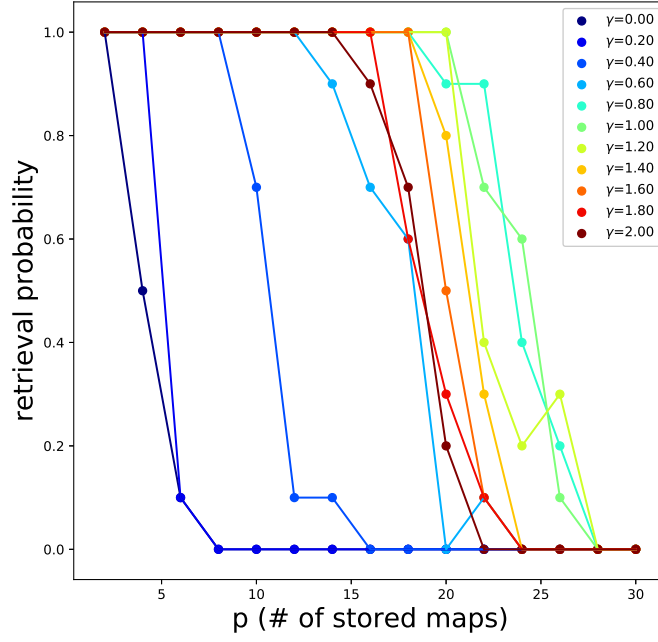


Figure 4.9: Retrieval - no retrieval phase transition for different values of  $\gamma$ . Obtained from simulations with  $N = 1000$ ,  $D = 10$  and  $L = 10$ . The non-monotonic dependence of the capacity from  $\gamma$  can be appreciated here: the transition point moves towards the right with increasing  $\gamma$  up to  $\gamma \sim 1$ , then back to the left.

The number of maps  $p_c$  at which the probability reaches zero defines as the storage capacity  $\alpha_c(\gamma, L) = p_c(\gamma, L)/N$ . Repeating this procedure for a range of values of  $\gamma$  and  $L$ , we obtain the plots shown in Fig. 4.10, for networks encoding one dimensional and two dimensional dynamical memories.

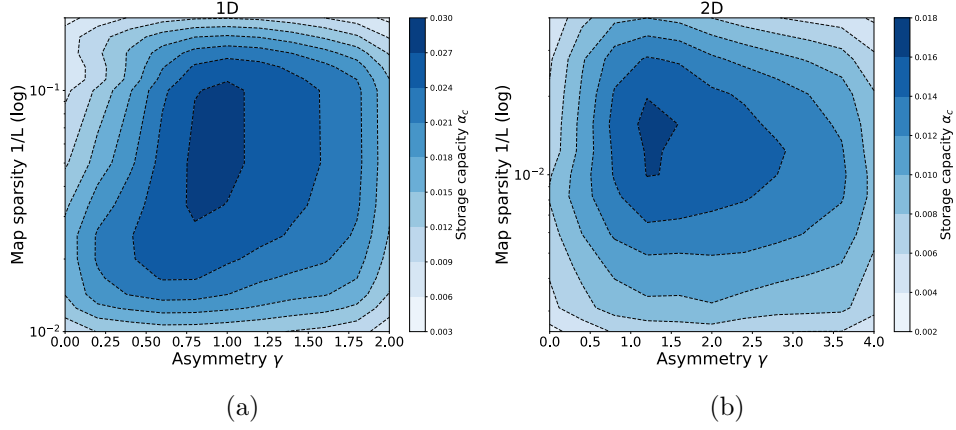


Figure 4.10: Storage capacity as a function of map sparsity  $1/L$  and asymmetry strength  $\gamma$ , for **(a)** one dimensional dynamic continuous attractors, **(b)** two dimensional dynamic continuous attractors.

The first thing that can be noticed is that, also in the fully connected case, the network can store a large number of maps, for a wide range of  $\gamma$  and  $L$ . A network with size in the order of ten thousand neurons could store from tens up to hundreds of dynamical memories.

The capacity for one dimensional attractors is higher than the one for their two dimensional counterparts. This is in line with what was found for symmetric networks [38].

Finally, we see that the peak of the capacity is found not only for intermediate values of map sparsity – again in line with what is known from the symmetric case – but also for intermediate values of the coefficient  $\gamma$ . This shows that moderate values of asymmetry can be beneficial for the storage of multiple continuous attractors, a non-trivial phenomenon that may be crucial for the memory capacity of biological networks.

## 4.4 Discussion

The results presented here show how a continuous attractor neural network with memory-dependent asymmetric components in the connectivity can function as a dynamic memory. Our model is simple enough to be treated analytically, robustly produces dynamic retrieval for a large range of the relevant parameters and shows a storage capacity that is comparable to – and in some cases higher than – the capacity for static continuous attractors.

The analytical solution of the single attractor case shows that the interaction between the strength of the asymmetry and the velocity of the shift can be modulated by global features of the network activity such as its sparsity. This makes the network able to retrieve at different velocities in different regimes, without necessarily requiring short term synaptic modifications. The insensibility of the general features of the dynamics to the fine details of the shape of the interactions suggests that this mechanism could robustly emerge from learning or self organization processes in the presence of noise. The analysis of the storage capacity shows that the asymmetry does not heavily impair memory performance, and that, in densely connected networks, out of equilibrium effects can be beneficial for memory.

The storage capacity of out of equilibrium continuous attractors has been calculated, in a different scenario, by Zhong et al. [120]. The authors considered the case of an external signal driving the activity bump along the attractor, in a network of binary neurons, and proceeded to calculate the storage capacity with several assumptions that allowed to model the interference of multiple maps as thermal noise. Interestingly, their main result is broadly compatible with what we show here: in the highly diluted regime the velocity of the external signal has a mild – detrimental – effect on the capacity. This hints that out of equilibrium effects could show some form of universality across different network models and implementations of the shift mechanism.

The possibility of dynamic retrieval makes attractor models suitable for the description and the quantification of complex memory phenomena such as hippocampal replay. The model we propose suggests that tendency of the activity to move in the neural population is a natural feature of networks with asymmetric connectivity, when the asymmetry is organized along a direction in a low dimensional manifold, and that static memories could be the exception rather than the rule. Indeed, Mehta et al. [121] have shown that place fields become more asymmetric in the course of spatial learning, demonstrating that the idea that symmetry emerges from an averaging of trajectory-dependent effects [122] does not always hold true. The model we



presented can be useful for the quantification of the effects that symmetry and asymmetry in the interactions have on the acquisition, retention and retrieval of memory.

Another interesting potential use of dynamic continuous attractors is the description of memory schemata. This will require further steps, such as an account of the interaction between hippocampus and neocortex, and a mechanism for the transition between different dynamical memories. Nevertheless, the idea of dynamic retrieval of continuous manifold and the integration of the model presented here with effective models of cortical memory networks [123] open promising perspectives.

Our model can describe continuous attractors with more than one dimension; however, it is worth noting that in the cases analysed here the asymmetry is constant along a single direction in each attractor. This can describe the situation in which the temporal evolution of the memory is structured along a certain dimension, and free to diffuse, without energy costs, in the remaining ones. The description of several one-dimensional trajectories, embedded in a two dimensional or three dimensional space and possibly intersecting, would instead require a position-dependent asymmetric component: this is an interesting direction that will be pursued in future work.

Finally, the full analytical description of a densely connected, asymmetric attractor network is a challenge that remains open, and can yield valuable insights on the workings of the neural circuits underlying memory.

## Chapter 5

# Conclusions

The work presented in this thesis aims to provide some novel ideas, quantitative analyses and experimental evidence on the role of the hippocampus and the surrounding cortices in the coding of space, time and memory. The results presented try to shed some light on open problems and suggest possible directions for future investigation.

The theoretical analysis presented in chapter 2 shows how grid coding is optimal for the storage of many memory in the dense regime. The two models considered allow us to investigate the effect of both the density of the activity (the number of active neurons in each memory) and the density of the connectivity (the number of connections per neuron). In both cases, hexagonal symmetry allows for the storage of a far greater number of independent memories than the simple square symmetry, when density increases. This is due to non-trivial effects of the hexagonal shape on the reverberation of noise through the loops of the networks, in the rate neuron model, and to the larger metastability of the hexagonal grid pattern, in the binary case.

The optimality of grids suggests that this kind of code could have a broader role than a metric for space during navigation. Indeed some evidence of grid-like coding has been found in regions beyond MTL [124], and grid cells seem to be involved also in the representation of abstract spaces [70][69][125]. The fact that the capacity for independent grid maps is largely above unity in a biologically plausible network suggests that grid representations can be useful for memory, a theoretical argument that challenges the standard view of the grid map as universal. This memory capacity is not exploited in the representation of simple, flat environments [57], but could be useful when the metric spaces represented differ more significantly. Boc-

cara et al. [66] and Butler et al. [67] have recently shown that grid patterns are deformed by the presence of salient features in the environment, and the first study shows some evidence of the presence of two distinct grid maps, one in the presence and one in the absence of rewards. A large number of independent grid representations could be of use for the encoding of abstract or cognitive spaces: there is no a priori reason to assume that spaces with possibly very different properties should share the same, universal representation.

Moreover, the universality of the grid representation is incompatible with curved surfaces, rarely encountered by laboratory animals, but very common in the natural environments of rodents [126]. Curved environments do not only challenge the universal grid hypothesis, but pose interesting questions for the theory of attractor neural networks: the representation of curved manifolds in a continuous attractor network is challenged by the impossibility of a full coherence between the representations [80]. The development of a computational description of a memory for curved environments is an interesting open perspective for future theoretical investigation. On the experimental side, the study of the effects of curvature on the grid pattern can probe interesting features of spatial memory in naturalistic settings.

The analysis of electrophysiological recordings presented in chapter 3 addresses the problem of the update of the spatial representation. The construction of multimodal representations of space has to be accompanied by a mechanism to integrate the variation of the encoded parameters. In the framework of attractor network theory, this mechanism has been hypothesized to be constituted by layers of cells sensitive simultaneously to the static correlate and its variation [23], appropriately connected to the population responsible for the encoding of the positional/orientational information. Independently of attractor network theories, the nature and extent to which self-motion information is represented in the parahippocampal cortices is a matter of open investigation.

Our analysis reveals that cells coding for linear speed, previously discovered in MEC and CA1, are also present in similar quantity in the pre and parasubiculum. This code for linear speed is complemented by a code for angular head velocity, which is also spread across these areas. These results suggest that the parahippocampal region could have a more general role in the representation of “derivative signals”, i.e. variations of behaviourally relevant quantities. Further investigation is required to establish the dependence of successful navigation on an intact population of speed and angular head velocity cells, as well as to clarify the role of these populations in the encoding of the variation of non-spatial signals.

The analysis of the overlap of speed and angular velocity cells with “static” populations (grid and head direction cells) shows a rich structure of conjunctive coding. The lack of correlation between different scores, and the magnitude of the overlap between different cell populations suggests the absence of a precise, finely tuned structure. The scenario is indeed compatible with the idea of independent assignment of the different coding properties: the number of cells that find themselves to be sensitive to both head direction and angular velocity – for example – is the one that would be expected if these two labels were distributed among the population independently one from the other. This fact opens interesting challenges for mechanistic models, that will have to account for this flexibility in the distribution of coding properties in their description of connectivity architectures. Together with the theoretical arguments presented in chapter 2, these results challenge the view of a parahippocampal region whose role is rigidly established by pre-wired, genetically determined neural circuits. They suggest, instead, the predominance of flexible, self-organizing mechanisms, able to shape with some freedom the firing properties and the interactions between neurons in response to the behavioural context.

An interesting feature of “derivative” coding is the monotone dependence of the firing rate on the value of the correlate (speed or angular velocity), that is strikingly different from the place-like coding found in both grid and head direction cells. We speculate that this could be related to the nature of the quantity being represented: derivative signals are scalar quantities, and have their range effectively bounded by the fact that position and direction cannot change too suddenly. Quantities of this kind can be effectively represented by simple proportionality of the firing rate. On the other hand, multidimensional and potentially unbounded quantities require a different representation, in which a preferred location in the space is represented by each cell. We hypothesize that this difference in the nature of coding is also at the origin of the differences between the average firing of “static” and “derivative” cells: lower the former as the result of a place-like, sparse coding, higher the latter, proportional to signals that are almost always strong during locomotive behaviour.

Finally, our analysis shows that speed and angular velocity coding are independent from the theta rhythm. This can have important implications for theories of spatial cognition. It could suggest, for example, that the strong correlation between speed and theta power and frequency does not reflect a causal relationship, but could be driven by external factors. Additional investigation is required to clarify this aspect, that remains for now nothing more than speculation.

In chapter 4 we addressed the problem of the temporal structure of memories. The model we propose shows that the temporal structure of a continuously varying stimulus can be incorporated in a continuous attractor network with the simple mechanism of asymmetric synaptic connections. The model can robustly retrieve dynamical continuous attractors and its behaviour is largely independent of the precise shape of the synaptic interactions. This robustness is important if one thinks that these interactions, encoding memories of past events, should be the result of a learning process that is subject, in general, to a large amount of noise. Importantly, dynamic retrieval is not limited to a single manifold: the network is able to perform cued retrieval in the presence of multiple dynamic memories encoded in its connectivity. This does not impair the correct unfolding of the temporal structure of the retrieved memory up to a certain number of stored memories, that defines the storage capacity of the system. The calculation of the storage capacity for dynamic continuous attractors shows that, for sparsely connected network, the capacity is high enough for the storage of hundreds of dynamic maps, and degrades gently with the strength of the asymmetric component. Interestingly, in densely connected networks, numerical simulations show that moderate values of the asymmetry can even enhance the capacity, a non-trivial effect whose analytical understanding is an important open direction. These results together suggest that asymmetry in the synaptic connectivity could be crucial for memory function, allowing the encoding of the temporal structure of memories and even improving the storage capacity.

The dynamical retrieval of the model generalizes, in the framework of attractor networks, the idea of cognitive maps, incorporating a temporal organization in the low-dimensional manifold encoding the structure of the memory. This feature is reminiscent of the idea of memory schemata – constructs that can guide and constrain our mental activity when we reminisce about the past or imagine future or fictional scenarios. The use of dynamic attractor networks for the study of memory schemata would require the integration of a mechanism describing the interaction between MTL and other cortical areas, and a way to model transition between dynamical memories, and is an interesting direction for future studies.

sec

# Appendices

## Appendix A

# Appendices to chapter 2

### A.1 Mean field equations: Binary Model

The free-energy can be written, in the large  $N$  limit, in terms of macroscopic quantities:

$$\begin{aligned} \mathcal{F} = & \frac{\alpha\beta}{2}r(f-q) - \frac{\alpha}{\beta}\Omega(q, \beta) + \int \mu(\vec{x})\rho(\vec{x}) \\ & - \frac{1}{2} \int d\vec{x}d\vec{y} \rho(\vec{x})K(|\vec{x}-\vec{y}|)\rho(\vec{y}) \\ & - \frac{1}{\beta} \int d\vec{x} \int Dz \ln[1 + e^{\beta z\sqrt{\alpha r} + \beta\mu(\vec{x})}] \end{aligned} \quad (\text{A.1})$$

where  $\beta$  is an inverse temperature or noise level, and the function  $\Omega(q, \beta)$  is given by

$$\begin{aligned} \Omega(q, \beta) = & 2 \sum_{k_1=1}^N \sum_{k_2=1}^N \left\{ \frac{\beta(q-f^2)}{1/\lambda_{k_1, k_2} - \beta(f-q)} \right. \\ & \left. - \ln[1 - \lambda_{k_1, k_2}\beta(f-q)] \right\} \\ & + \sum_{k=1}^N \left\{ \frac{\beta(q-f^2)}{1/\lambda_{0, k} - \beta(f-q)} - \ln[1 - \lambda_{0, k}\beta(f-q)] \right. \\ & \left. + \frac{\beta(q-f^2)}{1/\lambda_{k, 0} - \beta(f-q)} - \ln[1 - \lambda_{k, 0}\beta(f-q)] \right\}. \end{aligned} \quad (\text{A.2})$$

The order parameters minimizing the free energy functional are the average activity  $\rho(\vec{x})$  (see main text) and

$$q = \int d\vec{x} \int Dz \left[ 1 + e^{-\beta z \sqrt{\alpha r} - \beta \mu(\vec{x})} \right]^{-2} \quad (\text{A.3})$$

$$\begin{aligned} r = & 4(q - f^2) \sum_{k_1=1}^N \sum_{k_2=1}^N \left[ \frac{1}{\lambda_{k_1, k_2}} - \beta(f - q) \right]^{-2} \\ & + 2(q - f^2) \sum_{k=1}^N \left\{ \left[ \frac{1}{\lambda_{0, k}} - \beta(f - q) \right]^{-2} \right. \\ & \quad \left. + \left[ \frac{1}{\lambda_{k, 0}} - \beta(f - q) \right]^{-2} \right\} \end{aligned} \quad (\text{A.4})$$

where  $\lambda$  enforces the constraint  $\int d\vec{x} \rho(\vec{x}) = f$  and  $\lambda_{k_1, k_2}$  are the eigenvalues of the kernel  $K$  and are given by

$$\begin{aligned} \lambda_{k_1, k_2} &= w \frac{J_1(2\sqrt{z_{k_1, k_2}})}{\sqrt{z_{k_1, k_2}}} \\ z_{k_1, k_2} &= w\pi \left( k_1^2 + \left( \frac{k_2 - k_1 \cos \theta}{\sin \theta} \right)^2 \right) \end{aligned} \quad (\text{A.5})$$

where  $J_1$  is the Bessel function of the first kind of order 1.

In the text we focus on the limit of vanishing stochastic noise  $\beta \rightarrow \infty$ , and the term  $\beta(q - f)$ , which remains finite in such limit, can be identified with the parameter  $\psi$  of the threshold-linear model, quantifying the reverberation through the loops of the network of the *quenched* noise, which is due to the interference of the other maps.

## A.2 Mean field equations: Threshold-linear Model

When an energy functions can be defined (with full or in any case symmetric connectivity) the thermodynamics of the system is dominated by the minima



of the free energy *density*

$$\begin{aligned}
\mathcal{F} = & -T \left\langle \left\langle \int Dz \ln \text{Tr}(h, h_2) \right\rangle \right\rangle - \frac{1}{2} \sum_{\sigma, l} |\mathbf{m}^{\sigma, l}|^2 \\
& -B(m) - \sum_{\sigma, l} (\mathbf{m}^{\sigma, l})^2 + mB'(m) - r_0 y_0 + r_1 y_1 \\
& + \frac{\alpha d}{2\beta} \left( \ln[1 - T_0 \beta(y_0 - y_1)] - \frac{\beta y_1}{1 - T_0 \beta(y_0 - y_1)} \right)
\end{aligned} \tag{A.6}$$

where we have maintained a notation consistent with [78] and [38], for example

$$\begin{aligned}
\text{Tr}(h, h_2) = & k + \left( \frac{\pi g'}{2\beta} \right)^{1/2} \exp \left( \frac{\beta g'}{2} (h_0 - h)^2 \right) \\
& \left\{ 1 + \left[ \frac{\beta g'}{2} (h_0 - h) \right] \right\}
\end{aligned} \tag{A.7}$$

$$\begin{aligned}
h = & \sum_{\sigma l} \mathbf{m}^{l\sigma} \cdot \boldsymbol{\eta}^{l\sigma} + B'(m) \\
& - z \sqrt{(-2T_o r_1)}
\end{aligned} \tag{A.8}$$

$$h_2 = r_1 - r_0 \tag{A.9}$$

$$1/g' = 1/g - 2h_2 \tag{A.10}$$

$$Dz = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz, \tag{A.11}$$

while  $\langle\langle \cdot \rangle\rangle$  denotes an average over the quenched noise (the field centers in all other stored maps, distinct from the one which is currently expressed); and  $B(x)$ , together with the gain  $g$ , can be used to constrain the mean activity and the sparsity of the activity pattern [78], analogous to the parameter  $\lambda$  in the binary model.

The minima are given, in the limit  $T \rightarrow 0$ , by the saddle point equations

$$\mathbf{m}^{l\sigma} = g' \left\langle \left\langle \boldsymbol{\eta}^{l\sigma} \int_{h>Th} Dz(h-Th) \right\rangle \right\rangle \quad (\text{A.12})$$

$$m = g' \left\langle \left\langle \int_{h>Th} Dz(h-Th) \right\rangle \right\rangle \quad (\text{A.13})$$

$$y_0 = g'^2 \left\langle \left\langle \int_{h>Th} Dz(h-Th)^2 \right\rangle \right\rangle \quad (\text{A.14})$$

$$r_0 = \frac{\alpha T_0}{2} \frac{1 - T_0 \beta(y_0 - 2y_1)/d}{(1 - T_0 \beta(y_0 - y_1)/d)^2} \quad (\text{A.15})$$

$$y_1 = g'^2 \left\langle \left\langle \int_{h>Th} Dz(h-Th)^2 \right\rangle \right\rangle \quad (\text{A.16})$$

$$r_1 = \frac{\alpha T_0}{2d} \frac{T_0 \beta y_1}{(1 - T_0 \beta(y_0 - y_1)/d)^2} \cdot \left( -T_0 g' \left\langle \left\langle \int_{h>Th} Dz \right\rangle \right\rangle \right) \quad (\text{A.17})$$

Introducing the variables

$$\rho^2 = \frac{\alpha T_0^2 y_0}{d(1-\psi)^2} \quad (\text{A.18})$$

$$\psi = g' \frac{T_0}{d} \left\langle \left\langle \int^+ Dz \right\rangle \right\rangle \quad (\text{A.19})$$

we can write the free energy as a function of macroscopic quantities

$$\begin{aligned} \mathcal{F} = & -\frac{g'}{2} \left\langle \left\langle \int_{h>Th} Dz(h-Th)^2 \right\rangle \right\rangle + \sum_{\sigma l} \frac{(\mathbf{m}^{l\sigma})^2}{2} \\ & + mb(m) - B(m) + \frac{T_0 \rho^2 \psi d}{2} \end{aligned} \quad (\text{A.20})$$

with now

$$g' = \frac{1}{\frac{1}{g} - \frac{\alpha T_0}{1-\psi}}. \quad (\text{A.21})$$

To calculate the storage capacity, we focus on the case in which a single environment is retrieved by the network,

$$\mathbf{m}^{1l} > 0$$

$$\mathbf{m}^{\pi^l} = 0, \forall \pi \neq 1,$$

although the analysis can be extended to the retrieval of bump states that are localized in multiple environments. Without loss of generality, we assume therefore that environment  $\pi = 1$  is retrieved. With this assumption, and introducing the two signal-to-noise ratios

$$\mathbf{v}^l = \frac{\mathbf{m}^l}{\rho} \quad (\text{A.22})$$

$$w = \frac{b(m) - Th}{\rho} \quad (\text{A.23})$$

that represent respectively the *environment specific* component of the signal and the *uniform* background inhibition acting on each unit, the saddle point equations can then be reduced to a system of two equations in two variables

$$E_1(v, w) = A_1^2(v, w) - d\alpha A_3(v, w) = 0 \quad (\text{A.24})$$

$$E_2(v, w) = A_1(v, w) \left( \frac{d}{gT_0} - A_2(v, w) \right) - d\alpha A_2(v, w) = 0 \quad (\text{A.25})$$

where  $A_1(w, v)$ ,  $A_2(w, v)$  and  $A_3(w, v)$  are the averages:

$$\begin{aligned} A_1(w, v) &= \frac{1}{v^2 T_0} \left\langle \left\langle \sum_l \mathbf{v}^l \cdot \boldsymbol{\eta}^l \int^+ Dz \left( w + \sum_l \mathbf{v}^l \cdot \boldsymbol{\eta}^l - z \right) \right\rangle \right\rangle \\ &\quad - \left\langle \left\langle \int^+ Dz \right\rangle \right\rangle \end{aligned} \quad (\text{A.26})$$

$$A_2(w, v) = \frac{1}{v^2 T_0} \left\langle \left\langle \sum_l \mathbf{v}^l \cdot \boldsymbol{\eta}^l \int^+ Dz \left( w + \sum_l \mathbf{v}^l \cdot \boldsymbol{\eta}^l - z \right) \right\rangle \right\rangle \quad (\text{A.27})$$

$$A_3(w, v) = \left\langle \left\langle \int^+ Dz \left( w + \sum_l \mathbf{v}^l \cdot \boldsymbol{\eta}^l - z \right)^2 \right\rangle \right\rangle \quad (\text{A.28})$$

Solutions to equations (A.24) and (A.25) give the minima of the free energy that correspond to the retrieval of one of the stored environments.  $E_1(v, w) = 0$  describes a closed curve in the  $w - v$  plane, and these solutions are the intersections with  $E_2(v, w) = 0$ , which depends on the gain  $g$ .

As the storage load  $\alpha = p/C$  increases, this closed curve shrinks and eventually disappears. The value  $\alpha = \alpha_c$  at which the curve vanishes marks

a phase transition: for  $\alpha > \alpha_c$  retrieval solutions do not exist. The storage capacity  $\alpha_c$  can therefore be calculated by finding the vanishing point of  $E_1 = 0$ , and in this way one automatically selects the optimal value of the gain  $g$ , which therefore

### A.3 Finite connectivity and noise reverberation

Equations A.24 and A.25 can be extended to arbitrary value of connectivity density  $C/N$  following the self-consistent signal-to-noise analysis developed in [79]. This gives

$$E_1 = A_2^2 - \left(1 + \frac{C}{N} \left(\frac{(2-\psi)\psi}{(1-\psi)^2}\right)\right) d\alpha A_3 = 0 \quad (\text{A.29})$$

$$E_2 = \left(\frac{d}{gT_0} - d\alpha \frac{C\psi}{N(1-\psi)}\right) - A_2 = 0 \quad (\text{A.30})$$

These equations interpolate, as the free parameter  $C/N$  varies, between the two limiting cases of a fully connected network ( $C/N = 1$ ) and the extremely diluted case ( $C/N \rightarrow 0$ ) studied in [127]. We see that the reverberation factor  $\psi$  enters in the equation for the storage capacity as a correction on the loopless equation  $A_2^2 - d\alpha A_3 = 0$ , modulated by the connectivity density  $C/N$ , and that the lower the  $\psi$ , the higher the storage capacity.

For the fully connected network this correction gives

$$\frac{\psi}{1-\psi} = \sum_{k=1}^N \psi^k \quad (\text{A.31})$$

which is the sum over all the  $k$ -loops contributions to the reverberation of the noise.

Note, finally, that for ease of comparison with the binary model we have written in the main text

$$\begin{aligned} \bar{\mu} &= A_2 \\ r &= A_3. \end{aligned} \quad (\text{A.32})$$

### A.4 Free-energy barriers in the binary model

Free-energy values for the different metastable states are calculated using (A.4) after order parameters have been computed by solving the saddle-point equations. These equations are solved iteratively, starting from an

initial condition for order parameters, and iterating the values of the order parameters until convergence to fixed values. The free-energy values of the different metastable states are obtained by initializing  $\rho(\vec{x})$  as  $\rho_{Bp}(\vec{x})$  for Bump States (Fig.2.6(a,e)) and  $\rho_{BaE}(\vec{x})$  for Band Edge states (Fig.2.6(b,f)) or  $\rho_{BaD}(\vec{x})$  for Band Diagonal states (Fig.2.6(c,g)). In order to estimate the size of the barrier that must be jumped over in order to go from one state  $X$  to another state  $Y$ , we proceed as follows. The activity profile is initialized as  $\rho^{k=0,z}(\vec{x}) = z\rho_X(\vec{x}) + (1-z)\rho_Y(\vec{x})$ , with  $z$  chosen such that  $\rho^{k \rightarrow +\infty,z}(\vec{x}) = \rho_X(\vec{x})$  and  $\rho^{k \rightarrow +\infty,z-\epsilon}(\vec{x}) = \rho_Y(\vec{x})$  for  $\epsilon \ll z$ . When solving equations from such an initial condition, the network state goes close to a saddle-point lying at the boundary between the two basins of attraction associated to states  $X$  and  $Y$ , before sliding into state  $X$  as shown in Fig.2.7(c). The size of the barrier is then given by the difference between the free-energy of the saddle-point and that of the meta-stable state  $X$ .

## Appendix B

# Methods for Chapter 3

### B.1 Estimation of the behavioural correlates

#### *Position*

The position of the animal was estimated from the coordinates of two LEDs on the head of the animal. The X and Y coordinates of the both the LEDs were smoothed with a Gaussian filter with a 250 ms standard deviation, chosen to match the smoothing performed on the firing rate (see below), and the average between the two LED positions was used as the position of the animal.

#### *Head Direction*

HD was calculated as the angle between the line connecting the small LED to the big one and the x axis. HD is expressed in radians, 0 meaning that the rat head is lined to the x-axis, facing right.

#### *Linear speed and angular head velocity*

Speed was calculated as the modulus of the vector difference between the position at time  $t$  and the position at time  $t+1$ . Angular head velocity was calculated as the signed difference between the head direction at time  $t$  and the head direction at time  $t+1$ . The absolute value of the angular head velocity was used for the scoring of bidirectional angular head velocity cells (see below). No further smoothing was applied.

## B.2 Firing rate calculation

Instantaneous firing rate was obtained dividing the whole session in bins of 20 ms, coinciding with the frames of the tracking cameras. The spike count in each time bin was then calculated and divided by the temporal width to obtain the rate. The rate profile was smoothed with a 250 ms wide Gaussian filter.

## B.3 Speed filtering

The analysis on speed and angular velocity was performed on movement periods, defined as the ones in which the animal speed was  $>2$  cm/s. A speed filter was applied on the time series of each correlate, discarding the time points for which the instantaneous speed was below 2 cm/s, that were excluded in the subsequent analysis.

## B.4 Rate maps and tuning curves

### *Spatial rate maps*

The histograms for spike count and time spent in each location were constructed using equally spaced bins of 2-cm linear size. Each bin of the rate map was obtained as the ratio between spike count and time spent, smoothed with a Gaussian filter with standard deviation of 4 cm.

### *Directional rate maps*

The histograms for spike count and time spent facing each direction were constructed using equally spaced bins of size 6 degrees. Each bin of the rate map was obtained as the ratio between spike count and time spent, smoothed with a Gaussian filter with standard deviation of 6 degrees.

### *Speed and angular velocity tuning curves*

For tuning curve construction, the correlate was divided in equally spaced bins. For speed, 20 bins spanned the range between 2- and 50-cm/s (bin width 2.4 cm/s), for angular velocity the range -3-, +3-rad/s was again divided into 20 bins (bin width 0.15 rad/s). The firing rate in each bin was calculated as the average of the instantaneous firing rate values falling in the each given bin. A Gaussian smoothing window with standard deviation 0.15 rad/s for angular velocity and 2.4 cm/s for speed was applied.

## B.5 Shuffling

Chance-level statistics was calculated for a given variable W through a shuffling procedure. For each repetition, the firing rate time series was time shifted of a random interval of at least 30 seconds, with the end of the trial wrapped to the beginning. This procedure was repeated 100 times for each cell, and the shuffled score for variable W was calculated for each instance to compose the chance level statistics. For cell classification, all shuffled data from the same region were pooled together and the 99th percentile of the distribution was used as a classification criterion.

## B.6 Measure used for cell type classification

### *Speed Score*

The speed score was defined as the Pearson product-moment correlation between the cell's instantaneous firing rate and the instantaneous speed of the animal, across the whole recording session. This yields a score ranging from -1 to +1.

### *Unidirectional angular velocity score*

The unidirectional angular velocity score was defined as the Pearson product-moment correlation between the cell's instantaneous firing rate and the instantaneous angular velocity of the animal. Positive values of angular velocity correspond to clockwise head movement. Cells that had a score greater than the 99th percentile of the shuffled distribution were classified as clockwise modulated (CW), while cells whose score was lower than the 1st percentile were classified as counterclockwise modulated (CCW): they significantly code for head movement in the counterclockwise direction. CW and CCW populations are mutually exclusive by construction.

### *Bidirectional angular velocity score*

The unidirectional angular velocity score was defined as the Pearson product-moment correlation between the cell's instantaneous firing rate and the absolute value of the instantaneous angular velocity of the animal. Cells in this population increase their firing rate in response to head movement regardless of the direction.

### *Mean vector length (head-direction score)*

The mean vector length score is calculated from the head-direction tuning



map of a given cell as the sum:

$$\left| \frac{\sum \lambda_i e^{i\theta_i}}{\sum \lambda_i} \right|$$

Where  $\theta_i$  is the orientation in radians associated with bin  $i$  and  $\lambda_i$  is the firing rate in the bin. The sums run over all  $N$  directional bins, and the modulus of the resulting complex number is taken. Head direction was binned in bins of 6 degrees, and smoothed with a Gaussian filter with a standard deviation of 6 degrees.

#### *Grid score*

The grid score was calculated from the spatial autocorrelogram of a given cell. After exclusion of the centre of the autocorrelogram, the Pearson correlation of the autocorrelogram rotated by 30,60,90, 120 and 15 degrees (+- 3 degrees offsets) was considered. Only bins closer to the centre than an outer radius  $s$  were included in the calculation of the correlation. Given  $s$ , the grid score was defined as the difference between the average of the maximum correlations around 60 and 120 degrees (+- 3 degrees offsets) and the average of the minimum correlations around 30,90 and 150 degrees (+- 3 degrees offsets). The final grid score of the cell was then defined as the maximum grid score over values of  $s$  ranging from twenty to forty bins, computed at intervals of one bin.

#### *Theta index*

Theta modulation of individual cells was estimated from the frequency power spectrum of the spike-train autocorrelation histogram of the cell. A cell was defined to be theta modulated if the mean power in a 2 Hz window centred in the peak in the 5- to 11-Hz frequency range was at least fivefold greater than the mean spectral power in the 0- to 125-Hz range.

## **B.7 Estimation of the significance of overlaps between cell populations**

The observed overlaps between cell populations were compared to the ones that would result from the statistical null hypothesis of independent random assignment with a two-sided binomial test. The probability of observing an overlap of size  $k$  between two populations of sizes  $N_a$  and  $N_b$ , independently

drawn from a total number of cell  $N$  is given by:

$$p(k) = \frac{N!}{k!(N-k)!} p_{ab}^k (1 - p_{ab})^{N-k}$$

Where  $p_{ab} = p_a p_b$  and  $p_a = N_a/N, p_b = N_b/N$ .

## B.8 Information analysis

The information per spike conveyed by each cell about the correlate of interest (speed or angular head velocity) was calculated using the formula:

$$I = \frac{1}{\lambda} \sum_i \lambda_i p_i \log_2 \left( \frac{\lambda_i}{\lambda} \right)$$

Where  $i$  is the index of the correlate bin,  $p_i$  is the probability of observing the correlate in bin  $i$  (i.e. the normalized occupancy),  $\lambda_i$  is the average firing rate of the cell in bin  $i$ , and  $\lambda$  is the average firing rate of the cell. Speed was divided in 2 cm/s bins in the range 2-50 cm/s (as in all analysis, stillness periods were excluded), while angular head velocity was divided in 0.5 rad/s bins, in the range (-5,5) rad/s. Cells were considered to carry significant information about the correlate if the observed information rate exceeded the 99th percentile threshold of the null distribution obtained by shuffling the cell firing rate values (1000 shuffles per cell).

## B.9 GLM analysis

We analyzed the effect of each correlate (speed and angular head velocity) with a linear-nonlinear Poisson spiking GLM model. This model assumes that the firing rate of the cell depends on the value of the correlate as

$$r(t) = \exp\left(\sum_i X_i(t) w_i\right)$$

where  $X_i(t)$  is a one-hot vector (i.e. a vector with only one non-null element) indexing which value the correlate is taking at time  $t$ , and  $w_i$  are the coefficients of a linear filter quantifying the contribution of each value of the correlate to the firing rate of the cells. The model is fitted using the python module statsmodel.api, which finds the set of parameters  $w_i$  maximising the log-likelihood of the observed spikes, subject to an elastic-net regularization constraint. To perform the fitting procedure, the speed values have been

binned in 10 bins in the range 2-50 cm/s, and the angular velocity values divided in 10 bins in the range -3-3 rad/s. Cells were considered significantly modulated by a correlate if the log-likelihood of the best fit was significantly larger than the value obtained with only the average firing rate as a predictor. Significance was estimated with a 10-fold bootstrapping procedure to extract the confidence interval of the observed log-likelihood.

## B.10 Tuning curve fitting

Two different functional forms were fitted and compared to the tuning curve of modulated cells. A linear model

$$r = ax + b$$

And a sigmoid model

$$r = \frac{1}{1 + e^{-a(x-b)}}$$

where  $x$  is the value of the correlate (speed or angular head velocity) and  $r$  is the average firing rate of the cell at that value of  $x$ . Tuning curves were rescaled by their maximum value, in order to match the two model by number of parameters. The R2 fitting scores were then compared for each cells. Cells with a linear R2 greater than the sigmoid R2 were classified as linear, and vice versa.

## Appendix C

### Supplementary figures for Chapter 3

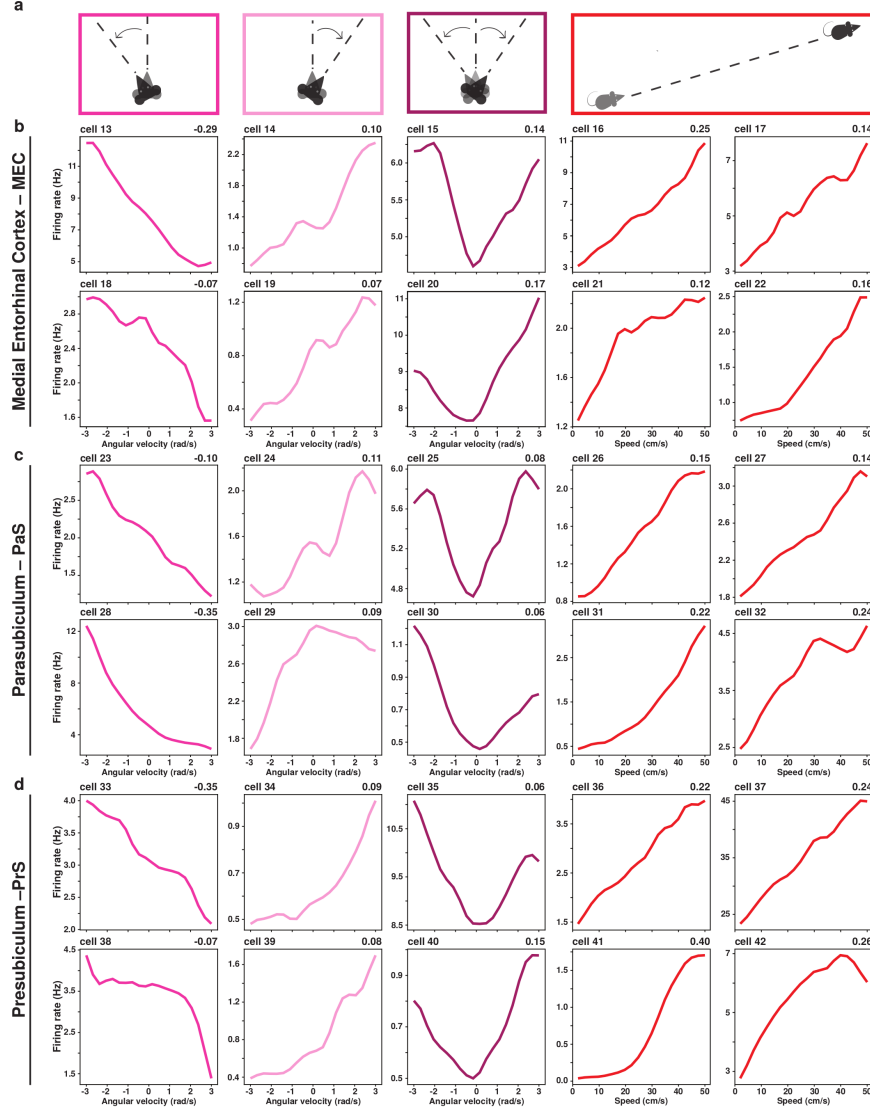


Figure C.1: Extended examples of speed and Angular head velocity cells  
 (a) Schematic representation of three type of angular head velocity (AHV) movement and linear speed, from left to right: counterclockwise (CCW, dark pink), clockwise (CW, light pink), bidirectional (BiDir, purple) and linear speed (red). 30 additional examples of self-motion cells: 6 AHV and 4 speed cells in each region; medial entorhinal cortex (b), parasubiculum (c) and presubiculum (d). The firing rate is represented as a function of angular velocity (in rad/s) or speed (in cm/s). AHV or speed scores are reported in the upper right corner. Cell ID are reported in the upper left corner. From left to right: CCW-AHV (dark pink), CW-AHV (light pink), BiDir-AHV (purple) and speed (red, last two columns). Note that the high values of the rate of cell 37 suggest that it may be an interneuron.

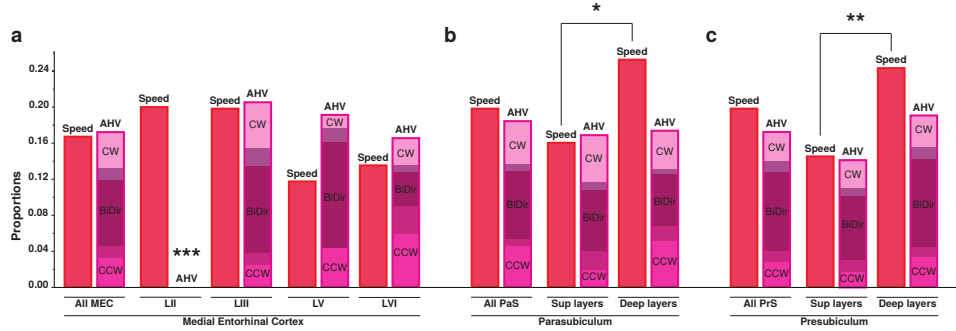
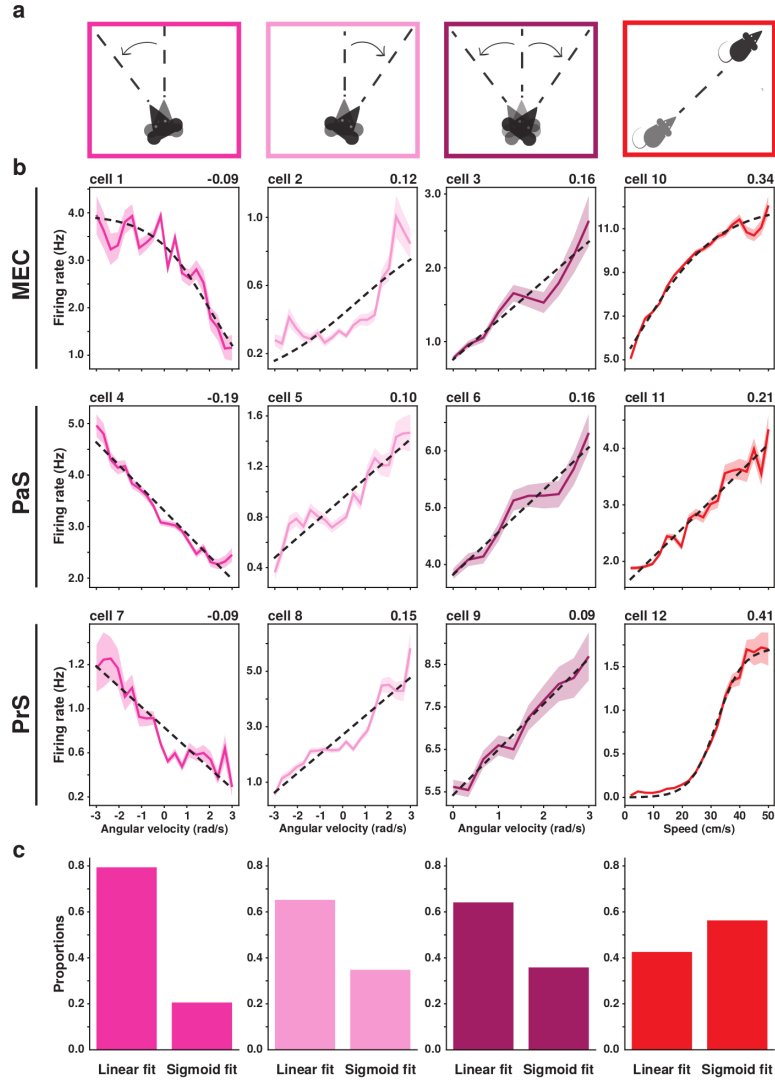


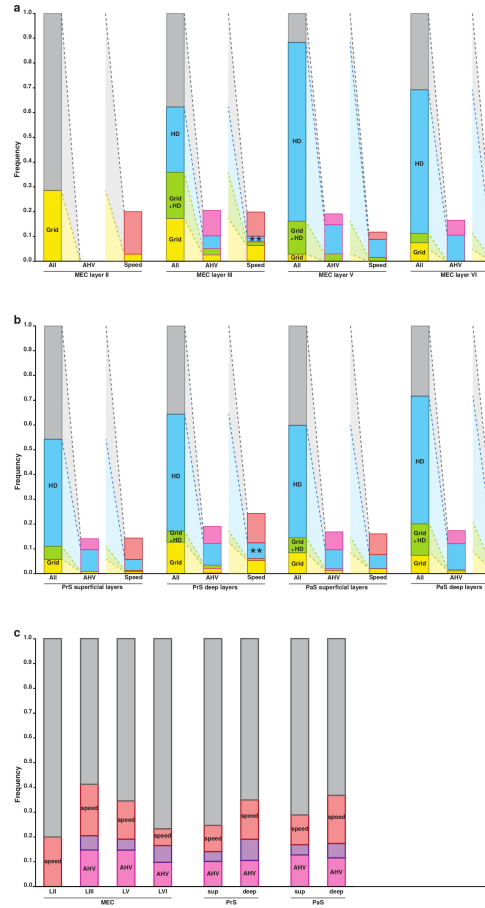
Figure C.2: Distribution of angular head velocity and linear speed modulation across parahippocampal layer

Proportion of speed (red) and AHV cells (CW light pink, CCW dark pink, BiDir purple) across layers. Shaded areas represent overlap between CW (CCW) and BiDir cells. **(a)** Proportions in medial entorhinal cortex: MEC LII (speed: 20.0%; AHV CW: 0%; AHV CCW: 0%, AHV BiDir: 0%) , MEC LIII (speed: 19.8%; AHV CW: 7.1%; AHV CCW: 3.8%, AHV BiDir: 12.8%), MEC LV (speed: 11.8%; AHV CW: 2.9%; AHV CCW: 4.4%; AHV BiDir: 13.2%) and LVI (speed: 13.5%; AHV CW: 3.7%; AHV CCW: 9%; AHV BiDir: 7.5%). Stars denote the significant absence of AHV cells in MEC LII (t-test, p-value < 0.001). **(b)** Proportions in the parasubiculum: superficial layers (speed: 16.1%; AHV CW: 6.0%; AHV CCW: 4%; AHV BiDir: 9.6%) and deep layers (speed: 25.3%; AHV CW: 4.79%; AHV CCW: 6.86%; AHV BiDir: 7.9%). **(c)** Proportions in the presubiculum: superficial layers (speed: 14.5%; AHV CW: 3.9%; AHV CCW: 3.1%; AHV BiDir: 9.3%) and deep layers (speed: 24%; AHV CW: 4.7%; AHV CCW: 4.5%; AHV BiDir: 12.2%). Stars denote the significant difference in speed cells between superficial and deep layers both in PrS and PaS (t-test, \*\* p-value < 0.01 and \* p-value < 0.05 respectively).

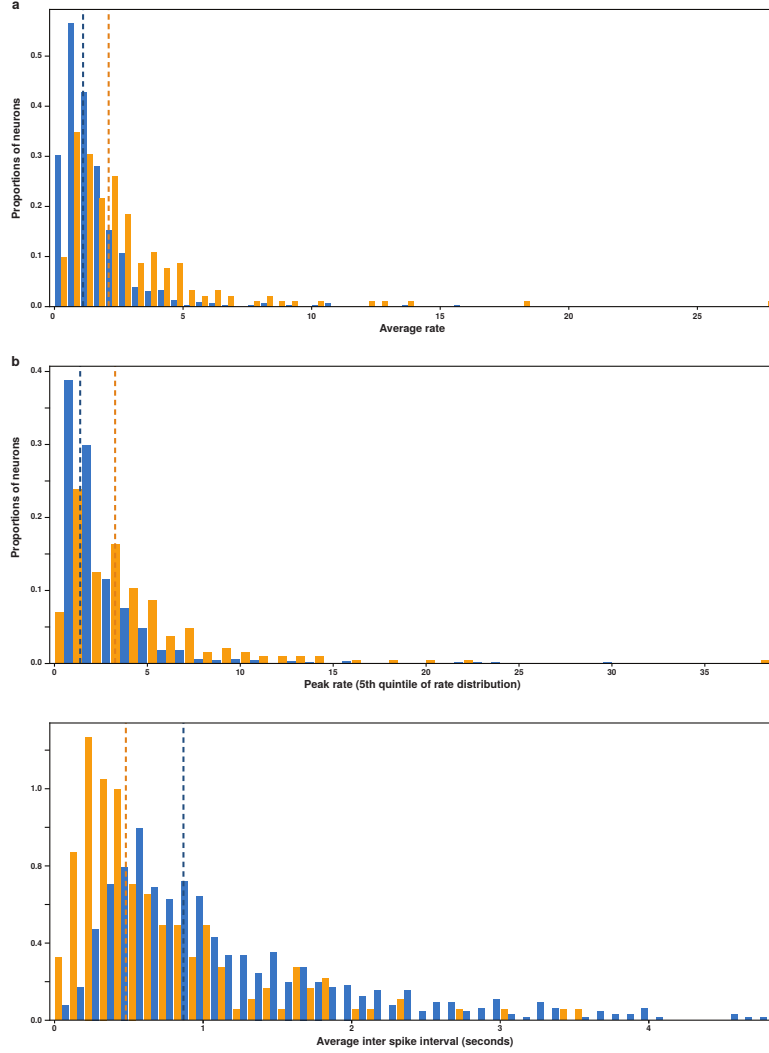








**Figure C.5: Distribution across area and layers of conjunctive coding**  
(a–b) Proportions of grid (yellow), HD (blue), grid x HD (green) cells in the whole layer population (left bar, black outline), within the AHV cell population (central bar, pink outline) and within speed cell population (right bar, red outline). Grey bars represent cells that are neither coding for grid nor HD. Pink bars in the AHV population histograms represents AHV cells that are neither coding for grid nor HD. Red bars in the speed population histograms represents speed cells that are neither coding for grid nor HD. Stars denote a significant change in proportions of a specific type of conjunctive cells within either the AHV or the speed population from what would be expected from the layer proportions within the general population. All proportions were as expected, except for an underrepresentation of speed x HD in MEC L III and in deep Prs (t-test, p-value < 0.01). MEC **(a)**: MEC layer II (grid: 28.6%; HD: 2%; grid x HD: 0%), layer III (grid: 35.9%; HD: 44.9%; grid x HD: 18.6%), layer V (grid: 16.2%; HD: 85.3%; grid x HD: 13.2%), and layer VI (grid: 11.3%; HD: 61.7%; grid x HD: 3.7%). PrS **(b, left)**: PrS superficial layers (grid: 11%; HD: 48.5%; grid x HD: 5.3%), and PrS deep layers (grid: 17.2%; HD: 51.6%; grid x HD: 4.5%). PaS **(b, right)**: PaS superficial layers grid: (14.5%; HD: 51.4%; grid x HD: 6%), PaS deep layers (grid: 20%; HD: 64.2%; grid x HD: 12.6%). Note the quasi-absence of HD and AHV cells in MEC LII. **(c)** Pink and red bars here represent the whole population of AHV (pink) and speed (red) cells in a given layer. Note that those are different populations than in (a–b). Purple bars represent cells whose activity is conjunctively modulated by speed and by AHV. From left to right: MEC layer II (AHV: 0%, speed: 20%, AHV x speed: 0%), MEC layer III (AHV: 20.5%, speed: 19.9%, AHV x speed: 5.8%), MEC layer V (AHV: 19.1%, speed: 11.8%, AHV x speed: 4.4%), MEC layer VI (AHV: 16.5%, speed: 13.5%, AHV x speed: 6.8%), PrS superficial layers (AHV: 14.1%, speed: 14.5%, AHV x speed: 3.9%), PrS deep layers (AHV: 19.1%, speed: 24.3%, AHV x speed: 8.5%), PaS superficial layers (AHV: 16.9%, speed: 16.1%, AHV x speed: 4%) and PaS deep layers (AHV: 17.4%, speed: 25.3%, AHV x speed: 5.8%).



**Figure C.6: Self-motion and static correlates have different firing properties**  
Comparison of firing properties between cells coding for self-motion (speed and AHV cells: orange bars) and static (HD and grid cells: blue bars) correlates. Vertical dashed lines indicate median values. **(a)** Distribution of the average firing rates of cells coding for motion and static signals (bin width: 0.5Hz). Median average rate for motion cells: 2.86 Hz (orange dashed line). Median average rate for static cells: 1.52 Hz (blue dashed line). **(b)** Distribution of the peak firing rates calculated as the 5th percentile of the firing rate distribution of each cell (bin width: 1Hz). Median peak rate for motion cells: 4.51 Hz (orange dashed line). Median peak rate for static cells: 1.99 Hz (blue dashed line). **(c)** Distribution of the average inter-spike interval (bin width: 0.1s). Median average inter-spike interval for motion cells: 0.71 s (orange dashed line). Median average inter-spike interval for static cells: 1.13 Hz (blue dashed line). Note that motion cells show a larger average and peak firing rate, as well as a lower average inter spike interval. This difference may be related to the nature of the coding: static cells use “place like”, sparse coding, while motion cells have a monotonic, dense response profile.

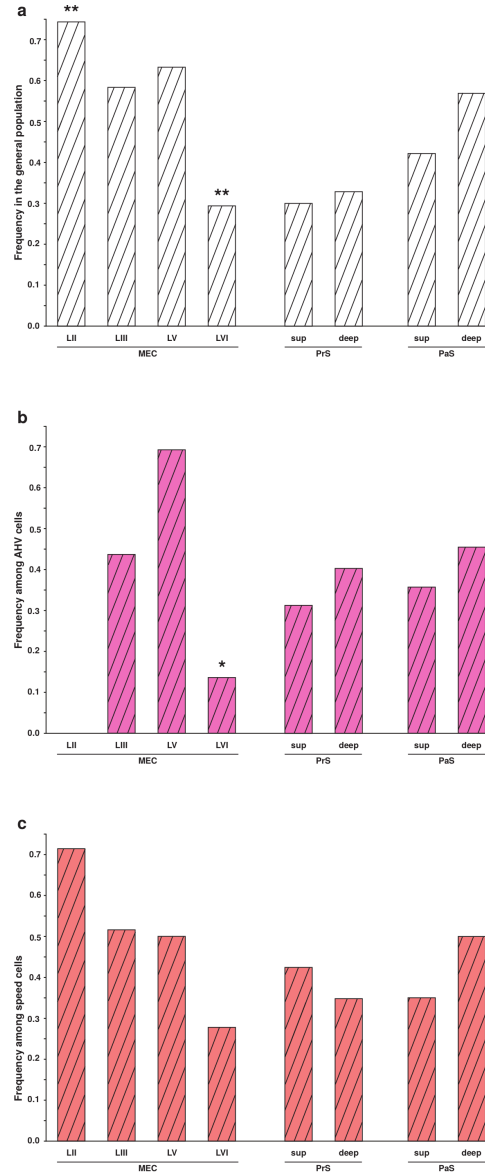


Figure C.7: Distribution of theta modulation by layer

Dashed bars represent proportions of theta modulated cells for each category of cells considered. Cells were considered theta modulated when their mean power in a 2 Hz window centred in the peak in the 5- to 11-Hz frequency range was at least fivefold greater than the mean spectral power in the 0- to 125-Hz range. (a) Percentages of theta modulated cells in the whole population (white dashed bars). From left to right: MEC LII (74.3%), MEC LIII (58.3%), MEC LV (63.2%), MEC LVI (29.3%), PrS superficial layers (29.9%), PrS deep layers (32.8%), PaS superficial layers (42.2%) and PaS deep layers (56.8%). Stars denote a significant difference in proportion of theta modulated cell in the layer considered, compared to the average theta modulation across all layers (t-test, \*\* p-value < 0.01). (b) Percentages of theta modulated cells in the AHV population (pink dashed bars). From left to right: MEC LII (0%), MEC LIII (43.8%), MEC LV (69.2%), MEC LVI (13.6%), PrS superficial layers (31.3%), PrS deep layers (40.2%), PaS superficial layers (35.7%) and PaS deep layers (45.5%). Stars denote a significant difference in proportion of AHV theta modulated cell from what would be expected given the average theta modulation in that specific layer (t-test, \* p-value < 0.05). (c) Percentages of theta modulated cells in the speed population (red dashed bars). From left to right: MEC LII (71.4%), MEC LIII (51.6%), MEC LV (50%), MEC LVI (27.8%), PrS superficial layers (42.4%), PaS deep layers (34.7%), PaS superficial layers (35%) and PaS deep layers (50%).

## Appendix D

# Appendices to chapter 4

### D.1 Solution of the equation for the activity profile in the case of many maps

We illustrate here the procedure for the numerical solution of the equation 4.26:

$$v(x + \Delta x) = g\mathcal{N}\left(\int dx' K(x - x')v(x') + w\right) \quad (\text{D.1})$$

We consider the one dimensional case and the exponential kernel

$$K(x - x') = e^{-|x-x'|} - \gamma\text{sign}(x - x')e^{-|x-x'|} \quad (\text{D.2})$$

First, following [38] we rewrite it with the transformation

$$u(x) = \mathcal{N}^{-1}\left(\frac{v(x)}{g}\right) \quad (\text{D.3})$$

obtaining

$$u(x + \Delta x) = g \int dx' K(x - x')\mathcal{N}(u(x')) + w \quad (\text{D.4})$$

We then transform this integral equation in a differential one, by differentiating twice. We obtain

$$u''(x + \Delta x) + 2g\gamma\Phi(u(x))u'(x) + 2g\mathcal{N}(u(x)) - u(x + \Delta x) + w = 0 \quad (\text{D.5})$$

where we used the fact that  $\mathcal{N}'(x) = \Phi(x)$ . Eq.D.5 is a second order, nonlinear delayed differential equation. To solve it, it is not sufficient to impose an initial condition on a single point for the solution and the first

derivative (i.e. something like  $u(x_0) = u_0, u'(x_0) = u'_0$ ): we have to specify the value of the function and its derivative in an interval  $[x_0, x_0 + \Delta x]$ .

To do so, we reason that, if we want a bump solution,  $u(x)$  has to be finite in for  $x \rightarrow \pm\infty$  and cannot diverge. We then require the function to be constant ( $u(x) = u_0, u'(x) = 0$ ) before a certain value  $x_0$ , whose value can be set arbitrarily without loss of generality.

The value  $u_0$ , at  $\gamma = 0$  and  $\Delta x = 0$  determines the shape of  $u(x)$ , as shown by the numerical solution presented in Fig. D.1 for  $u_0 < u^*$  the solution will diverge for  $x \rightarrow \infty$ , for  $u_0 > u^*$  it will oscillate. We are then left with a single value  $u_0(g, w) = u^*(g, w)$  for which the solution has the required form.

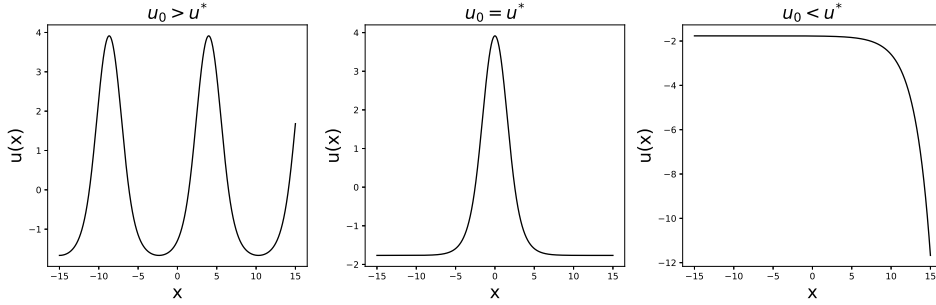


Figure D.1: Solutions to Eq.D.5 for  $g = 1, w = -1.8, \gamma = 0, \Delta x = 0$ .

Then, keeping  $u_0$  fixed, we can repeat a similar procedure to find  $\Delta x$  for different values of  $\gamma$ . Also in this case, the solution either diverges or oscillates a part from a single value  $\Delta x^*$ , for which the solution has the desired shape (see Fig.D.2). This eliminates the arbitrariness in the choice of  $\Delta x$  since it imposes, for given  $g$  and  $w$ , a relation  $\Delta x = \Delta x^*(\gamma)$ .

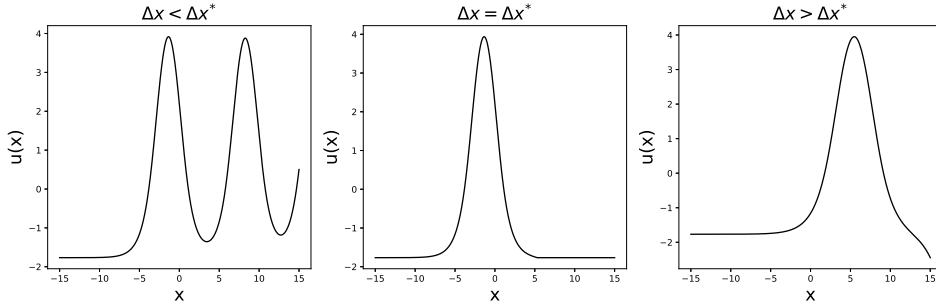


Figure D.2: Solutions to Eq.D.5 for  $g = 1, w = -1.8, \gamma = 0.2$ .

We can then find the shape of the bump  $u(x)$  for given values of  $g, w$  and  $\gamma$ , from which we can obtain the profile  $v(x) = g\mathcal{N}(u(x))$  that we need for the calculation of the storage capacity. Some examples of the obtained profiles, for different values of  $\gamma$ , are shown in Fig. D.3.

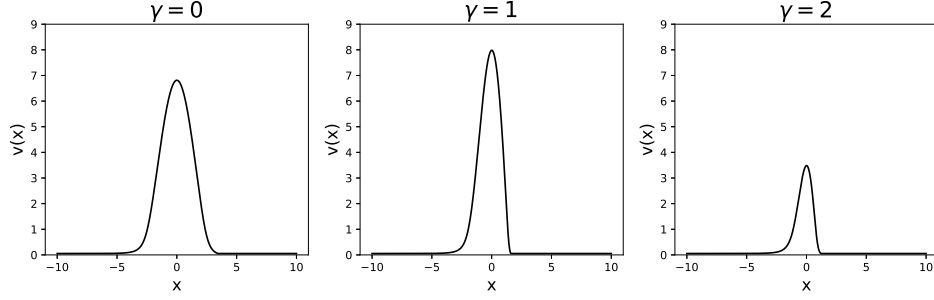


Figure D.3: Activity profile  $v(x)$ , obtained for the same  $g = 0.7$  and  $w = -1.3$  at different values of  $\gamma$ .

# Bibliography

- [1] Davide Spalla, Alexis Dubreuil, Sophie Rosay, Remi Monasson, and Alessandro Treves. Can grid cell ensembles represent multiple spaces? *Neural Computation*, 31(12):2324–2347, 2019.
- [2] Charlotte N Boccara, Francesca Sargolini, Veslemøy Hult Thoresen, Trygve Solstad, Menno P Witter, Edvard I Moser, and May-Britt Moser. Grid cells in pre-and parasubiculum. *Nature neuroscience*, 13(8):987–994, 2010.
- [3] Daniel L. Gebo. Vertebrate paleontology and evolution. by Robert L. Carroll, New York: W.H. Freeman and company. 1988. xiv + 698 pp., figures, appendix, index. \$52.95 (cloth). *American Journal of Physical Anthropology*, 77(1):135–135, 1988.
- [4] Alessandro Treves. Neural phase transitions that made us mammals. In Péter Érdi, Anna Esposito, Maria Marinaro, and Silvia Scarpetta, editors, *Computational Neuroscience: Cortical Dynamics*, pages 55–70, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.
- [5] Bradford C Dickerson and Howard Eichenbaum. The Episodic Memory System: Neurocircuitry and Disorders. *Neuropsychopharmacology*, 35(1):86–104, January 2010.
- [6] Larry R Squire, Arthur P Shimamura, and David G Amaral. Memory and the hippocampus. In *Neural models of plasticity*, pages 208–239. Elsevier, 1989.
- [7] Prasad SSV Vannemreddy and James L Stone. Sanger brown and edward schäfer before heinrich klüver and paul bucy: their observations on bilateral temporal lobe ablations. *Neurosurgical focus*, 43(3):E2, 2017.

- [8] Brenda Milner. Intellectual function of the temporal lobes. *Psychological bulletin*, 51(1):42, 1954.
- [9] Brenda Milner, Suzanne Corkin, and H-L Teuber. Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of hm. *Neuropsychologia*, 6(3):215–234, 1968.
- [10] Larry R Squire. Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological review*, 99(2):195, 1992.
- [11] Larry R Squire. The neuropsychology of human memory. *Annual review of neuroscience*, 5(1):241–273, 1982.
- [12] Larry R Squire, Neal J Cohen, and Joyce A Zoukounis. Preserved memory in retrograde amnesia: Sparing of a recently acquired skill. *Neuropsychologia*, 22(2):145–152, 1984.
- [13] James L McClelland, Bruce L McNaughton, and Randall C O’Reilly. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 102(3):419, 1995.
- [14] Lynn Nadel and Morris Moscovitch. Memory consolidation, retrograde amnesia and the hippocampal complex. *Current opinion in neurobiology*, 7(2):217–227, 1997.
- [15] Morris Moscovitch and Lynn Nadel. Consolidation and the hippocampal complex revisited: in defense of the multiple-trace model. *Current opinion in neurobiology*, 8(2):297–300, 1998.
- [16] D Marr. A theory of archicortical function. *Proc R Soc Lond B Biol Sci*, 262:23–81, 1971.
- [17] Bruce L McNaughton and Richard GM Morris. Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in neurosciences*, 10(10):408–415, 1987.
- [18] Alessandro Treves and Edmund T Rolls. Computational analysis of the role of the hippocampus in memory. *Hippocampus*, 4(3):374–391, 1994.



- [19] John O’Keefe and Jonathan Dostrovsky. The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 1971.
- [20] Torkel Hafting, Marianne Fyhn, Sturla Molden, May-Britt Moser, and Edvard I Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, 2005.
- [21] Trygve Solstad, Charlotte N Boccara, Emilio Kropff, May-Britt Moser, and Edvard I Moser. Representation of geometric borders in the entorhinal cortex. *Science*, 322(5909):1865–1868, 2008.
- [22] JB Ranck Jr. Head direction cells in the deep layer of dorsal pre-subiculum in freely moving rats. In *Society of Neuroscience Abstract*, volume 10, page 599, 1984.
- [23] Bruce L McNaughton, Francesco P Battaglia, Ole Jensen, Edvard I Moser, and May-Britt Moser. Path integration and the neural basis of the ‘cognitive map’. *Nature Reviews Neuroscience*, 7(8):663–678, 2006.
- [24] Emilio Kropff, James E Carmichael, May-Britt Moser, and Edvard I Moser. Speed cells in the medial entorhinal cortex. *Nature*, 523(7561):419–424, 2015.
- [25] Edvard I Moser, Emilio Kropff, and May-Britt Moser. Place cells, grid cells, and the brain’s spatial representation system. *Annu. Rev. Neurosci.*, 31:69–89, 2008.
- [26] Richard Morris. Developments of a water-maze procedure for studying spatial learning in the rat. *Journal of neuroscience methods*, 11(1):47–60, 1984.
- [27] Robert U Muller and John L Kubie. The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience*, 7(7):1951–1968, 1987.
- [28] Charlotte B Alme, Chenglin Miao, Karel Jezek, Alessandro Treves, Edvard I Moser, and May-Britt Moser. Place cells in the hippocampus: eleven maps for eleven rooms. *Proceedings of the National Academy of Sciences*, 111(52):18428–18435, 2014.
- [29] Edward C Tolman. Cognitive maps in rats and men. *Psychological review*, 55(4):189, 1948.

- [30] John O’keefe and Lynn Nadel. *The hippocampus as a cognitive map*. Oxford: Clarendon Press, 1978.
- [31] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.
- [32] Daniel J Amit, Hanoch Gutfreund, and Haim Sompolinsky. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Physical Review Letters*, 55(14):1530, 1985.
- [33] Kechen Zhang. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *Journal of Neuroscience*, 16(6):2112–2126, 1996.
- [34] Alexei Samsonovich and Bruce L McNaughton. Path integration and cognitive mapping in a continuous attractor neural network model. *Journal of Neuroscience*, 17(15):5900–5920, 1997.
- [35] Misha Tsodyks. Attractor neural network models of spatial maps in hippocampus. *Hippocampus*, 9(4):481–489, 1999.
- [36] Yoram Burak and Ila R Fiete. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput Biol*, 5(2):e1000291, 2009.
- [37] Mark C Fuhs and David S Touretzky. A spin glass model of path integration in rat medial entorhinal cortex. *Journal of Neuroscience*, 26(16):4266–4276, 2006.
- [38] Francesco P Battaglia and Alessandro Treves. Attractor neural networks storing multiple space representations: a model for hippocampal place fields. *Physical Review E*, 58(6):7738, 1998.
- [39] Rémi Monasson and Sophie Rosay. Crosstalk and transitions between multiple spatial maps in an attractor neural network model of the hippocampus: Phase diagram. *Physical review E*, 87(6):062813, 2013.
- [40] KiJung Yoon, Michael A Buice, Caswell Barry, Robin Hayman, Neil Burgess, and Ila R Fiete. Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature neuroscience*, 16(8):1077–1084, 2013.

- [41] Rishidev Chaudhuri, Berk Gerçek, Biraj Pandey, Adrien Peyrache, and Ila Fiete. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature neuroscience*, 22(9):1512–1520, 2019.
- [42] Karel Jezek, Espen J Henriksen, Alessandro Treves, Edvard I Moser, and May-Britt Moser. Theta-paced flickering between place-cell maps in the hippocampus. *Nature*, 478(7368):246–249, 2011.
- [43] John O’Keefe and Michael L Recce. Phase relationship between hippocampal place units and the eeg theta rhythm. *Hippocampus*, 3(3):317–330, 1993.
- [44] Céline Drieu and Michaël Zugaro. Hippocampal sequences during exploration: Mechanisms and functions. *Frontiers in cellular neuroscience*, 13:232, 2019.
- [45] H Freyja Ólafsdóttir, Daniel Bush, and Caswell Barry. The role of hippocampal replay in memory and planning. *Current Biology*, 28(1):R37–R50, 2018.
- [46] Matthew A Wilson and Bruce L McNaughton. Reactivation of hippocampal ensemble memories during sleep. *Science*, 265(5172):676–679, 1994.
- [47] György Buzsáki, Cornelius H Vanderwolf, et al. Cellular bases of hippocampal eeg in the behaving rat. *Brain Research Reviews*, 6(2):139–171, 1983.
- [48] Gyorgy Buzsaki, Zolt Horvath, Ronald Urioste, Jamille Hetke, and Kensall Wise. High-frequency network oscillation in the hippocampus. *Science*, 256(5059):1025–1027, 1992.
- [49] William E Skaggs and Bruce L McNaughton. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science*, 271(5257):1870–1873, 1996.
- [50] Zoltán Nádasdy, Hajime Hirase, András Czurkó, Jozsef Csicsvari, and György Buzsáki. Replay and time compression of recurring spike sequences in the hippocampus. *Journal of Neuroscience*, 19(21):9497–9507, 1999.

- [51] David J Foster and Matthew A Wilson. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440(7084):680–683, 2006.
- [52] Jadin C Jackson, Adam Johnson, and A David Redish. Hippocampal sharp waves and reactivation during awake states depend on repeated sequential experience. *Journal of Neuroscience*, 26(48):12415–12426, 2006.
- [53] David Dupret, Joseph O’neill, Barty Pleydell-Bouverie, and Jozsef Csicsvari. The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nature neuroscience*, 13(8):995, 2010.
- [54] Brad E Pfeiffer and David J Foster. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, 497(7447):74–79, 2013.
- [55] George Dragoi and Susumu Tonegawa. Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature*, 469(7330):397–401, 2011.
- [56] U Farooq and G Dragoi. Emergence of preconfigured and plastic time-compressed sequences in early postnatal development. *Science*, 363(6423):168–173, 2019.
- [57] Marianne Fyhn, Torkel Hafting, Alessandro Treves, May-Britt Moser, and Edvard I. Moser. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, (7132):190–194, 03 2007.
- [58] Hanne Stensola, Tor Stensola, Trygve Solstad, Kristian Frøland, May-Britt Moser, and Edvard I Moser. The entorhinal grid map is discretized. *Nature*, 492(7427):72, 2012.
- [59] Michael M Yartsev, Menno P Witter, and Nachum Ulanovsky. Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature*, 479(7371):103, 2011.
- [60] Stefan Leutgeb, Jill K Leutgeb, Carol A Barnes, Edvard I Moser, Bruce L McNaughton, and May-Britt Moser. Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science*, 309(5734):619–623, 2005.

- [61] Jonathan J Couey, Aree Witoelar, Sheng-Jia Zhang, Kang Zheng, Jing Ye, Benjamin Dunn, Rafal Czakowski, May-Britt Moser, Edvard I Moser, Yasser Roudi, et al. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nature neuroscience*, 16(3):318, 2013.
- [62] Tor Stensola, Hanne Stensola, May-Britt Moser, and Edvard I. Moser. Shearing-induced asymmetry in entorhinal grid cells. *Nature*, (7538):207–212, 02 2015.
- [63] Martin Hägglund, Maria Mørreaunet, May-Britt Moser, and Edvard I Moser. Grid-cell distortion along geometric borders. *Current Biology*, 2019.
- [64] Julija Krupic, Marius Bauza, Stephen Burton, Caswell Barry, and John O’Keefe. Grid cell symmetry is shaped by environmental geometry. *Nature*, (7538):232–235, 02 2015.
- [65] Julija Krupic, Marius Bauza, Stephen Burton, and John O’Keefe. Local transformations of the hippocampal cognitive map. 359(6380):1143–1146, 2018.
- [66] Charlotte N Boccara, Michele Nardin, Federico Stella, Joseph O’Neill, and Jozsef Csicsvari. The entorhinal cognitive map is attracted to goals. *Science*, 363(6434):1443–1447, 2019.
- [67] William N Butler, Kiah Hardcastle, and Lisa M Giocomo. Remembered reward locations restructure entorhinal spatial maps. *Science*, 363(6434):1447–1452, 2019.
- [68] Benjamin Dunn, Daniel Wennberg, Ziwei Huang, and Yasser Roudi. Grid cells show field-to-field variability and this explains the aperiodic response of inhibitory interneurons. *arXiv preprint arXiv:1701.04893*, 2017.
- [69] Dmitriy Aronov, Rhino Nevers, and David W Tank. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, 543(7647):719, 2017.
- [70] Alexandra O. Constantinescu, Jill X. O’Reilly, and Timothy E. J. Behrens. Organizing conceptual knowledge in humans with a gridlike code. *Science*, 352(6292), 2016.

- [71] Benjamin R Kanter, Christine M Lykken, Daniel Avesar, Aldis Weible, Jasmine Dickinson, Benjamin Dunn, Nils Z Borgesius, Yasser Roudi, and Clifford G Kentros. A novel mechanism for the grid-to-place cell transformation revealed by transgenic depolarization of medial entorhinal cortex layer ii. *Neuron*, 93(6):1480–1492, 2017.
- [72] Emilio Kropff and Alessandro Treves. The emergence of grid cells: Intelligent design or just adaptation? *Hippocampus*, 18(12):1256–1269, 2008.
- [73] Federico Stella, Bailu Si, Emilio Kropff, and Alessandro Treves. Grid cells on the ball. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(03):P03013, 2013.
- [74] Eugenio Urdapilleta, Francesca Troiani, Federico Stella, and Alessandro Treves. Can rodents conceive hyperbolic spaces? *Journal of the Royal Society Interface*, 12(107):20141214, 2015.
- [75] R. Monasson and S. Rosay. Crosstalk and transitions between multiple spatial maps in an attractor neural network model of the hippocampus: Collective motion of the activity. *Phys. Rev. E*, 89:032803, Mar 2014.
- [76] Kathryn R Hedrick and Kechen Zhang. Megamap: flexible representation of a large space embedded with nonspatial information by a hippocampal attractor network. *Journal of neurophysiology*, 116(2):868–891, 2016.
- [77] Ila R Fiete, Yoram Burak, and Ted Brookings. What grid cells convey about rat location. *Journal of Neuroscience*, 28(27):6858–6871, 2008.
- [78] Alessandro Treves. Threshold-linear formal neurons in auto-associative nets. *Journal of Physics A: Mathematical and General*, 23(12):2631, 1990.
- [79] Yasser Roudi and Alessandro Treves. An associative network with spatially organized connectivity. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(07):P07010, 2004.
- [80] Federico Stella, Eugenio Urdapilleta, Yifan Luo, and Alessandro Treves. Partial coherence and frustration in self-organizing spherical grids. *Hippocampus*, 30(4):302–313, 2020.

- [81] Noga Mosheiff and Yoram Burak. Velocity coupling of grid cell modules: stable embedding of a low dimensional variable in a high dimensional neural attractor. *bioRxiv*, page 651513, 2019.
- [82] Alessandro Treves, Ayumu Tashiro, Menno P. Witter, and Edvard I. Moser. What is the mammalian dentate gyrus good for? *Neuroscience*, 154(4):1155–1172, 2008.
- [83] Howard Eichenbaum. A cortical–hippocampal system for declarative memory. *Nature Reviews Neuroscience*, 1(1):41–50, 2000.
- [84] Edvard I Moser, May-Britt Moser, and Bruce L McNaughton. Spatial representation in the hippocampal formation: a history. *Nature neuroscience*, 20(11):1448, 2017.
- [85] May-Britt Moser, David C Rowland, and Edvard I Moser. Place cells, grid cells, and memory. *Cold Spring Harbor perspectives in biology*, 7(2):a021808, 2015.
- [86] Jeffrey S Taube, Robert U Muller, and James B Ranck. Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis. *Journal of Neuroscience*, 10(2):420–435, 1990.
- [87] Malcolm G Campbell and Lisa M Giocomo. Self-motion processing in visual and entorhinal cortices: inputs, integration, and implications for position coding. *Journal of neurophysiology*, 120(4):2091–2106, 2018.
- [88] Kiah Hardcastle, Niru Maheswaranathan, Surya Ganguli, and Lisa M Giocomo. A multiplexed, heterogeneous, and adaptive code for navigation in medial entorhinal cortex. *Neuron*, 94(2):375–387, 2017.
- [89] Case H Vanderwolf. Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalography and clinical neurophysiology*, 26(4):407–418, 1969.
- [90] Willard L McFarland, Herman Teitelbaum, and Elizabeth K Hedges. Relationship between hippocampal theta activity and running speed in the rat. *Journal of comparative and physiological psychology*, 88(1):324, 1975.
- [91] Robert W Stackman and Jeffrey S Taube. Firing properties of head direction cells in the rat anterior thalamic nucleus: dependence on vestibular input. *Journal of Neuroscience*, 17(11):4349–4358, 1997.

- [92] Joshua P Bassett and Jeffrey S Taube. Neural correlates for angular head velocity in the rat dorsal tegmental nucleus. *Journal of Neuroscience*, 21(15):5740–5751, 2001.
- [93] Sung Soo Kim, Hervé Rouault, Shaul Druckmann, and Vivek Jayaraman. Ring attractor dynamics in the drosophila central brain. *Science*, 356(6340):849–853, 2017.
- [94] Daniel Turner-Evans, Stephanie Wegener, Herve Rouault, Romain Franconville, Tanya Wolff, Johannes D Seelig, Shaul Druckmann, and Vivek Jayaraman. Angular velocity integration in a fly heading circuit. *Elife*, 6:e23496, 2017.
- [95] William E Skaggs, Bruce L McNaughton, and Katalin M Gothard. An information-theoretic approach to deciphering the hippocampal code. In *Advances in neural information processing systems*, pages 1030–1037, 1993.
- [96] Max L Mehlman, Shawn S Winter, and Jeffrey S Taube. Functional and anatomical relationships between the medial precentral cortex, dorsal striatum, and head direction cell circuitry. ii. neuroanatomical studies. *Journal of neurophysiology*, 121(2):371–395, 2019.
- [97] Longtang L Chen, Lie-Huey Lin, Edward J Green, Carol A Barnes, and Bruce L McNaughton. Head-direction cells in the rat posterior cortex. *Experimental brain research*, 101(1):8–23, 1994.
- [98] Yave Roberto Lozano, Hector Page, Pierre-Yves Jacob, Eleonora Lomi, James Street, and Kate Jeffery. Retrosplenial and postsubicular head direction cells compared during visual landmark discrimination. *Brain and neuroscience advances*, 1:2398212817721859, 2017.
- [99] Mateo Vélez-Fort, Edward F Bracey, Sepiedeh Keshavarzi, Charly V Rousseau, Lee Cossell, Stephen C Lenzi, Molly Strom, and Troy W Margrie. A circuit for integration of head-and visual-motion signals in layer 6 of mouse primary visual cortex. *Neuron*, 98(1):179–191, 2018.
- [100] Patricia Preston-Ferrer, Stefano Coletta, Markus Frey, and Andrea Burgalossi. Anatomical organization of presubicular head-direction circuits. *Elife*, 5:e14592, 2016.
- [101] Shawn S Winter, Max L Mehlman, Benjamin J Clark, and Jeffrey S Taube. Passive transport disrupts grid signals in the parahippocampal cortex. *Current Biology*, 25(19):2493–2502, 2015.



- [102] Jeffrey L Calton and Jeffrey S Taube. Degradation of head direction cell activity during inverted locomotion. *Journal of Neuroscience*, 25(9):2420–2428, 2005.
- [103] Neil Burgess, Caswell Barry, and John O’keefe. An oscillatory interference model of grid cell firing. *Hippocampus*, 17(9):801–812, 2007.
- [104] Michael E Hasselmo, Lisa M Giocomo, and Eric A Zilli. Grid cell firing may arise from interference of theta frequency membrane potential oscillations in single neurons. *Hippocampus*, 17(12):1252–1271, 2007.
- [105] Ipshita Zutshi, Mark P Brandon, Maylin L Fu, Macayla L Donegan, Jill K Leutgeb, and Stefan Leutgeb. Hippocampal neural circuits respond to optogenetic pacing of theta frequencies by generating accelerated oscillation frequencies. *Current Biology*, 28(8):1179–1188, 2018.
- [106] Joseph O’Neill, Barty Pleydell-Bouverie, David Dupret, and Jozsef Csicsvari. Play it again: reactivation of waking experience and memory. *Trends in neurosciences*, 33(5):220–229, 2010.
- [107] Mattias P Karlsson and Loren M Frank. Awake replay of remote experiences in the hippocampus. *Nature neuroscience*, 12(7):913–918, 2009.
- [108] Federico Stella, Peter Baracska, Joseph O’Neill, and Jozsef Csicsvari. Hippocampal reactivation of random trajectories resembling brownian diffusion. *Neuron*, 102(2):450–461, 2019.
- [109] George Dragoi and Susumu Tonegawa. Distinct preplay of multiple novel spatial experiences in the rat. *Proceedings of the National Academy of Sciences*, 110(22):9100–9105, 2013.
- [110] Igor Gridchyn, Philipp Schoenenberger, Joseph O’Neill, and Jozsef Csicsvari. Assembly-specific disruption of hippocampal replay leads to selective memory deficit. *Neuron*, 2020.
- [111] Kenneth Kay, Jason E Chung, Marielena Sosa, Jonathan S Schor, Mattias P Karlsson, Margaret C Larkin, Daniel F Liu, and Loren M Frank. Constant sub-second cycling between representations of possible futures in the hippocampus. *Cell*, 180(3):552–567, 2020.
- [112] Alex P Vaz, John H Wittig, Sara K Inati, and Kareem A Zaghloul. Replay of cortical spiking sequences during human memory retrieval. *Science*, 367(6482):1131–1134, 2020.

- [113] Vanessa E Ghosh and Asaf Gilboa. What is a memory schema? a historical perspective on current neuroscience literature. *Neuropsychologia*, 53:104–114, 2014.
- [114] Pierre Baraduc, J-R Duhamel, and S Wirth. Schema cells in the macaque hippocampus. *Science*, 363(6427):635–639, 2019.
- [115] Haim Sompolinsky and I Kanter. Temporal association in asymmetric neural networks. *Physical review letters*, 57(22):2861, 1986.
- [116] Emilio Kropff and Alessandro Treves. The storage capacity of potts models for semantic memory retrieval. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(08):P08010, 2005.
- [117] Eleonora Russo, Vijay MK Nambodiri, Alessandro Treves, and Emilio Kropff. Free association transitions in models of cortical latching dynamics. *New Journal of Physics*, 10(1):015008, 2008.
- [118] Chol Jun Kang, Michelangelo Naim, Vezha Boboeva, and Alessandro Treves. Life on the edge: latching dynamics in a potts neural network. *Entropy*, 19(9):468, 2017.
- [119] Henry Markram, Joachim Lübke, Michael Frotscher, and Bert Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297):213–215, 1997.
- [120] Weishun Zhong, Zhiyue Lu, David J Schwab, and Arvind Murugan. Nonequilibrium statistical mechanics of continuous attractors. *Neural Computation*, 32(6):1033–1068, 2020.
- [121] Mayank R Mehta, Carol A Barnes, and Bruce L McNaughton. Experience-dependent, asymmetric expansion of hippocampal place fields. *Proceedings of the National Academy of Sciences*, 94(16):8918–8921, 1997.
- [122] Patricia E Sharp. Computer simulation of hippocampal place cells. *Psychobiology*, 19(2):103–115, 1991.
- [123] Vezha Boboeva, Romain Brasselet, and Alessandro Treves. The capacity for correlated semantic memories in the cortex. *Entropy*, 20(11):824, 2018.
- [124] Christian F Doeller, Caswell Barry, and Neil Burgess. Evidence for grid cells in a human memory network. *Nature*, 463(7281):657–661, 2010.

- [125] Jacob LS Bellmund, Peter Gärdenfors, Edvard I Moser, and Christian F Doeller. Navigating cognition: Spatial codes for human thinking. *Science*, 362(6415), 2018.
- [126] John B Calhoun. *The ecology and sociology of the Norway rat*. Number 1008. US Department of Health, Education, and Welfare, Public Health Service, 1963.
- [127] Alessandro Treves and Edmund T Rolls. What determines the capacity of autoassociative memories in the brain? *Network: Computation in Neural Systems*, 2(4):371–397, 1991.