POLITECNICO DI TORINO

Corso di Laurea Magistrale
in Physics of Complex Systems

Tesi di Laurea Magistrale

A model of joint hippocampal and cortical representation of space and context

Relatori
prof. Johnatan Aljadeff
prof. Andrea Pagnani

Candidato
Fabio Veneto

Anno Accademico 2021-2022

Summary

Manipulating sensory cues in the environment or changing the animal's task sometimes leads to hippocampal remapping of the spatial representation: a phenomenon whereby the preferred location of all place-cells changes [global remapping (Muller and Kubie 1987)]. Recent evidence in humans suggests that hippocampal remapping is correlated with switches between different behaviours (Julian and Doeller 2021). Remapping can thus be viewed as evidence of recognizing that "something has changed enough" to require using a different spatial representation. Despite extensive research into the conditions and mechanisms of pattern separation in the hippocampus, the rules that determine whether a specific manipulation will cause remapping are currently unknown. Recently, it was discovered that a phenomenon akin to global remapping occurs even when no manipulation is made to the environment, to the state of the animal or to the behaviour it performs (Sheintuch et al. 2020).

The aim of this thesis is to uncover principles underlying mixed of contextual information with hippocampal spatial representations.

In the first part of the thesis, we studied a well-known model frequently used to describe remapping phenomena in the hippocampus: the Hopfield model. We initially analysed this model because, even though the steady state description has been well studied (Hopfield 1982), the transient dynamics was analysed only in particular cases. The aim was to develop a useful method that would have allowed us to efficiently study an extension of the Hopfield model.

We analysed the dynamics initialised on a line connecting two attractors, where, in this scenario, each attractor represented a different environment. We developed an analogous 1-dimensional model to describe the dynamics along this line.

In the second part, we built a Hopfield-type model which couples together different patterns in order to represent the same environment in different contexts. Our assumption was that if the cortical inputs to the hippocampus are precise enough to initialize the network in the "correct" basin of attraction, then map multiplicity, discovered by Sheintuch et al. 2020, was not random but rather represented contextual variables.

Firstly, we analysed the steady-state of the generalized Hopfield model. We showed the emergence of multiple fixed points within the same environment. Ultimately, we modelled the dynamics along different lines connecting all the fixed points within the same environment and across different environments with a stochastic machine.

This model allowed us to obtain information on the precision and the dimensionality of inputs, which the hippocampus receives from cortical areas, that can induce transitions between attractors and, analogously, between different contextual representations.

Our model of transitions between different representations can be interpreted as a first step to understand the minimal mechanisms supporting the process of learning mixed spatial and contextual representations.

Acknowledgements

I wish to express my sincere gratitude to my supervisor, Professor Johnatan Aljadeff, for offering me the amazing opportunity to join his lab and for being so patient and supportive during this exciting work. Thank you for leading me into the beauty of science and research.

I would also like to thank my family that always believed and encouraged me in my adventures. None of this would have been possible without you.

Contents

Chapter 1

Introduction

Memory and sensory cues from multiple modalities allow real life navigation which involves dynamic transitions between multiple environments. The seminal work of O'Keefe and Dostrovsky 1971 and O'Keefe and Nadel 1978 demonstrated that the hippocampus maintains a "cognitive map": the network exhibits neuronal activity patterns in a location specific manner. Multiple experimental results (Aronov, Nevers, and Tank 2017, Radvansky et al. 2021, Ito et al. 2015, O'Keefe 1999, Sheintuch et al. 2020) describe a rich phenomenology where spatial and contextual representations are mixed ("multiplexed") in the hippocampus. The switches made by the brain between spatial maps, as a result of changes in non-spatial variables, make these representations distinctive. These may underlie animals' ability to perform different tasks while also navigating, e.g. working memory and sensory discrimination (Ito et al. 2015, Nieh et al. 2021). The way cross-talk between cortical and hippocampal networks fuses spatial and contextual information is considered a major knowledge gap. Specifically the factors, which determine whether hippocampal spatial representations will remain either stable to support perceptual permanence of the environment or remap to support identification of changes to it, have limited understanding.

1.1 Experimental results

1.1.1 Place cells

O'Keefe and Dostrovsky 1971 discovered the existence of "place cells". They claimed that the hippocampus was the anatomical locus of a "cognitive map", i.e. a holistic neural representation of the environment that permits rats and mice to efficiently solve spatial problems. Anatomically, place cells are pyramidal neurons of the hippocampus in rats (O'Keefe 1979) and mice (Rotenberg et al. 1996, McHugh et al. 1996). They are found in both the CA1 and CA3 regions of the hippocampus (O'Keefe 1979). Recordings of place cells have been performed in both the dorsal (septal) hippocampus and the ventral (temporal) hippocampus, suggesting that the entire structure participates in mapping (Poucet, Thinus-Blanc, and Muller 1994, Jung, Wiener, and McNaughton 1994).

Functionally, each place cell is selectively active only when the rat is in a certain part of the environment, called “firing field” or “place field". In other words, they are characterized by location-specific firing. When the rat is outside the field, the firing rate of that place cell is zero or indistinguishable from the baseline activity (Muller 1996).
Figure 1.1 shows an illustration of the firing rate properties of a representative place cell from region CA1 of the hippocampus. Firing fields are cell specific. In a fixed environment,

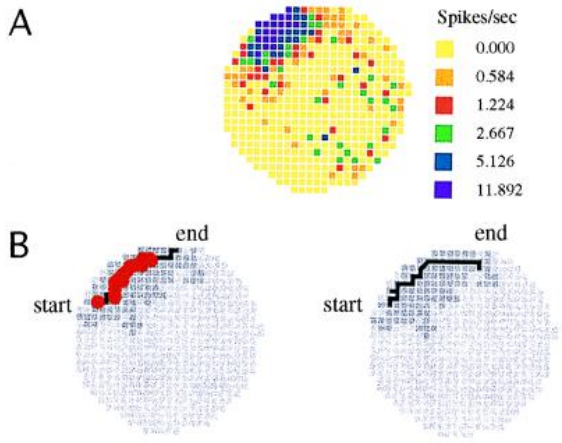


Figure 1.1. (A) Color-coded map of the positional firing rate of a place cell for a 16 minutes session; the circular area is an overhead view of a cylindrical apparatus 76 cm in diameter with a 51 cm high wall. The firing field is the dark region at the top left of the figure. Outside the firing field, the cell fires sporadically. The yellow pixels encode firing rate equal to zero.
(B) The two diagrams show individual passes of the rat through the field; the paths are black lines and action of the cell are depicted by red dots (image from Fenton and Muller 1998).

each place cell has a stable field that is characteristic of that particular place cell. Taking into account all the place cells together, O’Keefe and Dostrovsky 1971 hypothesized the existence of a "cognitive map" in the hippocampus.

1.1.2 Remapping

It has been discovered that the firing patterns of place cells were frequently altered in response to changes in sensory or cognitive inputs (Muller, Kubie, et al. 1991, Lever et al. 2002, Leutgeb et al. 2005). Such changes in firing activity constitute a ‘remapping’ of the place cell representation of space. This phenomenon is usually named "global remapping" (Latuske et al. 2018). Place fields and firing rates change drastically such that the activity patterns of hippocampal place cells observed in two different environments are not correlated (Muller and Kubie 1987). Remapping can thus be viewed as evidence of recognizing that “something has changed enough in the environment” to require using a different spatial representation, i.e. a different set of associations between sensory inputs and neuronal outputs.

Figure 1.2 shows an example simultaneous recording of place cells that remapped in response to a change in a sensory cue. It has recently been discovered that a phenomenon

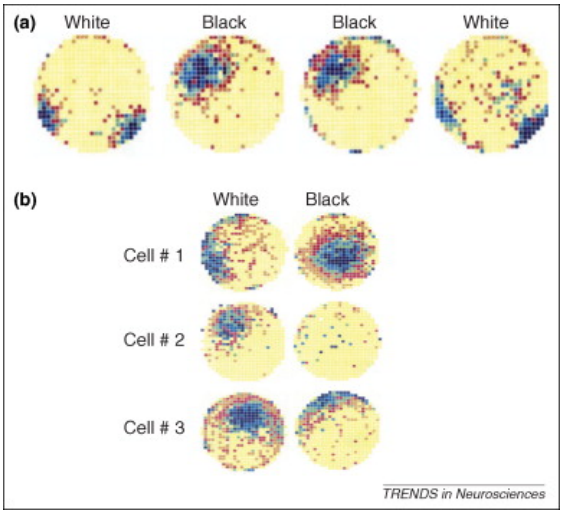


Figure 1.2. Firing rate maps for hippocampal place cells recorded in a cylinder containing either a white or a black intra-maze cue card.
(a) An example place cell recorded across four sessions in a cylinder with either the white or black cue card (as indicated above).
(b) Three additional cells recorded in the cylinder for white and black cue card sessions. Place fields changed location or disappeared in response to the cue card substitution (Colgin, E. Moser, and M. Moser 2008,Bostock, Muller, and Kubie 1991).

akin to global remapping occurs even when no manipulation is made to the environment, to the state of the animal or to the behavior it performs (Sheintuch et al. 2020). In other words, these reversible transitions between population-level modes of activity appear to be random. Subsequent studies from other labs have reproduced similar findings for different areas of the brain (Low et al. 2021 in the entorhinal cortex).
In the work by Sheintuch et al. 2020, hippocampal place cells were imaged as mice explored the same environment, over many repeated visits, both within the same session, and across multiple sessions spanning weeks. More precisely, each imaging session consisted of multiple separate three-minute visits to the same environment. During each particular visit, each mouse ran freely back and forth to collect water rewards at both ends of the track. With a three-minute inter-trial interval, the mice were placed in an opaque bucket near the track, in order to disconnect them from the testing environment. Overall, for each mouse, they collected data from $25 - 40$ trials within the same environment.
They found that multiple stable maps coexistence over weeks in the mouse hippocampus. Once a given map was retrieved, it persisted throughout the trial until the mouse was dissociated from the environment. In this way, the network reset, which possibly allowed it to switch to a different spatial representations upon entering the linear track on the next visit.
Furthermore, they claimed that a low variability in the network's initial conditions could

result in convergence to the same attractor on different visits, which might lead to the stabilization of a single attractor state. Conversely, an higher variability could result in convergence to different attractors, which might lead to the stabilization of multiple separate attractor states, corresponding to multiple stable maps of the same spatial context.

1.1.3 Hypothesis

Despite the extensive research into spatial maps in the hippocampus, the conditions underwhich remapping occurs are still not well understood. Recent evidence in humans suggests that hippocampal remapping is strongly correlated with switches between different behaviors (Julian and Doeller 2021). Remapping can thus be interpreted as a major change in the environment that requires using a different spatial representation.
Nonetheless, Sheintuch et al. 2020 experiment suggests that remapping may still occur without any spatial manipulation of the environment. Subsequently, we hypothesize that remapping occurs due to a major change in the animal's internal state, which is represented in the pre-frontal cortex. We interpret it as a switch of contextual information which leads to a joint hippocampal and cortical representation of space and context. We further hypothesize that attractor dynamics in the hippocampus support efficient multiplexed representations of spatial and behaviorally relevant contextual information.
Furthermore, Sheintuch et al. 2020 experiment suggests that the variability on the network initial state might be responsible for the existence of multiple representation in the hippocampus. Consequently, we hypothesize that cortical inputs endow hippocampal neuronal populations with a reliable encoding of contextual variables. That is to say, the pre-frontal cortex is responsible for setting the network in the 'correct' initial condition that allows the attractor dynamics to converge on the desired spatial and contextual representation of the state.
We built a theoretical attractor model of mixed hippocampal representations, constrained with Sheintuch et al. 2020 experimental data (chapter 3).

1.2 Previous modelling

Evidence from experimental studies suggests that hippocampal circuits generate attractor dynamics, i.e. specific patterns of neuronal activity which are: stable to perturbations; correlated with variables in the external world (e.g., the animal's position); persistent in the absence of sensory input (Knierim and Neunuebel 2016). Despite this progress, the link between attractor dynamics and hippocampal representations of non-spatial information is missing. In other words, the rules that determine whether a specific manipulation will cause remapping ("jumping to a different attractor") are currently unknown.
The most adopted model to describe hippocampal attractor dynamics is the Hopfield model. It was originally proposed by Hopfield 1982 and it provides a useful and successful formalism for understanding attractor dynamics. We will describe in more details this model in chapter 2. However, it is important to mention that the Hopfield model is not consistent with the results obtained by Sheintuch et al. 2020. For example, it is indeed

impossible to guarantee that changing the animal's task without changing the environment will result in a state representing that specific environment. Furthermore, each environment is reduced to a single point. In chapter 4, we will discuss how to extend the fixed-point attractor to a continuous attractor model by following the formalism proposed by Battaglia and Treves 1998. To solve this discrepancy, in previous literature, different modification of the Hopfield model were considered. We mention the work by Haga and Fukai 2019 where they introduced cross-coupling between every pair of "neighboring states", based on their position in a sequence. By contrast, in this thesis, we will propose a different approach which cross-couples only the "within-environment" states (chapter 3).

Chapter 2

Hopfield network

Our aim is to understand whether the switches between neuronal spatial representations can arise from attractor dynamics of a generalized Hopfield model. Before studying the extended model describing the coexistence of multiple maps for each single environment representing different contexts, we studied a well-known model frequently used to describe remapping phenomena in the hippocampus: the **Hopfield model**.
We initially analyze this model because, even though the steady state description has been well studied (Hopfield 1982, Wilson and McNaughton 1993, Hopfield 1984), the transient dynamics was analyzed only in particular cases.
As we discussed in subsection 1.1.3, we assume that cortical inputs to the hippocampus are precise enough to initialize the network in the "correct" basin of attraction. We hypothesize that relatively small amounts of noise in specific directions can be sufficient to induce transitions between different contexts. Conversely, in other directions of the activity space, we assume that the network is more robust to noise. Therefore, we analyze the dynamics initialised on a line connecting two attractors. If the pre-frontal cortex initialization of the network is reliable, then we expect it to induce transitions between different contexts.
Nonetheless, as we will discuss in the following chapter, the Hopfield network does not explain the coexistence of multiple contexts for each single environment. The attractors are uncorrelated with one another and, therefore, they produce uncorrelated representations of the same environment, which is consistent with the lack of correlations between spatial representations of different environments. Consequently, in the Hopfield model, we interpret each pattern as a population level representation of a single environment. To remedy this discrepancy, we latterly extend the same approach when attractors are coupled with each other (chapter 3).

2.1 Steady State description

The steady state description of the Hopfield network has been extensively studied. In the following subsections, we will briefly analyze all the known results, such as the storage capacity and the energy function of the model; subsequently, in the next section, the

unknown transient dynamics on the line connecting two attractors.

2.1.1 The Model

The goal is to store a set of p patterns ξ_i^μ, where $\mu = 1,2,\ldots,p$. When introduced with a new pattern, the network responds by producing the stored pattern that most closely resembles it. Each pattern in the network is composed by units which we label by $i = 1,2,\ldots,N$, where N is the total number of neurons in the network. Each one of the units x_i can be either $+1$ (representing an active neuron whose firing rate is above a certain threshold) or -1 (inactive neuron).
Within the configuration space described by all the possible states of the network, the stored patterns ξ^μ are attractors of the network.
The dynamics of the network is described by the following update rule

$$x_i(t+1) = sgn\left(\sum_{j=1}^{N} J_{ij}x_j(t)\right) \tag{2.1}$$

where $sgn(x)$ is the sign function defined as

$$sgn(x) = \begin{cases} 1 & if\ x \geq 0 \\ -1 & if\ x < 0 \end{cases}$$

and $x_i(t) \in \{-1,+1\}$ represents the state of neuron i at time t. The update will be done *asynchronously*, i.e. updating each x_i one at the time. Asynchronous update is preferred since not all the neurons have the same fixed delay ($t \to t+1$), nor they are updated synchronously by a central clock. At each time step, we select at random a unit x_i to be updated, and apply the rule in Equation 2.1. The network eventually settles into a stable configuration, one for which no x_i changes (Hertz, Krogh, and Palmer 1991). Technical details on how this is accomplished can be found in section 2.2.1.
To store the patterns ξ^μ , the connectivity matrix J_{ij} is defined based on the following rule

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu \tag{2.2}$$

This mimics "Hebbian learning", since it can be interpreted as a weight term for which the connection between neuron i and neuron j switches from excitatory to inhibitory and vice versa as more patterns are added.
This model presents a well-defined energy function as well

$$E_{ij} = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} J_{ij}x_i x_j \tag{2.3}$$

The combination of the update rule Equation 2.1 and 2.2 is usually named **Hopfield network**.

2.1.2 Storage Capacity

We defined a dynamical system that has attractors at the desired points ξ_i^μ, but those are not the only attractors. The network also exhibits: reversed states, mixture states and spin-glass states.
The reversed states $-\xi_i^\mu$ are minima and have the same energy as the original patterns since both the dynamics and the energy function have a perfect symmetry $\forall i$. Mixture states correspond to linear combinations of an odd number of patterns (Amit, Gutfreund, and Sompolinsky 1985a). For instance, for $p = 3$, the network will also store

$$\xi_i^{mix} = sgn\left(\pm\xi_i^{\mu_1} \pm \xi_i^{\mu_2} \pm \xi_i^{\mu_3}\right)$$

Similarly any number of odd patterns may be combined. The system does not choose an even number of patterns because they can add up to zero on some sites, whereas the units x_i have to be ± 1 (Hertz, Krogh, and Palmer 1991).
For large p, spin-glass states are local minima that are not correlated with any finite number of the original patterns ξ^μ (Amit, Gutfreund, and Sompolinsky 1985b).
Generally, mixture states and spin-glass states are called spurious states. They tend to have rather small basins of attraction compared to the retrieval states and it can be reduced even more when $N \gg p$, i.e. for small α (Hertz, Krogh, and Palmer 1991).
It is useful to define a load parameter

$$\alpha = \frac{p}{N} \tag{2.4}$$

which describes how many patterns can be stored in the network. It can be evaluated analytically, using a mean field approximation (Amit, Gutfreund, and Sompolinsky 1985a), that, for large α, the network does not retrieve correctly the stored memories anymore. The theoretical value obtained by Amit, Gutfreund, and Sompolinsky 1985a, for the Hopfield network with the Hebbian learning rule, in the large N limit, is $\alpha_{max} = 0.138$. Numerically, we can obtain the same result by imposing perfect memory retrieval, setting a threshold $P_{err} < \frac{1}{N}$ to describe the average difference between the retrieved and the stored pattern (Krotov and Hopfield 2016). Nonetheless, the network has finite size N therefore P_{err} would not reproduce the theoretical result. Consequently, we set an arbitrary small threshold $P_{err} = 0.02$ to evaluate α_{max} (Hertz, Krogh, and Palmer 1991). The result is shown in Figure 2.1. Here, we show the numerical tool to evaluate the maximum capacity since it will be applied again in chapter 3.

2.2 Transient dynamics

We are interested in studying the dynamics on the line connecting two attractors. All attractors are built statistically equivalent, therefore we analyze the dynamics in pairs, for simplicity 'Pattern 1' and 'Pattern 2'.
The following simulations are all performed with a fixed number of neurons $N = 1000$. We study the effect of the various free parameters of the model and obtain an equivalent 1-dimensional dynamical system. The single dimension is the line connecting the two

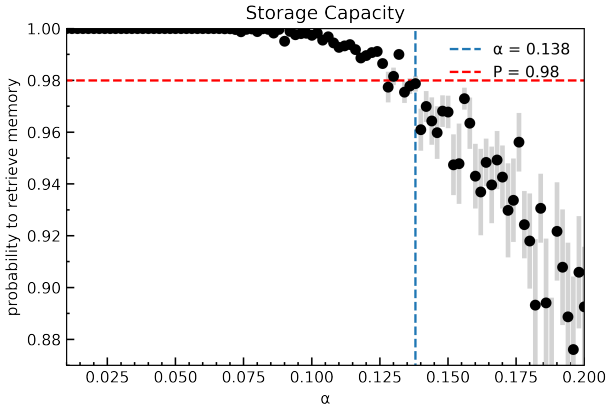


Figure 2.1. Probability to retrieve a memorized pattern depending on parameter α (2.4). The blue dashed line is the theoretical value of maximum capacity $\alpha = 0.138$. Numerically the memorized pattern was considered incorrect when the pattern obtained was different from the stored memory by a factor 0.02 (red dashed line).

patterns. For the future application in chapter 3, this can be qualitatively interpreted as the line where the cortical inputs to the hippocampus act to initialize the network in the "correct" basin of attraction.

2.2.1 Dynamics parameters dependence

Initial condition

As mentioned before, we want to study the dynamics on the line connecting two attractors. We define it as

$$x_i\,(t=0) = (1-\epsilon)\,\xi_i^{\mu_1} + \epsilon\,\xi_i^{\mu_2} \qquad \forall i \tag{2.5}$$

where $\epsilon \in [0,1]$. Differently from Equation 2.1, where $x_i \in \{-1,+1\}$, here $x_i \in [-1,+1]$. Introducing a non-linearity to map all x_i into binary values would not have allowed us to explore all the values along the line connecting the two attractors. Changing ϵ is then equivalent to move the initial configuration $x(t=0)$ between the two attractors. Explicitly, for $\epsilon \sim 0$ and $\epsilon \sim 1$, the probability to converge, respectively, to ξ^1 and ξ^2 is large. For $\epsilon = 0.5$, the probability to converge to ξ^1 is equal to the probability to converge to ξ^2. In order to precisely measure this effect, we evaluate the projection of the evolving configuration $x(t)$ on the stored patterns ξ^μ. We can define it as

$$z(t) = \frac{1}{N}\,(x(t)\cdot\xi^\mu) \tag{2.6}$$

Even though we fixed the initial condition, the network has an intrinsic noise which is due to *asynchronous* update. The order with which we update the single neurons x_i leads to different dynamics, as shown in Figure 2.2. The first approach, studied to carry out the update specified by Equation 2.1, is selecting a certain random permutation for each

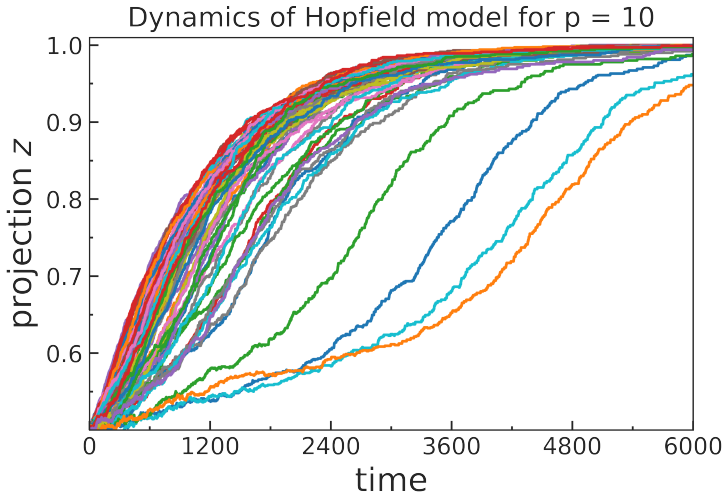


Figure 2.2. Multiple trajectories starting in the same pattern configuration ξ^μ and initial condition defined by ϵ. The effect of the asynchronous update is shown above. The order with which the neurons are updated influence the dynamics but not the retrieved memory.

simulation and then update the neurons in that order. This approach is leading to a linear-like dynamics of z with respect to t. In other words, these asynchronous updates without replacement adds variability similar to synaptic delay.
We then introduce replacements in the update rule. At each time step, we select at random a unit x_i to be updated, and apply the rule 2.1. This leads to the dynamics shown in Figure 2.2. Since the probability to update a certain neuron x_i is lower with replacements, the simulation time is longer as well. These asynchronous updates with replacement are like synaptic delays with a heavier tail.
All the simulations we discuss from now on are performed using an asynchronous update rule with replacements.
For the sake of completeness, another strategy to obtain a similar dynamics to the one shown in Figure 2.2 is modifying the update rule as follows

$$x_i(t+1) = \tanh\left(g \sum_{j=1}^{N} J_{ij} x_j(t)\right)$$

The update is performed again asynchronously but without the need of replacements. However, this approach introduces a new parameter $g \in \mathbb{R}$. In order to get an exponential behaviour which we can model, we need to fit, for each p, g accordingly. Therefore, the previous approach is more convenient.
We also study whether there is an effect due to orthogonal directions (with respect to the main direction) that are not included in Equation 2.5. Explicitly

$$x_i\,(t=0) = \left(\frac{1-\epsilon_1}{2}\right)\xi_i^{\mu_1} + \frac{\epsilon_1}{2}\,\xi_i^{\mu_2} + \frac{\epsilon_2}{p-2}\sum_{k=3}^{p}\xi_i^k \qquad \forall i$$

where $\epsilon_1 \in [0,1]$ and $\epsilon_2 \in [0,1]$. Here, ϵ_1 has the same role as ϵ in the previous initial condition, while ϵ_2 is now keeping into account the other $N-1$ directions. For any value of ϵ_2, there are no meaningful effect on the dynamics. We assume that the direction between the two attractors is indeed the principal direction. This somehow falls in the known picture of Hopfield basin of attraction for which each attractor has a certain "pulling area".
All the simulations shown from now on are performed with the initial condition defined in Equation 2.5 and the asynchronous update rule with replacements.
As shown in subsection 2.1.2, α characterizes the memory load of the network. Since we fixed the number of neurons N, the number of stored patterns p will be our main focus. This will be studied in the next section.

2.2.2 1-dimensional approach

Our goal is to create an equivalent dynamical system capable of capturing the main features of the transient dynamics of the Hopfield network. Subsequently, apply the obtained results to study the extension of the Hopfield model.

In the Hopfield model, the state of the network is an N dimensional vector quantity. A set of initial conditions along orthogonal directions had no effect on convergence, which justifies the assumption that the system is effectively one dimensional. Consistently, we study the projection z defined in Equation 2.6. The convergence of the neural state to the attractor is analogous to a ball subject to a potential energy function. We can write the energy conservation for that kind of classical mechanics problems as

$$U \coloneqq mgh = \frac{1}{2}mv^2$$

where m and g are emptied of their classical meaning, as mass and acceleration, and are only used as constants to fit the data. We want to obtain an equivalent representation for the velocity v from which derive a potential energy and, ultimately, an equivalent ODE.

First of all, we notice that $z(t) \sim \frac{\epsilon}{N}$. For instance at $t = 0$, assuming perfect orthogonality between patterns (true in the limit of $N \to \infty$ and p finite or, vice versa, for $p \to \infty$ and N finite), the projection on Pattern 2 is the following

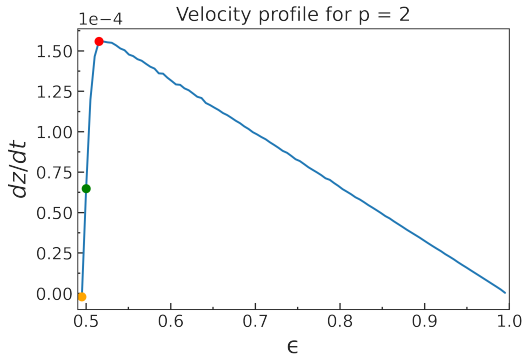
$$\begin{aligned} z(t=0) &= \frac{1}{N}\left(x(t=0)\cdot\xi^{\mu_2}\right) \\ &= \frac{1}{N}\left((1-\epsilon)\xi^{\mu_1} + \epsilon\xi^{\mu_2}\right)\cdot\xi^{\mu_2} \\ &= \frac{\epsilon}{N} \end{aligned}$$

where we used the definition in Equation 2.5 and the definition of orthogonality

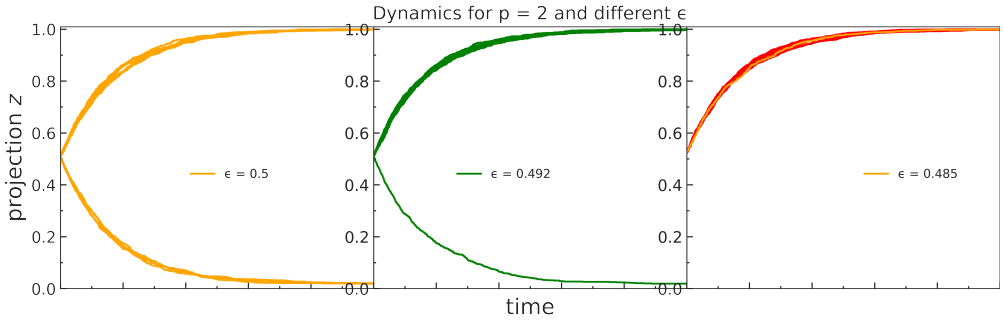
$$\xi^{\mu_i}\cdot\xi^{\mu_j} = \delta_{ij} = \begin{cases} 1 & if \quad i = j \\ 0 & otherwise \end{cases}$$

Numerically, it can be shown that this holds at any time t. Therefore, given a certain ξ^{μ} and an initial condition on the direction defined by Equation 2.5, the dynamics to

reach convergence follows a similar average path independently of the initial condition. Therefore we can consider $z(t)$ and ϵ as equivalent variables.
Furthermore, if we consider ϵ as a position, then $\frac{d\epsilon}{dt} \propto \frac{dz}{dt} := v$. The result is shown in Figure 2.3



(a) velocity profile



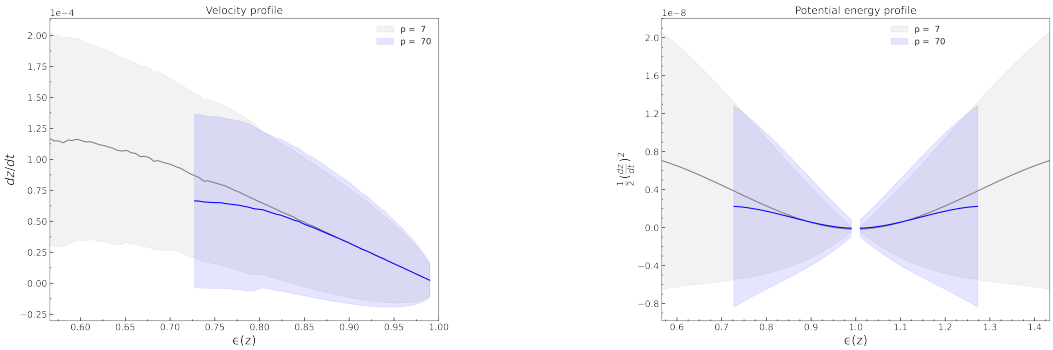
(b) multiple dynamics

Figure 2.3. (a) Average velocity profile for $p = 2$. The colored dots represent specific initial conditions for which the dynamics is shown below.
(b) Multiple dynamics for different values of initial condition chosen accordingly to their position on the velocity profile. When the velocity is maximum (red curves), there is $Prob = 0$ to escape from the closest pattern. As for velocities slightly below, the probability to escape convergence from the closest pattern increases further (green curves). For $\epsilon = 0.5$, 50% of the trajectories converge to pattern 1 and 50% to pattern 2 (orange curves).

Observation 1 *We do not observe spurious states in Figure 2.3 since, in this case,* $\alpha = \frac{2}{1000} = 0.002$*, therefore two order of magnitude smaller than maximum capacity.*

Even though it falls outside the experimental neuroscience relevance, where the animals are usually trained in a small number of environments, for sake of completeness, we extend this approach from any value of p, even close to maximum capacity. In Figure 2.4(a), it

is shown the velocity profile both for a small p and for p close to maximum capacity.
In Figure 2.4, since we are interested in the velocity profile leading to convergence to the



(a) velocity profile

(b) energy landscape

Figure 2.4. (a) Velocity profile for a small number of pattern $p = 7$ and a number of pattern for which the model is closer to maximum capacity $p = 70$. Independently of p, closer to the pattern to be retrieved, the velocity's magnitude is very similar. Farther from the attractor, the velocity is strongly dependent on the number of patterns.
(b): Potential energy with error bars

closest pattern, we are only considering the values after the maximum ϵ_{max} in Figure 2.3. In this context, ϵ_{max} represents the largest ϵ that still assures convergence to the closest pattern.
When p increases, both the maximum value of the curve, the basin of attraction and the velocity $\frac{dz}{dt}$ decrease, leading to a smaller value of ϵ_{max}.

As a first approximation, we assume $m = 1$. The potential energy, i.e. the energy landscape of the Hopfield network on the line connecting two attractors, is shown in Figure 2.4(b).

Generalizing $\forall p$, we obtain fitting values listed in Table 2.1. At last, we obtain an equivalent 1-dimensional problem described by the following stochastic differential equation (SDE)

$$\frac{dz(\epsilon, p, t)}{dt} = -\nabla U(\epsilon, p) + \sigma(\epsilon, p) dW_t \tag{2.7}$$

$$= 2az - 3bz^2 + (2a - 3b) + (\sigma_1 z^2 + \sigma_2 z + \sigma_3) dW_t \tag{2.8}$$

where dW_t describes a Wiener process, $a(p)$ and $b(p)$ are the fitted parameters for the average behaviour while $\sigma_1, \sigma_2, \sigma_3$ are the fitted parameters for the noise. Explicitely, $a(p) = a_1 p^2 + a_2 p + a_3$ and $b(p) = b_1 p^2 + b_2 p + b_3$.
The model has been solved using the Euler method. In Figure 2.5, it is shown a fit for $p = 15$.

An important tool given by the 1-dimensional model described in Equation 2.7 is that the unstable fixed point of the model is fitted in such a way to be similar to ϵ_{max}.

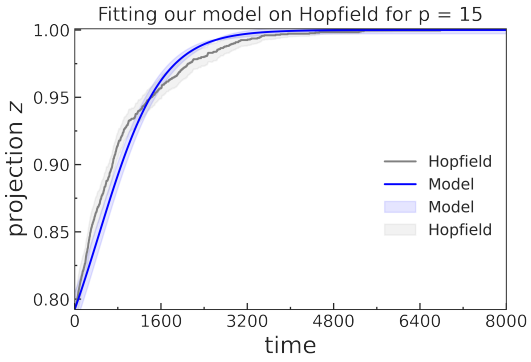


Figure 2.5. Comparison between the obtained SDE and the Hopfield model with error bars for $p = 15$. The model is consistent with the original Hopfield network for any value of p.

a_1	a_2	a_3	b_1	b_2	b_3
−1.21e−9	1.78e−7	1.72e−6	−4.17e−10	5.96e−8	8.18e−7

σ_1	σ_2	σ_3
−1.25e−6	1.29e−6	−5.12e−8

Table 2.1. Values of the fitting parameters

Physical interpretation

The convergence of the neural state to the attractor is analogous to a ball subject to a potential energy function. The 1 dimensional approximation we created can thus be interpreted as a simple mechanics problem of a ball rolling downhill with position dependent friction. At the minimum energy, the velocity is equal to 0 and the acceleration is constant (Figure 2.4). In other words, in order to stop at the minimum, the dynamics over ϵ needs friction, which is enforced by $U(\epsilon)$.
This result is consistent with how the Hopfield energy landscape is usually depicted. More precisely, the Hopfield network has a well-defined energy function defined in Equation 2.3 whose central property is to be (not strictly) decreasing. The attractors are its local minima. This is usually sketched in a similar way, as a ball which slides downhill until it comes to rest at one of the local minima, thanks to the decreasing energy function.
The main difference is that Equation 2.3 describes a complex energy landscape on a N dimensional space which only, as a first approximation, can be thought as a mechanics problem.
For the sake of this thesis, formalizing a precise reduction to a 1 dimensional dynamical system is a good approximation that allows a better understanding of the generalized Hopfield network described in chapter 3.

Chapter 3

The extension of the Hopfield's model

Our hypothesis is that map multiplicity (Sheintuch et al. 2020) is not random but rather represents contextual variables (subsection 1.1.2). This implies that cortical inputs to the hippocampus are precise enough to initialize the network in the "correct" basin of attraction. Experiments suggest that indirect projections from the medial prefrontal cortex, via the thalamic nucleus reuniens, to CA1 in the hippocampus indeed enable the hippocampus to encode contextual variables robustly (Ito et al. 2015, Wirt and Hyman 2019, Griffin 2021). While coding of space is robust to noise (subsection 2.1.1), relatively small amounts of noise can be sufficient to induce transitions between different contexts. We hypothesize that the choice of context by external inputs is robust to noise in specific directions of the activity space, and less robust in other directions.

Our goal is to show that it is indeed possible to transition between different contexts within the same environment. In order to accomplish this, we will work with a generalized Hopfield network which couples together different patterns in order to represent the same environment in different contexts. Such phenomena were previously identified in Hopfield-type networks using mean-field-theory techniques (Pereira and Brunel 2018,Tirozzi and Tsodyks 1991), but these techniques have not been applied to the generalized Hopfield network we propose here.

3.1 Steady-State description

In the following section, we provide a general description of the steady-state properties of the generalized Hopfield network. Differently from its counterpart described in chapter 2, we approach the model without prior knowledge.

3.1.1 The Model

In its original formulation, expressed by Equation 2.2, the Hopfield network attractors are uncorrelated with one another and orthogonal to each other, consistently with the

lack of correlations between spatial representations of different environments (Wilson and McNaughton 1993). Conversely, it is impossible to guarantee that, for example, changing the animal's task without changing the environment will result in a state representing that specific environment. Based on the results obtained by Sheintuch et al. 2020, we generalize the Hopfield model to remedy the discrepancy by introducing the cross-context coupling coefficient, denoted a.
The goal is to store a set of p patterns in k contexts $\xi_i^{\mu,\nu}$, where $\mu = 1,2,\ldots,p$ and $\nu = 1,2,\ldots,k$, representing, in this scenario, p environments in k contexts.
Each pattern in the network is composed as before by units labelled by $i = 1,2,\ldots,N$, where N is the total number of neurons in the network.
The connectivity matrix J_{ij} between neuron i and neuron j now reads

$$J_{ij} = \frac{1}{N}\left(\sum_{\mu=1}^{p}\sum_{\nu=1}^{k}\xi_i^{\mu,\nu}\xi_j^{\mu,\nu} + a\sum_{\mu=1}^{p}\sum_{\nu=1}^{k}\sum_{\substack{\rho=1\\ \rho\neq\nu}}^{k}\xi_i^{\mu,\nu}\xi_j^{\mu,\rho}\right) \tag{3.1}$$

where k is the number of contexts for each environment p. $\xi_i^{\mu,\nu}$ is now the stored pattern representing environment μ in context ν. The first term is the "Hebbian learning" we previously discussed and applied in the Hopfield model. The second term corresponds to binding of states across contexts since ρ is summed over all contexts not equal to ν.

For simplicity, we reduce the number of contexts for each environment μ to $k = 2$. The connectivity matrix can now be written in an equivalent form as

$$J_{ij} = \frac{1}{N}\left(\sum_{\mu=1}^{2p}\xi_i^{\mu}\xi_j^{\mu} + \frac{a}{2}\sum_{\mu=1}^{p}\left(\xi_i^{2\mu-1}\xi_j^{2\mu} + \xi_j^{2\mu-1}\xi_i^{2\mu}\right)\right) \tag{3.2}$$

The update rule remains unmodified (Equation 2.1).
Within the configuration space described by all the possible states of the network, the attractors of the network are not well-defined anymore, given the presence of the cross-context coupling term.

3.1.2 Dependence on the coupling parameter

In this subsection, we study the effect of the cross-context coupling term a on the network. We aim at finding a range of values for the parameter a that is consistent with our previous assumption. We hypothesize that, for small a, the network will behave similarly to the Hopfield network, while, for large a, the contextual representation will collapse and each environment representation will not be differentiated across contexts.

In Figure 3.1, it is shown the overlap between two attractors as a function of a. It exists a narrow range of the parameter a, that we call a_{range}, identified by the two blue dashed lines, which is consistent with our assumption. We found that the range defined by a_{range} is dependent on α. We therefore selected a value of $overlap = 0.2$ and chose a accordingly. We call this value a_{cross}. In Table 3.1, we display the dependence of a_{cross} with respect to α.

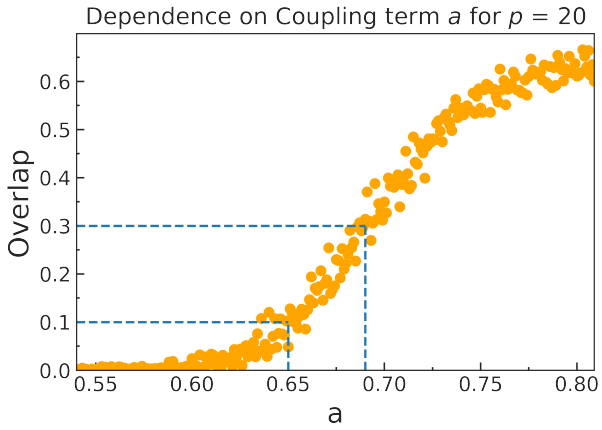


Figure 3.1. We can identify three regions: the leftmost region, for which we have $overlap \simeq 0$, can be interpreted as a region with similar behaviour as a Hopfield network. The rightmost region, where $overlap > 0.3$, leads to the contextual representation collapse and the merging of the two coupled fixed point into a superposition state. The central region, where $0.1 < overlap < 0.3$, is indeed where we measure a moderate coupling effect that leads to the emergence of coupled fixed points. Simulation obtained for $p = 20$ and $N = 3000$

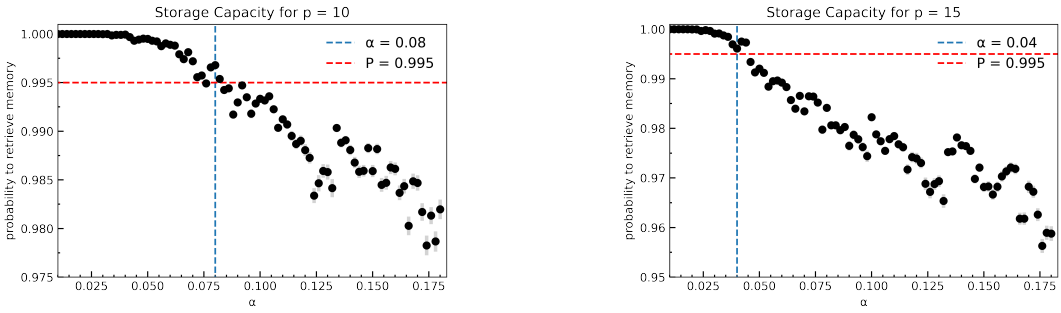
α	a_{cross}
0.0010	0.984
0.0020	0.894
0.0033	0.786
0.0066	0.693
0.0130	0.607

Table 3.1. Values of α and a_{cross} when $overlap = 0.2$

3.1.3 Storage Capacity

As we mentioned in subsection 2.1.2, above a certain value of $\alpha_{max} = \frac{p}{N}$, the network does not retrieve correctly the stored memories. We studied that, in the Hopfield network, α_{max} is well defined and the transition between correctly and incorrectly stored patterns occurs sharply (Hopfield 1982). However, a priori, we cannot make the same claim about the generalised Hopfield model. In Figure 3.2, it is shown the transition depending on different p. Unlike the Hopfield model, the transition appears to be smoother and with a strong dependence on the number of patterns p. Unlike our previous approach, we are not aware of the maximum capacity for this network. Therefore, we perform multiple simulations with large N and set a threshold value $P_{err} < \frac{1}{N}$ in order to obtain perfect memory retrieval. Specifically, we set $P_{err} = 0.005$. We observe that the coupling modifies the fixed point continuously, therefore defining memory retrieval is indeed problematic.

As it is shown in Figure 3.2, the value of α_{max} for which the patterns are retrieved



(a) Storage capacity $p = 10$

(b) Storage capacity $p = 15$

Figure 3.2. (a) Probability to retrieve a memorized pattern depending on parameter α for $p = 10$ and (b) for $p = 15$

correctly is lower than its Hopfield counterpart for any value of p. More specifically, it is quickly decreasing the larger p gets. It is shown in Figure 3.2 that, for $p = \{10,15\}$, α_{max} halves its value. As a result, in order to maintain correct memory retrieval, N must scale incredibly fast with respect to p.
Nonetheless, even for small p, the analyses of the model becomes computationally expensive since we continuously modify the fixed point due to the introduction of the coupling term. Subsequently, we analyze the model exclusively for small p. In particular, we mainly focus on the case $p = 4$ for which α_{max} is similar to the Hopfield model, namely $\alpha_{max} \approx 0.13$.

3.1.4 Coexistence of orthogonal and non-orthogonal fixed points

As we discussed in subsection 2.1.1, in the Hopfield network, the patterns are defined as orthogonal to each other (Figure 3.3(a)). By contrast, the addition of the cross-coupling term a, defined in subsection 3.1.1, partially breaks this symmetry. In particular, even though the stored patterns ξ^μ are always defined and initialized orthogonal to each other, the network can converge to fixed points which are different from ξ^μ $\forall\mu$. Henceforth, following the definition of the connectivity matrix J_{ij} (Equation 3.1), we hypothesize that the introduction of the cross-coupling term a would lead to the appearance of new non-orthogonal, with respect to the patterns initialized by the network, fixed points within the same environment while, at the same time, maintaining the orthogonality between fixed points representing different environments.
Based on Equation 2.6, we define

$$z_\mu(t) = x(t) \cdot \xi^\mu$$

where $x(t)$ is the configuration at time t and ξ^μ represents the stored patterns. We define t_c as the time for which the network reaches convergence. Henceforth, we can write

$$z_\mu(t_c) = x(t_c) \cdot \xi^\mu := \xi_{new} \cdot \xi^\mu$$

where $z_\mu(t_c)$ now represents the correlation between the patterns initialised by the network ξ^μ and the fixed point the network actually converges to ξ_{new}. We then compute

$$c = \max_{\mu \in \{1,\dots,p\}} z_\mu(t_c)$$

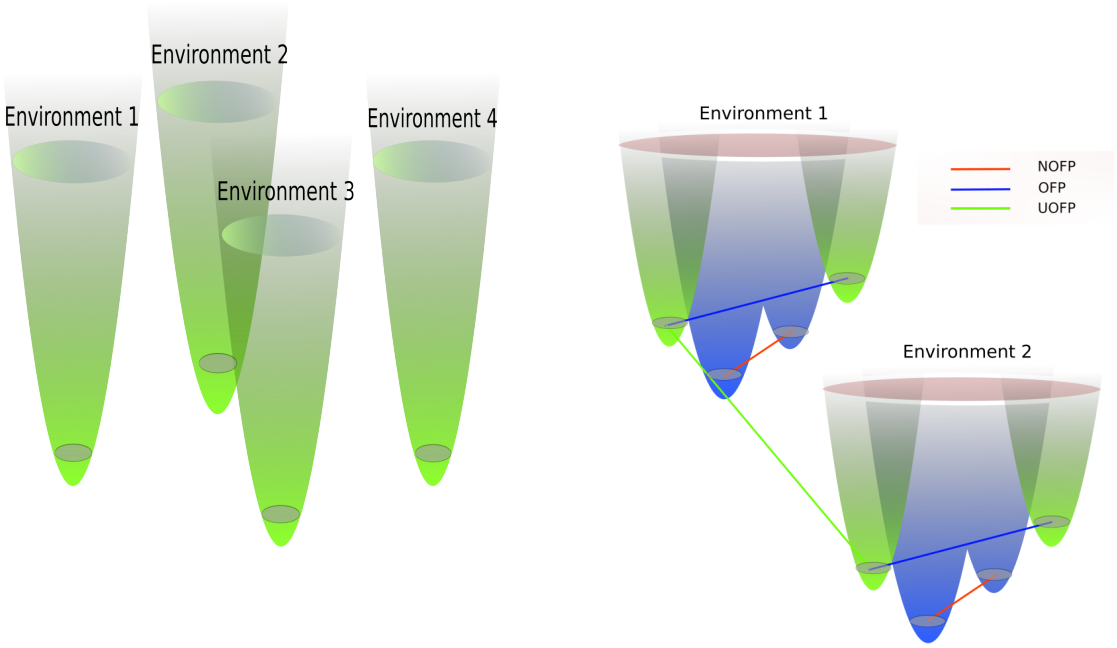
$$\rho = \arg \max_{\mu \in \{1,\dots,p\}} z_\mu(t_c)$$

If $c = 1$ then $\exists \mu : \ \xi_{new} \cdot \xi^\mu = 1$, i.e. the network converges to the stored pattern ξ^ρ. If $0.7 \leq c < 1$ then ξ_{new} represents a new fixed point which differs from all the stored patterns ξ^μ. The stored pattern which mostly resembles ξ_{new} is by definition ξ^ρ. To keep into account this information, we use a compact notation ξ^ρ_{new}. From now on, we will use ξ^μ_{new} only to denote the new fixed point generated by the introduction of the cross-coupling term which differs from all the stored patterns ξ^μ. Consequently, ξ^μ will simply be the stored patterns initialized by the network.
For example, we can identify $\mu = \{1,2\}$ as the representation of the same environment in two different contexts, while, $\mu = \{3,4\}$ as the representation of a different environment in two different contexts. Since J_{ij} is symmetric for attractors representing different environments, we would obtain the same results for any other μ. Anyhow, we reduced our description to $p = 4$ therefore the description is comprehensive.
Our goal will be to describe the dynamics on the line connecting two attractors; consequently, we define separately different pairs of fixed points. More specifically, we define as '**NOFP**' the line connecting ξ^1_{new} and ξ^2_{new} (or, equivalently, ξ^3_{new} and ξ^4_{new}). Similarly, we define as '**OFP**' the line connecting ξ^1 and ξ^2 (or, equivalently, ξ^3 and ξ^4). Ultimately, we define as '**UOFP**' the line connecting ξ^1_{new} and ξ^3_{new} (or, equivalently, ξ^2_{new} and ξ^4_{new}, or ξ^1_{new} and ξ^4_{new}, or ξ^2_{new} and ξ^3_{new}). We will also call UOFP the line connecting ξ^1 and ξ^3 (or, equivalently, ξ^2 and ξ^4, or ξ^1 and ξ^4, or ξ^2 and ξ^3). We will explain in the following section why we are not distinguishing the last two cases.

The results produced by the model partially confirm our hypothesis. More specifically, each context is now represented by two fixed points: the stored pattern ξ^μ, enforced by the Hebbian learning rule, and a new fixed point ξ^μ_{new}, enforced by the cross-coupling term. Differently than before, the dynamics within the same environment is now characterized by the coexistence of four different fixed points. In Figure 3.3(b), a schematic illustration of the fixed points and the lines connecting them is shown. We will study the dynamics in the next section.

3.2 Transient dynamics

After studying the steady state of the model, we will now analyze the dynamics along different lines connecting all the fixed points within the same environment and across different environments.
As we discussed in the section above, the model generates a new set of coupled fixed points ξ^μ_{new} within the same environment. We define a fixed point as a global minimum based on which fixed point is more likely that the network converges to. Our hypothesis is that those fixed points are stable to perturbations and represent the new global minima of the



(a) Illustration of the energy landscape of the Hopfield network

(b) Illustration of the energy landscape of the extension of the Hopfield network

Figure 3.3. Schematic illustration of (a) the energy landscape we sketch for network dynamics of the Hopfield network and (b) for the extension of the Hopfield network. In (b), we also show the lines connecting the fixed points defined before.

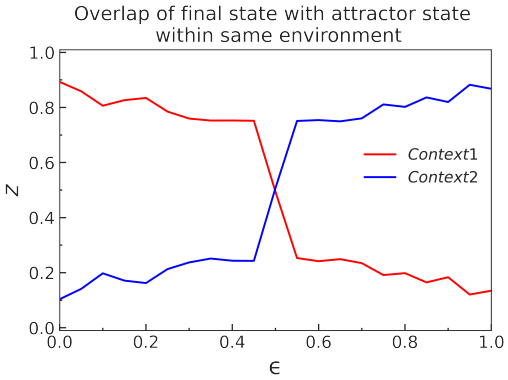


Figure 3.4. We computed the final state z the network settles on as a function of the initial condition ϵ on the line connecting the representation of the same environment in context 1 (blue line) and context 2 (red line). Due to the coupling, the network converges on average to a non-orthogonal fixed point therefore $z(\epsilon) \neq 1$.

dynamics. We also expect that it is possible to escape these new global minima, which we can interpret as a small probability of not recognizing the environment correctly.
We cannot assume a priori that ξ^{μ}_{new} are the global minima of the network therefore, in the next subsections, we study each case independently. Subsequently, we create a 1-dimensional approach based on the same method we developed for the Hopfield network.

3.2.1 Dynamics on the line between fixed points

We are considering $p = 4$. As we previously mentioned, we will call for simplicity $\mu = \{1,2\}$ the representation of the same environment in two different contexts, and $\mu = \{3,4\}$ the representation of a different environment in two different contexts.

Dynamics on the line between coupled orthogonal fixed points (OFP)

The first case we distinguish is by moving the initial condition on the line OFP. In Figure 3.5, we are showing the probability to convergence to any fixed point while moving on the line OFP.

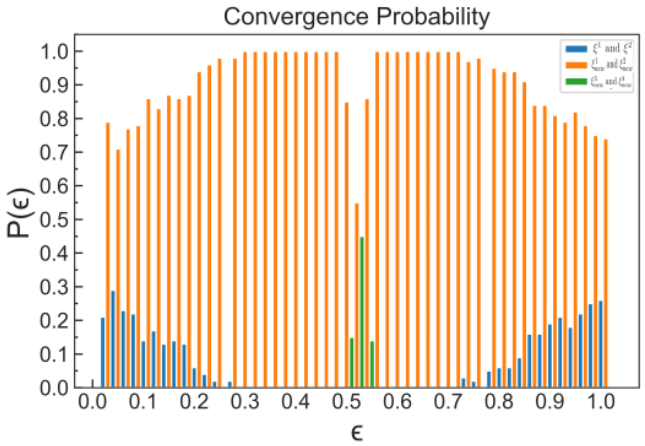


Figure 3.5. Convergence probability while moving along the line OFP. The orange bars are describing the probability to convergence to one of the two ξ^{μ}_{new} within the same environment. The same can be said for the blue bars describing ξ^{μ}, which are only present for the extreme values of ϵ. The green bars, which are representing convergence to any other fixed point which is not within the same environment, are present for $\epsilon \approx 0.5$

ξ^{μ} are now relative minima of the dynamics. Their convergence probability is negligible everywhere except when we are close to their attractor basins (for $\epsilon \simeq 0$ and $\epsilon \simeq 1$).
We can also notice that, for $\epsilon \approx 0.5$, there is a non-negligible probability to convergence to a fixed point of a different environment (green bars).
The above behaviour is compatible with what we hypothesized. ξ^{μ} are local minima while ξ^{μ}_{new}, describing the existence of two different contexts within the same environment, are now global minima; they are also stable to perturbation which is given by moving on a line

OFP. Furthermore, there is a small probability of not recognizing the correct environment, shown by the green bars.

Dynamics on the line between coupled non-orthogonal fixed points (NOFP)

Subsequently, we proceed on analyzing the dynamics along the line NOFP. We obtain a behaviour, shown in Figure 3.6, not coherent with our hypothesis and underlying a more complicated energy landscape. Even though we are moving on the line NOFP,

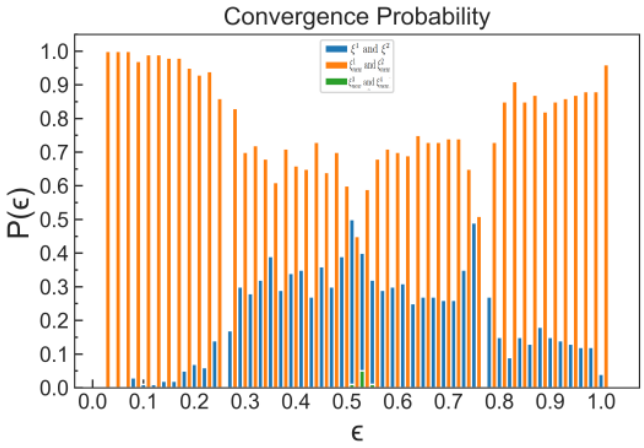


Figure 3.6. Convergence probability while moving along the line NOFP. The blue bars, describing ξ^μ, are now present for any value of ϵ. The green bars, representing convergence to any other fixed point which is not within the same environment, are now almost zero everywhere.

the probability to convergence to ξ^μ is not negligible. This behaviour can be explained considering that the N dimensional space, in which the dynamics is evolving, is not easily reducible to a lower dimensional space as we previously manage to obtain for the Hopfield model. This underlying complexity allows a relative minimum with a smaller basin of attraction to be the fixed point of the dynamics, contrary to what we hypothesized. We also notice how the probability to convergence to a fixed point representing a different environment drops to zero.
Following this result, we cannot disregard the existence of ξ^μ since they are stable fixed points of the dynamics with a non-negligible probability.

Dynamics on the line between uncoupled orthogonal fixed points (UOFP)

Ultimately, we study the behaviour on the line UOFP, i.e. fixed points belonging to different environments. Here, the model is following again our hypothesis. ξ^μ_{new} are the global minima of the dynamics while ξ^μ are not stable (Figure 3.7). This result is obtained independently whether the line connects two ξ^μ_{new} or two ξ^μ (assuming that they are representing different environments).

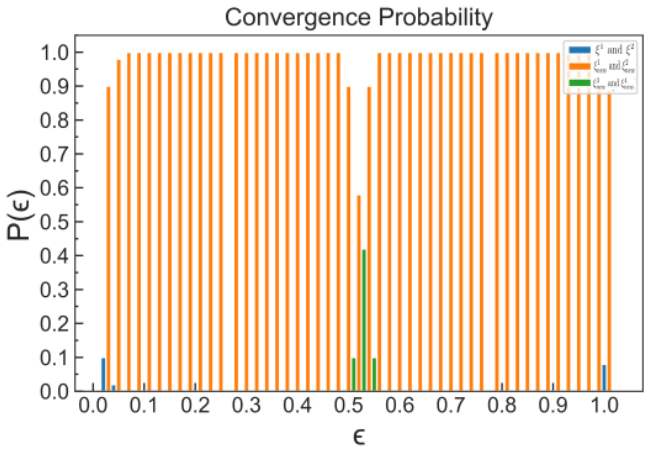


Figure 3.7. Convergence probability while moving along the line UOFP. The blue bars, describing ξ^μ, are now absent for any value of ϵ. The green bars, representing convergence to any other fixed point outside the same environment, are now present again for values of $\epsilon \approx 0.5$.

3.2.2 Stochastic 1-dimensional approach

We now want to summarize our results in a stochastic 1-dimensional model. We cannot proceed as before by outlining a single SDE. Therefore, we create a stochastic machine. The states of the machine are the fixed points of the network. We introduce an initial state I and a final state F connected to all the other states X, namely $UOFP$, $NOFP$ and OFP. Each state represents the evolution on the respective line. The transition probability from the initial state to any other state $a_{(I,X)}$ is assigned randomly. After the first steps, the network will likely transit to the state with the highest probability. Within all the states, the transition probabilities $a_{(X,Y)}$ are obtained from the heatmaps shown in Figure 3.9. They describe the number of times the value of the projection z at time t emerged for each state. Self loops, representing the probability to remain in the current state $a_{(X,X)}$, are allowed.
At each time step t, we update the transition probabilities accordingly to the value of the heatmap at $z\,(t-1)$ for each state, and then either update to a new state or remain in the current one depending on the values of $a_{(X,Y)}$ at that time t. When convergence is reached, i.e. the probability value for one of the states W is greater than 0.98, we move to the final state F with transition probabilities $a_{(X,F)} = \delta_{X,W}$ where $\delta_{X,W}$ is the Kronecker delta.
A state machine similar to the one described above is depicted in Figure 3.8.
The heatmaps shown in Figure 3.9 were used to infer the transition probabilities $a_{(X,Y)}$ for each state at each t. Since the number of time steps t is incredibly high before convergence is reached, we are not showing their values for each t. To obtain $a_{(X,Y)}$ from the heatmaps, we superpose the three heatmaps and normalize them obtaining for each value of $z(t)$ the transition probabilities $a_{(X,Y)}$ with respect to one of the fixed points, e.g. a ξ^μ_{new} since it has the highest probabilities among all the states to be the fixed point the dynamics converges

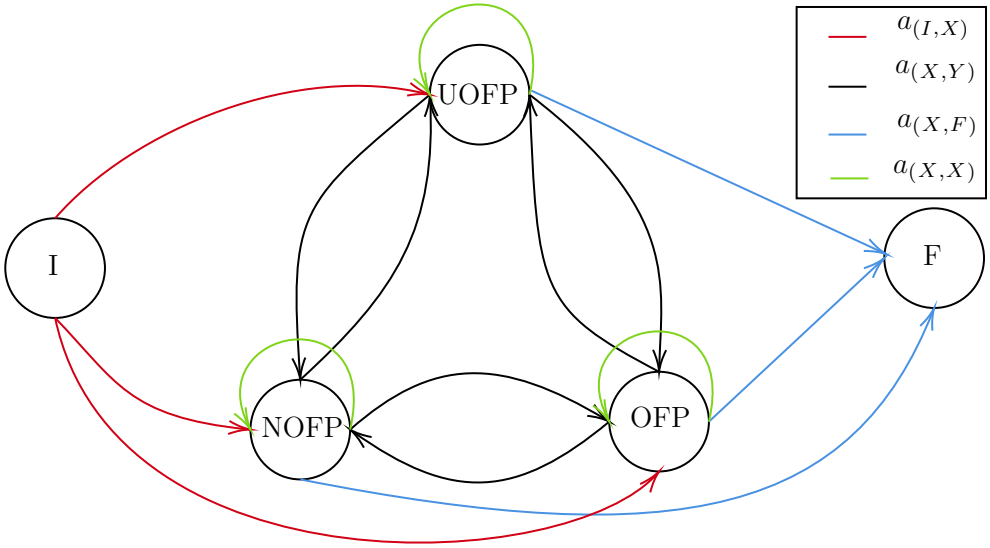


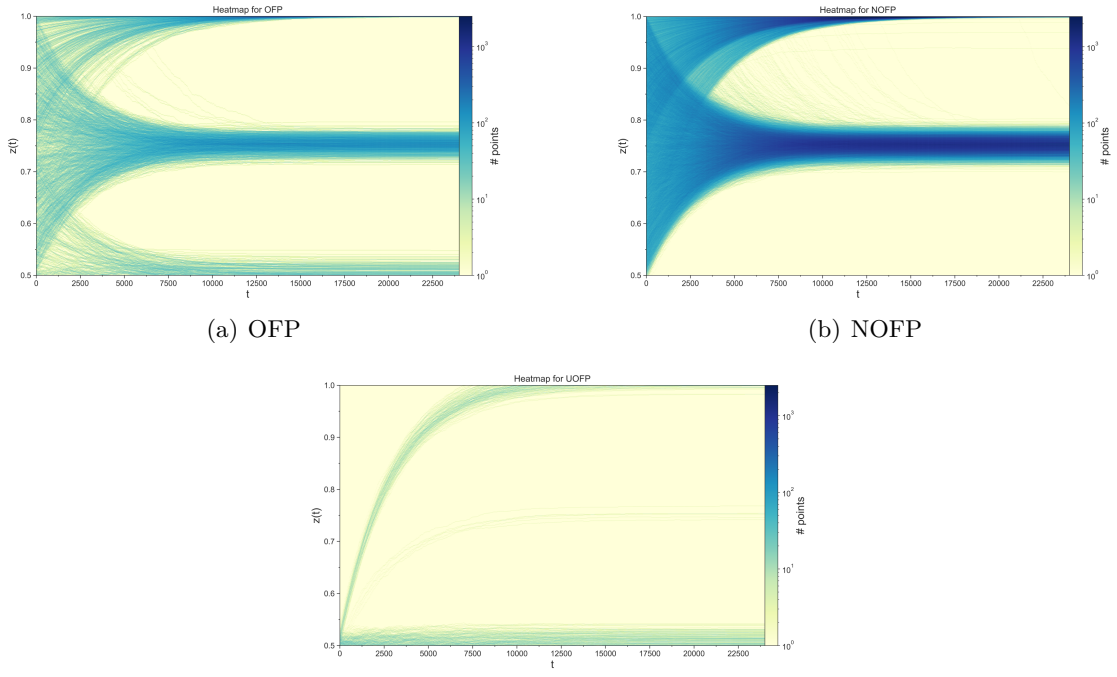
Figure 3.8. State machine describing the dynamics evolution

to. The heatmaps are symmetrical on the z-axis with respect to $z = 0.5$. Henceforth, $z = [0,0.5]$ is equivalent to $z = [0.5,1]$ since we are representing the convergence to the other fixed point on the line. In Table 3.2, we list a few values of $z(t)$ (evaluated with respect to one of the ξ^{μ}_{new}) to convey a better estimate of the order of magnitude of the transition probabilities $a_{X,Y}$. Each one of the states has emission probability $e_{X,z(t)}$,

$z_{\xi^{\mu}_{new}}$	t	$NOFP$	OFP	$UOFP$
0.52	0	0.596	0.115	0.288
0.75	0	0.925	0.075	0
0.95	0	0.894	0.106	0
0.52	22500	0	0.8	0.2
0.75	22500	0.981	0.019	0
0.95	22500	0.937	0.063	0

Table 3.2. Values of transition probability $a_{X,Y}$ for different values of $z(t) = x(t) \cdot \xi^{\mu}$. The values above shows how the NOFP state has the highest probabilities. Nonetheless, for $z(t = 0) \simeq 0.5$, i.e. when we are near the midpoint of the line connecting the two fixed points, the probability to transit to a different environment is no longer negligible. For $t = 22500$, the network has almost reached convergence therefore having probability values similar to 1, independently of the state.

i.e. corresponding to the emission of $z(t)$ when the network is in state X. Each emission probability is described by a SDE fitted exactly as in subsection 2.2.2 which represents the



(a) OFP

(b) NOFP

(c) UOFP

Figure 3.9. Heatmaps describing the number of times the value of the projection z at time t emerged for each state (a) OFP, (b) $NOFP$ and (c) $UOFP$. An appropriate normalization of these values leads to the transition probabilities $a_{X,Y}$

evolution on the different lines. Each state is therefore locally described by a 1-dimensional model.
The model described here is a good approximation of the dynamics of the extension of the Hopfield network. In the final chapter, we will discuss possible approaches to improve this description.

Physical interpretation

Due to the higher level of complexity of the N dimensional space, it was impossible to map the model directly to an analogous 1-dimensional problem. We therefore created a state machine that would allow us to locally approximate the problem as 1-dimensional. The complexity of the space was considered by inferring the parameters for the state machine. It is indeed harder to sketch a physical interpretation.

Chapter 4

Conclusion

We started our discussion by analyzing the classical Hopfield model. Based on the remapping phenomenon, the network converged to attractors which represented a collection of neuronal population states we interpreted as environments. Moreover, based on the results obtained by Sheintuch et al. 2020, our goal was to show that it was indeed possible for the hippocampus to support efficient multiplexed representations of spatial and behaviorally relevant contextual information, and, consequently, to safely transit between representations of different contexts within the same spatial representation. Accordingly to previous literature, we hypothesized that the contextual variables were arising from the pre-frontal cortex. Henceforth, we proposed a generalized Hopfield network which couples together different patterns in order to represent the same spatial representation in different contexts.
Before analyzing the generalized Hopfield network, we focused on the transient dynamics of the Hopfield model along a line connecting two different attractors. We developed an analogous 1 dimensional model to describe the dynamics along this line. Subsequently, we analyzed the steady-state of the generalized Hopfield model. We showed the emergence of multiple fixed points within the same environment. Ultimately, we modelled the dynamics along different lines connecting all the fixed points within the same environment and across different environments with a stochastic machine.

4.1 Results interpretation

We hypothesized that remapping occurs due to a major change in the animal's internal state arising from the pre-frontal cortex. We interpret it as a switch of contextual information which leads to a joint hippocampal and cortical representation of space and context. Our aim was to obtain a quantitative description of how switches are induced inside the brain, in order to understand the minimal manipulations that trigger remapping and allow to transit between different contextual representation of the same environment. We developed an analogous 1 dimensional stochastic model to efficiently study the generalized Hopfield model we proposed, which describes mixed representations of spatial and behavioral contextual information.

The model allowed us to obtain information on the precision and the dimensionality of inputs, which the hippocampus receives from cortical areas, that can induce transitions between attractors and, analogously, between different contextual representations.
Our model of transitions between different representations can be interpreted as a first step to understand the minimal mechanisms supporting the process of learning mixed spatial and contextual representations. Nonetheless, as we mention in the next section, the model can further be improved.

4.2 Future works

In the model investigated thus far, each pattern corresponded to a fixed-point attractor, i.e., a specific combination of firing-rates of neurons in the population. Based on previous literature on ring attractor models (Battaglia and Treves 1998), our dynamical system can be extended to a continuous attractor model with multiple environments and a cross-context coupling a, similarly to Equation 3.1. More specifically, we can assume that the animal navigates in a one-dimensional ring-shaped environment where its position is described by an angle θ. We do not know whether the cross-context coupling a has the same effect on the dynamics of the continuous attractor network as it did on the network with fixed-points. It is indeed possible that the sensitivity to input noise we found in the fixed-point case is remedied here by the smoothness imposed by the continuous attractor model.

References

[1] D. Amit, H. Gutfreund, and H. Sompolinsky. "Spin-Glass Models of Neural Networks". In: *Physical Review* A 32 (1985), pp. 1007–1018.

[2] D. Amit, H. Gutfreund, and H. Sompolinsky. "Storing Infinite Numbers of Patterns in a Spin-Glass Model of Neural Networks". In: *Physical Review* 55 (1985), pp. 1530–1533.

[3] D. Aronov, R. Nevers, and D.W. Tank. "Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit". In: *Nature* 543 (2017), pp. 719–722.

[4] F.P. Battaglia and A. Treves. "Attractor neural networks storing multiple space representations: A model for hippocampal place fields". In: *Phys. Rev. E* 58 (6 1998), pp. 7738–7753.

[5] E. Bostock, Muller, and J.L. Kubie. "Experience-dependent modifications of hippocampal place cell firing". In: *Hippocampus* 1.2 (1991), pp. 193–205.

[6] L.L. Colgin, E.I. Moser, and M. Moser. "Understanding memory through hippocampal remapping". In: *Trends in Neurosciences* 31.9 (2008), pp. 469–477. ISSN: 0166-2236.

[7] A.A. Fenton and Muller. "Place cell discharge is extremely variable during individual passes of the rat through the firing field". In: *Proc Natl Acad Sci U S A.* 95(6) (1998), pp. 3182–3187.

[8] A.L. Griffin. "The nucleus reuniens orchestrates prefrontal-hippocampal synchrony during spatial working memory". In: *Neurosci. Biobehav. Rev.* 128 (2021), pp. 415–420.

[9] T. Haga and T. Fukai. "Extended Temporal Association Memory by Modulations of Inhibitory Circuits". In: *Phys. Rev. Lett.* 123 (7 2019), p. 078101.

[10] J. Hertz, S. Krogh, and R. Palmer. *Introduction To The Theory Of Neural Computation.* Reading. MA. USA: Addison-Wesley, 1991. ISBN: 9780429968211.

[11] J.J. Hopfield. "Neural Networks and Physical Systems with Emergent Collective Computational Abilities". In: *Proceedings of the National Academy of Sciences* USA 79 (1982), pp. 2554–2558.

[12] J.J. Hopfield. "Neurons with graded response have collective computational properties like those of two-state neurons". In: *Proceedings of the National Academy of Sciences* 81 (10) (1984), pp. 3088–3092.

[13] H.T. Ito et al. “A prefrontal–thalamo–hippocampal circuit for goal-directed spatial navigation”. In: *Nature* 522 (2015), pp. 50–55.

[14] J.B. Julian and C.F. Doeller. “Remapping and realignment in the human hippocampal formation predict context-dependent spatial behavior”. In: *Nature neuroscience* 24.6 (2021), pp. 863–872. ISSN: 1097-6256.

[15] M.W. Jung, S.I. Wiener, and B.L. McNaughton. “Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat”. In: *Journal of Neuroscience* 14.12 (1994), pp. 7347–7356. ISSN: 0270-6474.

[16] J.J. Knierim and J.P. Neunuebel. “Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics”. In: *Neurobiology of learning and memory* 129.18 (2016), pp. 38–49.

[17] D. Krotov and J.J. Hopfield. “Dense Associative Memory for Pattern Recognition”. In: (2016).

[18] P. Latuske et al. “Hippocampal Remapping and Its Entorhinal Origin”. In: *Frontiers in Behavioral Neuroscience* 11 (2018). ISSN: 1662-5153.

[19] S. Leutgeb et al. “Independent Codes for Spatial and Episodic Memory in Hippocampal Neuronal Ensembles”. In: *Science* 309.5734 (2005), pp. 619–623.

[20] C. Lever et al. “Long-term plasticity in hippocampal place-cell representation of environmental geometry”. In: *Nature* 416.6876 (2002), pp. 90–94.

[21] I. Low et al. “Dynamic and reversible remapping of network representations in an unchanging environment”. In: *Neuron* 109.18 (2021), 2967–2980.e11. ISSN: 0896-6273.

[22] T.J. McHugh et al. “Impaired hippocampal representation of space in CA1-specific NMDAR1 knockout mice”. In: *Cell* 87(7) (1996), pp. 1339–1349.

[23] Muller. “A quarter of a century of place cells”. In: *Neuron* 17.5 (1996), pp. 813–822.

[24] Muller and J.L. Kubie. “The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells”. In: *Journal of Neuroscience* 7.7 (1987), pp. 1951–1968.

[25] Muller, J.L. Kubie, et al. “Spatial firing correlates of neurons in the hippocampal formation of freely moving rats”. In: *J. Paillard (Ed.), Brain and space* (1991), pp. 296–333.

[26] E.H. Nieh et al. “Geometry of abstract learned knowledge in the hippocampus”. In: *Nature* 595 (2021), pp. 80–84.

[27] J. O’Keefe. “A review of the hippocampal place cells”. In: *Prog Neurobiol.* 13(4) (1979), pp. 419–439.

[28] J. O’Keefe. “Do hippocampal pyramidal cells signal non-spatial as well as spatial information?” In: *Hippocampus* 9 (1999), pp. 352–364.

[29] J. O’Keefe and J. Dostrovsky. “The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat”. In: *Brain Res.* 34 (1971), pp. 171–175.

[30] J. O'Keefe and L. Nadel. *The hippocampus as a cognitive map. Preliminary evidence from unit activity in the freely-moving rat.* Oxford university press, 1978.

[31] U. Pereira and N. Brunel. "Attractor Dynamics in Networks with Learning Rules Inferred from In Vivo Data". In: *Neuron* 1699 (2018), 227–238.e4.

[32] B. Poucet, C. Thinus-Blanc, and Muller. "Place cells in the ventral hippocampus of rats". In: *Neuroreport* 5 (Nov. 1994), pp. 2045–8.

[33] B.A. Radvansky et al. "Behavior determines the hippocampal spatial mapping of a multisensory environment". In: *Cell Rep.* 109444 (2021).

[34] A. Rotenberg et al. "Mice expressing activated CaMKII lack low frequency LTP and do not form stable place cells in the CA1 region of the hippocampus". In: *Cell* 87(7) (1996), pp. 1351–1361.

[35] L. Sheintuch et al. "Multiple Maps of the Same Spatial Context Can Stably Coexist in the Mouse Hippocampus". In: *Curr. Biol.* 30 (2020), 1467–1476.e6.

[36] B. Tirozzi and M. Tsodyks. "Chaos in Highly Diluted Neural Networks". In: 14 (1991), p. 727.

[37] M.A. Wilson and B.L. McNaughton. "Dynamics of the hippocampal ensemble code for space". In: *Science* 261 (1993), pp. 1055–1058.

[38] R.A. Wirt and J.M. Hyman. "ACC Theta Improves Hippocampal Contextual Processing during Remote Recall". In: *Cell Rep.* 27 (2019), 2313–2327.e4.