NTNU – Trondheim
Norwegian University of
Science and Technology

Examining Remapping Phenomena in a Computational Model of the Hippocampus

Vegard Edvardsen

Master of Science in Computer Science
Submission date: June 2014
Supervisor: Keith Downing, IDI

Norwegian University of Science and Technology
Department of Computer and Information Science

Abstract

The hippocampus is a brain structure known to be involved in high-level cognitive features such as memory and navigation. Spatially responsive cells in the hippocampal region, known as place cells and grid cells, have given neuroscientists an opportunity to try to decipher the neural implementation of computations performed by the hippocampus. Important insights toward this goal might be offered by studying the phenomena of global remapping and rate remapping, known from hippocampal experiments, that relate changes in the environment to changes in neural activity. This project implements a biologically plausible neural network that models grid cells, place cells and other parts of the hippocampus in order to computationally reproduce the remapping phenomena. The implemented model successfully exhibits both global and rate remapping, and thus offers a potential framework for a better understanding of why these phenomena occur.

Sammendrag

Hippocampus er en hjernestruktur som er kjent for å være involvert i høyere kognitive funksjoner som hukommelse og navigasjon. Romlig påvirkede celler i området rundt hippocampus, kjent som stedceller og gitterceller, har gitt hjerneforskere en anledning til å prøve å forstå den nevrale implementasjonen av beregninger utført i hippocampus. Viktig innsikt for å oppnå dette målet kan være å hente ved å studere fenomenene global omkartlegging og ratemessig omkartlegging. Disse fenomenene, kjent fra hippocampus-eksperimenter, relaterer endringer i omgivelsene til endringer i den nevrale aktiviteten. Dette prosjektet implementerer et biologisk plausibelt nevralt nettverk som modellerer gitterceller, stedceller og andre deler av hippocampus for å gjenskape omkartsleggingsfenomenene i en beregningsmessig modell. Den implementerte modellen kan vellykket forevise både global og ratemessig omkartlegging, og den utgjør dermed et potensielt rammeverk for en bedre forståelse av hvorfor disse fenomenene inntreffer.

Contents

Glossary

afferent neuron Presynaptic neuron, an upstream neuron that sends signals to the current neuron.

CAN Continuous Attractor Network.

DG Dentate Gyrus, the first submodule of the hippocampus, which receives inputs from the Entorhinal Cortex and projects outputs to CA3.

efferent neuron Postsynaptic neuron, a downstream neuron that receives signals from the current neuron.

MEC Medial Entorhinal Cortex.

1 Introduction

1.1 The Hippocampus and the Entorhinal Cortex

The hippocampus is a brain structure found in vertebrates, including mammals such as humans. In humans, the hippocampus is located in the temporal lobes in the two hemispheres of the brain, situated underneath the neocortex, which is the wrinkled outer layer of the brain. Figure 1.1 shows the location of the hippocampus in a cut-out view of the human brain.

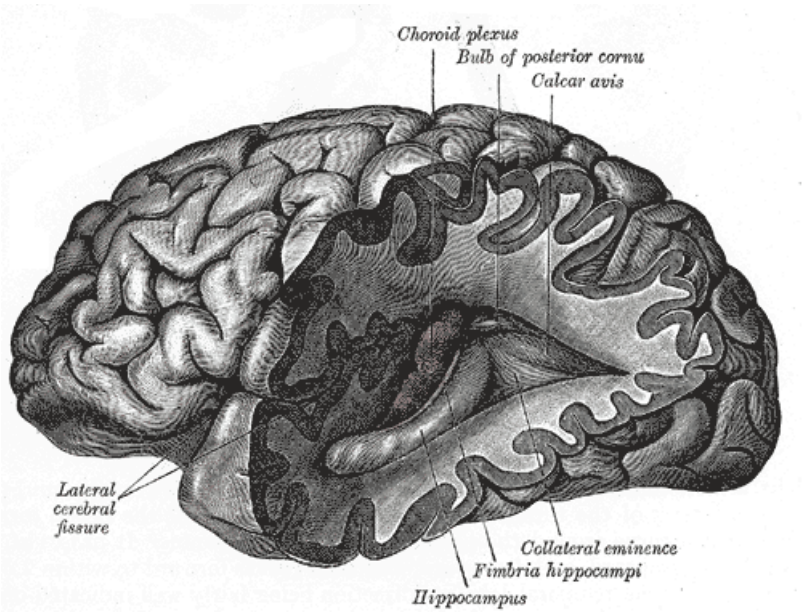


Figure 1.1: The location of the hippocampus inside the human brain. The curved structure in the center of the cut-out is the hippocampus. Figure reprinted from a work in the public domain [9].

The function of the hippocampus remained unknown until the middle of the previous century, when evidence from persons with damage to the hippocampus came to light.

A notable patient, H.M., had both of his hippocampal areas deliberately lesioned in 1954 in an attempt to alleviate his suffering from severe epileptic seizures. While the seizures stopped, he also lost his ability to form new long-term memories, a condition known as anterograde amnesia. His ability to recall previously formed memories was however not impacted [18].

Anatomically, the hippocampus is considered to consist of the four major subareas Dentate Gyrus (DG), CA3, CA1 and subiculum. The primary connectivity of the hippocampus with the rest of the brain is through the entorhinal cortex, an adjacent brain structure, which connects further to different parts of the neocortex. The aforementioned areas and the connectivity between them are shown in Figure 1.2.

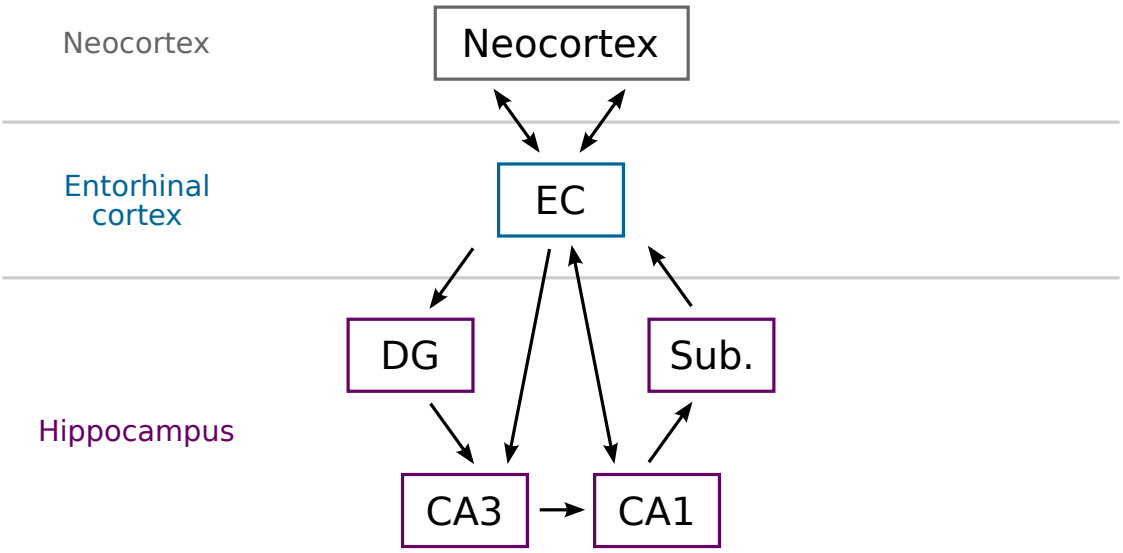


Figure 1.2: High-level overview of areas in the hippocampus and their connectivity with each other and with the entorhinal cortex, as well as the connectivity of the entorhinal cortex with the neocortex. Entorhinal cortex (EC), dentate gyrus (DG), subiculum (Sub.).

The neocortex receives sensory signals of many different modalities, such as sight, hearing and touch. The sensory information is processed by the neocortex into ever more abstract and high-level representations, in a neocortical information processing hierarchy [12]. The neocortex is also responsible for generating motor command outputs for voluntary movement. The fact that the hippocampus primarily exchanges information with a brain region that operates on highly processed and abstract representations, provides further substantiation that the hippocampus is involved in high-level features of cognition.

1.2 Remapping and Cognitive Maps of Space

Another high-level cognitive function of the hippocampus became evident with the discovery of *spatially responsive cells* in the hippocampal region, i.e. cells whose activity depends on where the animal is located in the environment.

The direct relationship between externally available parameters such as the animal's position in an experimental enclosure, and the firing activity of these high-level hippocampal cells, provides an important opportunity for neuroscientists to try to decipher the inner workings of cognitive features of the hippocampus on a neural level. The spatially responsive cells enable scientists to correlate high-level neural activity with the influencing external parameters and to experimentally observe changes that occur in the neural activity in response to induced changes in the external environment.

Place cells are hippocampal neurons that typically are active in only a single, contiguous area in the animal's environment. *Grid cells* are a type of neurons in the entorhinal cortex that are active at multiple locations in the environment, in a repeating, grid-like pattern. The collective activity to the grid cells appears to encode the two-dimensional spatial coordinates of the animal in a general way. The information encoded by place cells and grid cells thus comprises a "cognitive map" of the space navigated by the animal [16].

Gathering neuroscientific evidence provides an ever better understanding of the hippocampus. This includes results from animal experiments, where neural activity is recorded from brains of animals as they explore different environments, as well as theoretical results, where the empirically observed results are modeled in various ways.

One type of experiments that are performed in the study of the hippocampus are called *remapping* experiments. These experiments seek to measure the changes that occur in the hippocampal neurons in response to *changes in the environment*, for example when the wall color of the animal's enclosure suddenly changes or when the shape of the enclosure morphs from a square box into a cylinder. The term "remapping" refers to the change in the hippocampal spatial maps that happens in response to the environmental changes.

The different types of response seen from the hippocampus have eventually been categorized into *rate remapping* and *global remapping*. The distinction between the two concerns whether the *same* subset or a *different* subset of neurons is active before and

after the environmental change. Studying these phenomena of global and rate remapping, for example working to uncover the exact conditions that determine whether a global remapping or a rate remapping response is seen, might give insights into how the cognitive map works and in general how information is acquired, processed and stored in the hippocampus.

1.3 Research Questions

This project seeks to explore whether the phenomena of global and rate remapping can be replicated in a computational model of the hippocampus, potentially offering insights into the circumstances of these phenomena.

For my preliminary project to the Master's thesis in the fall of 2013, I implemented a computational model of grid cell and place cells [5]. This model will be extended to create a more complete hippocampal model that should be able to show features of remapping.

A desired property of this model is that it should be biologically plausible, so that the results of the model can be interpreted in light of known results from neuroscientific remapping experiments.

The following research questions have thus been formulated for this project:

Research Question 1: Can place cells and grid cells be modeled to alter their firing activity in response to sensory information from the environment?

Research Question 2: Can such a computational model be made to be compatible with the phenomenon of global remapping?

Research Question 3: Can such a computational model be made to be compatible with the phenomenon of rate remapping?

Research Question 1 concerns whether the spatially responsive cells of the computational model can be made to consider the sensory signals that are received from the environment. This is an important precondition for the model to be able to perform remapping in response to environmental changes. Research Questions 2 and 3 then

ask whether the computational model can be made to show global remapping and rate remapping phenomena, respectively.

1.4 Literature Search

This section gives a summary of how the main literature for this project was sought out.

Two articles in the March 2013 issue of Nature Neuroscience, by researchers at NTNU's Centre for the Biology of Memory, used computational models of grid cells to verify the viability of the hypotheses they proposed as explanations for their reported experimental results [1, 4]. In both cases, the articles based their simulations on a model of grid cells described by Y. Burak and I.R. Fiete in 2009 [3]. The articles also point to a review article about computational models of grid cells by L.M. Giocomo, M.-B. Moser and E.I. Moser from 2011, which in turn also discusses Burak & Fiete's model [8].

This review article cites a review article on grid cells from 2006, by McNaughton et al., that gives a good overview of grid cell research in general [16]. The review article also cites the work of E.T. Rolls, S.M. Stringer and T. Elliot from 2006 on modeling place cells as competitive networks [21]. E.T. Rolls coauthored a text book on biologically plausible neural networks with A. Treves in 1998 [22]. This book includes references to a computational model of the hippocampus described by E.T. Rolls in 1995 [19].

In a conversation with T. Solstad from the Centre for the Biology of Memory, two articles on remapping were brought up; Leutgeb et al., 2005 [15] and Wills et al., 2005 [26]. Both of these articles are also referred by McNaughton et al.'s 2006 review article on grid cells [16]. In 2014, S. Leutgeb and J.K. Leutgeb published a book chapter about remapping, that gives a current overview on the research into remapping [14].

1.5 Report Structure

Chapter 2 presents background material on place cells, grid cells and models thereof, remapping phenomena, biologically plausible neural networks and models of the hippocampus. Chapter 3 describes the model that was implemented for this project, as well as the training and testing procedures that were used to examine the research

questions of this project. Chapter 4 presents and analyzes the results of the model and seeks to answer the research questions. Chapter 5 discusses the contributions in, the possible uses of and the future work regarding the model and makes a conclusion to the report.

2 Background

This chapter provides background material on several concepts encountered later in this report. Section 2.1 describes background knowledge about place cells and grid cells. Section 2.2 describes characteristics of the global and rate remapping phenomena and considers why they might occur. Section 2.3 describes some common biologically plausible neural networks that will be seen later in the report, while Section 2.4 discusses models of the hippocampus. Section 2.5 describes the continuous attractor model of grid cells, while Section 2.6 describes the competitive network model of place cells.

2.1 Place Cells and Grid Cells

Evidence of the hippocampus' role in navigation was first found when researchers discovered spatially responsive cells in the hippocampus of rats in the 1970s [17]. As these cells typically respond when the animal is within the boundaries of a single, contiguous area in the environment, the name "place cells" was motivated for these neurons. Place cells are found in the DG, CA3 and CA1 subareas of the hippocampus.

A key point of interest concerns the nature of the input signals that enable the place cells to determine whether the animal is inside or outside of the boundaries of the respective place fields of the neurons. An interesting finding was that the place fields remain stable even when the animal is tested in a pitch dark environment, i.e. the place cells are able to determine the current location even in the absence of visual input. This suggests that there is a *path integrator* somewhere in the brain, that keeps track of the total displacement of the animal based on self-motion cues.

When the Centre for the Biology of Memory at NTNU discovered grid cells in 2005 [10], these emerged as a candidate for the proposed path integrator. Grid cells are found in the Medial Entorhinal Cortex (MEC), a part of the entorhinal cortex adjacent

to the hippocampus.

Contrary to the singular place fields of the place cells, grid cells have multiple peaks of activity. These peaks are laid out in a highly regular hexagonal grid that repeats throughout the environment, thus gaining the grid cells their name. An illustration of the activity of a grid cell is shown in Figure 2.1.

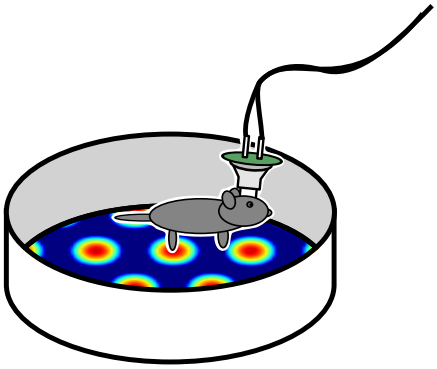


Figure 2.1: Example of grid cell activity. A rat is seen exploring a cylindrical enclosure, with an electrode inserted into the brain to measure the response of a grid cell. The grid cell's activity is plotted on the floor of the enclosure.

The grid-like maps traced out by the grid cells, present patterns of different sizes—measured by the spacing between adjacent peaks—as well as of different rotations. There are often found multiple grid cells with the same size and rotation of their respective grid maps, but with different offsets of the maps relative to each other. These grid cells are said to belong to the same *grid module*, but to have different *phases*. Within a given grid module, only a subset of the neurons will be active at the same time, namely the cells whose respective grid maps have peaks that align with the current position of the animal.

As all of the cells within the same grid module will have the same spatial response maps, except for the offsets of these maps relative to each other, a sufficient way of describing a grid cell is by expressing the grid map's offsets along the x and the y axes, i.e. the phase. Because the maps repeat for every shift of a distance equal the spacing between the map peaks, these phases should necessarily be given modulo the spacing distance. The specific subsets of cells that are active within a grid module thus reveal information about the animal's 2D coordinates *modulo the grid size* of the grid module.

The grid modules found in the rat brain have a wide range of grid sizes, ranging

from centimeters to meters, and this correlates with the location within the entorhinal cortex—the smallest-spaced grid modules are located dorsally (toward the back), while the largest-spaced grid modules are located ventrally (toward the front) [16]. Given inputs from multiple grid modules with a range of grid sizes, it is possible to pinpoint unique, non-repeating locations in the environment. This mechanism combines the resolution of the smaller-spaced grids with the "range of unambiguity" of the larger-spaced grids. The concept is illustrated in Figure 2.2, where a unique place field is obtained by summing the inputs from multiple grid modules and then thresholding the result appropriately.

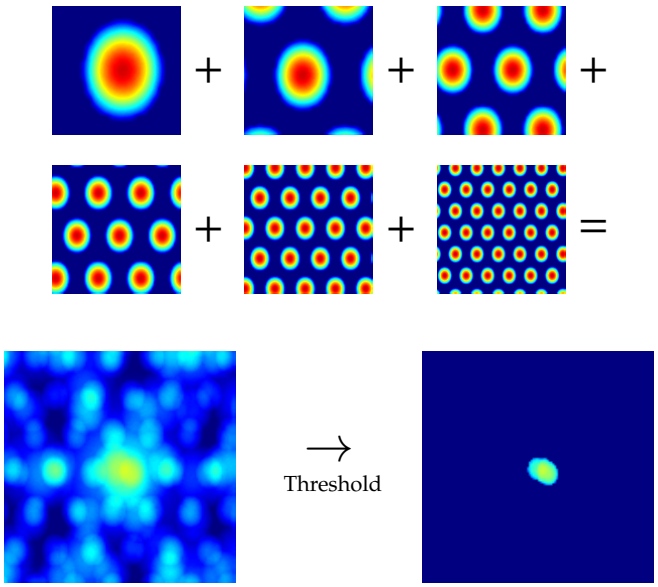


Figure 2.2: The sum of grid maps of different sizes generates a place field when thresholded appropriately. Figure inspired by [16].

The conclusion is that the grid cells collectively are able to represent arbitrary global coordinates in 2D space, by activating appropriate phases in the different grid modules. This makes the grid cell system a viable candidate for a path integrator, as the self-motion signals can be used to shift between these grid phases directly. Because the place cells in the hippocampus are located downstream from the entorhinal cortex, it is conceivable that the signals from the entorhinal cortex grid cells are what enables place cells to form in the hippocampus.

2.2 Remapping Phenomena

In many neuroscientific studies that attempt to uncover information about the wiring of the hippocampus and important properties of the hippocampal architecture, animal experiments are used where the neural responses are analyzed under various conditions. Often in these studies, the animals are tested in experimental environments that change over the course of the experiment. Examples of these modifications include changing the wall color and gradually morphing the shape of the box that encloses the animal—see Figure 2.3.

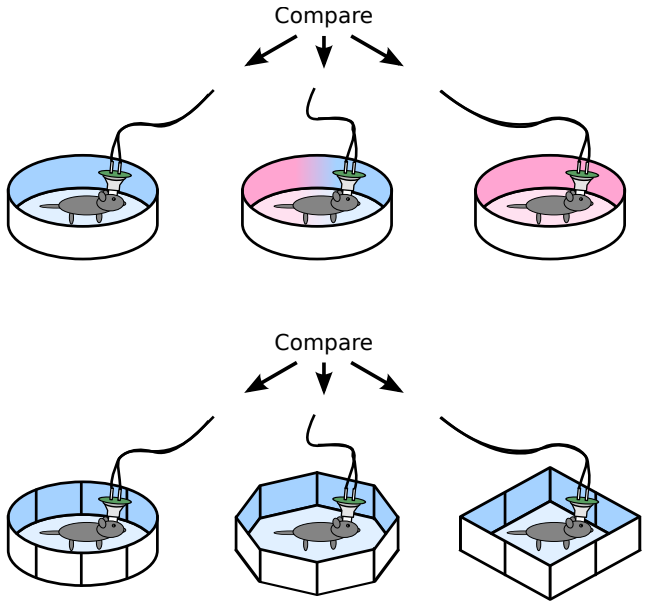


Figure 2.3: Examples of environmental changes performed during remapping experiments. A rat is seen exploring various enclosures, with an electrode inserted into the brain to measure neural responses in the hippocampal area. Above: Changing the color of the enclosure. Below: Changing the shape of the enclosure.

The changes that occur in the neural firing of the animal in response to these environmental modifications, are called *remapping*. The name implies that the brain has switched over to different spatial maps in the changed environment.

These changes in neural response that have been observed experimentally over time, can largely be grouped into two categories—rate remapping and global remapping. During rate remapping, the same set of place cells remains active in both environments, and the boundaries of the neurons' place fields stay intact. However, a dif-

ference is seen in the *firing rates* of these neurons—some place cells might fire more intensely in the modified environment, while other place cells might fire more weakly. There can be as much as tenfold changes in the peak firing intensities. Rate remapping is thus characterized by a persistence in the spatial extents of the place fields but with a possibly dramatic change in the distribution of peak firing rates across the population of place cells.

Global remapping, on the other hand, is characterized by the place cells changing over to using completely different place fields. Usually this entails that place cells that were active in the first environment go silent in the second environment, and vice versa. When comparing the activity of the cell population between the two environments—which subsets of cells are active and what their firing values are at corresponding locations—there is found to be no such correlation. In other words, the activity patterns in the hippocampus before and after global remapping are completely orthogonal to each other [14].

Global remapping and rate remapping also differ in the responses seen to testing situations that are *intermediate* to the environments the animal has been trained in, for example the mixed-color or octagonal enclosures seen in Figure 2.3. With global remapping, there is an *abrupt change* in response that happens at some midway point between the two training environments, while the rate remapping response is a *gradual transition* between the responses of the training environments.

The two contrasting types of response seen to environmental changes, raises the question of what it is that determines the type of remapping seen. Is it simply the magnitude of the environmental change that triggers a global remapping response instead of a rate remapping response, or are there different mechanisms at play?

Experimental evidence has given clues as to how this might be understood. When global remapping is seen in the hippocampus, the activity of the grid networks in MEC changes simultaneously. During rate remapping, however, the activity in MEC remains the same as before. As explained in the previous section, the activity in MEC can be seen as encoding the animal's global coordinates in 2D space. It thus appears as though global remapping is accompanied by a change in the animal's *belief about its current location*, whereas there is no such change for rate remapping.

Therefore, the view has eventually emerged that the MEC state is an important part of the mechanism that selects the spatial maps used by the hippocampus and that then also determines whether global or rate remapping is seen. The actual locations of the experimental enclosures within the laboratory *during training* has thus received atten-

tion as an important parameter in remapping experiments, as the specific coordinates of these locations will determine the MEC state that the animal will associate with its first impressions of the training enclosures. These place-to-event associations, or MEC-state-to-event associations, will be committed to memory and then later affect the remapping outcome in the test trials.

This view will be the premise for this project too. The model will be trained in simulated enclosures that are distinguishable by the sensory signals that reach the hippocampus, for example due to the boxes having different wall colors. During the test trials, the model will be "confused" as to its current location, by having the MEC states reset to silent initial conditions. The model will then receive sensory inputs corresponding fully or partly to one or more of the previously seen training environments, and the response of the model will then be measured.

By training in environments where the MEC states are different between the training environments, we expect that a global remapping response will be seen during testing. By the same logic, we expect to see a rate remapping response from the model if the model is trained with the *same* MEC state in the various training environments.

2.3 Biologically Plausible Neural Networks

2.3.1 Learning in the Brain

Artificial neural networks seen in the fields of Artificial Intelligence and Machine Learning are derived from knowledge about neural networks found in the brain, but they are often not biologically plausible themselves. Artificial neurons and real-life neuron have in common the integration of inputs from afferent neurons across synapses of varying strength, the transformation of the input sum through an activation function and the propagation of this activation value to efferent neurons. Artificial neural networks are commonly arranged in *layers* of neurons, which is not uncommon in the brain either. Where the two differ, however, is usually in the schemes employed to achieve *learning* in the neural networks.

Backpropagation has been a popular learning method within artificial neural networks in the recent decades. This method works by propagating error values backwards from the output layer of the network, using the differential on the error value of each

synaptic weight as a guide in adjusting these synapses. While backpropagation has been a successful algorithm, there is no evidence that this is a biologically plausible mechanism. Learning in the brain occurs *locally* in each synapse—there is only available local information about the presynaptic and postsynaptic firing rates, not any backpropagated measure of global error. More generally, there is only *local interaction*, i.e. between adjacent components, requiring information to physically travel in order to affect distant parts of the system.

Donald O. Hebb postulated a plausible local learning rule in 1949 [13], that has later been known as *Hebbian learning*. He suggested that synapses are strengthened whenever the presynaptic and postsynaptic neurons fire about simultaneously, the rationale being that this pattern indicates a possible causal relationship between the two neurons. If, however, there is activity in one of the neurons without activity in the other, the causality seems less probable and the synapse will thus be weakened. Hebb's learning rule has often been summarized as "Cells that fire together, wire together".

Later discoveries has confirmed the biological plausibility of Hebb's rule. *Long Term Potentiation* (LTP) is a learning mechanism first discovered in the 1960s, in experiments where neurons in the hippocampus of rabbits were subjected to high-frequency electrical stimulation. The stimulated neurons would afterwards produce stronger postsynaptic responses, and this effect would last for periods as long as several months. This long-term increase in synaptic efficacy can be seen as an example of Hebbian learning in the brain.

The remainder of this section will describe some common biologically plausible neural network architectures that build on the principles of local learning and local interactions, that have been described by E.T. Rolls and A. Treves in their 1998 text book on this topic [22]. These network types are useful to consider when examining neural networks found in the brain. They will also be seen later in this report, as parts of the computational model of the hippocampus.

2.3.2 Pattern Association Networks

Pattern association networks receive two kinds of inputs, *event* axons and *context* axons. The goal of the pattern association network is to learn to recognize the contexts that are associated with each event. In other words, if the pattern of inputs on the context axons closely correspond to patterns previously seen when a given event axon was active, the pattern association network should produce the same output response

as if the event axon itself were active.

Pattern association networks often have a structure as shown in the schematic in Figure 2.4. The triangles represent pattern associative neurons, with the bold, vertical lines above them representing the dendritic inputs to the neurons, and the arrows extending below them representing the output axons of the neurons. The vertical arrows pointing down on the dendrites represent incoming event axons from afferent neurons, that each strongly drive the respective pattern associative neuron through an unmodifiable synapse. The horizontal incoming arrows represent the context axons, where each axon forms synapses with all of the pattern associative neurons. These synapses are modifiable, indicated by the bulbs on the dendritic spines.

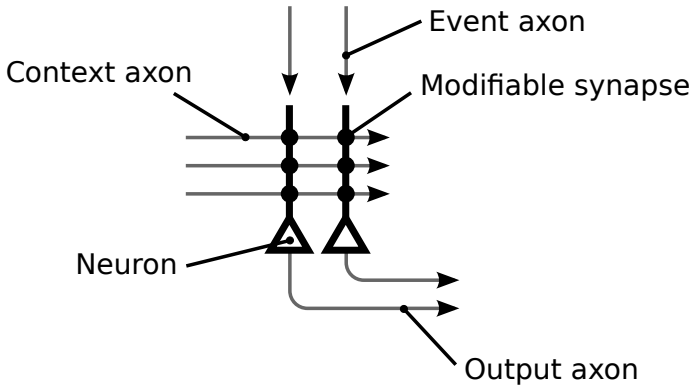


Figure 2.4: Schematic representation of a pattern association network.

The output firing r_i of each neuron i is determined by the sum of the incoming axon values scaled by the weights of their respective synapses:

$$r_i = f\left[e_i + \sum_j r'_j w_{ij}\right] \quad [22]. \tag{2.1}$$

In the formula above, e_i is the event input to neuron i, r'_j is the value of the j-th context axon and w_{ij} is the synaptic weight between neuron i and context axon j. The result is mapped through a function $f(x)$, which is an activation function that is used to introduce some non-linearity in the response of the neuron, e.g. by forcing the output to zero when the input sum is below a certain threshold.

The key point is how the synaptic weights w_{ij} are updated in order to learn to recognize the context patterns that correlate with the firing of the neuron. The weight

update rule

$$\delta w_{ij} = k r_i r'_j \quad [22], \tag{2.2}$$

where k is a *learning rate*, can be used for this purpose. As seen from the formula, each synaptic weight is updated in proportion to the product of the presynaptic and the postsynaptic firing rates. This is an example of the Hebbian learning rule described above.

The values of all of the context axons r'_j can be viewed as a vector $\mathbf{r}'$ of context information, and the same applies to the synaptic weights w_{ij} for a given neuron i, which can be written as the vector $\mathbf{w}_i$. The sum in Equation 2.1 can then be written as the dot product between these two vectors: $\mathbf{r}' \cdot \mathbf{w}_i$. The job of a pattern associative neuron is to maximize this dot product for the context vectors $\mathbf{r}'$ that correlate with the firing of the event axon e_i, which is the same as aligning the weight vector $\mathbf{w}_i$ in the direction of the relevant context vectors $\mathbf{r}'$. This is the *vector interpretation* of the operation of pattern association networks.

2.3.3 Competitive Learning Networks

Competitive learning networks detect *clusters* in the set of input vectors. After a competitive learning network has been trained, each of the neurons will have learned to respond to input vectors that lie within the boundaries of the cluster it is responsible for. These clusters are non-overlapping, and together cover all of the input vectors.

An example of a competitive learning network is shown in Figure 2.5. Competitive learning networks are similar to pattern association networks in the output value calculations and weight update rules, but they don't receive any event axons as the pattern association networks do.

As seen in the figure, the neurons are affected by two kinds of signals. The first kind of input is the input axons, with modifiable synapses as for the pattern association networks. The second kind of input is a recurrent signal feeding back from all of the neurons in the competitive network. These signals are *inhibitory*, so that the neuron that is the most strongly activated by the input axons, will prevent the other neurons in the network from firing. This is called mutual inhibition.

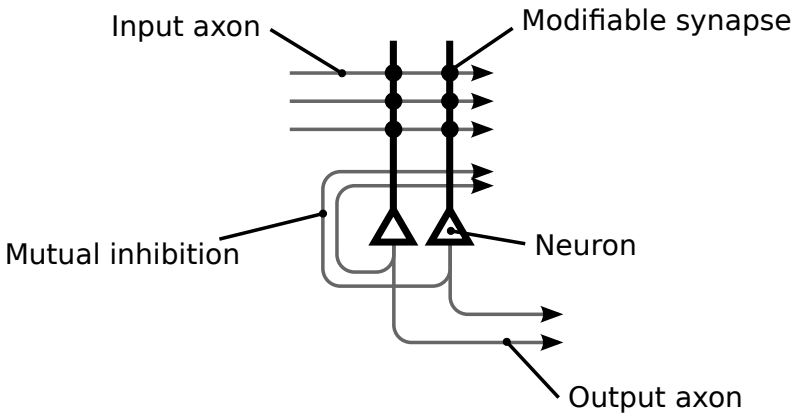


Figure 2.5: Schematic representation of a competitive learning network.

The inhibition is implemented as an additional step after the output firing rates have been calculated as in Equation 2.1, where the most strongly activated neurons will have their firing strengthened, while the less strongly activated neurons will have their firing suppressed. A common mechanism is the "winner-takes-it-all" rule, where the most strongly activated neuron gains an output value of 1, while the rest of the neurons get output values of 0. The learning rule takes these adjusted output values into account, so that the inhibited neurons will not have their synaptic weights updated.

An important feature of competitive learning networks is that they *orthogonalize* the firing vectors. The output vectors of a competitive network will only have one component active at the same time, due to the winner-takes-it-all rule. This firing vector will thus be orthogonal to all other possible firing vectors, and there will therefore be no correlation between the different output vectors from the network. This holds true even for inputs vectors that are closely correlated, assuming that they have been assigned to different clusters by the competitive network. The competitive network will thus either strengthen the correlation between two input vectors or make it disappear, depending on whether the vectors end up in the same cluster or in different clusters.

This effect is called *pattern separation* and is an important feature in neural networks for information processing. Pattern association networks, for example, perform better when the inputs have been orthogonalized by a competitive network beforehand. The well-known issue for neural networks of discriminating between inputs that are not linearly separable can be solved by first passing the signals through a competitive network.

2.3.4 Autoassociation Networks/Attractor Networks

Autoassociation networks are, as the name suggests, pattern association networks that receive inputs *from themselves*. An example of an autoassociation network is shown in Figure 2.6. As seen in the figure, the neurons receive event axons in the same way as the pattern associative neurons do. There is also recurrent connectivity from all of the neurons feeding back into the network, similar to the competitive networks. The recurrent axons in the autoassociation network are not inhibitory, however, but have modifiable synapses that work in the same way as for the context axons in the pattern association networks.

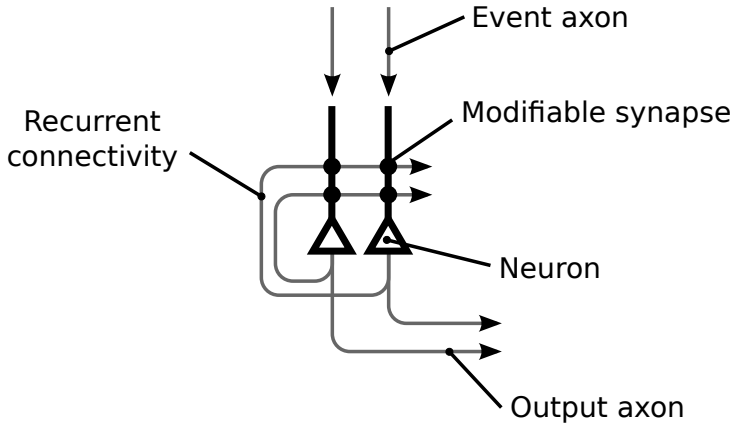


Figure 2.6: Schematic representation of an autoassociation network.

The recurrent connectivity enables neurons in the network to be activated by familiar patterns of firing in the network itself. This means that once a trained pattern has been activated in the network through the event axons, the recurrent axons will ensure that the state of firing is maintained even after the event input recedes. This stable state is known as a *basin of attraction*, as the network activity "falls into" this state and remains there. The autoassociation network is thus an example of an *attractor network*, characterized by the stable maintenance of firing patterns in basins of attraction.

Autoassociation networks, and the more general attractor networks, have two important features that make them interesting objects of study. The maintenance of stable states, in the absence of input that forces a change in the current state, implements a form of *short term memory*. The autoassociation network "remembers" the current state until the network is updated with a different state of firing. An analogy from computer science would thus be that of a CPU register.

Additionally, due to the attractor nature of the stable network states, the network can recall a complete pattern of firing throughout the network based on incomplete inputs. Given parts of a previously trained input vector, the network will fall into the closest basin of attraction and reconstruct the full output pattern. This is known as *pattern completion* and is an important feature of memory systems in the brain.

2.4 Modeling the Hippocampus

2.4.1 Introduction

This project will implement a biologically plausible model of the hippocampus, in order to examine how remapping phenomena occur. It is thus important to have background knowledge about the anatomy of the hippocampus, and it is also useful to consider how this has been represented as biologically plausible neural networks by others.

This section starts with an anatomical overview of the principal neurons in the hippocampus and how these are connected. Afterwards a computational model of the hippocampus, using concepts introduced in Section 2.3, is presented. The models described in this section are based upon Rolls & Treves' 1998 text book [22].

2.4.2 Anatomical Overview

Figure 1.2 in Section 1.1 showed a high-level overview of the connectivity between the entorhinal cortex and the various subareas of the hippocampus. Figure 2.7 expands upon this by showing a more detailed schematic of some of the principal neurons in the hippocampus [20]. The figure shows the principal neurons in the four main subareas DG, CA3, CA1 and subiculum and how these are connected to the entorhinal cortex.

The major source of inputs to the hippocampus is the *perforant path*, originating in the entorhinal cortex and projecting outputs to DG and CA3. The dentate gyrus implements mutual inhibition through inhibitory interneurons, indicated by the recurrent connection in the figure. The outputs from DG to CA3, called the *mossy fibers*, have

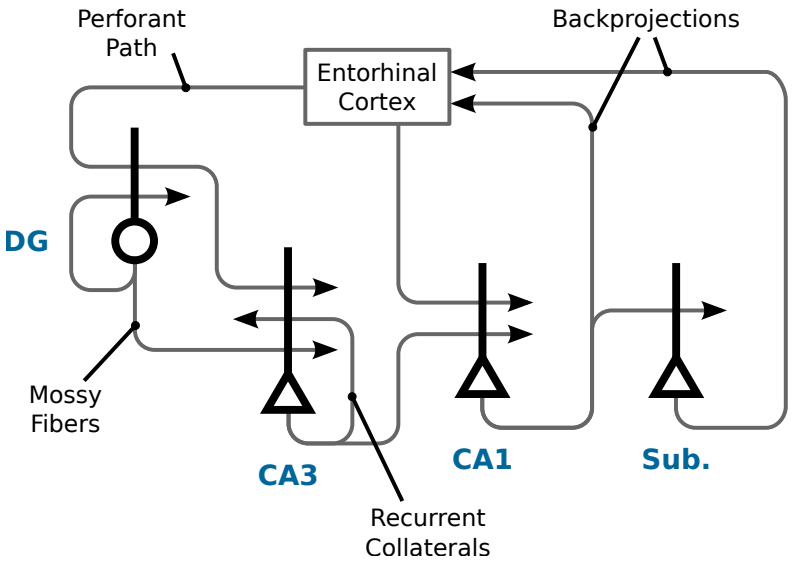


Figure 2.7: Schematic representation of some of the principal neurons in the hippocampus and their connectivity [20].

a highly *diluted* connectivity with CA3, i.e. each mossy fiber makes contact with relatively few CA3 neurons. However, the mossy fibers have a strong influence on the CA3 neurons that they do connect with.

CA3 neurons, in addition to receiving inputs from the mossy fibers and the perforant path, also receive inputs from the *recurrent collaterals*, which are recurrent circuitry within CA3. CA3 neurons project to CA1, which also receives inputs from the entorhinal cortex. CA1 projects to the subiculum and back to the entorhinal cortex. The signals from CA1 to the entorhinal cortex are an example of hippocampal *backprojections*, which are the means by which the hippocampus influences the rest of the brain. The subiculum also originates backprojections to the entorhinal cortex.

2.4.3 Neural Network Model for Hippocampal Information Processing

Rolls & Treves describe a biologically plausible neural network model of the hippocampus, built using the concepts introduced in Section 2.3 and based on the anatomical knowledge outlined above, among other sources of information [22]. The model was trained and tested on arbitrary vectors of data, and the study was thus concerned

with the general information processing abilities and memory related properties of the hippocampal model.

DG is implemented as a competitive learning network, orthogonalizing the perforant path inputs from the entorhinal cortex. The competitive aspect is motivated by the mutual inhibition known to be implemented by the inhibitory interneurons in DG.

The orthogonalized signals from DG are passed on to CA3, along the mossy fibers. In addition to the mossy fiber inputs, the CA3 neurons also receive perforant path inputs independently of DG. CA3 is implemented as an autoassociation network, where the mossy fiber inputs drive CA3 neurons strongly, while the perforant path inputs drive more weakly. This allows attractor states in CA3 to be formed based on the orthogonalized DG inputs, but to later be recalled based on the original perforant path inputs.

CA1 is implemented as a competitive learning network, receiving inputs from CA3 and from the entorhinal cortex. The task of CA1 is to reconstruct previous states of firing in the entorhinal cortex based on recall performed by CA3. Accordingly, CA1's output signals are sent to the entorhinal cortex as backprojections.

The high-level view of this model is thus as follows: The entorhinal cortex is both the main source of inputs and the main target of outputs from this model, but during separate phases—information flows from the entorhinal cortex to the hippocampus during training, but from the hippocampus to the entorhinal cortex during recall testing, based on retrieval cues on the perforant path inputs.

DG performs pattern separation on the hippocampal inputs, to enable CA3 to distinguish between input clusters that have some degree of correlation between them, and to learn different concepts for these clusters. CA3 has the ability to learn arbitrary associations between these concepts, through the autoassociation network implemented by the recurrent collaterals. This can be seen as an implementation of episodic memory, where each episode contains associations to sensory stimuli from a wide range of modalities.

During recall, the recurrent collaterals enable CA3 to activate all parts of an episodic memory based on perforant path cues that initially activate only a part of the memory. This is an example of pattern completion. The various activated parts of an episodic memory are combined in neurons in the competitive network of CA1 as a precursory stage to backprojections to the cortex. CA1 performs a *recoding* where the activity of various subsets of different episodic memories in CA3 are represented by unique

neurons in CA1. These neurons, representing conjunctions of neurons in CA3, provide an efficient way of projecting signals back to the cortex.

The backprojections from CA1 and the subiculum reach the entorhinal cortex, where they reactivate neurons that were coactive with the CA1 neurons are the time of learning. These entorhinal cortex neurons represent the original inputs that the hippocampus should restore, so the hippocampal loop has thus been closed when the backprojections reach the entorhinal cortex. In the case of a successful recall, the full previous pattern of activity in the entorhinal cortex should have been restored, based on a partial input to the hippocampus.

2.4.4 Summary

The model presented above offers a suggestion for how the memory features of the hippocampus might be implemented. However, the model does not consider how the spatial features of the hippocampus, such as place cells, come about.

As mentioned in Section 2.1, input signals from path integrating networks of grid cells in MEC are sufficient for the hippocampus to be able to form place cells. The next section thus introduces a computational model of grid cells that will be used to model the entorhinal cortex for this project. The next section thereafter discusses how these signals from the entorhinal cortex then can be used to form place cells in the hippocampus.

2.5 Continuous Attractor Model of Grid Cells

2.5.1 Introduction

Several types of computational models for grid cells have been proposed, that attempt to account for both the path integrative abilities of the grid cells as well as the characteristic hexagonal firing patterns [8]. The two types of model that have received the most attention are the *oscillatory-interference models* and the *continuous attractor models*.

The oscillatory-interference models propose that grid cells receive inputs from multiple directionally tuned neurons that produce very precise oscillations. The directionally tuned neurons respond to movement along their preferred axes by temporarily adjusting the frequencies of their output oscillations. These frequency differences will result in a drift in the peaks of the sum of the oscillations, and it is proposed that it is these shifting peaks that become visible through the grid cells.

A disadvantage of these models is that they are sensitive to timing discrepancies, e.g. due to the noisy signals inherent in the brain. Continuous attractor models, on the other hand, do not depend on these precise timing relationships, but instead view the different states of the grid modules as different basins of attraction in specially crafted attractor networks.

An attractor model of grid cells will thus be used for this project. Attractor models are well-suited for integration into hippocampal models such as the one described above, e.g. due to the emphasis on firing rates instead of on precise timing values. The rest of this section presents some of the theory behind these continuous attractor models of grid cells.

2.5.2 Ring Attractor Model of the Head-Direction System

Attractor networks were first studied as a possible mechanism in navigation as a model for how the *head-direction system* works. The head-direction system consists of neurons that are active only when the animal's head is turned in a specific direction. Attractor networks provide an explanation for how these head-direction cells can keep track of the current direction of the head, given only turning speeds as input.

In the attractor model of the head-direction system, head-direction specific neurons are recurrently connected to each other depending on the difference in their preferred head angles. Neurons that respond to similar directions are strongly connected in an excitatory fashion, while neurons with different directional preferences inhibit each other. This is illustrated in Figure 2.8. The connection strength between the neurons as a function of the difference in their preferred head angles is called the *connectivity profile* of the network. Because the connection strength is a function of angle difference only, the connectivity between every pair of neurons will be symmetric.

Given ongoing activity in one of these head-direction cells, the excitatory recurrent connectivity with itself and similarly tuned neurons—which excite symmetrically in

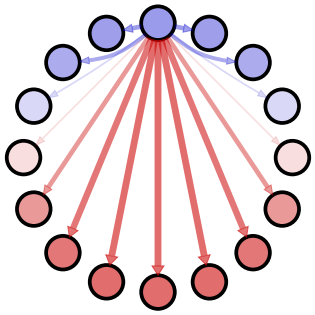


Figure 2.8: An example of how a neuron in a continuous attractor model of the head-direction system might receive excitation and inhibition from the other neurons along the ring network. Line thickness and neuron color represents connectivity with the topmost neuron. Red indicates inhibition and blue indicates excitation. Figure inspired by [16].

return—will ensure that the activity is maintained over time. Moreover, the inhibitory connections to differently tuned head-direction cells will deny any other peaks of activity in the network. Thus, the connectivity profile imposes a structure on the network wherein only a single "packet" of activity can be present at a time. This activity packet will keep itself active through recurrent excitation.

2.5.3 Stable States and the Attractor Manifold

The patterns of neural activity in the head-direction network that adhere to this structure of a single packet of activation, are thus the stable states/attractor states of the network. All other patterns of activity are unstable, and these patterns will eventually subside to one of the attractor states.

To get the network to change state—to move the packet of activity—the network requires external input. By exciting and/or inhibiting certain neurons in the network, a new activity pattern can be forced. When this external input eventually ceases, the network will settle into the attractor state most closely matching the imposed network activity, as explained previously for autoassociation networks.

In the ring attractor model of the head-direction system, this means that an external input representing e.g. a right-turning movement of the animal's head, should cause

neurons clockwise of the current activity packet to be excited. Likewise, a left-turning movement should excite neurons counter-clockwise. These "shifted excitations" will cause the attractor network to convert to new stable states.

The *attractor manifold* is the abstract space that consists of all of these possible attractor states of the network. Some such states are more easily transformed between than others—in the head-direction system, states that represent similar directions require much smaller external inputs to change between than states representing dissimilar directions. The states that are easily transformed between, are connected in the attractor manifold. The head-direction system, allowing the stable network activity to easily shift either clockwise or counter-clockwise, is thus said to have a *one-dimensional attractor manifold*.

In the head-direction system, the various attractor states of the network represent values of a continuous property, in this case head angle. The network is thus called a Continuous Attractor Network (CAN) to signify that the attractor network encodes a continuous variable.

2.5.4 Spontaneous Pattern Formation

How does the aforementioned single packet of network activity form in the first place? Consider an initial state of the network where all of the neurons are activated uniformly—for example, experiencing no activation at all. The connectivity profile causes all of the neurons to have the same set of input weights, only in a different order for each neuron. Thus, all of the neurons will receive the same amount of input at every moment. In every new time step of the network, all of the neurons will update to the same new output value, and the network activity remains uniform.

To resolve this indefinite situation of uniform network activity, a small amount of randomness has to be introduced to the network. In nature, this is taken as a given due to the inherent noise in biological systems. The randomness will "break the symmetry", and the patterns imposed by the structure of the system will start to form. This mechanism is related to *Turing instability*, a phenomenon of pattern formation in chemical systems seen in biology [24].

In computer models of attractor networks, the required randomness can be presented either as weak noise in the initial activation values of the neurons or as stochasticity in the firing outputs of the neurons.

2.5.5 Continuous Attractors for Two-Dimensional Variables

Soon after the discovery of grid cells in 2005, it was proposed that the grid cells in the MEC comprise continuous attractor networks that encode the animal's position as two-dimensional variables [8]. M.C. Fuhs and D.S. Touretzky extended the one-dimensional ring in the CAN model of the head-direction system to a *2D sheet* of interconnected neurons to represent 2D locations in the environment [7].

While the connection strengths between neurons in the ring attractor networks depend on distance along the ring, the 2D attractor networks of Fuhs & Touretzky use the Euclidean distance between neurons in the sheet to determine their connection strengths. See Figure 2.9 for an illustration of how such a neuron might be connected to the other neurons in the 2D sheet.

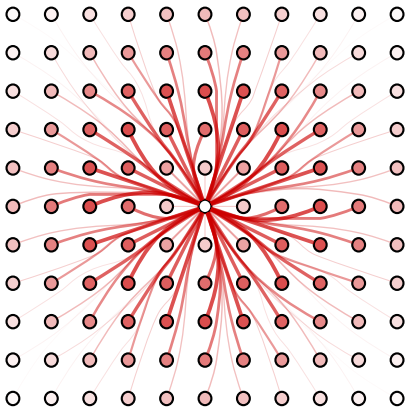


Figure 2.9: An example of how a neuron in a 2D continuous attractor network might be inhibited by the other neurons in the sheet. Line thickness and neuron color represents strength of inhibition on the center neuron. Figure inspired by [16].

Due to the connectivity profile and the symmetry breaking discussed above, the 2D sheet of neurons will gain packets of activity similar to those in the head-direction network. Given that the inhibition in the connectivity profile fades off sufficiently quickly at longer distances, the sheet will have room for multiple such packets of activity. These packets will self-organize to fill the available space as much as possible, which happens to result in a hexagonal pattern similar to the patterns known from grid cell recordings in animals.

2.5.6 From Network Activity Patterns to Single-Neuron Responses

The fact that the activity pattern in the neural sheet resembles the hexagonal pattern recorded from grid cells is significant for the model to match the experimental data. It is important to note, however, that the pattern seen in the sheet of neurons in the CAN model is of a different nature than the activity patterns seen from experimental neuron recordings. In these experiments, the activity level—the rate of spiking—of single neurons is plotted against the animal's position within the environment. The plots thus show *how individual neurons respond spatially*. The measured response is referred to as the *single neuron response* (SN response).

The patterns seen in the plots of CAN models' neural sheets, on the other hand, represent the *collective activity* of entire population of neurons at given points in time. Obtaining a plot similar to these from empirical data would not be an easy exercise. First, the orderly 2D layout of neurons in a CAN model's neural sheet is merely a convenient simplification. Within single grid modules in animals, the locations of the individual neurons do not appear to have any correlation with their spatial phases [10]. Second, a prohibitively large number of neurons would have to be simultaneously recorded from in order to get all of the point-in-time information conveyed in these plots.

What is then the relation between the network activity plots from the computational models and the single-neuron spatial maps obtained in experiments? Recall how the head-direction CAN model uses turning speed inputs to drive the activity packet around the ring of neurons. Similarly, the 2D CAN models of grid cells use velocity inputs (direction and speed) to shift the entire network pattern around the neural sheet. The direction and distance of the offset is proportional to the corresponding movement in the real environment.

Viewed from the perspective of a single neuron, this coordinated movement of the entire pattern will cause the neuron to "see" different parts of the global pattern at different times. This concept is illustrated in Figure 2.10. When the single neuron's response is plotted on a line that tracks the animal's movement—upon which the shifting of the network pattern is based in the first place—the resulting picture will reveal the total pattern of network activity.

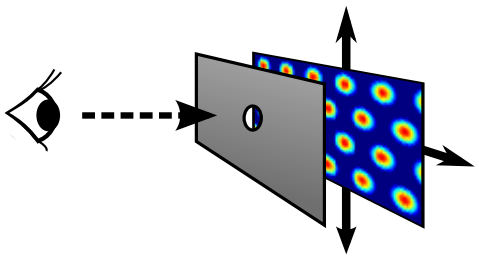


Figure 2.10: A single neuron in a continuous attractor network will "see" different parts of the global pattern as the pattern is shifted around the network, analogous to watching an image moving around behind a screen with a tiny hole through it.

2.5.7 Shifting the Network Activity Pattern

One way of achieving the shifting of the activity pattern described in the previous section, is to assign directional preferences to the neurons in the sheet [3]. The neurons whose directional preferences align with the direction of movement will receive extra external input. By then introducing offsets in the connectivity of the neurons depending on their directional preferences, the activity pattern can be made to shift in the intended direction.

An illustration of how this offset might be implemented is shown in Figure 2.11. This figure shows the four center neurons colored according to their directional preference, for example north, south, west and east respectively for the purple, green, blue and red neurons. The red neuron receives its incoming signals from the other neurons in the sheet *as if it were* the neuron immediately to its left. Similarly, the blue neuron is connected as if it were the neuron to its right. The connectivity profiles for the purple and green neurons are omitted for the sake of clarity.

In the absence of movement, none of the neurons will receive any extra input. Assuming that all of the four directional preferences are uniformly distributed across all of the neurons in the sheet, the differently offset connectivity profiles will cancel each other out and the activity pattern will stay in the same place. During movement, however, the neurons with their directional preferences aligned with the direction of movement will receive extra input that causes these neurons' offsets of the connectivity profiles to be reinforced, thus inducing movement in the activity pattern.

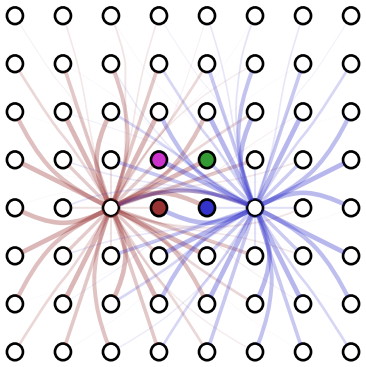


Figure 2.11: An example of how neurons in a 2D CAN sheet can have different directional preferences. The red and blue neurons are shown to receive their input signals in an offset fashion that corresponds with their preferred input velocity direction.

2.5.8 Grids of Different Scales and Rotations

As described in Section 2.1, grid cells appear to belong to distinct grid modules that exhibit different scales and rotations of their grid maps [10]. This is an important feature of grid cells, as the collective activity of an assortment of grid patterns is required to generate unique place fields in the hippocampus.

These variations of the grid pattern are accounted for in the CAN model by *modulating the velocity inputs* differently in the various grid modules. Scaling the speed input by 50 %, for example, will cause the network pattern to respond more slowly during the same displacement of an animal. The spatial response of single neurons will thus show grid patterns that are about 200 % the size of the unscaled patterns. Similarly, rotation of the grid pattern is achieved by offsetting/adding a bias to the input angle.

CAN grid modules of different scales and rotations will thus appear identical when looking at the network patterns, but the spatial responses of single neurons will reflect the different velocity modulations.

2.5.9 Network Topology

Fuhs & Touretzky's model uses a sheet of neurons where neurons near the edges of the sheet are left unconnected in the directions pointing away from the center of the

sheet, where there would otherwise be neurons had the sheet been infinitely large. The edge regions of the sheet thus lack some of the inhibitory input that the more central regions receive. Because of this, the neurons along the border of the sheet will have a higher level of activity than in the central regions. This disrupts the rest of the activity pattern, allows the activity pattern to rotate and prevents the CAN from responding to slow velocity inputs [2].

Y. Burak and I.R. Fiete improved upon Fuhs & Touretzky's model by introducing a *periodic connectivity* in the network [3]. Neurons along the left edge of the sheet connect to neurons along the right edge, as if duplicates of the sheet were tiled around the edges to surround the entire sheet. The same applies to the top and bottom edges. Geometrically speaking, connecting the top and bottom edges of a sheet can be visualized as forming a horizontal cylinder. Connecting the left and right edges of this cylinder will then result in a torus shape. This type of "wrap-around" connectivity is thus called a *torus topology*.

With the periodic connectivity of Burak & Fiete's model, any border effects are eliminated—the activity pattern behaves the same along the edges as in the center of the sheet. This mitigates the pattern rotation and speed sensitivity issues of Fuhs & Touretzky's model, thus improving the navigational performance of the model [3].

2.6 Competitive Network Model of Place Cells

With the grid cell signals from the entorhinal cortex in place, based on the model presented in the previous section, it is interesting to see whether hippocampal architectures similar to the one outlined in Section 2.4.3 can implement place cells based on grid cell inputs. The DG, CA3 and CA1 areas of the hippocampus are all known to contain place cells. Given the unidirectional connectivity of the hippocampus, with the DG as the first submodule and the major originator of signals to the rest of the hippocampus, it is natural to first look to the DG for a possible implementation of place cells.

DG is thought to implement a competitive learning network. The desired functionality of a place cell is to recognize a contiguous area in the environment. This is equivalent to responding to a cluster of vectors in the input space, assuming a grid nature of the inputs from MEC. As this is precisely the function implemented by competitive learning networks, this is thus a viable network type to explore for a network of place

cells.

Rolls et al. successfully used a competitive learning network, as a model of DG, to generate place cells from grid cell inputs in 2006 [21]. In this work, the grid cells inputs were represented by an explicit formula using Gaussian functions to create the peaks in the grid pattern, instead of modeling the MEC grid networks directly. For my preliminary project to this thesis, however, I used a similar competitive learning network to generate place cells based on inputs from continuous attractor networks of grid cells in an integrated model. The combination of continuous attractor networks for grid cells and a hippocampal network inspired by the principles of the model presented in Section 2.4.3, is thus a viable framework for modeling remapping phenomena.

3 Model and Implementation

3.1 Overview

Figure 3.1 shows an overview of the model implemented for this project in order to examine the phenomena of remapping.

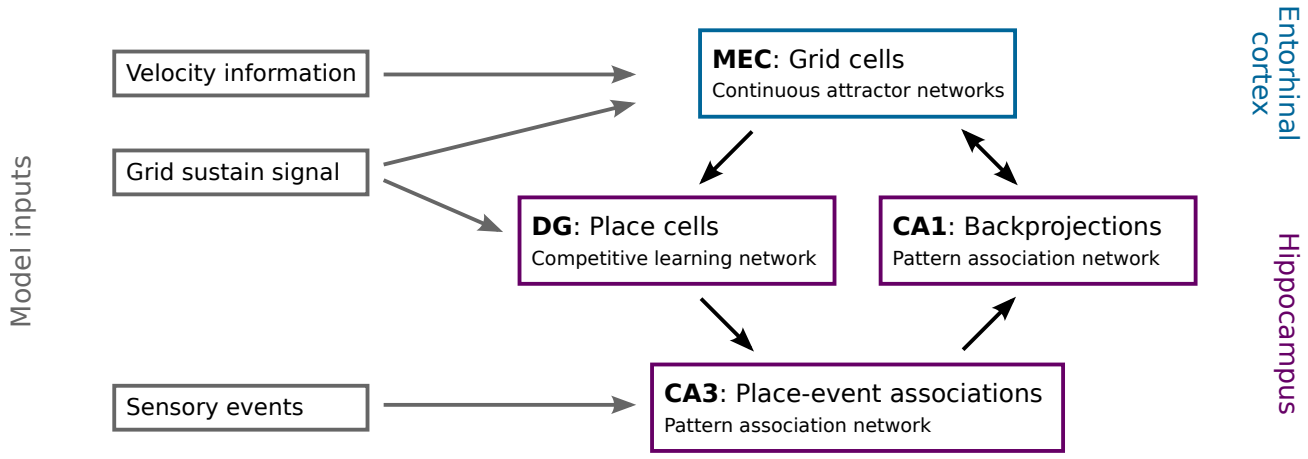


Figure 3.1: Schematic overview of the implemented model.

A series of continuous attractor networks, based on the Burak & Fiete model, represents networks of grid cells in the medial entorhinal cortex. The collective activity of these grid modules encodes the animal's belief about its 2D coordinates in the environment at any given point in time. The output of these grid modules is passed on to a competitive learning network, representing the dentate gyrus of the hippocampus, that will recognize clusters in the information from the grid cells and thus generate place cells.

Signals from these place cells are then sent to a pattern association network representing the CA3 subarea of the hippocampus. The CA3 network also receives sensory signals about the current location in the environment. These signals are of an unspecified

kind, but can be considered to contain information about e.g. smell, surface texture or color at the current location. The task of the CA3 network is to associate the visited places with the sensory experiences at these locations.

CA3 passes its information on to CA1, which is also a pattern association network. CA1 associates firing of CA3 cells with firing of MEC grid cells. In other words, CA1 can reconstruct old states in the MEC grid modules based on previously seen firing patterns in CA3, as a form of hippocampal backprojections. Before reinstatement of MEC state from CA1, the attractor networks in MEC should first have their current internal states cleared out. This is facilitated by a "grid sustain signal" to MEC/DG that determines whether the current states should be maintained.

The network will be trained in different environments, where CA3 will learn to associate places with sensory events. The goal of the model is to later then be able to recall information about the location from sensory inputs alone. The network will be supplied with sensory information in both unambiguous and ambiguous configurations, and the response of the network will then be compared with the response seen during training. Can the model distinguish between the trained environments based on the sensory information alone, and how does this relate to the real-world phenomena of global remapping and rate remapping?

Section 3.2 describes the inputs to the model, before the continuous attractor model of the MEC grid cells is described in Section 3.3, the competitive learning model of the DG place cells is described in Section 3.4 and the pattern association networks representing the CA3 and CA1 areas are described in Section 3.5. Section 3.6 lists the parameters used, while Section 3.7 explains how the model is trained and tested. Section 3.8 discusses the biological plausibility of the model, while Section 3.9 makes a few notes about the implementation and some optimizations employed. Finally, a summary of the complete neural network is given in Section 3.10.

3.2 Model Inputs

3.2.1 Animal Velocity

During training and testing, the MEC grid cells receive information about velocity—animal speed and direction—as the animal moves around in the environment. The

path that the animal travels is predetermined, and is kept the same for all training and testing rounds.

The path used for the simulations in this report is based on a ten minute recording of a real laboratory rat foraging in a cylindrical box with a radius of 90 cm. A plot of the ten minute path is shown in Figure 3.2. This recording was made public as part of a dataset released from the study of grid cells by Hafting et al. in 2005 [10, 11].

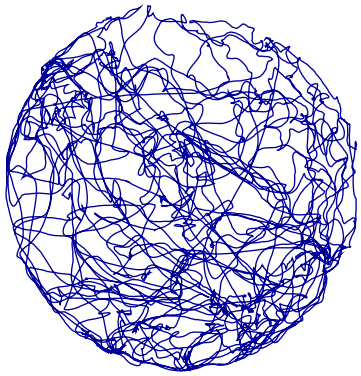


Figure 3.2: The animal path used during training and testing of the model [11].

3.2.2 Spatially Contingent Sense Signals

A key objective of the model is for the CA3 neurons to learn associations between places in the environment and sensory events that occur at these locations, in order to later be able to recognize the current location based on sensory inputs alone. The details of these sensory inputs are thus an important aspect of the model. As mentioned previously, the sensory signals are not defined to be of any specific kind, but can be considered to represent information about for example color, surface texture or smell.

The model supports a configurable amount of these sensory input axons. For each training and testing configuration, the user can specify for each sensory axon how it should respond at the different locations in the environment. This is done by supplying a grayscale image file to act as a map for the sensory response. Each pixel value represents the axon's firing value at the corresponding location in the 2D environment. Black and white pixels represent firing intensities of one and zero, respectively.

These concepts are illustrated in Figure 3.3, where an animal is seen training in two

different cylindrical environments. The animal is configured with six sensory axons, where the first half of the axons is active only in the blue environment and the other half is active only in the red environment. The two environments are thus distinguishable to the animal through the sensory inputs, and this is signified by different wall colors in the figure. The animal is positioned in the center of the two boxes, so the firing intensities of the sensory axons are thus determined by the center pixel of the respective sensory maps, as indicated by the crosses in the figure.

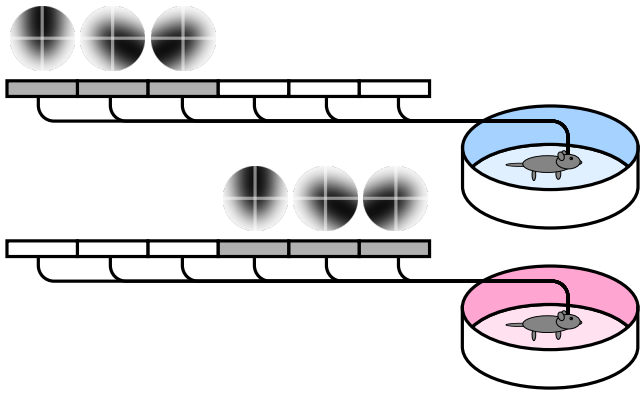


Figure 3.3: Sensory response maps can be configured per axon, per environment.

Figure 3.4 shows the different sensory maps used for the simulations in this project. The twelve maps each consist of a contiguous area of high intensity that is located radially out from the center of the environment, in various multiples of a 30° angle. The exact configurations that are used during the simulations will be listed later in the chapter.

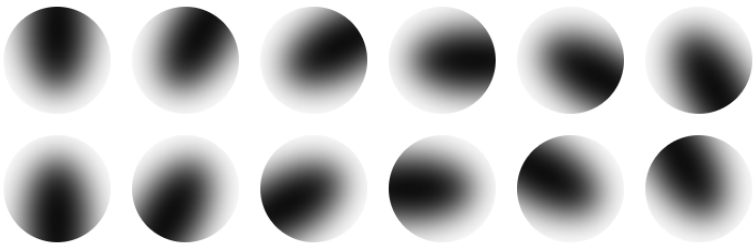


Figure 3.4: The twelve maps that are used to determine sensory inputs.

3.2.3 Grid Sustain Signal to select Mode of Operation

The model receives a signal that controls whether the grid modules in MEC should maintain their current states of firing or allow the attractor networks to seek new basins of attractions based on CA1 input. This "grid sustain signal" is represented by a variable G that will later appear in the equations specifying MEC and DG activity. G always has the value 1 during training, but during testing the variable is faded in from 0 to enable the MEC grid modules to be influenced by CA1. Specifically, the value develops as

$$G(t) = \begin{cases} 0.5 \cdot \cos((t+1) \cdot \pi) + 0.5, & t < 1. \\ 1, & t \geq 1, \end{cases} \tag{3.1}$$

where t is the current simulation time in seconds. The value of $G(t)$ is plotted for the first second in Figure 3.5.

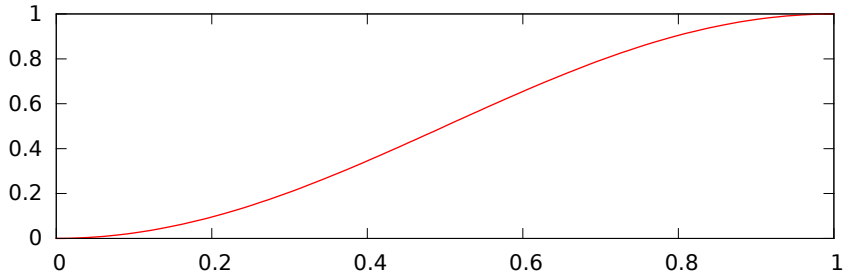


Figure 3.5: Grid sustain signal, $G(t)$, from $t = 0$ to $t = 1$.

The G input essentially selects between two modes of operation for the model. When $G = 1$, the grid modules maintain their patterns of activity undisturbed by CA1, only updating their states in response to velocity input—the model is thus in *path integration mode*. When $G = 0$, the grid modules lose their states of activity and are instead influenced by the CA1 backprojections—the model is thus in *recall mode*. These two modes are illustrated in Figure 3.6, which shows the flow of information in the model during the different modes of operation.

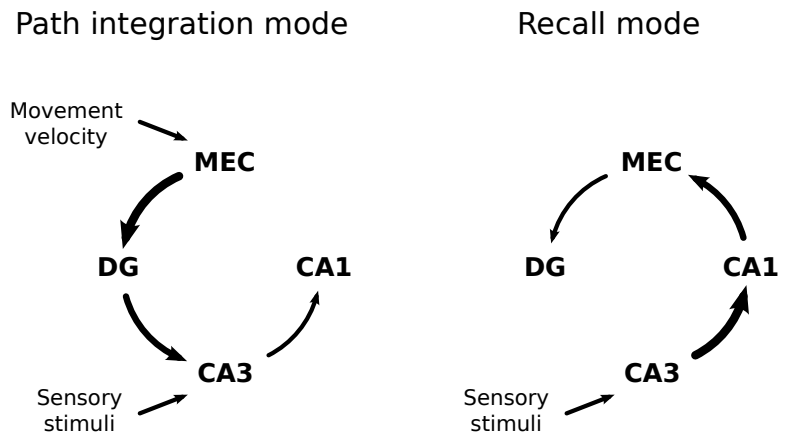


Figure 3.6: The two modes of model operation, selected by the grid sustain signal.

3.3 Continuous Attractor Networks of Grid Cells

3.3.1 Model Description

The implemented system supports a variable number of MEC grid modules. These grid modules operate independently of each other, but they receive the same velocity inputs and project their outputs to the hippocampal areas as one common bundle of axons.

The CAN model implemented for the grid cells is based on Burak & Fiete's model [3]. Neurons are laid out in a 2D sheet, with connectivity wrapping around the left/right and top/bottom edges to get a torus topology. The firing rate s_i of each neuron is determined by

$$\tau \frac{ds_i}{dt} + s_i = f\left[\sum_j W_{ij} s_j + B_i\right], \tag{3.2}$$

where W_{ij} is the synaptic strength between neurons i and j, and B_i is the external input to neuron i [3]. dt specifies the length of the time intervals used for the numerical

integration, and thus determines the accuracy of the integration, while τ is a time constant that influences the magnitude of response to velocity inputs. f is the activation function of the neurons, defined as

$$f(x) = \begin{cases} x, & x \geq 0, \\ 0, & x < 0. \end{cases} \tag{3.3}$$

The attractor network nature of the model comes about because of the weights W_{ij} between the neurons in the network. These weights are defined as

$$W_{ij} = W_0(|\mathbf{x}_i - \mathbf{x}_j - \mathbf{e}_{\theta_i}|), \tag{3.4}$$

that is, W_{ij} is a function of the Euclidean distance $\mathbf{x}_i - \mathbf{x}_j - \mathbf{e}_{\theta_i}$. The expression $\mathbf{x}_i - \mathbf{x}_j$ represents the distance between neurons i and j in the 2D neural sheet, taking into consideration that the shortest distance between neurons i and j might be across wrap-around edges in the torus topology. $\mathbf{e}_{\theta_i}$ introduces an offset of the otherwise symmetric weight matrix, that depends on neuron i's directional preference θ_i. This offset enables the activity pattern to be shifted across the neural sheet in response to velocity input, as described in Section 2.5.7.

The connectivity profile W_0 is defined as

$$W_0(x) = e^{-\gamma x^2} - e^{-\beta x^2}, \tag{3.5}$$

where γ and β are related to the parameter λ as $\beta = 3/\lambda^2$, $\gamma = 1.05 \times \beta$ [3]. λ approximately represents the *periodicity* of the activity packets in the network pattern, that is, the distance between two adjacent packets. A plot of the connectivity profile with $\lambda = 15$ can be seen in Figure 3.7.

During path integration mode, in the absence of velocity inputs, the external input B_i to neuron i has the value of 1 for all neurons. The value changes during animal movement, depending on the movement velocity and each neuron's directional preference θ_i. Movement directly along the direction of θ_i will give an extra excitation above the baseline value of 1, while movement in the opposite direction will inhibit the external inputs to neuron i so that B_i goes below 1. During recall mode, B_i contains the backprojected signals from CA1. B_i can thus be written

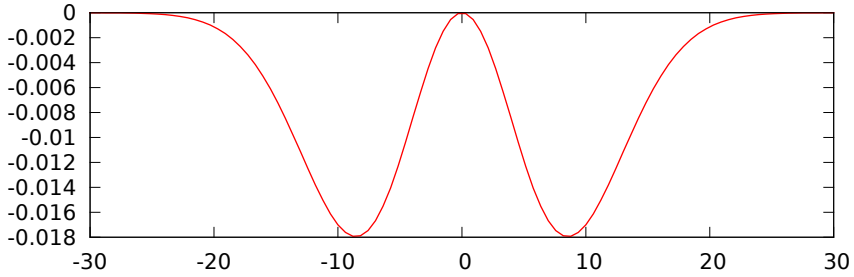


Figure 3.7: $W_0(x)$ when $\lambda = 15$.

$$B_i = I_i + G(t) \cdot (1 + \alpha \mathbf{e}_{\theta_i} \cdot \mathbf{v}) \,, \tag{3.6}$$

where I_i is the backprojection to neuron i and the velocity dependent term is the dot product between the input velocity vector $\mathbf{v}$ and the unit vector $\mathbf{e}_{\theta_i}$ pointing in neuron i's direction of preference θ_i [3].

The distribution of the directional preferences θ_i in the neural sheet repeats in 2×2 sized blocks. Within each such block, the four neurons are assigned their unique directional preference from the list north, south, west and east. The formula used to assign directional preferences is

$$2 \times (Y \textbf{ mod } 2) + (X \textbf{ mod } 2), \tag{3.7}$$

where X and Y are the sheet coordinates of the neuron. The resulting integer is used as an index into the list $[W, E, S, N]$.

The backprojections I_i are only active when the grid sustain signal $G(t)$ is low, i.e. during recall mode. Within each 2×2 block, the four neurons receive the same back-projected signal I_i. Given a matrix P_{xy} of backprojections for each 2×2 block (x, y), the backprojection signal is thus

$$I_i = P_{X/2,X/2} \cdot (1 - G(t)) \,, \tag{3.8}$$

where again X and Y are the sheet coordinates of the neuron.

3.3.2 Calculation of Outputs

When the grid modules receive velocity inputs, the activity pattern will be distorted because of how the different neurons within each 2×2 block have dissimilar directional preferences. It is desirable to avoid these disturbances in the output signals from the grid modules to the hippocampal areas of the model.

Because the four directions N, S, W and E are equally represented within each 2×2 block, the velocity dependent terms cancel each other out. Averaged over the four neurons and assuming the model is in path integration mode ($G(t) = 1$), the external input to neurons in such a block is thus always 1:

$$\sum_{\substack{\text{Neurons in}\\ 2\times2 \text{ block}}} \frac{B_i}{4} = \sum_{\substack{\text{Neurons in}\\ 2\times2 \text{ block}}} \frac{(1 + \alpha \mathbf{e}_{\theta_i} \cdot \mathbf{v})}{4} = \frac{4 + \alpha\left(\mathbf{e}_N + \mathbf{e}_S + \mathbf{e}_W + \mathbf{e}_E\right) \cdot \mathbf{v}}{4} = 1. \tag{3.9}$$

This can be used to approximate an undistorted output value for each 2×2 block, by simply using the average firing rate over the four neurons in each block. The matrix of output values O_{xy} from a grid module can thus be described as

$$O_{xy} = \frac{1}{4}\left(s_{2x,2y} + s_{2x+1,2y} + s_{2x,2y+1} + s_{2x+1,2y+1}\right). \tag{3.10}$$

As an example, for a grid network of 40×40 neurons, this will give an output matrix of 20×20 values.

3.4 Competitive Learning Network of Place Cells

The competitive learning network, representing a model of the dentate gyrus, is implemented with a Hebbian learning rule and with soft competition among the neurons, using formulas described by E.T. Rolls and A. Treves [22].

All of the neurons in the competitive learning network receive the same vector of inputs $\mathbf{r}'$. Each neuron i has its own weight vector $\mathbf{w}_i$ for the synapses made with the input signals. The activation h_i of a neuron i is the sum of the input signals weighted by

the neuron's weight configuration, mapped through the activation function described in Equation 3.3:

$$h_i = f\left[\sum_j r'_j w_{ij}\right] \quad [22]. \tag{3.11}$$

The resulting vector of activation values, $\mathbf{h}$, then needs to go through a process of competition to produce the final output firing values of the competitive network. The implemented model uses *soft competition*, wherein the most strongly activated neurons will receive strong levels of output, while the less activated neurons will be inhibited from outputting as strongly as they would do without any competition. The output firing values r_i produced by the soft competition process are calculated according to

$$r_i = \frac{e^{h_i/T}}{\sum_j e^{h_j/T}} \cdot G(t), \tag{3.12}$$

where T specifies the *temperature* of the competition. The strength of the competition increases as the temperature decreases toward zero. Thus, at very low values of T only one neuron may be active at a time, while larger values of T allow for multiple neurons to be coactive.

Note that the outputs are scaled by the grid sustain signal $G(t)$. This is done in order to silence the DG activity when the grid modules, which provide the inputs to DG, are in recall mode.

Once the output vector $\mathbf{r}$ has been determined, the synaptic weights w_{ij} are updated for the associative learning to take place. The weights are modified according to the Hebbian learning rule

$$\delta w_{ij} = k r_i r'_j, \tag{3.13}$$

where k is the learning rate. Finally, the weight vectors onto each neuron i are normalized to be of unit length, to prevent the synaptic weights from growing unbounded:

$$\sum_j (w_{ij})^2 = 1. \tag{3.14}$$

3.5 Associative Neural Networks in CA3 and CA1 Areas

3.5.1 Overview

Both the CA3 and the CA1 areas are modeled as pattern association networks, introduced in Section 2.3.2. Given the vector of input values $\mathbf{r}'$ to the CA network in question and the weight vector $\mathbf{w_i}$ of each neuron i, the neuron's firing value is calculated as for DG neurons in Equation 3.11 and mapped through the same activation function $f(x)$ described in Equation 3.3. In contrast to DG neurons, the CA neurons' firing values are not scaled afterwards.

3.5.2 CA3 Connectivity

Neurons in CA3 receive inputs from DG neurons and from the sensory axons. The input from the dentate gyrus comes in two copies, excitatory and inhibitory. This mechanism will be further explained below. All of the CA3 inputs are scaled by a constant multiplier that depends on the input source. Additionally, some of the inputs are *diluted*, that is, each of the CA3 neurons does not make contact with all of the afferent neurons, but only a given fraction of them. A specification of the scaling multipliers and the dilutions of the CA3 input sources is shown in Table 3.1.

Table 3.1: Inputs to CA3, with multipliers and dilutions.

SOURCE	TYPE	MULTIPLIER	DILUTION
DG neurons	Excitatory	2.1	$1/(\text{DG count})$
DG neurons	Inhibitory	-2.0	1
Sensory axons	Excitatory	0.5	0.5

As seen in the table, each sensory axon only has a $50\ \%$ chance that a given CA3 neuron has a synapse with it. The inhibitory signals from DG have full connectivity with CA3,

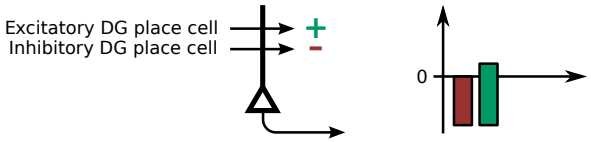
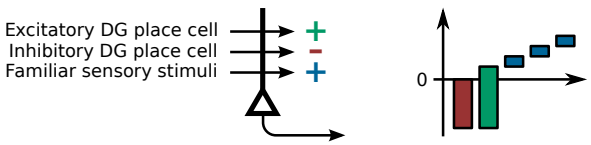
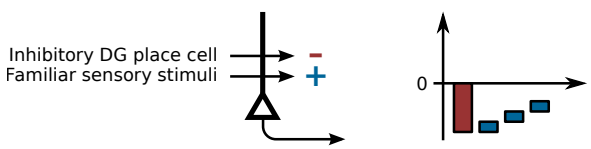
while the excitatory DG signals have a dilution such that each CA3 neuron only makes contact with one DG neuron.

The "excited-by-one, inhibited-by-all" mechanism for DG-CA3 connectivity is used to ensure that each CA3 neuron exhibits a place field that

1. initially fires weakly, when only excited by the DG cell,
2. can later be strengthened multiple times, by training the sensory synapses, and
3. will not fire when the animal is in another location, even if the incoming sensory signals closely match the ones learned by the neuron.

This scheme allows CA3 neurons to code for sensory familiarity in their output firing values without compromising the spatial integrity of their place fields. A demonstration of how the proposed DG-CA3 mechanism solves the three numbered requirements listed above, is shown in Table 3.2. The leftmost figure in each row indicates which of the three input sources are active for the example CA3 neuron, while the rightmost figure shows what the neuron's output firing value will be based on the sum of the contributions of each input source.

Table 3.2: Details of the DG-CA3 mechanism.

SITUATION	CA3 RESPONSE
1. Weak CA3 field without familiar sensory inputs.	
2. Strong CA3 field with familiar sensory inputs.	
3. CA3 field with familiar sensory inputs but in the wrong location.	

3.5.3 CA3 Learning

The CA3 input synapses from DG are not trained, but remain hard-wired the way they are initially set up by the random dilution process. As each CA3 cell is excitatorily connected to exactly one DG cell, it is up to the DG cell to determine the place field for the CA3 neuron.

The synapses between the sensory axons and the CA3 neurons are trained using the *Oja rule*, a Hebbian learning rule for pattern association networks. Under the Oja rule, the synaptic weights are updated according to

$$\delta w_{ij} = k r_i (r'_j - w_{ij} r_i), \tag{3.15}$$

where again k is the learning rate. The difference between the Oja rule and the plain Hebbian learning rule in Equation 3.13 lies in the last factor of the equation. Instead of multiplying the output value r_i by the synapse's input value r'_j, the input value is first subtracted a "forgetting term" $w_{ij} r_i$ that grows with the strength of the synapse. This ensures that the synapse will not grow unbounded through continued reinforcement [6].

3.5.4 CA1 Connectivity and Learning

CA1 neurons receive inputs from CA3 and from MEC neurons, specifically from the grid output axons O_{xy} described in Equation 3.10. A specification of all CA1 input sources, with scaling multipliers and dilutions, is shown in Table 3.3.

Table 3.3: Inputs to CA1, with multipliers and dilutions.

SOURCE	TYPE	MULTIPLIER	DILUTION
MEC output axons	Excitatory	1.0	1/(MEC output axon count)
CA3 neurons	Excitatory	0.2	1

For the CA3 synapses, the CA1 neurons use the same learning rule as CA3, the Oja rule. Similarly to how each CA3 neuron is strongly driven by exactly one DG neuron, each CA1 neuron is strongly driven by exactly one MEC output axon. This excitatory

connection from MEC to CA1 is only active during training—during testing phases, CA1 only receives inputs from CA3.

Each CA1 neuron has a reciprocal relationship with the MEC neurons it receives inputs from, as the output value from the CA1 neuron is sent back to these same MEC neurons as a form of hippocampal backprojections. Given that CA1 neuron i receives its MEC inputs from grid output axon O_{xy}, the CA1-input to the 2×2 block of grid cells (x, y) is thus

$$P_{xy} = r_i, \tag{3.16}$$

i.e. the output value of CA1 neuron i. As described in Section 3.3, the grid cells are only affected by these backprojections P_{xy} when the grid sustain signal is weak and the grid modules thus are in recall mode.

3.6 Model Parameters

3.6.1 Network Parameters

The model parameters used for the MEC, DG, CA3 and CA1 networks are listed in Tables 3.4, 3.5, 3.6 and 3.7, respectively. "Settle time" refers to the amount of time allowed for the grid pattern to settle into a stable state before the rest of the simulation proceeds. Δt specifies the simulation time intervals between weight updates.

Table 3.4: Parameters used for MEC.

PARAMETER	VALUE
Sheet size	40×40
τ	10 ms
dt	1 ms
λ	15
α	0.10315
Initial s_i	$random(0, 10^{-4})$
Settle time	2 s

Table 3.5: Parameters used for DG.

PARAMETER	VALUE
k	0.01
T	0.2
Δt	1 s
Initial $\mathbf{w}_i$	Random, normalized
Count	32 (global remapping) 16 (rate remapping)

Table 3.6: Parameters used for CA3.

PARAMETER	VALUE
k	0.01
Δt	1 s
Initial w_{ij}	$random(0, 10^{-3})$
Count	256

Table 3.7: Parameters used for CA1.

PARAMETER	VALUE
k	0.01
Δt	1 s
Initial w_{ij}	$random(0, 10^{-3})$
Count	MEC output axon count

3.6.2 Grid Sizes used during Simulation

As described in Section 2.1, multiple grid modules of different grid sizes should be used as input to the hippocampus to enable the formation of single-peaked place fields. Four grid modules were chosen to be used in MEC for these simulations.

Grid modules found experimentally in animals seem to have a discrete range of grid sizes [23]. Furthermore, the ratio between successive grid sizes appears to be approximately the square root of two ($\sqrt{2} \approx 1.41$). In line with these findings, the selection of grid sizes was chosen to be 0.7, 0.5, 0.35 and 0.25 (approximately $\sqrt{2}^{-1}$, $\sqrt{2}^{-2}$, $\sqrt{2}^{-3}$ and $\sqrt{2}^{-4}$, respectively).

3.7 Training and Testing Procedures

3.7.1 Environments

The model is first trained in a series of *training environments*, denoted E_A, E_B, etc., and afterwards tested for recall ability in a series of *testing environments*, denoted T_1, T_2, etc. Figure 3.8 illustrates a sequence of such training and testing environments. The two training environments E_A and E_B in the figure are similar to the two cylindrical environments seen in Figure 3.3 in Section 3.2.2—the blue and red wall colors indicate that different sets of sensory axons are active in the two environments.

The training environments E_A and E_B in Figure 3.8 are also marked with Greek letters α and β on the floor underneath them. These letters indicate that the two boxes are placed at two different locations in the global environment—two different *training sites*—and are thus represented by different values in the global coordinate system that

the MEC comprises. In other words, the same relative location inside the two boxes will be represented by different sets of global coordinates, or network states, in the MEC grid modules of the animal. The place cells activated in DG will thus also be different.

The thought bubbles above the animal indicate whether the animal is aware of which environment it is currently navigating in—during training, the MEC states in the model are explicitly set to either α or β, but during testing rounds the animal starts out with empty states in MEC and thus needs to determine its location based on the sensory stimuli.

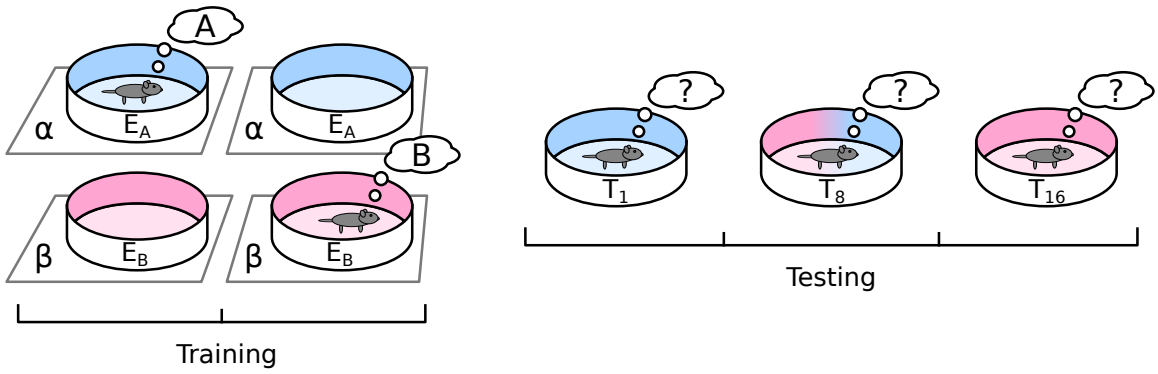


Figure 3.8: Global remapping experiment. The animal is trained in two training environments E_A and E_B at two different training sites α and β, respectively. Later the animal is tested in a range of testing environments T_{1-16}, where the sensory stimuli either unambiguously indicate one of the training environments, or ambiguously indicate a mix of the training environments.

Because the training environments in Figure 3.8 are situated in two different training sites, the figure is an example of how the global remapping simulations are performed. Figure 3.9 shows a similar sequence of training and testing environments for rate remapping simulations. Note the difference between the two figures—in the rate remapping example, the two training environments E_A and E_B are situated at the same training site α. The sensory configurations for E_A, E_B and T_{1-16} are however the same as for the global remapping experiment.

The specific sensory configurations used for the simulations in this project are shown in Table 3.8. There are 24 sensory input axons to the model—two axons of each of the twelve sensory response maps shown in Figure 3.4. This is shown in the table by the 24 images in the header row, whose respective columns represent the 24 axons.

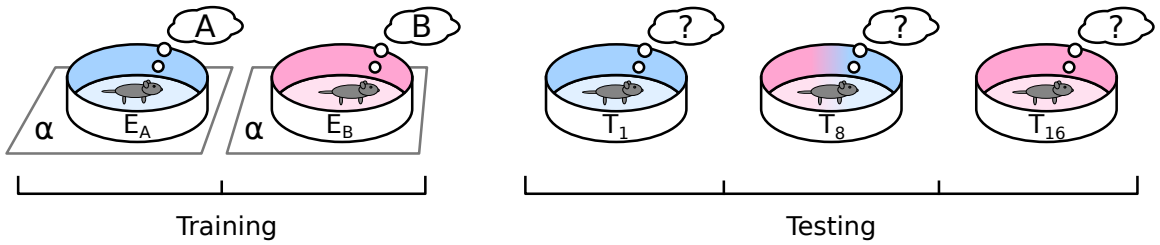


Figure 3.9: Rate remapping experiment. The animal is trained in two training environments E_A and E_B, both situated at the same training site α. Later the animal is tested in a range of testing environments T_{1-16}, where the sensory stimuli either unambiguously indicate one of the training environments, or ambiguously indicate a mix of the training environments.

The rest of the rows specify the sensory configurations for E_A, E_B and T_{1-16}. A filled circle (•) indicates that the given sensory axon is active in the given environment, while a blank circle (○) indicates otherwise. The training environments E_A and E_B each have a full set of the twelve sensory maps from Figure 3.4 active. The first and the last testing configurations, T_1 and T_{16}, correspond to E_A and E_B respectively. The configurations in between comprise a gradual transition from T_1 to T_{16}, where given sensory maps are turned off in the half of the sensory axons that correspond to E_A and correspondingly turned on in E_B's half of the sensory axons.

Table 3.8: Sensory configurations used during remapping experiments.

Training	E_A	●	●	●	●	●	●	●	●	●	●	●	●	○	○	○	○	○	○	○	○	○	○	○	○
	E_B	○	○	○	○	○	○	○	○	○	○	○	○	●	●	●	●	●	●	●	●	●	●	●	●
Testing	T_1	●	●	●	●	●	●	●	●	●	●	●	●	○	○	○	○	○	○	○	○	○	○	○	○
	T_{2-5}	●	●	●	○	●	●	●	○	●	●	●	○	○	○	○	●	○	○	○	●	○	○	○	●
	T_{6-8}	●	●	○	●	●	○	●	●	○	●	●	○	○	○	●	○	○	●	○	○	●	○	○	●
	T_{9-11}	○	○	●	○	○	●	○	○	●	○	○	●	●	●	○	●	●	○	●	●	○	●	●	○
	T_{12-15}	○	○	○	●	○	○	○	●	○	○	○	●	●	●	●	○	●	●	●	○	●	●	●	○
	T_{16}	○	○	○	○	○	○	○	○	○	○	○	○	●	●	●	●	●	●	●	●	●	●	●	●

The configurations from T_2 to T_{15} are grouped together as T_{2-5}, T_{6-8}, T_{9-11} and T_{12-15} to save space in the table. These *trial groups* represent different shifted configurations of the same amount of active/inactive axons in each half of the axon columns. The trial group T_{2-5} consists of different configurations characterized by having $^3/_4$ of E_A's axons turned on and $^1/_4$ of E_B's axons turned on. Likewise, T_{6-8}, T_{9-11} and T_{12-15} have

$^2/_3$, $^1/_3$ and $^1/_4$ of E_A's axons turned on, respectively. This shows a gradual transition in similarity between E_A and E_B as the tests progress from T_1 to T_{16}.

3.7.2 Training Phase

The training phase of the program proceeds through the following steps in order:

1. Pattern formation in the grid networks.
2. Training of DG place cells.
3. Training of CA3 place-event associations and CA1 backprojections.

The first step is to enable formation of the hexagonal grid patterns in the MEC grid networks. This is done by initializing the grid cells to random firing values and waiting a sufficiently long amount of time for the patterns to form, with parameters as specified previously in Table 3.4. As described above, there might be multiple sites for the training environments to be situated—for example, in the global remapping case, there are the two training sites α and β. The program thus repeats the process of generating novel grid patterns for each of these globally distinct training sites, storing the finished grid patterns in internal matrices for reuse later during training.

The next step is to train the DG neurons to form place fields. This is done separately before training the rest of the hippocampal model, to have stable place fields available when training CA3 and CA1. The ten minute path described in Section 3.2.1 is simulated repeatedly, and the weight update rule for the competitive learning network is invoked at fixed time intervals $\Delta t = 1$s. At the beginning of each training round, the MEC grid modules are initialized with the stored grid patterns from the pattern formation phase that correspond to the training site (α or β) of the current training environment. The ten minute path is repeated ten times in total for each training environment, alternating between the two environments for each training round.

Lastly, the CA3 and CA1 networks are trained. This follows the same procedure as for DG training, by alternating between the different training environments for a total of ten training rounds each, and in each training round invoking the learning rule at fixed time intervals Δt along the ten minute path.

3.7.3 Testing Phase

Contrary to the training environments, the testing environments do not have any preset information about training sites attached—the goal of the model is precisely to recall this information based on the sensory inputs alone. At the beginning of each testing round, the network states of the MEC grid modules are thus reset, i.e. all neuron values are set to zero. The grid sustain signal G is also set to zero, as described by Equation 3.1.

The sensory axons then send signals to CA3 that originate from the center point of the testing environment. CA3 neurons will respond to these sensory signals undisturbed by DG, as the DG outputs are modulated by G. CA3 passes its signals on to CA1, which in turn activates MEC neurons through backprojections. As the simulation proceeds, G is then increased toward 1, gradually forcing the MEC grid modules into a regular pattern of firing and restoring normal firing behavior in DG.

When G has reached 1, i.e. after one second, the MEC grid networks have settled into their final states, and it is time to analyze the response of the model. This is done by simulating how the model would respond at all pixels in the environment, given the final states of the MEC grid networks. Further details about these analyses are given in the next chapter.

3.8 Biological Plausibility

The implemented model is considered to be biologically plausible. The overall architecture of the model is comparable to that of the real hippocampus, as evidenced by comparing Figure 3.1 to Figure 1.2. Each part of the model is represented by biologically plausible neural networks. The MEC grid modules are represented by attractor networks, requiring recurrent inhibitory connectivity—this is supported by neuroscientific data [4]. The DG place cells are implemented as a competitive learning network, which is supported as a model of the real dentate gyrus by the existence of inhibitory interneurons in that part of the hippocampus [21].

CA3 and CA1 are implemented as pattern association networks. These require LTP in the synapses of the neurons, which is known to be the case in these hippocampal areas [18]. Note that CA3 has not been implemented as an autoassociation network, as suggested by Rolls & Treves among others [22]. An autoassociation network was not

needed to perform the task of place–event association required of CA3 in this project, and the simpler network architecture of a pattern associator was chosen instead. If needed, it should be possible to extend the model to use an autoassociator for CA3.

The model assumes accurate velocity inputs and sensory stimuli that provide dependable spatial information. Because the hippocampus mainly receives highly processed information from the neocortex, it is not unreasonable to assume a certain level of sophistication from these signals.

The grid sustain signal is used to choose between path integration mode and recall mode in the hippocampal model. While there is probably no directly equivalent signal in the real hippocampus, the hippocampus does appear to oscillate between externally and internally driven modes in step with the oscillations of the theta brain waves. These externally and internally driven modes might be considered similar to the path integration and recall modes, respectively.

3.9 Implementation Details and Optimizations

The implementation of the model for this project was done in the C programming language. The implementation uses the *C Standard Library* for most of its input/output handling and for mathematical functions such as trigonometry and pseudo-random number generation. It uses the *libpng* library for reading image files containing simulation data, such as the sensory response maps. *Gnuplot* is used for plotting results from the model, while *Python* is used for some of the offline data analysis.

The *GCC* compiler was used for the parts of the project written in C, and notably a feature of the GCC compiler called *compiler intrinsics* was used to enable AVX vectorization for some of the floating point calculations involved in the CAN grid cell model. AVX is a class of Single-Instruction Multiple-Data (SIMD) instructions for the x86 CPU architecture, that enable up to eight floating point operations to be performed in parallel in each CPU core. Additionally, *OpenMP* was used to parallelize performance-critical loops across multiple CPU cores.

A key performance optimization that was employed in the implementation of the model is the emulation of grid module outputs during training of DG, CA3 and CA1. This optimization was also used when measuring the model response at arbitrary coordinates in the environment, used for some of the plots and analyses in the next chap-

ter. During path integration mode, the behavior of the MEC grid modules—according to the CAN model—is highly predictable. The network patterns in the grid modules will shift around according to the velocity inputs that are received by the grid networks. By calculating these offset network patterns directly based on the known coordinates of the animal, instead of performing the significant amount of floating point operations needed to produce the same result, the execution time of the model can be improved considerably. It is not believed that these emulated network outputs have any major impact on the results of the model.

3.10 Summary

A summary of the hippocampal neural network model described in this chapter is shown in a schematic form in Figure 3.10, where all of the parts of the model are shown by their principal neurons, as well as the connectivity between these neurons. By comparing this schematic with the hippocampal model described by Rolls & Treves, shown in Figure 2.7, there can be seen to be a resemblance between the two models at the architectural level.

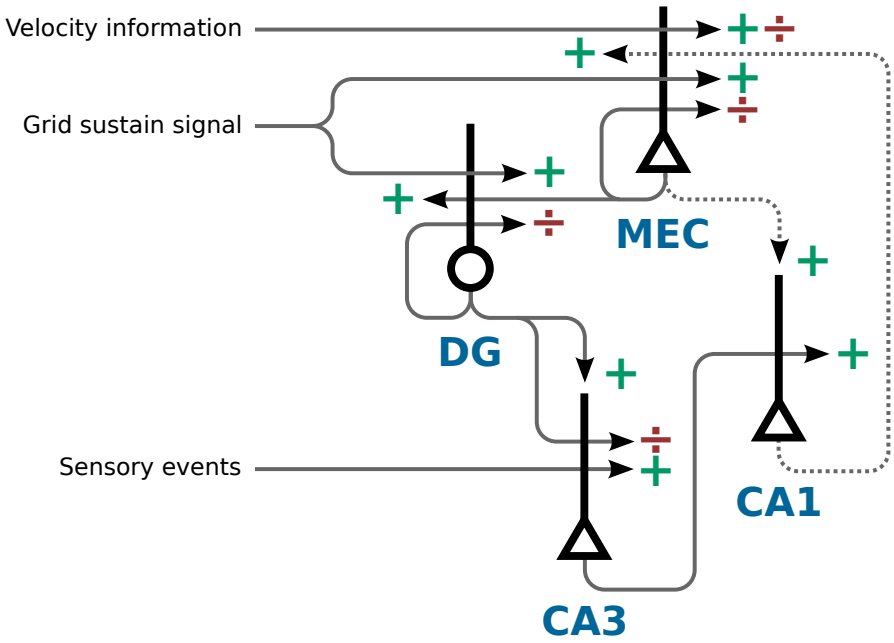


Figure 3.10: The schematic of the full neural network.

As described in Section 3.2.3, the model has two modes of operation during testing—path integration mode and recall mode. The model also behaves in a third way during training, where the CA1–MEC connectivity goes from MEC to CA1 instead of the other way around. The flow of information in the model during these three modes of network operation is summarized in Figure 3.11.

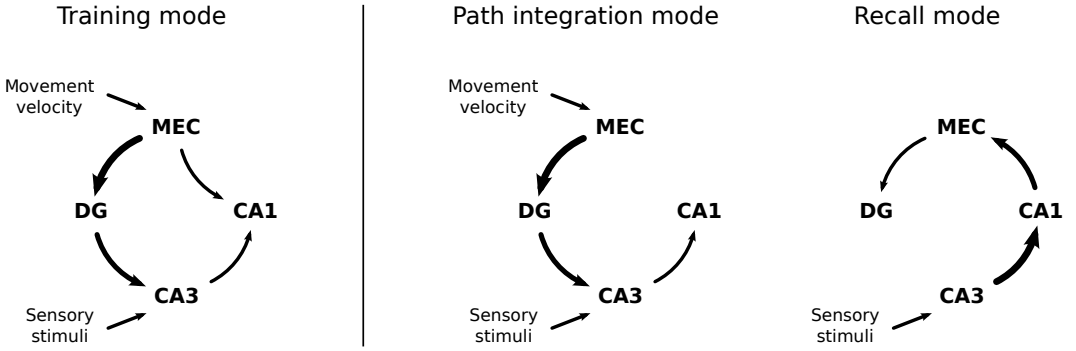


Figure 3.11: Information flow in the model during training and for the two modes of operation during testing.

4 Results and Analysis

This chapter shows the results of the implemented model and discusses the implications of these results. Section 4.1 shows how the different parts of the hippocampal model work, first each part in isolation and then all of the parts together. The section starts with a demonstration of the grid cell modules in the entorhinal cortex, and then goes on to show how the place cells in the dentate gyrus work, how the neurons in CA3 form place–event associations and how the signals from CA1 cause backprojections to MEC, inducing previous states of grid cell activity. The last part of the section shows how all of these parts of the model work together to enable recall of previous states in MEC based on sensory stimuli.

Section 4.2 introduces the remapping experiments and their methods of analysis. Section 4.3 shows the results of the global remapping experiment, while Section 4.4 shows the results of the rate remapping experiment. Finally, Section 4.5 discusses the results seen in this chapter and relates them to the research questions of this project.

4.1 Operation of the Hippocampal Model

4.1.1 Grid Cells in the Entorhinal Cortex

Figure 4.1 shows the formation over time of a grid pattern in a MEC network of 64×64 grid cells. The pattern formation happens because of the symmetry breaking process brought about by random initial values, described in Section 2.5.4, and because the network connectivity encourages the firing activity to collect in hexagonally distributed packets of activity.

The individual grid output axons are shown as pixels, where the color represents the firing rate. Because the output signals from the MEC modules are averages of blocks of

2×2 neurons, as described in Section 3.3.2, the figure shows a plot of 32×32 values. The neurons' coordinates in the figure represents their location in the attractor network's two-dimensional neural sheet, which determines their connectivity to other neurons and the offset of their spatial responsivity in the environment.

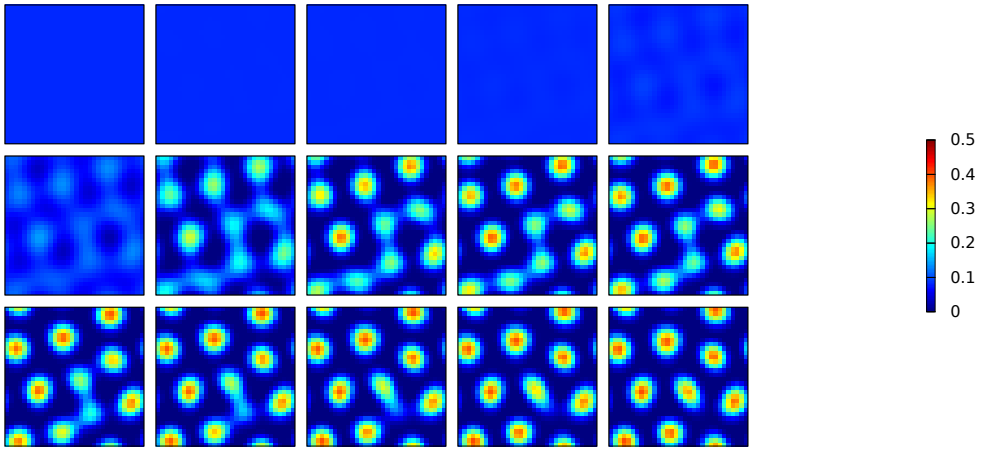


Figure 4.1: Formation of grid pattern in a network of 64×64 neurons, shown as the values of the 32×32 output axons. Top left plot shows simulation state at $t = 0.05$ s, increasing in steps of $\Delta t = 0.05$ s going from left to right, top to bottom.

As seen in the figure, distinct packets of activity have formed in the network by the time of the last plot in the time sequence, $t = 0.75$ s. However, the packets have not yet been distributed out in a strict hexagonal pattern. To verify that this indeed does happen eventually, Figure 4.2 shows the network activity at $t = 2.00$ s. In this figure, the packets have self-organized to form a precise hexagonal pattern.

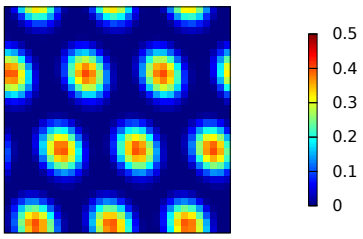


Figure 4.2: Network activity from the same simulation as shown in Figure 4.1, at $t = 2.00$ s.

As described in Section 3.6.1, the sheet size used for the rest of the simulations in this

project is 40×40 neurons in each grid module. For an impression of how a MEC network of this size looks, a plot sequence similar to Figure 4.1, for a 40×40 grid module, is shown in Figure 4.3.

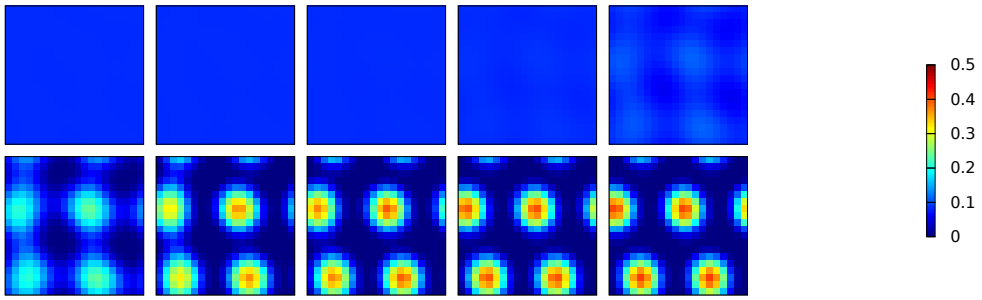


Figure 4.3: Formation of grid pattern in a network of 40×40 neurons, shown as the values of the 20×20 output axons. Top left plot shows simulation state at $t = 0.05$ s, increasing in steps of $\Delta t = 0.05$ s going from left to right, top to bottom.

Although all of the activity peaks that comprise a full copy of the hexagonal pattern do no longer fit in the neural sheet shown in these plots, a hexagonal pattern can still be realized by the grid module due to the torus topology that enables the pattern to wrap around the horizontal and vertical edges. Figure 4.4 shows the last plot in the time sequence of Figure 4.3 tiled 3×3 times, to illustrate that the there is a complete hexagonal pattern in the torus networks of the grid modules.

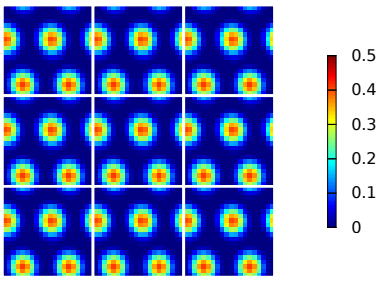


Figure 4.4: Network activity from the simulation shown in Figure 4.3 at $t = 0.50$ s, tiled 3×3 times to reveal the hexagonal pattern.

To see how individual grid cells work when the animal moves around in the environment, the output value of a single grid output axon was plotted as the color of a line tracing out the animal's current position. This is shown in Figure 4.5. The network pattern shown above is visible through the single output axon's value shown plotted

in this figure, because of how the network pattern is shifted around the neural sheet of grid cells in response to velocity input, as described in Section 2.5.6.

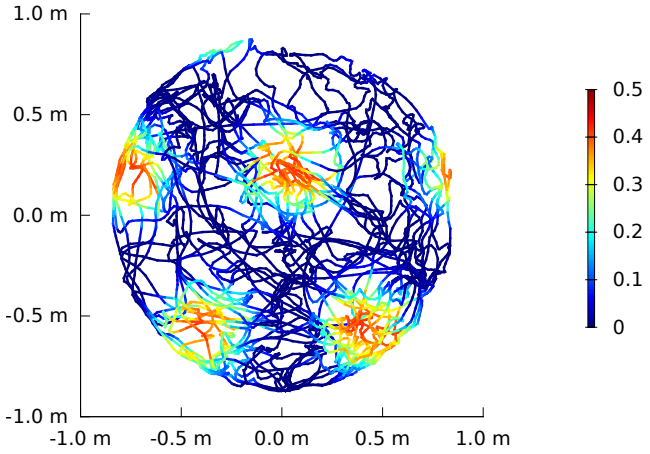


Figure 4.5: Single neuron response during a simulation of the entire ten minute path described in Section 3.2.1. The value of output axon $O_{0,0}$ is shown plotted as the color of a line tracing out the animal's current position.

4.1.2 Place Cells in the Dentate Gyrus

Based on spatial signals from the entorhinal cortex grid cells that contain information about the animal's coordinates, the competitive learning network in DG is able to form place fields.

Trained in One Environment

The model was trained as specified in Section 3.7.2, using 16 DG neurons; after formation of grid patterns in the entorhinal cortex, the animal's movements were simulated according to the path traced out in Figure 3.2, applying the Hebbian learning rule to the DG neurons every $\Delta t = 1$ s along the way. This procedure was repeated ten times, in a single training site.

Figure 4.6 shows the *spatial response plots* for each of the 16 DG neurons, that were generated after the training phase was completed. These plots show the circular outline of the simulated animal enclosure viewed from above. Each pixel within the enclosure is colored according to how strongly the DG neuron responded at the corresponding location. The plots were generated by emulating the appropriate MEC outputs at each sampling location, updating the hippocampal simulation (without performing any weight updates) and then recording the output response from each DG neuron.

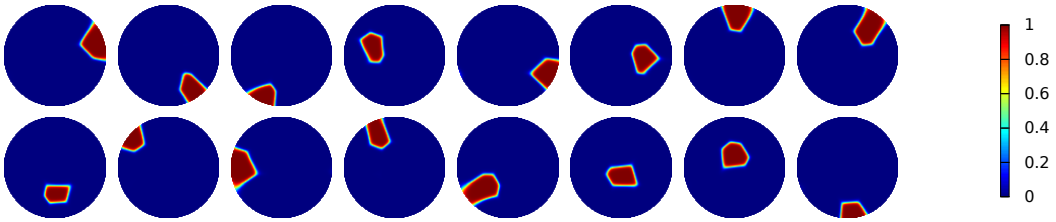


Figure 4.6: Spatial responses of the 16 DG neurons after finished training phase.

As seen in the spatial response plots, the DG cells have acquired a single, contiguous place field each. When the animal is within the boundaries of these place fields, the DG cells respond at maximum level, while staying silent outside of the place fields—this is as expected from a competitive network.

A competitive learning network also predicts that the place fields will be non-overlapping. To verify this, the 16 place fields are shown plotted together in the same coordinate system in Figure 4.7. The plot confirms that the place fields are non-overlapping, and also demonstrates that the 16 DG place cells together cover the entirety of the visited enclosure with place fields.

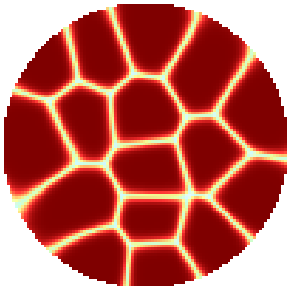


Figure 4.7: Place fields from Figure 4.6 plotted in the same coordinate system.

Trained in Two Environments

The model was then trained in two different training environments, using the training procedures specified for the global remapping experiment in order to make the model train in two different training sites α and β. As part of the model parameters for the global remapping experiment, DG was now configured to have 32 neurons. As described in Section 3.7.2, the ten minute training path was repeated ten times for each of the two training environments, alternating between the two environments for each training round. The resulting spatial response plots for eight example neurons are shown in Figure 4.8. Each neuron's response is now shown as a pair of circular plots, showing how the neuron responds in each of the two training environments.

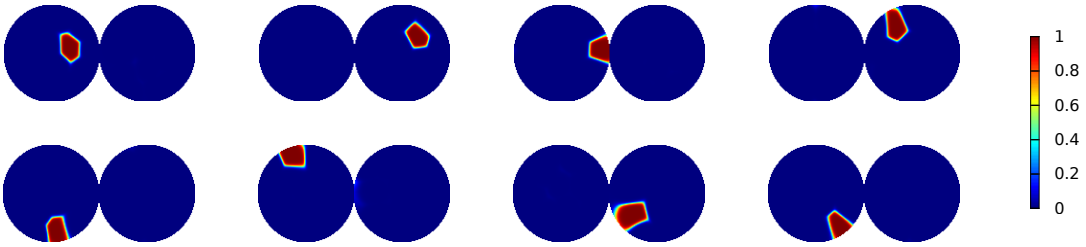


Figure 4.8: Spatial responses of eight example DG neurons after training at two different training environments. Left and right half of each pair is the spatial response in training environment E_A and E_B, respectively.

As seen in the figure, the DG neurons have each acquired a place field, and the neurons are active only in one of the training environments. It should be noted that there sometimes are occurrences of DG neurons that respond to place fields in both environments, but these are an exception.

Figure 4.9 shows the merged spatial responses of the 32 DG neurons, similar to Figure 4.7 above. The figure confirms that the DG model has succeeded in creating place fields that together cover both of the training environments.

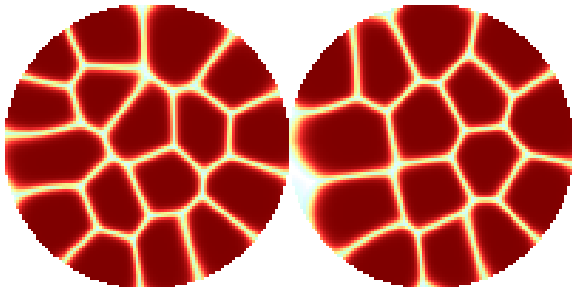


Figure 4.9: Place fields from all 32 DG neurons plotted together in the same coordinate system, with one plot for each of the two training environments.

4.1.3 Place–Event Associations in CA3

With the DG neurons upstream of CA3 successfully forming place fields, neurons in CA3 should be able to use this information to form place–event associations. As described in Section 3.5.2, CA3 neurons receive strong excitation from randomly selected DG neurons, as well as a slightly weaker inhibition from all DG neurons. This mechanism is in place to enable CA3 cells to form strong associations to sensory stimuli that occur within their place fields, without responding to these stimuli elsewhere.

Trained in One Environment

The model that was simulated in the first part of the previous section also included 256 neurons in CA3, that were trained as specified in Section 3.7.2. Spatial response plots of 24 of these CA3 neurons are shown in Figure 4.10.

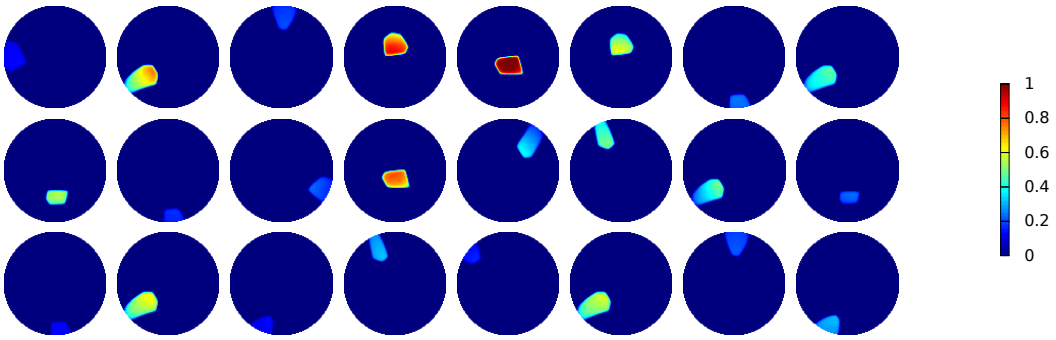


Figure 4.10: Spatial response plots of 24 example CA3 neurons.

The plots show that all of the CA3 cells have gained place fields, but the peaks in activation strength varies widely between the neurons. Notice that all of the CA3 place fields are derived from the DG place fields shown in Figure 4.6. Neurons in DG determine the spatial extents of the CA3 neurons' firing, but local learning in the sensory CA3 synapses determines *how strongly* the CA3 neurons fire. This can be seen in Figures 4.11 and 4.12, where examples are shown of weakly and strongly activated CA3 cells, respectively.

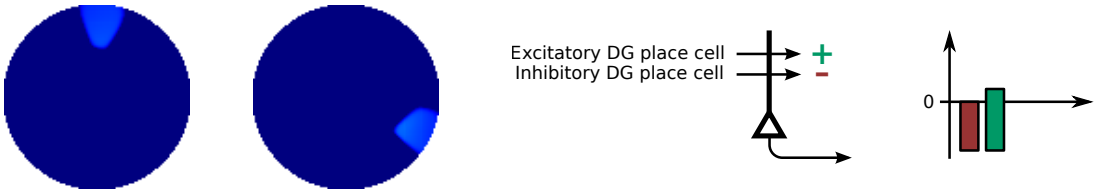


Figure 4.11: Spatial response plots of two weakly firing CA3 neurons (left), and an illustration of the input sources that influence these neurons (right).

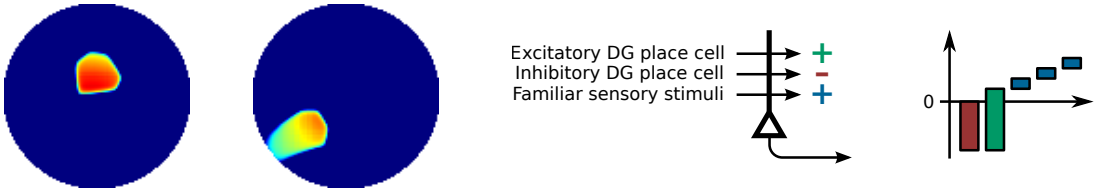


Figure 4.12: Spatial response plots of two strongly firing CA3 neurons (left), and an illustration of the input sources that influence these neurons (right).

The weakly firing neurons in Figure 4.11 receive almost all of their activation from the DG signals and possibly some from weak sensory influences. The strongly firing neurons in Figure 4.12 show that the CA3 neurons can be made to fire several times more strongly by training their sensory synapses to respond to familiar stimuli. Because of the dilution between the sensory axons and the CA3 neurons, some CA3 cells will not be able to form these strong sensory synapses, and this contributes to the variety of both weakly and strongly activated CA3 cells seen.

Whereas in DG only one neuron is active at a time, CA3 has more interesting dynamics in its firing. Depending on the activity in DG, smaller or larger subsets of cells in CA3 will be active. Additionally, how strongly each of these neurons fire depends on the incoming sensory stimuli. In order to fully compare CA3 activity from a given point in time with previously recorded activity, one thus needs to consider the *population*

vectors involved—vectors of simultaneous firing rates from all of the CA3 neurons. CA3 activity is interesting on the level of entire assemblies of cells, not just on the level of individual ones.

CA3 has two modes of operation—strongly influenced by DG when the model is in path integration mode, and influenced by sensory inputs undisturbed by DG when the model is in recall mode. These two modes are demonstrated in Figures 4.13 and 4.14 respectively. The figures include plots of the firing rates of all CA3 neurons—the CA3 population vectors—for each of the two cases.

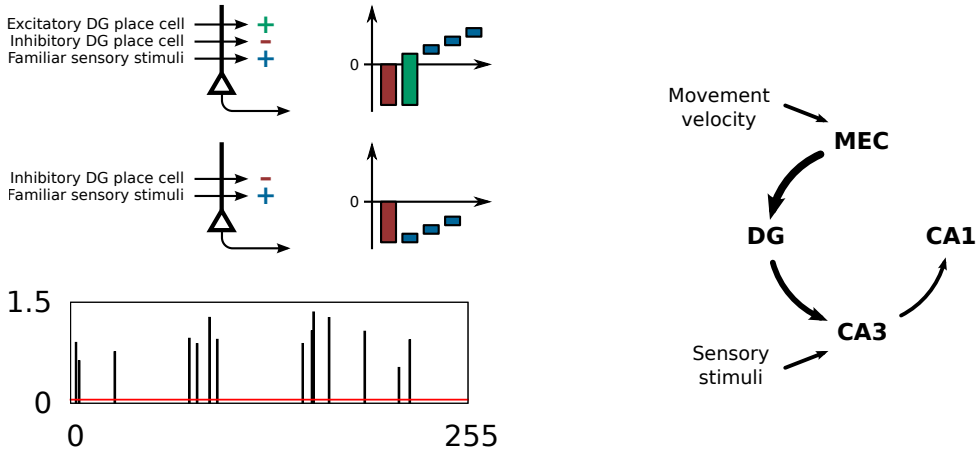


Figure 4.13: CA3 activity in path integration mode: Example population vector of firing rates for the 256 CA3 neurons (left, bottom). Explanations for strongly firing CA3 neurons (left, top) and for silent CA3 neurons (left, middle). Illustration of signal flow in the model during path integration mode (right).

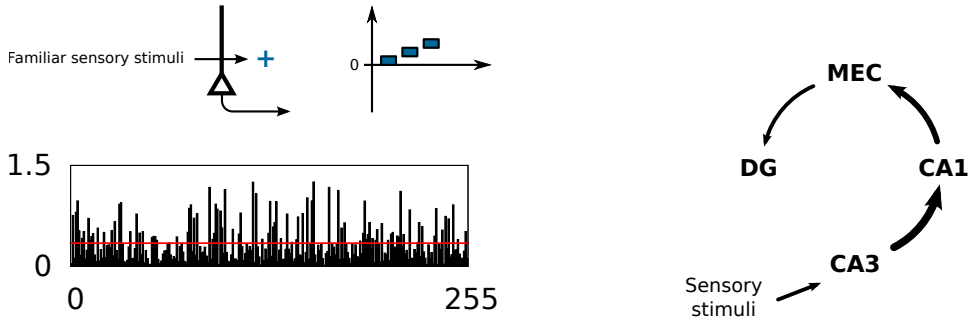


Figure 4.14: CA3 activity in recall mode: Example population vector of firing rates for the 256 CA3 neurons (left, bottom). Explanation for active CA3 neurons (left, top). Illustration of signal flow in the model during recall mode (right).

The firing in CA3 during path integration mode is considerably more sparse than during recall mode, as seen in the two figures above. In path integration mode, only the CA3 neurons whose associated DG cells are active are allowed to fire, whereas all CA3 neurons may fire in recall mode.

In order to understand what information the firing vectors in Figures 4.13 and 4.14 contain, the vectors must be compared with corresponding vectors recorded from CA3 under various conditions previously, most interestingly the population vectors at different coordinates in the training enclosure. In other words, for each set of (x, y) coordinates in the environment, there should be a population vector of the firing rates of each of the 256 CA3 neurons. By viewing the element index into these population vectors as a third coordinate, z, it becomes apparent that the information needed for this analysis can be represented as a 3D matrix of all the different z neurons' firing values at all (x, y) coordinates in the environment.

This information is already available from the spatial response plots seen e.g. in Figure 4.10, which contain given neurons' response values at all (x, y) coordinates. Assuming that these spatial response plots are available for all neurons, the plots can be "stacked" on top of each other to create the three-dimensional matrix described above. This idea is illustrated in Figure 4.15. The figure shows a selection of the plots from Figure 4.10 stacked on top of each other, with an arrow passing through them at fixed (x, y) coordinates to extract z values for the corresponding population vector.

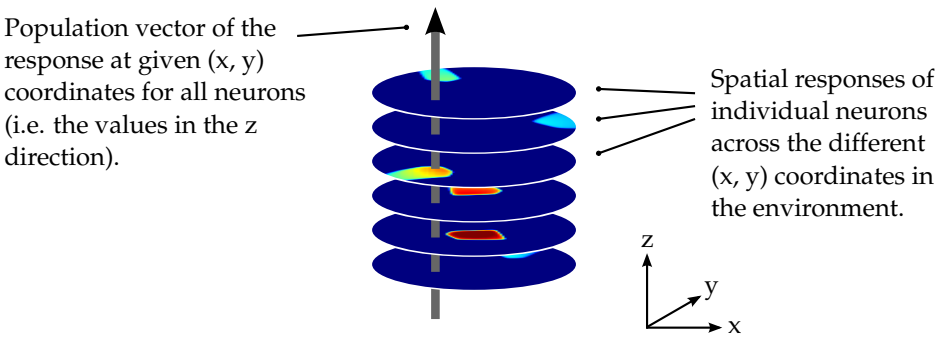


Figure 4.15: Population vectors for given (x, y) coordinates can be obtained by extracting the vector of values along the z direction in a 3D stack of the 2D spatial response plots.

With population vectors available for all (x, y) coordinates in the training enclosure, it is then possible to calculate the correlation between all these vectors and the pop-

ulation vectors shown in Figures 4.13 and 4.14. A simple measure for the correlation between two vectors is the angle between them, e.g. as calculated by

$$\alpha = \text{acos}\left(\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|\cdot|\mathbf{b}|}\right), \tag{4.1}$$

where $\mathbf{a}\cdot\mathbf{b}$ is the dot product of $\mathbf{a}$ and $\mathbf{b}$ and $|\mathbf{v}|$ is the length of $\mathbf{v}$. Figure 4.16 uses this correlation measure to show the correlation for the population vectors from Figures 4.13 and 4.14 with the recorded CA3 assembly response at different coordinates.

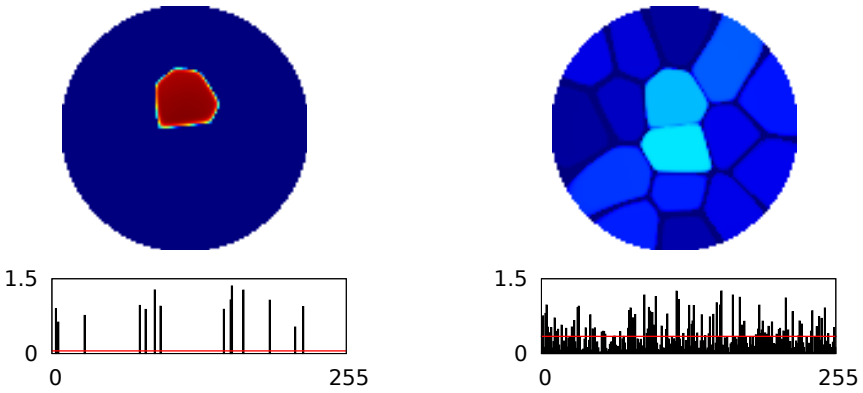


Figure 4.16: Correlation with recorded CA3 activity for the population vectors from Figure 4.13 (left) and from Figure 4.14 (right). Correlation plots are color-coded on a scale from blue for no correlation (orthogonal vectors), to red for full correlation (parallel vectors).

The leftmost correlation plot shows a strong correlation inside of the boundaries of the active DG cell, and no correlation outside. This is as expected from the model during path integration mode. The population vector in the rightmost correlation plot was recorded during recall mode, when the model received sensory stimuli corresponding to the central area of the enclosure. The plot shows varying degrees of correlation around the entire environment, but the correlation is noticeably stronger in the center. This indicates that CA3 is able to associate sensory stimuli with locations in the environment.

Trained in Two Environments

The simulation run from Section 4.1.2, where DG was trained in two different training environments, also included 256 CA3 neurons that were trained afterwards. Spatial response plots of eight example CA3 neurons from this simulation run are shown in Figure 4.17.

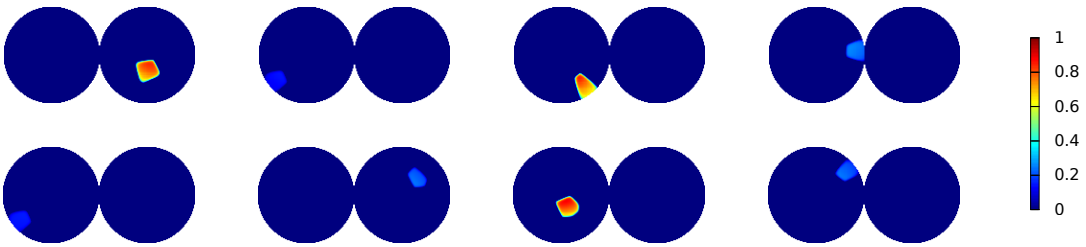


Figure 4.17: Spatial response plots of eight example CA3 neurons after training at two different training environments. Left and right half of each pair is the spatial response in training environment E_A and E_B, respectively.

As the DG neurons mostly formed place fields in only one of the training environments, it is to be expected that the CA3 neurons do the same. From the examples in Figure 4.17, this appears to be the case. However, to verify that CA3 is able to distinguish between the two environments, the model was tested in recall mode twice—once with sensory inputs corresponding to the center of training environment E_A and once with inputs corresponding to the center of training environment E_B. The resulting population vectors and correlation plots are shown in Figure 4.18.

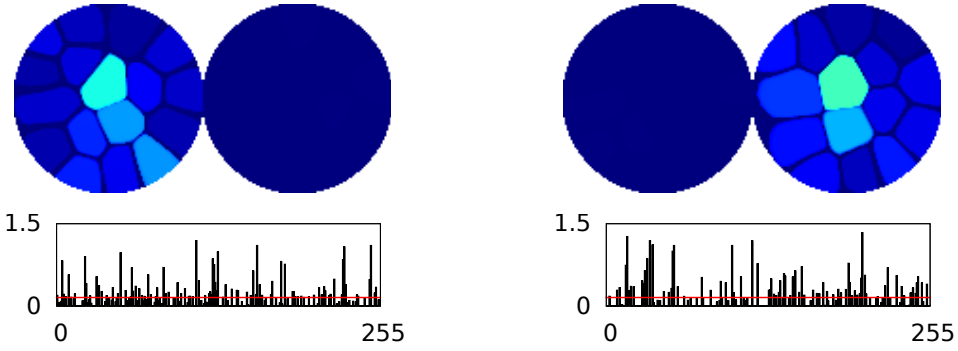


Figure 4.18: CA3 population vectors obtained by sending sensory inputs corresponding to the center of training environment E_A (left, bottom) and E_B (right, bottom). Correlations with recorded CA3 activity shown above population vectors. Left and right half of each pair of correlation plots is the spatial correlation with training environment E_A and E_B, respectively.

The correlation plots show that CA3 is able to recognize the correct training environment. When sensory stimuli from training environment E_A is received, the population vector is only correlated with recordings from training environment E_A, and correspondingly for training environment E_B.

4.1.4 Backprojections to MEC from CA1 Signals

The model was tested to see whether the backprojections from CA1 to MEC are able to reinstate previous patterns of firing in the grid modules. After training the model as usual, the neurons in DG were forced to stay silent, with the exception of one arbitrary DG place cell that was made to fire at maximum rate. Sensory inputs to CA3 were disabled, so that the activation of the DG place cell was allowed to propagate through CA3 on to CA1—and from there to the MEC grid networks—unaffected by sensory signals.

Figure 4.19 shows the pattern of the backprojected signals from CA1 to one of the MEC grid modules. A grid pattern is clearly visible, which is a positive sign that the signal originating in the DG place cell has triggered the backprojection of a meaningful signal from CA1 to MEC. The dark pixels in the pattern is explained by the dilution of the connectivity between MEC and CA1—not all grid cells are connected with CA1.

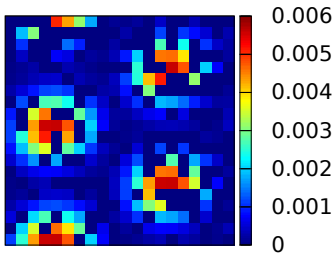


Figure 4.19: Incoming pattern of the backprojections from CA1 to the grid cells in one of the grid modules.

The response of the MEC grid networks was then tested. The grid modules were initially set to almost empty states, with only weak, random activation values. They were then allowed to fall into basins of attraction under the influence of the CA1 backprojections. The grid sustain signal was gradually turned up, as described in Section 3.2.3, while the influence of the CA1 signals was correspondingly turned down.

Figure 4.20 shows the development over time of the network pattern in the grid module that received the backprojections shown in Figure 4.19. From the almost silent initial state, the network activity gradually develops into a pattern similar to that of the backprojected signals.

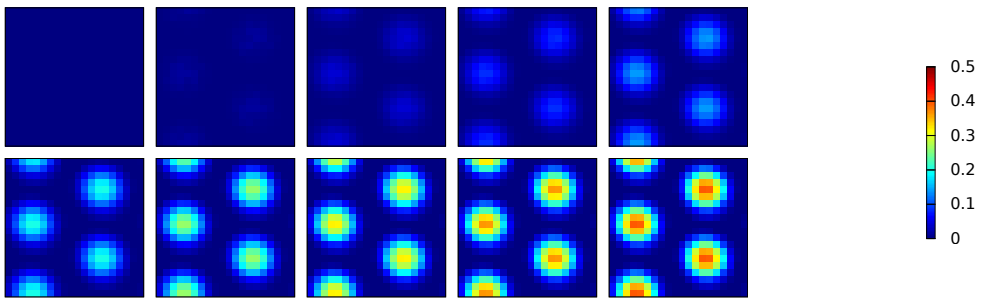


Figure 4.20: Grid pattern forming in the grid module receiving the backprojections shown in Figure 4.19. Network pattern at the start of the simulation ($t = 0$ s) shown in the top left, increasing in steps of 0.1 s going from left to right, top to bottom.

For a qualitative comparison of the backprojected pattern to the final grid pattern in the MEC attractor network, plots of the two are shown side-by-side in Figure 4.21.

As seen by the similar orientation and offset of the two patterns, there appears to have been a successful backprojection of MEC activity from the hippocampus. In other words, the hippocampal backprojections have successfully induced a given set of global coordinates in the MEC path integrator.

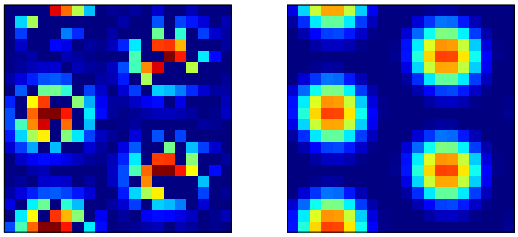


Figure 4.21: Side-by-side comparison of the pattern of the backprojected signals (left) and the final network pattern after the backprojection test was finished (right). Left is from Figure 4.19 and right is at $t = 1.0$ s from the same simulation as in Figure 4.20.

4.1.5 Recalling Hippocampal State from Sensory Inputs

With place–event associations in CA3 and hippocampal backprojections to MEC confirmed to work, the model was tested for the ability to recall the previous state in the entorhinal cortex and the hippocampus for a given spatial location, based on sensory inputs to CA3 alone. The model was trained as usual, and then tested as specified in Section 3.7.3—the MEC grid networks were reset, sensory signals corresponding to the central location in the training enclosure were sent to CA3, and the grid modules were allowed to be influenced by CA1 backprojections while the grid sustain signal was gradually turned up from zero.

Figures 4.22, 4.23, 4.24 and 4.25 show situational overviews of the model at different points in time during the recall simulation. For CA3 and DG, the figures show the momentary distributions of firing rates (below) and the correlations of these firing vectors with the recorded firing activity from the training environment at different coordinates (above). For MEC, the values of the output axons—the network patterns—are shown for each of the four grid modules. Arrows from CA3 to MEC, from MEC to DG and from DG to CA3 show how strongly these areas influence each other at the various points in time, as indicated by the different shades of gray on the arrows.

Figure 4.22 shows the state of the model right after the test has been initialized. CA3

does not receive any DG input, neither excitation not inhibition, so it is fully influenced by the sensory inputs. As the CA3 firing rate distribution plot shows (bottom left), many of the CA3 neurons are triggered to small or large extents by the sensory inputs. When the activity is compared to the historical activity at different coordinates within the cylindrical enclosure (top left), there is seen to be some level of correlation with many different locations, but the center locations have the strongest correlations.

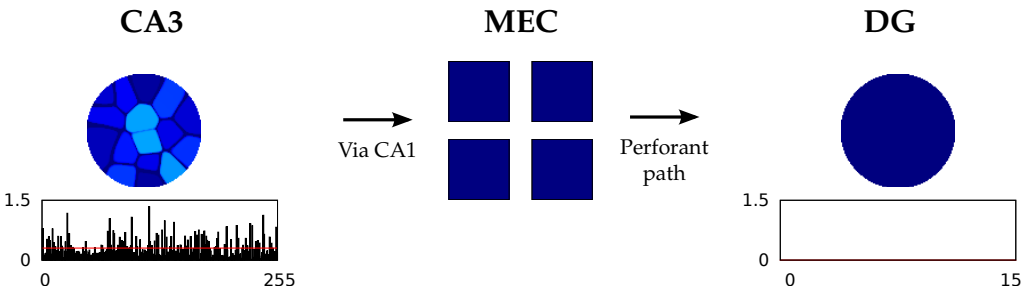


Figure 4.22: Network situation at time $t = 0.0$ s.

In Figure 4.23, 0.1 seconds have passed in the simulation. The sensory induced CA3 activity has propagated to CA1 and further on to MEC as backprojections. As seen in the MEC activity plots (middle), some of the grid cells are now weakly activated due to these backprojections. Even though the network patterns in the MEC grid modules have not yet strengthened to the canonical grid patterns, the input vectors to DG from MEC are familiar enough for the correct DG place cell to be activated. However, the grid sustain signal is still too weak for the place cell to fire strongly, and the activity in CA3 is thus still mostly unaffected.

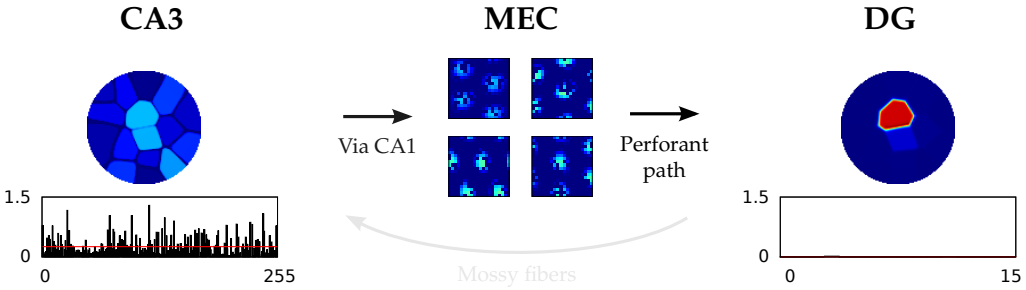


Figure 4.23: Network situation at time $t = 0.1$ s.

The situation after another 0.2 seconds is shown in Figure 4.24. The patterns in the MEC grid modules are strengthening, and the activated DG place cell is exerting more

influence over the firing activity in CA3 than previously. This can be seen by the increasingly sparse firing distribution in CA3, as the inhibition from DG disables many of the weakly activated CA3 neurons. Furthermore, the CA3 cells that are associated with the triggered DG place cell receive additional excitation from DG, contributing to an increased spatial correlation for the CA3 activity with the place field in question.

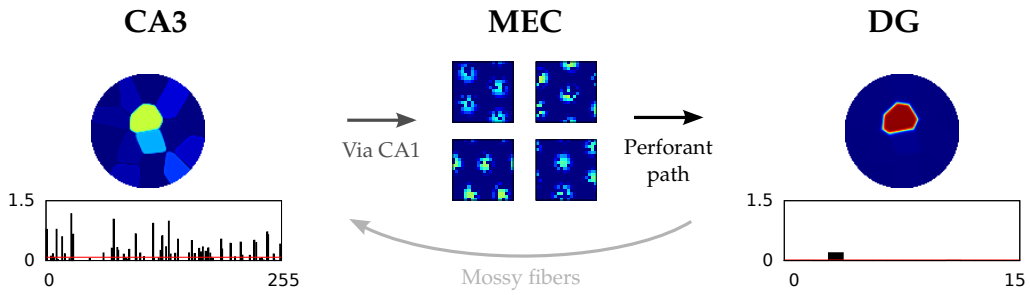


Figure 4.24: Network situation at time $t = 0.3$ s.

Figure 4.25 shows the situation after 1.0 seconds, when the grid sustain signal has been turned up maximally and the backprojections to MEC have ceased. The activity in the network is now fully controlled by the MEC grid modules, as indicated by the arrows. CA3 receives strong inhibition from DG, so that only CA3 neurons with the correct place fields are active. With the combined excitation from DG and the sensory inputs, the CA3 activity is strongly correlated with that seen during the training session.

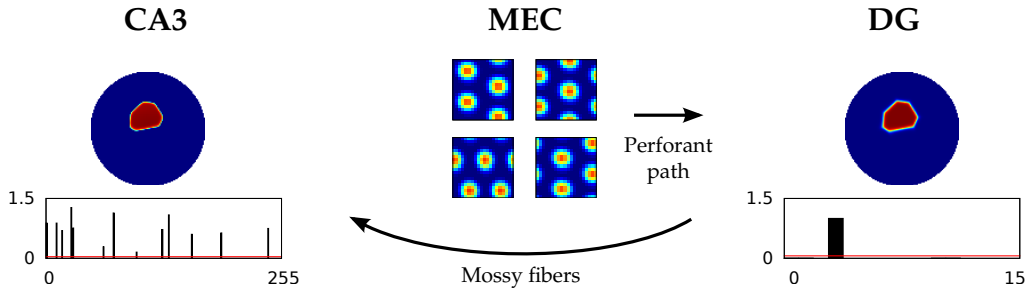


Figure 4.25: Network situation at time $t = 1.0$ s.

4.2 Methods of Analysis for Remapping Experiments

The mechanisms documented in the previous section, of place cell formation, place–event associations and backprojections to the path integrator—were then combined

in order to try to reproduce the phenomena of global remapping and rate remapping known from animal experiments.

The model was trained and tested as described in Section 3.7. As indicated in these procedures, the model is trained in two training environments, E_A and E_B. Depending on whether a rate or global remapping experiment is performed, the training environments will be located at the same site or at two different sites, respectively. In other words, the MEC coordinates will be the same for the two training environments during the rate remapping experiment, while during the global remapping experiment the two training environments will see different coordinates represented in the MEC grid networks.

To account for the extra MEC states that need to be represented by DG place cells in the global remapping experiment, there are twice as many DG neurons in the global remapping simulations as in the rate remapping simulations—32 and 16, respectively.

When the networks are done training, the model is tested in 16 different test configurations for the sensory inputs, as described in Section 3.7. At the end of the recall phase for each of these test configurations, i.e. when $G(t)$ reaches 1, the responses of the neurons in CA3 are analyzed in two different ways.

The first kind of analysis is to correlate the final population vector in CA3 yielded by each of the 16 test runs with the activity recorded at various spatial coordinates in the two training environments E_A and E_B, i.e. the same type of plot as the correlation plots shown e.g. in Figures 4.16 and 4.18. These plots will allow for a visual interpretation of what kind of state the hippocampal model has ended in—correlated only to E_A, only to E_B or to a mixture of both?

A more quantitative method of analysis was inspired by Wills et al.'s study from 2005 on attractor-like hippocampal responses during remapping experiments [26]. Their approach was to calculate *similarity* scores between pairs of three-dimensional spatial response stacks, such as the stack illustrated in Figure 4.15. In other words, given two trials, e.g. T_1 and T_8, a similarity score between the hippocampal response in the two trials can be calculated by supplying the spatial response maps from all neurons in both of the trials.

In practice, this means that spatial response plots from each of the CA3 neurons need to be extracted at the end of each test run T_{1-16}. This was done in the same way as for the spatial response plots shown previously in this chapter. At the end of a simulation run of all 16 test environments, the recorded data thus consisted of a four-dimensional

data set containing the firing rate for all combinations of (t, i, x, y)—the firing rate of neuron i at coordinates (x, y) after performing the recall test in test environment T_t.

Two test trials were designated as the A and B trials, to represent the start and end trials in the morph sequence between the two training environments. For a given neuron i and a given intermediate test trial K, the similarity $s'_i(K, A)$ to A and $s'_i(K, B)$ to B can then be calculated as follows:

$$s'_i(K, A) = \frac{r_{KA}(i) - r_{AB}(i)}{r_{AA}(i) - r_{AB}(i)}, \qquad s'_i(K, B) = \frac{r_{KB}(i) - r_{AB}(i)}{r_{BB}(i) - r_{AB}(i)}. \tag{4.2}$$

$r_{MN}(i)$ is the correlation of neuron i's spatial response map after trial M to the same neuron's map after trial N. r is calculated as the Pearson product-moment correlation coefficient, using

$$r = \frac{\sum_j \left(X_j - \bar{X}\right)\left(Y_j - \bar{Y}\right)}{\sqrt{\sum_j \left(X_j - \bar{X}\right)^2}\sqrt{\sum_j \left(Y_j - \bar{Y}\right)^2}} \quad \text{[25]}, \tag{4.3}$$

where X_j and Y_j are vectors containing firing values from corresponding coordinates in trials N and M respectively.

Some of the test configurations belong naturally together in *groups of trials*, for example the four test configurations T_{2-5}, because they have the same ratios of similarity to the training environments. It would be useful to consider the results from these similar test configurations as one common trial group I. For this purpose, $\langle r(i)\rangle_{IN}$ is defined to be the average of $r_{MN}(i)$ for all trials M belonging to trial group I. This yields the following definitions for the similarity of neuron i's responses in trial group I to trials A and B:

$$s_i(I, A) = \frac{\langle r(i)\rangle_{IA} - \langle r(i)\rangle_{AB}}{\langle r(i)\rangle_{AA} - \langle r(i)\rangle_{AB}}, \quad s_i(I, B) = \frac{\langle r(i)\rangle_{IB} - \langle r(i)\rangle_{AB}}{\langle r(i)\rangle_{BB} - \langle r(i)\rangle_{AB}} \quad \text{[26]}. \tag{4.4}$$

Taking the averages of $s_i(I, A)$ and $s_i(I, B)$ over all neurons i in CA3, yields the final similarity scores for trial group I's similarity to test trials A and B, respectively.

4.3 Global Remapping Experiment

Figure 4.26 shows the network situation at the beginning and end of the test run with test configuration T_1, which in sensory configuration is equal to training environment E_A, as seen in Table 3.8. The figure is similar to the network figures in Section 4.1.5, with the exception that the spatial correlation plots for CA3 and DG now show the correlation with both training environment E_A (left) and training environment E_B (right). The figure shows a strong initial correlation to training environment E_A in CA3. In the final state of the network, after the recall procedure was finished, the network has settled into a strong correlation with training environment E_A and no correlation with training environment E_B.

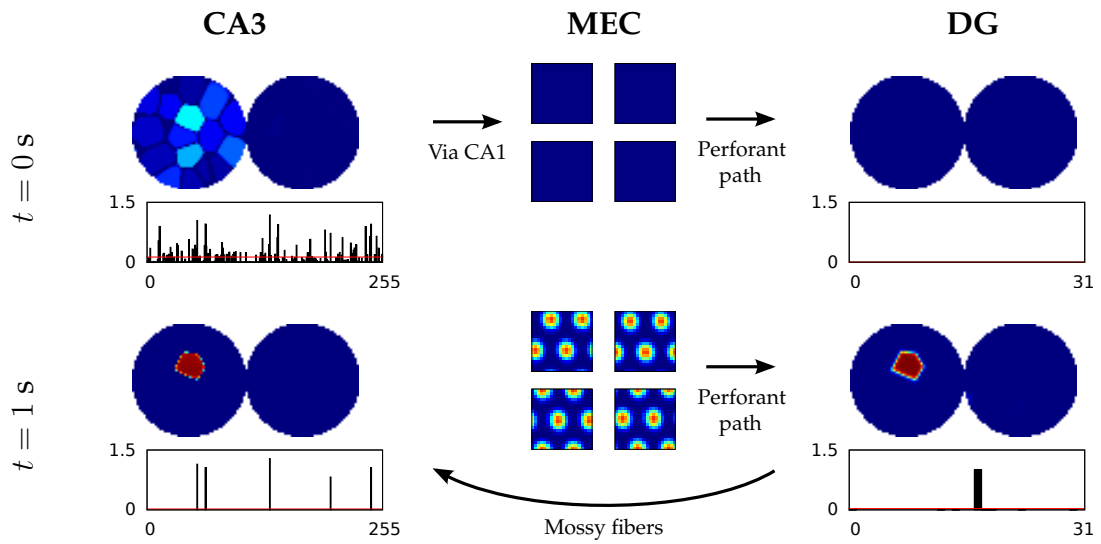


Figure 4.26: Global remapping experiment: Initial (above) and final (below) network state for test configuration T_1, which resembles training environment E_A.

Figure 4.27 shows the same data as Figure 4.26, but this time for a test run with test configuration T_{16}, which is equivalent to training environment E_B. Again there is a strong initial CA3 correlation with the correct training environment—this time, training environment E_B. When the network has settled, the correlation to training environment E_B has strengthened, and there is no correlation with training environment E_A.

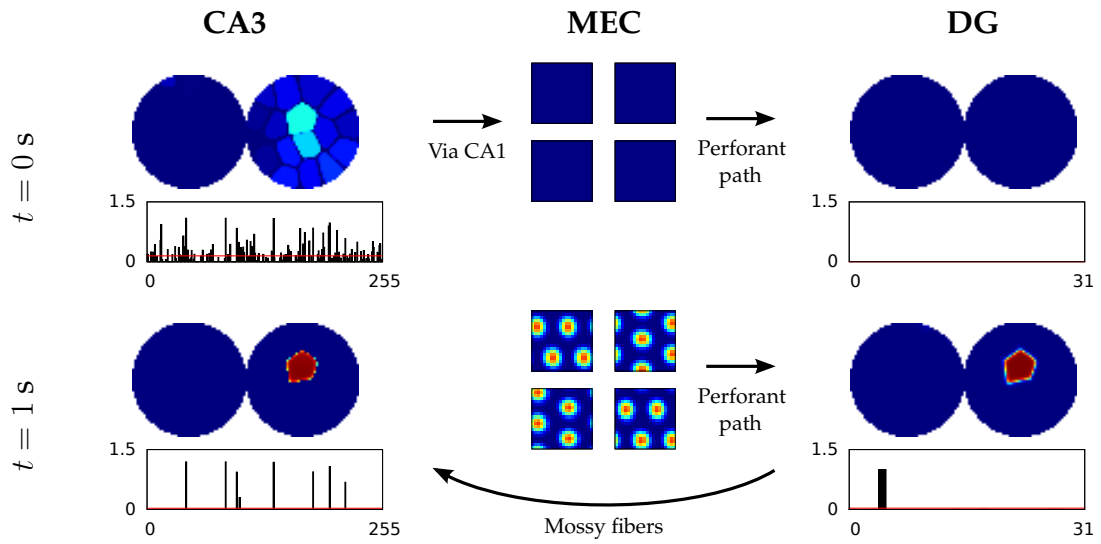


Figure 4.27: Global remapping experiment: Initial (above) and final (below) network state for test configuration T_{16}, which resembles training environment E_B.

Figures 4.26 and 4.27 show results that would be expected for a global remapping situation. The networks are able to settle to states that are consistent with the environment indicated by the sensory stimuli. The population vectors once the networks are in their settled states, are orthogonal to the saved responses from the other environment, and this holds true for both DG and CA3.

These two test runs consist of sensory configurations that unambiguously indicate one of the two training environments. That is, the sensory configuration is either identical to training environment E_A or to training environment E_B. It is, however, also important to examine how the networks respond when the sensory stimuli are ambiguous and indicate a mix of the two training environments. This will reveal whether the model implements attractor dynamics in the hippocampal response, as found to be true in rat experiments by Wills et al. [26], or whether the hippocampal response will show a mixture of the responses for training environments E_A and E_B.

Figure 4.28 gives a condensed overview of the CA3 responses for all of the 16 test configurations listed in Table 3.8. The vertical columns contain the network responses for individual test configurations, as specified by the numbers on top. The first and the second rows contain the spatial correlation plots with training environments E_A and E_B, respectively, for each of the 16 test runs. The upper half of the figure shows the network responses at the beginning of the test runs, and the lower half shows the responses at the end of the test runs.

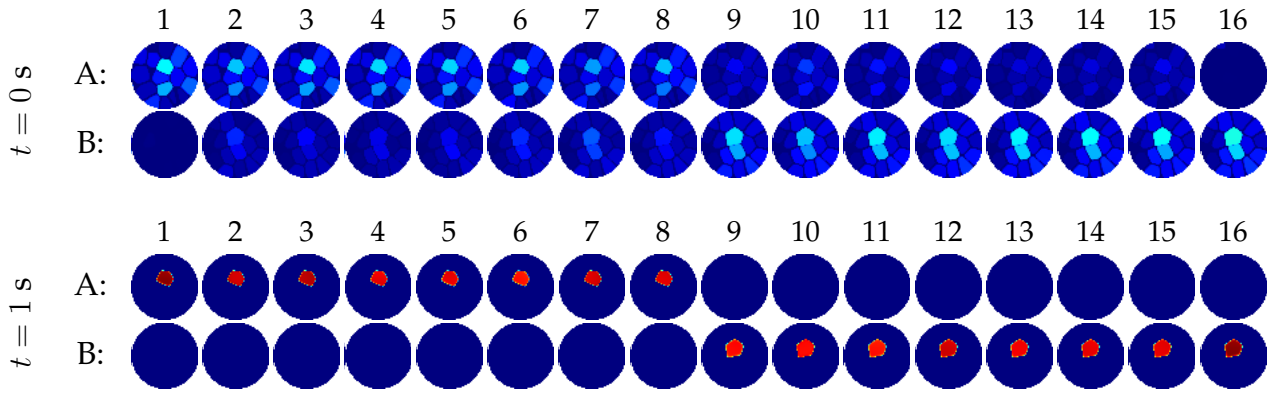


Figure 4.28: Correlation plots of CA3 responses for each of the 16 test configurations (columns) with training environments E_A and E_B (rows), at simulation times $t = 0$ s (upper half) and $t = 1$ s (lower half).

The upper half of Figure 4.28 shows how the initial CA3 response, only influenced by the sensory stimuli, changes as the sensory configurations become gradually less similar to training environment E_A and gradually more similar to training environment E_B. While the responses always show a dominance in the correlation with one of the two training environments, there is usually a weak correlation also with the other environment. In the lower half of the figure, however, it can be seen that the final response from the model, after the recall phase has been completed, strictly belongs to either environment E_A or E_B—the correlation with the non-dominant training environment has become orthogonal. This shows that there is an attractor dynamic in the remapping response of the model.

The change-over from environment E_A to E_B happens between the 8th and the 9th test configuration. This is at the same point in the test configuration sequence when the sensory stimuli change over from having a majority coming from environment E_A to coming from environment E_B—in the 8th configuration, $^2/_3$ of the active axons come from environment E_A, while in the 9th configuration, $^2/_3$ come from environment E_B.

To provide a sense of what these results mean for individual CA3 neurons, a series of spatial response plots for five example neurons is shown in Figure 4.29. The columns represent the 16 test configurations, while the rows represent the five different neurons. The five neurons in the figure were selected out of the 256 CA3 neurons to be the ones with the largest difference in peak firing rate between the first and the last test configurations.

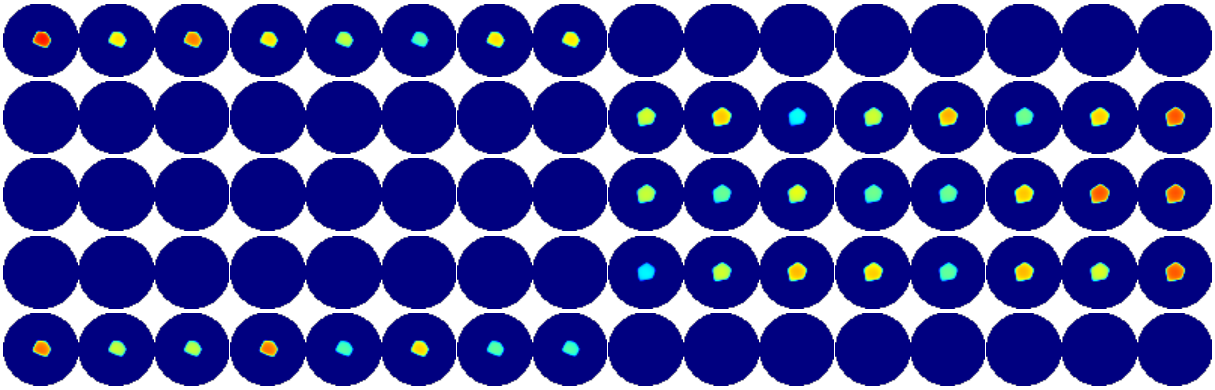


Figure 4.29: Spatial response plots for five example CA3 neurons (rows) in each of the 16 test runs (columns).

As seen in the figure, each neuron is active in only half of the 16 test runs—the first eight or the last eight. This coincides with the change-over from the environment E_A-response to the environment E_B-response that happens between the 8th and the 9th test configuration, as noted above. During all of the test runs that a given neuron is active, it retains the same place field—the only variance between the test runs is in the distribution of firing rates across the place field. This is consistent with the expectations for a global remapping response.

For a quantitative assessment of the degree to which the network responses from the various intermediate test configurations are similar to the results from the two test trials T_1 and T_{16}, corresponding to the training environments E_A and E_B respectively, the similarity measure from Wills et al. [26] was used. The results are shown in Figure 4.30.

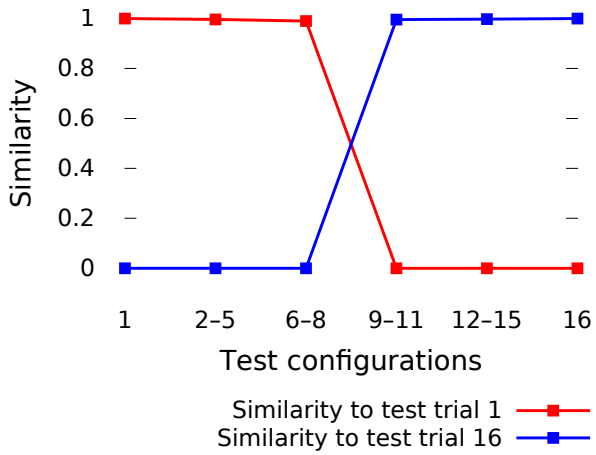


Figure 4.30: Similarity score of the response from intermediate trial groups to the test results from test trials T_1 and T_{16}, corresponding to the training environments E_A and E_B respectively.

The similarity measure chart emphasizes the global remapping aspects of the results shown and discussed above. When the majority of the sensory inputs correspond to training environment E_A, the CA3 response is highly similar to T_1, representing training environment E_A, and not similar to T_{16}, representing the opposite training environment. When the majority changes over to training environment E_B, there is an abrupt change in the CA3 response, to a strong similarity to the environment E_B-response and no similarity to the environment E_A-response.

To verify that the successful demonstration of global remapping seen in Figure 4.30 is replicable, the full simulation of the global remapping experiment was repeated 20 times. Figure 4.31 summarizes these runs by showing the final similarity measure chart, equivalent to Figure 4.30, for each of the 20 simulations.

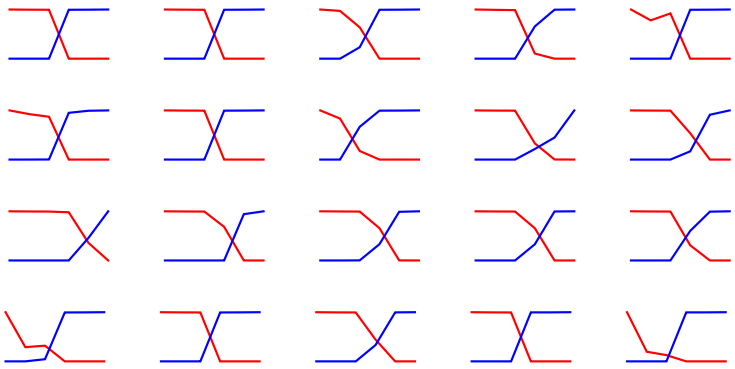


Figure 4.31: Final similarity measure chart from 20 different runs of the global remapping experiment, to verify the replicability of the results shown above.

While there are anomalies in some of these 20 simulations, the majority of the simulations show results that look appreciatively similar to the similarity chart seen in Figure 4.30.

Calculating the median for each value across the 20 simulations seen in Figure 4.31 yields the chart shown in Figure 4.32. The resulting chart expresses the same features as discussed for Figure 4.30, with a high degree of similarity to one test environment up until the middle of the morph sequence, when the similarity scores abruptly change over. This demonstrates that the implemented hippocampal model shows distinct characteristics of global remapping.

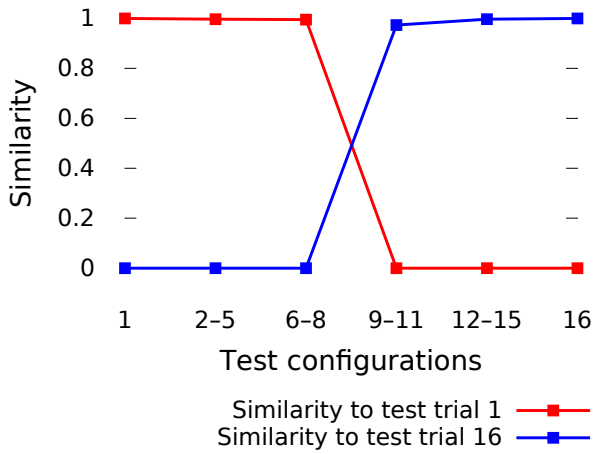


Figure 4.32: Median similarity chart based on the results from Figure 4.31.

4.4 Rate Remapping Experiment

The experiment from the previous section was repeated, but with a change in the training environments to elicit a rate remapping response instead of the global remapping response seen above. Specifically, the training environments E_A and E_B were set up to use the same training site α, instead of having separate training sites α and β as in the previous section. As discussed in Section 2.2, it is believed that this will cause the same place fields to be active in both training environments, but that the CA3 cells will vary their peak firing rates differently in the two environments in response to the sensory stimuli at their respective locations, i.e. a rate remapping response.

Figure 4.33 shows the initial and final network situations for a test run with test configuration T_1 from Table 3.8, which is equivalent to the sensory configuration seen during training in environment E_A. Figure 4.34 shows the same information for a test run with test configuration T_{16}, equivalent to training environment E_B. These two figures are comparable to Figures 4.26 and 4.27, respectively, from the global remapping experiment.

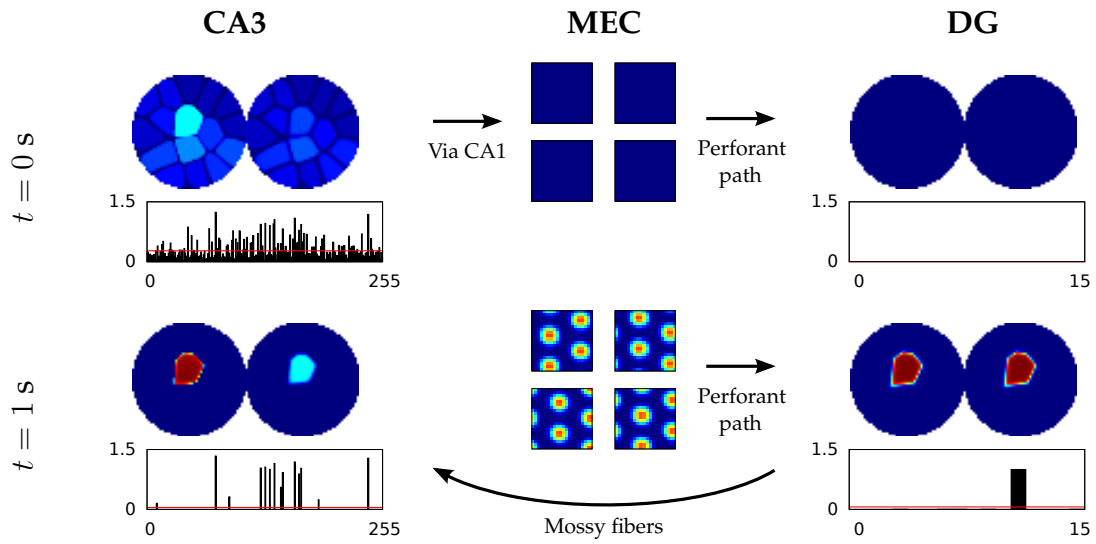


Figure 4.33: Rate remapping experiment: Initial (above) and final (below) network state for test configuration T_1, which resembles training environment E_A.

Figure 4.33 shows that the initial CA3 response is correlated to both of the training environments, but most strongly to environment E_A. After the network has settled—at the end of the recall phase—the CA3 response is strongly correlated to the central area

of training environment E_A (left half in the figures). The network has thus performed a successful recall of the correct network state for the location indicated by the sensory inputs.

However, there is also a noticeable correlation to a place field of an identical shape in the other training environment (right half in the figures), but with a significantly weaker correlation. This is reflected in DG, where the same place field correlation is seen in both environments. This is as expected, as the two training environments E_A and E_B are located at the same training site α and thus use the same MEC patterns to represent equivalent locations in the two environments. The DG response will thus be identical between the two environments, but the CA3 neurons can distinguish between the two based on the sensory stimuli they receive.

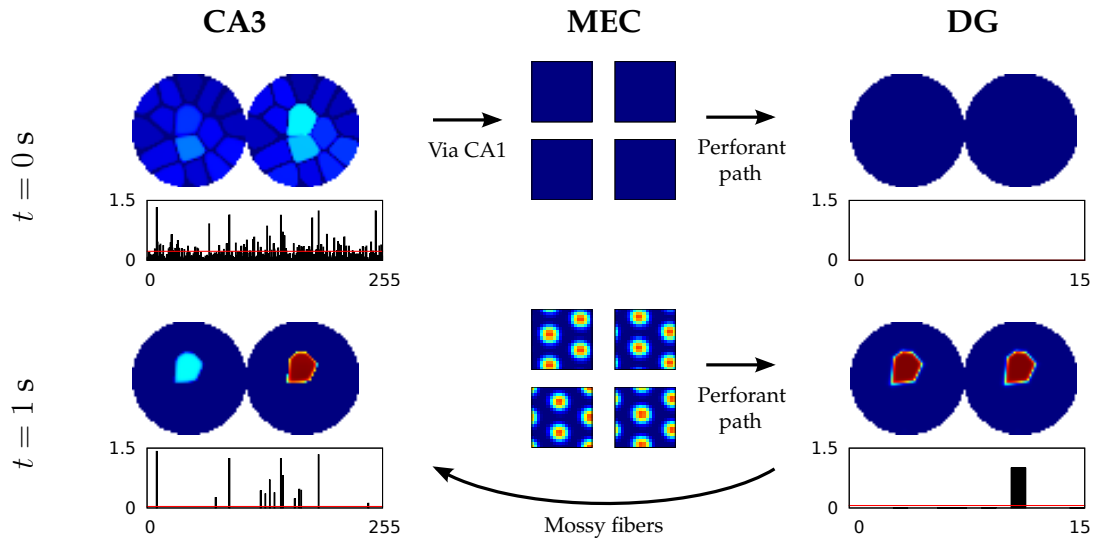


Figure 4.34: Rate remapping experiment: Initial (above) and final (below) network state for test configuration T_{16}, which resembles training environment E_B.

Figure 4.34, from the test run with configuration T_{16}, shows a response similar to Figure 4.33, but with dominance in the correlations with training environment E_B instead of E_A. The DG state is the same as in the previous figure, as expected due to the identical MEC state between the two training environments, but CA3 shows a stronger correlation with E_B instead of E_A. This is again consistent with rate remapping—the place fields are retained but the differences in sensory inputs are coded into differences in the CA3 neurons' peak firing rates.

Figure 4.35 shows sequences of correlation plots for the CA3 responses in the initial

and final states of the 16 different test runs, equivalent to Figure 4.28 from the global remapping experiment. As in the global remapping experiment, the spatial correlation of the initial CA3 response is gradually changing from predominantly E_A to predominantly E_B. In contrast to global remapping, however, the same gradual change in the correlations is also visible for the *final* CA3 responses. In other words, there is no longer any abrupt change-over between the 8th and the 9th test configurations, but a gradual decrease in correlation to E_A and a gradual increase in correlation to E_B. The CA3 responses are not orthogonal to the non-dominant training environments, as during global remapping.

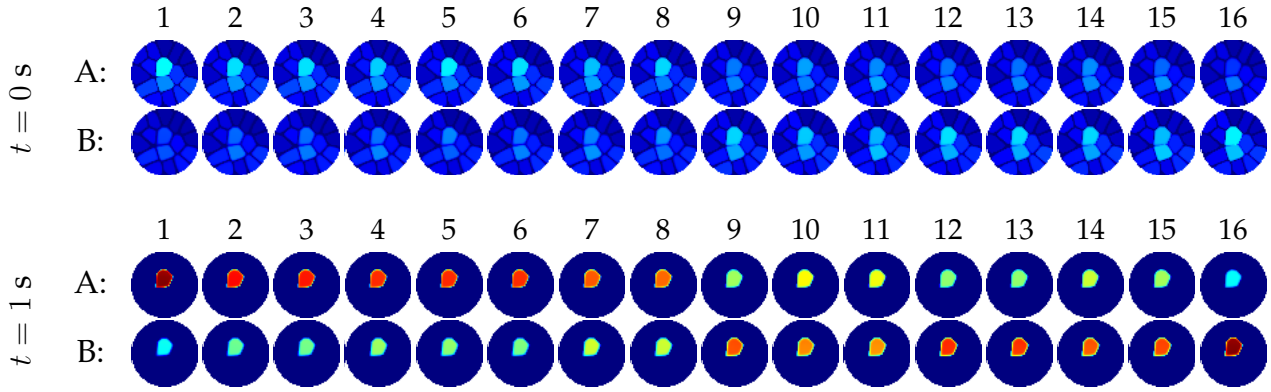


Figure 4.35: Correlation plots of CA3 responses for each of the 16 test configurations (columns) with training environments E_A and E_B (rows), at simulation times $t = 0$ s (upper half) and $t = 1$ s (lower half).

Figure 4.36 shows sequences of spatial response plots across the 16 different test configurations for five example CA3 neurons, as in Figure 4.29 for the global remapping experiment. The figure indicates the same development in the response of individual neurons as for the entire population as demonstrated in the previous figure—a gradual change in peak firing rate instead of an abrupt switch between the 8th and the 9th test configurations.

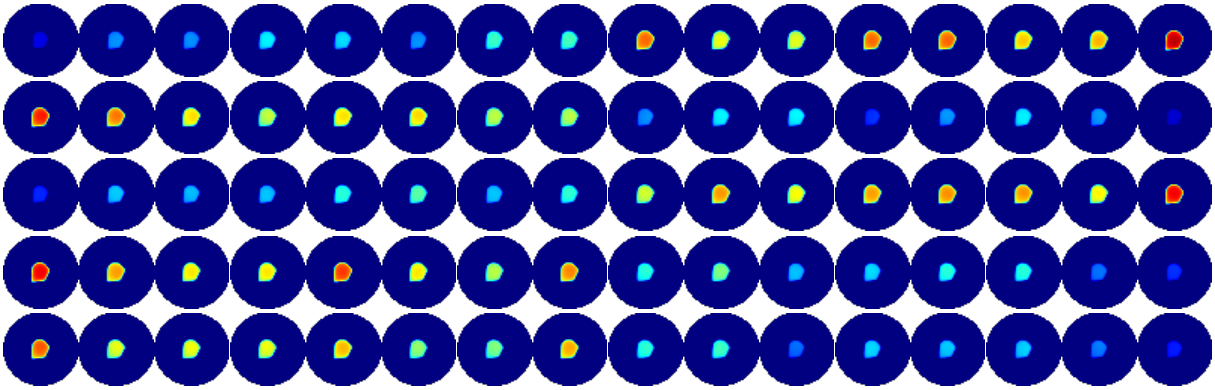


Figure 4.36: Spatial response plots for five example CA3 neurons (rows) in each of the 16 test runs (columns).

The same similarity measure as calculated for the global remapping experiment was also calculated for the data from the rate remapping experiment. The results are shown in Figure 4.37. The chart echoes the results seen from the plots sequences above, where the change resulting from environmental modifications is gradual, almost linear, instead of the abrupt change that characterizes global remapping.

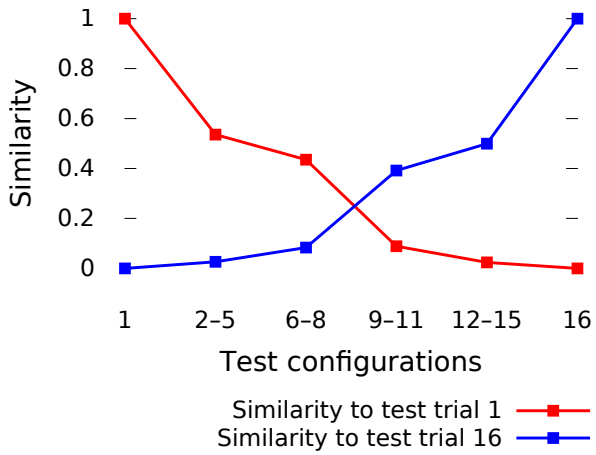


Figure 4.37: Similarity score of the response from intermediate trial groups to the test results from test trials T_1 and T_{16}, corresponding to the training environments E_A and E_B respectively.

As with the global remapping experiment, the rate remapping experiment was repeated 20 times to verify that the results are replicable. The similarity charts from each

of these 20 runs are shown in Figure 4.38.

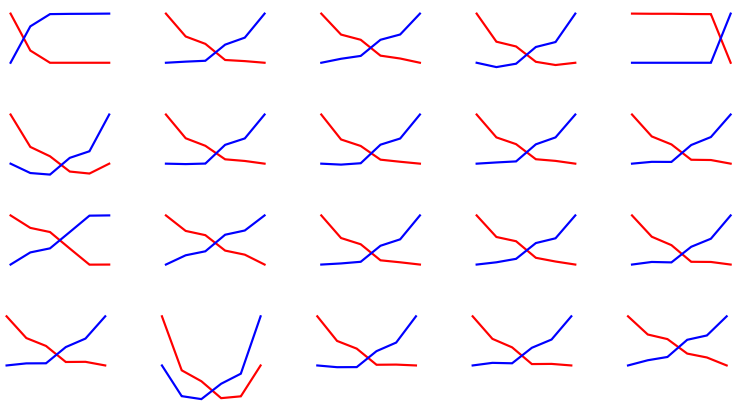


Figure 4.38: Final similarity measure chart from 20 different runs of the rate remapping experiment, to verify the replicability of the results shown above.

Again some of the simulation runs show deviating results, but the majority of the similarity charts are comparable to the one shown in Figure 4.37. To show the trend from these 20 runs, the median similarity scores are plotted in Figure 4.39. This chart shows results that are highly similar to the ones discussed above, which confirms that the implemented model consistently shows characteristics expected of rate remapping.

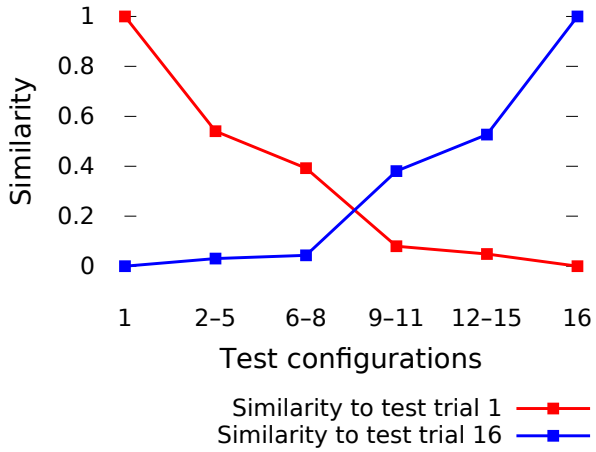


Figure 4.39: Median similarity chart based on the results from Figure 4.38.

4.5 Discussion of Results

For the convenience of the discussion, the research questions given in the introduction chapter of this report are repeated below:

Research Question 1: Can place cells and grid cells be modeled to alter their firing activity in response to sensory information from the environment?

Research Question 2: Can such a computational model be made to be compatible with the phenomenon of global remapping?

Research Question 3: Can such a computational model be made to be compatible with the phenomenon of rate remapping?

Research Question 1 essentially asks whether sensory stimuli can cause backprojections to MEC grid modules that make these networks of grid cells alter their firing patterns. Section 4.1.3 showed that CA3 neurons are able to associate sensory stimuli with locations in the environment, as place–event associations. Section 4.1.4 showed that firing activity in the hippocampal submodules can induce given patterns of firing in the MEC grid modules through backprojections from CA1.

These two features of the model were combined in Section 4.1.5 to show the recall of a previous state in the entire model, including the entorhinal cortex, based on sensory inputs alone. A clear example of this recall process occurring was offered by Figure 4.23, which has been repeated in Figure 4.40 below. The figure shows MEC and DG restoring a previous state based on sensory inputs to CA3, demonstrating a positive answer to Research Question 1.

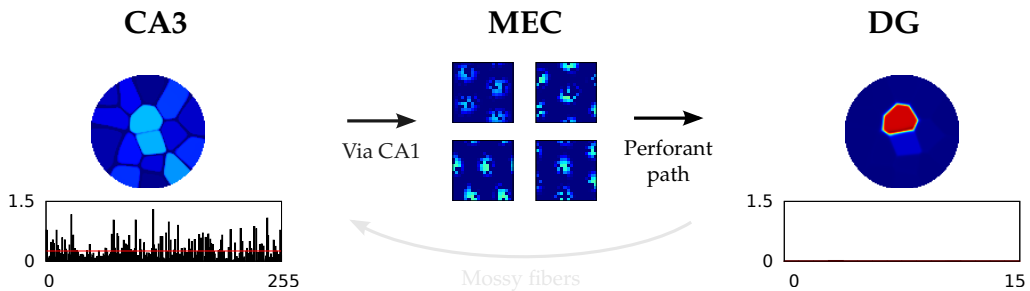


Figure 4.40: Copy of Figure 4.23, showing the network situation of the hippocampal model at $t = 0.1$ s into a recall test.

The recall mechanism postulated by Research Question 1 is an important prerequisite to Research Questions 2 and 3, which ask whether the model exhibits the features of global and rate remapping, respectively. By testing the network response after recall simulations in various sensory configurations, the remapping features of the model could be examined. Figure 4.41 shows the sequences of final correlation plots for example runs of the global and rate remapping experiments, originally shown in Figures 4.28 and 4.35.

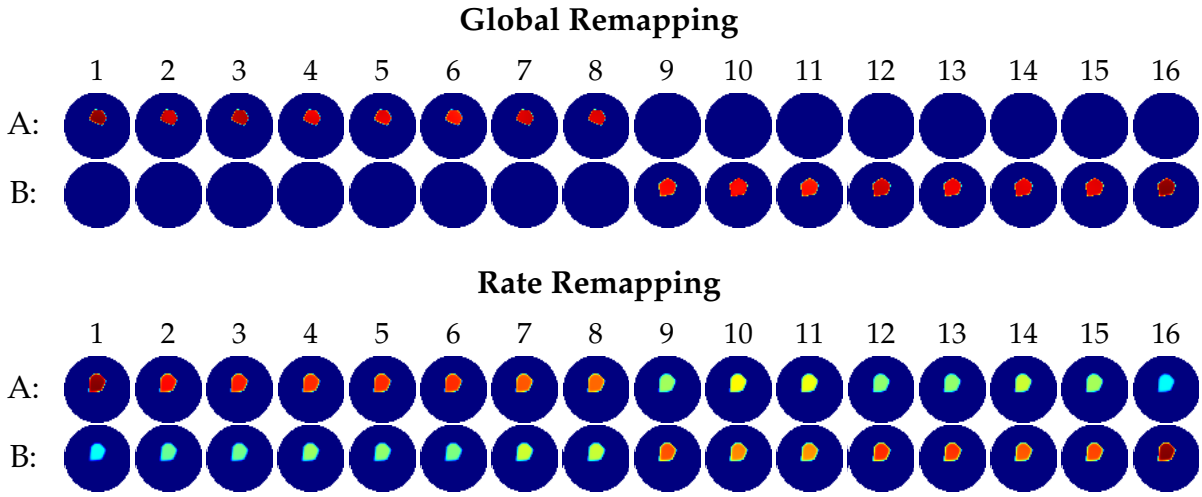


Figure 4.41: Global and rate remapping results from Figures 4.28 and 4.35, respectively. Shows final correlation plots of CA3 responses for each of the 16 test configurations with training environments E_A and E_B.

The two experiments differed only in the number of different sites used during training, as well as the number of DG cells. With a different site used for the two training environments, the model showed a global remapping response. With the same site used for the training environments, a rate remapping response was seen instead. The global remapping response is visible as the abrupt change-over between the 8th and the 9th test configuration in the sequence above, while the rate remapping response is visible as the gradual change from high correlations to low correlations and vice versa.

A similarity measure due to Wills et al. [26] was used to quantitatively describe the similarity of intermediate test responses to the results from the two test trials T_1 and T_{16}, corresponding to the training environments E_A and E_B respectively. The median values of these similarity measures, from 20 different runs of each of the two experiments, were used to create the similarity charts shown in Figure 4.42.

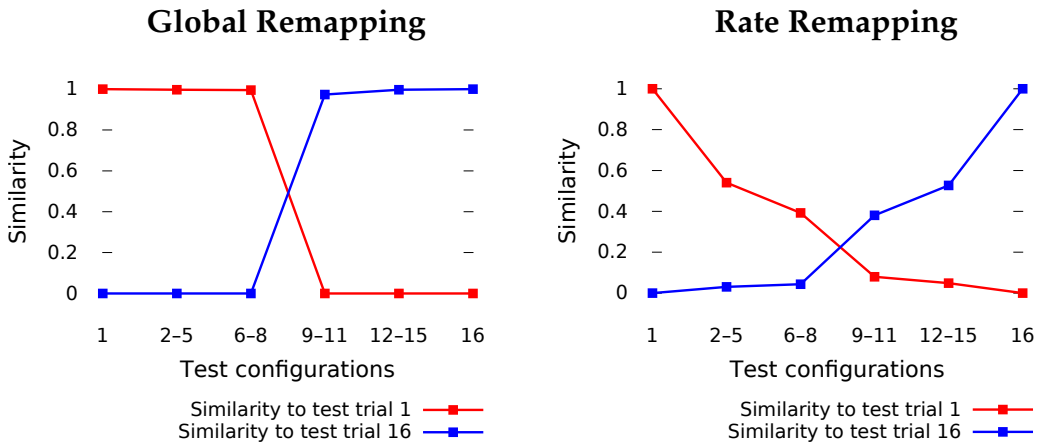


Figure 4.42: Median similarity charts from the global and rate remapping experiments, repeated from Figures 4.32 and 4.39, respectively.

The similarity charts reflect the results seen in the correlation plots—the global remapping experiments show an abrupt change in environmental correlations, while the rate remapping experiments show a gradual change. This is consistent with the expectations from neuroscientific experiments. Remapping experiments with animals show attractor-like change between spatial maps in the hippocampus during global remapping, thought to be caused by the training enclosures being located in different places in the laboratory, while there is only a change in the rate distribution within the same set of place fields during rate remapping.

These results point to positive answers for Research Questions 2 and 3 as well. This should, however, not be interpreted as a claim of biological validity for the implemented model. Instead, the main idea has been to show that the proposed mechanism behind the distinction between rate and global remapping—the mechanism of one versus multiple training sites—can be modeled in a simplified but plausible way to produce the expected remapping phenomena.

5 Discussion

This chapter discusses some novel aspects and possible uses of the implemented model, points out areas for future work and seeks to make conclusions to the research questions pursued in this project.

5.1 Contributions

This section points to some aspects of the implemented model that are believed to be novel contributions to the area of modeling spatially responsive hippocampal neurons.

The implemented hippocampal neural network, summarized in Figure 3.10, bears an architectural resemblance to the hippocampal model described by E.T. Rolls and A. Treves [22]—the dentate gyrus, represented as a competitive network, receives inputs from the entorhinal cortex and is a major gateway of signals to the rest of the hippocampal areas, which pass through the CA3 and CA1 areas before backprojecting to the entorhinal cortex.

However, the model described by Rolls & Treves is not concerned with spatial information, but was instead used to measure the ability for recall of arbitrary bit vectors in the entorhinal cortex [19]. In the implemented model, on the other hand, the exact values of the backprojections to the entorhinal cortex are not interesting, as much as the spatial responses that are elicited from the hippocampal neurons after a recall operation. The implemented hippocampal model uses a series of continuous attractor networks of grid cells in the entorhinal cortex as the main source of inputs to the hippocampus and as the recipient of hippocampal backprojections, and it is believed that this is a novel construction.

The different networks of the implemented model are largely based on models de-

scribed by others, such as the attractor networks of grid cells due to Burak & Fiete [3], and the competitive learning and pattern association networks described by Rolls & Treves [22]. CA3, however, is not implemented as an autoassociative network as suggested among others by Rolls & Treves [22], but is implemented as a pattern association network without any recurrence. The varied firing dynamics of CA3 in familiar and unfamiliar environments, that gives rise to the two different remapping phenomena in the model, is aided by the "excited-by-one, inhibited-by-all" mechanism in the DG–CA3 connectivity that was described in Section 3.5.2. This mechanism was devised particularly for this project.

Another aspect of the implemented model, that has been specifically devised for this project in order to examine remapping phenomena, is the nature of the training and testing procedures—training the model in various training environments with different sensory inputs, potentially involving multiple different training sites, and then testing the model in testing environments that gradually transform from resembling one training environment into resembling the other. These configurations and procedures have enabled the analysis of global and rate remapping-like responses from the model.

5.2 Possible Uses

When viewed on a reasonably high level, the implemented model is biologically plausible. The overall architecture and the major connectivity pathways of the model, as well as the network properties of the different areas, are based on knowledge about the hippocampus and about biologically plausible neural networks. There is thus also the potential for a neuroscientific application of the implemented model, given a sufficiently high level of abstraction. For example, the implemented model might be used to validate high-level hypotheses or to make high-level predictions about the interplay between and the importance of different hippocampal areas in navigational tasks.

In the realm of computer science and artificial intelligence, the model can potentially be used as a starting point for a subsymbolic robot controller. The implemented model is already able to perform path integration, associate sensory stimuli to visited locations and to reinstate previously visited coordinates in the path integrator based on familiar sensory stimuli. Localization of the robot within its environment and learning to know this environment are important tasks for a robot controller, and these abilities of the implemented model could potentially form part of a navigational subsystem

within a larger robot controller.

5.3 Future Work

Future work on the implemented model in the direction of computational neuroscience, could be to iteratively refine the biological realism of the model. The framework is in place of a biologically plausible hippocampal model that exhibits remapping phenomena, but the model is far from representing an accurate and complete computational simulation of the hippocampus. By updating the model e.g.

- with more anatomically correct data for the number of neurons, axons and synapses in each part of the hippocampus,
- by increasing the realism of the firing characteristics of the individual neurons by introducing spiking neurons and noisy signalling, or
- by placing a greater emphasis on the time dimension of the firing activity by modeling theta wave patterns and differences in timing synchrony between different parts of the model,

to name a few suggested areas for exploration, the model could potentially be used to investigate other kinds of phenomena from experimental neuroscience in the future.

In the direction of artificial intelligence and robotics, the viability of the model as a component for a robot controller could be further explored. The hippocampal model could be integrated in a larger model of the cerebral cortex, which comprises the neocortex in addition to the hippocampus. The neocortex processes sensory inputs to highly abstract representations before they reach the hippocampus. It is also the target for most of the hippocampal backprojections, and some of these backprojections might eventually propagate to the motor areas of the neocortex and end up as motor outputs from the brain.

As a part of a larger model of the cerebral cortex, the implemented hippocampal model could be adapted to use the sensory representations generated by the neocortical part of the system, instead of the abstract, simulated sensory inputs used during this project. These sensory inputs could potentially come from different sensors

attached to the robot controller, enabling the hippocampal model to perform place–event associations involving sensory events from the robot's real environment.

A different avenue of exploration for the hippocampal part of a cerebral cortex model is to generate motor outputs to solve navigational tasks. Given a learned map of the environment, given a specific place that the robot controller wants to move to and assuming that the hippocampal model has localized the robot in the environment, the hippocampus could potentially calculate the necessary bearing for the robot controller and send this information to the neocortex as backprojections that eventually help drive the motor outputs of the controller.

5.4 Conclusion

This project has sought to replicate the phenomena of global and rate remapping, known from neuroscientific experiments into hippocampal function. An early assumption of the model was the hypothesis that the global coordinates of the experimental enclosure during training determines the outcome of the remapping experiment. Given this assumption, the model has shown an ability to account for both global remapping and rate remapping.

Although the model is not biologically realistic in all of its aspects, the overall architecture, the pathways of information flow and the local network functionality of each area of the model are biologically plausible. The model should thus possess some degree of explanatory utility when considering various hypotheses for remapping behavior.

The motivation for modeling remapping phenomena was largely concerned with explaining results from neuroscience, but as discussed above, the knowledge gained from this model could also potentially be useful for robot controllers in the field of artificial intelligence. This goes to show the benefit of cooperation between these two fields, both toward understanding the brain and toward achieving machine intelligence.

Bibliography

[1] BONNEVIE, T., DUNN, B., FYHN, M., HAFTING, T., DERDIKMAN, D., KUBIE, J. L., ROUDI, Y., MOSER, E. I., AND MOSER, M.-B. Grid cells require excitatory drive from the hippocampus. *Nature neuroscience* (2013).

[2] BURAK, Y., AND FIETE, I. R. Do we understand the emergent dynamics of grid cell activity? *The Journal of neuroscience 26*, 37 (2006), 9352–9354.

[3] BURAK, Y., AND FIETE, I. R. Accurate path integration in continuous attractor network models of grid cells. *PLoS computational biology 5*, 2 (2009), e1000291.

[4] COUEY, J. J., WITOELAR, A., ZHANG, S.-J., ZHENG, K., YE, J., DUNN, B., CZAJKOWSKI, R., MOSER, M.-B., MOSER, E. I., ROUDI, Y., AND WITTER, M. P. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nature neuroscience* (2013).

[5] EDVARDSEN, V. *Modelling Grid Cells and Place Cells using Artificial Neural Networks*. Report, TDT4501 "Computer Science, Specialization Project", NTNU, 2013.

[6] FLOREANO, D., AND MATTIUSSI, C. *Bio-inspired artificial intelligence: theories, methods, and technologies*. The MIT Press, 2008.

[7] FUHS, M. C., AND TOURETZKY, D. S. A spin glass model of path integration in rat medial entorhinal cortex. *The Journal of Neuroscience 26*, 16 (2006), 4266–4276.

[8] GIOCOMO, L. M., MOSER, M.-B., AND MOSER, E. I. Computational models of grid cells. *Neuron 71*, 4 (2011), 589–603.

[9] GRAY, H. *Anatomy of the human body*. Lea & Febiger, 1918. Copyright expired. Figure downloaded from `http://commons.wikimedia.org/wiki/File:Gray739.png`, last accessed 2014-06-11.

[10] HAFTING, T., FYHN, M., MOLDEN, S., MOSER, M.-B., AND MOSER, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature 436*, 7052 (2005), 801–806.

[11] HAFTING, T., FYHN, M., MOLDEN, S., MOSER, M.-B., AND MOSER, E. I. `Hafting_Fig2c_Trial1.mat` from `hafting_et_al_2005.tgz`, the dataset from [10]. Downloaded from `http://www.ntnu.edu/kavli/research/grid-cell-data`, last accessed 2013-12-11.

[12] HAWKINS, J., AND BLAKESLEE, S. *On Intelligence*. St. Martin's Press, 2004.

[13] HEBB, D. O. *The Organization of Behavior*. Wiley, New York, 1949, p. 62.

[14] LEUTGEB, S., AND LEUTGEB, J. K. Remapping to discriminate contexts with hippocampal population codes. In *Space, Time and Memory in the Hippocampal Formation*. Springer, 2014, pp. 227–251.

[15] LEUTGEB, S., LEUTGEB, J. K., BARNES, C. A., MOSER, E. I., MCNAUGHTON, B. L., AND MOSER, M.-B. Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science 309*, 5734 (2005), 619–623.

[16] MCNAUGHTON, B. L., BATTAGLIA, F. P., JENSEN, O., MOSER, E. I., AND MOSER, M.-B. Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience 7*, 8 (2006), 663–678.

[17] O'KEEFE, J., AND DOSTROVSKY, J. The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain research 34*, 1 (1971), 171–175.

[18] PURVES, D., AUGUSTINE, G., FITZPATRICK, D., HALL, W., LAMANTIA, A., MCNAMARA, J., AND WHITE, L. *Neuroscience*, 5th ed. Sinauer Associates, 2012, pp. 703–708.

[19] ROLLS, E. T. A model of the operation of the hippocampus and entorhinal cortex in memory. *International Journal of Neural Systems 6* (1995), 51–70.

[20] ROLLS, E. T. The mechanisms for pattern completion and pattern separation in the hippocampus. *Frontiers in systems neuroscience 7* (2013).

[21] ROLLS, E. T., STRINGER, S. M., AND ELLIOT, T. Entorhinal cortex grid cells can map to hippocampal place cells by competitive learning. *Network: Computation in Neural Systems 17*, 4 (2006), 447–465.

[22] ROLLS, E. T., AND TREVES, A. *Neural networks and brain function*. Oxford university press, 1998.

[23] STENSOLA, H., STENSOLA, T., SOLSTAD, T., FRØLAND, K., MOSER, M.-B., AND MOSER, E. I. The entorhinal grid map is discretized. *Nature 492*, 7427 (2012), 72–78.

[24] TURING, A. M. The chemical basis of morphogenesis. *Bulletin of mathematical biology 52*, 1 (1990), 153–197.

[25] WALPOLE, R. E., MYERS, R. H., MYERS, S. L., AND YE, K. *Probability and statistics for engineers and scientists*, 8 ed. Pearson Prentice Hall, 2007.

[26] WILLS, T. J., LEVER, C., CACUCCI, F., BURGESS, N., AND O'KEEFE, J. Attractor dynamics in the hippocampal representation of the local environment. *Science 308*, 5723 (2005), 873–876.