Memory Capacity in the Hippocampus: Influence of Inhibition and Remapping

Axel Kammerer
Munich, July 20, 2015

Dissertation der
Graduate School of Systemic Neurosciences der
Ludwig-Maximilians-Universität München

Memory Capacity in the Hippocampus: Influence of Inhibition and Remapping

Dissertation der
Graduate School of Systemic Neurosciences der
Ludwig-Maximilians-Universität München

Submitted by

Axel Kammerer

Munich, July 20, 2015

First Reviewer and Supervisor:
Prof. Christian Leibold
Second Reviewer:
Dr. Martin Stemmler

Date of Oral Defense:
June 29, 2015

Contents

List of abbreviations

EC	**E**ntorhinal **C**ortex
MEC	**M**edial **E**ntorhinal **C**ortex
LEC	**L**ateral **E**ntorhinal **C**ortex
HC	**H**ippo**c**ampus
DG	**D**entate **G**yrus
CA1, CA3	**C**orniu **A**mmonis areas 1 and 3
PP	**P**erforant **P**ath
MF	**M**ossy **F**ibres
SC	**S**chaffer **C**ollaterals
SWR	**S**harp **W**ave **R**ipple
LFP	**L**ocal **F**ield **P**otential
LTP	**L**ong **T**erm **P**otentiation
STDP	**S**pike **T**iming **D**ependent **P**lasticity

Summary

Neural assemblies in hippocampus encode positions. During rest, the hippocampus replays sequences of neural activity seen during awake behavior. This replay is linked to memory consolidation and mental exploration of the environment. Recurrent networks can be used to model the replay of sequential activity. Multiple sequences can be stored in the synaptic connections. To achieve a high memory capacity, recurrent networks require a pattern separation mechanism. Such a mechanism is global remapping, observed in place cell populations. A place cell fires at a particular position of an environment and is silent elsewhere. Multiple place cells usually cover an environment with their firing fields. Small changes in the environment or context of a behavioral task can cause global remapping, i.e. profound changes in place cell firing fields. Global remapping causes some cells to cease firing, other silent cells to gain a place field, and other place cells to move their firing field and change their peak firing rate. The effect is strong enough to make global remapping a viable pattern separation mechanism.

We model two mechanisms that improve the memory capacity of recurrent networks. The effect of inhibition on replay in a recurrent network is modeled using binary neurons and binary synapses. A mean field approximation is used to determine the optimal parameters for the inhibitory neuron population. Numerical simulations of the full model were carried out to verify the predictions of the mean field model. A second model analyzes a hypothesized global remapping mechanism, in which grid cell firing is used as feed forward input to place cells. Grid cells have multiple firing fields in the same environment, arranged in a hexagonal grid. Grid cells can be used in a model as feed forward inputs to place cells to

produce place fields. In these grid-to-place cell models, shifts in the grid cell firing patterns cause remapping in the place cell population. We analyze the capacity of such a system to create sets of separated patterns, i.e. how many different spatial codes can be generated. The limiting factor are the synapses connecting grid cells to place cells. To assess their capacity, we produce different place codes in place and grid cell populations, by shuffling place field positions and shifting grid fields of grid cells. Then we use Hebbian learning to increase the synaptic weights between grid and place cells for each set of grid and place code. The capacity limit is reached when synaptic interference makes it impossible to produce a place code with sufficient spatial acuity from grid cell firing. Additionally, it is desired to also maintain the place fields compact, or sparse if seen from a coding standpoint. Of course, as more environments are stored, the sparseness is lost. Interestingly, place cells lose the sparseness of their firing fields much earlier than their spatial acuity.

For the sequence replay model we are able to increase capacity in a simulated recurrent network by including an inhibitory population. We show that even in this more complicated case, capacity is improved. We observe oscillations in the average activity of both excitatory and inhibitory neuron populations. The oscillations get stronger at the capacity limit. In addition, at the capacity limit, rather than observing a sudden failure of replay, we find sequences are replayed transiently for a couple of time steps before failing. Analyzing the remapping model, we find that, as we store more spatial codes in the synapses, first the sparseness of place fields is lost. Only later do we observe a decay in spatial acuity of the code. We found two ways to maintain sparse place fields while achieving a high capacity: inhibition between place cells, and partitioning the place cell population so that learning affects only a small fraction of them in each environment. We present scaling predictions that suggest that hundreds of thousands of spatial codes can be produced by this pattern separation mechanism.

The effect inhibition has on the replay model is two-fold. Capacity is increased, and the graceful transition from full replay to failure allows for higher capacities when using short sequences. Additional mechanisms not explored in this model could be at work to concatenate these short sequences, or could perform more complex

operations on them. The interplay of excitatory and inhibitory populations gives rise to oscillations, which are strongest at the capacity limit. The oscillation draws a picture of how a memory mechanism can cause hippocampal oscillations as observed in experiments. In the remapping model we showed that sparseness of place cell firing is constraining the capacity of this pattern separation mechanism. Grid codes outperform place codes regarding spatial acuity, as shown in Mathis et al. (2012). Our model shows that the grid-to-place transformation is not harnessing the full spatial information from the grid code in order to maintain sparse place fields. This suggests that the two codes are independent, and communication between the areas might be mostly for synchronization. High spatial acuity seems to be a specialization of the grid code, while the place code is more suitable for memory tasks.

In a detailed model of hippocampal replay we show that feedback inhibition can increase the number of sequences that can be replayed. The effect of inhibition on capacity is determined using a meanfield model, and the results are verified with numerical simulations of the full network. Transient replay is found at the capacity limit, accompanied by oscillations that resemble sharp wave ripples in hippocampus. In a second model

Hippocampal replay of neuronal activity is linked to memory consolidation and mental exploration. Furthermore, replay is a potential neural correlate of episodic memory. To model hippocampal sequence replay, recurrent neural networks are used. Memory capacity of such networks is of great interest to determine their biological feasibility. And additionally, any mechanism that improves capacity has explanatory power. We investigate two such mechanisms.

The first mechanism to improve capacity is global, unspecific feedback inhibition for the recurrent network. In a simplified meanfield model we show that capacity is indeed improved.

The second mechanism that increases memory capacity is pattern separation. In the spatial context of hippocampal place cell firing, global remapping is one way to achieve pattern separation. Changes in the environment or context of a task cause global remapping. During global remapping, place cell firing changes in

unpredictable ways: cells shift their place fields, or fully cease firing, and formerly silent cells acquire place fields. Global remapping can be triggered by subtle changes in grid cells that give feed-forward inputs to hippocampal place cells. We investigate the capacity of the underlying synaptic connections, defined as the number of different environments that can be represented at a given spatial acuity. We find two essential conditions to achieve a high capacity and sparse place fields: inhibition between place cells, and partitioning the place cell population so that learning affects only a small fraction of them in each environments. We also find that sparsity of place fields is the constraining factor of the model rather than spatial acuity. Since the hippocampal place code is sparse, we conclude that the hippocampus does not fully harness the spatial information available in the grid code. The two codes of space might thus serve different purposes.

Chapter 1

Introduction

1.1 Patient HM and clinical studies

The importance of the hippocampus for memory was discovered by accident. Some patients that suffered from epilepsy or psychosis and developed resistance to medicative treatment decided to try experimental surgery. In epilepsy, this entailed removal of the hippocampal area. In some patients both sides were removed, or one side was removed and the other destroyed by a stroke or similar natural cause. The bilateral removal of the hippocampal structure had very particular memory deficits that allowed to disambiguate the different memory systems in humans. One of these patients became a focus of research. We will revisit his story.

In 1953, the patient Henry Gustav Molaison (henceforth referred to as HM) had a bilateral removal of his medial temporal lobes to alleviate epileptic seizure. A precise account of the surgery and the ensuing impairments is given in (Scoville and Milner, 1957). HM agreed to such radical surgery after medicative treatment had failed. The lesion included hippocampus (HC), medial entorhinal cortex (MEC), as well as the amygdala. HM's treatment resulted in severe amnesia. His condition did not allow him to form new memories of everyday events (anterograde amnesia). Memories preceding the surgery were affected as well (retrograde amnesia). His perception and intellect, however, were unaffected. To give two striking examples, he could not remember what activity he performed a minute ago, no matter if it was something trivial like a conversation or eating, or a memory task set up by

researches. He could also not remember the doctors he saw almost daily. Each day, he acted as if he was meeting them for the first time.

The fact that only memory was afflicted by the surgery was surprising in itself. Before HM, it was believed that memory is distributed over the cortex, integrated in other systems responsible for intellectual or perceptual tasks (Squire, 2009). Work with HM established the idea that a specific brain structure plays an important role for memory. More precisely, some kinds of memory are dependent on temporal lobe structures.

Over the following years, researchers posed a wide range of memory tasks to HM. These tasks revealed the existence of multiple memory systems. In some tasks HM performed as well as a healthy person, while in other's he had very specific difficulties or could not accomplish them at all. These very differentiated effects of his memory impairment gave a clue about the memory systems connected to the lesioned areas. One task in which he performed well was retaining a sequence of three digits for up to 15 minutes, but only if he stayed focused on it (Squire, 2009). As soon as his attention was diverted, he forgot it, together with the fact he was given a task at all. The retention of the three digits used a memory system that was termed working memory or immediate memory. The fact that his working memory was intact was also the reason he could carry on conversations and understand the memory tasks given to him.

Another task was his ability to learn motor skills. It was tested by drawing a five pointed star (Milner, 1962), with the added difficulty of only seeing the drawing hand in a mirror. Over several days, HM improved and the lines in his drawing became more straight, his movements more fluid until he reached levels comparable to a healthy subject. Successful retention was shown over the following 3 days. In each session, HM had lost all recollection of the previous sessions and assured he had never done this task before.

The memories HM could not retain were termed declarative and it was assumed they depended on the temporal lobe to a large extent. These are further subdivided into episodic knowledge (memories that involve a spatial and temporal context)

and semantic knowledge (general knowledge about the world). The term non-declarative is used to cover all other memory systems, like motor skills, simple conditioning, priming and perceptual learning. Brain areas required for these are the cerebellum, basal ganglia, amygdala and neocortex. The idea to even search for these systems was born when HM showed retention of motor and other skills (Squire, 2009).

The most striking effect of HM's impairment of course was the inability to form new memories of everyday events (anterograde amnesia). In addition, most events that happened in the years before the surgery were forgotten as well (retrograde amnesia). Despite this, older memories were still there. This retrograde amnesia that was graded in time is sometimes described as the first direct evidence of memory consolidation (Lechner et al., 1999; Frankland and Bontempi, 2005): new memories might be initially dependent on the hippocampus, and over time are moved to other areas of the brain. The details of this mechanism, however, are still under debate (Nadel and Moscovitch, 1997; Frankland and Bontempi, 2005; Winocur et al., 2010).

The knowledge that declarative memory depends on the temporal lobe together with the idea of different memory systems and how to test for them in behavioral tasks was successfully replicated in animal models (Squire, 2009) of human memory impairment. Careful analysis of animal models (Mishkin, 1978; Squire and Zola-Morgan, 1991) and further human patients with impairments similar to HM (Scoville and Milner, 1957) indicated the structures relevant for declarative memory to be the hippocampus as well as adjacent, anatomically related areas, including entorhinal, perirhinal, and parahippocampal cortices (Squire and Zola-Morgan, 1991).

The findings in HM and similar patients later converged with an independent line of research focusing on rodent navigation (Squire, 1992). In rodents, the HC is important for navigational tasks as well as memory. These findings will be covered in a later section. Before that, to have a common ground, the next section will review hippocampal anatomy and neural circuits which are common to all mammals.

1.2 Hippocampal anatomy

We will review both the internal circuitry of HC and its connections to other brain areas.

The hippocampus is very similar in rats and in humans, Fig. 1.1. Compared to the rat, the human hippocampus represents a smaller fraction of the cortex. Luckily, the internal connectivity is comparable across mammals (Insausti, 1993).

The connections it makes with other areas, in particular all sensory areas and their processing stages suggest a central role for hippocampal function in memory. It has access to sensory information of all modalities (seeing, hearing, etc.), a prerequisite to form memories that combine all these sensory impressions. HC also interacts with neocortex, where higher functions like working memory in prefrontal cortex rely on the detailed memories of HC. The internal circuitry of HC is of great interest for detailed models of memory mechanisms. In this section, both will be introduced in brief.

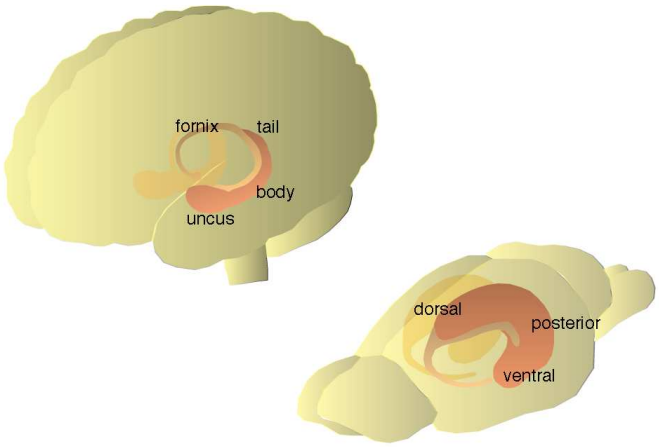

FIGURE 1.1: Position and shape of the hippocampus in humans (left) and rats (right). Adapted from the Scholarpedia page on the hippocampus by Buzsaki.

As can be inferred from HM's and other patients' cases, the connections of HC and neocortex are of particular importance for memory. The hippocampus is strongly connected to all of neocortex, mostly through the entorhinal cortex (Van Hoesen and Pandya, 1975; Swanson and Kohler, 1986; Amaral and Lavenex, 2006). Neocortex contains the sensory areas and prefrontal cortex, which is responsible for

higher cognitive functions. This particular connectivity suits the proposed functions of the hippocampus: the formation of episodic memories requires inputs from all sensory modalities. The extraction of semantic knowledge from several episodic memories (e.g. the fact that fire is hot from several instances of burning yourself) is hypothesized to happen in HC (Eichenbaum, 2004). A transformation of rapidly formed memories from HC to long lasting memories in neocortex seems possible, and would be the basis of memory consolidation (Frankland and Bontempi, 2005). Both the extraction of semantic memory as well as memory consolidation likely require active recall triggered by neocortex, hence the bidirectionality is thought to be important (Eichenbaum, 2000).

The internal circuits of HC themselves are well described (Amaral and Witter, 1989). The anatomy is best understood by looking at a slice of hippocampus that is cut perpendicular to its long and bent dorso-ventral axis, seen in Fig. 1.2a. Comparing these transversal slices from different heights along the dorso-ventral axis, the same anatomical structures can be seen.

Focusing on a transversal slice, Fig. 1.2a, one finds two structures distinguished by a high concentration of neurons. The first is cornu ammonis (CA) or ammon's horn. It contains pyramidal neurons and interneurons, and can be further subdivided into 4 areas CA1 to CA4. Of these, two stand out. The first is CA3, which can be identified by the high recurrent connectivity (between 2% and 4% in the rat, (Amaral et al., 1990)) and the incoming connections from dentate gyrus (mossy fibers). The second is CA1, which has lower recurrent connectivity and receives inputs from CA3 and also EC. CA1 receives mostly feed forward inputs from CA3. In addition, CA1 is easily accessible in electro physiological experiments in rodents, and most data available is from this area. CA3 is likely to contain the mechanisms for memory and association, since recurrent connectivity has proven to be essential in many connectionist models.

The other area with high concentration of neurons is the dentate gyrus (DG). This structure consists of granule cells with tiny cell bodies and strong excitatory connections to CA3 (other brain areas also have cells called granule cells, yet the small cell body is the main common feature they share).

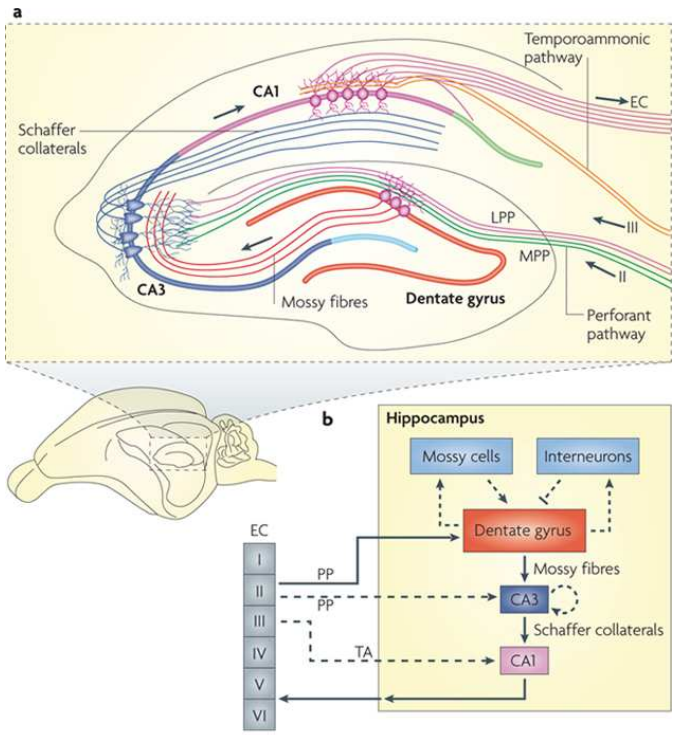

FIGURE 1.2: a) Hippocampal circuitry, including the trisynaptic loop (perforant pathway, mossy fibers, Schaffer collaterals). b) Scheme of the hippocampal neural network. The trisynaptic loop is shown with solid arrows. The perforant pathway (PP) connects layer II neurons in entorhinal cortex (EC) to dentate gyrus (DG). Mossy fibers connect DG granule cells to pyramidal cells in CA3. Schaffer collaterals link pyramidal cells in CA3 to pyramidal cells in CA1. CA1 pyramidal neurons project back to neurons in the deeper layers of the EC. The perforant path also sends projections from EC directly to CA3, skipping DG. Similarly, the temporoammonic pathway (TA) sends information directly to CA1 from EC layer III neurons. Adapted from Deng et al. (2010) with permission.

The connections between EC, DG, CA3 and CA1 form the famous trisynaptic loop, Fig. 1.2b. Starting at EC layer II, signals are passed on to DG and CA3 via the perforant path (PP). Then mossy fibers (MF) project from DG to CA3. The Schaffer collaterals (SC) project from CA3 to CA1. Finally CA1 sends inputs to layers V and VI of EC, closing the loop. The three stages of PP, MF and SC are called trisynaptic loop. In addition, the hippocampal structures of both brain hemispheres are connected.

The interplay of recurrent structures (CA3, EC layers V and II) with purely feed forward structures (DG, CA1) is believed to be the neural basis of effects like pattern completion, pattern separation, and decorrelation of neural code for storage in memory networks.

The similarity of the hippocampal structure across mammals allows to compare animal models with observations in human patients as presented before. The most important results regarding the hippocampal structure in rodents is presented in the next section.

1.3 A cognitive map of space

The hippocampus (HC) and entorhinal cortex (EC) are essential for navigation tasks in rodents. The neurons in HC and EC maintain an internal representation of the animals position, as has been demonstrated by electrophysiological recordings. Place cells and grid cells are the most prominent examples of cells that represent space. The properties of space encoding cells are discussed in this section.

The idea that animals have a map-like representation of space was raised by Tolman (1948) after he performed behavioral experiments that showed their ability to find alternate routes to a goal, or short cuts. This is by no means self evident, and it opposed the common behaviorist view: that complex behaviors like finding short cuts can be explained by a series of sensory-motor responses. A navigational strategy like that would always be centered on the animal, and is called egocentric. For example, a tick navigates based solely on the heat and odor gradients around it, without ever having a full map of the heat and odor distribution of a wider area. The latter would be a map-like representation, detached from the existence of the tick within that map. This is called an allocentric representation of space. The advent of micro wire implantation in freely moving animals made it possible to search for cells that represent space. Such place cells were found in the hippocampus in 1971 (O'Keefe and Dostrovsky, 1971) and they are a neural substrate of the allocentric map-like representation suggested by Tolman. A variety of other cell types encoding aspects of space was found in the wake of this discovery, and in parallel to it.

A milestone in understanding hippocampal function was reached with the discovery of hippocampal cells that encode space (O'Keefe and Dostrovsky, 1971). Such place cells fire only in distinct positions and are found in CA1, CA3 and DG.

Assemblies of place cells recorded simultaneously have been shown to cover the environment with their firing fields. Most place cells in CA have one firing field in a box of 1 square meter, whereas DG place cells have an average number of fields close to two (Leutgeb et al., 2007; Fenton et al., 2008). Another type of space encoding cell was found three decades later (Fyhn et al., 2004). So called grid cells have multiple firing fields, laid out in a triagonal lattice over the environment they encode. Assemblies of grid cells cover the environment with firing fields, and using differently sized triangular lattices, it is possible to encode space unambiguously. This thesis focuses on these two cell types.

There are additional cell types that encode aspects of space. More precisely, there are border cells that fire at certain borders of the enclosure, like one wall of a square box. Head direction cells fire only if the animal's head is turned in a direction specific to the cell and do not depend on position itself. Speed cells' firing rate scales linearly with movement speed of the animal. Yet decoding the position from these cells requires more effort. The grid and place cells are a more refined representation that is invariant to many manipulations of the environment and is easy to decode. And indeed models have been presented that produce place or grid cell firing using these other cell types as inputs. This suggests that more complex tasks like navigating to target locations and finding short cuts relies on the higher level representations of grid and place cells.

Regarding the term navigation, it is important to be aware that all space encoding cells have been observed only in very particular environments. Usually the space encoded in experiments is a linear track of length between one meter up to 18 meters or boxes with square or circular shape with linear dimension of one to three meters. Also mazes of many kinds are used. No observations in natural habitats exist due to experimental limitations. Navigation in large habitats might include entirely different neural processing. For example, an area that spans a square kilometer might be too big to be covered with place fields. This could require a different strategy of navigating between known spots, aided by directional cues from distant landmarks. The grid and place codes might then only be used for decisions on crossing points or for more detailed tasks at the target location, like finding a stash of food hidden away earlier.

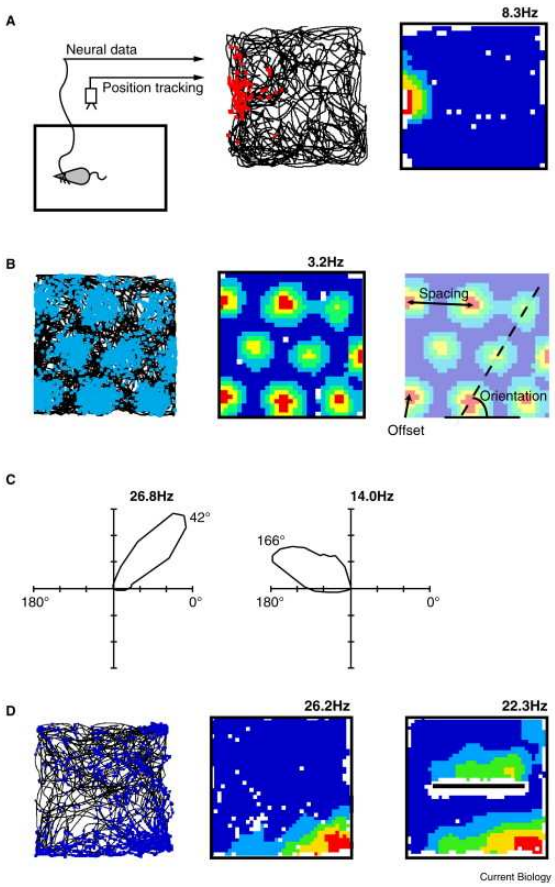

FIGURE 1.3: Neural representations of self-location in the hippocampal formation. (A) Left, setup of single unit recording. A rodent forages in an open environment while extracellular electrodes record neural activity. An overhead camera records the trajectory. Middle, raw data for a place cell. Shown are place cell spikes (red dots) overlaid on rodent trajectory (black line). Right, a firing rate map of the raw data; maximum firing rate depicted in red, its numerical value above the map. White bins are unvisited. The place cell shown is only active when the animal occupies a small area on the left side of the environment. (B) Raw data (left) and firing rate map (middle) for a grid cell in medial entorhinal cortex (MEC). Firing fields are laid out in a hexagonal lattice. Right, the grid-like firing pattern is mainly characterized by its orientation, spacing, and offset (or phase). (C) Two head direction cells recorded from the deep layers of MEC. Head direction cells are also found in rat's dorsal presubiculum, mammillary nuclei, anterior thalamus, and retrosplenial cortex. The polar plots show firing rate as a function of head direction. Both cells show a preferred direction. When the animal's head points in this direction, the cell fires strongest. (D) A boundary vector cell in the subiculum, showing the raw data (left) and firing rate map (middle). Boundary vector cells fire whenever there is an environmental boundary a short distance away in a preferred direction. Placing an additional wall triggers firing (right). Adapted from Barry and Burgess (2014) with permission.

In the context of this thesis, I focus on place cells and grid cells and the interaction of their spatial codes. The other cell types are believed to be used by the mechanisms that produce place cell and grid cell firing fields. In the following, all cell types will be outlined.

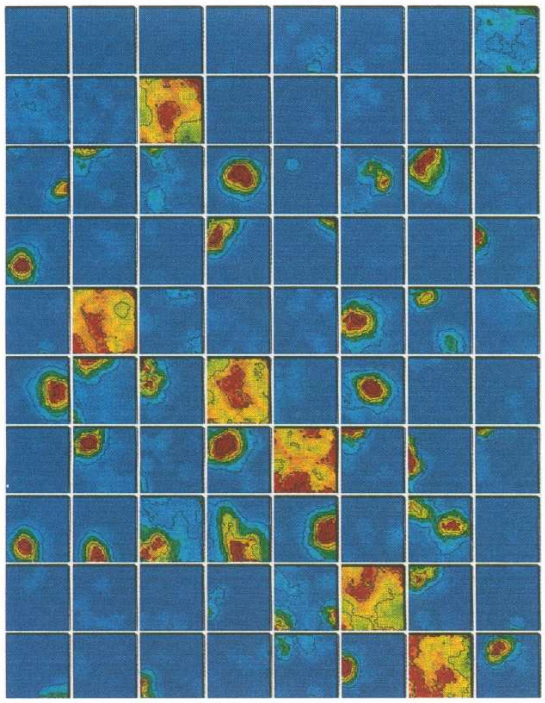

FIGURE 1.4: Spatial firing distributions of 80 simultaneously recorded pyramidal and inhibitory neurons in the hippocampal formation of a rat. The rat freely explored a square box. Each panel represents the firing rate distribution over the square shaped environment for one cell. Red color indicates maximal firing rates, but only for cells that fire significantly related to space. No firing is indicated by dark blue. Inhibitory cells lack defined place fields, they fire over the whole environment while still having spatial modulation. Adapted from Wilson and McNaughton (1993) with permission.

1.3.1 Place cells

O'Keefe and Dostrovsky (1971) found cells that encode space in the hippocampal area of rats. The cells fired only when the animal was at a particular position, Fig. 1.3a, coining the name place cells. Further work (O'Keefe, 1976) showed that the place specific firing was very robust. It is not caused by a simple sensory stimulus, e.g. it persisted when lights were turned off (independent of visual stimuli) or when the arm of the maze was exchanged with an unused arm changing olfactory stimuli. In addition, the cells did not react to direct sensory stimulation, like shining a light at the animal. The experiments established that place cells encode an abstract notion of space, decoupled from primary sensory input. Place cells have been found in dentate gyrus (DG), hippocampal areas CA1 and CA3, and in subiculum (O'Keefe, 1979; Kim et al., 2012).

If enough place cells are recorded they usually cover the whole environment,

Fig.1.4, (O'Keefe, 1976; Wilson and McNaughton, 1993). Indeed, from activity of those cell assemblies over time one is able to decode the rat's trajectory (Wilson and McNaughton, 1993). These two facts, the encoding of abstract space rather than sensory stimuli that coincide with a location, and the coverage of an environment with place fields, justify calling the place cell assemblies a "cognitive map of space" (O'Keefe and Nadel, 1978). In particular, it is an allocentric one, in contrast to an egocentric representation in which the position of the animal would be central. The neural substrate of the cognitive map postulated by Tolman (Tolman, 1948) had been found.

In addition to being a cognitive map, the rat hippocampus, like in humans, plays a role in memory (Lever et al., 2002; O'Keefe and Conway, 1978; O'Keefe and Speakman, 1987). In particular it was found that the same place cell assemblies that were active during awake behavior were active during sleep (Wilson and McNaughton, 1994; Skaggs and McNaughton, 1996; Lee and Wilson, 2002). When decoding from these cells, the same trajectory as in awake behavior was obtained, but replayed on a much faster scale. These replays coincide with so called hippocampal sharp wave ripples (SWR), synchronized bursts of pyramidal cells which happen during slow wave sleep or moments of rest while awake (Buzsaki et al., 1983). An example of replay happening during awake SWR is shown in Fig. 1.5. Sharp wave ripples are thought to be capable of potentiating downstream synapses and are a candidate mechanism for the transfer of information to neocortex, i.e. memory consolidation (Buzsaki, 1989; Nakashiba et al., 2009). Replay has also been observed in reverse order (Foster and Wilson, 2006; Diba and Buzsaki, 2007). Animals are also able to replay trajectories that are relevant to a decision: while the animal is at a junction of a maze it may replay the possible trajectories including the one leading to the reward (Johnson and Redish, 2007). This shows that HC is not only representing space, but also carries out active computations using the cognitive map. A third finding about replays challenges the idea that they have anything to do with memory: (Dragoi and Tonegawa, 2011) found that replay events happen even before the animal visits the relevant part of a linear track. It is hence called preplay, and it is yet to be determined why this happens. Either it is a consequence of fixed hippocampal circuitry, meaning that sequences

of assembly firing are pre-wired in HC and external events are only mapped (or associated) to their firing by a learning process, or the sequences are learned but similar to existing ones, and HC fits them into an existing schema for accelerated learning (the preplay being the existing schema) (Tse et al., 2007).

All these cases have one thing in common, the existence of sequentially firing cell assemblies in HC. Several studies suggest that replay in HC is important for memory consolidation as well as mental exploration (Buhry et al., 2011). Concerning the underlying mechanism, the CA3 region is a prime candidate that could produce the sequences due to its recurrent connectivity. From a theoretical perspective, it is interesting to construct a network that produces such sequences, and to test how many can be stored and successfully replayed.

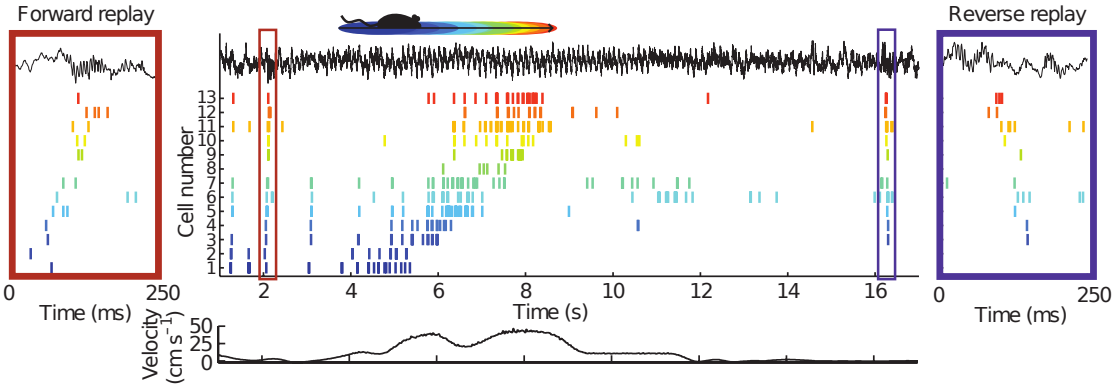

FIGURE 1.5: Place cell activity while a rat runs on a linear track is replayed during sharp wave ripples (SWRs) when the rat rests. Spike trains for 13 place cells are shown. The actual traversal of the linear track is marked by the times with speed significantly above zero (bottom graph) as well as by theta activity in the LFP signal (black line above spike trains). Place fields are visible in the spike trains during traversal and the cells have been ordered according to place field position. The red and blue box mark awake SWRs. Both before and after traversal, awake SWRs are accompanied by place cells replaying sequences seen during running in the same order or reverse order. Forward replay (left inset, red box) occurs before traversal of the environment and reverse replay (right inset, blue box) after. Adapted from Diba and Buzsaki (2007) and Carr et al. (2011) with permission.

1.3.2 Grid cells

After the discovery of place cells, more space encoding cells were found in medial entorhinal cortex (MEC). Recordings in layers II and III of MEC (Fyhn et al., 2004; Hafting et al., 2005) showed cells with a very particular spatial tuning: multiple sharp firing fields existed, and they were lying on a triangular lattice or grid. The

grid proofed to be a robust phenomenon, appearing in environments of all shapes and sizes. Grids are almost fully described by three parameters (Hafting et al., 2005), as depicted in Fig. 1.3b:

1. The spatial period, or distance between vertices of the triangular lattice.
2. The orientation or angle of the grid.
3. The phase of the grid.

The width of the fields is not a parameter of its own, since it scales with spatial period (Hafting et al., 2005; Brun et al., 2008). Spatial period and orientation is mostly shared by nearby grid cells. Usually most grid cells recorded from a single tetrode have the same periodicity. It also changes along the dorso-ventral axis of MEC, showing larger periods the more ventral the recording site lies. Smallest periods recorded were around 20cm, largest around 10m (Brun et al., 2008; Fyhn et al., 2004; Hafting et al., 2005; Stensola et al., 2012). More recently it has been shown that this change in scale is not continuous along the dorso-ventral axis, but happens stepwise in a roughly geometric progression (Stensola et al., 2012). These subgroups of grid cells with different periodicity are called modules, and it is estimated that a rat has 5 to 9 of them. Cells in a module also share the same orientation. The one parameter not shard in a module is the phase. It is distributed over space, allowing coverage of the whole environment by firing fields. It has been shown that this setup, in particular due to the geometric progression, optimally encodes space (Mathis et al., 2012). Optimal in this case means that decoding position from grid cell spikes is possible with high acuity using as few cells as possible. The decoding scheme used is called “hierarchical”, since spatial periods significantly differ in size and the progression from large to small is used to remove ambiguities. An alternative coding scheme is based on the Chinese remainder theorem (Fiete et al., 2008). The periods are close to each other but are incommensurable, allowing unambiguous encoding of vast spaces. Yet this decoding scheme requires the spatial periods to be very precise over the whole space.

1.3.3 Head direction cells, conjunctive cells and speed cells

Head direction cells have a maximum in firing rate when the animal's head is pointing in a certain direction in the ground plane (yaw angle). This direction depends on the head and environments alone, independent of the body's orientation. They were originally discovered in the rat's dorsal presubiculum, mammillary nuclei, anterior thalamus, and retrosplenial cortex (Taube et al., 1990a;b). More recently they have been found in MEC (Sargolini et al., 2006). It is speculated that they are used for path integration, together with velocity signals.

Cells that combine the tuning to head direction and the grid cell firing fields are called conjunctive cells (Sargolini et al., 2006). These cells have hexagonal firing fields that only fire for certain head-directions.

While grid cells, head direction cells, and conjunctive cells are all modulated by speed (Sargolini et al., 2006), it was recently reported that some cells in MEC explicitly encode velocity (Kropff Causa et al.). Together with head direction cells, this is the most simple input to create a neural representation of space.

Of all the space encoding cells, place cells in CA3 have a special importance. Before knowing of the existence of place cells in CA3, the area was already hypothesized to be central to memory functions due to its recurrent connectivity. The finding of place cells in this area showed that not only sensory representations are available to CA3, but also a cognitive map of space. This was followed by the observation that the spatial map changes drastically based on small changes in the environment or behavioral context, which greatly helps most of the suggested memory models to perform better. This phenomenon of changing the cognitive map of space is called remapping. The next section will give an overview on experimental findings related to remapping.

1.4 Remapping

Remapping describes the observation that small changes in the environment or behavioral context can drastically change the place cell firing fields. Rate remapping refers to the firing rates of individual place fields changing, and for cells with several fields this means they can change independently. Global remapping causes a switch of the space encoding cell population: some place cells cease firing while formerly silent cells become place cells; place cells that do not cease move their firing fields and change firing rate in an unpredictable manner. Here, the detailed situations that trigger both types of remappings are described. The case of global remapping is of particular interest to us, since it directly benefits memory models. The behavior of entorhinal grid cells during global remapping is described at the end of this section, since it is the basis of a hypothesized mechanism of global remapping in hippocampus.

Changes in the environment can trigger changes in place cell firing: Putting a wall in the location of a firing field can abolish that field, while putting only the lead base that held it leaves it intact (Muller and Kubie, 1987). Some place cells show a firing field where a new object was placed, or a known one removed, while other cells are independent of the same objects placed or removed in their firing field (O'Keefe, 1976). Experiments like this give clues about what place cells exactly encode and how they do it.

While many of these effects can be subtle and intriguing, we will particularly focus on two phenomena: rate remapping and global remapping. They are of special interest since they usually encompass all or many recorded cells, hence they have implications for the HC as a memory structure. During rate remapping (Fig. 1.6 left panel) place field positions stay the same, but the firing rates of individual place fields can change (Leutgeb et al., 2005). This might allow to encode space by position of place fields, and behavioral context or other things in the rates. The fact that place fields of the same cell can adjust rate differently (e.g. one can increase, the other decrease), allows a huge space of representations for contexts of positions. An even more radical change occurs during global remapping (Fig. 1.6 right panel), a complete change in almost all recorded place cells (Leutgeb et al.,

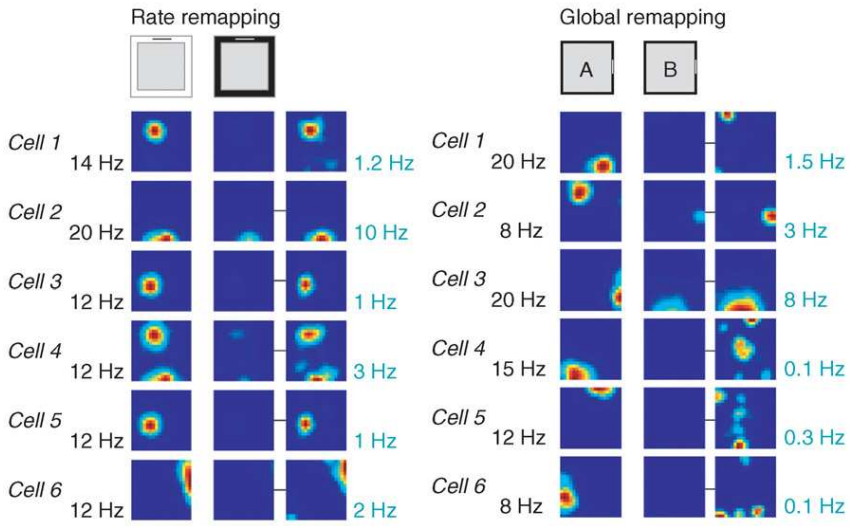

FIGURE 1.6: Rate remapping and global remapping in six CA3 place cells (firing rate maps show dark blue for 0 Hz and red for maximum firing rate as indicated left and right of each row). Left, rate remapping is observed when rats explore two environments with different color configuration but identical distal cues (same position of black and white box within the laboratory). Right, global remapping is triggered when the distal cues change while the box stays the same (box is moved to a different room). In both panels, the left column shows the ratemap of the cell in the condition that caused the highest peak firing rate (either white or black color for the left panel, or room A or B for right panel). The middle column shows the ratemep of the other condition that caused lower peak firing, using the same firing rate scale as in the first column. The final column shows the same data as the middle column, but scaled so that red corresponds to their own maximum firing rates. For rate remapping, the locations of place fields stayed identical, but firing rates changed. For global remapping, both field locations and firing rates changed. Adapted from Colgin et al. (2008) and Leutgeb et al. (2005) with permission.

2005). Some move their place fields to new positions, others cease firing altogether, and some formerly silent cells start showing place fields.

Both are triggered by rather small changes in environments. Furthermore, the environments are usually very well known to the animal. A systematic study of remapping started was first done by Leutgeb et al. (2005). A comparison between the most renowned studies can be found in (Colgin et al., 2008). In one of the experiments, the animal was foraging for randomly dropped food within a square box with white walls and a black cue card on a wall. Rate remapping could be induced by changing the colors to black walls and white cue card. It was crucial to position the box in the same location of the room, to ensure distal cues are identical. Global remapping was induced by changing the location of the box and keeping the same colors. Another way to induce rate remapping was to change the shape of the box from rectangular to circular (Fig. 1.6).

Comparison between multiple studies (Colgin et al., 2008) shows that remapping greatly depends on the training history of the animals. For example, a switch from rectangular environment to circular surrounding in the same location can induce both rate remapping (Leutgeb et al., 2005) or global remapping (Fyhn et al., 2007). One interesting case shows that global remapping can often be obtained between a rectangular and circular environment in the same location if a pre-training occurred in which the round and square enclosures were in different locations in the room. The final sessions had them tested in a new room in one location and showed global remapping (see personal communication of R.U. Muller and J.L. Kubie in (Colgin et al., 2008)). The line between rate and global remapping is truly a thin one: slight changes like the material of the boxes, or even the food used during training in the boxes, can make the difference (Colgin et al., 2008)).

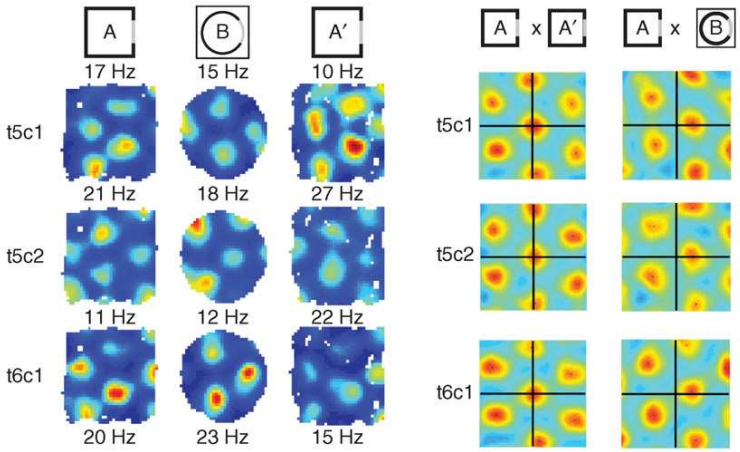

FIGURE 1.7: Remapping and grid realignment in medial entorhinal cortex (MEC). Rate maps (left) of MEC grid cells recorded in square and circle enclosures with identical distal cues. Changing the enclosure simultaneously causes global remapping in place cells (not shown) and the shown changes in grid cell firing patterns. Grid cells show coherent realignment (the same shift in phase for all cells) between the square and circle enclosures, as visualized by the spatial cross-correlations (Pearson product-moment correlations) seen on the right for the same cells. Adapted from Colgin et al. (2008) and Fyhn et al. (2007) with permission.

In the first experiments it seemed surprising that minor changes can cause an absolute change in the spatial fields (Muller and Kubie, 1987; Colgin et al., 2008). The training phase explains why subtle changes can trigger such a large change in coding cell populations: the animal has time to differentiate the codes. Now if the only purpose of the code was to represent space, this would not be necessary.

It is hence conceivable that remapping serves the memory function of HC. The completely different codes that are produced even for similar situations help to reduce memory interference.

Further insight into remapping was gained when grid cells and place cells were observed simultaneously in several remapping paradigms (Fyhn et al., 2007). During rate remapping in HC, grid cell firing patterns remain virtually unchanged. For the cases in which global remapping occurs in HC, the grid firing fields in MEC either change in phase and keep their orientation (Fig. 1.7), or change both phase and orientation. All observed cells undergo the same change: the same translation and rotation is applied to all firing fields. Yet only cells close to each other were measured, and it is believed that cells in other modules along the dorso ventral axis undergo different translations and rotations. This assumption leads to a simple mechanism of hippocampal global remapping (Monaco et al., 2011). Another model explains the observed rate remapping by adding LEC inputs (Renno-Costa et al., 2010).

The next section introduces the hypothesis of two stages of memory encoding seen in the hippocampus. One stage inscribes memories acquired during behavior, the second consolidates them during rest. Both remapping and the two stage model are considered essential for memory.

1.5 The two stage model of memory

The hippocampus has two modes of operation, easily differentiated by the local field potential. One stage incorporates sensory information in a neural substrate and another stage shuts down inputs to let the neuronal networks process or adjust to the new information. The switch between states is triggered by release of neuromodulators like acetylcholine (Giocomo et al., 2007; Hasselmo and Bower, 1993; Hasselmo et al., 1992). A separation in two-stages is hypothesized to be of great importance in several brain areas (Buzsaki, 1996).

In hippocampus, the first stage is defined by theta activity, an oscillation at 4–12 Hertz. Theta activity is interwoven with gamma oscillations (at 30-130 Hz).

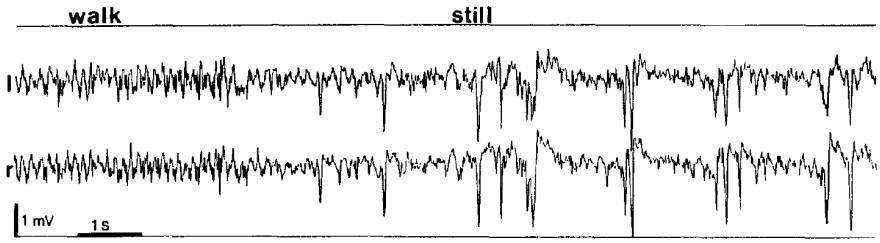

FIGURE 1.8: Local field potential recorded in the stratum radiatum of the left (1) and right (r) CA1 region of the hippocampus during walk-immobility (still) transition. During walking, regular theta waves can be observed. When the animal is immobile, large monophasic sharp wave ripples (SWRs) replace the theta activity. SWRs are bilaterally synchronous. Adapted from Buzsaki (1989) with permission.

Gamma rhythms are the result of fast inhibition and are also observed in other brain areas (Buzsaki and Wang, 2012). Neurons firing together in one gamma cycle are believed to be cell assemblies encoding information (Harris et al., 2003). The second stage is marked by sharp wave ripples (SWR). SWRs are a compact wave shape modulated by high frequency ripples (at 140–200 Hz). Theta oscillations usually happen during exploratory movement and REM sleep, whereas sharp wave ripples occur during immobility, consummatory behaviors, and slow-wave sleep, see Fig. 1.8.

Theta oscillations and sharp wave ripples strongly influence information flow between neocortex and HC. Theta activity favors information transfer from neocortex to HC (Buzsaki, 1989), and sharp wave ripples encourage transfer from HC to neocortex (Chrobak and Buzsaki, 1996). The correspondence of theta oscillation with memory encoding and the correspondence of sharp wave ripples with memory consolidation is in agreement with many observations.

To understand how the two stage model and remapping help memory, the next section will introduce several classes of models for memory and grid and place field formation.

1.6 Models of the hippocampal area

Two types of models of the hippocampal area are important for this work: models of memory and models of navigational aspects. We will shortly introduce the

concepts of memory networks, both attractor networks that can hold a single memory state over time as well as associative networks that can loop through a sequence of memories. Regarding navigation, there are three classes of model that are relevant to place cells, grid cells and remapping. First, generative models suggest mechanisms that produce place fields and grid fields from lower level signals like head direction cells, border cells, as well as recurrent network connectivity. Other models explore how place cell firing can emerge from grid cell firing, and vice versa. And finally we will look at models of remapping. The most relevant one for us relies on simple feed forward connections from grid to place cells. Finally, the models of remapping will be discussed.

1.6.1 Memory networks

The first concept to understand is what constitutes a memory in most network models: objects, people and facts are represented by an assembly of neurons that activate at the same time. For example, a dear person is represented by one assembly, the family dog by another. When a neural network is able to keep those assemblies firing, it is considered memory (working memory in this case, as it keeps a representation available). Alternatively, associative memory is realized by giving the network a cue representation, and having it respond with an associated representation.

In attractor networks (Hopfield, 1982), assemblies of neurons can be activated externally (by sensory cues for example) and keep firing after the external input ceased. This ability to sustain firing is one of the proposed mechanisms for memory. The system sustains the firing, which in turn can be accessed by other brain areas without sensory input.

An additional benefit of attractor networks is pattern completion. Incomplete external cues are automatically completed by the system. Then firing of the full assembly is sustained normally. In the visual system, this might be the mechanism that allows us to identify an object we do not fully see. Like a tiger hiding behind vegetation.

The basis of this system lies in potentiation of the right synapses. When an external cue causes an assembly of neurons to fire, the synapses between them are strengthened. After enough repetitions (learning phase), the synapses are so strong that the firing neurons can sustain each others activity after the external input stopped. They become each other's input. This also explains pattern completion. If some neurons in the assembly are not firing, they will be activated by the many other neurons of the assembly. There is of course a critical threshold of neurons necessary. Not all patterns can be completed.

The name attractor network stems from the pattern completion. A marble on a plane with some depressions will stay in the depressions. Similarly, the memory system will stay in the memory states. And if it is in a state close to a memory state, it will converge to that memory. Just like the marble rolling to the bottom of a valley.

The mechanism of pattern completion is also what ultimately limits the network. So far we neglected the interaction of multiple memories and their neuron assemblies. If two memory assemblies share too many neurons, they will activate neurons from both assemblies, and form a new attractor that does not represent any external cue learned. The two memories are said to interfere. If enough memories are stored, there will be such overlaps. At some point, the network fails in a catastrophic way. Failure happens at a critical number of memories, called the capacity of the network. Beyond that capacity, all memories are no longer accessible, and either lead to activation of the whole network, or spurious attractor states that do not represent a memory. This sudden transition is called catastrophic forgetting. It would be more desirable to have graceful failure instead, meaning that memories are still accessible, but carry more noise. The concepts of attractor states, pattern completion, and catastrophic forgetting are important beyond this specific model. Many recurrent networks can show these phenomena.

To avoid interference and store as many memories as possible, all memory assemblies should have as little overlap as possible. There are two suggested mechanisms in the mammalian brain. The first is the dentate gyrus which possesses sparse and powerful projections to CA3 pyramidal cells (Treves et al., 2008). During learning,

the DG inputs by themselves can trigger CA3 cells to fire, and thereby strengthen their recurrent connections during learning. During recall, the weaker inputs from layer II of EC can reactivate assemblies. This allows DG to select sparse cell assemblies with little chance of overlapping with existing memory assemblies. The precise mechanism of selecting assemblies is not known. Nevertheless, even if DG does nothing else but selecting random sparse assemblies, this would be a powerful mechanism. In addition, DG does not convey new information to CA3 since both CA3 and DG receive the same inputs from layer II of entorhinal cortex. All together, this suggests DG acts as a decorrelator of neural assemblies during learning. A second mechanism that decorrelates assemblies takes advantage of remapping in entorhinal grid cells (Colgin et al., 2008; Fyhn et al., 2007). Grid cells project onto CA3 pyramidal cells. These grid-to-place models produce place cell firing fields from grid cell activity. Accordingly, the rather simple remapping in grid cells, as explained earlier, causes global remapping of place cells, much like it is observed in experiments (Monaco et al., 2011). And this global remapping in place cells is a pattern separation mechanism, since the memory of a certain place will be encoded using very distinct cell assemblies in CA3 for each remapping.

A memory network that cycles through a sequence of memories can be constructed by the same principles. In the most simple case, one assembly is activated as a cue and triggers a target assembly (Willshaw et al., 1969; Marr, 1971). This allows to associate memories with one another. To achieve this, the synapses are strengthened between neurons of the cue and neurons of the target assembly. Like this, activity flows from one assembly to the next. An associative network also has the property of pattern completion, albeit with reduced efficiency since it has to happen within a single time step. It is possible to create a sequence of memories in this way. Such a network will follow the chain until it reaches the end of the sequence and firing ceases (Leibold and Kempter, 2006). No steady state exists, since the sequence does not allow to dwell in one state. Also, an item may never repeat, or it would lead to failure since no disambiguation is possible. The viability of such a network depending on biological parameters can be found in (Leibold and Kempter, 2006). Its optimization in terms of memory capacity by adding inhibition is explored in this thesis.

One important class of error when recalling memories is interference. Memories are said to interfere with each other if the assemblies representing them share synapses, resulting in undesired activation of neurons. The worst case example is a network loaded with so many memories that virtually all assemblies are connected. For the associative network, this results in a state where all neurons fire or are silent, depending on how the threshold/inhibition was tuned. One strategy to reduce interference is to use sparse coding, e.g. few active neurons in each assembly. Additionally, it is desirable to have a mechanism that ensures that even similar memories can be stored using very distinct assemblies, a principle called pattern separation. An example of two similar memories that should be stored as very different assemblies is seeing a tiger in a zoo compared to seeing a tiger in your garden. The sensory inputs are very similar, but the context makes all the difference. Remapping could be a mechanism for separate the two patterns. An additional input, like fear, could trigger a recoding of the memory, which we observe as global remapping of place cells in experiments.

The requirement of memory networks to have pattern separation also has measurable predictions. Remapping in HC cells should be very unpredictable. This is found to be true. In particular, CA3 is more unpredictable than CA1 during remapping (Leutgeb et al., 2004). This is possibly due to a higher recurrent connectivity in CA3 (Amaral and Witter, 1989), which would cause more interference and hence require better orthogonalization of the neuronal assemblies used for memories.

1.6.2 Place cell and grid cell models from lower level cells

Here we review models that produce place and grid cell firing based on inputs from lower level cells like head direction cells, speed cells, and border cells.

One class of model uses continuous attractor dynamics to maintain an internal representation of position (Samsonovich and McNaughton, 1997; Redish and Touretzky, 1998; Tsodyks, 1999), and inputs from head direction cells and velocity dependent inputs to update this representation.

A continuous attractor network resembles the Hopfield memory network described earlier. There, we had discrete attractors, comparable to a marble on a plane with some depressions. In the continuous attractor model, we have grooves that make pathways, and the marble can move freely in a groove, but cannot leave it. With neurons, such a groove or a continuous attractor can be built by arranging all neurons on a square sheet, then connecting neighboring neurons with strong excitation, while far away neurons inhibit each other. Each neuron has equivalent connections to all others, giving the system translational symmetry on the sheet. The system can be tuned so that a random external input causes one neuron and its neighbors to win out: they fire while inhibiting every other neuron. This forms a bump of activity on the cell sheet. This bump stays in place, until head direction and velocity inputs move it around. It can move around due to the translational symmetry of synaptic weights. It cannot lose its shape since the mechanism of pattern completion would correct any deviation. We fooled the system into accepting translations, but no other change. Coming back to our analogy, the marble staying in the groove is equivalent to the bump keeping its shape, and moving inside the groove is equivalent to moving the bump on the cell sheet.

If we look at the activity of a single cell, it is not trivial that it will result in a typical place cell firing map. To make the bump on the cell sheet a near perfect representation of the position in space, velocity, and direction input have to work precisely when moving the bump. Luckily, this works out fine.

Tuning the system in a different way, a hexagonal pattern of bumps can be created (Fuhs and Touretzky, 2006; McNaughton et al., 2006; Burak and Fiete, 2009). Now each cell will become active at multiple spots in the environment. The hexagonal pattern in the cell sheet is transferred to the firing map of individual cells by moving the bumps.

While it is beautiful to achieve the hexagonal pattern as an emergent phenomenon, one can criticize that the translational symmetry in the connections is hard to achieve and upkeep.

Grid cell firing has also been modeled as the result of interfering oscillations (Burgess et al., 2007). To do so, a base oscillation interferes with an oscillation whose frequency depends on velocity. In one dimension, the velocity dependence allows for a repeating pattern with fixed spatial period. In two dimensions, it yields a stripe pattern, since the velocity vector needs to be projected onto an arbitrary direction to make the model essentially one dimensional again. This is then repeated using a total of three different directions, each producing striped firing maps. These stripe cells are angled at 60 degrees, and are used as inputs for a grid cell. The resulting pattern is a hexagonal grid.

1.6.3 Place cells from grid cells and vice versa

While both grid cell and place cell firing can be obtained from the lower level cells, they can also be obtained from each other. A very simple model to produce place cell firing maps from grid cell activity is to take feed forward connections from grid to place cells. These connections can be set in different ways. Hebbian learning, random connections, as well as optimization to yield sparse place cell firing and other methods have been shown to produce similar outcomes (Cheng and Frank, 2011).

The other direction is more difficult. To produce grid firing fields from place cells, two things are important (Kropff and Treves, 2008). First, grid cells need a firing rate adaptation mechanism, which will reduce their excitability if they fire continuously. Second, the mean grid cell activity has to be kept constant. This can be achieved by inhibitory feedback between grid cells. This allows competition between grid cells, resulting in few cells being active in the same locations, and more different phases that cover the environment. The inputs for the grid cells are place cells with feed forward connections, and Hebbian learning is used on these connections while the animal moves on a random trajectory. At one point in space, some place cells fire and trigger the firing of a grid cell, that by coincidence has a higher initial weight. This grid cell, in turn, inhibits the grid cells with weaker weights. As the rat moves, the grid cell tires due to adaptation, and other grid cells win the competition. Moving far enough, the grid cell recovers and might

start firing again. When the rat comes back to the same spot, the same grid cell will win out, due to the strengthened connections from the active place cells. But as it moves away, no matter what direction, the grid cell tires and stops firing. Like this, there will always be a ring shaped zone where the cell is silent. A dense packing of such firing fields with ring shaped zones of inactivity yields a hexagonal pattern.

Both types of models can function with imperfect inputs (Azizi et al., 2014; Kropff and Treves, 2008). In particular, the place-to-grid model requires only spatially modulated input, as can be found in other areas like LEC.

1.6.4 Models of remapping

The feed forward transformation of grid to place cell firing offers an easy way of remapping. By shifting or rotating the grid firing fields, completely new place cell firing fields emerge (Monaco et al., 2011). It was shown that the neural assemblies encoding locations (population vectors) are decorrelated by this procedure (Monaco et al., 2011). In particular, applying as little as four different shifts to sub groups of grid cells (the discrete modules that share a spatial period (Stensola et al., 2012)) results in a remapping that is as random as if shifts for individual cells had been used. This means that the very different remappings in HC could have a simple control mechanism that uses few parameters (shifts and rotations of 4 to 10 grid cell modules) and can address the HC remappings in a consistent way (Fyhn et al., 2007; Colgin et al., 2008).

Remapping in CA3 can also be modeled using a single recurrent network of CA3 cells (Samsonovich and McNaughton, 1997; Monasson and Rosay, 2013; 2014; Stringer et al., 2002). Each remapping is learned as a single continuous attractor. Within it, activity can flow smoothly between neuron assemblies to represent spatial position. To remap, the system jumps to another continuous attractor, in which position is represented by other assemblies in a continuous way. Using this combination of multiple continuous attractors, it is easy for external inputs to update position, and more difficult to switch between remappings.

Chapter 2

Inhibition Enhances Memory Capacity: Optimal Feedback, Transient Replay and Oscillations

Contributions

The authors Axel Kammerer (AK) and Álvaro Tejero (AT) contributed equally to this paper. The work was done under the supervision of Christian Leibold (CL); AK, AT and CL conceived and designed the research. AK, AT and CL derived the results for the mean field model. AK wrote the numerical code and carried out the simulations for the verification of mean-field results with cellular simulations. AT established the Bayes-optimal threshold derivation. AK, AT and CL discussed the results and wrote the paper.

The paper has been published in *Journal of Computational Neuroscience* Kammerer et al. (2013).

J Comput Neurosci (2013) 34:125–136
DOI 10.1007/s10827-012-0410-z

Inhibition enhances memory capacity: optimal feedback, transient replay and oscillations

Axel Kammerer · Álvaro Tejero-Cantero · Christian Leibold

Received: 17 February 2012 / Revised: 18 May 2012 / Accepted: 21 June 2012 / Published online: 11 July 2012

Abstract Recurring sequences of neuronal activation in the hippocampus are a candidate for a neurophysiological correlate of episodic memory. Here, we discuss a mean-field theory for such spike sequences in phase space and show how they become unstable when the neuronal network operates at maximum memory capacity. We find that inhibitory feedback rescues replay of the sequences, giving rise to oscillations and thereby enhancing the network's capacity. We further argue that transient sequences in an overloaded network with feedback inhibition may provide a mechanistic picture of memory-related neuronal activity during hippocampal sharp-wave ripple complexes.

Keywords Associative memory · Sequence memory · Memory capacity · Sequence replay · Inhibition · Neuronal network dynamics · Learning and plasticity · Sharp-wave ripples · Oscillations

Action Editor: David Golomb

Axel Kammerer and Álvaro Tejero-Cantero contributed equally to this paper.

A. Kammerer (✉) · Á. Tejero-Cantero · C. Leibold
Department Biologie II, Ludwig-Maximilians University Munich, Großhadernerstrasse 2, 82152 Planegg, Germany
e-mail: kammerer@bio.lmu.de

A. Kammerer · Á. Tejero-Cantero · C. Leibold
Graduate School for Systemic Neurosciences, Ludwig-Maximilians-University Munich, Großhaderner Str. 2, 82152 Planegg, Germany

Á. Tejero-Cantero · C. Leibold
Bernstein Center for Computational Neuroscience Munich, Großhaderner Str. 2, 82152 Planegg, Germany

1 Introduction

The hippocampus is a brain structure crucially involved in the formation of autobiographic, episodic memories. Electrophysiological recordings of hippocampal neurons in behaving rodents have revealed the existence of *place cells*, which are active at only a few particular locations in a known environment, and silent elsewhere (O'Keefe and Dostrovsky 1971). The response of a population of place cells hence encodes the position of an animal in an environment and a spatial trajectory of this animal is represented by a sequence of active place cells (e.g. Dragoi and Buzsáki 2006; Davidson et al. 2009).

During sleep and resting states, spontaneous hippocampal activity bursts have been observed in which neurons were activated in an order similar to exploratory phases (Wilson and McNaughton 1994; Nádasdy et al. 1999; Lee et al. 2002; Diba and Buzsáki 2007; Gupta et al. 2010; Dragoi and Tonegawa 2011). The activity sequences observed in slow-wave sleep mostly correspond to previously experienced trajectories, whereas under awake rest, they have been found to even code for future trajectories. Therefore, these network bursts have been hypothesized to reflect memories and imaginations of spatial episodes (for review, see Buhry et al. 2011).

The activity sequences are correlated with the occurrence of sharp-wave ripple complexes (Lee and Wilson 2002; Diba and Buzsáki 2007), a brief (∼100 ms) field potential deflection that is superimposed with a high-frequency oscillation in the hippocampal CA1 pyramidal cell layer (∼200 Hz; Buzsáki et al. 1992). Although it is still debated how these spontaneous sequences are related to memory in the psychological sense, they

constitute a fascinating example of a biophysical memory phenomenon realized by local network mechanisms (Csicsvari et al. 2000; Sullivan et al. 2011; Maier et al. 2011).

In this paper, we extend on a model of a sequence memory network using a dynamical systems approach. The model network is operated in a regime at which it can robustly reproduce all stored sequences. The summed length of all stored sequences is called the capacity of the network. It is known that neuronal inhibition can improve this capacity and the robustness of sequence retrieval (e.g. Tsodyks and Feigel'man 1988; Treves 1990; Hirase and Recce 1996). Here we show that, as the network is operated close to its capacity limit, inhibition gives rise to oscillations. Beyond maximum capacity, inhibition enables transient replay, i.e. the partial reproduction of stored sequences—just as observed during sharp-wave-ripple-associated replay in the hippocampus.

2 Model

We model neuronal sequence generation in a network of binary neurons with binary synapses (Willshaw et al. 1969; Golomb et al. 1990; Nadal 1991; Hirase and Recce 1996; Leibold and Kempter 2006). The network consists of a randomly connected network of N excitatory neurons $i = 1, \ldots, N$. The neurons are simplified as binary with state $x_i = 1$ if neuron i fires and state $x_i = 0$ if it is silent. In the hippocampus, neuronal firing during the spontaneous sequences is phase-locked to ripple oscillations (Maier et al. 2011) at a frequency around 200 Hz. We therefore formulate the dynamics in discrete time t indicative of the oscillation cycle. A model neuron i receives an input from neuron j if there is a connection by an active "synapse". Following Gibson and Robinson (1992), the synapses are described by two independent binary stochastic processes. One stochastic variable indicates the presence ($w_{ij} = 1$) or absence ($w_{ij} = 0$) of a morphological connection, with probability $\mathrm{prob}(w_{ij} = 1) = c_m$. The constant c_m thereby denotes morphological connectivity. The other stochastic variable s_{ij} describes the synaptic state, which will be used to store memories. In the potentiated state ($s_{ij} = 1$) a synapse translates a presynaptic spike into a postsynaptic potential whereas in the silent state ($s_{ij} = 0$) it does not influence the postsynaptic neuron. The model neuron i fires a spike at cycle $t + 1$ if the sum of its inputs $h_i(t)$ in the previous cycle t exceeds a threshold θ. In summary, the network dynamics is described by the equation $x_i(t+1) = \Theta\big[\sum_{j=1}^{N} w_{ij} s_{ij} x_j(t) - \theta\big]$, with Θ denoting the Heaviside step function.

The memories stored in the network are sequences of activity patterns $\boldsymbol{\xi}$ described by binary vectors of dimension N, $\boldsymbol{\xi} \in \{0, 1\}^N$. A memory sequence of length Q is an ordered occurrence of activity patterns $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_Q$ (Fig. 1). The number M of active neurons in each pattern is called pattern size, and is the same for all patterns. Memory sequences are stored in the synapses using the learning rule by Willshaw et al. (1969): A synapse is potentiated *only* if it connects two neurons that are activated in sequence at least once. Then the number P of stored associations between a cue and a target pattern is related to the fraction c/c_m of activated synapses by

$$P = \frac{\ln(1 - c/c_m)}{\ln(1 - f^2)} \,. \tag{1}$$

The coding ratio $f = M/N$ is the fraction of active neurons and fixes the firing rate. If all P stored associations can be replayed, the number $\alpha \equiv P/(N\, c_m)$ is called the capacity of the network, and counts the stored associations per number of synapses at a neuron. Note that the sequence length Q is generally much smaller than the total number P of associations; the number of retrievable sequences of length Q is given by $\lfloor P/(Q-1) \rfloor$. Whereas P reflects the combinatorics of the synaptic matrix (determined by N, M, c_m, c), the requirement of a minimum Q implies a stability constraint for the network as a dynamical system.

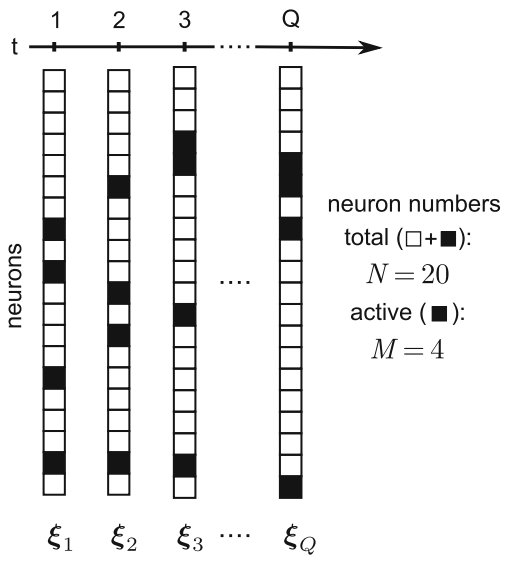

Fig. 1 Sequence of activity patterns. A network of N neurons (*boxes*) is considered in discrete time t. At each time step M neurons are active (*filled boxes*). A sequence that lasts for Q time steps is described by the binary vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_Q$, where $(\xi_k)_i = 1$ if neuron i fires in pattern k and $(\xi_k)_i = 0$ if not. If P transitions between activity patterns are stored in the network, the number of stored sequences of length Q is $\lfloor P/(Q-1) \rfloor$

In the biologically relevant scenario of low coding ratio f, capacity grows like $\alpha \propto M^{-2}$ for fixed c, c_m. Maximum capacity thus corresponds to the minimum pattern size M_{opt} at which the last (Q-th) element of the sequence can still be recalled. The minimum pattern size has been shown to scale like $\ln N$ (Willshaw et al. 1969; Gardner 1987), and hence $\alpha \propto N/(\ln N)^2$.

Sequence retrieval is described by two macroscopic dynamical variables: the number $m_t \in [0, M]$ of correctly activated neurons (*hits*) and the number $n_t \in [0, N-M]$ of incorrectly activated neurons (*false alarms*). For large network sizes N and large pattern sizes M, we can assume Gaussian distributions for the number of inputs $h(t)$, and reinterpret the variables m and n in a mean-field sense as their respective expectation values over realizations of the connectivity matrix. The distributions of inputs are thus characterized by the means $\mu \equiv \langle h(t) \rangle$ and variances $\sigma^2 \equiv \langle h(t)^2 \rangle - \langle h(t) \rangle^2$; for "hit" neurons

$$\begin{aligned} \mu_{\text{On}} &= c_m\, m + c\, n\,, \\ \sigma^2_{\text{On}}(m,n) &= c_m\,(1-c_m)\, m \\ &\quad + c\left[(1-c) + c\,\text{CV}^2_{\text{q}}(n-1)\right] n, \end{aligned} \tag{2}$$

and for "false-alarm" neurons,

$$\begin{aligned} \mu_{\text{Off}} &= c\,(m+n)\,, \\ \sigma^2_{\text{Off}}(m,n) &= c\left[(1-c) + c\,\text{CV}^2_q\,(m+n-1)\right] \\ &\quad \times (m+n)\,. \end{aligned} \tag{3}$$

The terms proportional to

$$\text{CV}^2_q = (1-f^2)^P\, \frac{\left(1 - \frac{f^2}{1+f}\right)^P - (1-f^2)^P}{[1-(1-f^2)^P]^2} \tag{4}$$

originate from correlations in the synaptic states that are induced by Willshaw's learning rule and are computed in the Appendix following Gibson and Robinson (1992).

The network dynamics is then implemented as an iterated map:

$$(m_{t+1}, n_{t+1}) = [T_{\text{On}}(m_t, n_t),\, T_{\text{Off}}(m_t, n_t)]\,. \tag{5}$$

Since only those neurons fire whose input $h(t)$ exceeds the threshold θ, we estimate the expectation values of hits m_{t+1} and false alarms n_{t+1} from the cumulative distribution function (cdf) of the normal distribution, $\Phi(z) \equiv [1 + \text{erf}\,(z/\sqrt{2})]/2$,

$$\begin{aligned} T_{\text{On}}(m,n) &= M\,\Phi[(\mu_{\text{On}} - \theta)/\sigma_{\text{On}}] \\ T_{\text{Off}}(m,n) &= (N-M)\,\Phi[(\mu_{\text{Off}} - \theta)/\sigma_{\text{Off}}]\,. \end{aligned} \tag{6}$$

The nullclines (e.g. $n - T_{\text{On}}(n,m) = 0$) of this dynamical system are shown in Fig. 2(A) for a case of stable retrieval, i.e., there exists an asymptotically stable fixed point $(m_\infty, n_\infty) = [T_{\text{On}}(m_\infty, n_\infty),\, T_{\text{Off}}(m_\infty, n_\infty)]$ with many hits $m_\infty \simeq M$ and few false alarms $n_\infty \ll N - M$.

If the firing threshold is too low or the pattern size is too large (Fig. 2(C)), the nullclines do not cross in a retrieval regime: After initialization at the condition of perfect retrieval $(m_0, n_0) = (M, 0)$, all neurons immediately start to fire and the network falls into an all-active state, $(m,n) \simeq (M, N-M)$. If the firing threshold is too high or the pattern size is too low (Fig. 2(B)) only an unstable fixed point exists in the retrieval region. After initialization at perfect retrieval, the network immediately falls into an all-silent state $(m,n) \simeq (0,0)$.

The phase diagram reveals the three phases of our model sequence memory network (Fig. 2(D)): all silent, all active, and retrieval. The region in which retrieval is possible is wedge-shaped with a thin tip at low pattern sizes M. It turns out that the dynamics usually converges to the fixed points in only a few iterations, meaning that if sequence retrieval is stable for some finite length $Q \gtrapprox 10$, it is likely to be stable for $Q \to \infty$ (see thin grey area in Fig. 2(D) which indicates transient replay between 4 and 99 iterations). Note that, technically, $Q \leq P$, and thus the limit $Q \to \infty$ should be interpreted as having stored a cyclic sequence with $\boldsymbol{\xi}_{Q+1} = \boldsymbol{\xi}_1$.

The region of retrieval obtained from the mean-field equations can be validated with computer simulations of the corresponding networks of binary neurons. As expected, owing to the finite size of the simulated network, the region of retrieval is overestimated by mean-field theory, yet the deviations are relatively small (white discs in Fig. 2(D)). According to Eq. (1), the number P of stored associations increases with decreasing coding ratio $f = M/N$, and thus the network obtains the highest memory capacity at the wedge tip $M = M_{\text{opt}}$. There the stability of the fixed point is particularly sensitive to noise and thus the high capacity is not accessible unless the dynamics can be stabilized. A natural way to stabilize replay is to include feedback inhibition (see Section 3).

2.1 Optimal firing threshold

A different view on the task of the excitatory neurons during sequence replay is that of an optimal detector:

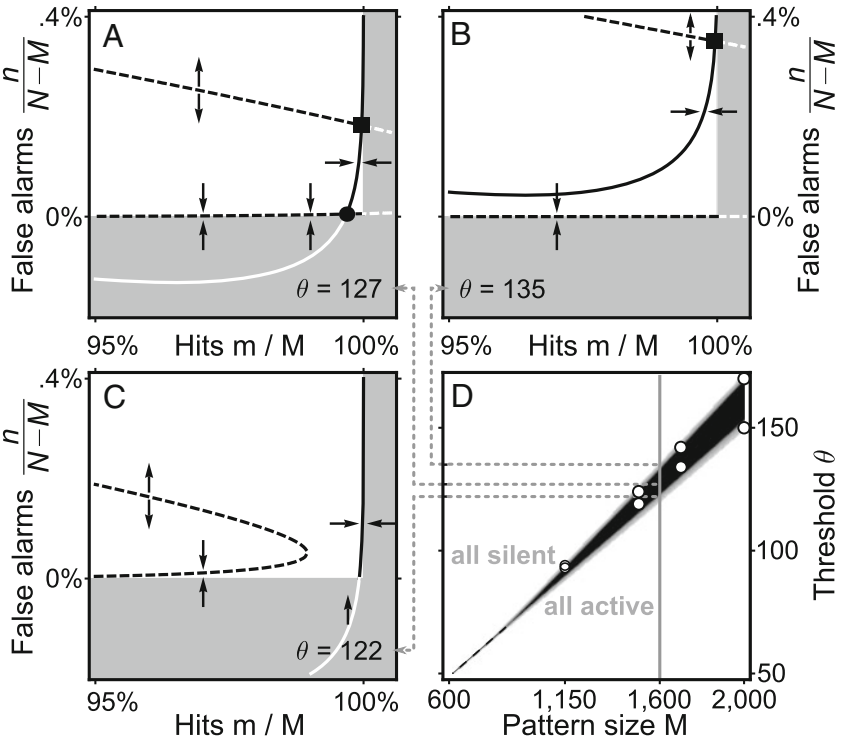

Fig. 2 Phase space and phase diagram. (**A**) Phase space is spanned by the numbers m of hits and n of false alarms. The m nullcline (*solid line*) intersects twice with the n nullcline (*dashed*), producing stable (*disc*) and unstable (*square*) fixed points. *Arrows* indicate attractive or repulsive character of the nullcline; *gray* areas correspond to unphysical values $n < 0$ or $m > M$. (**B**) Same as A for higher threshold θ. Only the unstable fixed point remains. (**C**) Same as A for lower θ. Both fixed points disappear. (**D**) Phase diagram. M-θ space consists of three areas: All silent; the sequence dies out. All active; all neurons fire at maximum rate. Sequence retrieval (*black*); the fraction of hits is much larger than the fraction of false alarms for infinitely many time steps (here tested as $m_t/M > 0.9, n_t/(N-M) < 0.1$ for $t \leq 100$). The *dashed line* separates the all-silent and all-active phases for M-values at which no retrieval phase exists. Areas in *light gray* correspond to transient retrieval of at least 4 time steps. *White discs* mark the boundary of the retrieval region as obtained from simulations of N binary neurons for exemplary values of M. Parameters here and elsewhere are $N = 10^5$, $M = 1{,}600$, $c_m = 0.1$, and $c = 0.05$, unless specified otherwise

the detector neuron is supposed to fire if it belongs to the hit population $\mathcal{A}_{\text{On}}$ of the current time step, or not, in which case it belongs to the false-alarm-population $\mathcal{A}_{\text{Off}}$. The prior probabilities for a neuron to belong to either of these populations, $\Pr(\mathcal{A}_{\text{On}}) = f$ and $\Pr(\mathcal{A}_{\text{Off}}) = 1 - f$, are given by the coding ratio f, which stipulates how many active neurons code for a pattern at any one time step. The basis for the decision whether to fire or not is a one-shot sample from the distributions of synaptic input levels. We again approximate these distributions as Gaussians whose mean and variance depend on whether the detector neuron is target in a pattern or not.

A Bayesian strategy to solve this problem ideally seeks to maximize the *probability of success* S, i.e. the probability of taking the right decision:

$$S = \Pr\left(\text{spike}|\mathcal{A}_{\text{On}}\right)\Pr(\mathcal{A}_{\text{On}}) + \Pr(\text{silence}|\mathcal{A}_{\text{Off}})\Pr(\mathcal{A}_{\text{Off}})\,.$$

Given the spike generation model (spike $\equiv h \geqslant \theta$), the conditional probabilities of spike or silence correspond to integrals of the respective probability densities over regions of synaptic input separated by the threshold θ:

$$S(\theta) = \Pr(h \geqslant \theta|\mathcal{A}_{\text{On}})\Pr(\mathcal{A}_{\text{On}}) + \Pr(h < \theta|\mathcal{A}_{\text{Off}})\Pr(\mathcal{A}_{\text{Off}})\,.$$

The mean-field considerations leading to Eq. (6) allow to rewrite the success probability in terms of Gaussian cdfs:

$$S(\theta) = \Phi(z_{\text{On}})\ f + (1 - \Phi(z_{\text{Off}}))\ (1 - f)\ ,$$

with

$$z_{\text{On/Off}} = \frac{\mu_{\text{On/Off}} - \theta}{\sigma_{\text{On/Off}}}\,.$$

The threshold θ_{opt} that maximizes the success probability can be readily obtained by demanding

$\mathrm{d}S(\theta)/\mathrm{d}\theta = 0$. Since $\partial\Phi(z)/\partial z = \mathrm{e}^{-z^2/2}/\left(\sqrt{2\pi}\sigma\right)$ we have

$$\frac{\mathrm{d}S}{\mathrm{d}\theta} = \frac{1}{\sqrt{2\pi}}\left(-f\frac{\mathrm{e}^{-z_{\mathrm{On}}^2/2}}{\sigma_{\mathrm{On}}} + (1-f)\frac{\mathrm{e}^{-z_{\mathrm{Off}}^2/2}}{\sigma_{\mathrm{Off}}}\right),$$

i.e. the optimal threshold is at the crossing point of the weighted Gaussians. The resulting equation is quadratic in θ,

$$z_{\mathrm{Off}}^2 - z_{\mathrm{On}}^2 = 2\log\left(\frac{1-f}{f}\frac{\sigma_{\mathrm{On}}}{\sigma_{\mathrm{Off}}}\right),$$

and has roots

$$\theta_{\pm} = \left(\sigma_{\mathrm{On}}^2 - \sigma_{\mathrm{Off}}^2\right)^{-1}\left[c(m+n)\sigma_{\mathrm{On}}^2 - (c_m m + cn)\sigma_{\mathrm{Off}}^2 \pm \sigma_{\mathrm{On}}\sigma_{\mathrm{Off}}\sqrt{(c-c_m)^2 m^2 + \left(\sigma_{\mathrm{Off}}^2 - \sigma_{\mathrm{On}}^2\right)\log\left(\frac{M^2}{F^2}\frac{\sigma_{\mathrm{Off}}^2}{\sigma_{\mathrm{On}}^2}\right)}\right]. \tag{7}$$

Generally, one of the thresholds is positive and the other negative when $m \simeq M, n \ll N - M$, which enables heuristic identification of the sign leading to maximization of $S(\theta)$.

2.2 Optimal threshold adaptation

Foreshadowing our interest in adaptive regulation of the threshold, we ask how the threshold should change with the excitatory activities m and n. Figure 3 displays the optimal threshold $\theta_{\mathrm{opt}}(m,n)$ from Eq. (7): In the phase-space region of retrieval (large m, small n), the level curves of $\theta_{\mathrm{opt}}(m,n)$ can be very well approximated by a linear function, which, using Taylor expansion, is

$$\theta_{\mathrm{opt}}(m,n) = \theta_{\mathrm{opt}}(M,0) + \partial_m\theta_{\mathrm{opt}}(M,0)(m-M) + \partial_n\theta_{\mathrm{opt}}(M,0)\,n\,.$$

Figure 3(A) thus demonstrates that the optimal threshold derived from analytic optimality considerations linearly increases with activity levels m and n as has been found numerically by Hirase and Recce (1996).

The coupling coefficients $\partial_m\theta_{\mathrm{opt}}$ and $\partial_n\theta_{\mathrm{opt}}$ are plotted in Figure 3(B). Both partial derivatives have positive values, which, as intuitively expected, corresponds to an increase in threshold for growing activity. Two things should be noted further: First, the two coefficients depend on M only little (at least for $M >$ 880 at which replay is stable; see Figure 2). Second, they are of similar value (between c and $2c$ in our example network). Together this indicates that the optimal adapting threshold may be approximately realized by a single multiplicative coupling constant $b \approx \partial_m\theta_{\mathrm{opt}} \approx \partial_n\theta_{\mathrm{opt}}$ that is the same for both hits m_t and false alarms n_t (see Section 3.1).

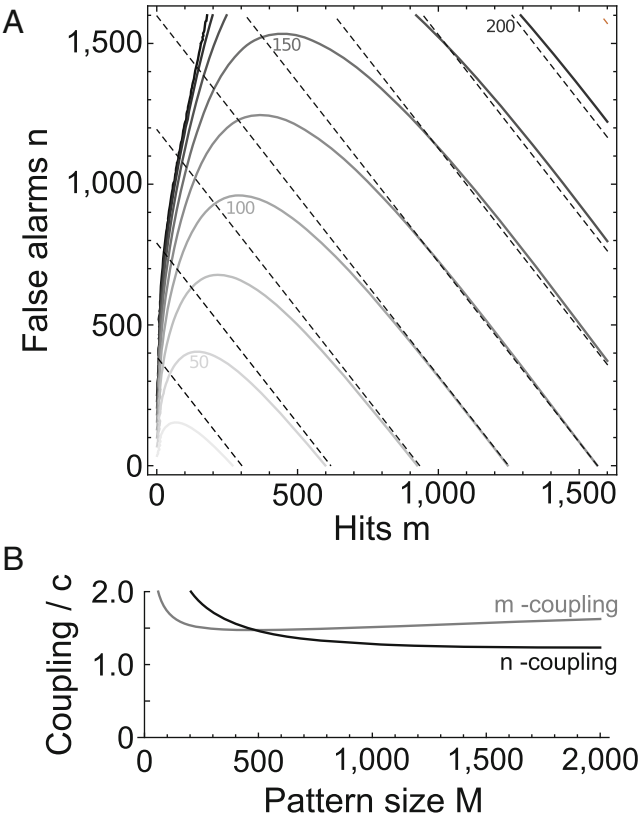

Fig. 3 Optimal threshold is linear in activity. (**A**) Level curves (*grey*) of the optimal threshold as a function of the numbers of hits m and false-alarms n. In this exemplary network with $c = 0.05$, $c_m = 0.1$, $N = 10^5$ and $M = 1{,}600$, the linear approximation (*dashed*; $\theta(m,n) \simeq 1.118 + 0.079\,m + 0.062\,n$) is very good in the retrieval region ($m \approx M, n/N \approx 0$) of the phase plane. (**B**) Coefficients $\partial_{m/n}\theta_{\mathrm{opt}}$ that couple the optimal detector threshold θ_{opt} and excitatory population activity of On (*grey*) and Off cells (*black*)

3 Role of inhibition

It has been shown previously that adaptive thresholds can be interpreted as instantaneous global feedback inhibition and can improve the capacity of associative memory networks (Golomb et al. 1990; Treves 1990; Hirase and Recce 1996). We therefore have investigated the effect of inhibition with respect to its phase-space behavior in our model. First, we consider inhibition to provide an instantaneous negative feedback. Second, and unlike previous approaches, we treat global inhibition as an additional dynamical variable.

3.1 Instantaneous global inhibition

Motivated by previous results on optimal thresholds (Section 2.2 and Hirase and Recce 1996), we introduce an instantaneous negative feedback proportional to the total number $m+n$ of active neurons. The dynamics is derived from Eq. (6) by substituting $\theta \to \theta + b\ (m_t + n_t)$, where the positive b acts as a feedback gain.

The main effect of inhibition is as follows. When the threshold θ is too low (as in Fig. 2(C)), inhibition moves the n-nullcline rightward; when θ is too high (as in Fig. 2(B)), inhibition moves the m-nullcline downward. Finally, in cases for which M is below the optimal pattern size M_{opt} of the purely excitatory model, and no threshold exists for which replay is stable (as in Fig. 4(A)), inhibition moves both nullclines at the same time. Thus, inhibition restores the stable fixed point and therefore effectively enlarges the retrieval phase (Fig. 4(B)). In particular, inhibition lowers the optimal pattern size M_{opt}, thereby enhancing memory capacity $\alpha \propto M_{\text{opt}}^{-2}$ (by a factor of about 2 in the example of Fig. 4(C)). Interestingly, the optimal range for the feedback gain ($b \gtrsim c$) fits well to that for the Bayes-optimal threshold in Fig. 3(B). For such optimal values of b the lower border of the wedge becomes roughly horizontal and the threshold θ is close to zero (not shown). Physiologically, the feedback gain b may be adjusted into this range by plasticity of inhibitory synapses.

To investigate the scaling behavior of the memory capacity, we determined the minimum pattern size M_{opt} for different network sizes N and found the well-known logarithmic dependence $M_{\text{opt}} \propto \ln N$ regardless of the inhibitory gain b (Fig. 4(D)). The capacity thus still grows with network size as $\alpha \propto N/(\ln N)^2$.

An alternative view on instantaneous global feedback inhibition can be derived from the mean values in Eqs. (2) and (3), viz., the substitution $\theta \to \theta + b\ (m_t + n_t)$ effectively reduces the connectivities c_m, and c to $c_m - b$ and $c - b$. The ratio $r = (c_m - c)/c$ between non-activated and activated synapses can be interpreted as the plasticity resources of the network and was shown in Leibold and Kempter (2006) to critically define the signal-to-noise ratio at the post-synaptic neurons (with maximal capacity at $r \approx 10$ for fixed N, M, c_m). By substituting $c_m \to c_m - b$ and $c \to c - b$, instantaneous global inhibition with $b < c$ can be formally interpreted to increase the signal-to-noise ratio like $r \to r/(1 - b/c)$. Since, in the present paper, we initially assume $r = 1$, a feedback gain of $b > 0$ thus generally increases the signal-to-noise ratio for fixed c and thereby enhances capacity (Fig. 4(C)).

3.2 Dynamic global inhibition

We next asked whether the effects observed with instantaneous global inhibition can be reproduced in a more physiological scenario in which inhibitory activity

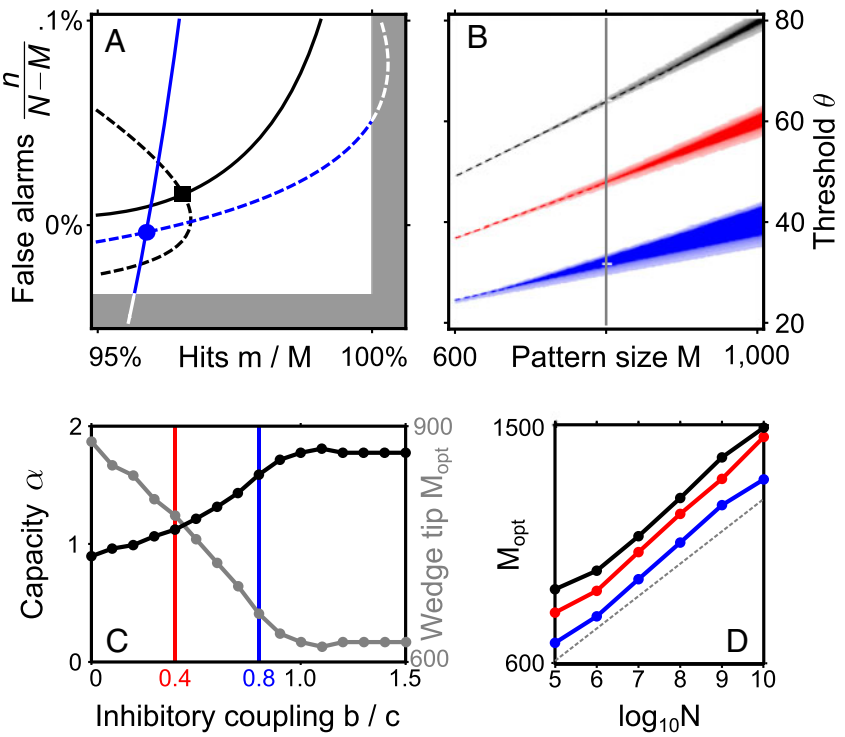

Fig. 4 Retrieval with instantaneous global inhibition $b = 0$ (*black*), $b = 0.4\,c$ (*red*), and $b = 0.8\,c$ (*blue*). (**A**) Nullclines (*red* omitted). Firing threshold θ for the *blue* nullclines is offset by $-b$ M to account for lower mean input h. (**B**) Phase diagram. *Grey vertical line* and *white dashes* indicates M, θ values used in A. *Light colors* show transient retrieval of at least 4 time steps. (**C**) Capacity (*black*) and minimum pattern size M_{opt} (*gray*) as a function of b. (**D**) The minimum pattern size grows sublinearly with N so that the capacity shows an overall increase. The *dashed grey line* indicates a logarithmic dependence $M \propto \ln N$

has its own dynamics, and what additional features such inhibitory dynamics would give rise to. To this end, we extended the model from Eqs. (5) and (6) by including a third dynamical variable k_t that accounts for the number of spikes in an inhibitory pool of K neurons. Each neuron in this pool is assumed to project to and to receive input from excitatory neurons with probabilities c_{IE} and c_{EI}, respectively. Analogous to Eq. (5), the dynamics of k_t is implemented as the map $k_{t+1} = T_{\mathrm{Inh}}(m_t, n_t)$ with

$$T_{\mathrm{Inh}}(m,n) = K\,\Phi[(\mu_{\mathrm{Inh}} - \eta)/\sigma_{\mathrm{Inh}}]. \tag{8}$$

The mean synaptic input and its variance are

$$\mu_{\mathrm{Inh}}(m,n) = c_{EI}\,w_{EI}\,(m+n) \tag{9}$$

$$\sigma^2_{\mathrm{Inh}}(m,n) = w^2_{EI}\,c_{EI}\,(1-c_{EI})\,(m+n). \tag{10}$$

The parameter w_{EI} denotes the synaptic weight of the connections from excitatory to inhibitory neurons. The inhibitory action on the sequence-related variables m and n is implemented by replacing the thresholds in Eq. (6) by $\theta \to \theta + w_{IE}\,c_{IE}\,k_t$, and the variances by $\sigma^2_{\mathrm{On/Off}}(m,n) \to \sigma^2_{\mathrm{On/Off}}(m,n) + k\,w^2_{IE}\,c_{IE}\,(1-c_{IE})$. Again, w_{IE} is the corresponding synaptic weight.

To test for sequence retrieval, the map is initialized with a perfect pattern and matching inhibition, $(m_0, n_0, k_0) = [M, 0, T_{\mathrm{Inh}}(M,0)]$. The resulting phase diagram reveals again regions of stable and transient retrieval (Fig. 5). In agreement with the linear instantaneous inhibition model, the retrieval region in the phase diagram extends to lower pattern sizes M (higher capacities). However, the non-linearity of the sigmoidal Gaussian cdf in Eq. (8) introduces a shearing of this region that can be explained as follows: The Gaussian cdf is roughly linear in the vicinity of the inhibitory threshold η and virtually flat elsewhere. Hence, as an approximation, inhibition has no effect at low total activities $m+n$, it adds a constant $w_{IE}\,c_{IE}\,K$ to the threshold θ at high total activities and establishes a nearly linear coupling for intermediate regimes, similar to the instantaneous-inhibition model from Figure 4. During sequence retrieval, total activity is approximately constant, $m+n \simeq M$, and therefore the retrieval region of the dynamic-inhibition model can be understood as a combination of three retrieval regions of the instantaneous-inhibition model for different feedback gains b and thresholds θ.

The broadening in θ of the region of retrieval (for constant M) with both instantaneous and dynamic inhibition suggests that sequence memory becomes more robust. Mean field theories, however, generally overestimate the regions of stable sequence retrieval (Latham and Nirenberg 2004). To assess the predictive power

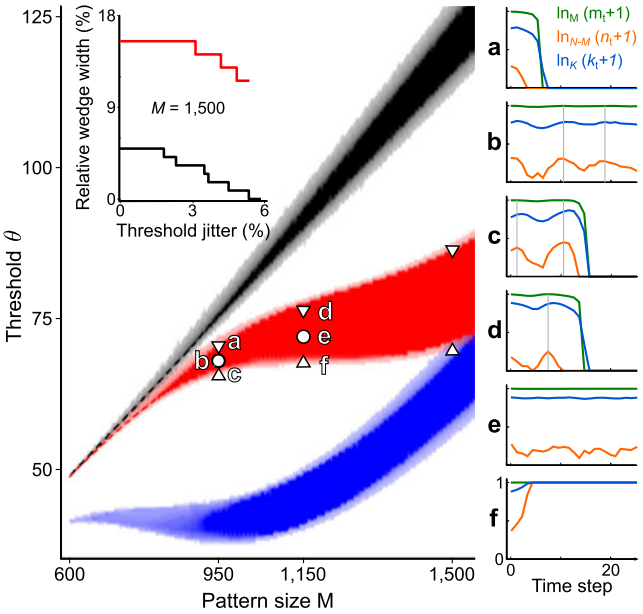

Fig. 5 Comparison of the retrieval regions in M-θ space for the 2-dimensional model without inhibition (*black*) vs. the 3-dimensional model with increasing dynamical feedback inhibition (*red, blue*). *Triangles* on the *red* region mark the first values of θ (in integer steps) for which sequence retrieval is unstable in simulations. *Light colors* show transient retrieval of at least 4 time steps. (**a**–**f**): Example trajectories from network simulations for M-θ pairs as indicated by white markers a–f (m_t: *green*; n_t: *orange*, k_t: *blue*). The *grey vertical lines* indicate Hilbert phase zero for the false-alarm neuron activity n_t. Parameters are $K = 5{,}000$, $w_{IE} = 0.012$, $c_{IE} = 1$, $w_{EI} = 1$, $c_{EI} = 0.01$, and $\eta = 13$ for the *red* region; $\eta = 8.8$ for the *blue* region. *Inset*: Robustness of sequence retrieval against threshold jitter with (*red*, $K = 5{,}000$) and without inhibition (*black*) at $M = 1{,}500$ for simulated networks with threshold noise

of our mean field results, we ran further simulations where neuronal thresholds θ were jittered according to a Gaussian process. The results show that the increase of the relative range of thresholds by inhibition indeed withstands threshold noise (Fig. 5 Inset). At high capacities, the demand of robustness against threshold noise implies that the area of retrieval should be broadest at minimum $M = M_{\mathrm{opt}}$.

We suggest two heuristic criteria for the parameters of dynamic inhibition. First, to achieve maximum sensitivity of the inhibitory feedback, the linear region of $T_{\mathrm{Inh}}(m,n)$ should be centered at the average total input $m+n \simeq M$ during retrieval. This requirement is granted by setting the inhibitory threshold to $\eta = \mu_{\mathrm{Inh}}(M,0)$. Second, the slope at this average total input should yield maximum capacity according to the instantaneous inhibition model (Figs. 3(B) and 4(C)), i.e., it should take a value of at least c. This requirement can be met by appropriately choosing the coupling factor $w_{IE}\,K$. The blue region in Fig. 5 illustrates the outcome

of such an optimization at $M = 880$ with an effective slope of $1.6\,c$ (the red region is obtained at $M = 1{,}300$ and slope $1.3\,c$). The region of stable retrieval is almost flat in M-θ space, suggesting that replay is robust against variability in pattern size M. To the left of the region of stable retrieval, we observe in lighter color a substantial region of transient sequences. Such large regions of transient retrieval only occur for slopes larger than c (not shown), which corresponds to the optimal gain factors for the threshold adaptation from Fig. 3(B). The minimum pattern size M_{opt} of stable retrieval, however, does not decrease further for slopes above c (as in Fig. 4(C)).

Simulations confirm the shape of the fundamental regimes all active, all silent, and retrieval predicted by the three-dimensional mean-field model. Figure 5(a–f) displays simulated trajectories (m_t, n_t, k_t) for typical situations. Interestingly, all-silent states can also sometimes be observed for low threshold values, where inhibition overcompensates the false alarms and transiently allows for sequence retrieval before the network falls back into silence (Fig. 5(c)).

In the retrieval phase the network typically exhibits oscillatory behavior (Fig. 5(b–d)) arising from the interplay between excitatory neurons m_t, n_t and inhibition k_t that manifests itself in oscillations of the two with the phase of inhibition slightly delayed (by about one timestep). The periods of these oscillations are about 5 to 10 time steps corresponding to gamma-range frequencies of 20 to 40 Hz, under our initial assumption that one time step corresponds to a ripple cycle of 5 ms. The oscillatory activity components are present during both transient (Fig. 5(c), (d)) and ongoing replay (Fig. 5(b)). We further analyzed the oscillations based on the inhibitory activities k_t during ongoing replay from cellular simulations (Fig. 6). As a measure for the oscillation amplitude we computed the standard deviation over time std(k/K), and found that it increases towards the edges of the region of replay. As a consequence, the oscillations are particularly strong at the low-M tip of the replay wedge, where the network realizes its maximum capacity. From this, we conclude that gamma oscillations herald the onset of dynamical instability as it is the case at the capacity limit.

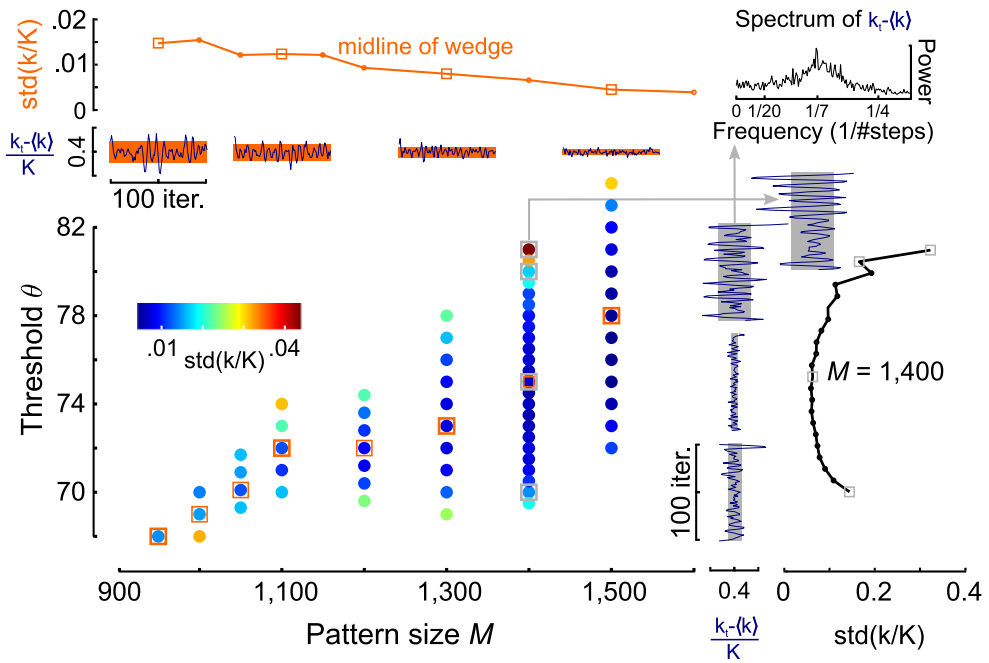

Fig. 6 Transition to instability is marked by increase in amplitude of gamma oscillations. At *lower left*, *colored discs* mark combinations of M and θ for which numerical simulations revealed stable retrieval. The standard deviation over time of the inhibitory activity k_t (normalized to K) is represented by the color code as indicated. At *top left*, oscillation amplitudes (measured as std(k/K) are shown for networks with M, θ along the midline of the wedge. Examples of k_t/K are given with std(k/K) as *orange bars*. At right, oscillation amplitudes std(k/K) are shown for networks of fixed $M = 1{,}400$ and different thresholds; the corresponding k_t/K are given with std(k/K) as *grey bars*. The *top right* panel shows an exemplary power spectrum of the inhibitory activation $k - \langle k \rangle$ for a simulation with M-θ values as indicated. The peak of the spectrum at around 1/7 per time step corresponds to 30 Hz for a time step of 5 ms

4 Discussion

This paper presents a dynamical-systems extension of time-discrete sequence memory models with inhibitory feedback. The effect of instantaneous global feedback inhibition in memory networks has been well studied (e.g. Tsodyks and Feigel'man 1988; Golomb et al. 1990; Treves 1990; Hirase and Recce 1996; Amit and Huang 2010). Our model shows that also dynamical feedback inhibition can stabilize the retrieval of memory sequences and thereby increase both memory capacity and robustness. The optimal instantaneous global inhibitory feedback is a roughly linear function of the total network activity as numerically found by Hirase and Recce (1996) and semi-analytically confirmed by a probabilistic approach in the present paper. Extending the model to dynamic global inhibition, we find that, at the edges of stable replay, inhibition induces strong oscillations, which can be interpreted as gamma oscillations.

Gamma oscillations are ubiquitous in the brain and their origin is generally attributed to local inhibitory networks (Wang 2010). Several cognitive functions have been related to increased gamma power and coherence, such as sensory integration, attention, and memory (Jutras and Buffalo 2010). Specifically, gamma coherence between subregions in the hippocampal formation and prefrontal cortex has been shown to correlate with involvement in a short-term memory task (Sigurdsson et al. 2010). This finding fits well into the general view of gamma rhythms as a mechanism that facilitates communication between brain areas (e.g. Colgin 2011). In our model, gamma occurs as a side effect of feedback stabilization during replay. In combination with these findings, our model suggests that memory networks may have to be critically loaded to be able to transfer information to other brain areas.

Our model also reveals parameter regions in which transient retrieval occurs that lasts for only a few time steps. These regions of transient retrieval (light color in Fig. 5) extend far into low pattern sizes M for strong inhibitory feedback, and thus correspond to the regimes of largest memory capacity. Neuronal networks exhibiting activity sequences hence operate with optimal memory performance if they are in a regime of hyperexcitability that is stabilized by delayed inhibition. This transient retrieval regime is consistent with the dynamic features of sequence replay during sharp wave ripple complexes in the hippocampus, which typically extends over 5 to 10 cycles of an approximately 200 Hz oscillation, and that are accompanied by delayed inhibitory feedback (Maier et al. 2011).

In large environments, sequence replay *in vivo* can span several ripple episodes (Davidson et al. 2009), showing that long sequences can be constructed by concatenating multiple transient replay episodes. Our model argues that such fragmentation solves the dilemma between stability and capacity by not having to trade capacity for stability. Instead it uses dynamic feedback inhibition to break sequence replay into short stable fragments of transient replay. It remains open though, how information transfer between these fragments is realized and whether it occurs intra- or extrahippocampally.

Throughout the paper, we consider the connectivity parameters c and c_m as constants, based on the assumption that the morphological connectivity c_m is mainly determined by geometrical constraints such as the size of the cell surface, or the volume requirement of wiring. The functional connectivity c is assumed to result from the specific learning rules that ensure that the network always remains plastic: In order to store new memories a large fraction of synapses has to be able to change its state. In the parameter regime used for our analysis, this requirement is fulfilled by fixing $c/c_m = 0.5$. Moreover, the connectivities employed are small, since experiments indicate that hippocampal networks are sparsely connected (Miles and Wong 1986).

Limitations of our model arise from specific assumptions underlying our analysis. One of them is that of a constant pattern size M. In reality pattern sizes may be variable (as discussed for a different learning rule in Amit and Huang 2010), leading to a decreased capacity. Another significant simplification of our model is the discreteness in time. Dynamical interactions of synaptic currents and membrane processes during sharp-wave ripples may also reduce capacities. In this sense the capacity values derived in this paper can only be considered as upper bounds and for determining scaling behavior.

Extending the model to more realistic dynamics is necessary to investigate how close to the upper bound of capacity a real spiking network can get. Such a translation to biophysically more realistic neuron models, however, raises difficult problems. The massive bombardment by synaptic inputs (specifically inhibition) sets the cells into a high-conductance regime, in which the effective time constants become short and spike generation very sensitively depends on the timing of inputs. Further, the interplay between excitation and inhibition not only has to keep the cell in a balanced state in which spiking is sparse, but also has to ensure synchrony of spiking in the time slots of roughly 5 ms.

Other models of hippocampal sharp-wave ripples focused on dynamic features of sharp-wave ripples as a network phenomenon mostly disregarding functional aspects of sequence replay. In the model of

Memmesheimer (2010) sharp waves arise from spontaneous bursts of excitatory activity that are shaped by non-linear dendritic integration. Such a scenario requires a relatively high level of background activity (high n) and it is not yet clear how well this can work together with sequence replay at high memory capacities, where false alarms n are not desired. In another model by Vladimirov et al. (2012) synaptic integration plays no role in evoking action potentials. Spiking is propagated across axons by axo-axonal gap junctions (Schmitz et al. 2001). Also in this model the relation of these axonal spike patterns to memory-related processes has not been evaluated. Moreover, it's unclear how inhibition could physiologically be realized in such a scenario. We thus conclude that, despite these considerable efforts, we still lack a model of sharp wave ripples that combines realistic physiological phenomenology with functional hypotheses of the hippocampal memory networks.

Acknowledgements The authors are grateful to Alexander Mathis for comments on a previous version of the manuscript. This work was funded by the German Ministry for Education and Research (BMBF) under grant number 01GQ0440 (BCCN Munich) and the German Science Foundation (DFG) under grant numbers Le-2250/2 and GRK 1190.

Appendix: First and second moments

The dynamics underlying neuronal activity sequences is formulated as a two-dimensional iterated map in Eqs. (2)–(6). This time-discrete dynamics is simplified using Gaussian approximations for the distributions of the number h of synaptic inputs to a specified neuron. The Gaussian approximation therefore requires expressions for the means and variances of the input sums h.

Inputs can be of two kinds, hits m and false alarms n. The input sum $h = \sum_{j=1}^{m+n} w_j s_j$ thus runs over all $m+n \leq N$ active (firing) neurons in the network and depends on two binary random variables for each potential input: $w \in \{0,1\}$ indicating the presence of a synaptic connection, and $s \in \{0,1\}$ indicating its state (Gibson and Robinson 1992). The stochasticity of s is inherited from the randomness of the activity patterns underlying the memory sequences via Willshaw's learning rule.

The distribution of w is given by the morphological connectivity such that $\mathrm{prob}(w=1) = c_m$. The probability $\mathrm{prob}(s=1)$ of a synapse having been potentiated depends on whether it connects or not neurons that should fire in sequence at the particular point in time.

The Willshaw rule ensures that synapses that connect sequentially firing neurons are in the potentiated state, i.e. $\mathrm{prob}(s=1)=1$, and thus for this subset of synapses the input sum depends on a binomial process with probability $\mathrm{prob}(w=1) = c_m$.

For the other synapses, the probability $\mathrm{prob}(s=1) = q_x$ depends on the the number $x \leq P$ of associations the specific postsynaptic neuron is involved in. Note that if the postsynaptic neuron is never supposed to fire, the Willshaw rule will activate none of its synapses and thus $q_0 = 0$. In general, the probability that a neuron is not a target in one specific step of the sequence (association) is $1-f$, and thus the probability that it is not a target in any one of x associations is $(1-f)^x$. Conversely, the probability of such a synapse being potentiated is $q_x = 1-(1-f)^x$. Hence, assuming independence of the two binomial processes, the input sum h for this subset of synapses is binomial with probability

$$\mathrm{prob}(w_i s_i = 1) = c_m\, q_x \,. \tag{11}$$

The probability distribution of the input h can then be determined as

$$p(h) = \sum_{x=0}^{P} p(h|x)\; p(x) \,, \tag{12}$$

in which the conditional probability $p(h|x) = \binom{m+n}{h} (c_m\; q_x)^h\, (1 - c_m\; q_x)^{m+n-h}$ is derived from Eq. (11), and the probability $p(x)$ that a neuron is involved in x associations is also binomial, viz. $p(x) = \binom{P}{x} f^x\, (1-f)^{P-x}$.

To compute expected values of h, we have to discern between neurons that should be active at time step $t+1$ (and are supposed to generate the hits) and those that should be silent (and potentially give rise to false alarms). For the potential false alarms, we obtain

$$\begin{aligned}
\langle h\rangle_{\mathrm{Off}} &= \sum_{h=0}^{m+n} h \sum_{x=0}^{P} p(h|x)\; p(x) = \sum_{x=0}^{P} p(x) \sum_{h=0}^{m+n} h\; p(h|x) \\
&= \sum_{x=0}^{P} p(x)\; (m+n)\; (c_m\, q_x) \\
&= (m+n)\, c_m \sum_{x=0}^{P} [1-(1-f)^x] \binom{P}{x} f^x (1-f)^{P-x} \\
&= (m+n)\, c_m \left[1 - (1-f)^P \sum_{x=0}^{P} \binom{P}{x} f^x \right] \\
&= (m+n)\, c_m \left[1 - (1-f)^P\, (1+f)^P \right] \\
&= (m+n)\, c \,.
\end{aligned}$$

Note that the last step makes use of the capacity of the Willshaw rule, Eq. (1). Similarly, for the potential hits, we obtain

$$\langle h\rangle_{\mathrm{On}} = \sum_{h'=0}^{n} h' \sum_{x=0}^{P} p(h'|x)\, p(x) + \sum_{h''=0}^{m} h''\, p(h'') = n\,c + m\,c_m\ .$$

Here the expected value sums over two independent subsets of neurons, the first one (h') representing the false alarms, and the second (h'') representing the hits during the previous time step.

The corresponding variances can be obtained analogously employing the formula of the geometric series several times, and introducing the abbreviation $\mathrm{CV}_q^2 = \mathrm{var}_x\, q/\langle q\rangle_x^2$ with expected values according to the distribution $p(x)$:

$$\sigma^2_{\mathrm{On}}(m,n) = c_m\, m\,(1-c_m) + n\,c\left[(1-c) + c\,\mathrm{CV}_q^2\,(n-1)\right]$$

$$\sigma^2_{\mathrm{Off}}(m,n) = (m+n)\,c \times\left[(1-c) + c\,\mathrm{CV}_q^2\,(m+n-1)\right].$$

Note that $\mathrm{CV}_q \to 0$ for $f \to 0$, and, in this limit, the variance formulas $\sigma^2_{\mathrm{On}} \to m\,c_m\,(1-c_m) + n\,c\,(1-c)$, $\sigma^2_{\mathrm{Off}} \to (m+n)\,c\,(1-c)$ from the present theory approximate those in Leibold and Kempter (2006).

References

Amit, Y., & Huang, Y. (2010). Precise capacity analysis in binary networks with multiple coding level inputs. *Neural Computation, 22,* 660–688.

Buhry, L., Azizi, A. H., & Cheng, S. (2011). Reactivation, replay, and preplay: How it might all fit together. *Neural Plasticity, 2011*, 203462.

Buzsáki, G., Horvath, Z., Urioste, R., Hetke, J., & Wise, K. (1992). High-frequency network oscillation in the hippocampus. *Science, 256*, 1025–1027.

Colgin, L. L. (2011). Oscillations and hippocampal-prefrontal synchrony. *Current Opinion in Neurobiology, 21*, 467–474.

Csicsvari, J., Hirase, H., Mamiya, A., & Buzsáki, G. (2000) Ensemble patterns of hippocampal CA3-CA1 neurons during sharp wave-associated population events. *Neuron, 28*, 585–594.

Davidson, T. J., Kloosterman, F., & Wilson, M. A. (2009). Hippocampal replay of extended experience. *Neuron, 63*, 497–507.

Dragoi, G., & Tonegawa, S. (2011). Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature, 469*, 397–401.

Diba, K., & Buzsáki, G (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience, 10*, 1241–1242.

Dragoi, G., & Buzsáki, G. (2006). Temporal encoding of place sequences by hippocampal cell assemblies. *Neuron, 50*, 145–157.

Gardner, E. (1987). Maximum storage capacity in neural networks. *Europhysics letters, 4*, 481–485.

Gibson, W. G., & Robinson, J. (1992). Statistical analysis of the dynamics of a sparse associative memory. *Neural Networks, 5*, 645–661.

Golomb, D., Rubin, N., & Sompolinsky, H. (1990). Willshaw model: Associative memory with sparse coding and low firing rates. *Physical Review. A, 41*, 1843–1854.

Gupta, A. S., van der Meer, M. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron, 65*, 695–705.

Hirase, H., & Recce, M. (1996). A search for the optimal thresholding sequence in an associative memory. *Network, 7*, 741–756.

Jutras, M. J., & Buffalo, E. A. (2010). Synchronous neural activity and memory formation. *Current Opinion in Neurobiology, 20*, 150–155.

Latham, P. E., & Nirenberg, S. (2004). Computing and stability in cortical networks. *Neural Computation, 16*, 1385–1412.

Lee, A. K., & Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron, 36*, 1183–1194.

Leibold, C., & Kempter, R. (2006) Memory capacity for sequences in a recurrent network with biological constraints. *Neural Computation, 18*, 904–941.

Maier, N., Tejero-Cantero, Á., Dorrn, A., Winterer, J., Beed, P., Morris, G., et al. (2011). Coherent phasic excitation during hippocampal ripples. *Neuron, 72*, 137–152.

Memmesheimer, R. (2010). Quantitative prediction of intermittent high-frequency oscillations in neural networks with supralinear dendritic interactions. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 11092–11097.

Miles, R., & Wong, R. K. (1986). Excitatory synaptic interactions between CA3 neurones in the guinea-pig hippocampus. *Journal of Physiology, 373*, 397–418.

Nadal, J.-P. (1991). Associative memory: On the (puzzling) sparse coding limit. *Journal of Physics. A. Mathematical and General, 24*, 1093–1101.

Nádasdy, Z., Hirase, H., Czurkó, A., Csicsvari, J., & Buzsáki, G. (1999). Replay and time compression of recurring spike sequences in the hippocampus. *Journal of Neuroscience, 21*, 9497–9507.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research, 34*, 171–175.

Schmitz, D., Schuchmann, S., Fisahn, A., Draguhn, A., Buhl, E., Petrasch-Parwez, E., et al. (2001). Axo-axonal coupling: A novel mechanism for ultrafast neuronal communication. *Neuron, 31*, 831–840.

Sigurdsson, T., Stark, K. L., Karayiorgou, M., Gogos, J. A., & Gordon, J. A. (2010). Impaired hippocampal-prefrontal synchrony in a genetic mouse model of schizophrenia. *Nature, 464*, 763–767.

Sullivan, S., Csicsvari, J., Mizuseki, K., Montgomery, S., Diba, K., & Buzsáki, G. (2011). Relationships between hippocampal sharp waves, ripples, and fast gamma oscillation: Influence of dentate and entorhinal cortical activity. *Journal of Neuroscience, 31*, 8605–8616.

Vladimirov, N., Tu, Y., & Traub, R. D. (2012). Shortest loops are pacemakers in random networks of electrically coupled axons. *Frontiers in Computational Neuroscience, 6*, 17.

Treves, A. (1990). Graded-response neurons and information encodings in autoassociative memories. *Physical Review A, 42*, 2418–2430.

Tsodyks, M. V., & Feigel'man, M. V. (1988). Enhanced storage capacity in neural networks with low level of activity. *Europhysics Letters, 6*, 101–105.

Wang, X.-J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews, 90*, 1195–1268.

Willshaw, D. J., Buneman, O. P., & Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature, 222*, 960–962.

Wilson, M. A., & McNaughton, B. L. (1994) Reactivation of hippocampal ensemble memories during sleep. *Science, 265*, 676–679.

Chapter 3

Hippocampal Remapping is Constrained by Sparseness rather than Capacity

Contributions

This work was done under the supervision of Christian Leibold (CL). Axel Kammerer (AK) and CL conceived and designed the research. AK wrote the code and carried out the simulations. AK and CL discussed the results and wrote the paper.

The paper has been published in *PLoS Computational Biology* Kammerer and Leibold (2014).

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Hippocampal Remapping Is Constrained by Sparseness rather than Capacity

Axel Kammerer[1,2], **Christian Leibold**[1]*

1 Department Biologie II, Ludwig-Maximilians-Universität München, Planegg, Germany, **2** Graduate School for Systemic Neurosciences, Ludwig-Maximilians-Universität München, Planegg, Germany

Abstract

Grid cells in the medial entorhinal cortex encode space with firing fields that are arranged on the nodes of spatial hexagonal lattices. Potential candidates to read out the space information of this grid code and to combine it with other sensory cues are hippocampal place cells. In this paper, we investigate a population of grid cells providing feed-forward input to place cells. The capacity of the underlying synaptic transformation is determined by both spatial acuity and the number of different spatial environments that can be represented. The codes for different environments arise from phase shifts of the periodical entorhinal cortex patterns that induce a global remapping of hippocampal place fields, i.e., a new random assignment of place fields for each environment. If only a single environment is encoded, the grid code can be read out at high acuity with only few place cells. A surplus in place cells can be used to store a space code for more environments via remapping. The number of stored environments can be increased even more efficiently by stronger recurrent inhibition and by partitioning the place cell population such that learning affects only a small fraction of them in each environment. We find that the spatial decoding acuity is much more resilient to multiple remappings than the sparseness of the place code. Since the hippocampal place code is sparse, we thus conclude that the projection from grid cells to the place cells is not using its full capacity to transfer space information. Both populations may encode different aspects of space.

Citation: Kammerer A, Leibold C (2014) Hippocampal Remapping Is Constrained by Sparseness rather than Capacity. PLoS Comput Biol 10(12): e1003986. doi:10.1371/journal.pcbi.1003986

Editor: Matthias Bethge, University of Tübingen and Max Planck Institute for Biologial Cybernetics, Germany

Received June 13, 2014; **Accepted** October 14, 2014; **Published** December 4, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: To support this work, CL received grants from the Federal Ministry for Education and Research (http://www.bmbf.de/) under grant numbers 01GQ1004A and 01GQ0981 and the German Research Foundation (www.dfg.de) under grant number RTG 1091. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: leibold@bio.lmu.de

Introduction

The neuronal representation of space that is necessary for navigation and orientation has been traditionally assigned to the hippocampal place cell system [1], where cells fire only at few distinct locations and are silent elsewhere. Since the discovery of grid cells in the medial entorhinal cortex (MEC) [2,3], which fire on a hexagonal spatial lattice, a second space representation is now known and it has become unclear what the functional differences of the two are. It is speculated that the MEC grid cells are predominantly used in path integration, whereas the place cells may connect position and context information [4]. From the coding perspective it is remarkable that the hippocampal place fields are considerably sparse, whereas the grid fields generate a much denser code with approximately one third of all grid cells active at any one time [3]. Since both networks are reciprocally connected anatomically [5,6] and functionally [7,8], the two space representations have to be synchronized. Understanding the interplay of both codes thus leads to the more general question of how a dense neuronal code can be efficiently transferred into a sparse code and vice versa.

In this paper, we focus on the mapping from grid to place cells. This extends previous coding approaches in so far as they studied the isolated grid cell system from a mainly information theoretic perspective [9,10]. Here, we discuss a coding theory by including the further constraint that the grid code has to be readable by the place code at a similar and behaviorally relevant resolution, since we assume that space information is only relevant for the brain if it can be read out by other neurons. Employing two population models, for grid cells and place cells, we show that a relevant resolution of the order of centimeters can be easily transferred from a relatively small grid-cell to a relatively small place-cell population. Larger numbers (particularly of place cells) can thus be used to encode multiple environments [11] at a similar spatial resolution. Our model also shows that may interference owing to multiple environments reduces the sparseness of the hippocampal code much faster than it reduces the space information of the population patterns measured by the number of different environments that can be encoded at a given spatial resolution. These findings argue against a pure feed-forward model of place field formation from grid cells, consistent with recent experimental findings [7,12–16].

Results

Here we briefly summarize the general structure of our model, whereas a detailed account is provided in the Materials and Methods Section. A population of N_g grid cells is connected to N_p

Author Summary

The mammalian brain represents space in the population of hippocampal place cells as well as in the population of medial entorhinal cortex grid cells. Since both populations are active at the same time, space information has to be synchronized between the two. Both brain areas are reciprocally connected, and it is unclear how the two codes influence each other. In this paper, we analyze a theoretical model of how a place code processes inputs from the grid cell population. The model shows that the sparseness of the place code poses a much stronger constraint than maximal information transfer. We thus conclude that the potentially high spatial acuity of the grid code cannot be efficiently conveyed to a sparse place cell population and thus propose that sparseness and spatial acuity are two independent objectives of the neuronal place representation.

place cells via a feed-forward synaptic matrix. The grid cells are organized in four modules that differ in the spatial period (or grid spacing) of the periodic hexagonal firing patterns [17]. The neuronal activities of the MEC and hippocampal populations are assumed to encode either linear tracks or square boxes both of length 1 m (Figs. 1 and 2). Different environments are represented by phase shifts of the grid fields that are identical for all cells in a module [18] but random between modules [19].

The spike count of the grid cells is assumed to follow Poisson statistics. For the place cells we first define place fields that optimally cover the whole environment but are only used as teacher patterns in a training step in which we construct synaptic weights between grid cells and place cells by supervised Hebbian learning. The teacher place fields are randomly assigned in each environment (shuffling of place cells) resembling the global remapping [20] of hippocampal place fields found in experiments. For each such remapping synaptic weights are incremented according to the Hebb rule such that all shifted grid patterns activate the corresponding remapped place code.

Realizations of grid field spikes are projected via the learned feed-forward connections to the place field population that employs a soft winner-take-all mechanism (E%-MAX rule) to emulate recurrent inhibition [21]. The activity from these simulations determines the actual firing fields and spike statistics of the place cells. The spatial acuity of both codes is measured by the empirical minimum mean square decoding error of single trial activity. The simulations are evaluated by a variety of measures including sparseness and the similarity between the place fields used during training and those obtained in the simulation.

The capacity of a spatial code consists of two components. First, the spatial resolution [9], or how precisely one can infer a spatial position. Second, how many different environments can be represented. Since different environments are obtained by MEC phase shifts and hippocampal remapping, all spatial information is conveyed by the same synaptic connections. Thus the multiple stored environments interfere at the cost of spatial resolution.

Resolution of the grid code

To assess the ground truth of our model, we first evaluate the coding capacity of the grid cell population on a one-dimensional linear track (Fig. 3). The spatial resolution (denoted as root-mean square estimation error; RMSE) non-trivially depends on the tuning width σ_g of the grid code and the number N_g of neurons [9,22]. Three examples of grid codes are shown in Fig. 3A–C for three different values of σ_g. Grids as usually observed in MEC are most similar to the situation in Fig. 3B, whereas Fig. 3A and C illustrate settings with extremely thin and broad tuning curves, respectively. Thus, the biological value of σ_g is about 1, which corresponds to a ratio between tuning width and spatial period of about 0.3 (see Fig. S4 of [3]). However, the RMSE non-monotonically depends on σ_g [22] with a minimum at rather thin tuning curves (Fig. 3D).

The resolution (RMSE) improves with N_g such that even for moderate cell numbers (several hundreds) it is easy to obtain spatial resolutions in the range of 1 mm and below. From a behavioral perspective, however, one may ask whether such a resolution is actually psychophysically reasonable, or even useful. We thus suggest that resolution is probably not the major objective of the grid code and test the alternative possibility that the grid

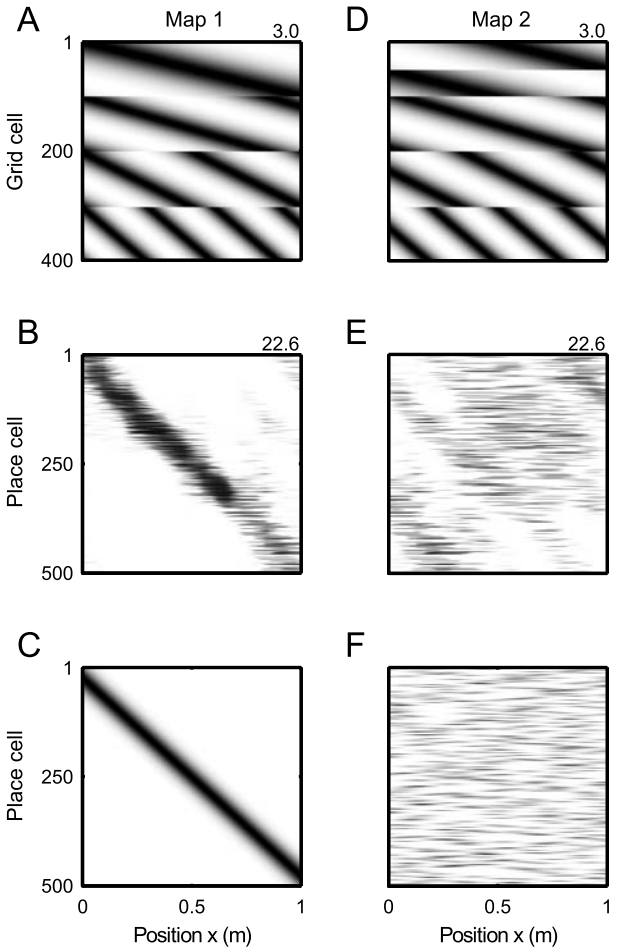

Fig. 1. Hebbian learning of multiple linear tracks. (A) Grid cell firing maps for 400 grid cells with width constant $\sigma=1$ on a 1 meter linear track ($S_g=1.5$, $S_p=2.56$, $\sigma_p=0.05$ m, $N_p=500$). The cells are organized in 4 modules, with a period ratio of 1.67 to achieve a spatial period of 30 cm in the lowest module. The numbers at top right corners indicate the maximal spike count C_g as a proxy for peak firing rate (see Materials and Methods). (B) Firing rates of place cells which received the grid field activity from A as an input. The corresponding synaptic connections were obtained from an Hebbian outer product rule based on the rate maps of the grid population in A and the ideal place field population (C). (D) To represent a second environment, the grid code from A is shifted by module-specific phases. (E) Globally remapped place code that is learned from the remapped rate maps in D and F. (F) Ideal place code in the second environment.
doi:10.1371/journal.pcbi.1003986.g001

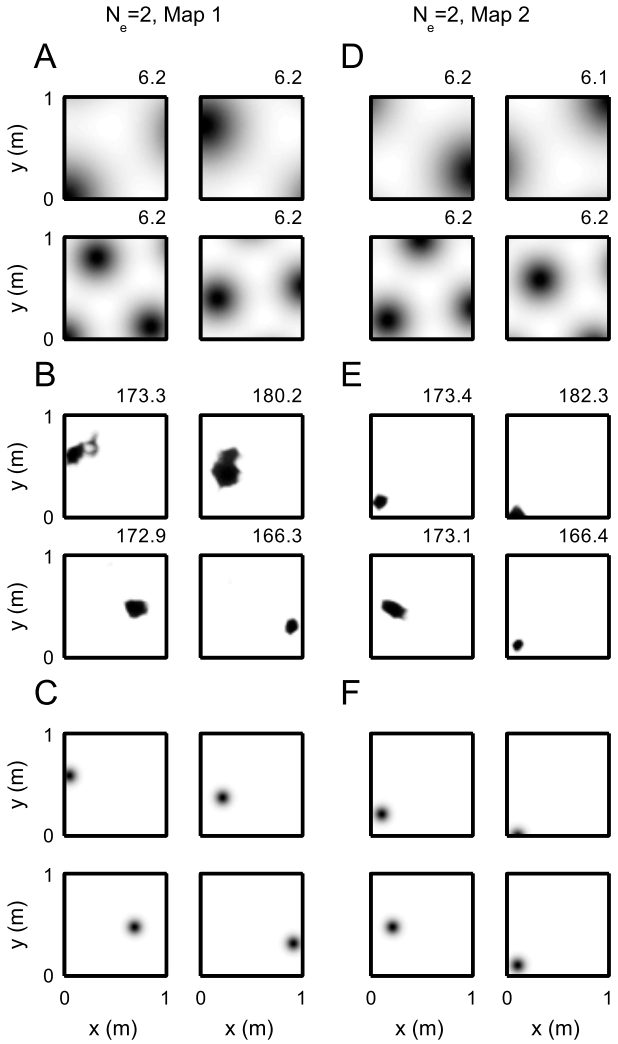

Fig. 2. Two-dimensional rate maps for grid cells and place fields in two environments ($N_e=2$). (A, D) Grid rates differ by module-specific phase shifts. Four example cells are shown, two from the first module (top) and two from the second (bottom). A total of four modules was used. Maximum spike counts C_g shown above each plot. (B, E) place cell rate maps for both remappings. Positions of place fields are set by Hebbian learning. (C, F) Desired place fields as used for Hebbian learning. Firing fields in C are distributed in a square lattice equidistantly across the environment. Fields in F are obtained by shuffling cell identities from C, which ensures equal coverage. Parameters are $N_p=500$ place cells and $N_g=400$ grid cells and $\sigma_p=0.05$ m, $\sigma_g=0.3$, $S_p=2.56$, $S_g=1.5$. All other parameters are as for the one-dimensional case.
doi:10.1371/journal.pcbi.1003986.g002

code may be designed to display a reasonable spatial resolution in as many environments as possible. As a lower bound for such a reasonable resolution we postulate an RMSE of 0.5 cm (dashed line in Fig. 3D) and ask the question, which parameter setting in N_g,σ_g-space would actually result in this behaviorally relevant RMSE (Fig. 3E). The minimum N_g scales supra-linearly with σ_g, i.e. it flattens out for smaller σ_g. We thus argue that $\sigma_g \approx 1$ is a good choice because it still is in the super-linear regime requiring only relatively small cell numbers and at the same time results in tuning widths that are similar to biology (like Fig. 3B). For further analysis we thus fix the grid code to $\sigma_g=1$ and $N_g=400$.

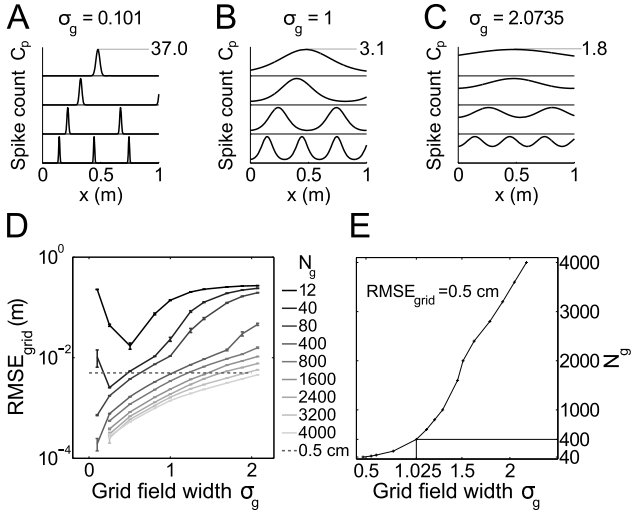

Fig. 3. Root mean square error of grid cells ($RMSE_{grid}$) on the linear track. (A–C) Example tuning curves for 4 cells from different modules and three choices of width constant σ_g. (D) $RMSE_{grid}$ as a function of cell number N_g and tuning width σ_g. (E) Scaling of N_g with σ_g for fixed $RMSE_{grid}=0.5$ cm. Parameters are $M=4$, $\lambda_{max}=(1+0.4\sigma_g)$ m, $S_g=1.5$.
doi:10.1371/journal.pcbi.1003986.g003

Resolution of the place code in a single environment

The spatial acuity of the population code of grid cells can only be made use of if it can be read out by downstream centers. We therefore asked under which conditions the resolution of grid cell network from the previous subsection can be preserved in the place cell network under the ideal conditions that only one environment has to be represented (number of environments $N_e=1$); Fig. 4.

Since the tuning curves are actually learned there exists a clear lower bound for the tuning widths that reflects the minimal width of the grid cell population (Fig. 4A–F). Narrower place fields cannot be achieved by the present model even if the fields used during training are much narrower than the smallest grid fields. Similar as for the grid cell code, a reduction in the place field width effectively improves the RMSE, however, the resolution is limited by that of the grid code (0.5 cm). Therefore an increase in the number N_p of place cells reduces the RMSE and the performance quickly converges to the minimum for $N_p \gtrapprox 100$; Fig. 4G. Only relatively few neurons are needed to achieve such a behaviorally relevant resolution, and thus we next asked how many different environments can be represented at this resolution.

Multiple environments

Storing multiple environments generates interferences of the place codes since each remapping taxes synaptic resources. Thus the spatial resolution of the place code is getting worse when storing multiple environments (Fig. 5). However, even for 21 remappings in our parameter regime ($N_p=500$) the decoding error is still relatively low ($<5\%$). Also the number N_e of remapped environments for which decoding is possible increases with the number of place cells (Fig. 6A), such that even for moderate place cell numbers N_p many environments can be easily decoded at physiological resolution.

Although space information is retained for considerably large values of N_e, the place code degenerates already for much smaller N_e. This degeneration is best described by a loss of sparseness (Fig. 6B, [23]) resulting from less localized firing fields, while the

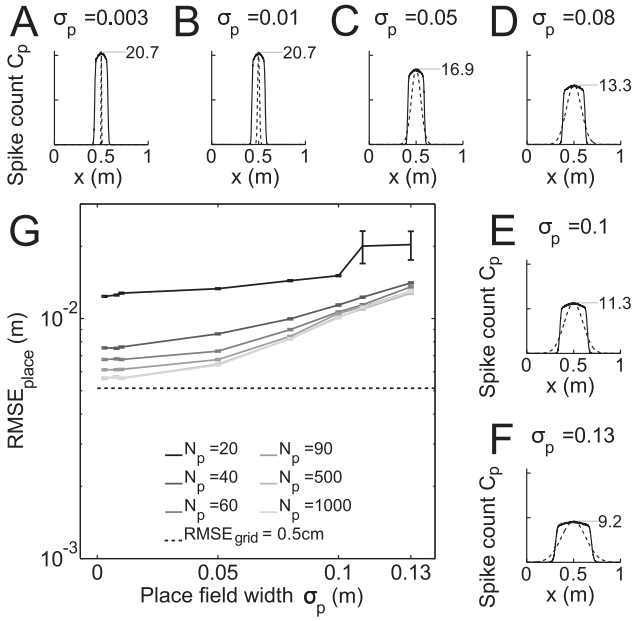

Fig. 4. Root mean square error of place cells $RMSE_{place}$ on linear track. (A–F) Place cell tuning functions (spike counts C_p as a function of space x). Dashed lines: teacher tuning curves used for training. Solid lines: tuning curves after learning averaged over 800 trials (σ_p is the width of the teacher curves). (G) Place cell resolution $RMSE_{place}$ as function of σ_p and N_p. Grid cell resolution is shown as dashed line. Parameters used were $N_g = 400$, $\sigma_g = 1.038$, $S_p = 2.56$, other parameters were as in Fig. 3.
doi:10.1371/journal.pcbi.1003986.g004

average spike count C_p remains constant (see Materials and Methods). This delocalization results in a reduction of the number of proper place cells (Fig. 6C) which exhibit an increased number of regular-sized firing fields (Fig. 6D, E) before they cease to be place cells and are active over almost the whole track as indicated by a mean population sparseness (average fraction of active cells at a position) close to 1 (Fig. 6F). Also the firing fields quickly loose their similarity to the trained firing fields (Fig. 6G). From these observations we conclude that although a large number N_p of putative place cells allow to reliably decode a large number of environments by remapping, the place field quality (i.e. the sparseness) of the encoding neurons disappears. Thus the observation of a sparse place code in the hippocampus must result from further objectives beyond decoding quality and remapping capacity.

Generalization to open fields

To test whether these observations are specific to the one-dimensional paradigm, we repeated the same simulations and analysis for a two-dimensional enclosure (see Materials and Methods and Fig. 2). As in the one-dimensional case, inspection of single examples for high numbers N_e of remappings reveals that the place-selectivity of the readout neurons (the putative place cells) deteriorates much faster than the decoding quality (Fig. 7). Even random spatial patches (for $N_e = 30$; Fig. 7 B) allow for almost perfect decoding (Fig. 7 E). Spatial estimation only breaks down, if hardly any space modulation is observable in the firing patterns (Fig. 7 C, F). These exemplary observations are corroborated by a systematic quantitative assessment of the code and the firing fields in Fig. 8.

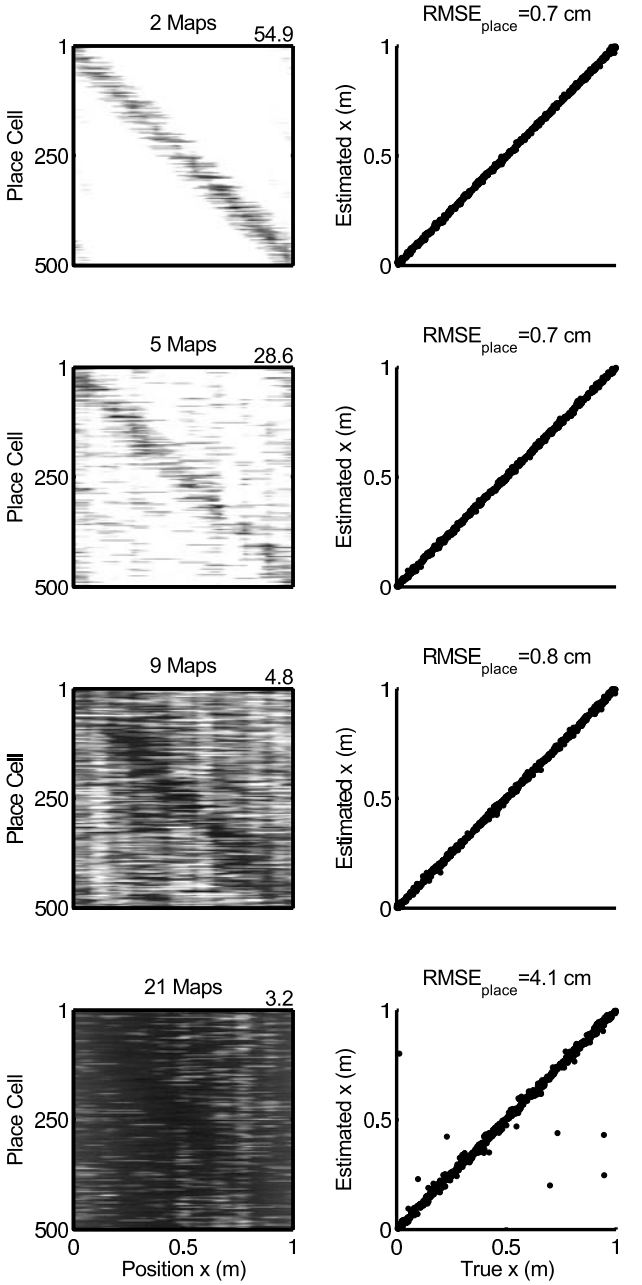

Fig. 5. Quality of 1-d place code for increasing number of maps in a network with $N_p = 500$ place cells and $N_g = 400$ grid cells and $\sigma_p = 0.01$ m, $\sigma_g = 1.038$, $S_p = 2.56$, $S_g = 1.5$. Left column: Rate map for environment 1. Right column: Position estimates from the place code as a function of real position.
doi:10.1371/journal.pcbi.1003986.g005

In analogy to the one-dimensional case, decoding quality increases with the number N_p of putative place cells and remains in the centimeter range for 40 and more remappings if $N_p \geq 500$ (Fig. 8A). At the same time, the place field characteristics deteriorate with increasing N_e as was described in the one-dimensional case (Fig. 6): sparseness decreases (Fig. 8B, F), place field number increases before no clear place fields are visible anymore (Fig. 8C, D, E), place fields loose their similarity to the trained patterns (Fig. 8G).

In the two-dimensional case for few place cells $N_p = 50$, we observe an improvement in resolution when going from one to

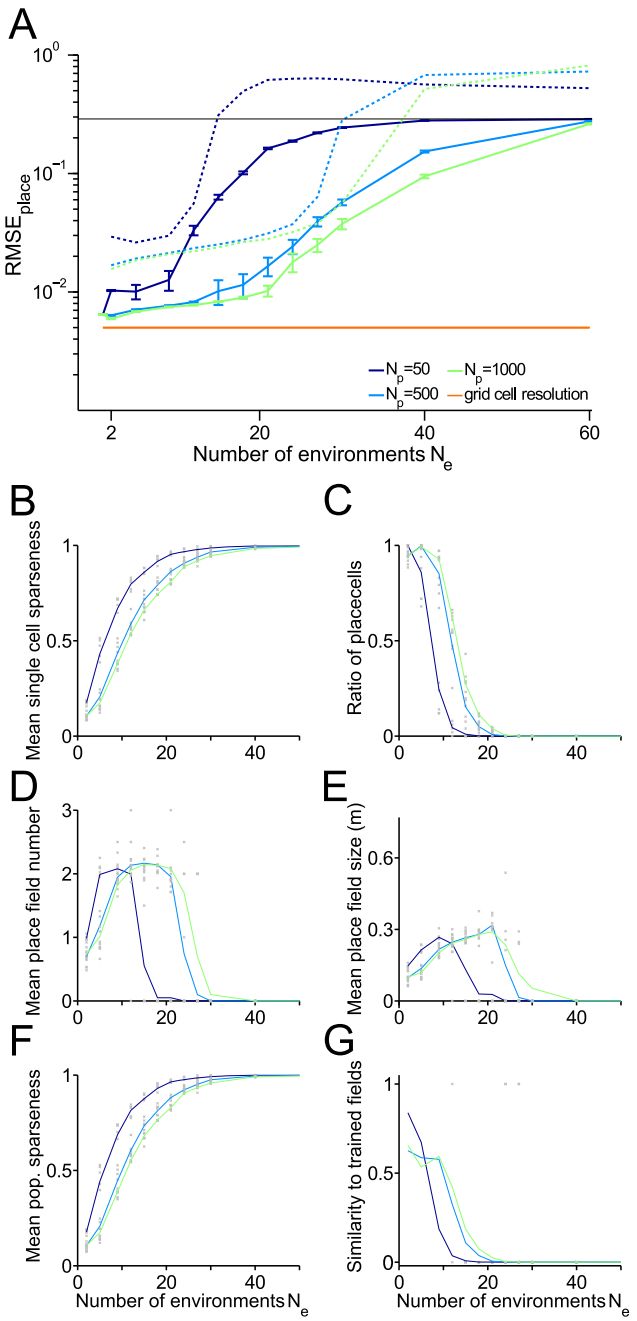

Fig. 6. Capacity for storing remappings on the linear track. Place cell resolution and further measures as functions of the number N_e of remappings stored. (A) Root mean square error (RMSE) of place cells. Blue and green solid lines: Mean over realizations. Dashed lines: 99% quantiles. Red line RMSE of the grid cell input. (B) Mean single cell sparseness. (C) Ratio of proper place cells. (D) Mean number of place fields for proper place cells. (E) Mean size of place fields for proper place cells. (F) Mean population sparseness. (G) Ratio of cells for which Hebbian learning was successful (according to the three similarity criteria defined in the Materials and Methods section). Parameters were $N_g = 400$, $\sigma_p = 0.01$ m, $\sigma_g = 1.038$, $S_p = 2.56$, $S_g = 1.5$, 4 modules, 20 realizations.

doi:10.1371/journal.pcbi.1003986.g006

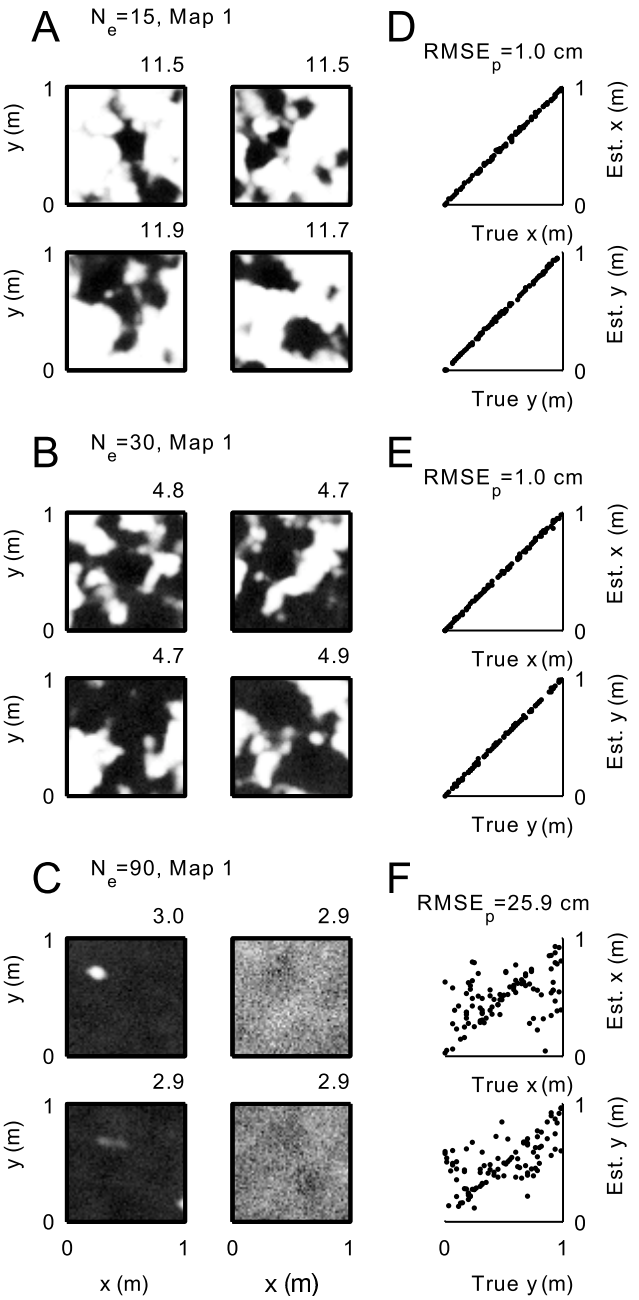

Fig. 7. Quality of 2-d place code, for increasing number of stored environments N_e. (A–C) Rate maps of four example cells for 15, 30, and 90 stored remappings. The desired place field positions (not shown) are identical to Fig. 2 C, but in this case are hardly achieved. (D–F): Minimum mean squared error estimates of position plotted against true position for 500 trials, again for $N_e = 15$, 30 and 90. Parameters as in Fig. 2.

doi:10.1371/journal.pcbi.1003986.g007

about 10 remappings before the decoding error again increases with N_e. Although counter-intuitive, this effect reflects that an increase in mean population sparseness at first provides a better coverage of the square box. To make the model work also for small N_e, the number N_p of place cells has to be large to overcome this finite size effect. It therefore imposes a constraint on a minimum number of N_p. This effect also exemplifies that decoding RMSE depends on many different aspects and thus it is generally difficult to use it as a single measure for comparing the "quality" of a population code.

We also assessed the robustness of our findings with respect to essential model parameters. We evaluated the place code for

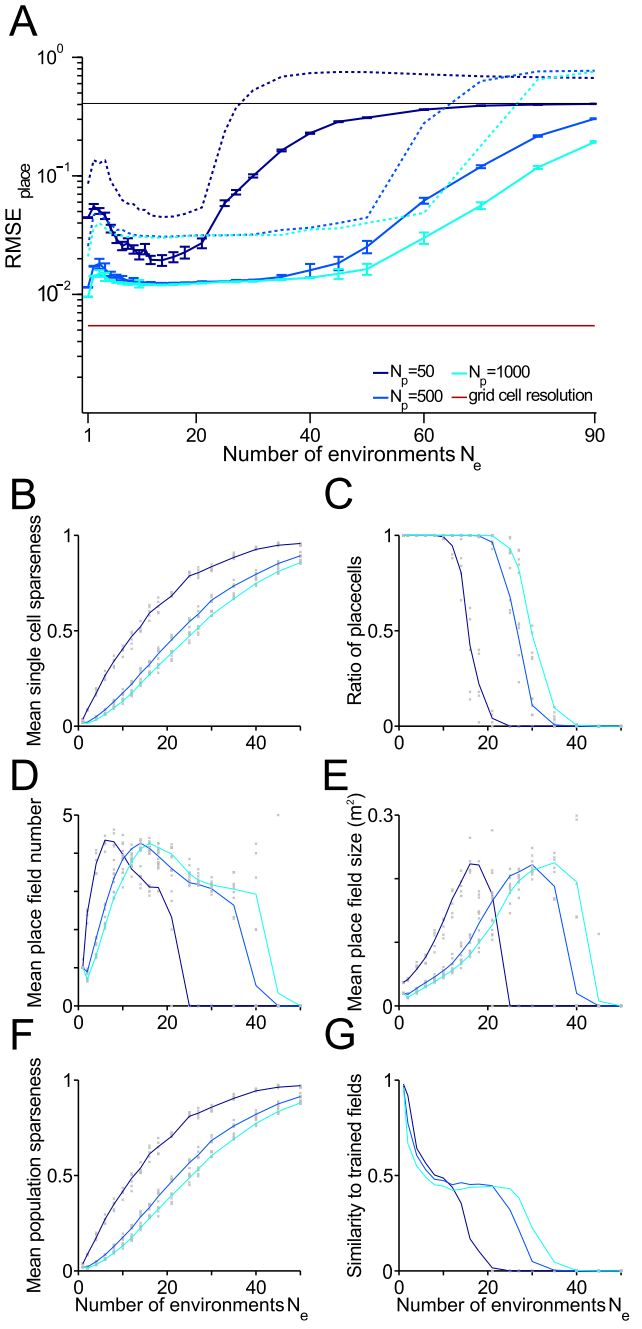

Fig. 8. Capacity for storing remappings in a square box. Place cell resolution and further measures as functions of the number N_e of remappings stored. (A) Root mean square error (RMSE) of place cells. Blue and green solid lines: Mean over realizations. Dashed lines: 99% quantiles. Red line RMSE of the grid cell input. (B) Mean single cell sparseness. (C) Ratio of proper place cells. (D) Mean number of place fields for the proper place cells. (E) Mean size of place fields for the proper place cells. (F) Mean population sparseness. (G) Ratio of cells for which Hebbian learning of place fields was successful (according to the three similarity criteria defined in the Materials and Methods section). Parameter used as before $N_g = 400$, $\sigma_p = 0.01$ m, $\sigma_g = 0.3$, $S_p = 2.56$, $S_g = 1.5$, 4 modules, 15 realizations.
doi:10.1371/journal.pcbi.1003986.g008

different number of grid cells N_g, while keeping a constant total number $S_g N_g$ of input spikes and found essentially no difference (S1 Figure). Also, a mere increase in the number S_p of place field spikes only improves the spatial resolution but does not alter any of the other place field characteristics (S2 Figure).

Direct control of sparseness

A substantial effect on the population code can be observed by altering the strength of feedback inhibition in the place field population by means of the E% value (Fig. 9). This parameter determines the firing threshold as the input strength E% below the maximum (see Methods and [21]). The E% value directly controls the sparseness of the code (Fig. 9B–G). For low E% values (sparse codes) and low numbers N_e of environments, we again observe the finite size effect of high RMSE, which then improves with increasing N_e (Fig. 9A). This initially high RMSE, however, can again be compensated for by using larger numbers N_p of place cells (as in Fig. 8 A). As a result, the decreasing E% generally allows to store more environments, however, at the cost of high N_p to achieve a sufficiently small RMSE for low N_e.

Partial learning

If one constrains the parameter space to biologically realistic mean population sparseness values for the hippocampal place fields about 0.1 to 0.2 (Supporting Information of [24] and [25], see Discussion) our simulations of the standard parameter regime (Fig. 8) show that such a regular place code can only be observed for up to about ten environments. Also for increased E% value the number of sparsely encoded environments is only increased to several tens (Fig. 9). A major factor limiting the number N_e of environments is that in our model the synapses to the place cells are updated in each remapping, i.e., the place cells experience maximal interference. One can considerably extend the number of remappings for a given sparseness if the synaptic changes from different remappings are distributed to varying subsets of place cells, thereby increasing the overall number of putative place cells (partial learning). This strategy is motivated by an experimental report showing that only a small subset of CA1 pyramidal cells shows intracellular determinants for being recruited as a place cell in a novel environment [26]. We illustrate the benefits of partial learning by a further set of simulations in which the synaptic weights to only a fraction f of the place cells are updated in each individual remapping (partial learning; Fig. 10). Using mean population sparseness as a criterion for the breakdown of the place code, partial learning increases the number of possible remappings (Fig. 10A) to over a hundred. As a measure for capacity, one can define a critical number of environments at which the mean population sparseness exceeds a (biologically motivated) threshold value of 0.12 (see Discussion). This critical N_e only weakly increases with the number N_p of place fields but strongly decreases with increasing fraction f of partial learning (Fig. 10B, C).

In rat hippocampus the number N_p of CA1 neurons is in the order of several 100 thousands and thus according to Fig. 10B, a sparse place representation may still be consistent with storing hundreds to thousands of remappings if each place cell is involved in only a small fraction of environments.

The encoding acuity (RMSE) is generally not affected by partial learning as long as N_p is not too small (Fig. 10D). Only for very small values of f, when a winner-take-all effect of the E%-MAX rule *decreases* sparseness for $N_e \to \infty$, spatial acuity deteriorates. However, this regime is biologically unrealistic, since there the number $N_p f$ of neurons encoding an environment tends to zero.

The geometry of the spatial firing patterns (place field size and number), is virtually unaffected by f (Fig. 10 D, E). The place field sizes we find in the model (up to 0.05 m^2) are within the range reported in the experimental literature [25,27], the mean number

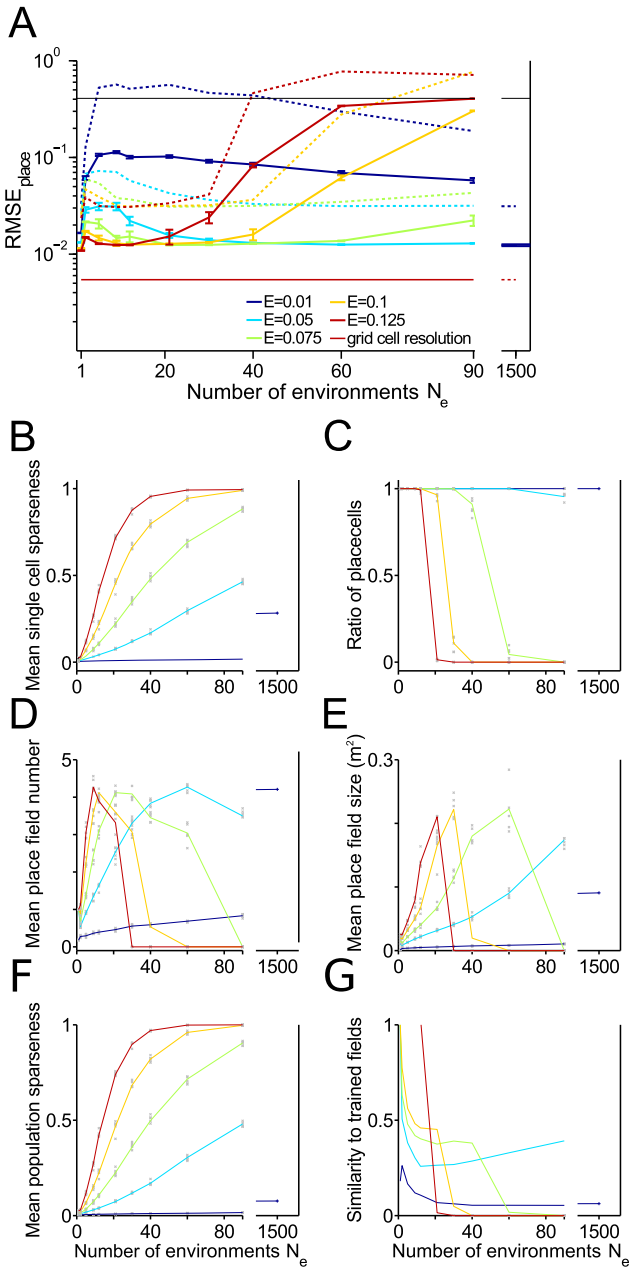

Fig. 9. Effect of the E% parameter on the capacity for storing remappings in a square box for $N_p = 500$ place cells. Place cell resolution and further measures as functions of the number N_e of remappings stored. (A) Root mean square error (RMSE) of place cells. Blue and green solid lines: Mean over realizations. Dashed lines: 99% quantiles. Red line RMSE of the grid cell input. (B) Mean single cell sparseness. (C) Ratio of proper place cells. (D) Mean number of place fields for the proper place cells. (E) Mean size of place fields for the proper place cells. (F) Mean population sparseness. (G) Ratio of cells for which Hebbian learning of place fields was successful (according to the three similarity criteria defined in the Materials and Methods section). Parameter used as before $N_g = 400$, $\sigma_p = 0.01$ m, $\sigma_g = 0.3$, $S_p = 2.56$, $S_g = 1.5$, 4 modules, 8 realizations. The curve for $E\% = 0.1$ is taken from Figure 8 and has 15 realizations.
doi:10.1371/journal.pcbi.1003986.g009

of place fields (about 3) is at the upper bound of the $1-3$ fields per m^2 experimentally found in the hippocampus and dentate gyrus [24,27], which indicates that the place code might in fact even be sparser than than the 0.12 threshold motivated by current experimental data (see Discussion).

Discussion

The hippocampal formation hosts two space representations. A sparse one in the hippocampus proper, in which the neurons have a limited number of distinct firing fields (place fields) and a dense one in the MEC, where grid cells exhibit multiple firing fields located on the nodes of a hexagonal lattice. If both brain regions encode the unique physical spatial position of the animal, the two codes have to be coherent. Anatomically both brain areas are reciprocally connected [5–8] and thus place cell activity will influence grid cell activity and vice versa.

In this paper, we focus on the connections from the medial entorhinal grid cells to the hippocampus, which anatomically correspond to the perforant pathway and the temporo-ammonic pathway. These pathways have initially been thought to predominantly underly the transformation from grid to place cells [19,28–32]. More recently, developmental studies [12,13] and pharmacological interventions that block grid cell firing [7,14–16], have shown that place cells can also be observed independently of grid-field firing (but see [33]). Thus, while the MEC-to-hippocampus connections seem to be unnecessary to generate place fields, they are likely important in synchronizing both codes. This view is further corroborated by the observation that place cell firing is less stable if MEC input is eliminated [34].

Although it is known from information theory that capacity and sparseness cannot be maximized simultaneously [35,36], our paper exemplifies this rule for a specific neuronal network example, in that it shows that maximization of capacity of MEC-to-hippocampal connections destroys the sparseness of the hippocampal place code.

From the theoretical perspective, if the synaptic matrix is know that transforms one code into another, reading out a dense code is more difficult than reading out a sparse code. This is because the synaptic matrix gives rise to a much noisier postsynaptic signal for dense input patterns [37]. Therefore the transformation from place cells to grid cells is less problematic than the other way round. The grid to place transformation provides an interesting test case to study information transfer between different brain areas in general.

Our model is largely based on experimental reports of grid and place cell remapping [18,20,38–40]. While place cells turn on, turn off, or show random relocation during global remapping [40], grid fields shift and rotate. In our model, we consider only shifts, since rotations were shown to be less efficient for remapping previously [19]. Although the grid modules seem to operate functionally independent [17], it is not yet clear whether the modules remap independently as proposed in [19]. A further finding from [19] was that a few ($\gtrsim 2$) modules suffice for strong remapping and data [17] suggest that MEC has only about 5 to 9 modules. Only a part of these modules innervate any one place cell, owing to the dorso-ventrally ordered topography of the input fibers. We therefore concluded that a biologically reasonable number of modules influencing any single place cell is about 4. We further assume that the number of cells per module is constant, which is optimal from a theoretical perspective [9] but might not necessarily be the case [17].

To connect our simulations to hippocampal physiology, we assume a population sparseness value of 0.12. This value can be estimated by combining data from the supporting information (Table S1 of [24]) (mean number of place cells: $1.1/(0.8\text{ m})^2$ for CA3, $2/(0.8\text{ m})^2$ for DG; percentage of place fields: 62/71 for

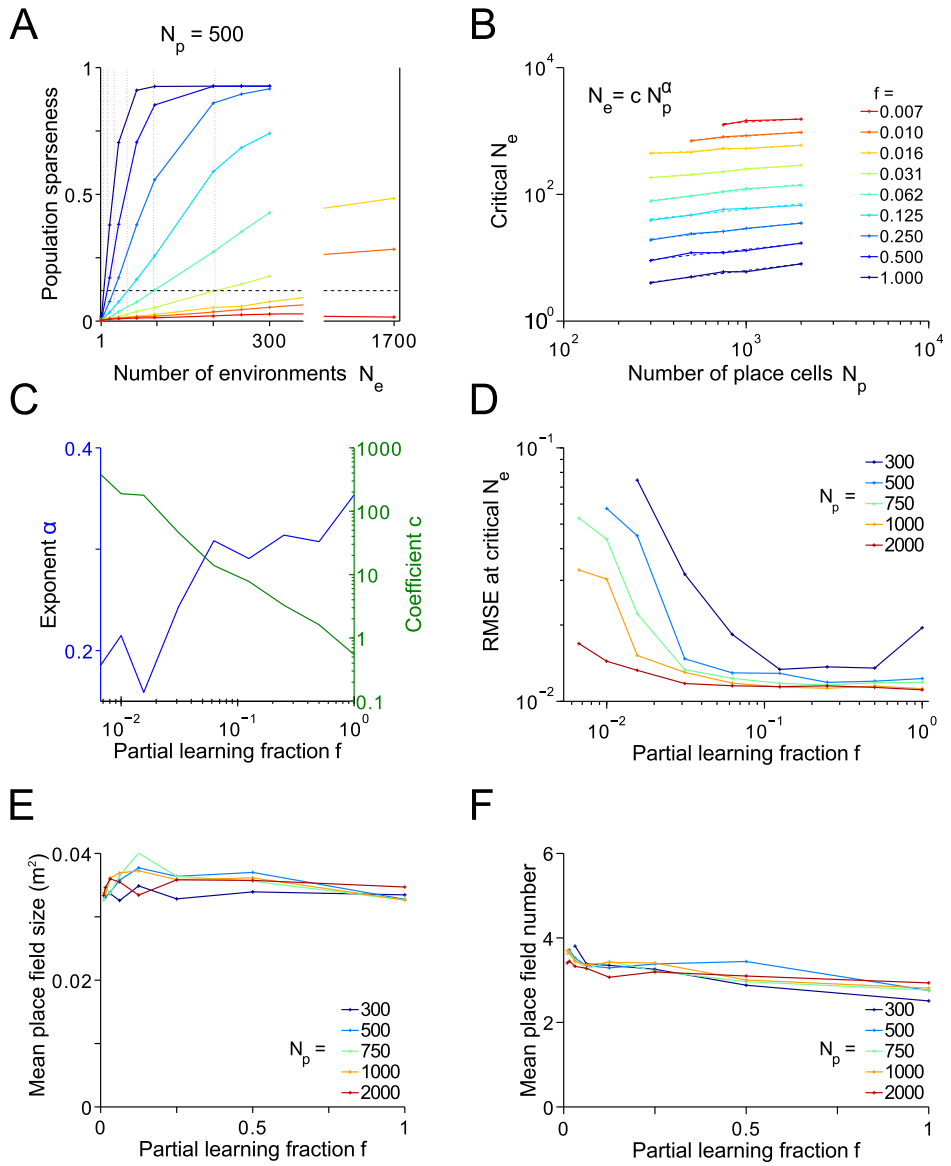

Fig. 10. Partial learning. Effect of place cell number N_p and of the fraction f that are trained to encode one environment on the number of environments N_e. (A) Population sparseness as function of environments N_e stored, for $N_p = 500$ place cells. Different colors represent different fractions for partial learning, see legend in B. The critical value N_e at which sparseness reaches a biologically realistic value of 0.12 is obtained by interpolation. (B) Critical values of N_e as function of place cell number N_p and partial learning fraction f. Data can be fitted by simple logarithmic functions $N_e = cN_p^{\alpha}$. (C) Exponent α and coefficient c of fit from B. (D) Root mean square errors (RMSE) at the critical N_e for the N_p and f in B. (E) and (F): Mean place field size (E) and number (F) at the critical N_e for the N_p and f in B. Averages are over proper place cells.
doi:10.1371/journal.pcbi.1003986.g010

CA3, 41/44 for DG) and place field areas measured in [25] in a circular enclosure of diameter 76 cm (field area: 0.08 m^2 for CA3, 0.06 m^2 for DG). The estimate of the population sparseness for a 1 m^2 enclosure (as in our simulations) thus follows from the product of these three values, i.e., we obtain about 0.12 for CA3 and 0.17 for DG. However, in our simulations, a sparseness value of 0.12 yields a number of place fields per place cell that is slightly higher than observed in experiments, and thus the above numbers may over-estimate the sparseness values in the real rodent brain.

Previous coding theories of MEC grid cells have extensively investigated spatial resolution. According to [9,41], hierarchical grid codes outperform place codes by far in terms of their scaling behavior. A main reason is that for a constant resolution, the number of place cells scales with area, whereas for grid cells only those with larger period have to be scaled up with area for disambiguation, however, the resolution mostly rests on the smallest grid periodicity and thus the size of the population with small periodicity is independent of spatial range to be encoded. The parameter regimes in which grid codes are particularly superior to place codes provide relative root mean square errors in the range of 10^{-3} and even far below [9]. For a one meter environment, this would correspond to (sub-)millimeter resolution which is biologically irrelevant for encoding but might be important for MEC models of path integration [42,43] where errors can accumulate over time. In the regime used for the present model (Figs. 3 and 4), the surplus in resolution of the grid

code is relatively small, consistent with a biologically relevant decoding situation of high noise and few modules [44].

A further noteworthy result of our simulations is that a population code still contains almost maximal space information (in terms of minimal RMSE), even if no clear spatial firing fields can be delineated anymore. On the one hand this shows that also brain areas like the lateral entorhinal cortex [45] and the subiculum [46] with only weakly space-modulated individual neurons can provide high-resolution space information on the population level and thus a superposition of such weakly modulated firing fields via synaptic inputs is sufficient to provide place information to any downstream structure. This means that also the hippocampus and the MEC may not generate their strongly spatially modulated firing fields de-novo but inherit them from weakly modulated populations as e.g. the lateral entorhinal cortex. On the other hand our findings show that sparseness of the hippocampal place representation is not due to coding precision requirements but must serve other purposes. Manifold advantages of sparseness have been proposed [47] including energy efficiency [48]. A further classical benefit of sparse representations arises for auto-associative memory networks, where it facilitates memory retrieval due to reduced interference [37,49–52].

Although our model includes lateral inhibition via the E% rule to limit the overall network activity the network cannot enforce sparseness except for unrealistically low values of f. So it is still possible that other assumptions about the recurrent connections may enforce sparseness more effectively, while allowing remappings. For example, in a model using a combination of recurrent excitation and inhibition [53,54] place fields arise from stable attractor states, where each attractor reflects the topology of place field positions for one remapping. The capacity (number of remappings per neuron) of this autoassociator is in the range of few percent and, thus for $N_p = 1000$ may end up slightly above the capacity derived from our model (≈ 10) (for fixed realistic sparseness). So, recurrent excitatory connections between place cells can potentially help to keep the place fields compact. The disadvantage of attractor-like solutions is that they show catastrophic forgetting, whereas our model exhibits a gradual decline of the order parameters (Figs. 6, 8 and 9).

The view on how space information is communicated between the reciprocally connected brain areas hippocampus and MEC has recently undergone a dramatic change from a completely feed-forward grid-to-place dogma [19,28–32] to an almost reversed place-to-grid picture [7,12–16]. We started out under the assumption that the spatial precision in the hippocampus mostly relies on inputs from MEC grid cells and remapping the MEC triggers remapping on the hippocampus. If this was the only function of the MEC-to-hippocampus connections, they should be filled with as much space information as possible and the representation would no longer be sparse. Our results thus show that functionally the classical pure grid-to-place hypothesis would only suboptimally use the coding resources. The required compact place fields and the MEC-to-hippocampus synapses thus do not seem to be optimized to transfer space information.

Since new experimental data [7,12–16] show that MEC is actually not essential for generating place cells, our findings suggest the possibility that hippocampal space information might actually primarily stem from other regions than the MEC. The grid field input to place fields thus likely imposes only modulatory or stabilizing effects. Conversely, no grid cells have been so far observed without place cell activity, and thus the place-to-grid hypothesis is still a possible candidate. However, it is unclear why hexagonal symmetry might emerge from the perspective of a transformation of a sparse place code to a dense code, and thus it might as well be that the two codes are generated independently for different computational purposes and the reciprocal connections are only required for synchronization and stabilization.

Materials and Methods

Grid cell firing rate maps in one dimension

The N_g grid cells are modeled as Poisson spikers with firing maps $R_i(x)$ that denote the mean spike count of cell $i = 1 \dots N_g$ conditioned on the position $x \in [0,1]$ on a 1 meter track. All cells have the same maximal spike count C_g and the same field width parameter σ_g. The cells differ in their spatial periods λ_i and grid phases φ_i. The specific model for the cells' Poisson spike counts follows a von Mises function:

$$R_i(x) = C_g \exp\left(\frac{\cos(2\pi/\lambda_i(x-\varphi_i))-1}{\sigma_g^2}\right).$$

Each cell belongs to one of M modules. Cells in a module share a spatial period λ_i. The phases φ_i in each module are chosen equidistantly such that the firing fields cover the linear track; Fig. 1A.

Though we have only one width parameter σ_g for all cells, the tuning width $\sigma_g \lambda_i/(2\pi)$ for the cells in one specific module scales with the period λ_i, as can be seen from expanding the cosine term in $R_i(x)$.

The spike count C_g is adjusted such that the whole grid cell population generates a constant given number S of spikes averaged over all positions x and cells i, i.e.,

$$S = \langle \mathrm{poiss}(R_i(x)) \rangle_{i \in \{1,\dots,N_g\}; x \in [0,1]} \approx \frac{\sum_{i,b} R_i(x_b)}{N_g B} \qquad (1)$$

Here, the locations x are discretized in $B = 10^4$ bins x_b. The value used for S is 1.5 spikes per cell. Since for Poisson spikers the spike count is a product of averaging duration, firing rate and number of cells with the same rate function R_i, the three factors cannot be distinguished. Although, for simplicity, we call N_g the number of grid cells, it is more correctly referred to as the number of grid cell channels (different rate functions R_i).

The different modules are defined by their grid period λ_i. In our grid cell population, the first module is assigned the largest spatial period, which we take $\lambda_1 = (1 + 0.4\,\sigma_g)$ m such that each cell in this module only has one unique firing field on the track. The smaller periods of the other modules are obtained via geometric progression, $\lambda_{m+1} = \frac{\lambda_m}{r_\lambda}$, with a period ratio r_λ, and $m = 1, \dots M$. The period ratio $r_\lambda = (\lambda_1/\lambda_M)^{1/(M-1)}$ is defined via the number M of modules and the smallest period λ_M, which is set to 30 cm, a lower bound suggested by experiments [3,17]. Thus the only remaining degrees of freedom for the grid code are the number M of modules, the width constant σ_g and the mean spike count per length S. We choose $\sigma_g = 1$, $M = 4$ and $S_g = 1.5$ unless otherwise mentioned.

Hebbian learning of place cells

The synaptic weights w_{ij} of the feed forward connections from grid to place cells are set by Hebbian learning based on the rate maps $R_i(x)$ of the grid cells from eq. (1) and the desired rate maps

$$D_i(x) = \exp\left(-\frac{(x-c_i)^2}{2\sigma_p^2}\right) \quad (2)$$

of the place cells with width σ_p and centers c_i that uniformly cover the interval $[-\sigma_p, 1m+\sigma_p]$; Fig. 1C.

With these idealized place fields, the weights are calculated according to outer product (Hebbian) rule: using discretized locations x_b, $b=1,\ldots,B$ we define

$$w_{ij} = \frac{\sum_{b=1}^{B} D_i(x_b)\, R_j(x_b)}{\sum_{b=1}^{B} D_i(x_b)} \ . \quad (3)$$

The denominator ensures that connections to place cells with fields at the borders are as strong as the ones to centered place fields.

Remapping

The two networks (grid and place cells) are supposed to encode N_e environments. Each environment has a new grid code generated by shifting each module's phases by a constant $s_m, \in [0, \lambda_m]$, $m=1,\ldots,M$. These shifts have experimentally been shown to be coherent within one module [18] and have been theoretically proposed to be uncorrelated between modules [19]. The shifted grid field patterns are denoted by $R_i^{(e)}(x)$. A new place code $D_i^{(e)}(x)$ is generated by randomly choosing the place field centers c_i. Hebbian learning as in eq. 3 is repeated N_e times and weights are added.

Place cell spikes and position decoding

The place cell spikes for cell i at a position x are produced by drawing Poisson spikes $k_j = \text{poisson}\left(R_j(x)\right)$ for the grid cells, then taking the weighted sum

$$U_i = \sum_{j=1}^{N_g} w_{ij} k_j$$

of those, to yield a membrane potential of the place cells. The activity is then generated following the E%-MAX rule [21], that emulates the effect of recurrent inhibition: after finding the maximum membrane potential $U_{\max} = \max_i(U_i)$, all $U_i < 0.9\, U_{\max}$ are set to zero and the ones above this threshold are multiplied with a constant C_p, and used as place cell firing rate from which spike counts q_i are derived according to Poisson statistics.

Decoding the place code via a minimum mean square estimator [55]

$$\hat{x} = \langle x \rangle_{\{q_i\}} = \int \mathrm{d}x\, x\, p(x|\{q_i\}) \quad (4)$$

requires a statistical model $p(x|\{q_i\})$ of place cell firing. Since in the model the single trial spike counts q_i are statistically independent the posterior can be obtained using Bayes' rule,

$$p(x|\{q_i\}) = \frac{\Pi_i\, p(q_i|x) p(x)}{\Pi_i \int \mathrm{d}x\, p(q_i|x) p(x)} \ .$$

The prior is taken as constant, $p(x) = 1/m$. The individual likelihoods $p(q_i|x)$ are obtained by repeating the above stochastic process 800 times for each cell and each sampled position and sampling the relative frequencies of spike counts q_i. This distribution is then fitted with a bimodal model function consisting of a probability A_i of cell i not firing, and probability of firing q_i spikes following a normal distribution with fit parameters mean $\mu_i(x)$ and variance $\sigma_i(x)$:

$$p(q_i|x) = A_i(x)\delta(q_i) + (1 - A_i(x))\,\text{norm}(q_i, \mu_i(x), \sigma_i(x)) \ . \quad (5)$$

Examples for such fits are shown in Fig. 11. Again, the constant C_p is obtained by fixing the number S of spikes per centimeter per cell in an iterative fashion. The resulting value is $S = 2.56$ unless otherwise mentioned.

Two-dimensional place code

For comparison we also implemented the model in two spatial dimensions $\vec{x} \in [0,1]^2$. There, the grid cell's firing maps are set as in [31]

$$R_i(\vec{x}, \lambda, \theta, \vec{c}_i) = C_g\, g\left(\sum_{k=1}^{3} \cos\left(\frac{4\pi\, \vec{u}(\theta_k + \theta) \cdot (\vec{x} - \vec{c}_i)}{\sqrt{3}\lambda_m}\right)\right),$$

with $u(\theta_k) = (\cos(\theta_k), \sin(\theta_k))$ being a unitary vector pointing into direction θ_k. Using $\theta_1 = -30$, $\theta_2 = +30$ and $\theta_3 = +90$, the three spatial waves add up to a hexagonal firing pattern with spatial period λ_m, a maximum at $\vec{c}_i$, and orientation θ (Fig. 2A). The nonlinearity $g(y) = \exp(0.3(y+1.5)) - 1$ both adjusts the minimal firing rate to zero and matches the spatial decay of the firing rate peaks to experiments [31]. Like for the one-dimensional simulations we use four modules. Cells in one module share spatial

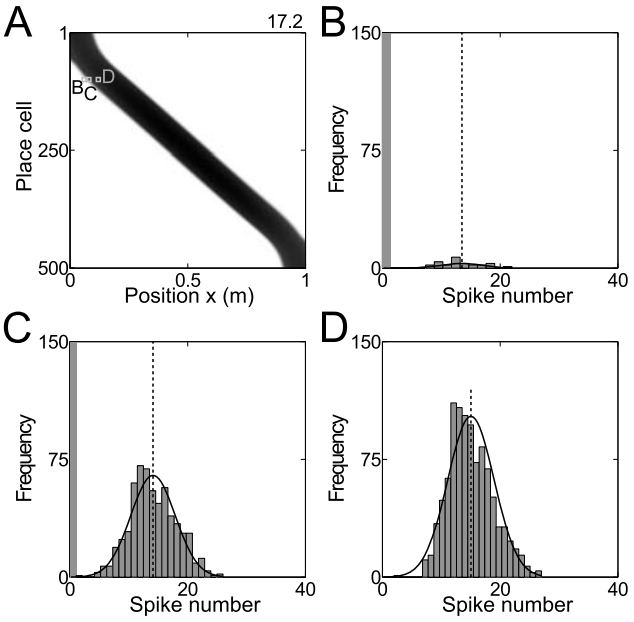

Fig. 11. Spike count likelihood of place cells. (A) Firing rates (gray code) of model place cells as a function of position. (B–D) Simulated spike counts and fits of the model function eq. (5) for examples indicated in A.
doi:10.1371/journal.pcbi.1003986.g011

period and orientation. The period for the first module is $\lambda_1 = 1.42$ m (larger than the box). The smallest period is set to 0.3 m. The two intermediate periods are again obtained by geometric progression $\lambda_{m+1} = \frac{\lambda_m}{r_\lambda}$. Orientation θ for each module is drawn at random. The "centers" $\vec{c}_i$ are uniformly distributed over the Wigner cell of size λ. For all computational purposes, we used 100×100 spatial bins to discretize the box.

To generate two-dimensional place fields we set feed-forward weights by Hebbian learning, using Gaussian tuning curves as firing maps for place fields as in eq. (2), but with x and c_i replaced by their two-dimensional counterparts (Fig. 2 B, C). The centers $\vec{c}_i$ cover the box uniformly on a square grid. Centers of teacher place fields for cell exceeding the number of nodes on the square lattice were distributed randomly. Weights are then calculated using eq. (3).

The spikes are produced as in the one-dimensional case. Decoding follows eq. (4) with one-dimensional quantities replaced by their two-dimensional counterparts.

For a remapping, each grid cell module is assigned one random spatial shift vector, added to all $\vec{c}_i$ from that module. The shift is obtained by drawing a vector from the Wigner cell of that module using a uniform distribution (Fig. 2 D). For remapping, the place cells are assigned new centers at random, which again cover the box equidistantly. Then Hebbian learning is repeated, adding to the existing weights (Fig. 2 E, F).

Partial learning

Partial learning as used in the simulations of Fig. 10 was implemented as follows. For each environment we selected a random set of $f N_p$ cells such that each cell is selected approximately the same amount of times across environments. This was achieved via random permutations of the cell indices. The sets of $f N_p$ cells were taken from such a random index sequence one after the other, and only if less than $f N_p$ items were left in the index sequence, a new random permutation was generated.

For each set of $f N_p$ selected cells we defined teacher place fields that cover the whole environment as uniformly as possible on a square grid with $\lfloor \sqrt{(N_p f)} \rfloor^2$ nodes (see previous section). Hebbian learning according to eq. (3) was applied to only the synapses between the grid field population and the selected set of postsynaptic cells.

By construction, some place cells will be used in more environments than others. We normalize the rows of w_{ij} after all environments have been learned to avoid that the cells that are involved in more environments (and thus have larger weights) are overly excited and exert too much inhibition on the remaining cells via the E%-MAX rule.

Single cell sparseness

According to [23], single cell sparseness is defined as $\langle R \rangle^2 / \langle R^2 \rangle$, where $R(x)$ denotes the firing rate of the specific cell as a function of position x and $\langle \cdot \rangle$ indicates the average over space.

Population sparseness

Population sparseness is defined as the percentage of place cells firing above a threshold of 20% of the maximum firing rate at any position.

Detection of (proper) place fields

The number and size of place fields was found by first thresholding the rate maps, discarding all bins below 20% of the maximal rate, and then applying the algorithm by Hoshen and Kopelman [56]. Bins were considered neighboring if they share an edge, hence diagonal bins were not neighbors. Place fields were only included in the analysis (proper place fields) if they were larger than 50 cm^2 and smaller than 60% of the total environment.

Success of Hebbian learning by similarity

Learning of place fields was considered successful in a cell if the learned field showed sufficient similarity to the training field according to three criteria: 1) the total area above a threshold of 20% peak rate has to be smaller than $0.6 m^2$, 2) the place field center has to be detected close to the desired location, i.e., no further away than the place field radius ($\sqrt{\text{area}/\pi}$), and 3) the desired place field has to have an area at least twice the size of all other place fields.

Supporting Information

S1 Figure Effect of the place cell spike number S_p on the capacity for storing remappings in a square box. Place cell resolution and further measures as functions of the number N_e of remappings stored for $N_p = 500$. (A) Root mean square error (RMSE) of place cells. Blue and green solid lines: Mean over realizations. Dashed lines: 99% quantiles. Red line RMSE of the grid cell input. (B) Mean single cell sparseness. (C) Ratio of proper place cells. (D) Mean number of place fields for the proper place cells. (E) Mean size of place fields for the proper place cells. (F) Mean population sparseness. (G) Ratio of cells for which Hebbian learning of place fields was successful (according to the three similarity criteria defined in the Materials and Methods section). Parameter used as before $N_g = 400$, $\sigma_p = 0.01$ m, $\sigma_g = 0.3$, $S_p = 2.56$, $S_g = 1.5$, 4 modules, 15 realizations, 10 for $S_p = 0.64$.
(EPS)

S2 Figure Effect of varying grid cell number N_g and grid cell spike count S_g with constant $N_g S_g$ on the capacity for storing remappings in a square box. Place cell resolution and further measures as functions of the number N_e of remappings stored for $N_p = 500$. (A) Root mean square error (RMSE) of place cells. Blue and green solid lines: Mean over realizations. Dashed lines: 99% quantiles. Red line RMSE of the grid cell input. (B) Mean single cell sparseness. (C) Ratio of proper place cells. (D) Mean number of place fields for the proper place cells. (E) Mean size of place fields for the proper place cells. (F) Mean population sparseness. (G) Ratio of cells for which Hebbian learning of place fields was successful (according to the three similarity criteria defined in the Materials and Methods section). Parameters used are as before $N_g = 400$, $\sigma_p = 0.01$ m, $\sigma_g = 0.3$, $S_p = 2.56$, $S_g = 1.5$, 4 modules, 7 realizations, 15 for $N_g = 400$, $S_g = 1.5$, data from Fig. 8.
(EPS)

Acknowledgments

We would like to thank Alexander Mathis for discussions.

Author Contributions

Conceived and designed the experiments: AK CL. Performed the experiments: AK. Analyzed the data: AK. Wrote the paper: AK CL.

References

1. O'Keefe J, Nadel L (1978) The Hippocampus as a Cognitive Map. Oxford University Press.
2. Fyhn M, Molden S, Witter MP, Moser EI, Moser MB (2004) Spatial representation in the entorhinal cortex. Science 305: 1258–1264.
3. Hafting T, Fyhn M, Molden S, Moser MB, Moser EI (2005) Microstructure of a spatial map in the entorhinal cortex. Nature 436: 801–806.
4. Hartley T, Lever C, Burgess N, O'Keefe J (2014) Space in the brain: how the hippocampal formation supports spatial cognition. Philos Trans R Soc Lond, B, Biol Sci 369: 20120510.
5. Canto CB, Wouterlood FG, Witter MP (2008) What does the anatomical organization of the entorhinal cortex tell us? Neural Plast 2008: 381243.
6. Zhang SJ, Ye J, Couey JJ, Witter M, Moser EI, et al. (2014) Functional connectivity of the entorhinal-hippocampal space circuit. Philos Trans R Soc Lond, B, Biol Sci 369: 20120516.
7. Bonnevie T, Dunn B, Fyhn M, Hafting T, Derdikman D, et al. (2013) Grid cells require excitatory drive from the hippocampus. Nat Neurosci 16: 309–317.
8. Zhang SJ, Ye J, Miao C, Tsao A, Cerniauskas I, et al. (2013) Optogenetic dissection of entorhinal-hippocampal functional connectivity. Science 340: 1232627.
9. Mathis A, Herz AV, Stemmler M (2012) Optimal population codes for space: grid cells outperform place cells. Neural Comput 24: 2280–2317.
10. Fiete IR, Burak Y, Brookings T (2008) What grid cells convey about rat location. J Neurosci 28: 6858–6871.
11. Kubie JL, Muller RU (1991) Multiple representations in the hippocampus. Hippocampus 1: 240–242.
12. Wills TJ, Cacucci F, Burgess N, O'Keefe J (2010) Development of the hippocampal cognitive map in preweanling rats. Science 328: 1573–1576.
13. Langston RF, Ainge JA, Couey JJ, Canto CB, Bjerknes TL, et al. (2010) Development of the spatial representation system in the rat. Science 328: 1576–1580.
14. Koenig J, Linder AN, Leutgeb JK, Leutgeb S (2011) The spatial periodicity of grid cells is not sustained during reduced theta oscillations. Science 332: 592–595.
15. Brandon MP, Bogaard AR, Libby CP, Connerney MA, Gupta K, et al. (2011) Reduction of theta rhythm dissociates grid cell spatial periodicity from directional tuning. Science 332: 595–599.
16. Brandon MP, Koenig J, Leutgeb JK, Leutgeb S (2014) New and Distinct Hippocampal Place Codes Are Generated in a New Environment during Septal Inactivation. Neuron 82: 789–796.
17. Stensola H, Stensola T, Solstad T, Froland K, Moser MB, et al. (2012) The entorhinal grid map is discretized. Nature 492: 72–78.
18. Fyhn M, Hafting T, Treves A, Moser MB, Moser EI (2007) Hippocampal remapping and grid realignment in entorhinal cortex. Nature 446: 190–194.
19. Monaco JD, Abbott LF, Abbott LF (2011) Modular realignment of entorhinal grid cell activity as a basis for hippocampal remapping. J Neurosci 31: 9414–9425.
20. Leutgeb S, Leutgeb JK, Barnes CA, Moser EI, McNaughton BL, et al. (2005) Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. Science 309: 619–623.
21. de Almeida L, Idiart M, Lisman JE (2009) A second function of gamma frequency oscillations: an E%-max winner-take-all mechanism selects which cells fire. J Neurosci 29: 7497–7503.
22. Bethge M, Rotermund D, Pawelzik K (2002) Optimal short-term population coding: when Fisher information fails. Neural Comput 14: 2317–2351.
23. Treves A, Rolls ET (1992) Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. Hippocampus 2: 189–199.
24. Leutgeb JK, Leutgeb S, Moser MB, Moser EI (2007) Pattern separation in the dentate gyrus and CA3 of the hippocampus. Science 315: 961–966.
25. Park E, Dvorak D, Fenton AA (2011) Ensemble place codes in hippocampus: CA1, CA3, and dentate gyrus place cells have multiple place fields in large environments. PLoS ONE 6: e22349.
26. Epsztein J, Brecht M, Lee AK (2011) Intracellular determinants of hippocampal CA1 place and silent cell activity in a novel environment. Neuron 70: 109–120.
27. Jung MW, McNaughton BL (1993) Spatial selectivity of unit activity in the hippocampal granular layer. Hippocampus 3: 165–182.
28. Solstad T, Moser EI, Einevoll GT (2006) From grid cells to place cells: a mathematical model. Hippocampus 16: 1026–1031.
29. Blair HT, Welday AC, Zhang K (2007) Scale-invariant memory representations emerge from moir interference between grid fields that produce theta oscillations: a computational model. J Neurosci 27: 3211–3229.
30. Molter C, Yamaguchi Y (2008) Impact of temporal coding of presynaptic entorhinal cortex grid cells on the formation of hippocampal place fields. Neural Netw 21: 303–310.
31. de Almeida L, Idiart M, Lisman JE (2009) The input-output transformation of the hippocampal granule cells: from grid cells to place fields. J Neurosci 29: 7504–7512.
32. Cheng S, Frank LM (2011) The structure of networks that produce the transformation from grid cells to place cells. Neuroscience 197: 293–306.
33. Azizi AH, Schieferstein N, Cheng S (2014) The transformation from grid cells to place cells is robust to noise in the grid pattern. Hippocampus 24: 912–919.
34. Schlesiger MI, Cannova CC, Mankin EA, Boublil BB, Hales JB, et al. (2013) The medial entorhinal cortex is required for hippocampal phase precession. In: Society for Neuroscience Meeting. 578.29/KKK68.
35. Cover T, A TJ (1991) Elements of information theory. New York: Wiley.
36. Rolls E, Treves A (1998) Neural Networks and Brain Function. Oxford: Oxford University Press.
37. Willshaw DJ, Buneman OP, Longuet-Higgins HC (1969) Non-holographic associative memory. Nature 222: 960–962.
38. Muller RU, Kubie JL (1986) The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. J Neurosci 7: 1951–1968.
39. Bostock E, Muller RU, Kubie JL (1991) Experience-dependent modifications of hippocampal place cell firing. Hippocampus 1: 193–205.
40. Leutgeb S, Leutgeb JK, Treves A, Moser MB, Moser EI (2004) Distinct ensemble codes in hippocampal areas CA3 and CA1. Science 305: 1295–1298.
41. Mathis A, Herz AV, Stemmler MB (2012) Resolution of nested neuronal representations can be exponential in the number of neurons. Phys Rev Lett 109: 018103.
42. Fuhs MC, Touretzky DS (2006) A spin glass model of path integration in rat medial entorhinal cortex. J Neurosci 26: 4266–4276.
43. Burgess N, Barry C, O'Keefe J (2007) An oscillatory interference model of grid cell firing. Hippocampus 17: 801–812.
44. Mathis A, Herz AV, Stemmler MB (2013) Multiscale codes in the nervous system: the problem of noise correlations and the ambiguity of periodic scales. Phys Rev E Stat Nonlin Soft Matter Phys 88: 022713.
45. Hargreaves EL, Rao G, Lee I, Knierim JJ (2005) Major dissociation between medial and lateral entorhinal input to dorsal hippocampus. Science 308: 1792–1794.
46. Kim SM, Ganguli S, Frank LM (2012) Spatial information outflow from the hippocampal circuit: distributed spatial coding and phase precession in the subiculum. J Neurosci 32: 11539–11558.
47. Olshausen BA, Field DJ (2004) Sparse coding of sensory inputs. Curr Opin Neurobiol 14: 481–487.
48. Levy WB, Baxter RA (1996) Energy efficient neural codes. Neural Comput 8: 531–543.
49. Tsodyks M, Feigel'man M (1988) Enhanced storage capacity in neural networks with low level of activity. Europhys Lett 6: 101–105.
50. Nadal JP (1991) Associative memory: on the (puzzling) sparse coding limit. J Phys A 24: 1093–1101.
51. Leibold C, Kempter R (2006) Memory capacity for sequences in a recurrent network with biological constraints. Neural Comput 18: 904–941.
52. Kammerer A, Tejero-Cantero A, Leibold C (2013) Inhibition enhances memory capacity: optimal feedback, transient replay and oscillations. J Comput Neurosci 34: 125–136.
53. Monasson R, Rosay S (2013) Crosstalk and transitions between multiple spatial maps in an attractor neural network model of the hippocampus: Phase diagram. Phys Rev E 87: 062813.
54. Monasson R, Rosay S (2014) Crosstalk and transitions between multiple spatial maps in an attractor neural network model of the hippocampus: Collective motion of the activity. Phys Rev E 89: 032803.
55. Lehmann E, Casella C (1998) Theory of Point Estimation. Springer, New York.
56. Hoshen J, Kopelman R (1976) Percolation and cluster distribution. I. cluster multiple labeling technique and critical concentration algorithm. Phys Rev B 14: 3438–3445.

Supporting materials

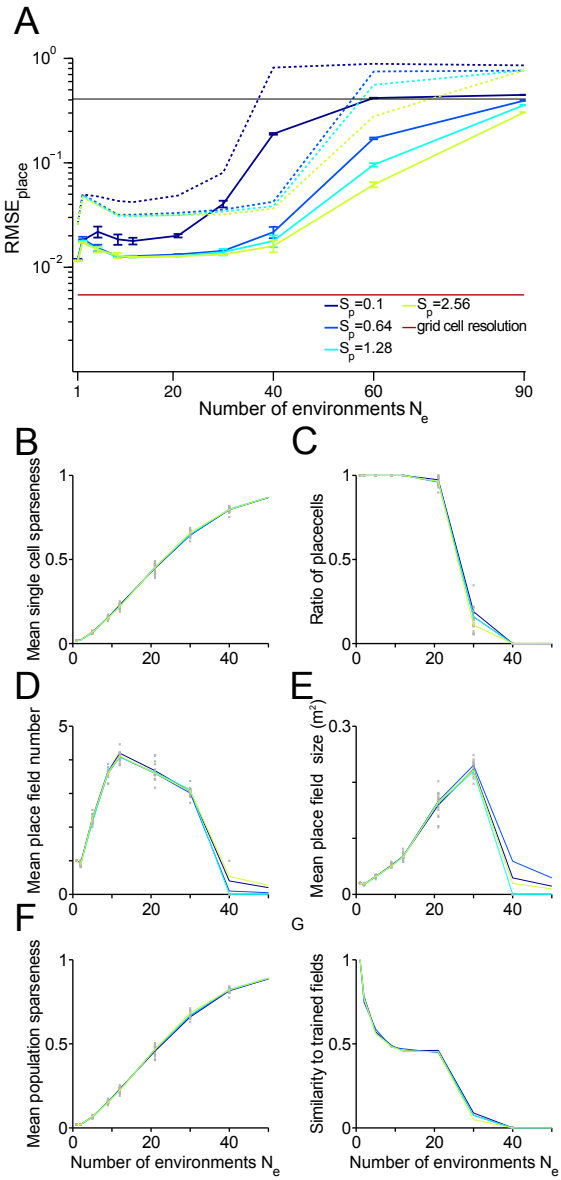

FIGURE 3.1: Effect of the place cell spike number S_p on the capacity for storing remappings in a square box. Place cell resolution and further measures as functions of the number N_e of remappings stored for $N_p = 500$. (A) Root mean square error (RMSE) of place cells. Blue and green solid lines: Mean over realizations. Dashed lines: 99% quantiles. Red line RMSE of the grid cell input. (B) Mean single cell sparseness. (C) Ratio of proper place cells. (D) Mean number of place fields for the proper place cells. (E) Mean size of place fields for the proper place cells. (F) Mean population sparseness. (G) Ratio of cells for which Hebbian learning of place fields was successful (according to the three similarity criteria defined in the Materials and Methods section). Parameter used as before $N_g = 400$, $\sigma_p = 0.01$m, $\sigma_g = 0.3$, $S_p = 2.56$, $S_g = 1.5$, , 4 modules, 15 realizations, 10 realizations for $S_p = 0.64$. Adapted from Kammerer and Leibold (2014) with permission.

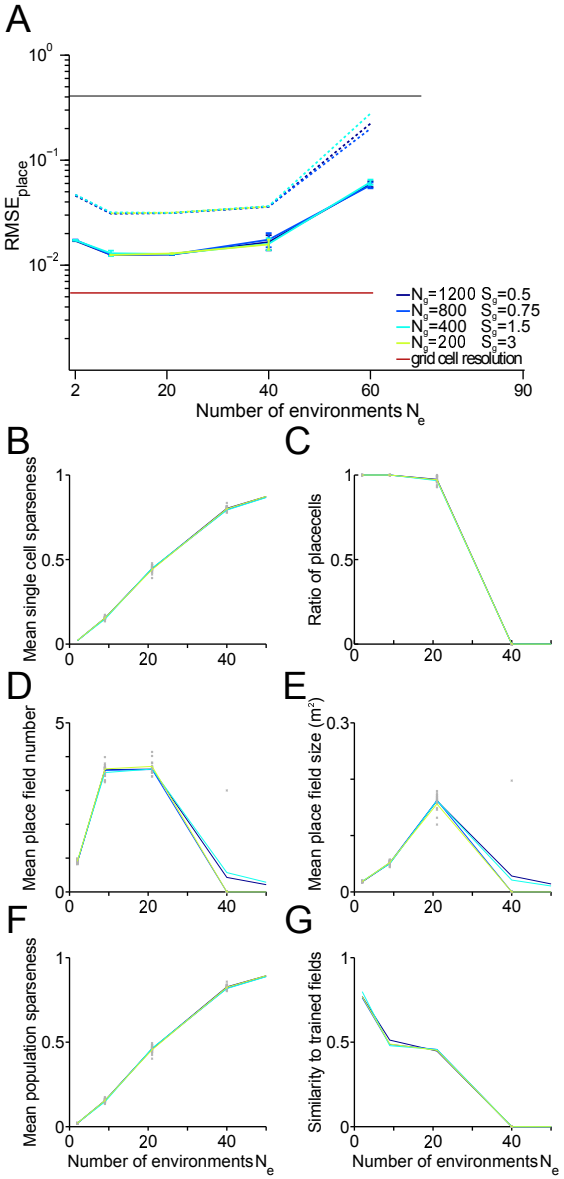

FIGURE 3.2: Effect of varying grid cell number N_g and grid cell spike count S_g with constant $N_g S_g$ on the capacity for storing remappings in a square box. Place cell resolution and further measures as functions of the number N_e of remappings stored for $N_p = 500$. (A) Root mean square error (RMSE) of place cells. Blue and green solid lines: Mean over realizations. Dashed lines: 99quantiles. Red line RMSE of the grid cell input. (B) Mean single cell sparseness. (C) Ratio of proper place cells. (D) Mean number of place fields for the proper place cells. (E) Mean size of place fields for the proper place cells. (F) Mean population sparseness. (G) Ratio of cells for which Hebbian learning of place fields was successful (according to the three similarity criteria defined in the Materials and Methods section). Parameters used are as before $\sigma_p = 0.01$m, $\sigma_g = 0.3$, $S_p = 2.56$, 4 modules, 7 realizations, 15 realizations for $N_g = 400$, $S_g = 1.5$, data from Fig. 8. Adapted from Kammerer and Leibold (2014) with permission.

Chapter 4

Discussion

In this chapter, the results are summarized and compared to the most relevant experimental findings. Furthermore, the shortcomings of the models used in this work are discussed. Other models are reviewed that give clues on how to resolve those issues. Fig. 4.1 shows a schematic of sequence replay in HC that puts the two mechanisms modeled in this thesis into context. During learning, pattern separation mechanisms select neural assemblies with little overlap used in the replay of a memory sequence. One such pattern separation mechanism takes advantage of shifts in MEC grid field phase. We investigated the limits and scaling behavior of the MEC pattern separation mechanism in chapter 3. The model, by its spatial nature, also tests aspects of the synchronization of spatial codes in MEC and HC.

For the replay phase (Fig. 4.1), we built on an existing model (Leibold and Kempter, 2006), further improving its memory capacity by adding an inhibitory population. We found that inhibition can improve the capacity of a network replaying sequences (Kammerer et al., 2013). At the capacity limit, replay of stored sequences was reduced to short transient sequences. Both the length of these sequences as well as the oscillations in inhibitory feedback are consistent with reactivation of place cell assemblies during sharp wave ripples.

In the following, we will compare the sequences of our sequence replay model (Kammerer et al., 2013), which use random neuron assemblies with few overlaps,

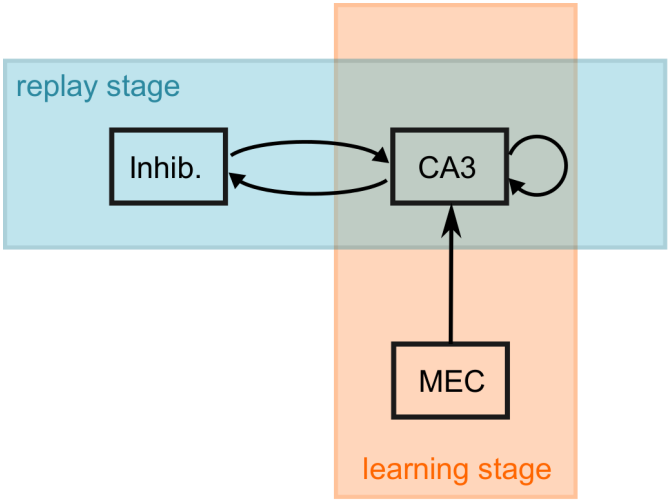

FIGURE 4.1: Schema of hippocampal replay. During learning (orange), new memory sequences are encoded by potentiation of recurrent connections in CA3. For spatial memories, MEC grid cell inputs are hypothesized to separate patterns used for memories (model from Kammerer and Leibold (2014), chapter 3). Replay (blue), can be facilitated by adding an inhibitory population, thereby increasing memory capacity (model from Kammerer et al. (2013), chapter 2).

with the sequences that are encoded in actual hippocampal replay, which inherently have overlaps. Additional models that address this and other problems are reviewed. Then, we will revisit experimental findings on hippocampal replay. We will argue that our model and its scaling predictions are applicable to the existing interpretations of sequence replay experiments.

We then give a short overview of how observations of remapping strengthen the view on how HC and EC encode space. Our own results regarding the synchronization of the two spatial codes in HC and EC are discussed.

4.1 Properties of model sequences and alternative models

In spite of all efforts, we are far away from an integrative model that resolves all problems that arise regarding replay of memory sequences. While our model of sequence replay allows to investigate scaling properties with biological parameters using high cell numbers, it has shortcomings. This paragraph summarizes the gravest differences to real sequences. Other classes of models are discussed that indicate how the brain might deal with several difficulties of sequence replay.

Realistic sequences of place cell firing

The patterns used in our sequence replay model are fully randomized, and hence have little overlap. Real sequences of place cell assemblies on the other hand have overlap since their place fields have spatial overlap and change continuously during movement. As a place field is traversed, the cell fires several times and hence participates in several assemblies. Successive assemblies will therefore share many cells. Only assemblies separated in time will have a chance to not overlap. Indeed, the overlap is thought to be necessary to ensure synapses between assemblies can be potentiated. Two place cells with overlapping fields fire over several theta cycles, and theta phase precession ensures they fire in correct order (Skaggs et al., 1996). This firing occurs on a time scale short enough to trigger short term synaptic plasticity (STDP), and supposedly sets up the weights for sequence replay during rest. If place cells did not overlap, the time scale would be dependent on the animal's speed, and STDP would not trigger reliably. The trade off is that patterns in the sequence will have overlaps.

Continuous attractor models

The overlapping patterns created by STDP and phase precession, together with a random trajectory would create a continuous attractor over time. The reason is that one assembly is not only connected to the next subsequent assembly, but to several, albeit more weakly. Nonetheless, this allows neural activity to flow smoothly between assemblies, and the random trajectory ensures there is no preferred direction. A continuous attractor model like this has been successfully used to create cognitive maps of multiple environments at the same time (Samsonovich and McNaughton, 1997). However, this model does not feature replay of any form. The attractor state can be moved by external cues, which equates to updating the internal representation of position by sensory inputs. Without sensory inputs, the representation stays in one and the same location, and even noise only generates a diffusive movement, not similar to real replay observations. To make the model replay sequences, spike-frequency adaptation was added to (Samsonovich and McNaughton, 1997) resulting in a model of mental exploration

and replay (Hopfield, 2010). The adaptation causes the firing neurons in the attractor assembly to fire less and less, making the attractor state move. This is interpreted as mental exploration. Random paths through the environment can be replayed like this. Even paths that were never experienced are possible, the only requirement is that enough paths were experienced to create a continuous attractor that covers the environment. Not only random paths can be replayed, it is possible to include a preference for formerly experienced or mentally explored paths. For this, the synapses connecting the assemblies of the experienced path are slightly changed by an STDP rule. Activity is then more likely to flow along the experienced path instead of randomly. This preferred path corresponds to the observed bias towards replay of over-trained trajectories (Louie and Wilson, 2001; Diba and Buzsaki, 2007).

Summarizing, there is a continuous attractor model that features replay of place cell assembly sequences (Hopfield, 2010). It uses adaptation to move the internal representation of space, creating replay. For replay of specific sequences, an additional STDP rule is necessary. While the mechanisms exist in HC, their precise role in replay has yet to be tested. Recently, a successful way of analyzing capacity for continuous attractor models was described in (Monasson and Rosay, 2013; 2014). However, mental exploration is more difficult than maintaining an attractor, since the energy landscape has to be very smooth. Hence the capacity could be drastically lower in this class of model. It has not been tested using a network that stores more than a few environments.

Wave-propagation path planning

Finally the mental exploration model can be altered to efficiently find the shortest path to a target location (Ponulak and Hopfield, 2013). In this model, during mental exploration, a target neuron assembly representing a target location is activated. Then, instead of creating sustained firing in an attractor assembly, an activity wave spreads from the target point, until it reaches the current position. The wave potentiates the synapses via anti-STDP, allowing activity to flow backwards from the current position to the target. From this activity a motor signal

can be constructed (Hopfield, 2010; Ponulak and Hopfield, 2013) to reach the target using the shortest path. Observing only few neurons in an experiment, such waves would look like reverse replay. As mentioned for the model in (Hopfield, 2010), capacity for this model is potentially small, and only few environments have been stored in the tested cases. In addition, it is prone to noise (Ponulak and Hopfield, 2013), which would further decrease capacity.

Models based on continuous attractors give a reason why place cells should have compact place fields, and ideally one place field. A constraint like spatial resolution of the place code alone fails at this. Even cells with multiple overlapping place fields spreading over the whole environment can produce high spatial resolution. A disadvantage of continuous attractors is catastrophic forgetting, they simply fail after a certain number of attractors has been stored. The continuous attractor models reviewed here also show how forward and reverse replay can happen in the same system. Forward replay could be mental exploration, and back ward replay might be the backflow along the trace created by STDP in the model of Hopfield (2010). Our model (Kammerer et al., 2013) cannot replay the same sequence forward and backward. In general, it fails as soon as two sequences have a single assembly in common. The ability to disambiguate at such a junction in the sequence requires information about the previous states of the network, a problem we discuss in detail in the next section.

Disambiguation of crossings in trajectories

Real sequences of place cell assemblies will likely share a complete assembly of neurons when two trajectories intersect at one position. In our model such a crossing point would lead to failure: as soon as either sequence reaches the crossing point, the network would try to replay the remainder of both sequences. The model has no mechanism to choose and replay only one. It has no track of its own history, only its last state is important for the next update. A simple solution is to add layers of neurons that delay the inputs by one or more time steps (Fukushima, 1973). A variety of models exists that use delays in more complex ways to also deal with the temporal dimension of inputs, e.g. (Coolen and Gielen, 1988; Heskes

and Gielen, 1992; Bartholomeus and Coolen, 1992; Herz et al., 1988), and several of these models can be quantified by statistical mechanics (Herz et al., 1991). A different solution (Minai et al., 1993) includes the history implicitly by adding a subset of neurons with fixed recurrent connectivity, and no explicit delays. The subnetwork's internal state contains the history of the inputs and permits disambiguations of sequences that share several items.

Reservoir computing models

The principle of implicitly keeping track of time by using recurrent network dynamics is taken to the extreme in the framework of 'reservoir computing' (Maass et al., 2002; Jaeger, 2001). Only the recurrent network is used, with static synaptic weights, and an output layer with plastic synapses. Learning happens in the synapses of the output layer using supervised learning techniques. If the reservoir parameters are chosen well, it works as a map from input space into a higher dimensional space of neuronal states. The higher dimension allows for two things: separation of similar patterns (separation property), and mapping of multiple inputs over time into a single pattern that can be read out in one time step (memory property). The read out is done by an output layer of neurons. The recurrent network essentially works as a black box. The only demand on it is to produce high dimensional trajectories and to have memory. The memory property of the dynamics can be controlled by capacitive effects in the neuron model, short term plasticity, and also recurrent connectivity. The separation property could be a mechanism for global remapping. In this scenario, few inputs encoding the context could enforce the use of a very different neuronal code in the reservoir, which results in global remapping. In addition, the memory property can potentially be used to evaluate different paths. A path, encoded by multiple place cells firing over time, can be mapped to a single neural code that summarizes information like path length, energy cost or other behaviorally relevant measures. Alternatively, an often used path can be tagged with a neural code that later can be used by other brain areas.

Next step for reservoir computing models

An understanding of how the memory property and separation property of a reservoir react to changes in network parameters is only the first step. To relate the model to real systems it is required to understand how synaptic plasticity can benefit the reservoir. Yet the addition of plasticity rules make it harder to create a working reservoir. One successful model that endows the reservoir with additional useful properties is SORN (Self Organizing Recurrent Network) (Lazar et al., 2009). By carefully combining three plasticity mechanisms, the recurrent connections learn unsupervised, followed by a supervised learning step in the output layer alone. This improves the performance in several tasks by creating more diverse trajectories in the recurrent network, which in turn makes it easier for the output layer to produce the correct output. One interesting effect is that the plasticity rules cause activity in the recurrent network to be more sparse (Fig. 4 in (Lazar et al., 2009)). Another observation is a generalization of a counting task, in which plasticity causes similar inputs to be mapped to similar states (the fifth letter in both series abbbbbbc and deeeeeef). These results make it seem feasible to find appropriate inputs that produce sparse place field firing from exploratory behavior. While other models have achieved this (Franzius et al., 2007), doing the same with a reservoir model would combine the abilities to represent space and to perform complex tasks including sequences of spatial and other memories.

4.2 Sharp wave ripples and replay events

Several experiments show sequential firing of neural ensembles. There are three cases to be considered (Buhry et al., 2011). Replay of sequences can be observed during sleep, in particular slow wave sleep (Lee and Wilson, 2002). Replay is also found during wakeful rest while a behavioral task is going on, and can happen in a forward as well as backward direction (Csicsvari et al., 2007; Foster and Wilson, 2006; Diba and Buzsaki, 2007). Finally, preplay has been observed during wakeful rest (Gupta et al., 2010; Dragoi and Tonegawa, 2011; 2013). Preplay refers to sequential firing of place cell ensembles that were not experienced during

preceding behavior. Preplay of was observed before movement along a linear track. One study (Dragoi and Tonegawa, 2011) found preplay of a linear track before the animal crossed the track, and replay afterwards. Presumably the preplay represented mental exploration or planning. One criticism is that the animal had experienced the whole environment beforehand. Another experiment found preplay in a formerly not experienced environment (Dragoi and Tonegawa, 2011), which suggests that the sequences observed are pre-wired (Buhry et al., 2011).

Replay during slow wave sleep

The case of replay during sleep is of special interest since memory consolidation is believed to happen during sleep (Marr, 1971; Crick and Mitchison, 1983; Diekelmann and Born, 2010). During sleep, it is observed that both SWR and replay correlate with memory performance (Axmacher et al., 2008; Dupret et al., 2010). The more SWR or replay events are observed, the better the memory performance. A causal link could be established as well (Girardeau et al., 2009; Ego-Stengel and Wilson, 2010). Disrupting replay when SWR are detected during SWS significantly affects memory performance. However, the memory performance is still satisfying from a behavioral point of view, indicating that either the disruption is not affecting all replay events, or that there are other mechanisms of consolidation at work (Buhry et al., 2011). In particular it is not clear if each SWR is accompanied by sequential activity or vice versa (Buhry et al., 2011), which could also explain the small effect.

In addition to the directly observed connection to memory consolidation, the SWR and replay events during sleep show all the necessary capacities to induce synaptic changes in the form of LTP and STDP (Bliss and Lomo, 1973; Gerstner et al., 1996; Markram et al., 1997; Magee and Johnston, 1997; Buzsaki et al., 1983).

Replay and reverse replay during wakeful rest

Replay events and SWR were also found during wakeful rest (Csicsvari et al., 2007; Foster and Wilson, 2006; Diba and Buzsaki, 2007). In addition to regular

replay, reverse replay was found. Replay does not always start at the rat's actual position. This can be seen as mental exploration. Alternatively it could mean that consolidation starts while the animal is still awake and performing the task (Buhry et al., 2011).

Preplay

A third type of sequential activity was found later and termed preplay. Preplay events represent trajectories never before realized by the animal. Finding these events requires more careful analysis than used in earlier replay experiments, and it was pointed out that these older experiments might show preplay if analyzed differently (Buhry et al., 2011). Arguably replay is an artifact, and the system's sequences are determined and not altered by experience. The connection to experiences is made by mapping the sensory inputs to these existing sequences. Different mappings might exist in parallel. This is in accordance with our model of remapping, that turned out to be uncontrollable by Hebbian learning. Sparsity could be controlled, but not the exact place field locations. Much like the place code generated by our model, the sequences are set in stone by some process, and their abundance should allow for sufficient mappings from experiences to sequences.

Relevance of our model

Our model of sequence replay is compatible with all possibilities implied by the observations of replay and preplay: the replay during sleep might be an inherent replay to consolidate the synaptic changes made during behavior. The awake behavior could be an inscription phase as well, or mental planning using memories. In both cases, inscription can mean two things. It can constitute changes in the synaptic matrix of the recurrent network, or instead, changes in the mapping from sensory areas to preexisting sequences of the recurrent network, leaving its internal synaptic weights unchanged. Correspondingly, replay can be seen in these two ways. Regarding our model, it has a learning phase in which the synapses of the recurrent network are changed, which seems incompatible with the idea of preexisting sequences. Yet we do not specify when this learning happens. For

the case of preexisting sequences, this learning phase might happen early during development, and all our results about capacity and transient replay are still valid. The case of experience based changes in the recurrent network is covered as well, since new sequences can be added at any time. However, in that case many models assume a mechanism of forgetting that favors new synaptic connections by slowly erasing older connections. In our model a mechanism of forgetting is implied by fixing the synaptic connectivity c, yet all sequences contribute equally to this connectivity, old or new. Indeed the model fits the case of preexisting sequences better.

The observation of forward and backward replay in the same experiment (Diba and Buzsaki, 2007) cannot be explained by our model alone. As mentioned earlier, to disambiguate between forward and backward replay, the model would need a means to identify context, e.g. by keeping track of the history of network states. Nonetheless, it is possible that forward and reverse replay are caused by different mechanisms.

4.3 Remapping

Remapping is a phenomenon that casts light on several hypothesis of hippocampal function. One is the hippocampal area CA3 implementing associative memory. The other is the idea that HC and MEC both contain cognitive maps of space, where MEC is more restricted to space alone whereas HC has the possibility to combine spatial information with content from many other sensory modalities and brain areas.

The hippocampus encodes episodic memories (Squire, 2004). In particular, CA3 is hypothesized to be crucial for memory functions, since its high recurrent connectivity (Amaral and Witter, 1989) allows it to work as an auto-associative memory network (McNaughton and Morris, 1987; Marr, 1971; Treves and Rolls, 1992). This hypothesis is backed up by the finding that remapping is stronger in CA3

than in CA1 (Leutgeb et al., 2004; 2005). Remapping decorrelates neural representations, which is important for memory networks, but not so much for networks that mostly receive feed forward connections, like CA1.

To encode episodic memories, CA3 needs to combine spatial and non-spatial information. From anatomical studies it is known that HC has access to all information in neocortex (Van Hoesen and Pandya, 1975; Swanson and Kohler, 1986; Amaral and Lavenex, 2006). In addition, remapping shows that CA3 not only relies on spatial cues, but also on context like wall color (Leutgeb et al., 2005) or details of the behavioral task (Wood et al., 2000). Hence the remapping experiments further corroborate that CA3 indeed uses more than spatial information.

Grid cells in entorhinal cortex on the other hand show more simple behavior during remapping (Fyhn et al., 2007). When CA3 undergoes rate remapping, the gird cell firing maps stay unchanged. Global remapping in CA3 is accompanied by shifts and in some cases also rotations of grid fields in EC. Yet these transformations are not as radical as in CA3, CA1 or DG. In grid cells, the mean firing rates stay the same and no cell ceases firing. Based on this contrast between CA3 and EC during remapping one can hypothesize that the grid code represents a second code for space, not influenced by context. Along this line, it has been argued that CA3 place fields fully draw spatial information from EC, and the CA3 network merely deals with context (Sharp, 1999), effectively separating the workload to increase memory capacity in CA3. This is disputed by experiments that show that place cell code is developed before grid cell code in young rats (Wills et al., 2010). Nevertheless there is enough spatial information in the grid code of the rat pups to create place fields (Azizi et al., 2014). Independent of HC, the EC has great relevance for navigation and is hypothesized to work as a path integrator (McNaughton et al., 2006). Path integration allows an animal to retrace its path or find the shortest way home in complete darkness in absence of any cues but proprioception and the vestibular system. The ability to find home in darkness is affected by damage to the EC (Parron and Save, 2004).

Taken together, the evidence draws a picture of EC holding a map of space,

whereas HC also holds spatial information, and in addition associates spatial content with context from all over the cortex. The idea that EC is specialized in encoding space without context is consistent with the fact that the grid code meets optimal requirements for encoding space and outperforms place codes (Mathis et al., 2012). Unraveling how the two codes interact is an important endeavor for experimentalists and modelers.

We investigated (Kammerer and Leibold, 2014) the feed forward connections from MEC to CA3 or CA1 in a grid to place cell model, under the hypothesis that the connections are learned by Hebbian learning. We find that it is not possible to learn place field positions in a controlled manner using Hebbian learning on the feed forward connections of a grid to place cell model, as was found before for a similar model (Cheng and Frank, 2011). Instead of showing up on the learned location, place fields mostly show up at multiple random locations. However, we show it is possible to control place field sparsity and to maintain a good spatial resolution by controlling parameters like inhibition. Indeed a good spatial resolution is easily obtained whereas sparse place fields require more careful adjustment of parameters. The most efficient way we encountered to control place cell sparseness is to only use a few place cells in each remapping. Hence we postulate that the finding that not all place cells participate in all remappings is necessary to successfully read out grid code while maintaining sparse and compact firing fields.

The little control over positions of place fields during learning in our model can be fixed in two ways. The first is to assume that the place field positions are not controllable, and CA3 acts as a random network, in which each assembly only is assigned a meaning in the read-out layer. This situation resembles the reservoir computing, a full recoding step in the read-out layer is required. Alternatively, additional mechanisms might help control place field positions. For example, a setup in CA3 using multiple continuous attractors (see Samsonovich and McNaughton (1997)) could stabilize place fields, and the grid to place model presented here could be used to move the attractor and address the different remappings.

4.4 Concluding remarks

The hippocampal formation participates in episodic memory and memory consolidation. It also encodes space in place cell and grid cell networks. Experimental evidence from anatomy, human patients, and electrophysiology in animals, gives an opportunity to constrain individual models to match various experimental observations.

The idea that remapping, which is observed in a spatial context, serves memory networks as pattern separator is very fruitful. In this thesis a pattern separation mechanism based on remapping was analyzed, and the results emphasize that the compact place fields we see in hippocampus likely are needed for more than just spatial resolution of the place code. As mentioned in the discussion, continuous attractor models can indeed take advantage of such compact place fields to efficiently navigate to a target location. Both models contribute to a more concise picture of why we find compact place fields.

Models of the hippocampal area go beyond explaining experimental observations. Many models illustrate powerful principles like pattern separation. Reservoir computing models for example showcase the principle of dimensional expansion of neural code. They also quantify how mechanisms like short term synaptic plasticity endow the network with memory. While reservoir computing models do not explain experimental findings in hippocampus, the principles they unveil might one day gain importance in a new experimental context. At the very least they can guide our intuition when building models.

References

Amaral D, Lavenex P (2006) *Hippocampal Neuroanatomy, in Andersen P., Morris R., Amaral D., Bliss T., O'Keefe J.: The Hippocampus Book.* Oxford University Press.

Amaral DG, Ishizuka N, Claiborne B (1990) Neurons, numbers and the hippocampal network. *Prog Brain Res* 83:1–11.

Amaral DG, Witter MP (1989) The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience* 31:571–591.

Axmacher N, Elger CE, Fell J (2008) Ripples in the medial temporal lobe are relevant for human memory consolidation. *Brain* 131:1806–1817.

Azizi AH, Schieferstein N, Cheng S (2014) The transformation from grid cells to place cells is robust to noise in the grid pattern. *Hippocampus* 24:912–919.

Barry C, Burgess N (2014) Neural mechanisms of self-location. *Curr Biol* 24:R330–339.

Bartholomeus M, Coolen A (1992) Sequences of Smoothly Correlated Patterns in Neural Networks with Random Transmission Delays. *Biological Cybernetics* 67:285–290.

Bliss TV, Lomo T (1973) Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J Physiol* 232:331–356.

Brun V, Solstad T, Kjelstrup K, Fyhn M, Witter M, Moser E, Moser M (2008) Progressive Increase in Grid Scale From Dorsal to Ventral Medial Entorhinal Cortex. *Hippocampus* 18:1200–1212.

Buhry L, Azizi AH, Cheng S (2011) Reactivation, replay, and preplay: how it might all fit together. *Neural Plast* 2011:203462.

Burak Y, Fiete I (2009) Accurate Path Integration in Continuous Attractor Network Models of Grid Cells. *PLoS Computational Biology* 5:1–16.

Burgess N, Barry C, O'Keefe J (2007) An Oscillatory Interference Model of Grid Cell Firing. *Hippocampus* 17:801–812.

Buzsaki G (1989) Two-stage model of memory trace formation: a role for "noisy" brain states. *Neuroscience* 31:551–570.

Buzsaki G (1996) The hippocampo-neocortical dialogue. *Cereb Cortex* 6:81–92.

Buzsaki G, Leung LW, Vanderwolf CH (1983) Cellular bases of hippocampal EEG in the behaving rat. *Brain Res* 287:139–171.

Buzsaki G, Wang XJ (2012) Mechanisms of gamma oscillations. *Annu Rev Neurosci* 35:203–225.

Carr MF, Jadhav SP, Frank LM (2011) Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nat Neurosci* 14:147–153.

Cheng S, Frank LM (2011) The structure of networks that produce the transformation from grid cells to place cells. *Neuroscience* 197:293–306.

Chrobak JJ, Buzsaki G (1996) High-frequency oscillations in the output networks of the hippocampal-entorhinal axis of the freely behaving rat. *J Neurosci* 16:3056–3066.

Colgin LL, Moser EI, Moser MB (2008) Understanding memory through hippocampal remapping. *Trends Neurosci* 31:469–477.

Coolen A, Gielen C (1988) Delays in Neural Networks. *Europhys Lett* 7:281–285.

Crick F, Mitchison G (1983) The function of dream sleep. *Nature* 304:111–114.

Csicsvari J, O'Neill J, Allen K, Senior T (2007) Place-selective firing contributes to the reverse-order reactivation of CA1 pyramidal cells during sharp waves in open-field exploration. *Eur J Neurosci* 26:704–716.

Deng W, Aimone JB, Gage FH (2010) New neurons and new memories: how does adult hippocampal neurogenesis affect learning and memory? *Nat Rev Neurosci* 11:339–350.

Diba K, Buzsaki G (2007) Forward and reverse hippocampal place-cell sequences during ripples. *Nat Neurosci* 10:1241–1242.

Diekelmann S, Born J (2010) The memory function of sleep. *Nat Rev Neurosci* 11:114–126.

Dragoi G, Tonegawa S (2011) Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature* 469:397–401.

Dragoi G, Tonegawa S (2013) Distinct preplay of multiple novel spatial experiences in the rat. *Proc Natl Acad Sci* 110:9100–9105.

Dupret D, O'Neill J, Pleydell-Bouverie B, Csicsvari J (2010) The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nat Neurosci* 13:995–1002.

Ego-Stengel V, Wilson MA (2010) Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus* 20:1–10.

Eichenbaum H (2000) A cortical-hippocampal system for declarative memory. *Nat Rev Neurosci* 1:41–50.

Eichenbaum H (2004) Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron* 44:109–120.

Fenton AA, Kao HY, Neymotin SA, Olypher A, Vayntrub Y, Lytton WW, Ludvig N (2008) Unmasking the CA1 ensemble place code by exposures to small and large environments: more place cells and multiple, irregularly arranged, and expanded place fields in the larger space. *J Neurosci* 28:11250–11262.

Fiete IR, Burak Y, Brookings T (2008) What grid cells convey about rat location. *J Neurosci* 28:6858–6871.

Foster DJ, Wilson MA (2006) Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* 440:680–683.

Frankland PW, Bontempi B (2005) The organization of recent and remote memories. *Nat Rev Neurosci* 6:119–130.

Franzius M, Sprekeler H, Wiskott L (2007) Slowness and sparseness lead to place, head-direction, and spatial-view cells. *PLoS Comput Biol* 3:e166.

Fuhs M, Touretzky D (2006) A Spin Glass Model of Path Integration in Rat Medial Entorhinal Cortex. *J Neurosci* 26:4266 – 4276.

Fukushima K (1973) A model of associative memory in the brain. *Kybernetik* 12:58–63.

Fyhn M, Hafting T, Treves A, Moser M, Moser E (2007) Hippocampal remapping and grid realignment in entorhinal cortex. *Nature* 446:190–4.

Fyhn M, Molden S, Witter M, Moser E, Moser M (2004) Spatial representation in the entorhinal cortex. *Science* 305:1258–64.

Gerstner W, Kempter R, van Hemmen JL, Wagner H (1996) A neuronal learning rule for sub-millisecond temporal coding. *Nature* 383:76–81.

Giocomo L, Zilli E, Fransén E, Hasselmo M (2007) Temporal frequency of subthreshold oscillations scales with entorhinal grid cell field spacing. *Science* 315:1719–22.

Girardeau G, Benchenane K, Wiener SI, Buzsaki G, Zugaro MB (2009) Selective suppression of hippocampal ripples impairs spatial memory. *Nat Neurosci* 12:1222–1223.

Gupta AS, van der Meer MA, Touretzky DS, Redish AD (2010) Hippocampal replay is not a simple function of experience. *Neuron* 65:695–705.

Hafting T, Fyhn M, Molden S, Moser M, Moser E (2005) Microstructure of a spatial map in the entorhinal cortex. *Nature* 436:801–6.

Harris KD, Csicsvari J, Hirase H, Dragoi G, Buzsaki G (2003) Organization of cell assemblies in the hippocampus. *Nature* 424:552–556.

Hasselmo ME, Anderson BP, Bower JM (1992) Cholinergic modulation of cortical associative memory function. *J Neurophysiol* 67:1230–1246.

Hasselmo ME, Bower JM (1993) Acetylcholine and memory. *Trends Neurosci* 16:218–222.

Herz A, Li Z, van Hemmen J.L. (1991) Statistical mechanics of temporal association in neural networks with transmission delays. *Phys Rev Lett* 66:1370–1373.

Herz A, Sulzer R, van Hemmen J (1988) The Hebb Rule: Storing Static and Dynamic Objects in an Associative Neural Network. *Europhys Lett* 7:663.

Heskes T, Gielen S (1992) Retrieval of Pattern Sequences at Variable Speeds in a Neural Network with Delays. *Neural Networks* 5:145–152.

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci* 79:2554–2558.

Hopfield JJ (2010) Neurodynamics of mental exploration. *Proc Natl Acad Sci* 107:1648–1653.

Insausti R (1993) Comparative anatomy of the entorhinal cortex and hippocampus in mammals. *Hippocampus* 3 Spec No:19–26.

Jaeger H (2001) The "Echo State" Approach to Analysing and Training Recurrent Neural Networks. *GMD Report 148. Bremen, GMD - German National Research Institute for Computer Science* .

Johnson A, Redish AD (2007) Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J Neurosci* 27:12176–12189.

Kammerer A, Leibold C (2014) Hippocampal remapping is constrained by sparseness rather than capacity. *PLoS Comput Biol* 10:e1003986.

Kammerer A, Tejero-Cantero A, Leibold C (2013) Inhibition enhances memory capacity: optimal feedback, transient replay and oscillations. *J Comput Neurosci* 34:125–136.

Kim SM, Ganguli S, Frank LM (2012) Spatial information outflow from the hippocampal circuit: distributed spatial coding and phase precession in the subiculum. *J Neurosci* 32:11539–11558.

Kropff E, Treves A (2008) The Emergence of Grid Cells: Intelligent Design or Just Adaptation ? *Hippocampus* 18:1256–1269.

Kropff Causa E, Carmichael J, Moser E, Moser MB Speed cells in the medial entorhinal cortex. *Program No. 094.04. 2014 Neuroscience Meeting Planner. Washington, DC: Society for Neuroscience, 2014. Online.* .

Lazar A, Pipa G, Triesch J (2009) SORN: a self-organizing recurrent neural network. *Front Comput Neurosci* 3:23.

Lechner HA, Squire LR, Byrne JH, Muller GE, Pilzecker A (1999) 100 years of consolidation–remembering Müller and Pilzecker. *Learn Mem* 6:77–87.

Lee AK, Wilson MA (2002) Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* 36:1183–1194.

Leibold C, Kempter R (2006) Memory Capacity for Sequences in a Recurrent Network with. *Neural Computation* 18:904–941.

Leutgeb J, Leutgeb S, Moser M, Moser E (2007) Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* 315:961–6.

Leutgeb S, Leutgeb JK, Treves A, Moser MB, Moser EI (2004) Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science* 305:1295–1298.

Leutgeb S, Leutgeb J, Barnes C, Moser E, McNaughton B, Moser M (2005) Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science* 309:619–23.

Lever C, Wills T, Cacucci F, Burgess N, O'Keefe J (2002) Long-term plasticity in hippocampal place-cell representation of environmental geometry. *Nature* 416:90–94.

Louie K, Wilson MA (2001) Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron* 29:145–156.

Maass W, Natschlager T, Markram H (2002) Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput* 14:2531–2560.

Magee JC, Johnston D (1997) A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. *Science* 275:209–213.

Markram H, Lubke J, Frotscher M, Sakmann B (1997) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* 275:213–215.

Marr D (1971) Simple memory: a theory for archicortex. *Philos Trans R Soc Lond, B, Biol Sci* 262:23–81.

Mathis A, Herz AV, Stemmler M (2012) Optimal population codes for space: grid cells outperform place cells. *Neural Comput* 24:2280–2317.

McNaughton BL, Battaglia FP, Jensen O, Moser EI, Moser MB (2006) Path integration and the neural basis of the 'cognitive map'. *Nat Rev Neurosci* 7:663–678.

McNaughton B, Morris R (1987) Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci* 10:408–415.

Milner B (1962) *Physiologie de l'hippocampe* Centre National de la Recherche Scientifique, Paris.

Minai A, Barrows G, Levy W (1993) Disambiguation of Pattern Sequences with Recurrent Networks. *WCNN* .

Mishkin M (1978) Memory in monkeys severely impaired by combined but not by separate removal of amygdala and hippocampus. *Nature* 273:297–298.

Monaco JD, Abbott LF, Abbott LF (2011) Modular realignment of entorhinal grid cell activity as a basis for hippocampal remapping. *J Neurosci* 31:9414–9425.

Monasson R, Rosay S (2013) Crosstalk and transitions between multiple spatial maps in an attractor neural network model of the hippocampus: phase diagram. *Phys Rev E Stat Nonlin Soft Matter Phys* 87:062813.

Monasson R, Rosay S (2014) Crosstalk and transitions between multiple spatial maps in an attractor neural network model of the hippocampus: collective motion of the activity. *Phys Rev E Stat Nonlin Soft Matter Phys* 89:032803.

Muller RU, Kubie JL (1987) The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *J Neurosci* 7:1951–1968.

Nadel L, Moscovitch M (1997) Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr Opin Neurobiol* 7:217–227.

Nakashiba T, Buhl DL, McHugh TJ, Tonegawa S (2009) Hippocampal CA3 output is crucial for ripple-associated reactivation and consolidation of memory. *Neuron* 62:781–787.

O'Keefe J (1976) Place units in the hippocampus of the freely moving rat. *Exp Neurol* 51:78–109.

O'Keefe J (1979) A review of the hippocampal place cells. *Prog Neurobiol* 13:419–439.

O'Keefe J, Conway DH (1978) Hippocampal place units in the freely moving rat: why they fire where they fire. *Exp Brain Res* 31:573–590.

O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain research* 34:171–175.

O'Keefe J, Nadel L (1978) *The Hippocampus as a cognitive map* Oxford University Press.

O'Keefe J, Speakman A (1987) Single unit activity in the rat hippocampus during a spatial memory task. *Exp Brain Res* 68:1–27.

Parron C, Save E (2004) Evidence for entorhinal and parietal cortices involvement in path integration in the rat. *Exp Brain Res* 159:349–359.

Ponulak F, Hopfield JJ (2013) Rapid, parallel path planning by propagating wavefronts of spiking neural activity. *Front Comput Neurosci* 7:98.

Redish AD, Touretzky DS (1998) The role of the hippocampus in solving the Morris water maze. *Neural Comput* 10:73–111.

Renno-Costa C, Lisman JE, Verschure PF (2010) The mechanism of rate remapping in the dentate gyrus. *Neuron* 68:1051–1058.

Samsonovich A, McNaughton BL (1997) Path integration and cognitive mapping in a continuous attractor neural network model. *J Neurosci* 17:5900–5920.

Sargolini F, Fyhn M, Hafting T, McNaughton B, Witter M, Moser M, Moser E (2006) Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* 312:758–62.

Scoville W, Milner B (1957) Loss of recent memory after bilateral hippocampal lesions. *J Neuropsychiatry Clin Neurosci* 20.

Sharp P (1999) Complimentary roles for hippocampal versus subicular/entorhinal place cells in coding place, context, and events. *Hippocampus* 9:432–43.

Skaggs WE, McNaughton BL (1996) Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* 271:1870–1873.

Skaggs WE, McNaughton BL, Wilson MA, Barnes CA (1996) Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* 6:149–172.

Squire LR (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychol Rev* 99:195–231.

Squire LR (2004) Memory systems of the brain: a brief history and current perspective. *Neurobiol Learn Mem* 82:171–177.

Squire LR (2009) The legacy of patient H.M. for neuroscience. *Neuron* 61:6–9.

Squire L, Zola-Morgan S (1991) The medial temporal lobe memory system. *Science* 253:1380–1386.

Stensola H, Stensola T, Solstad T, Frø land K, Moser MB, Moser E (2012) The entorhinal grid map is discretized. *Nature* 492:72–78.

Stringer SM, Rolls ET, Trappenberg TP, de Araujo IET (2002) Self-organizing continuous attractor networks and path integration: two-dimensional models of place cells. *Network* 13:429–446.

Swanson LW, Kohler C (1986) Anatomical evidence for direct projections from the entorhinal area to the entire cortical mantle in the rat. *J Neurosci* 6:3010–3023.

Taube J, Muller R, Ranck J (1990a) Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J Neurosci* 10:420–35.

Taube J, Muller R, Ranck J (1990b) Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. *J Neurosci* 10:436–47.

Tolman E (1948) Cognitive maps in rats and men. *Psychol Rev* 55:189–208.

Treves A, Rolls ET (1992) Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* 2:189–199.

Treves A, Tashiro A, Witter MP, Moser EI (2008) What is the mammalian dentate gyrus good for? *Neuroscience* 154:1155–1172.

Tse D, Langston RF, Kakeyama M, Bethus I, Spooner PA, Wood ER, Witter MP, Morris RG (2007) Schemas and memory consolidation. *Science* 316:76–82.

Tsodyks M (1999) Attractor neural network models of spatial maps in hippocampus. *Hippocampus* 9:481–489.

Van Hoesen G, Pandya DN (1975) Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. I. Temporal lobe afferents. *Brain Res* 95:1–24.

Wills T, Cacucci F, Burgess N, O'Keefe J (2010) Development of the hippocampal cognitive map in preweanling rats. *Science* 328:1573–6.

Willshaw DJ, Buneman OP, Longuet-Higgins HC (1969) Non-holographic associative memory. *Nature* 222:960–962.

Wilson MA, McNaughton BL (1994) Reactivation of hippocampal ensemble memories during sleep. *Science* 265:676–679.

Wilson M, McNaughton B (1993) Dynamics of the hippocampal ensemble code for space. *Science* 261:1055–8.

Winocur G, Moscovitch M, Bontempi B (2010) Memory formation and long-term retention in humans and animals: convergence towards a transformation account of hippocampal-neocortical interactions. *Neuropsychologia* 48:2339–2356.

Wood ER, Dudchenko PA, Robitsek RJ, Eichenbaum H (2000) Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* 27:623–633.

Acknowledgements

First and foremost, I want to thank my supervisor Prof. Dr. Christian Leibold, for his support and guidance during my PhD thesis. Likewise, I thank Dr. Martin Stemmler and Prof. Dr. Mark Huebener for helpful discussions during our TAC meetings. My special gratitude extends to my friends Dr. Alvaro Tejero and Dr. Alexander Mathis, who inspired me and many others with their passion for neuroscience. I am happy I had the chance to closely cooperate on two projects with them. I thank the Graduierten Kolleg GRK 1091 "Orientierung und Bewegung im Raum" for funding and for providing a forum for young scientists of various disciplines to share their knowledge. Similarly, I am greatly indebted to the graduate school of systemic neurociense (GSN) and their staff for providing a structured research program and bringing together scientists of many backgrounds. In particular I thank the GSN coordinators for their assistance and support: Dr. Alexandra Stein, Lena Bittl, Dr. Julia Brenndörfer and Maj-Catherine Botheroyd. I also thank Monika Volk for her assistance and her interest in our well-being at all times. I am very grateful to all the colleagues in the BCCN and GSN. I want to thank Mauro and Franzi for reading the manuscript of this thesis.

My special thanks go to my parents and my brother Niels. I could not wish for a more supportive family. And I am forever grateful to Mehrnoosh Jafari. Thank you for your support and love.

Curriculum vitae

Personal:

Name:	Axel Joachim Kammerer
Affiliation:	Graduate School of Systemic Neurosciences
	Ludwig-Maximilians-Universität München

Education:

Since Oct 2009:	PhD student at the GSN-LMU Munich,
	supervised by Prof. Dr. Christian Leibold
July 2008 - Feb 2009:	Research internship at Universitat Pompeu Fabra, Barcelona,
	in the group of Prof. Dr. Gustavo Deco
Sep 2002 - Feb 2008:	Studies in Physics at the LMU München (Diploma)
	Diploma Thesis on "Transport in inhomogeneous materials"
Sep 1993 - Jun 2002:	Secondary school: Gymnasium Olching (Abitur)

List of publications

Kammerer, A. and Leibold, C. (2014). "Hippocampal Remapping Is Constrained by Sparseness rather than Capacity." *PLoS Comput Biol* 10(12): e1003986, doi:10.1371/journal.pcbi.1003986

Kammerer, A., Tejero-Cantero, A , and Leibold, C. (2013). "Inhibition enhances memory capacity: optimal feedback, transient replay and oscillations." *J Comput Neurosci* 34(1):125–136. doi: 10.1007/s10827-012-0410-z

Kammerer, A. , Höfling, F., and Franosch, T. (2008). "Cluster-resolved dynamic scaling theory and universal corrections for transport on percolating systems." *EPL*, 84 66002, doi:10.1209/0295-5075/84/66002

Eidesstattliche Versicherung/Affidavit

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation 'Memory Capacity in the Hippocampus: Influence of Inhibition and Remapping' selbstständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

I hereby confirm that the dissertation 'Memory Capacity in the Hippocampus: Influence of Inhibition and Remapping' is the result of my own work and that I have only used sources or materials listed and specified in the dissertation.

Munich, July 20, 2015

Author contributions

Chapter 2 was published in *Journal of Computational Neuroscience* (Kammerer et al., 2013). The authors Axel Kammerer (AK) and Álvaro Tejero-Cantero (AT) contributed equally to this paper. The work was done under the supervision of Christian Leibold (CL); AK, AT and CL conceived and designed the research. AK, AT and CL derived the results for the mean field model. AK wrote the numerical code and carried out the simulations for the verification of mean-field results with cellular simulations. AT established the Bayes-optimal threshold derivation. AK, AT and CL discussed the results and wrote the paper.

Chapter 3 was published in *PLoS Computational Biology* (Kammerer and Leibold, 2014). The work was done under the supervision of Christian Leibold (CL). Axel Kammerer (AK) and CL conceived and designed the research. AK wrote the code, carried out the simulations and analyzed the data. AK and CL discussed the results and wrote the paper.

Munich, February 2015

Álvaro Tejero-Cantero Axel Kammerer Christian Leibold