Reactivation and Reinstatement of Hippocampal Assemblies

Gido M. van de Ven

Thesis submitted for the degree of Doctor of Philosophy

Wolfson College

University of Oxford

Hilary Term 2017

Abstract

Reactivation and Reinstatement of Hippocampal Assemblies

Gido M. van de Ven, Wolfson College
Thesis submitted for the degree of Doctor of Philosophy, Hilary Term 2017

New memories are labile, but over time some of them are stabilized. This thesis investigates the network mechanisms in the brain underlying the gradual consolidation of memory representations. Specifically, I performed a causal test of the long-standing hypothesis that the offline reactivation of new, memory-representing cell assemblies supports memory consolidation by stabilizing those assemblies and increasing the likelihood of their later reinstatement—and therefore presumably of memory recall.

I performed multi-unit extracellular recordings in the dorsal CA1 region of behaving mice, from which I detected short-timescale (25 ms) co-activation patterns of principal neurons during exploration of open-field enclosures. These cell assembly patterns appeared to represent space as their expression was spatially tuned and environment specific; and these patterns were preferentially reactivated during sharp wave-ripples (SWRs) in subsequent sleep. Importantly, after exposure to a novel—but not a familiar—enclosure, the strength with which an assembly pattern was reactivated predicted its later reinstatement strength during context re-exposure. Moreover, optogenetic silencing of hippocampal pyramidal neurons during on-the-fly detected SWRs during the sleep following exposure to a novel—but again not a familiar—enclosure impaired subsequent assembly pattern reinstatement. These results are direct evidence for a causal role of SWR-associated reactivation in the stability of new hippocampal cell assemblies.

Surprisingly, offline reactivation was only important for the stability of a subset of the assembly patterns expressed in a novel enclosure. Optogenetic SWR silencing only impaired the reinstatement of "gradually strengthened" patterns that had had a significant increasing trend in their expression strength throughout the initial exposure session. Consistent with this result, a positive correlation between reactivation and subsequent reinstatement was only found for these gradually strengthened patterns and not for the other, "early stabilized" patterns. An interesting interpretation is that the properties of the gradually strengthened patterns are all consistent with the Hebbian postulate of "fire together, wire together".

To enable investigation of the relation between interneurons and principal cell assembly patterns from extracellular recordings, as a final contribution this thesis describes a statistical framework for the unsupervised classification of interneurons based on their firing properties alone.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. David Dupret. Thanks to his unparalleled technical, practical and organizational skills I have been able to learn and perform a wide range of state-of-the-art experimental techniques. Throughout the DPhil, he has always been immediately available to solve problems, to comment on my latest graph or to discuss new ideas.

In addition, the lab built up by David has provided support in many ways, and I would like to thank every current and former member for making it such a productive and supportive environment. I specially thank Natalia Campo-Urriza for building microdrives and for help with the preparation and imaging of brain slices. Thanks to Dr. Stéphanie Trouche and Dr. Claire Bratley for help with some of the recordings for this thesis. I wish to thank Dr. Colin McNamara for tips and guidance throughout as well as before my DPhil studies, and Dr. Vítor Lopes dos Santos for several suggestions of data analysis protocols and stimulating discussions. I acknowledge Dr. Kevin Allen for setting up the on-the-fly SWR detection, and Prof. Jozsef Csicsvari and former members of his lab for the use of computer programs dating back to their time.

I would also like to thank the members of the wider unit for its fruitful scientific atmosphere. I express thanks to the unit's support staff, in particular to Jane Janson, Katharine Whitworth and Liz Norman. I acknowledge Dr. Linda Katona for suggesting the protocol for visualization of the tetrode tracks, and I thank Dr. Rafal Bogacz for useful discussions. Also thanks to Ben Micklem for his help with the production of a video abstract highlighting some of this thesis' findings, and to Pierangelo Pirak for preparing the animations therein (see https://youtu.be/USVDXlWQfRE).

I thank Prof. Peter Somogyi, Prof. Mark Woolrich and Dr. Vladyslav Vyazovskiy for helpful comments on my Transfer of Status report and/or my Confirmation of Status presentation.

My DPhil studies have been funded by a Medical Research Council Doctoral Training Award, and in their final phase by a Vice-Chancellor's Fund Award.

On a personal level, I express profound gratitude to my parents for providing me with every possible opportunity and supporting me throughout my education. Finally, I thank my friends and family, especially Madeline, for warm support all through this journey.

Contents

Figures

Table

Abbreviations

AAV	adeno-associated virus
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
ANOVA	analysis of variance
AP5	(2R)-amino-5-phosphonovaleric acid
CA	Cornu Ammonis
CaMKII	Ca2+/calmodulin-dependent protein kinase II
CPP	3-(2-Carboxypiperazin-4-yl)propyl-1-phosphonic acid
CV	coefficient of variation of the ISI distribution
DAPI	4',6-diamidino-2-phenylindole
dCA1	CA1 region of the dorsal hippocampus
DNMS	Delayed Non-Match to Sample
DNN	deep neural network
DPhil	Doctor of Philosophy
GFP	green fluorescent protein
ICA	independent component analysis
IEG	immediate-early gene
ISI	inter-spike-interval
LFP	local field potential
LTP	long-term potentiation
NDS	normal donkey serum
NMDA	N-methyl-D-aspartate
PB	0.1 M phosphate buffer
PBS	0.1 M phosphate-buffered saline
PBS-T	PBS with 0.25 % Triton X-100
PCA	principal component analysis
PFS	place-field similarity
PV	parvalbumin
REM	rapid eye movement
SCE	synchronous calcium event
SD	standard deviation
SEM	standard error of the mean
SE	standard error
SWR	sharp wave-ripple
Wfs1	Wolfram syndrome 1

CHAPTER 1: GENERAL INTRODUCTION

New memories are initially labile. This 'curious fact' of the mind has been reflected upon at least since classical antiquity (Dudai, 2004). Basing himself on introspection, the Roman orator Quintilian stated in the first century A.D. that "*[it] is astonishing how much strength the interval of a night gives [the memory]"* (Quintilian, 1856/2006). Almost two millennia later, two complementary lines of research revived the idea that memories are stabilized gradually. Based on clinical observations, the French psychologist Théodule Ribot (1882) noted that in amnesic patients recent memories are typically most impaired, while their remote memories—for example from childhood—tend to be spared. Around the same time, behavioural research with human participants learning lists of nonsense syllables demonstrated that there is a limited time window following acquisition during which new memories are sensitive to disruption (Müller and Pilzecker, 1900; Lechner et al., 1999). This last set of experiments led the German professor Georg Müller and his student Alfons Pilzecker to conclude that "*there is no alternative but to assume that (…) certain physiological processes, which serve to strengthen the associations induced (…), continue with decreasing intensity for a period of time*"[1]. To refer to these processes, Müller and Pilzecker introduced the term "consolidation" (Lechner et al., 1999).

Since the coining of the term "consolidation" at the start of the 20th century, the existence of time-dependent processes in memory storage has been firmly established (McGaugh, 1966). Brain damage, drugs, sensory interference and electroconvulsive shocks have all been shown to interfere with new memories during a limited time window following their acquisition (Ribot, 1882; Müller and Pilzecker, 1900; Duncan, 1949; Agranoff et al., 1966; Barondes and Cohen, 1966). Amnesic patients with localized brain lesions, later complemented by animal models of

[1] This quote from Müller and Pilzecker (1900; p.196) was taken from a translation of parts of their work provided by Lechner et al. (1999).

declarative memory and human brain imaging techniques, have over the last 60 years provided insights into where and how the processes of memory consolidation take place in our brain (reviewed by Milner et al., 1998; Squire et al., 2015). One hypothesis that has emerged from this growing body of research is that the reactivation of neuronal activity patterns representing recently experienced events contributes to memory consolidation by stabilizing these patterns, thereby enabling their reinstatement during memory recall. The main aim of this thesis is to perform a direct, causal test of this hypothesis.

1.1) Why is memory storage not instantaneous?

It seems appropriate to start with a brief reflection on the question *why* new memories need time to consolidate. Why has evolution shaped our memory system in such a way that memories are initially prone to interference? Why have we not been endowed with a memory system that deposits new memories into long-term storage instantaneously?

At first thought, it seems an inefficiency of our brain that it takes us time, and generally effort as well, to lastingly store new memories. This suggests that the evolution of our memory system might have been limited by earlier converged on biological constraints. In other words, the evolution of our memory system might have ended up in a local optimum. Given another set of hardware constraints (e.g., a genetic variation large enough to allow the evolution of new classes of neurons—or would they still be called neurons?—capable of different types of computations), an optimizing evolutionary process might converge to a different memory system, potentially closer to the assumed global optimum of effortless and instantaneous remembering.

In a short story called *Funes the Memorious* (first published in 1944), the Argentinean writer Jorge Luis Borges skilfully explored the consequences of such a flawless, immediate memory.

The story is about the fictional young man Ireneo Funes, who, after sustaining head injury due to falling off his horse, remembered and perceived everything: "*Funes remembered not only every leaf of every tree of every wood, but also every one of the times he had perceived or imagined it.*"[2] Although Funes himself considered his extraordinary memory a blessing at first, the storyteller recognized serious drawbacks soon after meeting him. He noticed that Funes was incapable of even the simplest generalizations. He described how Funes was not only unwilling to accept that the generic symbol 'dog' referred to so many dissimilar individuals of all shapes and sizes, but that it also "*bothered him that the dog at three fourteen (seen from the side) should have the same name as the dog at three fifteen (seen from the front).*" Towards the end of the story, Borges even goes so far to let the storyteller infer that Funes "*was not very capable of thought"*, and to finally conclude that *"[t]o think is to forget differences, [to] generalize, [to] make abstractions*".

The position arrived at by Borges' storyteller is very similar to a conclusion reached by the American philosopher William James several decades earlier. A famous quote from his book "The Principles of Psychology" states that "*[i]n the practical use of our intellect, forgetting is as important a function as recollecting*" (James, 1890; p.679). What James referred to here, is that for our memory to be useful, we need to be selective in what we remember. He argued that if too many memories were stored in our brain, the recall of each individual memory would be so time-consuming that no time would be left for thinking.

Note that the arguments put forward so far only highlight that there is a benefit in not remembering everything, but that this does not necessarily imply that remembering needs to be slow. Would it not be possible to quickly (or even instantaneously) remember important experiences, while discarding irrelevant ones? In daily life, however, it is often not immediately

[2] This and further quotes in this paragraph are from the English translation of *Funes the Memorious* by James Irby.

clear which experiences will turn out to be worth remembering. As solution for this, the British neuroscientist David Marr (1970, 1971) suggested a model, partly based on experimental results at the time, in which initially all encountered sensory information is deposited in a temporary storage. After it becomes clear which information is worth retaining, and also how that information fits in with previously stored knowledge, the relevant bits are transferred in a structured manner to the brain's long-term storage. An important point made by Marr was that he argued that it would be computationally inefficient to immediately store new information permanently.

Building on Marr's work, James McClelland and colleagues (1995) went a step further and suggested that even once it is clear which new information is relevant, there might be a computational benefit of incorporating this information into our long-term memory *slowly*. They argued that for any connectionist network, the successful extraction of shared structure between different experiences depends on a strategy they called "interleaved learning", in which learning about one experience is interspersed with learning about other experiences. Abrupt or immediate integration of new information into a connectionist network has the risk to severely disrupt the representation of existing knowledge in it, a phenomenon called "catastrophic interference" (McCloskey and Cohen, 1989). McClelland and colleagues proposed that interleaved learning is implemented in our brain in the form of two complementary learning systems, whereby one systems quickly stores new experiences and then gradually, and interspersed with each other, "replays" them to the other learning system. Over time, this second system extracts commonalities and learns statistical regularities from the events presented to it. That this approach to learning and memory is currently motivating the world's leading experts in the design of artificial intelligent agents (Kumaran et al., 2016) highlights that the gradual nature of memory consolidation might be a "necessary imperfection" of the brain.

1.2) Where are memories stored?

When attempting to pinpoint mechanisms in the brain underlying memory consolidation, a sensible first question to ask is where in the brain memories are stored. Perhaps the first to systematically address this was the German physician Franz Joseph Gall, who was interested in the broader quest of localizing all mental process to specific parts of the brain (Gall, 1822; Ackerknecht and Vallois, 1956; Zola-Morgan, 1995). Gall appreciated that the brain is not a homogenous organ, and he believed that by careful and systematic investigation of the correlation between the shapes of individual persons' skulls and the strength of their various mental faculties, the approximate locations in the brain of those mental faculties could be determined. Using this approach, Gall concluded that memory—which he believed to be divided into three different "faculties"—was located in several locations in the prefrontal cortex (Gall, 1822; Ackerknecht and Vallois, 1956). This way of investigating the brain became popular in the 19th century under the name "phrenology", mainly thanks to Gall's erstwhile collaborator Johann Spurzheim (Zola-Morgan, 1995).

In the first half of the 20th century, another famous attempt to identify specific brain sides important for the storage of memory was conducted by the American neuropsychologist Karl Spencer Lashley. He trained rats to perform memory tasks, while systematically performing lesions of different parts of the rat's cortex. Despite 30 years of careful experimentation, he failed to localize the side of memory storage: "*[t]his series of experiments has yielded a good bit of information about what and where the memory trace is not*" (Lashley, 1950; p.26). Generalizing from his overwhelming amount of collected data, Lashley induced two principles of brain function: "equipotentiality", which states that following damage to one part of a functional brain area its role can in some cases be taken over by another part, and "mass action", which suggests that the extent of memory impairment correlates with the size of brain injury rather than with its specific location (Lashley, 1929; Bartlett, 1960).

Despite that both Gall and Lashley are credited for important conceptual contributions to the study of memory, nowadays their inferences about the localization of memory are generally considered to be incorrect (McGaugh, 2007).

Amnesia following medial temporal lobe lesions

Striking evidence for the localization of memory to a specific part of the brain was provided in the 1950s, when the Canadian neuropsychologist Brenda Milner described memory impairments in a series of human patients after targeted surgical removal of parts of their medial temporal lobes by surgeons William Scoville and Wilder Penfield (Scoville and Milner, 1957; Penfield and Milner, 1958). Most remarkable was the case of patient H.M., who received bilateral medial temporal lobe resection as a radical experimental treatment for frequent and otherwise uncontrollable epileptic seizures. This intervention mostly cured H.M. from his worst seizures, but it had a dramatic effect on his memory. He appeared to have suffered "*a complete loss of memory for events subsequent to [his] bilateral medial-lobe resection (…), together with a partial retrograde amnesia for the three years leading up to his operation*", even though he had "*no impairment of personality or general intelligence*" (Scoville and Milner, 1957; p.17).

Although following his surgery H.M. was largely unable to remember recent experiences or to retain new factual information, he turned out to still be capable of some forms of learning. Using a task in which he had to draw a star while viewing his hand in a mirror, Brenda Milner (1962; Milner et al., 1998) reported that H.M. steadily improved in this motor skill task over 3 days of testing, even though he consistently reported to have no idea that he had ever done the task before. Another form of memory shown to be preserved in H.M. was "priming", a phenomenon in which exposure to a particular stimulus implicitly (i.e., without conscious realization) influences subsequent behaviour (Milner et al., 1968). These and other behavioural studies led to a distinction being made between declarative (or explicit) memories (e.g., for events, facts, people, places, objects) and non-declarative (or implicit) memories

(including motor skills and priming), with only the declarative memory system thought to be critically dependent on the medial temporal lobe (Sherry and Schacter, 1987; Squire, 2004).

The medial temporal lobe refers to a large part of the brain that includes the hippocampal formation (containing the dentate gyrus, the hippocampus proper consisting of the *Cornu Ammonis* [CA] fields and the subiculum), the entorhinal, perirhinal and parahippocampal cortices and the amygdaloid complex (Squire and Zola-Morgan, 1991). Initially it was the hippocampal formation that was cautiously suggested to have a central role in memory, because patients with temporal lobe lesions only developed serious memory impairments when the removed area extended far enough posterior to include the hippocampal region (Scoville and Milner, 1957; Milner et al., 1998). It proved however difficult to precisely and unambiguously locate the specific structures within the temporal lobe important for memory solely from human patients.

Animal models of declarative memory

A major step forward was the development of animal models for declarative memory (Squire, 1992). At first, however, experimentally induced lesions of the medial temporal lobe in monkeys provided puzzling results (discussed by Milner et al., 1998). The main reason was that it was unclear which tasks were appropriate to evaluate the monkey's equivalent of declarative memory. One task that proved useful was a simple object recognition test called "Delayed Non-Match to Sample" (DNMS). In the learning phase of this task a single stimulus was presented to the animal. After a variable delay, that same stimulus was presented again together with a stimulus the animal had never seen before. During this test phase, the animal had to choose the new stimulus in order to obtain a food reward. Mort Mishkin (1978) was the first to show that performance on versions of this task with delays of 10 seconds or longer was severely impaired in monkeys with temporal lobe lesions comparable to those that patient H.M. had incurred. Further work on this primate model, particularly involving more specifically targeted lesions, succeeded to slowly tear apart the various roles of the different parts of the medial temporal

lobe (reviewed by Squire and Zola-Morgan, 1991). Consistent with the suggestive evidence from patient studies, the hippocampal formation was found to be central to the acquisition of declarative memory. In addition, the entorhinal cortex, which provides the main input to the hippocampal formation, and the perirhinal and parahippocampal cortices, which together provide the main input to the entorhinal cortex, were found to have roles in declarative memory complementary to the hippocampus since lesions that included these extrahippocampal areas caused more severe memory deficits than lesions restricted to the hippocampal formation. The amygdala on the other hand was found not to be directly involved in declarative memory, but merely to have a modulatory role (for a recent review of the "anatomy of memory" see Van Strien et al., 2009).

In rats, performance on DNMS tasks was also shown to depend on the integrity of the medial temporal lobe memory system (Otto and Eichenbaum, 1992), although there has been disagreement over whether the rat's hippocampus itself is important for this task or not (e.g., compare Clark et al., 2001 and Forwood et al., 2005). There is, on the other hand, a wide agreement on the role of the rodent hippocampus in spatial memory (O'Keefe and Nadel, 1978; Morris et al., 1982; Buzsáki and Moser, 2013), which is also considered to be a form of declarative memory (Eichenbaum, 2000). As will be discussed in the next section, the study of memory in rodents has prominently contributed to insights into how memories are represented and stored in the brain.

Transfer of memories to the cortex?

Damage to the memory system of the medial temporal lobe not only prevents the acquisition of new memories, it also impairs the retrieval of memories that were acquired before the brain injury (Cohen and Squire, 1981). Importantly, recent memories tend to be more affected than remote ones (Zola-Morgan and Squire, 1990), although the length of this temporal gradient in retrograde amnesia was found to differ between species, between memory tasks and depending on the extent of the brain damage (Squire et al., 2001; Frankland and Bontempi, 2005; Winocur

et al., 2013). Based on these observations it was proposed that the hippocampus has a time-limited role in memory: memories are initially stored in the hippocampus, but they are gradually "transferred" to the cortex for their long-term storage (Squire and Alvarez, 1995; with important roots in the theory of Marr, 1970, 1971). This hypothesis is referred to as the *standard consolidation theory*. Although it has to be said that it was generally thought—also by proponents of this theory—that already during encoding, traces of memories were stored in both the hippocampus and the cortex. The important point of the standard consolidation theory is that memory retrieval critically depends on an intact hippocampus at first, but that over time this dependence diminishes so that memories can be recalled by the cortex on its own. The reorganization of memories between one structure (e.g., the hippocampus) and another (e.g., the cortex) is sometimes referred to as "systems consolidation", to set it apart from "local consolidation" which refers to the stabilization of a memory trace within a given brain structure (e.g., Dudai, 2004; Frankland and Bontempi, 2005; see also section 6.2 for further discussion of these terms).

However, there is evidence that while medial temporal lobe damage results in temporally graded retrograde amnesia for semantic memories (i.e., facts, general knowledge), this might not be the case for episodic memories (i.e., experiences, autobiographical events). Based on observations from both lesion (Fujii et al., 2000) and imaging studies (e.g., Gilboa et al., 2004), it has been suggested that the hippocampus always remains involved in the retrieval of episodic memories, especially if they are vivid and rich in contextual detail (Nadel and Moscovitch, 1997; Moscovitch et al., 2006). This hypothesis has been called the *multiple trace theory*, as it proposes that each time an episodic memory is recalled it is re-encoded, which leads to the formation of multiple overlapping but dissimilar memory traces for the same experience. This explains why remote memories are more widely distributed in the brain and why aspects of such memories (e.g., factual knowledge, the gist of an experience) can after a while be recalled by the cortex alone. To truly re-experience past events, however, the hippocampus is believed to always be required in order to bind together all details of an

episodic memory. A recent extension of the multiple trace theory, the *transformation hypothesis*, further stresses that there is a qualitative difference between the detailed memories depending on the hippocampus and the more abstract, gist-like memories supported by cortical areas (Winocur et al., 2010; Winocur and Moscovitch, 2011).

The complementary learning systems framework of McClelland and colleagues (1995) that was introduced earlier, although often associated with the standard consolidation theory (e.g., Dudai, 2012), can be seen as providing a computational model for the conceptual propositions of the transformation hypothesis. In this interpretation, memories are "transformed" when the slow learning system of McClelland et al. (1995) extracts shared structure and statistical regularities from the original event memories stored in the fast learning system. In line with the transformation hypothesis, McClelland et al. (1995) postulated the brain's fast learning system to be centred around the hippocampus—especially around the presumed auto-associative recurrent network of the CA3 area (McNaughton and Morris, 1987; Treves and Rolls, 1994)—and the slow learning system to be widely distributed across the cortex.

1.3) How are memories stored in the brain?

An important question related to where in the brain memories are stored, is how they are stored. Before discussing this question, it is imperative to highlight a distinction between memory "storage" and "representation", which can be a cause of confusion in the current literature. In this thesis, I will take a memory representation to be a specific neural activity pattern (either *locus*-specific or not, see next paragraph) that is present if and only if a particular mental concept is being perceived or being thought about. Storage is harder to define. Accepting the above definition of representation, memory storage can be interpreted as referring to two sets of configurations in the brain: (1) those enabling the appropriate activation of the neural pattern at times that the mental concept represented by it should be remembered (i.e., referring to the input

of a pattern); and (2) those ensuring that activation of the neural pattern triggers the responses appropriate for the particular mental concept that pattern represents (i.e., referring to the output of a pattern). Note that this definition of memory storage is static in the sense that it refers to the end result of the storage process. The process of memory storage, which encompasses both encoding and consolidation, necessarily involves changes in at least one of the above two sets of configurations. Another important note is that the term "mental concept" should be interpreted very broadly. For example, when a new association between two already existing mental concepts is made (e.g., between a food reward and a specific location), this newly established association itself should then be regarded as a mental concept as well.

Memory representation: activity of specific neurons or *locus*-independent patterns?

The modern era of neuroscience is often considered to have started when, at the end of the 19th century, the Spanish neuroanatomist Santiago Ramón y Cajal used the then recently discovered Golgi staining technique to convincingly demonstrate that the brain is made up of many individual nerve cells that connect to each other by contiguity rather than continuity (von Waldeyer-Harz, 1891; Jones, 1994). Another important contribution of Ramón y Cajal was his proposition that plasticity of the connections between these nerve cells could underlie learning and memory (Ramón y Cajal, 1894; DeFelipe, 2006). This hypothesis had a major influence on subsequent work by the British neurophysiologist Charles Sherrington, who is also credited for naming the connections between neurons "synapses" (Burke, 2007). Sherrington attempted to explain brain function in terms of reflexes, which he had shown to be mediated by unidirectional chains of neurons linked by synapses (e.g., Sherrington, 1892). He suggested that such "reflex arcs" are the basic building blocks of the central nervous systems and that combining many simple reflexes could form the basis for all mental processes (Sherrington, 1906). This theory predicted that learned behaviours depended on particular combinations of reflex arcs, thereby implying that memories were localizable to specific neurons and their synaptic connections. In the first half of the 20th century, an important source of doubt

concerning this hypothesis of localized memories was provided by Lashley's failed attempt to locate such memory traces (Lashley, 1929). Another argument against the localization of memories was based on the phenomenon of "perceptual generalization" or "stimulus equivalence": humans recognize a drawing of a triangle as a triangle regardless of where on the retina its image falls, and thus regardless of which specific retinal neurons are excited. This was taken as evidence that the concept of a triangle could not be represented by a specific set of neurons (e.g., Lashley, 1942). A different problem with the theory of compounded reflexes was computational, as it was unclear how complex mental processes could be supported merely by the serial computations afforded by reflex arcs (e.g., Lashley, 1951). An alternative hypothesis was provided by "field theory", which was put forward by the German psychologist Wolfgang Köhler (1940). According to this theory, mental concepts are represented by *locus*-independent electrical field patterns. That is, not so much which neurons are active, but the "shape" of the resulting electrical fields produced by their combined activity was postulated to be important for mental processing. However, exactly how such patterns are generated, or how and where they could be stored, was not clearly specified.

Cell assembly hypothesis

With the publication of his book *The Organization of Behavior* in 1949, the Canadian psychologist Donald Hebb provided an elegant and highly influential theory about how memories are stored and represented in the brain. He proposed that memories are represented by the combined activity of groups of neurons, which he called "cell assemblies". Importantly, he also suggested a mechanism of how such representations could be created: neurons repeatedly firing together in response to a particular stimulus could bind together through strengthened synaptic connections, resulting in (the activation of) those neurons together representing that stimulus. For this, he postulated the following "learning rule":

> "*When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change*

> *takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.*" (Hebb, 1949; p.62)

This postulate was later famously shortened to "*neurons wire together if they fire together*" (Löwel and Singer, 1992). The resulting strengthened connections would later enable the same neurons to also fire together in the absence of the stimulus, which Hebb believed to correspond to the memory of that stimulus.

Although this "cell assembly hypothesis" thus implies that representations of memories depend on specific (combinations of) neurons, it successfully circumvented the earlier raised criticisms of localized memories. First, the cell assemblies were proposed to be widely distributed across the brain, and to have a certain degree of redundancy, which Hebb argued could account for Lashley's failure to experimentally find the memory engram based on localized lesions: "*[t]he assembly is thought of as a system inherently involving some equipotentiality in the presence of alternate pathways each having the same function, so that brain damage might remove some pathways without preventing the system from functioning*" (Hebb, 1949; p.74). Second, although Hebb assumed memories to depend on the excitation of particular neurons at some point in the brain, such an assumption was not made or needed for the lower sensory areas. He predicted memory-representing cell assemblies to be created in higher processing areas due to the convergence of lower level inputs: "*the specificity of such an assembly of cells in 18 or 20* [i.e., higher level processing areas], *to a particular excitation in 17* [i.e., lower level visual area], *depends on convergences*" (Hebb, 1949; pp.72-73).

To counter the doubts about the computational ability of networks of synaptically connected neurons, Hebb heavily relied on two recent discoveries by the Spanish neurophysiologist Rafael Lorente de Nó, both of which he credits in his book as being instrumental for the development of his theory. Work by Lorente de Nó had revealed, first, that reciprocal connections between neurons are abundant in the central nervous system (e.g., Lorento de Nó, 1938) and, second, that typically excitation of multiple neurons was needed to excite a post-synaptic neuron, a phenomenon described as "optional" synaptic transmission

(Lorente de Nó, 1939). As Hebb recognized, these discoveries allowed neuronal networks to go beyond simple linear computations. In his book, he made a convincing case that his postulated learning rule of "fire together, wire together" could enable the storage and retrieval of information in such an extended neuronal network with reciprocal connections and optional synaptic transmission.

At the time Hebb wrote his book, it was not possible to use computer simulations to assess the computational feasability of his model. Even after technology had advanced to a stage that this became possible, however, Hebb himself allegedly never set finger to a computer to test his predictions (Milner, 1993). But others did and it was quickly pointed out that in its original form Hebb's model was too simple to work, most notably because inhibition was a crucial missing factor (see chapter 5; Rochester et al., 1956; Milner, 1957). Over the last 60 years, various revisions to the model have been proposed in the light of evolving neurobiological or computational insights. The modern interpretation of cell assemblies, for example, is quite different from how Hebb originally envisioned them (see section 3.1). But his fundamental idea, that the brain is capable of storing and retrieving information due to activity dependent changes in synaptic strength that create groups of connected neurons whose coordinated activity comes to represent mental concepts, proved to be visionary and is regarded as one of the main conceptual pillars of modern neuroscience.

In the 1970s, two initially separate lines of experimental research, both predominantly focussed on the hippocampus, would provide empirical support for Hebb's theory: synaptic plasticity and place cells.

Synaptic plasticity

In a landmark study published in 1973, Timothy Bliss and Terje Lømo demonstrated in the rabbit hippocampus that high-frequency stimulation of the perforant path fibres to the dentate gyrus results in a long-lasting potentiated population response of the dentate granule cells to subsequent stimulations (Bliss and Lømo, 1973; see also Lømo, 1966; Bliss and Gardner-

Medwin, 1973). Although this phenomenon, which was later termed long-term potentiation (LTP; Douglas and Goddard, 1975), was shown to be mediated by increases in synaptic strength, perhaps surprisingly no mention was made in these early papers about the possibility of this being the biological mechanism of Hebb's postulated learning rule. This changed when two papers shortly after each other reported that LTP had the property of "associativity" (McNaughton et al., 1978; Levy and Steward, 1979; although the first paper used the term "cooperativity", see McNaughton, 2003). These studies showed that while a weak stimulation applied in isolation does not result in LTP, the simultaneous application of multiple weak stimulations does. The molecular mechanisms underlying this associative property of LTP were discovered in quick succession throughout the 1980s (reviewed by Nicoll, 2017). It was consecutively shown that LTP depends on NMDA-receptors (Collingridge et al., 1983; Harris et al., 1984), that NMDA-receptors are voltage-dependent and therefore only capable of being activated when the post-synaptic cell is depolarized (Nowak et al., 1984; Mayer et al., 1984), that NMDA-receptors—unlike AMPA-receptors—are Ca^{2+} permeable (MacDermott et al., 1986; Ascher and Nowak, 1988) and that an increased Ca^{2+} concentration triggers LTP by autophosphorylation of the protein kinase CaMKII (Miller and Kennedy, 1986).

The discovery of NMDA receptor-dependent LTP provided evidence that the by Hebb postulated learning rule indeed exists in the brain, making his cell assembly hypothesis—besides theoretically appealing—biologically plausible. Moreover, using a spatial learning task called the "Morris water maze" (Morris, 1981), LTP was shown to be required for spatial memory. For this task, rats or mice are placed in a circular pool filled with opaque water in which an escape platform is hidden at a fixed location. Over multiple trials animals learn to locate the platform and swim straight towards it. Performance is then measured in a "probe test", during which the platform is removed, by the time animals swim around the location where the platform had been. Earlier it had been shown that rats with bilateral hippocampal lesions have impaired performance on this task (i.e., they spent less time around the location where the platform had been) compared to control rats (Morris et al., 1982). Blocking LTP—

either by infusion of the NMDA receptor antagonist AP5 (Morris et al., 1986), by selective knockout of a critical NMDA receptor subunit in CA1 pyramidal cells (Tsien et al., 1996a) or by genetically blocking autophosphorylation of CaMKII (Giese et al., 1998)—similarly impaired performance on this task.

Place cells

A second important development in the 1970s was the discovery of "place cells" by John O'Keefe and his student Jonathan Dostrovsky (O'Keefe and Dostrovsky, 1971). When performing single-unit recordings from the dorsal hippocampus of freely-moving rats, they found 8 units that "*responded solely or maximally when the rat was situated in a particular part of the testing platform facing a particular direction*" (O'Keefe and Dostrovsky, 1971; p.172). This remarkable finding has since been replicated by many labs (for reviews see O'Keefe, 1979; Muller, 1996) and in a wide variety of species (Ono et al., 1993; McHugh et al., 1996; Ekstrom et al., 2003; Ulanovsky and Moss, 2007). The part of the environment in which a place cell is maximally active—its "place field"—is stable within a single exposure session, but also upon re-exposure of the animal to the same environment hours, days or even months later (Muller et al., 1987; Thompson and Best, 1990). Large-scale recordings of ensembles of up to 148 hippocampal neurons demonstrated that the fields of place cells covered the entire explored environment, with in turn each location being associated with the co-activity of multiple place cells (Wilson and McNaughton, 1993, 1994). Between environments, place cells were found to "remap" in a seemingly random manner, in the sense that place cells with overlapping fields in one environment appeared not to have an increased likelihood of having overlapping fields in another environment (Muller and Kubie, 1987; Colgin et al., 2008). One implication of this is that in each spatial context, different groups of place cells are expected to be co-active.

These hippocampal place cells appear to have the main properties that Hebb postulated for his cell assemblies. The hippocampus is a "deep" area in the brain, so that the firing activity of its neurons is the result of multiple steps of convergences (Felleman and Van Essen, 1991).

Moreover, each group of co-active place cells signals a specific location and each spatial context has multiple of such co-active groups. But do these place cell assemblies indeed *represent* space? The evidence discussed so far indicates that place cell assemblies are active whenever the animal is in the location they presumably represent, but what about the reverse? Is there evidence that when a place cell assembly is active, the animal indeed thinks about or remembers the corresponding location?

Two studies by Clifford Kentros and colleagues provided evidence indicative of this. In a first study, they showed that pharmacologically blocking LTP—using the NMDA receptor antagonist CPP—impairs the long term stability of place cells, with no detectable effect on their initial expression or within-session stability (Kentros et al., 1998). As LTP had also been linked to spatial memory (e.g., Morris et al., 1986), this result indirectly associated long-term place cell stability and memory. Further correlative evidence was provided by a follow-up study in which they linked place cell stability to performance on a spatial memory task through attention (Kentros et al., 2004). Similarly, Barnes and colleagues (1997) had shown a correlation between place cell stability and spatial memory by comparing young and aging rats. Although all these studies are merely correlative, they suggest that a failure to reinstate a context's place cell map corresponds to an impaired ability to remember that context.

Recently, an elegant study by the lab of Karim Benchenane managed to go beyond correlative evidence (de Lavilléon et al., 2015). Conditioned upon the spiking activity of a selected place cell they performed electrical stimulations of the medial forebrain bundle in mice, which is known to induce a feeling of reward (Kobayashi et al., 1997). When performed during exploration, this stimulation protocol caused mice to develop a clear preference for the location corresponding to the field of the selected place cell. However, as during exploration it is not possible to dissociate between place cell activity and the location of the animal, from this experiment it was not clear whether the artificial place-preference was due to the coupling of the rewarding stimulation with the place cell activity or with the animal's location. Importantly, therefore, when the same stimulation protocol was performed while the animal slept (when

place cells are reactivated without the animal being in their place field, see next section), mice also developed a place preference for the field of the selected place cell. This result provides convincing evidence that place cell activity indeed represents (a memory of) the location of their place field.

Genetic tagging of active neurons

Perhaps the most recent evidence that memories are represented in the brain by the combined activity of groups of specific neurons comes from a fast growing line of research relying on the genetic tagging of active neurons. For this approach, promoter elements of immediate-early genes (IEGs; e.g., *cfos*, *arc* or *zif268*), a class of genes that is transiently up regulated in neurons in response to high firing activity (Cole et al., 1989; Guzowski et al., 2005), are engineered into transgenes to make the expression of specific proteins (e.g., fluorescence indicators and/or optogenetic actuators) under their control (Smeyne et al., 1992; Reijmers and Mayford, 2009). The combination of this technique with the tetracycline transactivator system, which limits a gene's potential expression to a period when Doxycycline is absent from the animal's diet (Aiba and Nakao, 2007), together with an addition that subsequently reinforces expression of the chosen protein(s) in cells in which the IEG was "turned on" during that period (Reijmers et al., 2007), enables the long-lasting tagging of neurons that are active during an experimenter controlled time window.

Reijmers and colleagues (2007) used this "TetTag" technique with the marker protein tau-LacZ to show that neurons in the basolateral amygdala of mice that are activated during context fear conditioning have an increased tendency to also be activated during subsequent re-exposure to the same context. Importantly, the extent to which neurons were reinstated during context re-exposure correlated with the behavioural expression of the fear memory (i.e., the amount of freezing). Similar results were later found for neurons in the hippocampus and cortex (Tayler et al., 2013). These correlational findings suggest that the activity of the tagged neurons represents the fear memory. To causally test this hypothesis, the lab of the Japanese scientist

Susumu Tonegawa adopted the TetTag technique to express channelrhodopsin into neurons of the dentate gyrus that were activated during context fear conditioning. Later optogenetic activation of those neurons was sufficient to induce freezing (i.e., recall of the fear memory), while similar activation of neurons that were tagged in a neutral context did not cause such a freezing response (Liu et al., 2012). This study indicates that activity of a specific group of neurons can be sufficient for the recall of a memory (i.e., if activated, the mouse remembers). That such activity can also be necessary (i.e., if not activated, the mouse can not remember), was demonstrated by two other groups that showed that optogenetically silencing hippocampal neurons that were active during context fear conditioning impaired recall of that fear memory (Denny et al., 2014; Tanaka et al., 2014).

The studies discussed above suggest that at least some memories are represented in the brain by the activity of specific groups of neurons. An elegant recent study used the ability to tag such memory-representing neuronal ensembles to probe the question how these memory representations are stored in the brain (Ryan et al., 2015). They tagged dentate gyrus neurons activated during fear conditioning and showed that these neurons had increased strength of their synaptic inputs and higher dendritic spine density compared to untagged neurons. Administering the protein synthesis inhibitor anisomycin immediately after fear conditioning both impaired subsequent recall of the fear memory and prevented the boosted synaptic input strengths and dendritic spine density of tagged neurons. This finding suggests that memories are stored in increased synaptic connections to, and maybe also from, the neurons involved in the memory representation. Strikingly, following anisomycin-induced amnesia, optogenetic activation of the tagged neurons restored the previously lost fear memory. This indicates that the amnesic effect of inhibiting protein synthesis was due to the prevention of changes in synaptic connections *to* the tagged neurons[3]. The relevant connections *from* those tagged neurons either were already in place (i.e., activation of those neurons would have always caused

[3] That is, in this experiment the memory impairment reflects an inability to excite the correct neuronal representation in the dentate gyrus, as opposed to a difference in the subsequent downstream effect of that neuronal representation.

freezing, also already before the fear conditioning) or they were formed upon fear conditioning in a protein synthesis independent manner. Future studies using this promising approach might be able to further elucidate the mechanisms underlying the storage of memory representations.

1.4) Reactivation as consolidation mechanism?

To briefly summarize the previous section, memories are thought to be *represented* by the activity patterns of cell assemblies and *stored* in strengthened synaptic connections to and/or from neurons involved in those patterns. The process of memory storage, therefore, should involve increases in the strength of specific synaptic connections, which generally require repeated pairing of pre- and post-synaptic activity. This suggests that for a pattern to be properly stored, it needs to occur multiple times. To increase the number of times a particular pattern is repeated, Donald Hebb (1949) hypothesized the existence of "reverberatory activity". He proposed that quick initial changes in synaptic strengths could create neuronal loops in which activity patterns representing a particular experience could be replayed for an extended period of time with as goal to "*prolong the time during which structural changes of learning can occur*" (Hebb, 1949; p.60).

Several decades later, David Marr (1970, 1971) also proposed that neuronal activity patterns representing a particular experience could persist in the brain beyond the experience itself. However, Marr thought that these reactivations did not need to be "reverberatory" in the sense that they had to be continuous from the moment of experience, but that they could also be initiated later, for example during sleep. In addition, Marr believed that the goal of such offline reactivation was not only to strengthen the memory-representing cell assemblies themselves, but also to "transfer" memories from temporary storage in one brain region (the hippocampus) to permanent storage in another (the cortex). This hypothesis of a hippocampal-cortical memory transfer was refined by later models of memory consolidation to a gradual reorganization of memory traces between both brain structures (McClelland et al., 1995; Nadel and Moscovitch,

1997; Winocur and Moscovitch, 2011). In these models, an important role ascribed to reactivation is enabling "interleaved learning" to allow the cortex to gradually extract shared structure and statistical regularities from individual experiences (McClelland et al., 1995).

Reactivation in rodent hippocampus

The first empirical evidence for reactivation was reported by Pavlides and Winson (1989) based on place cell recordings from the CA1 region of the rat hippocampus. They showed that if a rat was restricted to exploring a specific part of an enclosure, cells with their place field in the explored area had a higher firing rate during the subsequent sleep compared to cells with place fields in the restricted part of the enclosure. Although these first demonstrated differences in sleep firing rates were still far from conclusive evidence for the offline reactivation of specific memory traces, they demonstrated that hippocampal firing patterns during sleep are influenced by preceding waking experience. By simultaneously recording the activity of up to 74 hippocampal principal neurons, Wilson and McNaughton (1994) managed to next show that pairs of neurons with overlapping place fields during exploration of an open field enclosure had an increased tendency to fire together (in 100 ms time bins) during the subsequent sleep session (as compared to during a preceding sleep session), something that was not observed for pairs of neurons with non-overlapping place fields. To test whether temporal biases in firing relations between hippocampal neurons during waking experience would also be conserved during subsequent sleep, Skaggs and McNaughton (1996) allowed rats to run on a triangular track so that place cells would be activated in stereotyped sequences. They found that the temporal asymmetry of cross-correlations between pairs of neurons during track running indeed tended to recur during the subsequent sleep session (albeit compressed in time by a factor 10 to 20), while that same structure of temporal asymmetries was not detectable during a baseline sleep session recorded before the track running. Together, these three landmark studies provided the first evidence for the "replay" of temporally structured neuronal activity patterns of waking experience during subsequent sleep. Evidence for reactivation within the rodent hippocampus

has since quickly accumulated (for reviews see Sutherland and McNaughton, 2000; O'Neill et al., 2010). Among others, it has been demonstrated that replayed sequences of place cells are biased towards sequences that represent actual trajectories (Lee and Wilson, 2002), and that the more an animal visits a specific location, the more the place cells corresponding to that location will be reactivated during the subsequent sleep (O'Neill et al., 2008). While all of the above-discussed studies found reactivation (sometimes only) during periods of slow-wave sleep or awake rest, one study also detected hippocampal reactivation during REM sleep (Louie and Wilson, 2001). Finally, although the initial reactivation studies were all based on recordings from rats, reactivation has also been demonstrated in the mouse hippocampus (Nakashiba et al., 2009; McNamara et al., 2014).

In line with the hypothesis of reactivation being important for the consolidation of newly formed memories, it was reported that offline reactivation in the rodent hippocampus is increased after exploration of novel compared to familiar environments (O'Neill et al., 2008; McNamara et al., 2014) and that it is stronger after rewarded learning (Singer and Frank, 2009). Moreover, both hippocampal reactivation and performance on spatial memory tasks were found to be impaired in aging animals (Gerrard et al., 2008) and after genetically blocking CA3-output (Nakashiba et al., 2009). Finally, the offline reactivation of patterns related to specific goal-locations was shown to predict behaviour of rats in a subsequent memory task (Dupret et al., 2010).

Sharp wave-ripples (SWRs)

A consistent observation regarding reactivation in the rodent hippocampus is that it is strongest during sharp wave-ripples (SWR) events (Wilson and McNaughton, 1994; Kudrimoti et al., 1999; Nadasdy et al., 1999; O'Neill et al., 2008; Nakashiba et al., 2009). SWRs are short-lived (~40–120 ms) events in the hippocampal local field potential (LFP), best characterized by high-frequency (~125–250 Hz) 'ripple oscillations' in the CA1 pyramidal cell layer, occurring preferably (incidence rate: ~0.5–2 Hz) during periods of slow-wave sleep or awake rest

(O'Keefe, 1976; Buzsáki, 1986; Buzsáki et al., 1992; for review see Buzsáki, 2015a). SWRs are thought to be generated locally within the hippocampal formation by the synchronous discharge of CA3 cells (Buzsáki et al., 1983; Sullivan et al., 2011). These CA3 activity bursts are propagated by the Schaffer collateral fibres to the *stratum radiatum* of the CA1 region, where they cause a large amplitude, negative polarity deflection in the LFP called "sharp wave" (Buzsáki et al., 1983). This burst of excitation in turn triggers a near-synchronous discharge of cells in the CA1 region (Csicsvari et al., 2000), during which fast interactions between interneurons and principal cells cause the above described ripple oscillations in the pyramidal cell layer (Stark et al., 2014).

A couple of years before the empirical demonstration of reactivation, SWR events had been identified as ideal candidates for the offline stabilization of recently formed neuronal memory traces (Buzsáki, 1989). The origin of SWRs in CA3 was argued to be well suited for these SWRs to reflect previously expressed neuronal patterns, thanks to the pattern completion capability of the recurrent CA3 network (McNaughton and Morris, 1987). In addition, the high neuronal synchrony during SWRs (Buzsáki et al., 1983) and the similarity of the frequency of ripple oscillations to the electrical stimulation frequency used to experimentally induce LTP (Douglas, 1977), suggested SWRs to have favourable conditions for inducing synaptic plasticity, as was confirmed by a recent *in vitro* study (Sadowski et al., 2016).

Correlational evidence for an involvement of SWRs in memory consolidation came from studies showing that the incidence rate of SWRs increases after learning (Eschenko et al., 2008) and predicts subsequent memory performance (Ramadan et al., 2009). To causally test the role of SWRs, a closed-loop feedback system was used to trigger electrical stimulation of the ventral hippocampal commissure upon on-the-fly detected SWRs. Such stimulation induces a brief, highly synchronous discharge of a large number of pyramidal cells and interneurons, followed by a transient (~100 ms) inactivation of the hippocampal network (Zugaro et al., 2005). Two studies showed that disrupting SWRs in this way during the sleep following spatial learning causes small impairments in subsequent memory performance (Girardeau et al., 2009;

Ego-Stengel and Wilson, 2010), thus indicating a causal role for SWRs in memory consolidation. However, whether it is the reactivation during these SWRs that is important, or some other characteristic of SWRs, remains to be confirmed.

Hippocampal-cortical dialogue

As discussed above, offline reactivation is (also) proposed to be involved in the reorganization of memories, involving their transfer or transform from the hippocampus to distributed cortical areas (Marr, 1970, 1971; McClelland et al., 1995; Squire and Alvarez, 1995; Moscovitch et al., 2006; Winocur and Moscovitch, 2011). A direct prediction of this hypothesis is that hippocampal reactivation should be coordinated with reactivation in the cortex. Such joint hippocampal-cortical reactivation was indeed empirically demonstrated by Qin et al. (1997), who simultaneously recorded from the CA1 hippocampus and the posterior parietal cortex of freely-behaving rats. Later studies also found coordinated reactivation between hippocampal neurons and neurons recorded from the visual cortex (Ji and Wilson, 2007), the ventral striatum (Lansink et al., 2009), the auditory cortex (Rothschild et al., 2017) and the entorhinal cortex (Ólafsdóttir et al., 2016; but see O'Neill et al., 2017). Moreover, reactivation of learning-related assembly patterns of prefrontal neurons was found to be coordinated with the occurrence of hippocampal SWRs (Peyrache et al., 2009).

Causal evidence that a hippocampal-cortical coupling during sleep supports memory consolidation was provided by a recent study from the lab of Michaël Zugaro. They managed to improve the memory performance of rats on an object recognition task by delivering brief electrical pulses to the cortex triggered by on-the-fly detected SWRs during the sleep following encoding, while if those same pulses were delivered with a random 160–240 ms delay after on-the-fly SWR detection there was no such improvement (Maingret et al., 2016). The ideas underlying this study are that these brief electrical pulses evoke delta waves, followed by spindles, that propagate across the cortex (Vyazovskiy et al., 2009), and that hippocampal SWRs closely followed by such cortical delta spindles provide a window of opportunity for the

reorganization of memories between the hippocampus and cortex (Staresina et al., 2015). This study provides a vital starting point for the causal investigation of the hippocampal-cortical dialogue in memory consolidation, but more research is needed to unravel the involved neuronal mechanisms and underlying computational principles.

1.5) Overview

Based on the literature reviewed above, the following model for memory consolidation emerges. Upon encoding, new memories are represented by the coordinated activity of groups of neurons called cell assemblies and stored in (weakly) strengthened connections to and from those neurons, brought about by (an initial wave of) synaptic plasticity. Later recall of these memories involves the reinstatement of the activity of these cell assemblies. To enable this reinstatement, cell assemblies are thought to be reactivated during offline behavioural periods such as sleep or rest, which is believed to induce additional synaptic plasticity to further strengthen the connections needed for the assemblies' reinstatement.

Although this model is widely accepted and there is—as reviewed in this chapter—evidence for many of its parts, one central piece of evidence is missing. The hypothesis that the offline reactivation of "memory-representing" cell assemblies promotes the later reinstatement of those assemblies during memory recall has never been directly tested. To do so has been the main aim of my DPhil project.

In order to be able to test this hypothesis, there were two major challenges:

1. **It required the identification and tracking of "memory-representing" cell assemblies.** For this, I performed multi-unit recordings from the hippocampus of freely-behaving mice, a technique that is described in the first part of chapter 2. To identify and track cell assembly patterns in these recordings, I used a statistical assembly detection method that is described and discussed in chapter 3.

2. **In order to have a causal test, it required the selective disruption of offline reactivation.** To achieve this, I performed optogenetic silencing of hippocampal principal neurons during on-the-fly detected SWRs. This technique is described in the second part of chapter 2.

Chapter 4 then describes and discusses the results of applying these tools to test whether offline reactivation contributes to the stability of hippocampal cell assembly patterns. In chapter 5, I take another turn, and ask what different roles interneurons play in the formation, organisation and stabilisation of assemblies of principal neurons. Chapter 6 wraps up, discusses limitations of the experiments described in this thesis and highlights directions for further research.

CHAPTER 2: EXPERIMENTAL PROCEDURES

2.1) Multi-unit recordings from hippocampus of freely-behaving mice

Subjects

Except for chapter 5, all animals used to obtain the results described in this thesis were male adult transgenic CaMKIIa-Cre mice (4–7 months-old; Jackson Laboratories, B6.Cg-Tg(Camk2a-cre)T29-1Stl/J, stock number 005359; RRID: IMSR_JAX:005359). Animals had free access to water and food in a dedicated housing facility with a 12/12 h light/dark cycle. They shared a cage with their littermates until their first surgery, after which they were housed alone. Surgical procedures were performed under deep anaesthesia using isoflurane (0.5–2 %) and oxygen (2 l/min), with analgesia (0.1 mg/kg vetergesic or 5 mg/kg metacam) provided before and after. All experiments involving animals were conducted according to the UK Animals (Scientific Procedures) Act 1986 under personal and project licenses issued by the Home Office following ethical review.

Microdrive implantation

All recorded mice were implanted with a custom-made microdrive. The drive was designed to bilaterally target the pyramidal layer of dorsal CA1, with for each hemisphere one optic fibre (230 µm diameter, Doris Lenses) surrounded by five independently moveable tetrodes. Both the distance from the optic fibre to each of the tetrodes and the distance between neighbouring tetrodes was 0.4 mm. Tetrodes were constructed by twisting together four insulated tungsten wires (12 µm diameter, California Fine Wire) and shortly heating them to bind them together in a single bundle. Each tetrode was attached to a 6 mm long M1.0 screw to enable their independent movement. The drive was implanted under stereotaxic control (coordinates for optic fibre tips: 2 mm posterior, ±1.7 mm lateral, 1.1 mm ventral). Tetrodes were placed ~500 µm dorsal to the tip of the optic fibres so that they were initially above the CA1 pyramidal

layer. Following the implantation the exposed parts of the optic fibres and tetrodes were covered with paraffin wax, after which the drive was secured to the skull using dental cement. For extra stability, four stainless-steel anchor screws had first been inserted into the skull. Two of the anchor screws, which were inserted above the cerebellum, were attached to 50 µm tungsten wires (California Fine Wire) and served as ground and reference electrodes during the recordings. The placement of the tetrodes in dorsal CA1 was daily confirmed by the electrophysiological profile of the local field potential in the hippocampal ripple frequency band (Csicsvari et al., 1999a, 2000; see **Figure 2.1A,C,D**), and at the end of the procedure by the anatomical verification of the tetrode tracks (see below and **Figure 2.1B**).

Anatomical confirmation of tetrode placement

At the completion of the experiment, all implanted mice were deeply anesthetized with isoflurane/pentobarbital and transcardially perfused with 0.1 M phosphate-buffered saline (PBS) followed by cold 4 % paraformaldehyde (PFA) dissolved in PBS (McNamara et al., 2014). Implanted tetrodes were kept in the brain for at least 4 days of post-perfusion fixation in 4 % PFA before being removed and the brain being extracted. Brains were then kept in 4 % PFA for at least 24 h before slicing. In order to visualise tetrode tracks, brains were sliced into 100 µm thick coronal sections and stained using osmium. To do so, brain sections were rinsed extensively in 0.1 M phosphate buffer (PB) before being immersed in 4 % PFA with 0.05 % glutaraldehyde for 15 min. Sections were then rinsed three times in PB for 10 min and immersed in PB with 0.5 % osmium for 1 h. Sections were rinsed again in PB and dehydrated by successive immersion in alcohol and propylenoxid. Sections were finally mounted on slides and cover-slipped. The images were acquired on a microscope (Axio Imager M2; Zeiss) with 5x/0.1 NA objective and mounted colour camera (Retiga 2000R; QImaging) under bright-field illumination. They were converted to greyscale, inverted and contrasted to best illustrate the layers of the hippocampus and the tetrode tracks (see **Figure 2.1B**).

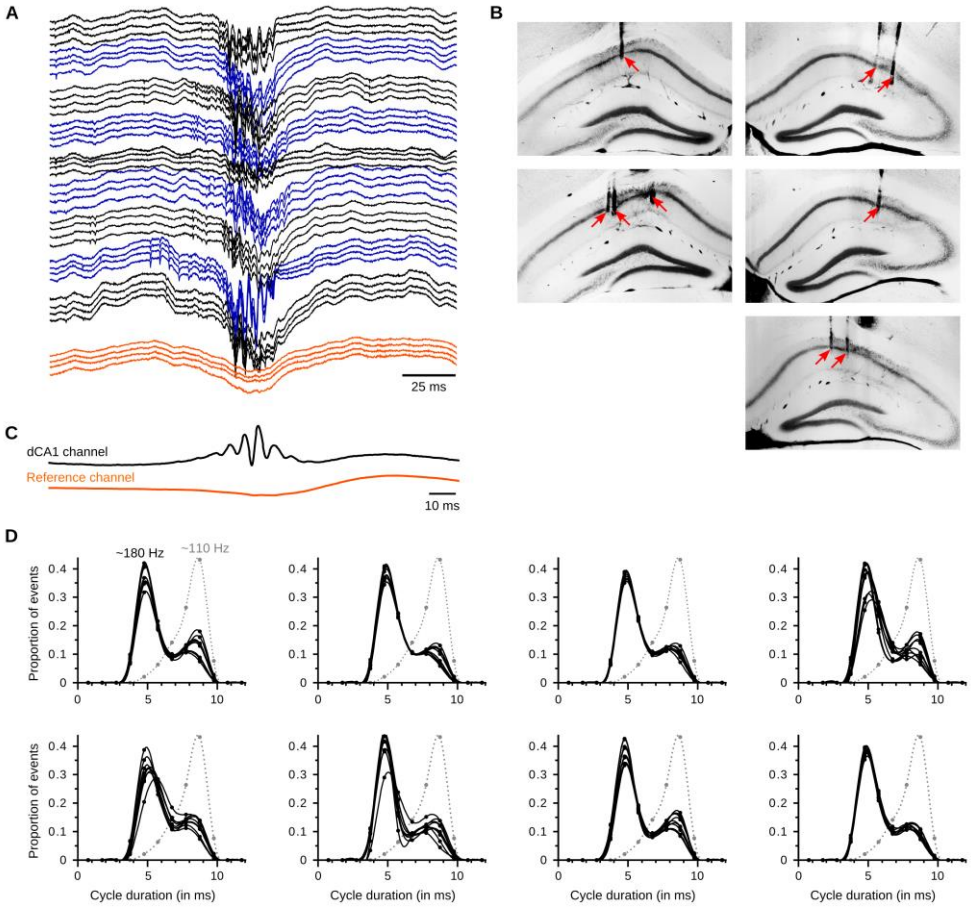

Figure 2.1 *Recordings were performed from the CA1 region of the dorsal hippocampus as shown from the tetrode tracks and the SWR electrophysiological profile.*

(A) Example wideband local field potential signals from all tetrodes of one of the recorded mice. Nine of these tetrodes were daily lowered to the pyramidal cell layer of the CA1 region of the dorsal hippocampus (dCA1), as indicated by the presence of SWR events. For clarity, the signal from all channels of the same dCA1 tetrode (4 wire-channels per tetrode) is colour coded in black or dark blue. The tenth tetrode (orange) was left above the hippocampus and used as a reference for SWR detection (see **Figure 4.4B**).
(B) Pictures of the osmium-stained coronal brain sections showing the tracks from the nine dCA1 tetrodes used to record the raw traces shown in (A). Red arrows indicate the tip of the tetrodes.
(C) Average SWR-triggered waveforms from a channel of a dCA1 tetrode and of the reference tetrode. Note the presence of ripples (~5–6 ms long cycles) in the dCA1, but not the reference, tetrode.
(D) Distribution of the cycle duration of the fast oscillatory events detected from the hippocampal tetrodes of each recorded mice (one panel per mouse; each tetrode from which principal neurons were recorded is represented by one black curve; top-left panel is of the same mouse as in (A-C)). The oscillatory frequency (in Hz) of the detected events can be calculated by $1000/x$, with x the cycle duration (in ms). For visualisation purposes, in each panel the consecutive data points of each tetrode are connected via a cubic spline curve after rendering the data monotonic. Note the bimodal distribution of the oscillatory events detected from all recorded hippocampal tetrodes (black data points and connecting curves) with ~180 Hz and ~110 Hz frequency peaks, which is consistent with their dCA1 location (Csicsvari et al., 1999, 2000). For comparison, the cycle duration distribution of the fast oscillatory events detected from a tetrode located in the CA3 region of the dorsal hippocampus of another mouse (not recorded by myself; gray data points and dotted curve) is shown on each panel; note in this case the unimodal distribution of the oscillatory events with a ~110 Hz frequency peak, consistent with previous studies (Csicsvari et al., 1999, 2000).

[Note that this figure was created by Natalia Campo-Urriza and David Dupret.]

Recording protocol

After recovery from the implantation surgery, each animal was connected to the recording apparatus and familiarized with a 12 x 12 cm high-walled box containing home cage bedding (the "sleep box") and with one of the open-field enclosures (the familiar enclosure) over a period of approximately seven days. During this period, tetrodes were gradually lowered to be positioned in the *stratum oriens*, above the CA1 pyramidal layer. On the morning of each recording day, tetrodes were lowered into the pyramidal cell layer in search of multi-unit spiking activity and SWR events (see **Figure 2.1A**). Tetrodes were not moved for at least 2 h before recordings started. At the end of each recording day, tetrodes were raised back to the *stratum oriens* to avoid damaging the pyramidal layer overnight. Note that although tetrodes were daily moved in and out of the pyramidal cell layer, it is possible that some cells were recorded on multiple days.

Each recording day, hippocampal network activity was continuously monitored as the animal completed one to three "recording blocks" (see **Figures 4.1B** and **4.5A**). Each day started with a 26 min period in the sleep box ("rest before"). For each recording block, the animal was then allowed to explore an open-field enclosure for 26 min ("exposure"), followed by 1 h in the sleep box ("rest after") before being allowed to again explore that enclosure for another 26 min ("re-exposure"). The open-field enclosure was either the familiar enclosure, which the animal had repeatedly been exposed to before, or a novel enclosure the animal had never seen before. The open-field enclosures differed in shape and in the cue-cards that lined some of the walls, but they all had an area of ~0.25 m^2. All enclosures were placed in the same position within the same recording room and were surrounded by a black curtain to prevent the animal from using distal cues as a common reference frame for all enclosures. In some of the recording blocks, the animal received optogenetic SWR silencing or optogenetic random silencing during the 1 h period in the sleep box between both exposures to the same enclosure (see below). The particular order of the "familiar" and "novel" recording blocks, as well as that of the optogenetic silencing and no silencing sleep/rest sessions, was varied between recording

days. The results described in this thesis are based on a total of 93 recording blocks: 17 with the familiar enclosure without light delivery, 13 with the familiar enclosure with SWR silencing, 26 with a novel enclosure without light delivery, 24 with a novel enclosure with SWR silencing and 13 with a novel enclosure with random silencing.

Data acquisition and position tracking

The extracellular signals picked up by each electrode were buffered directly above the head of the animal and sent over individual wires to a dual-stage amplifier (Sensorium Inc.). After amplification (1000x) and band pass filtering (0.1 Hz to 5 kHz) the signals were digitized at 20 kHz and saved to a computer along with synchronization signals from the position tracking and laser stimulation. To track the location (and estimate the locomotion speed) of the animal, three LED lights (red, green and blue) were attached to the cable directly above the head of the animal and traced by an overhead colour camera capturing 25 frames per second.

2.2) Optogenetic silencing of principal neurons during SWRs

Viral vector injection for ArchT expression

The light-controlled, outward proton-pump ArchT (Han et al., 2011) was expressed in hippocampal principal neurons using a Cre-loxP approach. Roughly two weeks before the implantation surgery, CaMKII-Cre mice (Tsien et al., 1996b) were injected with a Cre-recombinase-dependent adeno-associated viral (AAV) vector containing ArchT-GFP (rAAV2-CAG-FLEX-ArchT-GFP; UNC Vector Core) in the dorsal CA1 of their hippocampus, under stereotaxic control (Franklin and Paxinos, 2007). At four sites (coordinates: 1.7 / 2.3 mm posterior and ±1.7 mm lateral from bregma, and 1.1 mm ventral from the brain surface), 0.4 µl of the viral vector was injected at a rate less than 0.1 µl/min. Injections were performed with 5 µl calibrated glass micropipettes (Blaubrand), which were pulled and broken to have tip diameter ~17–20 µm.

Verification of ArchT-GFP expression

To confirm the injection site and specificity of the viral vector, 6 weeks after the injection surgery two CaMKII::ArchT mice were deeply anesthetized with isoflurane/pentobarbital and transcardially perfused with PBS followed by cold 4 % PFA and 0.1 % glutaraldehyde dissolved in PBS. The brains were extracted and sliced into 50 µm thick coronal sections. The sections were rinsed extensively in PBS with 0.25 % Triton X-100 (PBS-T) and blocked for 1 h at room temperature in PBS-T with 10 % normal donkey serum (NDS). Sections were then incubated at 4 °C for 72 h with primary antibodies (chicken anti-GFP, 1:1000, Aves Labs, cat# GFP-1020, RRID: AB_10000240 and rabbit anti-Wfs1, 1:500, Proteintech Group, cat# 11558-1-AP, RRID: AB_2216046) diluted in PBS-T with 3 % NDS ("blocking solution"). After that, sections were rinsed three times for 15 min in PBS-T and incubated for 4 h at room temperature in secondary antibodies (donkey anti-chicken Alexa Fluor 488, 1:500, Jackson ImmunoResearch, cat# 703-545-155, RRID: AB_2340375 and donkey anti-rabbit Cy3, 1:1000, Jackson ImmunoResearch, cat# 711-165-152, RRID: AB_2307443) diluted in the blocking solution. This step was followed by three rinses for 15 min in PBS. Sections were then incubated for 1 min with 4',6-diamidino-2-phenylindole (DAPI; 0.5 µg/ml, Sigma-Aldrich, cat# D8417) diluted in PBS to label cell nuclei, before undergoing three additional rinse steps of 10 min each in PBS. Sections were finally mounted on slides, cover-slipped with Vectashield mounting medium (Vector Laboratories) and stored at 4 °C. Images were acquired using an epifluorescence microscope (Axio Imager M2; Zeiss) and Stereo Investigator software (Virtual Tissue 2D module; MBF Bioscience). Cell counting was performed on *z*-stacks obtained with a 40x/1.3 NA objective using Stereo Investigator and ImageJ (http://rsb.info.nih.gov/) software and was conducted from every fourth coronal section containing the dorsal CA1 hippocampus. The percentage of principal neurons transfected with ArchT-GFP was evaluated by the percentage of cells immunopositive for Wolfram syndrome 1 (Wfs1), a marker specific for CA1 principal neurons (Lein et al., 2007; Valero et al., 2015; Cembrowski et al., 2016), that were immunopositive for the GFP reporter. This revealed a high-level of viral transfection of dorsal

CA1 principal neurons (90.8 ± 2.5 %, *n* = 8 sections from 2 mice, with 25 Wfs1-positive cells per section; see **Figure 4.4C**).

SWR detection and optogenetic silencing

A 561 nm diode-pumped solid-state laser (Laser 2000, Ringstead) was used to deliver light (~15 mW) to the dorsal CA1 pyramidal layer of CaMKII::ArchT mice through two lengths of optic fiber and a rotary joint with splitter (Doric Lenses). Light pulses were delivered either conditioned to the on-line detection of SWR events or with randomly selected intervals. For the on-line SWR detection (see also **Figure 4.4B**), one channel from the tetrode with the most pronounced ripples ("SWR detection electrode") and one channel from a tetrode left in the cortex ("reference electrode") were selected. The amplified and filtered electrophysiological signal from these two channels was sent to a computer with real-time kernel. To remove noise present on all tetrodes (e.g., muscle artefacts), the signal of the reference electrode was first subtracted from the signal of the SWR detection electrode. Then, every 5 ms, one copy of the last 20 ms of the differential signal was band-pass filtered (125–250 Hz, 19^{th} order Butterworth filter), and another copy was convolved with a Morlet wavelet (160 Hz center frequency). If in the last 10 ms of this time window the average power (root-mean-square) of the band-pass filtered signal was more than 4 SD above its baseline and the maximum of the wavelet-convoluted signal was more than 3 SD above its baseline, a SWR was said to be detected. For both the average power and for the maximum of the wavelet-convoluted signal, the baseline and SD were calculated using the first 25 sec of the recording session. Upon detection of a SWR, a 50 or 80 ms light pulse was triggered followed by a 30 ms refractory period during which no new light pulse could be triggered. To evaluate the accuracy of the on-line SWR detection, SWRs were identified offline as previously described (McNamara et al., 2014). For the random silencing control experiment, the intervals between subsequent pulses were independently sampled from a uniform random distribution between 30 ms and X ms, whereby X was chosen for each animal to ensure the total number of random pulses to be equal to or higher than the

average number of pulses delivered in the SWR silencing condition (number of pulses during 1 h sleep session: SWR silencing = 1939 ± 121, n = 37 recording blocks; random silencing = 2469 ± 169, n = 13).

"Network response" to first-time repeated optogenetic silencing

In one recording block with random silencing, the optogenetic silencing triggered a pronounced and unusual network response (see **Figure 2.2A**), as for example manifested by an abnormal high occurrence (3.6 Hz) of ripple-like events during the first 10 min of active exploration in the subsequent re-exposure session. As such an extreme network response was not observed on any other days (ripple occurrence rate was >6.4 interquartile ranges above the second highest observed rate), this recording block was not included in the analyses presented in this thesis.

Upon further investigation of my data set, I found that one-hour long repeated optogenetic silencing (both random and SWR-triggered) more often caused an increase in ripple-like events at the start of the subsequent re-exposure session (see **Figure 2.2B**), although generally not as extreme as for the excluded recording block described above. Strikingly, the three recording blocks with the strongest "network response" were recorded from three different mice and all on that animal's first recording day. This suggests that the first time an animal's hippocampus receives repeated optogenetic silencing might result in a different (more pronounced) response than successive times[4]. Unfortunately, my current data set is not well suited to properly test this hypothesis, as most mice received a varying number of undocumented "test" pulses before their first experiment. This could indeed explain why only for 3 out of 7 mice, I found such a strong network response to repeated optogenetic silencing on their first recording day.

[4] It has to be said that for one of the animals with a pronounced network response ("gv05"), it was actually the second application of one-hour long repeated optogenetic silencing that triggered the subsequent abnormal high rate of ripple-like activity. The first recording block with optogenetic silencing on that recording day had not caused such a network response. This suggests that the adaptation of the network to repeated optogenetic silencing takes place at a timescale between a couple of hours to one day, and that preceding optogenetic silencing within a period of at least an hour might instead prime a subsequent network response. The other two animals with a pronounced network response ("gv03" and "gv04") only had one recording block with optogenetic silencing on their first recording day.

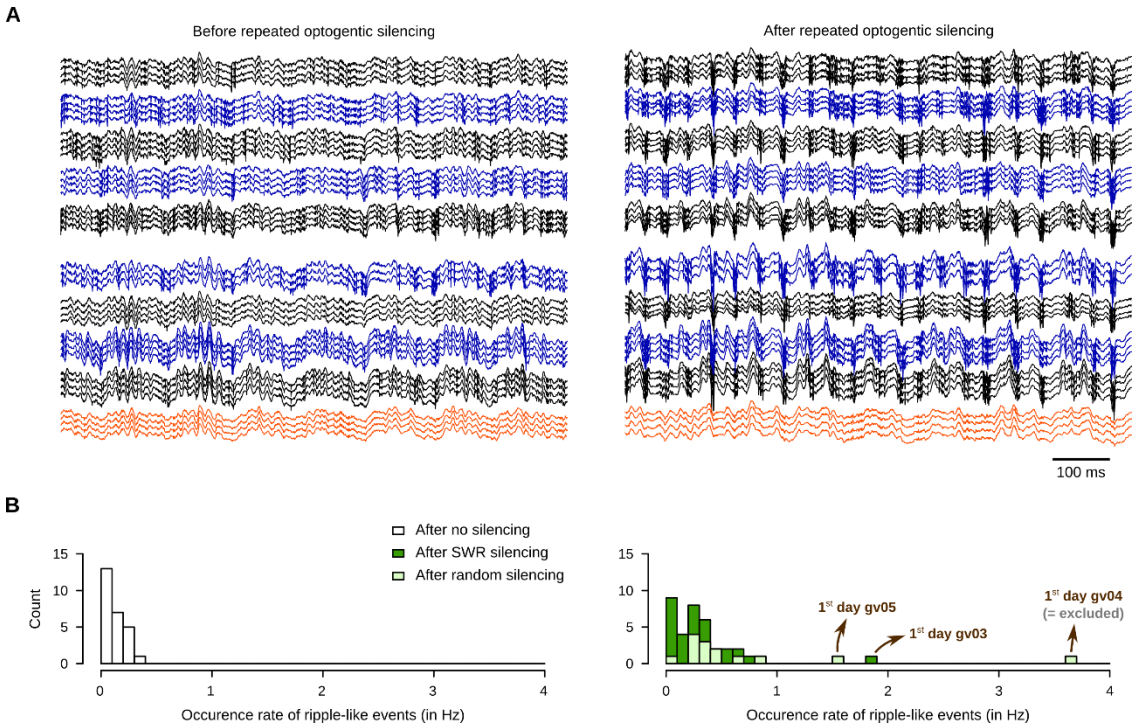

Figure 2.2 *Ripple-like events in the hippocampal LFP during active exploration directly following one hour of repeated optogenetic silencing.*

(A) Example traces of the recorded signals (of all 10 tetrodes) during active exploration before (left) and directly after (right) the one hour sleep/rest period with repeated optogenetic silencing from the excluded recording block (1st recording day of animal "gv04"). While before the repeated optogenetic silencing the recorded signals are dominated by theta oscillations characteristic for active exploration, afterwards the most noticeable features are the re-occurring short bouts of ripple-like activity. For clarity, signals from channels of the same hippocampal tetrode are colour coded in black or dark blue, with in orange the signals recorded from the reference tetrode which was left above the hippocampus.
(B) Distribution of the occurrence rate of ripple-like events during the first 10 min of the re-exposure session (during which animals actively explored the enclosure) after an one-hour long sleep/rest session either without (left) or with (right) repeated optogenetic silencing. The histogram on the right is further split up according to whether the light pulses were applied conditioned on on-line detection of SWRs (dark green) or with randomly selected intervals (light green). Note that repeated optogenetic silencing generally resulted in an increase in the number of ripple-like events. This effect was particularly pronounced for 3 recording blocks, which were from 3 different mice and which were all recorded on their animal's first recording day.

2.3) Data analysis protocols

Spike detection and unit isolation

For the offline detection of spikes, the recorded signals were first band-pass filtered (800 Hz to 5 kHz). Spikes were then detected based on the power (root-mean-square) of the filtered signal calculated in 0.2 ms sliding windows. Detected spikes of the individual electrodes were

combined per tetrode. To isolate spikes putatively belonging to the same neuron, spike-waveforms were first up-sampled to 40 kHz and aligned to their maximal through (Csicsvari et al., 1998). Principal component analysis (PCA) was applied to these waveforms ±0.5 ms from the through to extract the first three or four principal components per channel, such that each individual spike was represented by 12 waveform parameters. An automatic clustering program (KlustaKwik, http://klusta-team.github.io) was run on this principal component space[5] complemented with a variable indicating the time of each recorded spike. The resulting clusters were manually recombined and further isolated based on cloud shape in the principal component space, cross-channels spike waveforms, auto-correllogram and cross-correlation histograms (Harris et al., 2000). All sessions recorded on the same day were concatenated and clustered together. Each cluster used for further analysis showed throughout the entire recording day stable cross-channels spike waveforms, a clear refractory period in its auto-correllogram, well-defined cluster boundaries and an absence of refractory period in its cross-correlation histograms with the other clusters (McNamara et al., 2014). Yet, to control for possible mistakes in unit isolation, I repeated the central analysis of chapter 4 (i.e., regarding the effect of SWR silencing on assembly pattern reinstatement) after discarding all assembly patterns with two or more high contributing neurons recorded from the same tetrode (see **Figure 4.6A**). Moreover, for the neuron-pair analyses (see below), only pairs of neurons that were recorded from different tetrodes were included. For the analyses presented in chapter 3 and 4, principal neurons were manually identified from the resulting hippocampal units based on the shape of their auto-correllogram, their firing rate and their spike waveform (Csicsvari et al., 1998).

Spatial activity maps

To generate spatial activity maps from a point process of activation time stamps (e.g., a neuron's spike times or the activations of an assembly pattern), only times when the animal actively explored the environment were used (see McNamara et al., 2014). The horizontal plane

[5] If all four channels of a tetrode were "good", actually only the first two principal components of each channel were used for this automatic clustering step.

of the explored environment was divided into spatial bins of approximately 2 x 2 cm. Activation count maps (number of activations per spatial bin) and an occupancy map (time spent by the animal in each spatial bin) were generated, and all maps were spatially smoothed by convolution with a Gaussian kernel with standard deviation of one bin width. Smoothed activation count maps were then normalized by dividing them by the smoothed occupancy map to generate the spatial activity maps.

The coherence and sparsity of a spatial activity map (see **Table 4.1** and section 5.2) were calculated from the unsmoothed maps. Coherence, which reflects the similarity of the activation rate in adjacent bins, was calculated as Fisher's *r* to *z* transform of the Pearson correlation (across all visited bins) between the rate in a bin and the average rate of its eight nearest neighbours (Muller and Kubie, 1989). Sparsity, which reflects how tightly concentrated in space the activity is, was calculated as $(\sum P_i R_i)^2 / \sum P_i {R_i}^2$, where P_i is the propability of the animal occupying bin i and R_i is the activation rate in bin i (Skaggs et al., 1996). A spatial activity map's "field" (see **Table 4.1**) was calculated from the smoothed map and defined as those bins with an activation rate above its "field threshold", which was calculated as the rate of the bin with the highest rate minus 40 % of the difference in rate between the bins with the highest and lowest rate.

Neuron-pair and single-neuron measures

Although many of the analyses in this thesis are based on the "cell assembly patterns" described in the next chapter, some more conventional neuron-pair and single-neuron analyses are also performed. For each pair of principal neurons, the *co-firing coefficient* (see **Figure 3.2B**) was calculated as the Pearson correlation coefficient between the binned (25 ms time bins) spike-counts of the two neurons forming that pair (McNamara et al., 2014). To calculate the *place-field similarity* (PFS; see **Figures 3.2C**, **4.6B** and **4.8C**, and **Table 4.1**) scores between two neurons in a given exploratory session, spatial firing rate maps were first generated for both neurons. The PFS was then defined as the Pearson correlation coefficient from the direct

comparison of the spatial bins between the spatial firing rate maps of both neurons (McNamara et al., 2014). For the single-neuron analyses (see **Figures 4.6C** and **4.8D**, and **Table 4.1**), a given neuron's place field *similarity score* between two exploratory sessions was calculated as the Pearson correlation coefficient from the direct bin-wise comparison between that neuron's spatial firing rate maps from both sessions, whereby only spatial bins were included that were visited by the animal in both sessions (Kentros et al., 1998). For all neuron-pair and single-neuron analyses, only principal neurons with at least 100 spikes discharged during active exploration in each session considered were included. In addition, for the neuron-pair analyses, only pairs of neurons recorded from different tetrodes were included.

Statistical analyses

The means of two sets of observations were compared with a two-sample *t*-test or, if they were obtained from the same set of assembly patterns, with a paired *t*-test. To test for a differential effect of SWR silencing, a 2-way ANOVA was performed. Whether the average reactivation strength or the average environment-specificity index of a set of assembly patterns differed from zero was tested with a one-sample *t*-test. Testing whether a Pearson correlation coefficient (r) differed from zero was done with a standard *t*-test, with the standard error calculated as the square root of $(1 - r^2)/(n - 2)$, where n is the number of paired observations. Comparison of two Pearson correlation coefficients was done with a *z*-test, after application of Fisher's r to z transform (Zar, 1999; pp.386-387). Whether the proportion of patterns with higher expression strength during sleep/rest after than during sleep/rest before differed from the 50 % chance level was tested with a one-sample *z*-test for proportions. The linear regression models used contained an intercept and slope and were estimated using ordinary least squares. Whether the fitted slope differed from zero was tested with a *t*-test. All *t*-tests and *z*-tests were performed two-sided. Error bars are ± 1 standard error of the mean (SEM) or ± 1 standard error (SE) of the correlation coefficient, unless indicated otherwise in the legend. In the text reported group data are mean ± 1 SEM, unless stated otherwise.

CHAPTER 3: IDENTIFICATION AND TRACKING OF CELL ASSEMBLY PATTERNS

3.1) Introduction

The objective of this chapter is to identify cell assemblies from the multi-unit hippocampal recordings described in the previous chapter. As discussed in the first chapter of this thesis, Donald Hebb loosely defined a cell assembly as a group of neurons that tends to be active together and whose combined output represents a mental concept (Hebb, 1949). In addition, he hypothesized that the neurons making up these cell assemblies might form neuronal loops so that activity could reverberate in them.

Efforts to empirically pinpoint cell assemblies have predominantly focussed on the first aspect of Hebb's definition: finding groups of neurons that tend to be active together (e.g., Kudrimoti et al., 1999; Harris et al., 2003; Malvache et al., 2016). György Buzsáki recently formulated an elegant and compelling argument why it makes sense to focus on such co-active groups of neurons (Buzsáki, 2010). He reasoned that a neuron's spiking activity is relevant only in terms of its effect on downstream "readers", which most commonly are other neurons onto which that neuron makes synaptic connections. But since isolated activity of a single neuron typically is not enough to make a downstream neuron fire (Lorente de Nó, 1939), the relevant "effective unit" in the brain is the combined activity of groups of neurons. This "reader-centric" definition of cell assemblies also implies that the time scale of their expression (i.e., when should two spikes still be considered synchronous?) should correspond to the synaptic integration window of their target neurons, which is the time window over which neurons integrate their synaptic inputs. For principal neurons in the hippocampus and cortex, this integration window is in the range 10–30 ms (Spruston and Johnston, 1992; Koch et al., 1996).

Using tetrode recordings in the rat hippocampus, it was shown that the activity of principal neurons can be predicted from the activity of their peer neurons (atop the prediction from the location of the animal), and that this prediction is, indeed, optimal for time windows between 10 and 30 ms (Harris et al., 2003). This result is often cited as evidence for cell assemblies in the hippocampus, as it indicates the presence of groups of principal neurons that are active together at this time scale more often than expected by chance. This study did not, however, identify these putative cell assemblies nor did it provide evidence that their output represents mental concepts.

The approach to be taken in this chapter is to identify—from multi-unit hippocampal recordings—subsets of principal neurons that are repeatedly active together in 25 ms time windows (section 3.2), and then to validate those candidate cell assemblies by testing whether their output is consistent with them representing mental concepts (section 3.3). Section 3.4 deals with a potential criticism of the assembly detection method used in thesis. Section 3.5 discusses alternative approaches for cell assembly identification.

3.2) PCA/ICA-method

The challenge of this section is to find groups of hippocampal principal neurons that are active together during short (25 ms) time windows more often than expected by chance. Moreover, in order to be able to determine their behavioural correlates (and to evaluate the strength of their reactivation and reinstatement, see next chapter), a requirement is that it should be possible to track the expression of the detected assemblies over time. With these aims in mind, I selected a statistical framework for identification and tracking of assembly patterns that was recently proposed and tested on simulated hippocampal recordings by Vítor Lopes-dos-Santos and colleagues (2013).

Assembly pattern identification

To identify assembly patterns (see also **Figure 3.1A**) from a given multi-unit recording, the relevant spike train epochs are first divided into 25 ms time bins and for every neuron[6] its number of spikes in each bin is counted. In order not to bias the identification of assembly patterns towards neurons with higher firing rates, the binned spike counts are normalized for each neuron by a *z*-score transform:

$$z_{i,b} = \frac{x_{i,b} - \mu_{x_i}}{\sigma_{x_i}}$$

where $x_{i,b}$ is the spike count of neuron i in bin b, and μ_{x_i} and σ_{x_i} are respectively the mean and standard deviation of neuron i's spike counts across bins. This ensures that the normalised activity of each neuron has null mean rate and unitary variance. With the number of neurons denoted by n and the number of time bins by B, let $\mathbf{Z}$ be the $n \times B$ binned and *z*-scored spike count matrix with element (i, b) equal to $z_{i,b}$. From this matrix $\mathbf{Z}$, *assembly patterns* are identified in a two-step procedure (Lopes-dos-Santos et al., 2013; Trouche et al., 2016):

Determine number of significant patterns: a principal component analysis (PCA) is first applied to matrix $\mathbf{Z}$:

$$\sum_{j=1}^{n} \lambda_j \boldsymbol{p}_j \boldsymbol{p}_j{}^{\mathrm{T}} = \frac{1}{n} \mathbf{Z}\mathbf{Z}^{\mathrm{T}}$$

where $\boldsymbol{p}_j$ is the j^{th} principal component with corresponding eigenvalue λ_j (for $j = 1, \ldots, n$). Note that $\frac{1}{n}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}$ is the correlation matrix of $\mathbf{Z}$. To estimate the number of significant patterns embedded within the data, the Marčenko-Pastur law is used (Marčenko and Pastur, 1967; Götze et al., 2004). This law states that for a $n \times B$ matrix whose elements are independent and identically distributed random variables with zero mean and unit variance, all eigenvalues are asymptotically (i.e., when $n, B \to \infty$ such that B/n converges to a finite positive value)

[6] Note that for the assembly pattern detection in this thesis, only putative principal neurons were included.

bounded to the interval $\left[\left(1-\sqrt{n/B}\right)^2, \left(1+\sqrt{n/B}\right)^2\right]$. This suggests that if the firing activity of the neurons is independent from each other, then none of the eigenvalues is expected to exceed $\lambda_{\max} = \left(1+\sqrt{n/B}\right)^2$. Using simulated spike trains, this was indeed shown to be the case even for values of n and B considerably smaller than in my recordings (Peyrache et al., 2010; Lopes-dos-Santos et al., 2011, 2013). An eigenvalue above $\lambda_{\max}$ thus indicates that the pattern given by the corresponding principal component captures more correlation than any pattern would be expected to capture if the firing activity of all neurons was independent of each other. The number of eigenvalues above $\lambda_{\max}$ (denoted by N_A) therefore provides a lower bound for the number of distinct significant patterns.

Identify assembly pattern composition: the first N_A principal components each capture a significant amount of correlation between the firing activities of the neurons. However, principal components are restricted to be orthogonal to each other, while cell assemblies do not need to be (e.g., they can contain overlapping neurons). Principal components are also extracted from the data sequentially, which usually causes the first principal component to seemingly be a mixture of multiple assemblies (Laubach et al., 1999; Lopes-dos-Santos et al., 2011). Moreover, a PCA is solely based on pairwise correlations, but higher-order correlations could also inform assembly identification. To avoid these caveats, an independent component analysis (ICA) is used to identify the patterns. An ICA extracts patterns such that the linear projections of the data onto these patterns are as independent from each other as possible. However, an ICA directly applied to matrix $\mathbf{Z}$ without prior dimension reduction would extract as many patterns as there are neurons, which leads to spurious results (Lopes-dos-Santos et al., 2013). To restrict the number of patterns identified by ICA to N_A, the data is first projected onto the subspace spanned by the first N_A principal components:

$$\mathbf{Z}_{\mathrm{PROJ}} = \mathbf{P}_{\mathrm{SIGN}}{}^{\mathrm{T}}\mathbf{Z}$$

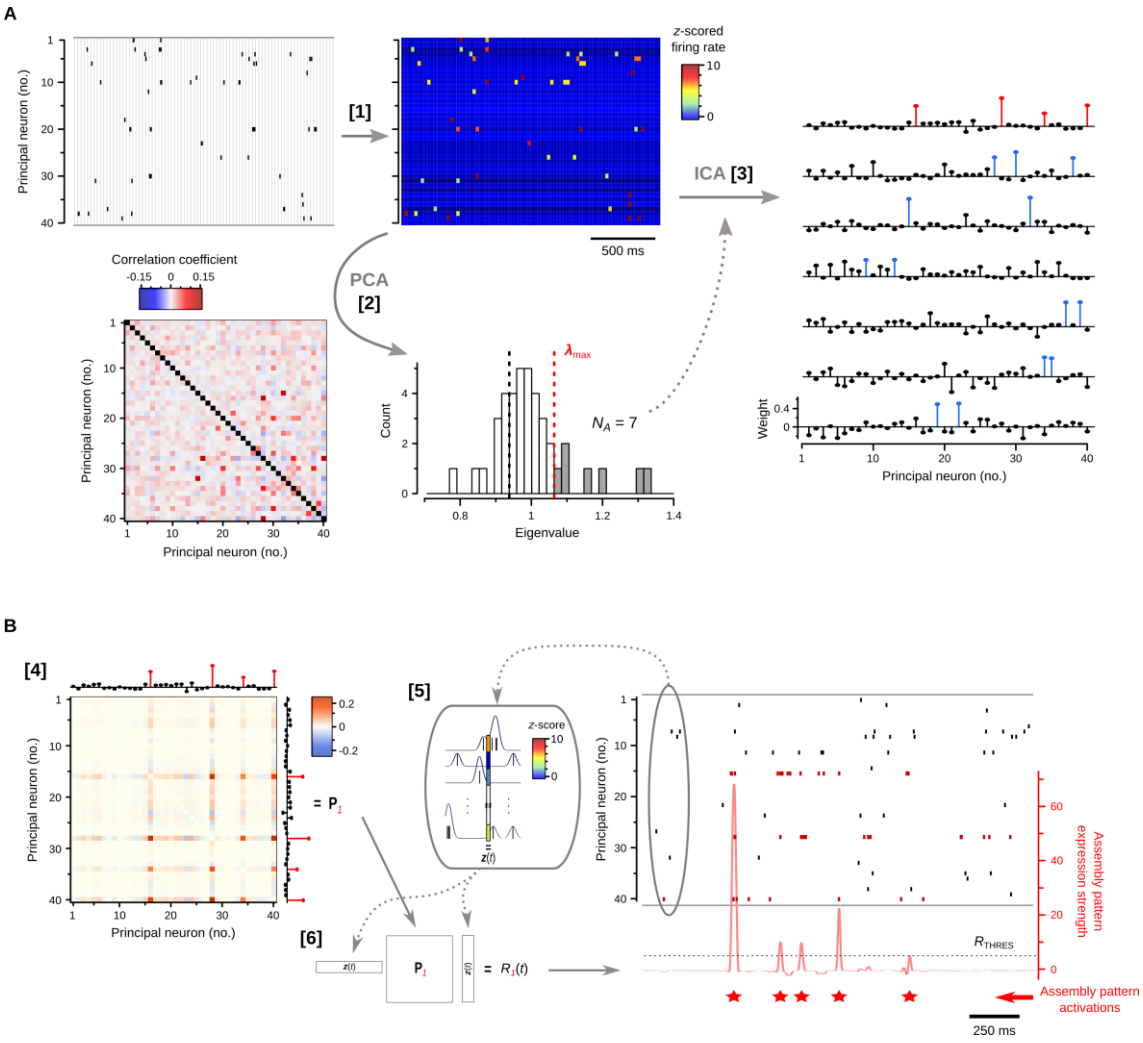

Figure 3.1 *Identification and tracking of assembly patterns.*

(A) **Assembly pattern identification**. The different steps to identify neuronal co-activation patterns are illustrated for 40 simultaneously recorded principal neurons. After binning (25 ms time bins) and normalizing (*z*-score) each neuron's spike counts **[1]**, principal component analysis (PCA) is applied to the correlation matrix of the resulting binned (and *z*-scored) spike count matrix to find the number of patterns describing statistically significant co-activations between neurons **[2]**, as estimated by the number of eigenvalues N_A exceeding the Marčenko-Pastur threshold $\lambda_{\max}$. Independent component analysis (ICA) is then used to identify N_A assembly patterns, given by weight vectors indicating the contribution of each neuron to that pattern **[3]**. For each pattern, neurons with a high contribution (i.e., weight exceeding two standard deviations above the mean) are coloured for display purpose (red for the first assembly pattern, blue for all the other patterns). For clarity, only a few second sample of the spike trains raster plot and the *z*-scored spike count matrix are shown; the assembly patterns are however identified using time bins spanning the entire session.
(B) **Tracking of assembly pattern expression strength.** For each assembly pattern, a projector matrix is constructed by taking the outer product of its weight vector and setting the diagonal to zero (to ensure only co-activations of neurons can contribute to the expression of an assembly pattern) **[4]**. The spike trains are convolved with a Gaussian kernel **[5]**. The assembly pattern expression strength is then taken as the quadratic form of the projector matrix with the convolved and *z*-scored spike trains **[6]**. The scale of the expression strength can thus be interpreted as projected *z*-scores onto the weight vector of the assembly pattern. Assembly pattern activations are defined as peaks in the expression strength exceeding $R_{THRES} = 5$.

where $\mathbf{P}_{\mathrm{SIGN}}$ is the $n \times N_A$ matrix with the first N_A principal components as columns. An ICA is then applied to the matrix $\mathbf{Z}_{\mathrm{PROJ}}$. That is, a $N_A \times N_A$ un-mixing matrix $\mathbf{W}$ is found such that the rows of the matrix $\mathbf{Y} = \mathbf{W}^{\mathrm{T}}\mathbf{Z}_{\mathrm{PROJ}}$ are as independent as possible. For this thesis, this optimization problem was solved using the fastICA algorithm (Hyvärinen, 1999) implemented in R (fastICA-package: Marchini et al., 2013). The resulting un-mixing matrix $\mathbf{W}$ is then expressed in the original basis spanned by all the neurons:

$$\mathbf{V} = \mathbf{P}_{\mathrm{SIGN}}\mathbf{W}$$

where the columns of $\mathbf{V}$ (i.e., $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{N_A}$) are the weight vectors of the *assembly patterns*. As both the sign and the scale of the ICA output is arbitrary, all weight vectors are scaled to unit length (i.e., $\sum_{i=1}^{n} {\boldsymbol{v}_k}^2(i) = 1$ for $k = 1, \ldots, N_A$) and their sign set such that the highest absolute weight of each assembly pattern is positive. Note that the detection of assembly patterns is solely based on neuronal co-activations within 25 ms time bins; the particular sequence of activations within these co-activation events is not considered here. Also note that for a co-activation pattern to be detected as significant, it need to be active in many different time bins.

An assembly pattern is thus described by a weight vector that contains for each recorded principal neuron its contribution to that pattern (e.g., see **Figures 3.1A** and **3.2A**). For each assembly pattern it would be possible to define a corresponding "cell assembly" as those neurons whose weight exceeds the mean weight by two standard deviation (*high contributing neurons*; with both mean and standard deviation calculated from only the weights of that pattern). Note that for the purpose of this thesis, this definition of cell assemblies is only used for an intuitive way of interpreting and discussing the presented results. Importantly, all analyses are performed directly on the assembly patterns themselves (i.e., using the weight vectors formed by the contribution of all simultaneously recorded principal neurons) and not merely on those neurons with a high contribution to a given pattern. I also refer the reader to the last part of section 3.5 for a discussion of the appropriateness to define a cell assembly as a

discrete group of neurons. Finally, note that the sparsity of an assembly pattern (see **Table 4.1**), which reflects to what extent the weight vector of a pattern is dominated by a small group of neurons, was calculated as:

$$1 - \frac{\sqrt{n} - \sum|v_i|}{\sqrt{n} - 1}$$

where n is the length of the weight vector (i.e., the number of principal neurons recorded that day) and v_i is the weight vector's i^{th} element (i.e., the contribution of neuron i to the pattern).

Tracking expression of assembly patterns over time

For example to determine their behavioural correlates (see section 3.3), or to evaluate their reactivation and reinstatement strength (see section 4.2), the expression of each pattern can be tracked over time as follows (see also **Figure 3.1B**):

$$R_k(t) = \boldsymbol{z}(t)^{\mathrm{T}} \mathbf{P}_k \boldsymbol{z}(t)$$

where $R_k(t)$ is the *expression strength* of assembly pattern k at time t and $\boldsymbol{z}(t)$ is a smooth vector function containing for each neuron its *z*-scored instantaneous firing rate. $\mathbf{P}_k$ is the projection matrix of assembly pattern k and is constructed from the outer product of its weight vector $\boldsymbol{v}_k$. The diagonal entries of this projection matrix are set to zero to prevent high expression strength caused by the isolated activity of a single neuron with high weight to that pattern (Peyrache et al., 2009, 2010, Lopes-dos-Santos et al., 2011, 2013). With this approach, only co-firing of neurons can contribute towards the expression of an assembly pattern. To increase the temporal resolution beyond the bin-size used to identify the assembly patterns, $\boldsymbol{z}(t)$ is obtained by convolving the spike train of each neuron with a kernel-function (Dupret et al., 2013; Lopes-dos-Santos et al., 2013) after which the resulting smooth curve is normalized by a *z*-score transform. As smoothing function, a Gaussian kernel is chosen with standard deviation $w/\sqrt{12}$, so that the kernel has the same standard deviation as fixed time bins of w ms (Kruskal et al., 2007). The value w determines the width of the "integration window" in which spikes are

still considered to be coincident. I set *w* to 25 ms to match the bin-size used to identify the patterns. The resulting expression strength time courses show a low baseline with sparse, transient peaks (e.g., see **Figures 3.1B** and **3.2A**). These peaks correspond to co-activations (within 25 ms) of at least two neurons with high contribution to that pattern, whereby the magnitude of these peaks is governed by both (1) the weights of the active neurons in that pattern and (2) the number of spikes discharged by each neuron within the window. Note that the scale of the expression strength can be interpreted as a "projected *z*-score". *Assembly pattern activations* are defined as peaks in the expression strength exceeding $R_{\mathrm{THRES}} = 5$. During the exposure session in which the patterns used in this thesis were detected (see next section), this threshold resulted in a mean assembly pattern activation rate of 0.95 ± 0.02 Hz (*n* = 521 assembly patterns). Note that each detected assembly pattern is thus activated many times.

3.3) Validation on hippocampal recordings

To assess the in the previous section described statistical framework for identification and tracking of assembly patterns (hereafter referred to as "PCA/ICA-method") on real data, I applied this method to the spike trains of hippocampal principal neurons simultaneously recorded from the dorsal CA1 region of mice while they spontaneously explored an open-field enclosure (44.2 ± 1.6 principal neurons per session, range: 22–90; see section 2.1). From 93 such "exposure" sessions (26 min each) recorded from 8 mice, a total of 521 assembly patterns (5.6 ± 0.2 per session) were identified. Each pattern is described by a weight vector containing the contribution of each recorded neuron to that pattern (**Figure 3.2A**). As a first sanity check, I confirmed that pairs of neurons with high contributions to the same assembly pattern had far stronger instantaneous rate correlations than other neuron pairs (**Figure 3.2B**).

I next assessed whether the detected assembly patterns carried behaviourally relevant information. When I tracked the expression of each pattern over time, their activations displayed a strong spatial tuning (**Figure 3.2A**). In line with this, the discharge of neurons with high

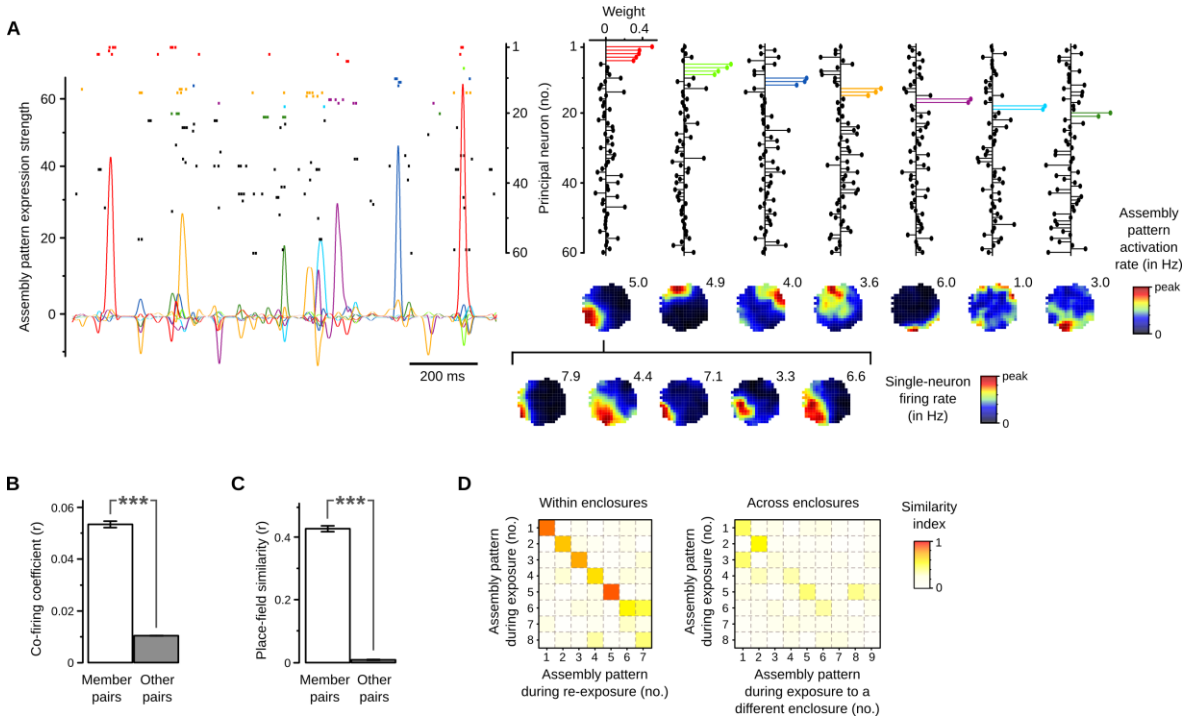

Figure 3.2 *Short-timescale hippocampal co-firing patterns are spatially tuned and environment specific.*

(A) Assembly patterns identified from repeated coincident neuronal discharges in 25 ms time bins spanning the exposure session. For visualization purposes, the 60 simultaneously recorded principal neurons are ordered and colour coded to highlight neurons with high contribution to the same pattern. Shown are a ~1.5 sec example raster plot of the spike trains (top-left; one neuron per row) along with the expression strength time course of each detected pattern (bottom-left), their weight vectors (top-right) and corresponding assembly spatial maps (bottom-right; numbers indicate peak assembly pattern activation rate). At the bottom, single-neuron firing rate maps (numbers indicate peak firing rate) are shown for the five neurons with high weight (highlighted in red) in the first pattern.
(B-C) Detected assembly patterns group together neurons with correlated firing activity and overlapping spatial tuning. Both the average co-firing coefficient (B) and place-field similarity (C) are much higher for pairs of neurons with high weight to the same pattern (n = 919 member pairs) than for other neuron pairs (n = 59,823 other pairs). Error bars represent ±1 SEM; *** P < 0.001.
(D) Detected assembly patterns are environment specific. Example from one recording day showing that the set of patterns expressed during an exposure session are more similar to those identified during re-exposure to that enclosure (left) than to those identified in another enclosure of a different recording block (right).

contributions to the same pattern substantially overlapped in space (**Figure 3.2A,C**). Thus, although the assembly pattern detection was blind to the animal's location and solely based on short-timescale co-activations, it successfully grouped together neurons encoding the same location. However, the detection appeared not to be only governed by the spatial overlap of neurons' discharge as different patterns simultaneously detected in the same enclosure could overlap in space (e.g., the light-green and orange patterns in **Figure 3.2A**). This suggests that a given location could be encoded by several assembly patterns (Jackson and Redish, 2007).

I finally tested whether the assembly patterns expressed during exploration are specific for a particular environment. For this, the similarity of two assembly patterns was quantified by a *similarity index* equal to the absolute value of the inner product of their weight vectors (Almeida-Filho et al., 2014). I found that sets of assembly patterns detected during the first exposure to an enclosure tended to be more similar to the set of patterns detected during re-exposure to that same enclosure than to a set of patterns detected in another enclosure (**Figure 3.2D**; within-enclosure *vs.* across-enclosure maximum assembly pattern similarity: 0.59 ± 0.01 *vs.* 0.40 ± 0.01, $n = 237$ patterns, $P < 0.0001$). This indicates that assembly patterns expressed during exposure to an enclosure are reinstated when the animal is re-exposed to the same environment.

To summarize, assembly patterns identified solely based on short-timescale (25 ms) neuronal interactions are shown to (1) be spatially selective, (2) be environment specific, (3) bind together co-active neurons and (4) bind together neurons with overlapping spatial tuning. These results are consistent with the interpretation of these assembly patterns as internal representations of space, and they constitute an important validation of the PCA/ICA-method on real hippocampal recordings.

3.4) Is the PCA/ICA-method biased by neurons' firing rates?

In one of the first steps of the PCA/ICA-method, the firing rate of each neuron is normalized by *z*-scoring its binned spike counts. The rationale behind this is to prevent a bias towards high firing neurons. Nevertheless, I found that neurons with relatively high firing rate (>1 Hz) were on average assigned larger weights in assembly patterns than relatively low firing neurons (<1 Hz; **Figure 3.3A**). Along the same line, neurons with a large contribution to at least one assembly pattern tended to have higher firing rates than neurons without any large contributions (**Figure 3.3B**).

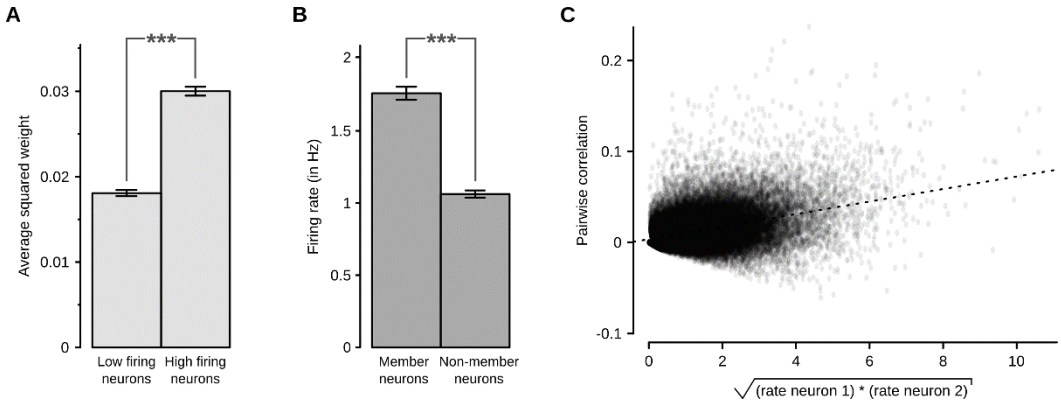

Figure 3.3 *High firing neurons contribute more to assembly patterns than low firing neurons.*

(A) Average contribution to assembly patterns (measured for each neuron as the average of its squared weights over all in that session detected patterns) of low firing neurons (<1 Hz; n = 2,544 neurons) and high firing neurons (>1 Hz; n = 1,767). Error bars represent ±1 SEM; *** P < 0.001.
(B) Average firing rate of neurons with high contribution to at least one assembly pattern (n = 1,340 member neurons) versus the average firing rate of neurons without any high contribution (n = 2,971 non-member neurons). Error bars represent ±1 SEM; *** P < 0.001.
(C) For each principal neuron pair, its pairwise spike count correlation is plotted against the geometric mean of the firing rate (in Hz) of both neurons. Dashed line is the ordinary least-squares regression line (slope = 0.007, P < 0.0001, R^2 = 0.18; correlation coefficient = 0.43), which indicates that pairs of neurons with high rates tend to have higher firing correlations than low firing neuron pairs. Note that one data point ([1.85, 0.38]) is outside this graph's axes.

These observations constitute a potential criticism of the PCA/ICA-method, as they indicate that the method might be biased by the firing rate of neurons. This issue is particular relevant in the light of an application of this method in a recent study of our lab (Trouche et al., 2016). In this study, we used TetTag mice (see last paragraph of section 1.4) to express ArchT in a subset of hippocampal neurons (the "tagged" ensemble) that were previously activated during exploration of an environment. Those tagged neurons were repeatedly silenced over the following 6 days while the mice were re-exposed to the same environment. The repeated silencing of the tagged neurons caused their firing rate to decrease also during periods when no stimulation was applied. Importantly, a different population of principal neurons (the "alternative" ensemble) increased their firing rate over this period. Using the PCA/ICA-method, we additionally showed that over these 6 days the tagged neurons also became less likely to contribute to assembly patterns while the alternative neurons became more likely to do so. If, however, the PCA/ICA-method would be biased by the firing rate of neurons, the additional

result obtained with this method would lose its value as the result could simply be explained by the earlier in the paper demonstrated rate changes.

However, because of the *z*-scoring, there is no mathematical necessity for high firing neurons to have higher contributions to assembly patterns than low firing neurons. This can be most simply explained by the hypothetical example of a neuron that doubles its amount of spikes in each 25 ms bin. This neuron would have a twice as high firing rate, but its *z*-scored spike-count vector—which is the measure based on which the assembly patterns will be detected—would remain unchanged. This means that the observed difference in contribution to assembly patterns between high and low firing neurons cannot simply be explained by their difference in firing rate. Therefore, I argue that the results presented in **Figure 3.3A,B**—instead of as a potential criticism—should be interpreted as the finding that in the mouse hippocampus, principal neurons with high firing rates tend to contribute more to the formation of assembly patterns than low firing neurons. This indicates that compared to a low firing neuron, a high firing neuron has on average a greater percentage of its spikes coinciding with spikes of its peer neurons. This finding is likely related to the also observed higher average pairwise correlation between high firing neuron pairs compared to pairs of low firing neurons (**Figure 3.3C**; see also de la Rocha et al., 2007).

"Population axis"

An insight in why it could be the case that higher firing neurons tend to have both stronger pairwise correlations and larger contributions to assembly patterns, was provided by the following—at first surprising—observation. As part of my attempt to better understand the properties and behaviour of the PCA/ICA-method, I used this method to (try to) detect assembly patterns after various ways of shuffling the recorded spike trains. In one shuffling procedure, the spike count vector of each neuron was independently shuffled (i.e., shuffling was performed after binning the spike counts). This shuffling procedure preserves the average firing rate of each neuron, but destroys any relations between the neurons. After shuffling in this way, most

of the time no principal components were found to be above the Marčenko-Pastur threshold (**Figure 3.4A**; bottom), which is consistent with the results based on simulations of independent neurons (see section 3.2; Peyrache et al., 2010; Lopes-dos-Santos et al., 2011, 2013). In another shuffling procedure, all spikes discharged by principal neurons were randomly re-assigned to the pool of recorded principal neurons, under the restriction that for each neuron the total number of spikes assigned to it remained the same. This shuffling procedure also preserves the average firing rate of each neuron and it additionally preserves—for any given time window—the "population rate" (i.e., the summed activity of all recorded principal neurons), but otherwise it destroys any firing relations between the neurons. This shuffling procedure is the same as the "SWAP"-method used in Jackson et al. (2006). After shuffling in this way, I consistently found one principal component to be far above the Marčenko-Pastur threshold (**Figure 3.4A**; top), indicating that there was still a pattern in the shuffled data that explained more firing rate correlation than would be expected from any pattern if the firing of all neurons would be independent from each other. Performing this shuffling with preservation of the population rate many times, revealed that each neuron's weight in this first principal component converged to a specific value (**Figure 3.4B**; top) that was directly related to its firing rate (**Figure 3.4C**; left). Such a convergence was not observed for the second principal component after shuffling with preservation of the population rate (**Figure 3.4B,C**; middle), or for the first principal component after shuffling without preserving the population rate (**Figure 3.4B,C**; bottom, right).

This indicates that given fluctuations in the population rate, simulated neurons with otherwise independent spiking activity have an expected pairwise firing rate correlation that is directly related to their firing rates. In other words, the existence of periods with relatively high and relatively low population firing, which has been demonstrated to be the case in the rodent hippocampus (Luczak et al., 2015), has the result that the "default" mode (i.e., spikes distributed across neurons under assumption of maximum entropy given the fluctuations in population rate and the neurons' firing rate) is that higher firing neurons tend to have higher pairwise correlations.

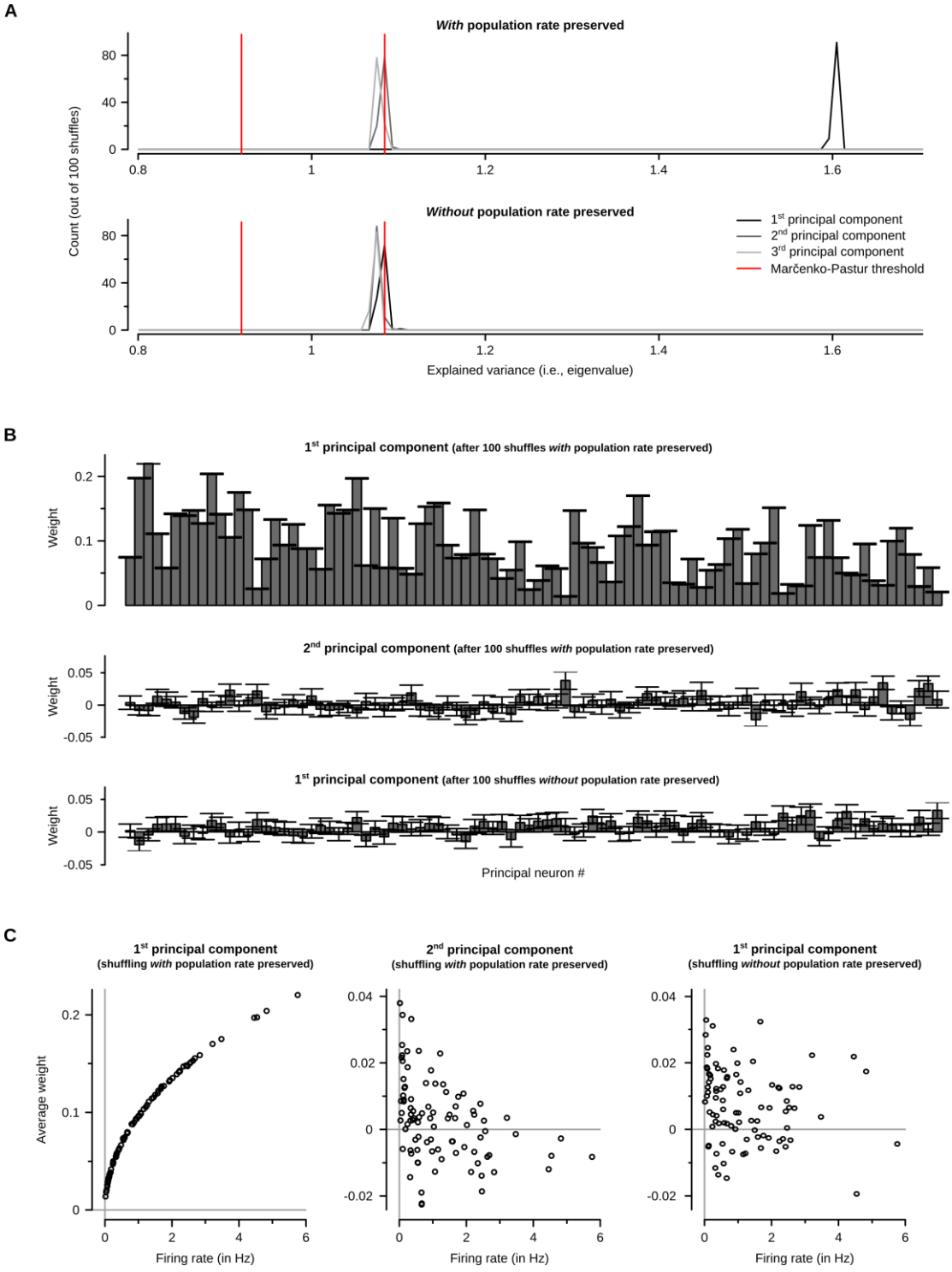

Figure 3.4 *"Population axis"*

After shuffling with preservation of the population rate, there are still significant correlations between the neurons. These significant correlations are restricted to one dimension—the "population axis"—and can be explained by an interaction between variations in the population rate and the neurons' firing rates.
(A) Example of one recording session illustrating that after shuffling by randomly re-assigning spikes to the pool of principal neurons under the restriction that each neuron keeps the same total number of

spikes (i.e., shuffling with preserving the population rate; top), consistently one principal component is far above the Marčenko-Pastur threshold. After shuffling of each neuron's spike count vector (i.e., shuffling without preserving the population rate; bottom), this is not the case. For these graphs, each shuffling procedure was repeated 100 times.
(B) After shuffling with preservation of the population rate, the first principal component converges to one axis (top). The second principal component on the other hand does not (middle), and neither does the first principal component after shuffling without preserving the population rate (bottom). Each shuffling procedure was repeated 100 times, data are represented as mean ±1 SEM.
(C) A neuron's weight in the axis to which the first principal component converges after shuffling with preservation of the population rate, is directly related to its firing rate. In these graphs, the average weight of each neuron after 100 shuffles—either in the first (left) or second (middle) principal component after shuffling with preserving the population rate or in the first principal component after shuffling without preserving the population rate (right)—is plotted against the neuron's mean firing rate (in Hz). Note the difference in scale of the *y*-axis between the first and last two graphs.

One should, however, be careful with—or indeed refrain from—interpreting this finding as artificially larger pairwise correlations (and consequently artificially higher assembly pattern contributions) for high firing neurons compared to low firing neurons. Firstly, a fixed population rate in combination with fixed firing rates would lead to larger pairwise correlations for higher firing neurons if the spikes would be randomly distributed over the neurons (i.e., with maximum entropy) given those two constraints. As empirical support that this is not the case, in my data I find that the variance explained by the "population axis" in the shuffled data is consistently greater than the variance explained by this axis in the actual recordings. Second, the fluctuations in the population rate should not be treated as a "given"; it is a property of how the brain is organized. In particular, it would not make sense to treat the population rate as given, but not the identity of the spikes that make up this population rate. Third, and perhaps most importantly, in terms of consequences (e.g., induced synaptic plasticity and effect on downstream neurons), an increased pairwise correlation between neurons is the same regardless of whether it can be (partly) explained by the combination of firing rates and fluctuations in the population rate or not.

3.5) Discussion

This chapter has described and examined a recently proposed statistical assembly detection method for the identification of assembly patterns from multi-unit recordings. It was shown how application of this PCA/ICA-method to recordings of hippocampal principal neurons during exploratory behaviour results in the detection of spatially tuned and environment specific assembly patterns. Especially since the PCA/ICA-method is blind to the location of the animal—the detection of patterns is solely based on short-timescale (25 ms) interactions, this result constitutes an important validation of this assembly detection method on *in vivo* recordings. Moreover, the spatial selectively of the detected assembly patterns indicates that their expression might contribute to the animal's internal representation of space. Since the main contribution to these patterns comes from place cells (e.g., see **Figure 3.2A**), the same arguments supporting the interpretation of place cell activity as representing space (see section 1.3, paragraph under header "Place cells") can be used here to make a strong case that the expression of these assembly patterns indeed represents space. This suggests that the here detected patterns not only bind together co-active neurons, but that their output represents mental concepts. I will therefore refer to these patterns as *cell assembly patterns*.

Another contribution of this chapter is that the here presented results open up the possibility for the PCA/ICA-method to be applied to recordings from other brain regions and/or species. An interesting first question would be whether behavioural correlates can also be found for assembly patterns detected in other brain regions. Note that depending on the width of the synaptic integration window of neurons in the chosen brain region, it might be appropriate to select a different time window for binning.

Cell assembly patterns *vs*. activity-dependent tagging of neuronal ensembles

A promising new approach for identifying (and manipulating) neuronal ensembles is the TetTag technique (see last paragraph of section 1.4). This technique allows the tagging of neurons activated during an experimenter controlled time window, which has been used to detect

neuronal ensembles whose activity seems to represent spatial contexts and/or associated fear memories (for reviews see Josselyn et al., 2015; Tonegawa et al., 2015). How do the neuronal ensembles identified with this technique relate to the here detected cell assembly patterns?

A major difference between both approaches is their widely different time scale. The PCA/ICA-method identifies cell assembly patterns based on neuronal co-activations within 25 ms time bins, while a TetTag ensemble is detected based on one much longer time window[7]. This explains why a cell assembly pattern typically corresponds to a specific location within an environment, while a TetTag ensemble would represent the entire spatial context (and/or its associated fear memory). Interestingly, and probably related to this low temporal specificity, it has been reported that the TetTag approach does not work equally well in all brain regions. In one of the studies of Susumu Tonegawa's lab, neurons in either the dentate gyrus or in the CA1 area that were activated during exposure to a neutral context A were tagged with channelrhodopsin (Ramirez et al., 2013). These tagged neurons were then optogenetically activated during fear conditioning in a different context B. When neurons had been tagged and stimulated in the dentate gyrus, this procedure resulted in a "false memory" in the sense that it caused mice to freeze upon re-exposure to context A. This is interpreted as evidence that optogenetic stimulation of the tagged dentate neurons activated the neuronal representation of context A, which as a result became associated with the fear memory. But when the same was done to neurons in the CA1 area, this procedure failed to create such a false memory. Simultaneous activation of all tagged CA1 neurons thus did not seem to activate the representation of context A. This suggests that in the CA1 area, representations of spatial context rely more on fine temporal coordination than in the dentate gyrus, where population activity of the involved neurons appears to be sufficient to activate them. The PCA/ICA-method might therefore be better suited than the TetTag approach for studying neuronal representations

[7] Because of the relatively slow dynamics of Doxycycline, this tagging window is at least several hours. Typically, during this period an animal will be allowed to briefly explore one spatial context (sometimes combined with emotional stimuli, such as foot shocks or a potential mate), but he will spend most of this time in his home cage to keep background tagging to a minimum.

in the CA1 area. Note that this discussion also highlights that if the PCA/ICA-method were to be applied to recordings from the dentate gyrus, it might be beneficial to select a (much) larger time window for the binning.

Other methods to identify cell assemblies from multi-unit recordings

The results presented in this chapter make a strong case that the with the PCA/ICA-method detected assembly patterns have properties akin to cell assemblies. However, it is not claimed—and it is not needed to do so—that these assembly patterns constitute the best possible identification of cell assemblies that can be extracted from the given recordings. Here, I will briefly highlight a selection of other approaches to identify cell assemblies from multi-unit recordings, and discuss their advantages and disadvantages.

Firstly, an important distinction that can be made between assembly detection methods is whether or not they take the temporal ordering of spikes into account. The ICA/PCA-method has no consideration for such temporal sequences, and it detects patterns solely based on co-occurrences of spikes within time bins of specified width. It ignores the particular order of spikes within these bins. Methods for detecting assemblies consisting of temporal spike sequences typically can take two forms: (1) identification of large population events (e.g., SWRs) combined with a decision for each such event whether the sequence of place cells in it reflects a path in a previously explored environment or not; or (2) identification of particular sequences that tend to occur more often than expected by chance. The rationale of the first type of method is different from that of the PCA/ICA-method in that it identifies unique events as opposed to recurring patterns. In the last decade, this approach has become popular for detecting reactivation—in particular awake replay. Although this approach has proven useful for analysing temporally structured reactivation (e.g., Foster and Wilson, 2006; Wikenheiser and Redish, 2015), it could not have been used for the purpose of this thesis as the expression of its detected assembly-events cannot be tracked over time. The second type of method is more similar to the ICA/PCA-method in the sense that both approaches detect neuronal patterns that

occur more often than expected by chance, except that one of them also takes the order of spikes into consideration for that decision. An example of such a method was recently proposed by Russo and Durstewitz (2017). The advantage of their method is that it is capable of detecting arbitrary complicated assembly structures, but disadvantage is the resulting steeply increased complexity. To deal with this complexity, their method makes simplifying assumptions in other regards, one of which is highlighted next.

Another distinction that can be made between assembly detection methods is whether the assemblies to be detected are restricted to being discrete (i.e., each neuron is either part of an assembly or not) or not (i.e., each neuron has a weight in an assembly that can vary along a continuum). Restricting the assemblies to be discrete has the advantage that it becomes possible to simply count the number of times a given group of neurons is active together, after which it can be tested whether that count is larger than expected based on the individual firing properties of the involved neurons (Grün et al., 2002a, 2002b). The above mentioned method of Russo and Durstewitz (2017) is an elegant extension of this approach that makes it possible to consider assembly activity occurring at various temporal scales and with arbitrary time lags. However, a disadvantage of restricting cell assemblies to being discrete is that—although this is not always realised—it is a simplifying assumption, and failure to appreciate this could lead to misinterpretations.

Cell assemblies: discrete building blocks of memory?

In a recent news article about a publication in the journal *Science*, it was announced that scientist had been able—for the first time—to reveal the brain's individual building blocks of memories[8]. In the scientific article referred to, the authors claim to have found cell assemblies in the mouse hippocampus that are distinct, orthogonal, independent and mutually exclusive (Malvache et al., 2016).

[8] News article by Emily Benson, posted on the New Scientist website at 15 September 2016 (https://www.newscientist.com/article/2106094-building-blocks-of-memories-seen-in-brains-for-the-first-time/).

To identify these cell assemblies, Malvache and colleagues used *in vivo* two-photon calcium imaging to capture the activation dynamics of a large population of neurons (~1000) with single-cell resolution. Recently developed fluorescence calcium indicators (Akerboom et al., 2012; Chen et al., 2013) allowed them to track the concentration of calcium in neurons' cell bodies, which closely relates to their spiking activity (Theis et al., 2016). They performed this technique on awake, head-fixed mice running on a treadmill in the dark, in the absence of any sensory cues. Despite the relatively low temporal resolution of calcium imaging (precision of 20–100 ms compared to <1 ms for electrophysiology; Harris et al., 2016), in a previous study with this set-up, the same group had elegantly shown that the activity of a subset of the imaged hippocampal neurons formed recurring, stereotyped sequences during running (Villette et al., 2015). In their most recent study (Malvache et al., 2016), they focussed on short-lasting events of high population activity—synchronous calcium events (SCE)—that occurred during periods of awake rest in between such running episodes. They clustered these SCEs according to which neurons participated in them, after which they defined for each significant SCE cluster a corresponding cell assembly. The clustering step was performed by applying the *k*-means algorithm on the columns of the cell participation covariance matrix of all SCEs, with the number of clusters selected based on the silhouette statistic. For every SCE cluster with a significant average silhouette value, a corresponding cell assembly was then defined as those cells that were significantly activated during SCEs assigned to that cluster. Importantly, however, if a neuron was significantly activated during multiple SCE clusters, it was only assigned to the cell assembly of one of those SCE clusters and not to the other cell assemblies. This step makes that the resulting cell assemblies are distinct from each other *by definition*, and it artificially reduces the dependence between the detected cell assemblies[9]. This is important to realize, as it means that the results reported by Malvache and colleagues do not provide

[9] Yet, the results reported by Malvache et al. (2016) seem to indicate that their detected cell assemblies are still dependent. Of all neurons assigned to a cell assembly, 12% was significantly correlated with another cell assembly, which is well above the presumed significance level of 5%. (Note that this does not mean that the remaining 88% is uncorrelated with or independent from other cell assemblies, but merely that their correlation is not significant.)

evidence that cell assemblies in the hippocampus are organized as distinct and independent entities, but merely that it is possible to find such discrete cell assemblies if one specifically looks for them. Note however that it is always possible to divide a continuum into discrete parts by thresholding.

An interesting question raised by the discussion above is whether cell assemblies are—or should be thought of as—"discrete" and/or "distinct". To interpret cell assemblies in this way appears to be the prevailing norm in the current literature. For example, in a recent influential review, György Buzsáki introduced the concept of a cell assembly as follows: "*Hebb hypothesized that a discrete, strongly interconnected group of active neurons, the "cell assembly", represents a distinct cognitive entity*" (Buzsáki, 2010, p.362; underlining not present in original). This quote highlights two important and prevalent assumptions: (1) that a cell assembly consists of a discrete group of neurons (i.e., a neuron is either part of a given cell assembly or not), and (2) that cell assemblies represent distinct cognitive entities (i.e., mental concepts represented by different cell assemblies are separable from each other). In addition, or perhaps as a result of these two assumptions, it is often assumed that the cell assemblies themselves are distinct from each other (i.e., there is a clear separation between any two cell assemblies). These assumptions make strong predictions about the organization of neuronal activity, which I believe Donald Hebb himself did not intend to make. Indeed, in his book *The Organization of Behavior*, I could not find the terms "discrete" or "distinct". Although the examples that Hebb used to explain the general principles of his theory often describe cell assemblies consisting of discrete groups of cells that represent more or less distinct mental concepts, he stressed that these examples were highly simplified and that the actual organization in the brain "*would be infinitely more complex than anything one could show with a diagram*" (Hebb, 1949; p.73).

The important question is of course not how Donald Hebb envisioned cell assemblies, but how they are organized in the brain. This question is much harder to answer, if not currently impossible. The main point I hope to make is that at present there is no convincing evidence

supporting the existence of discrete and/or distinct cell assemblies in the brain. At least not cell assemblies at a level at which they are "useful". To explain the need for addition of this last sentence, let me return to the above mentioned review of György Buzsáki. By taking the "reader-centric" definition of cell assemblies (see section 3.1) to the extreme, he argues that cell assemblies are distinct and absolute in the sense that they are either on (if the reader neuron fires) or off (if the reader neuron does not fire) (Buzsáki, 2010). Although the logic of this reasoning is valid, I would argue that in this extreme version of the reader-centric definition, the cell assembly concept loses its usefulness. The main value of the cell assembly concept is in changing the fundamental computational unit of the brain from individual neurons to the level of groups of neurons. The extreme reader-centric definition, however, defines for each individual neuron a corresponding cell assembly, thereby essentially bringing this fundamental unit back to the level of the individual neuron. Once a cell assembly concept goes beyond the level of individual neurons, I am not aware of any good reason—except when used *and* acknowledged as a simplifying approximation—for restricting those cell assemblies to be discrete and/or distinct.

CHAPTER 4: REACTIVATION STABILIZES RECENTLY FORMED HIPPOCAMPAL CELL ASSEMBLY PATTERNS

4.1) Introduction

As reviewed in the first chapter of this thesis, an increasing number of studies advocate hippocampal reactivation of neuronal memory traces as a network-level mechanism for memory consolidation. However, a causal relation between the (sleep) reactivation of new assembly patterns and their subsequent (awake) reinstatement during memory retrieval has not been demonstrated, even though such neuronal pattern reinstatement is thought to mediate memory retrieval. To fill this gap in the literature, the experiments described in this chapter aimed to directly test for a role of offline reactivation in the stabilization of neuronal traces of waking experiences.

Using the assembly detection method that was validated in the previous chapter, cell assembly patterns representing a never-before visited environment are identified in the mouse hippocampus, after which the expression strength of these patterns is tracked during the following sleep/rest and context re-exposure. This approach allows testing whether the strength with which an individual cell assembly pattern is reactivated predicts how well that pattern will later be reinstated when the animal is re-exposed to the same environment (section 4.2). Using closed-loop optogenetic silencing of principal neurons, it is then tested whether the selective disruption of SWR reactivation during sleep/rest impairs the context-dependent reinstatement of these patterns (section 4.3). Importantly, to test whether a potential role of reactivation would be time limited, which is an important characteristic of a consolidation process, assembly patterns representing a highly familiar enclosure are also detected, tracked and SWR-silenced. Furthermore, as an additional control experiment, optogenetic silencing is applied randomly instead of conditioned on the on-line detection of SWRs. Finally, in section 4.4, a fascinating

relation is demonstrated between the initial dynamics of newly-expressed cell assembly patterns and their dependence on offline reactivation.

4.2) Reactivation predicts reinstatement

As described in section 2.1, multichannel extracellular recordings were performed from the dorsal CA1 region of the hippocampus of freely-behaving mice (n = 8; **Figure 4.1A**) while they explored open-field enclosures alternating with periods of sleep/rest. Every day, principal neurons (44.4 ± 2.5 per day) were followed across multiple recording blocks (2.2 ± 0.1 per day). During each recording block (**Figure 4.1B**), mice were allowed to explore either a novel or familiar enclosure ("exposure") and were re-exposed to that enclosure ("re-exposure") after one hour in their sleep box ("rest after"). At the start of each day, a baseline sleep/rest session was recorded ("rest before").

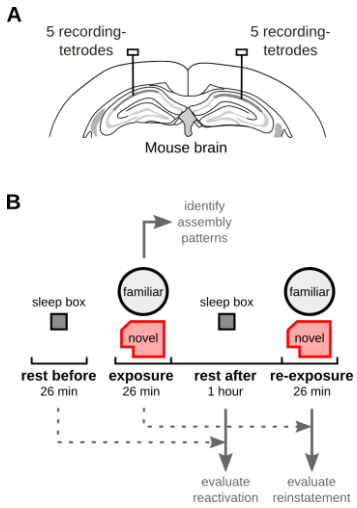

Figure 4.1 *Experimental protocol.*

(A) Ensemble recordings from the dorsal CA1 region of the mouse hippocampus.
(B) Schematic of a "light-OFF" recording block, with repeated exposure to either the familiar or a novel enclosure. Cell assembly patterns are detected during the initial exposure session ("exposure"), after which their expression strength is tracked throughout the full recording block in order to evaluate for each pattern its reactivation strength (i.e., average strength during "rest after" minus during "rest before") and its reinstatement strength (i.e., average strength during "re-exposure" minus during "exposure").

For each recording block, using the PCA/ICA-method of the previous chapter, I identified principal cell assembly patterns during the exposure session, and then tracked their expression strength throughout the full recording block (**Figure 4.1B**). In the exposure sessions of 43 recording blocks without optogenetic silencing, a total of 247 assembly patterns were identified

(108 patterns in 17 blocks with the familiar enclosure, 139 patterns in 26 blocks with a novel enclosure). As described in the previous chapter, these assembly patterns seemed to represent the enclosure in which they were identified as they were spatially tuned and environment specific (see **Figure 3.2**).

I found that 71.7% of these patterns had a higher average expression strength during the rest following the exposure session than during the rest preceding it (against chance: $P < 0.0001$; **Figure 4.2A**), confirming the presence of assembly pattern reactivation. To be able to compare the strength of such offline reactivation across patterns regardless of their differences in baseline strength (e.g., see **Figure 4.2A**), the *reactivation strength* of each assembly pattern was defined as its average expression strength during the rest after minus its average expression strength during the rest before. Note that a null score for reactivation strength thus indicates that a pattern was not reactivated (i.e., same strength during rest after as during rest before), while

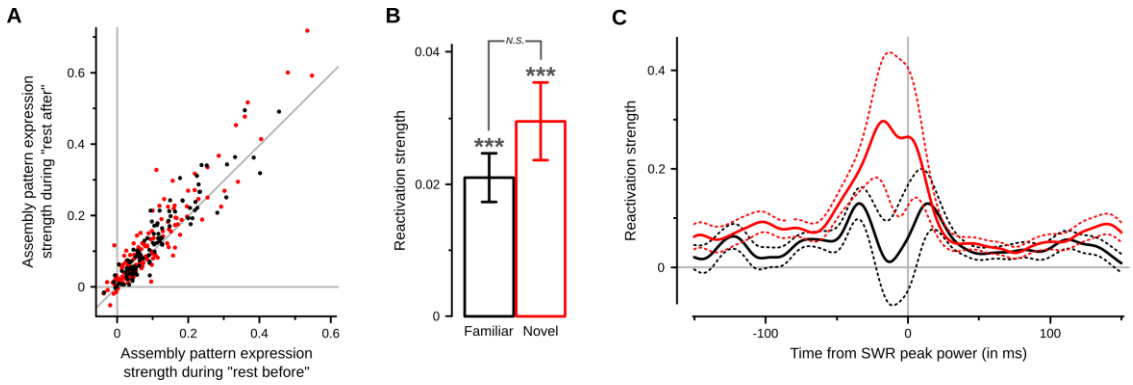

Figure 4.2 *Assembly pattern reactivation.*

(A) The average expression strength of most assembly patterns identified during exposure to either the familiar (black) or a novel (red) enclosure, is stronger during the sleep/rest session after this exposure than during the sleep/rest before. This indicates the presence of assembly pattern reactivation. Note that two data points (Novel: [0.85, 0.89] and [0.53, 1.15], both above the diagonal) are outside this graph's axes.
(B) Both after exposure to the familiar enclosure (n = 108 assembly patterns) and after exposure to a novel enclosure (n = 139), the average assembly pattern reactivation strength (expression strength during "rest after" minus expression strength during "rest before") is well above zero. Based on the full one-hour duration of rest after, there is no clear difference in reactivation strength after the familiar versus after a novel enclosure. Data are represented as mean ± SEM; *** $P < 0.001$, one-sample t-test versus zero; $^{N.S.}$ $P > 0.05$, unpaired two-sample t-test.
(C) Assembly pattern reactivation is emphasised during SWRs, particularly after exposure to a novel enclosure. Solid lines represent mean difference in expression strength between rest after and rest before, dashed lines ±1 SEM (black: familiar, red: novel). Both the mean- and SEM-traces are smoothed with a Gaussian kernel with SD = 4 ms.

the more positive this value the stronger the reactivation of that pattern was. When calculated over the full one-hour duration of the rest after, there was—perhaps somewhat surprising—no clear difference in assembly pattern reactivation strength after a familiar versus after a novel enclosure (but see **Figure 4.8E**), with in both cases the average reactivation strength being well above zero (**Figure 4.2B**). Further, in line with studies based on pairwise correlations (Wilson and McNaughton, 1994; O'Neill et al., 2008), I found that the reactivation of these assembly patterns was most pronounced time-locked to SWR events (**Figure 4.2C**), especially after exposure to a novel enclosure.

The *reinstatement strength* of an assembly pattern was similarly defined as its average expression strength during re-exposure minus its average expression strength during the first exposure. Note that a null score for reinstatement strength thus indicates that a pattern was "perfectly" reinstated (i.e., no loss in strength from initial exposure to re-exposure), while the more negative this value the worse the reinstatement of that pattern was (see the legend of **Figure 4.7B** for further discussion of this measure). Using these measures, I found that after exposure to a novel enclosure, the strength with which a cell assembly pattern was reactivated predicted how well that pattern would later be reinstated during re-exposure to the same

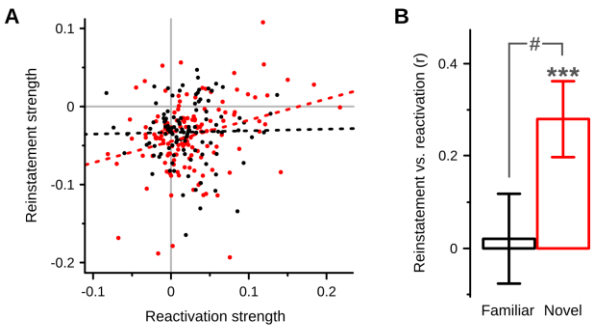

Figure 4.3 *Assembly pattern reactivation following a novel, but not the familiar, enclosure correlates with upcoming awake reinstatement.*

(A) Scatter-plot of the reinstatement strength (change in expression strength from exposure to re-exposure) versus the reactivation strength (change in expression strength from rest before to rest after) of assembly patterns detected in the familiar (black) or a novel (red) enclosure. Dashed lines are corresponding ordinary least squares regression lines (Familiar: slope = 0.02, P = 0.81, R^2 = 0.00; Novel: slope = 0.27, P < 0.001, R^2 = 0.08). (B) Correlation between reactivation and reinstatement strength is stronger after a novel enclosure than after the familiar one. Error bars represent ±1 SE of the correlation coefficient; Novel versus zero: *** P < 0.001; Familiar versus Novel: # P < 0.05.

[Note that one assembly pattern detected in a novel enclosure had both an unusual high reactivation and reinstatement strength. This pattern would be very influential for the here presented reinstatement-vs-reactivation correlation. I therefore left the data point corresponding to this pattern ([0.62, 0.14]) out from these graphs; its inclusion would further increase the correlation for novel blocks to $r = 0.41$*.]*

enclosure (**Figure 4.3**). As an important control, after exposure to a familiar enclosure, I found no such relation (**Figure 4.3**).

4.3) Selective disruption of reactivation impairs subsequent reinstatement

Having established a correlation between the reactivation and later reinstatement of assembly patterns representing a novel environment, the next aim was to causally test this relation. Motivated by the observation that the reactivation of assembly patterns is strongest during SWRs (**Figure 4.2C**; Wilson and McNaughton, 1994; O'Neill et al., 2008), it was reasoned that one way to selectively disrupt reactivation could be to temporarily silence hippocampal principal neurons during these SWR events.

Closed-loop optogenetic SWR silencing

SWRs were detected in real time based on the ripple power of the recorded signal from a tetrode in the pyramidal cell layer (**Figure 4.4B**). This way, 80.1 ± 1.0 % of the later offline identified SWRs were detected on-line with an average latency of 7.68 ± 0.30 ms before their peak power, while 52.8 ± 2.3 % of on-line detections did not correspond to an offline SWR identification (all based on n = 42 one-hour sleep/rest sessions). However, many of these "misdetections" still corresponded to ripple events, although they were less pronounced and below the (rather strict) threshold for offline identification. The detection of these smaller ripple events might actually be beneficial for the purpose of selectively disrupting as much reactivation as possible, since on-line detections without corresponding offline identified SWR indeed still signaled periods of increased assembly pattern reactivation (see ***Appendix A***). In order to be able to temporarily silence hippocampal principal neurons, the dorsal CA1 region of the hippocampus of CaMKII-Cre mice (n = 7) was injected with a flex-ArchT-GFP viral construct to express the light-driven proton-pump ArchT in those principal neurons (**Figure 4.4C**), after which these mice were implanted with optic fibres and tetrodes (**Figure 4.4A**). When light was then delivered upon

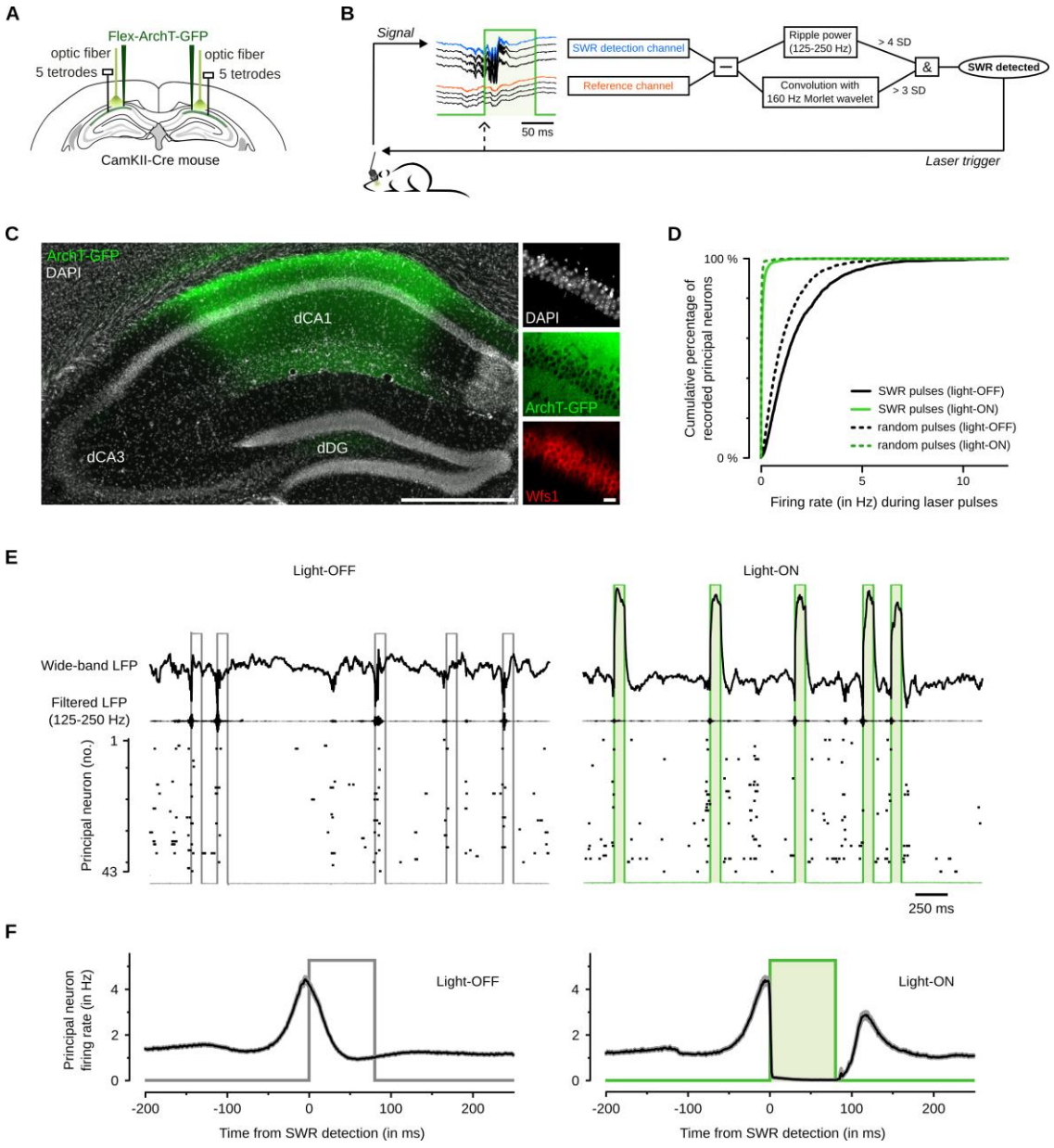

Figure 4.4 *Closed-loop optogenetic silencing of hippocampal principal neurons during on-line detected SWR events.*

(A) Ensemble recordings and optogenetic manipulation of CaMKII::ArchT mice.
(B) Schematic of the on-line SWR detection method (see section 2.2 for a full description).
(C) Representative image showing ArchT-GFP expression restricted to the CA1 region of the dorsal hippocampus (dCA1) from a CaMKII-Cre mouse (left; scale bar: 500 µm; dCA3 = dorsal CA3; dDG = dorsal Dentate Gyrus). The high-magnification images (right; scale bar: 20 µm) show the dCA1 pyramidal cell layer marked by DAPI labelling of cell nuclei (top) and with expression of both the GFP reporter (middle) and Wolfram syndrome 1 (Wfs1; bottom), a specific marker for CA1 principal neurons. Cell counting based on 8 such sections from 2 mice found 90.8 ± 2.5 % of Wfs1-positive cells to be transfected with ArchT-GFP.
(D) Cumulative frequency histogram of principal neuron firing rate calculated during SWR-triggered pulses without the laser powered (n = 1988 neurons), during SWR-triggered pulses with the laser powered (n = 1639), during randomly triggered pulses without the laser powered (n = 1988) and during randomly triggered pulses with the laser powered (random silencing; n = 459). Note that 94.7 % of principal neurons fire less than 0.25 Hz during laser-powered pulses, thus showing high-level of efficacy of ArchT silencing.

(E) Raw data example from a sleep/rest session without (left) and with (right) optogenetic SWR silencing. *Top trace*, wideband local field potential (LFP). *Bottom trace*, 125–250 Hz band pass filtered signal highlighting ripple frequency events. *Raster plot*, spike times from simultaneously recorded dCA1 principal neurons (one neuron per row).
(F) Firing rate response (mean ± SEM) of all recorded hippocampal principal neurons to 80 ms light pulses triggered by on-line SWR detections without (Light-OFF: n = 1988 neurons) and with (Light-ON: n = 1527) the laser powered.
[Note that panel C of this figure was created by Stéphanie Trouche.]

SWR detection, almost all recorded principal neurons were silenced within a few millisecond after light onset, with a quick return to baseline levels of firing following light offset (**Figure 4.4D-F**).

SWR silencing impairs stability of "novel" but not "familiar" assembly patterns

In 37 recording blocks with light pulses delivered conditioned on on-line SWR detections during the one-hour sleep/rest period, an additional 214 assembly patterns were identified from their exposure sessions (**Figure 4.5A**; 78 patterns in 13 blocks with the familiar enclosure, 136 patterns in 24 blocks with a novel enclosure). I found that optogenetic SWR silencing applied during rest following a novel enclosure impaired assembly pattern reinstatement during

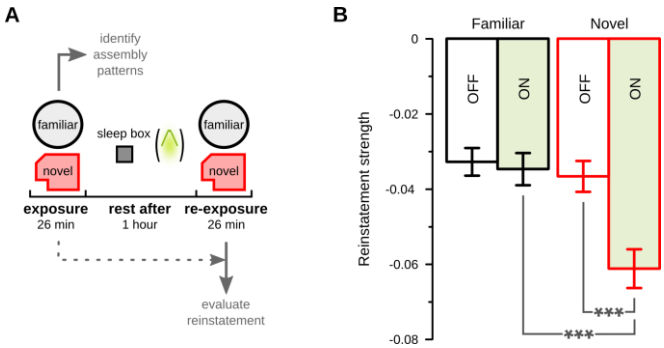

Figure 4.5 *Optogenetic SWR silencing impairs reinstatement of assembly patterns associated with a novel, but not the familiar, enclosure.*

(A) Schematic of a "light-ON" recording block, with repeated exposure to either the familiar or a novel enclosure. Cell assembly patterns are detected during the initial exposure session ("exposure"), after which their expression strength is tracked throughout the exposure and re-exposure session. A pattern's reinstatement strength is defined as its average expression strength during "re-exposure" minus its average expression strength during "exposure".
(B) After exposure to a novel enclosure, SWR silencing impairs the reinstatement of assembly patterns during context re-exposure (Light-OFF: n = 139 patterns; Light-ON: n = 136). This is not the case following exposure to the familiar enclosure (Light-OFF: n = 108 patterns; Light-ON: n = 78). Data are represented as mean ± SEM; *** $P < 0.001$.

subsequent context re-exposure (**Figure 4.5B**). Importantly, the same optogenetic SWR silencing during rest following the familiar enclosure did not alter the reinstatement of its assembly patterns (**Figure 4.5B**; with interaction SWR silencing x enclosure-type: $F(1,318) = 5.05$, $P < 0.05$). These results were replicated after discarding all assembly patterns with more than one high contributing neuron recorded from the same tetrode (**Figure 4.6A**), and further confirmed using conventional neuron-pair and single-neuron analyses (**Figure 4.6B,C**). Finally, an additional control experiment was performed in which light pulses were delivered with randomly selected intervals instead of conditioned on SWR detection (**Figure 4.7A**; an additional 60 assembly patterns were identified in 13 of these recording blocks with a novel enclosure). I found that after such random optogenetic silencing, cell assembly patterns expressed in a novel enclosure were reinstated stronger than after SWR silencing (**Figure 4.7B**).

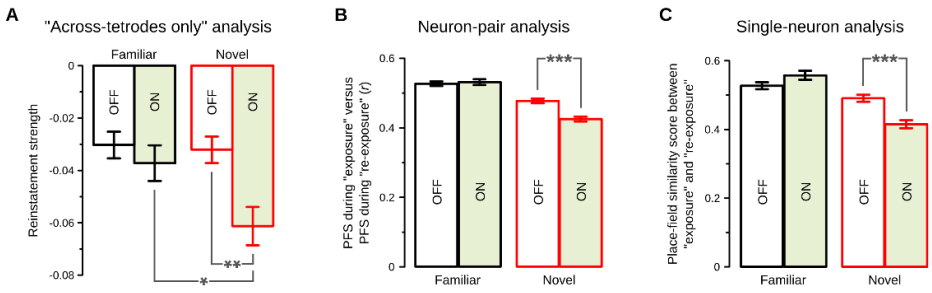

Figure 4.6 *Control analyses confirm the differential effect of SWR silencing on the reinstatement of neuronal representations of familiar versus novel space.*

(A) In order to restrict the central analysis of this chapter, namely the effect of SWR silencing on the reinstatement of "familiar" and "novel" assembly patterns, to neurons recorded exclusively from different tetrodes, here only assembly patterns are considered with for each tetrode at most one neuron with high contribution (>2 standard deviations above the mean) to that pattern. Such a conservative analysis also finds that SWR silencing impairs the reinstatement of "novel" assembly patterns (light-OFF: $n = 54$ assembly patterns; light-ON: $n = 39$), but not that of the "familiar" patterns (light-OFF: $n = 82$; light-ON: $n = 75$). This result is thus consistent with that shown for all assembly patterns in **Figure 4.5B**. Data are represented as mean ± SEM; * $P < 0.05$, ** $P < 0.01$, unpaired two-sample *t*-test.
(B) A pairwise spatial map similarity measure (e.g., see McNamara et al., 2014), which is calculated over all across-tetrodes principal neuron pairs as the Pearson correlation coefficient between their place-field similarity (PFS) scores during the exposure session and during the subsequent re-exposure (familiar, light-OFF: $n = 17{,}256$ neuron-pairs; familiar, light-ON: $n = 10{,}018$; novel, light-OFF: $n = 18{,}508$; novel, light-ON: $n = 16{,}335$). Error bars represent ± SE of the correlation coefficient; *** $P < 0.001$, *z*-test whether two correlations are different.
(C) A single-neuron place field similarity score (e.g., see Kentros et al., 1998), which is calculated for each principal neuron as the Pearson correlation between its spatial firing rate maps during the exposure session and during the subsequent re-exposure (familiar, light-OFF: $n = 788$ neurons; familiar, light-ON: $n = 519$; novel, light-OFF: $n = 1004$; novel, light-ON: $n = 927$;). Data are represented as mean ± SEM; *** $P < 0.001$, unpaired two-sample *t*-test.

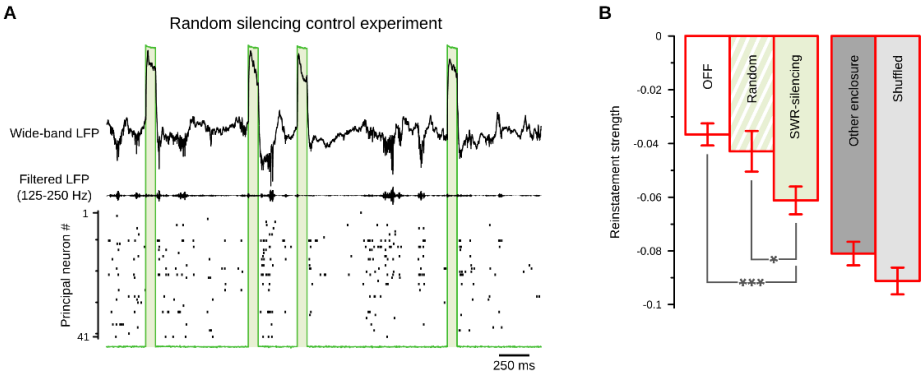

Figure 4.7 *Stronger assembly pattern reinstatement after random compared to SWR-triggered optogenetic silencing.*

(A) Raw data example from a random optogenetic silencing session. *Top trace*, wideband local field potential (LFP). *Bottom trace*, 125–250 Hz band pass filtered signal highlighting ripple frequency events. *Raster plot*, spike times from simultaneously recorded dorsal CA1 principal neurons (one neuron per row). Note the firing synchrony of principal neurons during SWR events that is spared by random optogenetic silencing performed independently of SWR occurrence.
(B) After exposure to a novel enclosure, SWR silencing (n = 136 assembly patterns, same data as in **Figure 4.5B**) impairs the upcoming reinstatement of assembly patterns during context re-exposure compared to when optogenetic silencing is not applied (light-OFF; n = 139, same data as in **Figure 4.5B**), but also compared to when principal neurons are randomly silenced (random; n = 60 assembly patterns from 5 mice). Note that because the reinstatement strength is defined as the change in a pattern's average expression strength from exposure to re-exposure, a null score would correspond to a "perfect" reinstatement while the more negative the worse the reinstatement. Further note that even in the baseline light-OFF condition, the average reinstatement strength is negative, which reflects that recalled "memory" patterns are not exact copies of those initially formed. For comparison purposes, also displayed are the reinstatement strength obtained in another enclosure (n = 258 assembly patterns) and the reinstatement strength obtained after shuffling the spike-identities during the re-exposure session (n = 335). The latter is displayed to show the expression strength level that could be expected with random coincidence of neuronal spike discharge. For the shuffling, all spikes discharged by principal neurons were randomly re-assigned to the pool of recorded principal neurons, under the restriction that for each neuron the total number of spikes assigned to it remained the same (this shuffling-procedure is the same as the "SWAP"-method used in Jackson et al., 2006; see also section 3.4). Data are represented as mean ± SEM; * $P < 0.05$, *** $P < 0.001$, unpaired two-sample t-test.

Are the "novel" enclosures really novel to the animal?

A crucial aspect of the above described study is the distinction between "familiar" and "novel" environments. For the interpretation of the results, it is therefore critical that the mice were able to distinguish between the familiar and the various novel enclosures. Throughout the experiment, each animal performed multiple recording days (5.3 ± 1.0 per mouse) with several recording blocks (2.2 ± 0.1 per day), meaning that each animal was exposed to many different "novel" enclosures (7.9 ± 2.2 per mouse; e.g., see **Figure 4.8A**). The familiar environment, to which the animal had repeatedly been exposed to before the recordings, was always the same

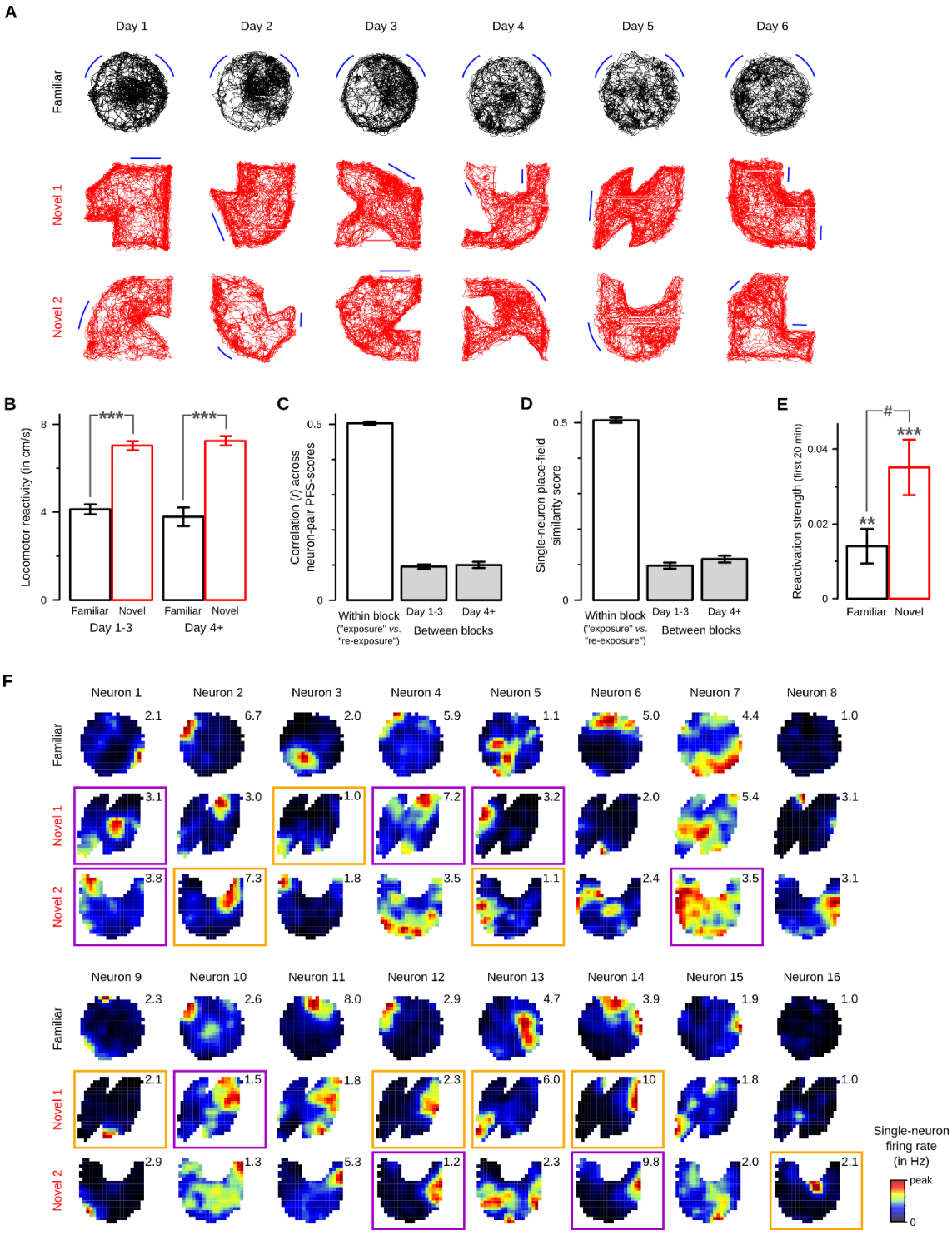

Figure 4.8 *Comparisons between "familiar" and "novel" recording blocks.*

(A) Example paths from one mouse recorded during various exposure sessions. Each path depicted corresponds to the open-field enclosure of one recording block. Besides differences in geometrical shape, the enclosures differed in the cue-cards (here schematically illustrated by blue lines) lining the walls. The particular order in which the "familiar" and "novel" recording blocks were presented varied between recording days.
(B) Locomotor reactivity to open-field enclosure, as quantified by the average locomotion speed during the first 5 min of the exposure session. The locomotion speed is substantially higher in a novel than in the

familiar enclosure, both during the first recording days (day 1–3, familiar: n = 19 sessions; day 1–3, novel: n = 28) and the last ones (day 4+, familiar: n = 11; day 4+, novel: n = 35). Data are represented as mean ± SEM; *** P < 0.001, unpaired two-sample t-test.
(C-D) Quantification of principal neuron remapping across enclosures. (C) Pearson correlation coefficient of the neuron-pair place-field similarity (PFS) scores between the indicated sessions, calculated over all principal neuron pairs (within recording block: n = 35,764 neuron-pairs; between blocks, day 1–3: n = 22,877; between blocks, day 4+: n = 13,004). Error bars represent ± SE of the correlation coefficient. (D) Single-neuron place field similarity score between the indicated sessions, averaged over all recorded principal neurons (mean ± SEM; within block: n = 1792 neurons; between blocks, day 1–3: n = 1207; between blocks, day 4+: n = 969).
(E) The reactivation strength (expression strength during "rest after" minus expression strength during "rest before") of assembly patterns is substantially stronger in the first 20 min after exposure to a novel enclosure (n = 139 assembly patterns) than in the first 20 min after exposure to the familiar enclosure (n = 108). Data are represented as mean ± SEM; ** P < 0.01, *** P < 0.001, one-sample t-test versus zero; $^{\#}$ P < 0.05, unpaired two-sample t-test.
(F) Example single-neuron firing rate maps expressed during the exposure sessions by a set of CA1 principal neurons simultaneously recorded throughout three recording blocks. The top right number of each map is the peak firing rate (in Hz). Warm colours (red) correspond to high firing rate regions (i.e., the place field) of the neuron in that enclosure. For both novel enclosures explored on that day by this mouse, colour coded squares indicate neurons with high contribution to an early stabilized (orange) and/or to a gradually strengthened (purple) assembly pattern detected in that enclosure (see section 4.4).

circular open-field enclosure with the same cue-cards lining its wall. Novel environments were created by arranging wooden blocks in a square box to achieve different shaped enclosures, with additional differences between them in the cue-cards lining some of the walls (e.g., see **Figure 4.8A**).

During the first 5 min of the exposure session, the animals' locomotion speed was substantially higher in novel environments than in the familiar (**Figure 4.8B**), indicating that mice successfully detected spatial novelty (e.g., Berke et al., 2008). Additionally, I found clear evidence that the recorded hippocampal principal neurons remapped between different enclosures (**Figure 4.8C-D,F**). Moreover, both the remapping and the difference in speed were present during the first (1–3) and the last (day 4+) recording days (**Figure 4.8B-D**), indicating that throughout the full duration of the experiment mice were capable of distinguishing between different enclosures and that also during later recording days they still treated "novel" enclosures as being novel. Finally, as would be predicted from the literature (O'Neill et al., 2008; McNamara et al., 2014), assembly pattern reactivation was substantially stronger in the first 20 min following exposure to a novel enclosure than in the first 20 min after exposure to the familiar one (**Figure 4.8E**).

4.4) Gradually strengthened, but not early stabilized, assembly patterns require reactivation

If repeated *offline* co-activation of neurons during sleep strengthens recently formed cell assemblies, I reasoned that repeated *on-line* co-activation of neurons during the initial exposure session might also strengthen such assemblies. To test for such a strengthening dynamic, a linear trend model (including intercept) was fitted to the expression strength (calculated in one second bins) of each assembly pattern during the first exposure to a novel enclosure. I found a significant positive slope for 134 out of 335 patterns (40.0%), compared to only 18 patterns (5.4%) with a significant negative slope. The patterns with a significant increasing linear trend are referred to as "gradually strengthened" (**Figure 4.9A**). Interestingly, the remaining patterns showed a similar strengthening only during the first few minutes (**Figure 4.9A**). This initial positive trend could reflect the rapid recruitment of these patterns during the first exposure to an enclosure, and these patterns are referred to as "early stabilized". Notably, gradually strengthened and early stabilized patterns had similar composition of their weight vectors and were equally spatially selective (**Figure 4.9B** and **Table 4.1**). A difference between both sets was that neurons with high contribution to gradually strengthened patterns had a somewhat higher average firing rate than high contributing neurons to early stabilized patterns (**Table 4.1**).

I finally tested whether these concurrently expressed assembly patterns equally required offline reactivation for their lasting expression. Both sets of patterns were reactivated, with the reactivation of the gradually strengthened patterns being stronger (**Table 4.1**). Importantly, only the reactivation strength of the gradually strengthened patterns, and not of the early stabilized ones, predicted their future reinstatement during context re-exposure (**Figure 4.9C**). Moreover, SWR silencing only impaired the reinstatement of the gradually strengthened patterns (**Figure 4.9D**; with interaction SWR silencing x pattern-type: $F(1,271) = 6.28$, $P < 0.05$). In the baseline condition (i.e., no optogenetic silencing), the reinstatement of both early stabilized and gradually strengthened patterns was context-dependent (**Figure 4.9D**; 'light-OFF' versus 'other

enclosure'). Optogenetic SWR silencing decreased the reinstatement of gradually strengthened patterns down to their non-specific strength level seen in a different enclosure, but it did not significantly affect the reinstatement of the early stabilized patterns (**Figure 4.9D**; 'light-ON' versus 'other enclosure').

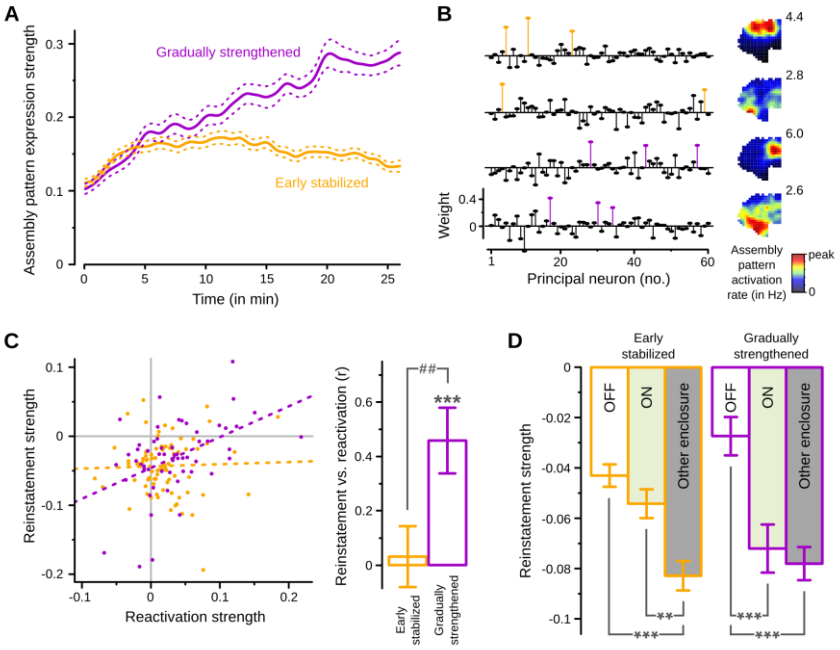

Figure 4.9 *Offline reactivation is required for reinstatement of gradually strengthened, but not of early stabilized, assembly patterns.*

(A) The expression strength of gradually strengthened patterns (purple; n = 134) continually increases during the exposure session, while that of early stabilized patterns (orange; n = 201) is more stable. Yet, both sets are equally strengthened in the first few minutes. For visualization-purposes, the individual expression strength time-courses were smoothed with a Gaussian kernel (SD = 5 sec) before calculating the presented mean ± SEM time-courses.
(B) Examples of two early stabilized (top) simultaneously expressed with two gradually strengthened (bottom) assembly patterns.
(C) In the light-OFF condition, the reactivation of gradually strengthened patterns, but not of early stabilized ones, correlates with their reinstatement strength during later context re-exposure. Dashed lines are corresponding ordinary least squares regression lines (Early stabilized: slope = 0.03, P = 0.78, R^2 = 0.00; Gradually strengthened: slope = 0.45, P < 0.001, R^2 = 0.21). Error bars represent ±1 SE of the correlation coefficient; Gradually strengthened versus zero: *** P < 0.001; Early stabilized versus Gradually strengthened: $^{\#\#}$ P < 0.01.
(D) SWR silencing does not impair the context-dependent reinstatement of early stabilized patterns (Light-OFF: n = 82 patterns; Light-ON: n = 83; Other enclosure: n = 155), but causes the reinstatement of gradually strengthened patterns to drop to the unspecific level at which they are expressed in a different enclosure of another recording block that day (Light-OFF: n = 57 patterns; Light-ON: n = 53; Other enclosure: n = 103). Data are represented as mean ± SEM; ** P < 0.01, *** P < 0.001.

[Note that one gradually strengthened assembly pattern had both an unusual high reactivation and reinstatement strength. This pattern would be very influential for the reinstatement-vs-reactivation correlation presented in panel C. I therefore left the data point corresponding to this pattern ([0.62, 0.14]) out from the graphs in panel C; its inclusion would further increase the correlation for gradually strengthened patterns to $r = 0.57$*.]*

Variable	Early stabilized	Gradually strengthened	P-value
Pattern composition			
Sparsity of weight vector	0.64 ± 0.01	0.63 ± 0.01	P = 0.44
Neurons with high contribution (%)	5.03 ± 0.14	5.27 ± 0.19	P = 0.29
Firing rate (in Hz) of high contributing neurons	**1.70 ± 0.09**	**2.11 ± 0.16**	**P < 0.05**
Expression strength			
First quarter of exposure session	0.148 ± 0.006	0.141 ± 0.009	P = 0.56
Last quarter of exposure session	**0.147 ± 0.006**	**0.279 ± 0.014**	**P < 0.0001**
Offline reactivation			
Reactivation strength	**0.019 ± 0.005**	**0.045 ± 0.013**	**P < 0.05**
Spatial selectivity			
Assembly-map coherence	0.85 ± 0.02	0.89 ± 0.03	P = 0.25
Assembly-map sparsity	0.56 ± 0.01	0.57 ± 0.02	P = 0.55
Assembly-field size (bins)	37.1 ± 1.9	37.5 ± 2.3	P = 0.89
Environment-specificity index	0.14 ± 0.02	0.17 ± 0.02	P = 0.34
Spatial remapping of high contributing neurons			
Single-neuron place field similarity score	0.16 ± 0.02	0.14 ± 0.03	P = 0.62
Correlation of member-pairs' PFS-scores	0.12 ± 0.07	0.15 ± 0.08	P = 0.80
Control variables			
Total number of neurons recorded	48.9 ± 1.3	47.5 ± 1.3	P = 0.44
Time spent in assembly-field (min)	3.02 ± 0.19	3.23 ± 0.23	P = 0.49

Table 4.1 *Characteristics of early stabilized and gradually strengthened assembly patterns.*

With the exception of the correlation of member-pairs' PFS-scores, reported data are mean ± SEM and the P-values refer to unpaired two-sample *t*-tests for difference in mean. A neuron is considered to have a high contribution to an assembly pattern if its weight to that pattern is more than 2 standard deviations above the mean.

The average reactivation strength is significantly greater than zero for both early stabilized ($t(81) = 4.16$, $P < 0.0001$, one-sample *t*-test) and gradually strengthened ($t(56) = 3.57$, $P < 0.001$) assembly patterns, indicating that both sets of patterns are preferentially reactivated during the sleep/rest session following the exposure session compared to a baseline sleep/rest session recorded before. A pattern's environment specificity index was defined as its maximum similarity index (see section 3.3) with any of the patterns detected during re-exposure minus its maximum similarity index with any of the patterns detected in a different enclosure from another recording block on the same day. The average environment-specificity index is significantly greater than zero for both early stabilized ($t(73) = 6.39$, $P < 0.0001$) and gradually strengthened ($t(54) = 7.05$, $P < 0.0001$) patterns, indicating that both sets of patterns are more similar to the patterns expressed during re-exposure to the same enclosure than to the patterns expressed in another enclosure.

The similar scores on the "single-neuron place field similarity score" indicate that high contributing neurons of early stabilized patterns remapped their place fields (compared to the enclosure of the previous recording block on that day) to a similar degree as high contributing neurons of gradually strengthened patterns. The similar scores on the "correlation of member-pairs' PFS-scores" indicate that there was no difference between early stabilized and gradually strengthened patterns in the extent to which pairs of neurons with high contribution to the same pattern already had overlapping place fields in the enclosure of the previous recording block. This last measure is reported ± SE of the correlation coefficient and the P-value refers to the comparison of both correlations with a *z*-test after application of Fisher's *r* to *z* transform.

4.5) Discussion

The idea that the stabilisation of newly formed cell assembly patterns involves their reactivation during resting behaviour has been a long-standing hypothesis central to many theories of memory consolidation, although it has never been directly tested. Here, by combining ensemble recordings with an unsupervised statistical framework, I identified short-timescale co-activation patterns of CA1 principal neurons, which had in the previous chapter been shown to be spatially selective. I demonstrated that for a specific set of these patterns, those with continued strengthening throughout their initial expression, the reinstatement during context re-exposure was both predicted by their reactivation and suppressed by optogenetic SWR silencing. These results indicate that the stabilization of recently formed, space-representing hippocampal cell assembly patterns depends on offline reactivation.

Time-limited role of SWR reactivation in the stability of neuronal representations of space could underlie memory consolidation

Previous studies had shown that post-learning disruption of sleep SWRs by electrical stimulation of the ventral hippocampal commissure impaired spatial memory performance (Girardeau et al., 2009; Ego-Stengel and Wilson, 2010). However, it was not possible in these studies to establish whether the observed impairment was caused by the disruption of the SWR-associated reactivation of waking firing patterns, or due to an unspecific effect of electrical stimulation coupled to SWRs. For example, electrical stimulation of the hippocampal commissure has been shown to trigger synaptic plasticity when applied with high frequency (Dragoi et al., 2003) and there is evidence suggesting that during SWRs the hippocampus is more susceptible to having synaptic plasticity induced (King et al., 1999; Sadowski et al., 2016). Moreover, it remained to be tested whether the effect of SWR disruption depends on such a manipulation being applied shortly after encoding. Indeed, to decisively demonstrate that a process has a role in consolidation, it is required to show that its disruption has a time-limited

effect, namely that its disruption only affects the persistence of traces of recent experiences and not those of remote ones (Dudai, 2004; Squire et al., 2015).

Here, SWR reactivation was directly silenced using an optogenetic approach. I found that this intervention disrupted the upcoming reinstatement of hippocampal assembly patterns when performed after the first exploration of a (thus novel) environment. Importantly, the same SWR silencing was ineffective on pattern reinstatement when performed after an environment that had been repetitively experienced before (hence familiar). This control condition rules out a generic effect of SWR disruption. Furthermore, the familiar environment is a "delayed-block condition" that establishes the time-limited role of SWR reactivation. These results, combined with the previously demonstrated behavioural effects, provide converging evidence that SWR reactivation supports memory consolidation by stabilizing the underlying cell assemblies.

Why are "familiar" patterns reactivated?

The lack of effect of SWR silencing on the reinstatement of assembly patterns representing a familiar environment raises the question why these patterns are still reactivated. One explanation could be that the repetitive explorations of a given environment lead to the formation of multiple "entry points" to the same assemblies. This is reminiscent of the idea that re-experiencing a given memory is associated with the formation of multiple neuronal traces (Nadel and Moscovitch, 1997; Moscovitch et al., 2006). In this scenario, reactivation following exploration of the familiar environment might still stabilize some of these additional traces, but SWR silencing is ineffective because previously stabilized traces are sufficient to retrieve the assembly patterns representing that environment. Another possibility is that reactivation no longer stabilizes "familiar" patterns within the hippocampus, but still contributes to their "transfer" to downstream circuits (Squire and Alvarez, 1995; see also discussion in section 6.2).

Early stabilized *vs*. gradually strengthened assembly patterns

The results presented in section 4.4 show that for those assembly patterns that had an increasing expression strength over continued experience in a novel environment, their subsequent reinstatement was predicted by their reactivation and disrupted by SWR silencing. The features of this set of patterns are consistent with the Hebbian postulate of "fire together, wire together". Conversely, the reinstatement of the other, concurrently expressed patterns that were no longer strengthened after the first few minutes of exploration was not predicted by their reactivation and unaffected by SWR silencing. Yet, both sets were equally spatially selective and thus appeared to carry a similar representational attribute. These findings indicate that concurrent space-representing assembly patterns can markedly differ in their plastic properties.

The early stabilized patterns might have gained independence from offline reactivation because they rapidly acquired the status of "familiar" patterns while stably expressed during the exposure session. In this scenario, their consolidation would take place "on-line" and the inefficacy of SWR silencing on these patterns would be an extreme reflection of the time-limited role of SWR reactivation. Perhaps, early stabilized patterns could be quickly consolidated because the place fields of their contributing neurons remap more coherently in the novel enclosure, for example according to a topographical transform, rather than unpredictably (*cf*. **Figure 4.8F**). Another, non-exclusive, possibility is that the early stabilized patterns are more "hard-wired" to represent the new spatial layout due to the specifics of their contributing neurons in terms of existing spatial inputs or intrinsic properties. The gradually strengthened patterns, in contrast, could gain their spatial selectivity and increased strength by more plastic changes throughout the exploration. Under this scenario, the difference in strength between both sets of patterns could reflect that such plasticity enables neurons to better synchronize with their peers. An interesting related hypothesis is that the early stabilized patterns could provide a "nearly-automatic" and yet stable representation of space, "ready-to-use" by downstream circuits, for instance for (immediate) navigational purposes. As the animal accumulates experiences in the environment, the strengthening of the gradually strengthened patterns could

reflect the formation of additional, perhaps richer, memory traces (Buzsáki, 2015b; Schiller et al., 2015).

Conclusions

Altogether, this chapter establishes that the lasting expression of recently formed hippocampal co-activation patterns that resemble classical "Hebbian" assemblies requires their offline reactivation. This finding supports the long-standing hypothesis of an instrumental role of offline SWR reactivation in the consolidation of memory-representing assemblies. However, reactivation-dependent assembly patterns were co-expressed with other space-coding patterns that did not require offline reactivation. As a pattern's dependency on offline SWR reactivation was related to its strengthening dynamics during spatial encoding, this chapter highlights functional heterogeneity within co-expressed representations of space.

CHAPTER 5: TOWARDS UNSUPERVISED CLASSIFICATION OF INTERNEURONS FROM EXTRACELLULAR RECORDINGS

5.1) Introduction

An important aspect that was missing in the original cell assembly model proposed by Donald Hebb was inhibition. Without a mechanism to counter the gradual strengthening of excitatory synaptic connections and the resulting build-up of activity, a connectionist network would converge to one giant, continually active cell assembly. Inhibitory interneurons were therefore proposed as addition to the cell assembly model to fulfil such a balancing role (Rochester et al., 1956; Milner, 1957). The contribution of inhibitory interneurons in maintaining a correct balance between excitation and inhibition is now widely acknowledged (e.g., Mariño et al., 2005; Yizhar et al., 2011).

Many different types of inhibitory neurons have been found in the brain, a much wider variety than for excitatory neurons (Markram et al., 2004). Moreover, anatomical features of interneurons (e.g., their post-synaptic targets) have been found to relate to their physiological properties (e.g., their firing pattern) (Klausberger et al., 2003). These observations suggest that interneurons might have additional, more specific roles than simply providing a correct overall balance of excitation and inhibition. Indeed, in the hippocampus, interneurons are thought to have important roles in organizing the general excitability of principal neurons (e.g., through the control of theta cycles or SWRs), but also more specific roles in the gating, filtering and selection of principal cell assembly patterns. For example, some interneuron types might play a role in selecting the principal neurons that will become part of a cell assembly, while others could be important for inhibiting neurons that do not belong to an active assembly.

That different types of interneurons have different roles in relation to principal cell assembly patterns was also suggested by a promising observation I made early during my DPhil studies. A figure capturing this observation was included in the Appendix of my Transfer of Status report and is reproduced here as **Figure 5.1**. This figure depicts three simultaneously recorded hippocampal interneurons—presumably of different types as they all have different firing rate, different shape of their auto-correllogram and different SWR-response; which couple differently to three spatially tuned hippocampal cell assembly patterns: one interneuron does not couple to any of the patterns, while the other two interneurons couple to one or two of the patterns but not to the other(s). This example has been the main motivation for the research presented in this chapter. To more formally investigate the coupling of different types of interneurons to principal cell assembly patterns, a first requirement is to be able to identify

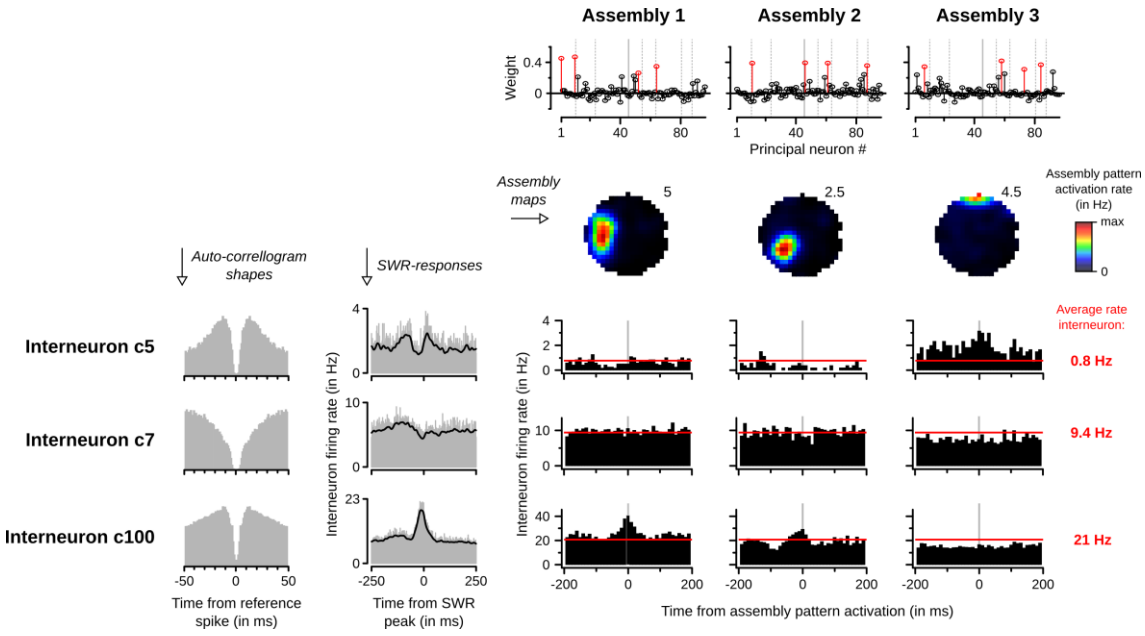

Figure 5.1 *Do different types of interneurons couple differently to principal cell assembly patterns?*

This example is based on a 26 min exploration session with 96 simultaneous recorded hippocampal principal neurons, on which the PCA/ICA-method identified 16 assembly patterns. The firing rate of three hippocampal interneurons is shown time-locked to the activations of three of these assembly patterns. Horizontal red lines indicate the average firing rate of the interneuron. The shape of the auto-correllogram and the SWR-response of the three interneurons, and the weight vectors and assembly spatial maps of the three assembly patterns are also depicted.

[This session was recorded by Natalia Campo-Urriza and Stéphanie Trouche, and was part of the data set of the McNamara et al. (2014) study.]

different types of interneurons from extracellular recordings, for which currently no generally accepted method exists. Therefore, a statistical framework to identify interneuron classes based on extracellular recordings alone is described in section 5.2 to 5.5. Section 5.6 uses the resulting classification to tentatively test for differences in the coupling to cell assembly patterns between the identified interneuron classes. Section 5.7 briefly discusses complementary approaches for investigating the roles of hippocampal interneurons.

5.2) Quantification of neuronal properties

Before neurons can be divided into different classes, the properties of each individual neuron need to be characterized. During my DPhil studies, I wrote a computer program—in the lab known under the name "neuronOverview"—that comprehensively characterizes a wide range of properties for each recorded neuron (see ***Appendix B***). It does so both qualitatively, in the form of graphs and figures, and quantitatively, in the form of numerical measures that can be used as input for the unsupervised clustering algorithms discussed in the next section. The names of these numerical measures extracted by this program are denoted in this section in $\boldsymbol{bold\ italic}$ font. This program is now widely used in the lab as a starting point for analysing newly obtained unit recordings.

Firing pattern

For each neuron, its firing rate (in Hz) is calculated over the full recording day ($\boldsymbol{firingRate}$), as well as separately for sleep/rest sessions ($\boldsymbol{firingRateSleep}$) and exploration sessions ($\boldsymbol{firingRateAwake}$). In addition, the firing rate is plotted individually for each session of the recording day (**Figure 5.2A**), for example to check the stability of a recorded neuron.

To visualize a neuron's firing pattern, the first plot generated is its auto-correllogram (**Figure 5.2B**). A neuron's auto-correllogram $\rho(t)$ is defined as its average firing rate (in Hz) triggered by its own spikes calculated in 1 ms time bins:

$$\rho(t) = \frac{\sum_{k=1}^{N} \sum_{(i=1, i \neq k)}^{N} \mathbf{1}_{(t_k + t - 1, t_k + t]}(t_i)}{N} \times 1000 \quad \text{for } t = \ \ldots, -2, -1, 0, 1, 2, \ldots \text{ ms}$$

with t_x the time (in ms) of spike x, N the total number of spikes fired by this neuron and $\mathbf{1}_A(x)$ an "indicator function" which equals 1 if $x \in A$ and 0 otherwise. Based on this plot, various numerical measures can be calculated to quantify a neuron's firing pattern. The mean of the auto-correllogram, which has been used—together with a neuron's average firing rate and a measure of its spike waveform—to separate hippocampal principal neurons from interneurons (Csicsvari et al., 1998, 1999b), is calculated as the time point (in ms) that divides the area under the auto-correllogram between 0 and 50 ms into half:

$$\boldsymbol{meanAuto} = \sum_{t=1}^{50} \rho(t)\, t \Big/ \sum_{t=1}^{50} \rho(t)$$

More recently, CA1 hippocampal principal neurons were separated from interneurons based on a measure for a neuron's refractory period and a measure for a neuron's burstiness (Royer et al., 2012). Following this paper, the index for a neuron's burstiness is defined as the difference between the peak of the auto-correllogram between 0 and 10 ms and the mean of the auto-correllogram between 40 and 50 ms, normalized by the maximum value of these two:

$$a = \max_{t=1,\ldots,10} \rho(t)$$

$$b = \frac{\sum_{t=41}^{50} \rho(t)}{10}$$

$$\boldsymbol{burstIndex} = \frac{a - b}{\max\{a, b\}}$$

With a small modification from the definition used by Royer and colleagues, the measure for a neuron's refractory period (in ms) is here defined as the smallest positive integer for which the slope of the auto-correllogram from zero to the corresponding bin's peak exceeds a threshold set at 0.15:

$$\boldsymbol{refractoryPeriod} = \min_{t \in \mathbb{Z}_+} \{t : \frac{\rho(t)}{t} > 0.15\}$$

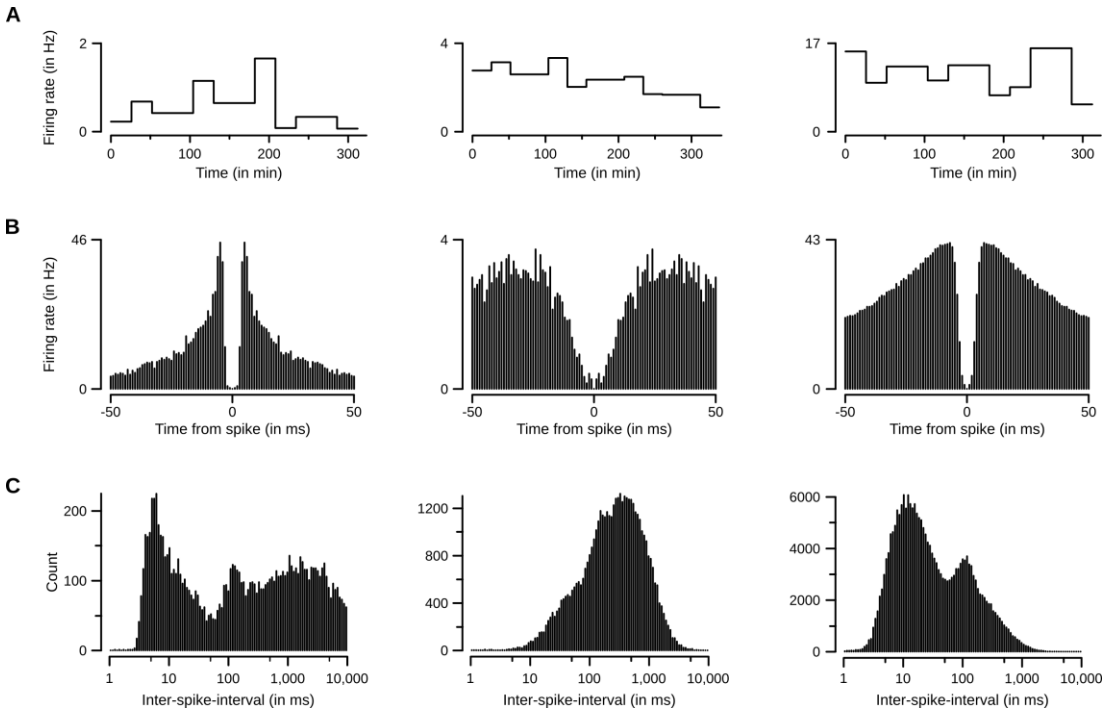

Figure 5.2 *Characterization of neurons' firing pattern.*

(A) The firing rate of three dorsal CA1 neurons is plotted for each session throughout the recording day. The characteristics of these three neurons are illustrated in this and the following four figures. They are selected examples of a putative principal neuron (left), a putative low-firing interneuron (middle) and a putative high-firing interneuron (right).
(B) The auto-correllograms of the same three neurons as in (A).
(C) The ISI-distributions, plotted with the *x*-axis at log-scale, of these three neurons.

The second plot to visualize a neuron's firing pattern is its ISI distribution (**Figure 5.2C**). If the spike times t_x $(1 \leq x \leq N)$ are ordered in time, the ISI of spike x is given by $\Delta t_x = t_x - t_{x-1}$ (for $x = 2, \ldots, N$). For visualization purposes, the distribution of these ISIs is plotted at log-scale. A classical measure for the variability of a neuron's spike train is the coefficient of variation of the ISI distribution (Holt et al., 1996), often abbreviated as CV, which is calculated as:

$$\boldsymbol{CV} = \frac{\sqrt{\frac{1}{N-2}\sum_{i=2}^{N}(\Delta t_i - \langle \Delta t \rangle)^2}}{\langle \Delta t \rangle}$$

whereby $\langle \Delta t \rangle = \frac{1}{N-1}\sum_{i=2}^{N} \Delta t_i$ is the mean ISI. To remove extreme outliers, all ISIs above 60 seconds are excluded for the calculation of the CV.

Spike waveform

For the analysis of a neuron's spike shape, the waveform of each spike is first band-pass filtered (800 Hz to 5 kHz), up-sampled to 40 kHz and aligned to its maximal through (see section 2.3). The waveforms of all a neuron's spikes are then averaged and the recording channel on which this average waveform has the highest amplitude is selected. The average spike waveform from this channel is plotted for visualization (**Figure 5.3**) and used to calculate the numerical measures described below. To visually check the stability of a neuron's spike shape throughout the recording day, the program "neuronOverview" additionally plots spike waveforms for the first and last session of each day: both the average waveforms from those sessions (in black) and 40 waveforms from randomly selected spikes of either session (in grey) are plotted for every channel of the tetrode that recorded the neuron (see ***Appendix B***).

Let $s(t)$ describe a neuron's average spike waveform on the selected channel (with t being the time from the maximal through and ranging from -0.8 ms to 0.8 ms). The spike amplitude is then defined as:

$$\boldsymbol{spikeAmplitude} = \max_t\{s(t)\} - \min_t\{s(t)\}$$

Note that because the spike amplitude depends for a large part on the distance between the recording tetrode and the recorded neuron, this measure is expected to carry relatively little information about the neuron itself. The duration and the shape of the spike waveform, on the other hand, are thought to be less depended on the recording distance and to contain more

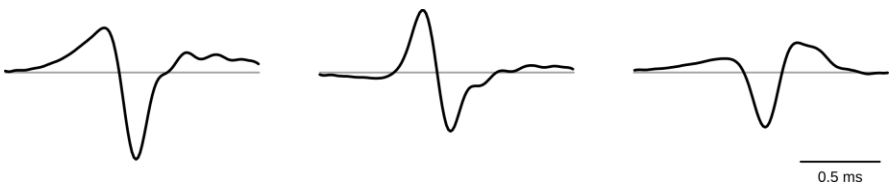

Figure 5.3 *Characterization of neurons' spike waveform.*

The average spike waveform (over the full recording day) of the same three dorsal CA1 neurons as in **Figure 5.2**.

information about the identity of the recorded neuron. Let me first define the following three time points:

$$t_{\text{min}} = \underset{t}{\text{argmin}}\{s(t)\}$$

$$t_{\text{first peak}} = \underset{t<t_{\text{min}}}{\text{argmax}}\{s(t)\}$$

$$t_{\text{second peak}} = \underset{t>t_{\text{min}}}{\text{argmax}}\{s(t)\}$$

The duration of the spike waveform is then characterized with the following measure (Weir et al., 2014):

$$\boldsymbol{spikeHalfwidth} = \min_{t>t_{\text{min}}}\{t: s(t) > 0.5 \min_{t}[s(t)]\} - \max_{t<t_{\text{min}}}\{t: s(t) > 0.5 \min_{t}[s(t)]\}$$

And the shape of the waveform is described by the measure (Royer et al., 2012; Weir et al., 2014):

$$\boldsymbol{spikeAsymmetry} = \frac{s(t_{\text{first peak}}) - s(t_{\text{second peak}})}{s(t_{\text{first peak}}) + s(t_{\text{second peak}})}$$

This last measure is bounded between -1 and 1, and is positive if the first peak is higher than the second peak and negative if the second peak is higher.

Spatial tuning

For each exploration session that lasted at least 10 min, a neuron's firing rate map (see section 2.3) is generated (**Figure 5.4**). To quantify a neuron's spatial tuning properties, sparsity and coherence (see section 2.3) are calculated for each session's map individually and then averaged to arrive at one value for each measure (***aveSparsity*** and ***aveCoherence***). Sparsity reflects how concentrated in space the firing activity of a neuron is, while coherence reflects how similar a neuron's average firing activity is between neighbouring spatial bins.

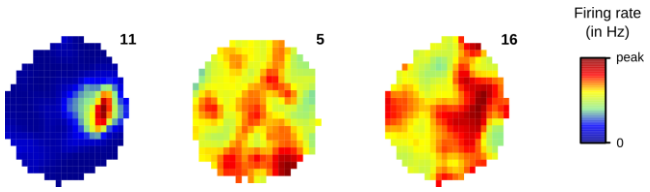

Figure 5.4 *Characterization of neurons' spatial tuning.*

The spatial firing rate maps (numbers indicate peak firing rate) of a 26 minute exploration session of the same three dorsal CA1 neurons as in **Figure 5.2**.

Phase coupling

The coupling of a neuron's spike to the phase of ongoing local field potential (LFP) oscillations in the CA1 pyramidal cell layer is calculated for four different frequency bands: theta (5–12 Hz), low gamma (30–50 Hz), high gamma (55–90 Hz) and ripple (120–250 Hz). For each of these frequency bands, the assessment of phase coupling is restricted to intervals in which the oscillations corresponding to that band are prominent.

Step 1: find "allowed intervals" for each frequency band

For the theta coupling, only spikes during "good" theta cycles are considered. To detect theta cycles on a given tetrode, the LFP signal recorded from one of its channels throughout all exploration sessions is first filtered in the theta band (5–12 Hz band-pass, 4^{th} order Butterworth filter). Peaks are detected in this filtered signal under the requirement that subsequent peaks are separated by at least 83 ms (i.e., ~12 Hz cycle; if there are two peaks within 83 ms from each other, only the highest one is kept) and that a peak needs to be higher than 0.5 SD of the filtered signal. Troughs are then detected similarly. For the next step, the instantaneous delta and theta power are calculated as the absolute value of the analytic signal obtained after taking the Hilbert transform on the recorded LFP signal filtered in the delta (2 Hz low pass, 4^{th} order Butterworth filter) and theta band (5–12 Hz band-pass, 4^{th} order Butterworth filter), respectively. Only peaks and troughs at times for which the instantaneous theta power exceeds the instantaneous delta power are kept. For every remaining peak that is both preceded and followed by a trough

between 33 ms and 167 ms, a "candidate" theta cycle is defined as ranging from the preceding trough to the subsequent trough. This procedure is repeated for all tetrodes in the CA1 pyramidal layer, and those theta cycles of the tetrode with the most theta cycles adhering to these criteria are selected. As a final step, only those theta cycles during which the average speed of the animal exceeds 2 cm/s are kept and called "good" theta cycles.[10]

Assessment of the coupling to low and high gamma oscillations is also restricted to time periods inside "good" theta cycles, but with the additional criterion that the instantaneous power in their respective bands needs to exceed 1.7 SD above the mean. Similar as for the delta and theta power, the instantaneous low gamma and high gamma power are calculated as the absolute value of the analytic signal obtained after taking the Hilbert transform on the recorded LFP signal (of the tetrode with the most "candidate" theta cycles) filtered in the low gamma (30–50 Hz band-pass, 4th order Butterworth filter) and high gamma band (55–90 Hz band-pass, 4th order Butterworth filter), respectively.

Assessment of the coupling to ripple oscillations is restricted to intervals identified as SWR events (see section 2.2 and McNamara et al., 2014) during sleep/rest sessions without any optogenetic light pulses.

Step 2: assign a phase to each spike

For every frequency band, each spike within its "allowed interval" is then assigned a phase. For the theta, low gamma and high gamma band, phase assignment is based on the LFP signal recorded from the tetrode with the most "candidate" theta cycles. For the ripple band, phase assignment is based on the LFP signal recorded from the tetrode on which that neuron was recorded. For each frequency band, the signal's instantaneous phase (in radians) was calculated at each point in time as the argument of the analytic signal obtained after taking the Hilbert

[10] Note that the procedure for detections of "good" theta cycles has largely been based on an analysis protocol suggested and tested by Vítor Lopes dos Santos.

transform of the signal filtered in the appropriate band (theta: 5–12 Hz, low gamma: 30–50 Hz, high gamma: 55–90 Hz, ripple: 140–200 Hz; all 4th order Butterworth band-pass filters).

Step 3: calculate statistics and generate graphs

The distribution of phases assigned to a neuron's spikes are plotted for each of the frequency bands (**Figure 5.5**). The mean phase (in radians) of a neuron's spikes to oscillations of a particular frequency band is calculated as:

$$\boldsymbol{phase}_{\textbf{[Band]}} = \text{atan2}\left(\frac{\sum_{i=1}^{N} \sin a_i}{N}, \frac{\sum_{i=1}^{N} \cos a_i}{N}\right)$$

with a_i the phase (in radians) assigned to spike i, N the total number of spikes that were assigned a phase and $\text{atan2}(x, y)$ the function that returns the angle (in radian) between a plane's positive x-axis and the point (x, y). Note that **[Band]** could be either **Theta**, **LowGamma**, **HighGamma** or **Ripple**. The mean resultant length, a measure of the strength of a neuron's phase coupling (on a scale from 0 to 1), is then calculated as:

$$\boldsymbol{r}_{\textbf{[Band]}} = \sqrt{\left(\frac{\sum_{i=1}^{N} \sin a_i}{N}\right)^2 + \left(\frac{\sum_{i=1}^{N} \cos a_i}{N}\right)^2}$$

Whether the distribution of phases assigned to a neuron's spikes differs from a uniform distribution around the circle can be tested with Rayleigh's test for circular uniformity (Zar, 1999; p.617). The p-value of this test is also reported ($\boldsymbol{p}_{[\boldsymbol{Band}]}$). For this thesis, the three numerical measures discussed above were calculated in R using the package "circular" (Agostinelli and Lund, 2013). Also, to be able to use them for clustering, the Polar coordinates $\boldsymbol{phase}_{\textbf{[Band]}}$ and $\boldsymbol{r}_{\textbf{[Band]}}$ were converted to Cartesian coordinates as follows:

$$\boldsymbol{x}_{\textbf{[Band]}} = \boldsymbol{r}_{\textbf{[Band]}} \times \cos\left(\boldsymbol{phase}_{\textbf{[Band]}}\right)$$

$$\boldsymbol{y}_{\textbf{[Band]}} = \boldsymbol{r}_{\textbf{[Band]}} \times \sin\left(\boldsymbol{phase}_{\textbf{[Band]}}\right)$$

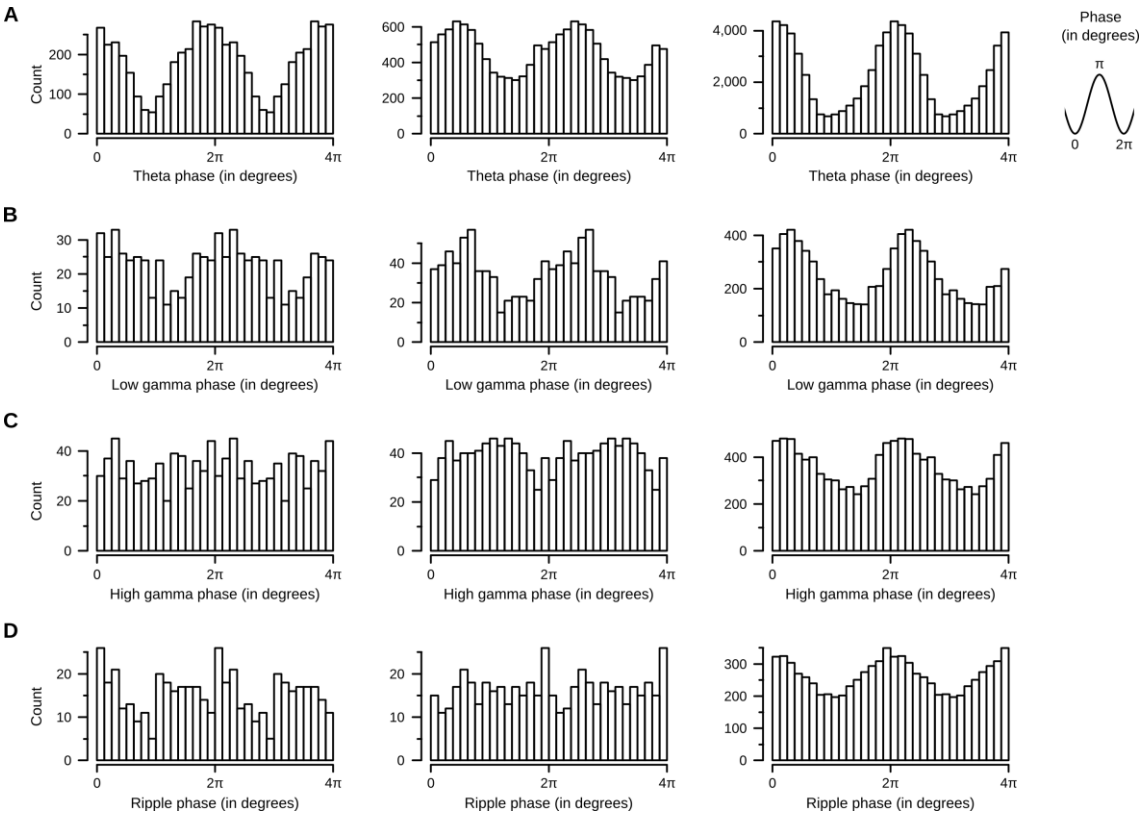

Figure 5.5 *Characterization of neurons' phase coupling.*

The phase distribution of the spikes of the same three dorsal CA1 neurons as in **Figure 5.2** for four different frequency bands: (A) theta (5–12 Hz); (B) low gamma (30–50 Hz); (C) high gamma (55–90 Hz); and (D) ripple (120–250 Hz).

Response to SWRs

A neuron's SWR-response $\sigma_{\mathrm{SWR}}(t)$ is defined as its average firing rate (in Hz) triggered by the peak power of offline identified SWR events (see section 2.2 and McNamara et al., 2014) calculated in 1 ms time bins:

$$\sigma_{\mathrm{SWR}}(t) = \frac{\sum_{m=1}^{M}\sum_{i=1}^{N}\mathbf{1}_{(T_m+t-1,T_m+t]}(t_i)}{M} \times 1000 \quad \text{for } t = \ldots,-2,-1,0,1,2,\ldots \text{ ms}$$

with t_i the time (in ms) of spike i, T_m the time (in ms) of the peak power of SWR event m, N the total number of spikes fired by this neuron, M the total number of SWR events and $\mathbf{1}_A(x)$ an "indicator function" which equals 1 if $x \in A$ and 0 otherwise. This SWR-response is plotted

for t ranging from -250 ms to 250 ms, both raw (**Figure 5.6**; grey histogram) and after smoothing with a Gaussian kernel with SD of 5 ms (**Figure 5.6**; black curve):

$$\widetilde{\sigma_{\mathrm{SWR}}}(t) = \sigma_{\mathrm{SWR}}(t) * g(t)$$

with t in ms, $g(t) = \frac{1}{\sqrt{50\pi}} e^{-\frac{t^2}{50}}$ the kernel function and $*$ denoting the "convolution operator".

Based on visual inspection of the SWR-response of many neurons, the following numerical measures were manually crafted to characterize the most salient features of a neuron's SWR-response. Firstly, defining a neuron's baseline firing rate (in Hz) as $baseRate = \frac{1}{250}\sum_{t=-500}^{-251} \widetilde{\sigma_{\mathrm{SWR}}}(t) + \frac{1}{250}\sum_{t=251}^{500} \widetilde{\sigma_{\mathrm{SWR}}}(t)$, the coupling of neuron's activity to SWR events is quantified as:

$$\boldsymbol{SWRcoupling} = \frac{\frac{1}{100}\sum_{-10}^{10} \widetilde{\sigma_{\mathrm{SWR}}}(t)}{baseRate}$$

$$\boldsymbol{SWRcouplingWide} = \frac{\frac{1}{100}\sum_{-50}^{50} \widetilde{\sigma_{\mathrm{SWR}}}(t)}{baseRate}$$

The magnitude of the peak firing of a neuron during SWR events (relative to its baseline firing), is quantified with the measure:

$$\boldsymbol{SWRpeak} = \frac{\max_{-100<t<100} \widetilde{\sigma_{\mathrm{SWR}}}(t)}{baseRate}$$

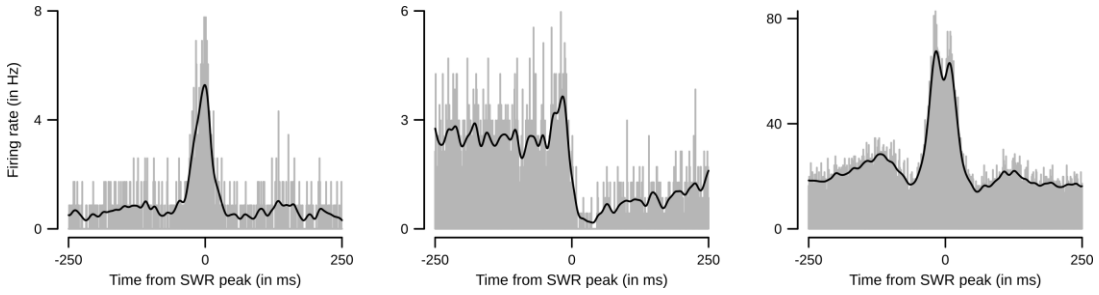

Figure 5.6 *Characterization of neurons' SWR-response.*

The SWR-response of the same three dorsal CA1 neurons as in **Figure 5.2**, both raw [grey histograms, $\sigma_{SWR}(t)$] and after smoothing with a Gaussian kernel with SD of 5 ms [black curve, $\widetilde{\sigma_{SWR}}(t)$].

A neuron's point of lowest firing during SWR events is similarly characterized by:

$$\boldsymbol{SWRtrough} = \frac{\min\limits_{-100<t<100} \widetilde{\sigma_{\mathrm{SWR}}}(t)}{baseRate}$$

The asymmetry between the average rate preceding and the average rate following the peak power of a SWR event is characterized by these two measures:

$$\boldsymbol{SWRassymetry} = \frac{\sum_{t=-50}^{-1} \sigma_{\mathrm{SWR}}(t) - \sum_{t=1}^{50} \sigma_{\mathrm{SWR}}(t)}{\sum_{t=-50}^{-1} \sigma_{\mathrm{SWR}}(t) + \sum_{t=1}^{50} \sigma_{\mathrm{SWR}}(t)}$$

$$\boldsymbol{SWRlongAssymetry} = \frac{\sum_{t=-200}^{-51} \sigma_{\mathrm{SWR}}(t) - \sum_{t=51}^{200} \sigma_{\mathrm{SWR}}(t)}{\sum_{t=-200}^{-51} \sigma_{\mathrm{SWR}}(t) + \sum_{t=51}^{200} \sigma_{\mathrm{SWR}}(t)}$$

These two measures are bounded between -1 and 1, and are positive if a neuron's rate preceding the SWR peak power is higher than the rate following it and negative if the rate following is higher. Finally, I noticed that some—but definitely not all—of the strongly SWR-coupled interneurons, instead of having a single peak in their SWR-response, seemed to have two separated peaks (e.g., **Figure 5.6**, right). To quantify this, all peaks in $\widetilde{\sigma_{\mathrm{SWR}}}(t)$ for t between -30 ms and 30 ms that are more than 3 SD_{base} above the $baseRate$ are extracted (with SD_{base} the standard deviation of $\widetilde{\sigma_{\mathrm{SWR}}}(t)$ for the intervals [-500,250] and [250,500]). If there are two such peaks, $\boldsymbol{SWRvalley}$ is defined as the height of the smallest peak minus the lowest point between both peaks, divided by the $baseRate$ (to remove direct dependence on a neuron's firing rate). If there are more than two such peaks, this subtraction is performed for all adjacent pairs, and $\boldsymbol{SWRvalley}$ is set to the maximum of all subtractions (divided by the $baseRate$). If there are less than two such peaks, $\boldsymbol{SWRvalley}$ is set to 0.

Characteristics of recording tetrode

From the characteristics of the recording tetrode, information can be gained about the location of a recorded neuron. Although in this thesis I only consider neurons recorded from tetrodes targeted to the pyramidal cell layer of the CA1, there are differences in the exact location these tetrodes recorded from. First, tetrodes were placed in either the left or right hemisphere. Second,

within the pyramidal cell layer, the location of the recording tip of tetrodes can vary along three axes: dorso-ventral, proximo-distal and superficial-deep. Information about the location along the dorso-ventral and proximo-distal axes could for example be gained from the drive lay-out, the implantation coordinates and/or post-mortem identification of the tetrode tracks, but for this thesis I have not attempted to gather information with regard to these two axes. I did, on the other hand, attempt to estimate a tetrode's recording location along the superficial-deep axis.

A first indication about how central a tetrode's recording location was with regard to the superficial-deep axis of the pyramidal cell layer is provided by the number of putative principal cells recorded on it (***nrPrinCells***). The higher this number, the more likely the tetrode's recording tip was located near the middle of the pyramidal cell layer. Another indication that a tetrode was well placed in the cell layer is the presence of high ripple power during SWR events. To visually assess this, a SWR-triggered average spectrogram is plotted (**Figure 5.7**; colour plot). For this plot, a morlet wavelet spectrogram is first calculated for a tetrode's recorded signal for frequencies ranging from 80 Hz to 260 Hz (in steps of 10 Hz), and *z*-scored for each frequency independently.[11] The average of each of these *z*-scored spectrograms is then calculated relative to offline identified SWR events in a 160 ms window around their peak ripple power and plotted in colour scale (**Figure 5.7**). To summarize the amount of ripple power during SWR events with one numerical value, a tetrode's signal is band-pass filtered in the ripple frequency band (140–200 Hz, 4th order Butterworth filter) after which the instantaneous amplitude in the ripple band is calculated as the absolute value of the analytic signal obtained after the Hilbert transform on this filtered signal. This instantaneous amplitude is then *z*-scored over all sleep/rest sessions of that recording day without any light pulses delivered. The ***rippleEnvelope*** is defined as the average of this *z*-scored instantaneous amplitude in a window of 40 ms around the peak ripple power of offline identified SWR events.

[11] All computations described in this paragraph are performed on the LFP signal recorded by one of the tetrode's channels during all sleep/rest sessions of that recording day without any light pulses delivered, after that signal has first been detrended and the signal of the reference tetrode has been subtracted from it.

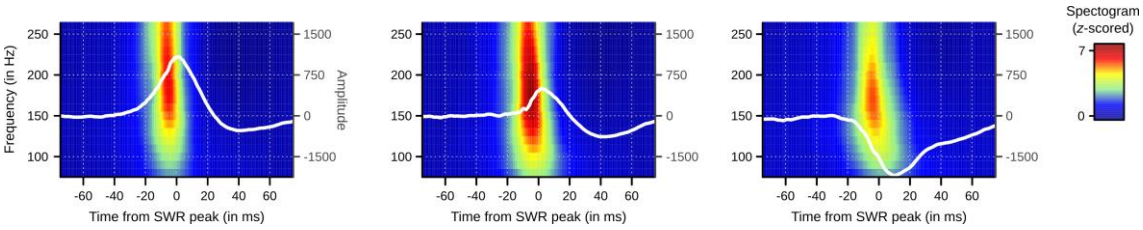

Figure 5.7 *Characterization of neurons' recording tetrode.*

The SWR-triggered average LFP (white curve) and spectrogram (colour plot) of three different recording tetrodes targeted to the CA1 pyramidal layer. The recording tips of these tetrode were likely located towards the top of the pyramidal cell layer (i.e., close to *stratum oriens*) [left; $\boldsymbol{rippleEnvelope} = 3.2$, $\boldsymbol{SWamplitude} = 839$], in the middle of the layer [middle; $\boldsymbol{rippleEnvelope} = 4.0$, $\boldsymbol{SWamplitude} = 103$] and towards the bottom of the layer (i.e., close to *stratum radiatum*) [right; $\boldsymbol{rippleEnvelope} = 3.6$, $\boldsymbol{SWamplitude} = -1096$].

An indication whether a tetrode's recording location was more towards the top of the pyramidal cell layer (i.e., towards *stratum oriens*, "deep") or more towards its bottom (i.e., towards *stratum radiatum*, "superficial"), can be gained from whether the deflection of the sharp wave accompanying the ripples during SWR events is positive (indicative of a location towards the *stratum oriens*) or negative (indicative of a location towards the *stratum radiatum*). To visually assess this, a tetrode's recorded signal is plotted averaged around the peak ripple power of offline identified SWR events (**Figure 5.7**; white trace). To quantify the average deflection of this sharp wave, a tetrode's signal is low-pass filtered (<25 Hz, 4th order Butterworth filter) after which its instantaneous amplitude is calculated as the absolute value of the analytic signal obtained after the Hilbert transform on this filtered signal. The $\boldsymbol{SWamplitude}$ is then defined as the average value over all SWR events of the original low-pass filtered signal at the time point during each SWR event for which this instantaneous amplitude is maximal.[12]

[12] Note that the plots in **Figure 5.7** and the measures $\boldsymbol{rippleEnvelope}$ and $\boldsymbol{SWamplitude}$ were largely inspired by an analysis protocol suggested by Vítor Lopes dos Santos.

5.3) Principal neurons *vs*. interneurons

In this section, the numerical measures described in the previous section are used to attempt to divide extracellularly recorded CA1 neurons into principal neurons and interneurons. In the next two sections, it will be attempted to further classify the putative interneurons into various types. In order to establish and evaluate these classification models, the full database of tetrode recordings from the dorsal CA1 region of the mouse hippocampus in our lab is used. In total, this data set consists of 7,429 neurons from 254 recording days from 47 animals (recorded by 8 investigators)[13,14].

As discussed in section 2.3, for the analyses presented in the previous two chapters, principal neurons were manually identified based on a visual assessment of the shape of their auto-correllogram, their firing rate and their spike waveform. To avoid dependence on the subjective judgement of different investigators, here an alternative way to divide CA1 neurons into principal neurons and interneurons based on some of the above extracted measures is proposed. Previous attempts to make such a division solely based on numerical measures were based on the firing rate and the mean of the auto-correllogram (Csicsvari et al., 1998, 1999b) or on the burstiness and the refractory period (Royer et al., 2012). Here, to combine the best of both approaches, a *k*-means clustering algorithm (see section 5.5) is used to divide the full data set of neurons into two clusters based on all four numerical measures ($\boldsymbol{firingRate}$, $\boldsymbol{meanAuto}$, $\boldsymbol{burstIndex}$ and $\boldsymbol{refractoryPeriod}$). Before clustering, the scale of these four measures is normalized by *z*-scoring them. This procedure results in an objective and seemingly satisfactory division between principal neurons (n = 6,628) and interneurons (n = 801) that is consistent with the current literature (**Figure 5.8**). The *k*-means algorithm, as well the clustering approaches

[13] The full data set actually contains 7,508 dorsal CA1 neurons, but 79 of those are excluded because they have missing values for one or more of the measures discussed in section 5.2.
[14] The seven other investigators that performed recordings included in this chapter are Claire Bratley, Natalia Campo-Urriza, Steven Cuell, David Dupret, Vadim Koren, Colin McNamara and Stéphanie Trouche.

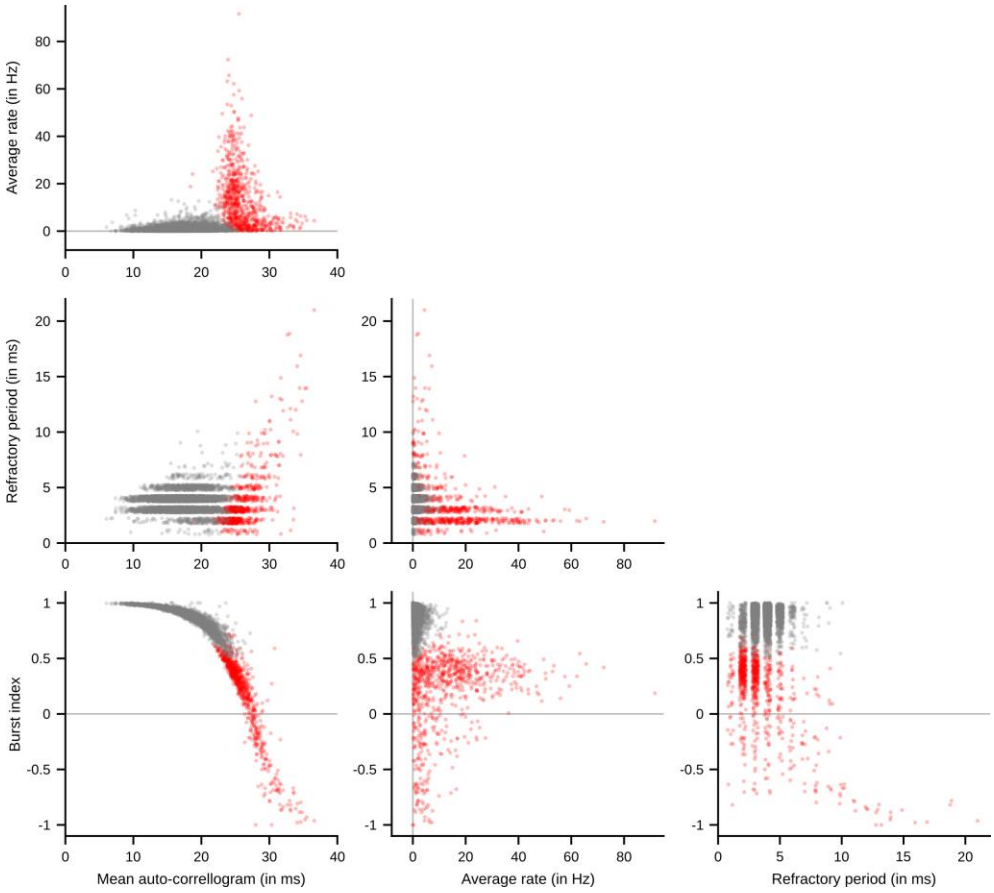

Figure 5.8 *Separation of principal neurons and interneurons.*

Using *k*-means clustering on their firing rate, the mean of their auto-correllogram, their burstiness and their refractory period, extracellular recorded CA1 neurons are divided into putative principal neurons (grey; n = 6628) and putative interneurons (red; n = 801). For visualization, jitter (uniformly distributed between -0.25 and 0.25 ms) is added to neurons' refractory period.

used in the remainder of this chapter, is implemented in the Python module *Scikit-learn* (Pedregosa et al., 2011).

5.4) Clustering interneurons based on SWR-response

The next aim is to subdivide the putative interneurons into different classes. It has been shown that anatomically identified interneuron types have characteristic responses to SWRs (Klausberger and Somogyi, 2008; Somogyi et al., 2014). Motivated by that, in this section it is

attempted to cluster interneurons solely based on their SWR-response. This approach allows to compare, and to validate, the resulting clusters based on other characteristics of the neurons (e.g., spike waveform, phase coupling, firing pattern, tetrode depth). The idea is that if interneuron clusters identified using only SWR-response measures would also differ in other characteristics, this would indicate that the interneuron data set has a structure that is suitable for clustering.

The input features based on which the interneurons are clustered in this section are the seven SWR-response measures described above: $\boldsymbol{SWRcoupling}$, $\boldsymbol{SWRcouplingWide}$, $\boldsymbol{SWRpeak}$, $\boldsymbol{SWRtrough}$, $\boldsymbol{SWRassymmetry}$, $\boldsymbol{SWRlongAssymmetry}$, $\boldsymbol{SWRvalley}$. Note that none of these measures directly depends on a neuron's firing rate, as they are all normalized by either the $baseRate$ or by the average rate during a specific interval of the SWR-response. Because these measures can have erratic, outlying values if a neuron has too few spikes during SWRs, neurons with an average rate below 0.5 Hz in a window of 500 ms around the SWR peak power are left out in this and the following three sections. This criterion excludes 44 interneurons, leaving 757 putative CA1 interneurons to cluster on. Based on the seven SWR-response measures, a *k*-means clustering algorithm (see section 5.5) is used to subdivide these interneurons into four clusters. Before clustering, to give all measures equal weight regardless of their scale, each measure was first normalized by *z*-scoring. In this section, since the main aim is merely to explore whether clustering of interneurons based on extracellular recordings is feasible and, if so, which measures should be included, no different number of potential clusters or alternative clustering algorithms are considered. This will, however, be done in the next section. Unsurprisingly, the four clusters found by the *k*-means algorithm differ in the shape of their average SWR-response (**Figure 5.9A**). Importantly, these interneuron clusters identified solely based on SWR-response measures also differ considerably in other characteristics, such as their firing pattern, their spike waveform and their phase coupling (**Figure 5.9B-M**). The suitability, or validity, of clustering on this data set is further highlighted by the application of

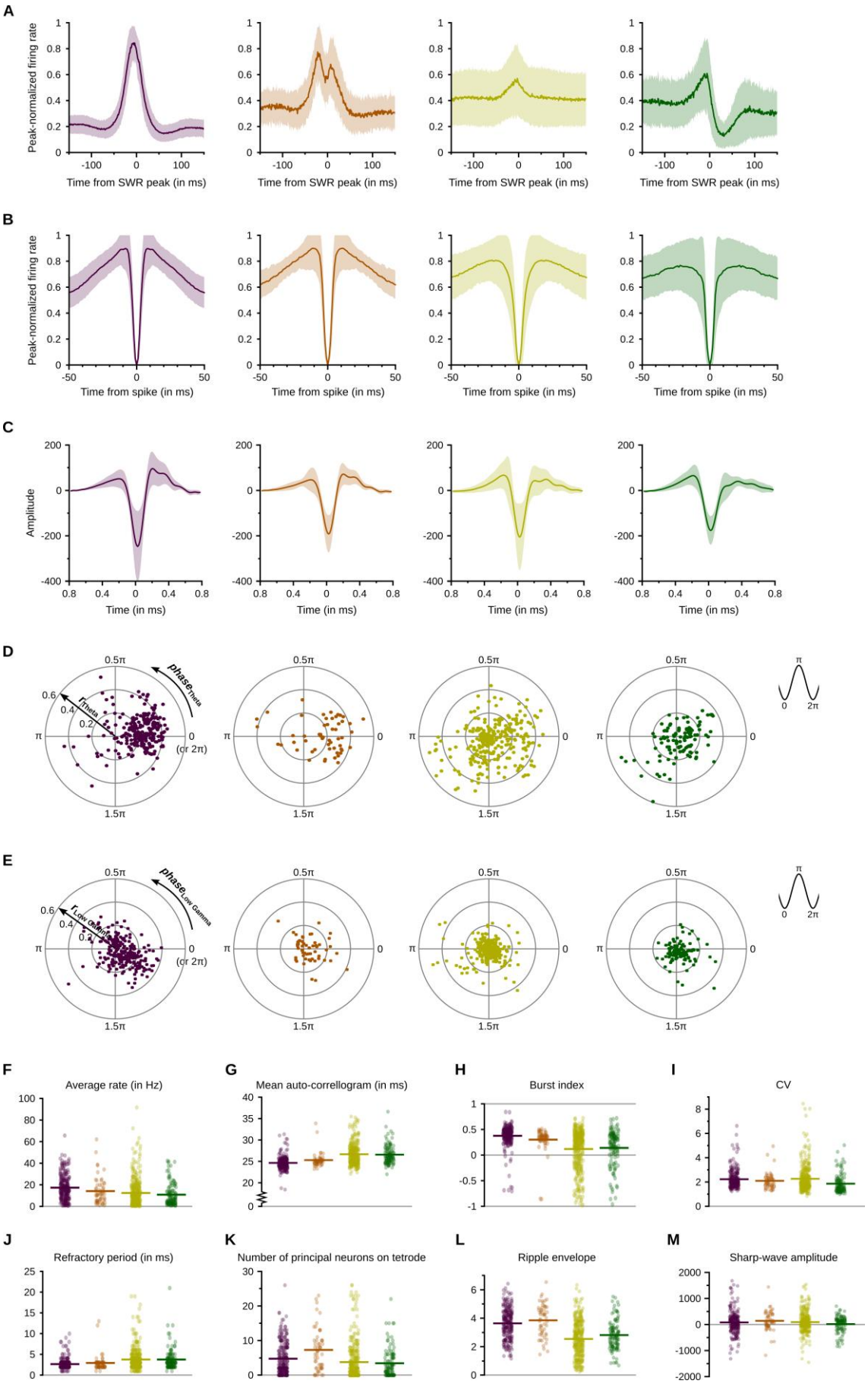
A
Peak-normalized firing rate
Time from SWR peak (in ms)
B
Peak-normalized firing rate
Time from spike (in ms)
C
Amplitude
Time (in ms)
D
phase Theta
r Theta
(or 2π)
E
phase Low Gamma
r Low Gamma
(or 2π)
F
Average rate (in Hz)
G
Mean auto-correllogram (in ms)
H
Burst index
I
CV
J
Refractory period (in ms)
K
Number of principal neurons on tetrode
L
Ripple envelope
M
Sharp-wave amplitude

Figure 5.9 *Interneuron clusters identified based on SWR-response measures.*

Using *k*-means clustering on seven manually crafted features describing the shape of their SWR-response, 757 putative interneurons are divided into four clusters. (A) Each cluster's peak-normalized SWR-response (mean ± SD), with the clusters ordered according to their average firing rate. The SWR-responses of these clusters are characterized by a single peak (dark purple; $n = 249$ neurons), a double peak (dark orange; $n = 56$), no clear coupling (dark yellow $n = 336$) and an increase in rate followed by a decrease (dark green, $n = 116$).
(B-M) These SWR-response clusters also considerably differ in some of their other characteristics: (B) peak-normalized auto-corrollogram (mean ± SD); (C) spike waveform (mean ± SD); (D) coupling to theta phase; (E) coupling to low gamma phase; (F) average firing rate; (G) mean of the auto-correllogram; (H) burst index; (I) coefficient of variation (CV); (J) refractory period; (K) number of principal neurons recorded on the same tetrode; (L) ripple envelope; (M) sharp-wave amplitude. The average firing rates during sleep/rest and during exploration sessions have similar distributions over these SWR-response clusters as the average overall firing rate, and are therefore not also shown. The coupling to high gamma and to ripple oscillations, and the average sparsity and coherence of the firing rate map, all do not clearly differ between these SWR-response clusters and are therefore not shown.
[One animal (three recording days) was recorded on a recording system with a different amplification of the recorded signal, causing for example the amplitude of the spike waveforms to be considerably larger. Putative interneurons (n = 23) recorded from this animal are therefore left out in the plot of the spike waveforms (C) and in the plot of the sharp-wave amplitudes (M). As the amplification of the recorded signal should not affect any of the other measures, these neurons were included everywhere else.]

dimensionality reduction techniques on a set of numerical measures that were not included in the initial clustering ($\boldsymbol{firingRate}$, $\boldsymbol{burstIndex}$, $\boldsymbol{refractoryPeriod}$, $\boldsymbol{CV}$, $\boldsymbol{spikeHalfwidth}$, $\boldsymbol{spikeAssymmetry}$, $\boldsymbol{rippleEnvelope}$, $\boldsymbol{x}_{\mathbf{Theta}}$ and $\boldsymbol{y}_{\mathbf{Theta}}$)[15]. Using multi-dimensional scaling (Kruskal, 1964; Mugavin, 2008), t-SNE (Maaten and Hinton, 2008), spectral embedding (Belkin and Niyogi, 2003) and isometric mapping (Tenenbaum et al., 2000)—all implemented in the Python module *Scikit-learn* (Pedregosa et al., 2011), it is shown that the identified SWR-response clusters also tend to group together based on this set of other measures (**Figure 5.10**).

5.5) Interneuron classes identified from extracellular recordings

Motivated by and building upon the results of the previous section, the goal of this section is to cluster hippocampal interneurons using any of the characteristics that can be extracted from

[15] Before the application of the manifold learning techniques listed in the following sentence, most of these measures were normalized by *z*-scoring. However, similar as in the next section, $\boldsymbol{x}_{\mathbf{Theta}}$ and $\boldsymbol{y}_{\mathbf{Theta}}$ were instead normalized by dividing $\boldsymbol{r}_{\mathbf{Theta}}$ by its standard deviation before the conversion from Polar to Cartesian coordinates.

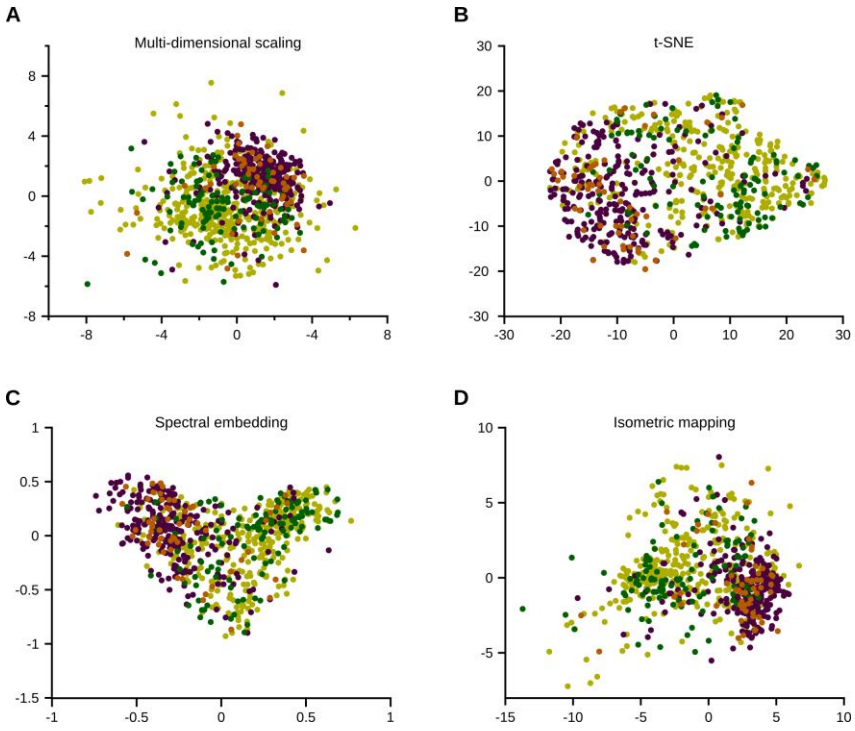

Figure 5.10 *Manifold learning reveals that interneuron clusters identified from SWR-response measures also tend to group together based on other characteristics.*

The manifold learning techniques applied to the interneurons (n = 757) are (A) multi-dimensional scaling, (B) t-SNE, (C) spectral embedding and (D) isometric mapping. The interneurons are colour coded according to which SWR-response cluster they were assigned to (same colour scheme as in **Figure 5.9**). Importantly, these manifold learning techniques are performed on measures that were not used for determining these clusters (see text for the list of included measures).

extracellular recordings. In addition to the seven SWR-response measures already used in the previous section, the measures $\boldsymbol{firingRate}$, $\boldsymbol{burstIndex}$, $\boldsymbol{refractoryPeriod}$, $\boldsymbol{CV}$, $\boldsymbol{spikeHalfwidth}$, $\boldsymbol{spikeAssymmetry}$, $\boldsymbol{rippleEnvelope}$, $\boldsymbol{x}_{\mathbf{Theta}}$ and $\boldsymbol{y}_{\mathbf{Theta}}$ are selected as input features for the clustering. Before clustering, all measures were again normalized by z-scoring, with the exception of the phase coupling measures $\boldsymbol{x}_{\mathbf{Theta}}$ and $\boldsymbol{y}_{\mathbf{Theta}}$. The scale of these measures was instead normalized by dividing $\boldsymbol{r}_{\mathbf{Theta}}$ by its standard deviation before the conversion from Polar to Cartesian coordinates took place. The selection of these measures was informed by the results of the previous section, with the restriction that measures that are too similar should not both be selected (e.g., that is why only $\boldsymbol{burstIndex}$ and not also $\boldsymbol{meanAuto}$ is included).

Clustering method / number of clusters

After selecting the input features on which to perform the clustering, the next questions are (1) which clustering algorithm to use and (2) how many clusters to look for. These questions are not independent, which is the reason they are discussed in the same paragraph.

The most common clustering method, which was already used in the previous two sections, is the *k*-means clustering algorithm (Lloyd, 1982; Friedman et al., 2001). This algorithm aims to partition a data set into a before-hand selected number of clusters K, in such a way that the total distance between points assigned to the same cluster is minimized. This minimization criterion, referred to as the *within-cluster dissimilarity*, is given by:

$$W_K(C) = \frac{1}{2}\sum_{k=1}^{K}\sum_{C(i)=k}\sum_{C(i')=k} d(x_i, x_{i'})$$

whereby $C(i)$ is a "cluster assignment rule" that returns for each observation i its assigned cluster and $d(x_i, x_{i'})$ is a "distance function" that returns for any two data points x_i and $x_{i'}$ the distance between them. The *k*-means algorithm uses squared Euclidean distance as distance metric (i.e., $d(x_i, x_{i'}) = \sum_{j=1}^{p}(x_{ij} - x_{i'j})^2 = \|x_i - x_{i'}\|^2$ with p the number of included features and x_{ij} the value of the j^{th} feature of observation i). To find the cluster assignment rule $C(i)$ that minimizes the within-cluster dissimilarity, *k*-means uses an iterative descent algorithm. This algorithm starts by randomly selecting K data points as "centroids" of the K clusters, after which the following two steps are repeated until convergence is achieved: (1) each data point is assigned to the cluster of its nearest centroid; and (2) the centroid of each cluster is updated by taking the mean of all data points assigned to that cluster.

For *k*-means clustering, the number of clusters K into which to divide the data set needs to be chosen before-hand. A classical way to do this is by the so-called "elbow-method" (Thorndike, 1953; Friedman et al., 2001). For this method, the optimal within-cluster dissimilarity [i.e., $W_K^* = \min_C W_K(C)$] is found and its logarithm plotted for values of K ranging from 1 to $K_{\max}$, which is here set to 15 (**Figure 5.11A**, red curve). As the optimal within-cluster

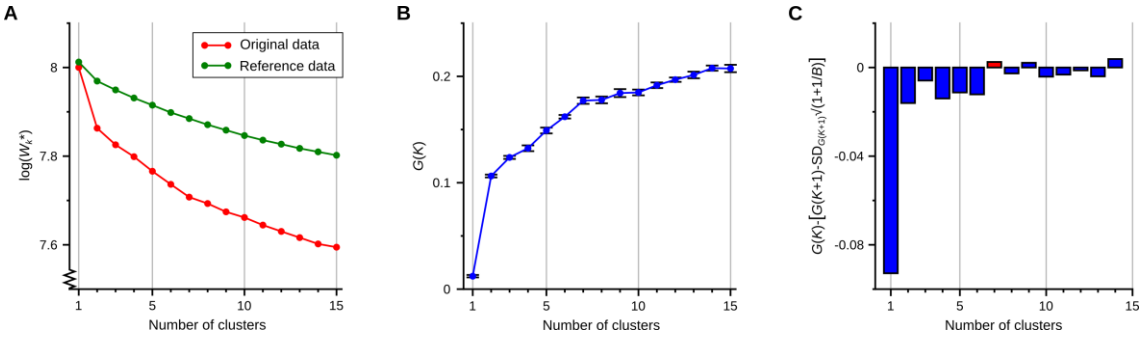

Figure 5.11 *Estimating the number of clusters using the gap statistic.*

(A) For *k*-means clustering, the logarithm of the optimal within-cluster dissimilarity is plotted as a function of the number of clusters, both for the original data set (red) and averaged over ten shuffled reference data sets (green).
(B) The gap statistic $G(K)$, which is equal to the difference between the two curves in (A), is plotted as function of the number of clusters. The error bars denote the standard deviation of $G(K)$ over the different reference data sets.
(C) According to the rule that selects the smallest K for which $G(K)$ is within one (sample size corrected) standard deviation from $G(K+1)$, seven is selected as the optimal number of clusters.

dissimilarity always decreases with an increasing number of clusters, it does not work to simply select the K that minimizes W_K^*. Instead, the optimal number of clusters is classically chosen as the value of K for which the graph of $\log(W_K^*)$ exhibits a "kink" or "elbow" (i.e., the value of K after which the decline of W_K^* slows down). However, as is more often the case in practice, the outcome of this heuristic is ambiguous: there seems to be a kink at two clusters, but also at seven and nine, and none of them are very convincing (**Figure 5.11A**, red curve). For situations like this, Tibshirani et al. (2001) proposed a formal procedure that is essentially an automated way of finding the kink. Their idea is to compare $\log(W_K^*)$ with its expected value under an appropriate reference distribution in which no clustering is present, and to select the number of clusters for which the gap between these two values is the largest. Tibshirani et al. (2001) suggested to use data uniformly distributed over a rectangle containing the original data as reference, but here I instead generate the reference data by randomly shuffling each input feature independently. This way of generating reference data also removes any clustering present in the data, and it has the advantage that the structure of the reference data is more similar to the original data than the structure of uniformly distributed reference data would be. The green curve in **Figure 5.11A** is the average optimal within-cluster dissimilarity based on

$B = 10$ such reference data sets: $\overline{\log(W^*_{K,B})} = \frac{1}{B}\sum_{b=1}^{B}\log(W^*_{K,b})$, whereby $\log(W^*_{K,b})$ is the value of $\log(W^*_K)$ for reference data set b. The difference between $\log(W^*_K)$ of the original data and $\overline{\log(W^*_{K,B})}$ is called the *gap statistic* $G(K)$, which is plotted in **Figure 5.11B** with the error bars denoting its standard deviation $SD_{G(K)} = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\log(W^*_{K,b}) - \overline{\log(W^*_{K,B})}\right)^2}$ (see Tibshirani et al., 2001). The number of clusters is then chosen as the smallest value K such that $G(K)$ is greater or equal to $G(K+1) - SD_{G(K+1)}\sqrt{1+1/B}$, a rule that has empirically been found to work well (Tibshirani et al., 2001; Friedman et al., 2001). For the here considered data set of 757 interneurons, this rule selects an optimal number of clusters of seven, as can be seen from **Figure 5.11C** where the histogram of $G(K) - \left[G(K+1) - SD_{G(K+1)}\sqrt{1+1/B}\right]$ is plotted. It needs to be stressed, however, that this estimate for the number of clusters in this particular data set is not very robust. For example, it sometimes changes drastically upon adding or removing a single input feature; or when using a different clustering method. Using agglomerative hierarchical clustering (see below), the optimal number of clusters estimated by the gap statistic is nine (not shown).

Another common way to select the number of clusters is based on the *silhouette coefficient* (Rousseeuw, 1987). For each observation, a silhouette score is calculated that indicates how well the selected clustering is suitable for that particular observation. Let observation i be assigned to cluster A. Then define $a(i)$ as the average distance from observation i to all other observations assigned to cluster A, $d(i, C)$ as the average distance from observation i to all observations assigned to cluster C, and $b(i)$ as the minimum of $d(i, C)$ over all clusters other than cluster A. The silhouette score of observation i is then:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

This score approaches 1 if observation i is similar to the other observations in its own cluster and dissimilar to observations in any other cluster, while a score below 0 indicates that

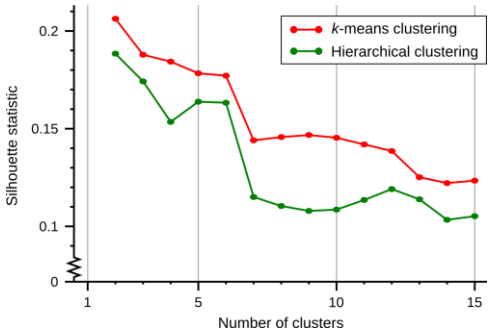

Figure 5.12 *Estimating the number of clusters using the silhouette statistic.*

The silhouette coefficient is plotted as a function of the number of clusters, both for *k*-means (red) and for agglomerative hierarchical clustering (green). For both clustering methods, the silhouette coefficient is maximized with two clusters.

observation *i* is actually more similar to a cluster other than it was assigned to. The silhouette coefficient is then simply the average silhouette score over all observations, and the optimal number of clusters is chosen as the number of clusters that maximizes this coefficient. Both with *k*-means and with agglomerative hierarchical clustering (see below), the optimal number of interneuron clusters selected by the silhouette coefficient is two (**Figure 5.12**).

Many more heuristics or indices have been proposed for determining the relevant number of clusters in a data set. Using the R package *NbClust* (Charrad et al., 2014), I executed an additional 24 methods for finding the number of clusters in the interneuron data set. Half of these methods suggest 3 as the best number of clusters, four methods suggest 2 clusters, three methods suggest 12 clusters, two methods suggest 11 clusters, and the last three methods suggest respectively 6, 14 and 15 clusters. These results are interpreted as indicating that there is not an unequivocal "best" number of clusters for this data set, and based on a visual assessment of the resulting cluster combinations it is decided to select *k*-means clustering with six clusters.

A popular alternative clustering method to *k*-means is hierarchical clustering. Advantage of hierarchical clustering is that it produces a hierarchical representation of clusters that can be visualized with a tree, which is called a *dendogram* (**Figure 5.13**). There exist different approaches (e.g., agglomerative/bottom-up and divisive/top-down) and different linkage strategies (e.g., average, complete, Ward) for constructing this hierarchy, and the resulting dendogram can be "cut" according to various criteria to arrive at different number of resulting clusters (Tibshirani et al., 2001; pp.520-528). Here, the dendogram produced by agglomerative clustering with Ward linkage is compared with the result of *k*-means clustering

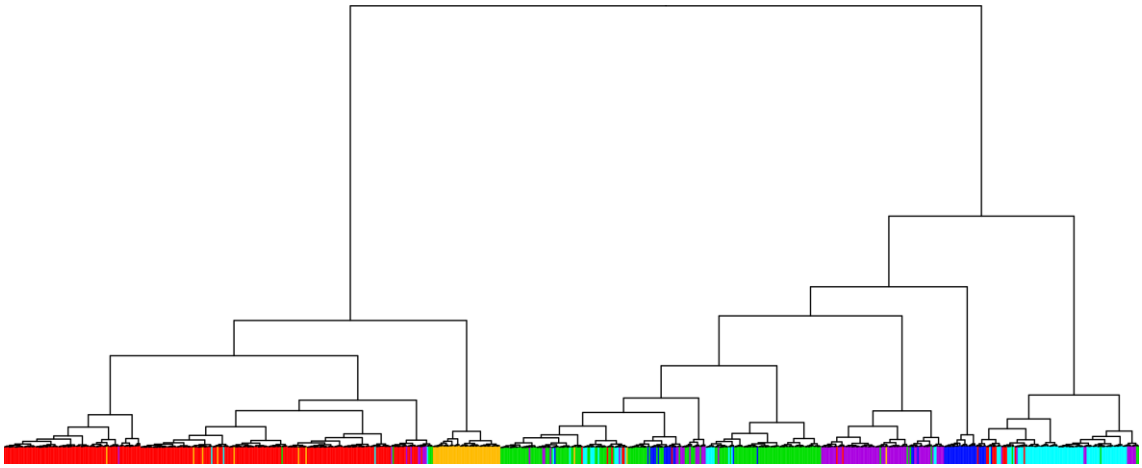

Figure 5.13 *Comparison of k-means and hierarchical clustering.*

Illustrated is the dendogram produced by agglomerative hierarchical clustering with Ward linkage, whereby individual interneurons are coloured according to the cluster they are assigned to by *k*-means clustering with six clusters (same colour code as in **Figure 5.14**). This tree plot was produced in Python using the ETE 3 toolkit (Huerta-Cepas et al., 2016).

with six clusters (**Figure 5.13**). Although there are some differences between the clustering produced by these different methods, that they are mostly similar gives confidence in the selected clustering model.

Characteristics of the identified interneuron classes

The interneuron classes found by *k*-means clustering with six clusters are illustrated in **Figure 5.14**. These six interneuron classes indeed differ in their SWR-response, their firing pattern, their spike waveform, their phase coupling and in the characteristics of their recording tetrode. However, it should be stressed that based on these firing properties, the interneurons are not as clearly sortable into distinct clusters as has been described for the sorting of interneurons when also their post-synaptic targets and their expression of molecular markers are known (Klausberger and Somogyi, 2008; Somogyi et al., 2014). It thus seems that these in anatomical terms distinct interneuron types have overlapping firing properties. Nevertheless, although the here presented interneuron classification based on firing properties does not find such clearly distinct clusters, the identified interneuron classes likely do correspond to these underlying distinct anatomical types to a certain degree. Indeed, the strong coupling to SWRs and to the

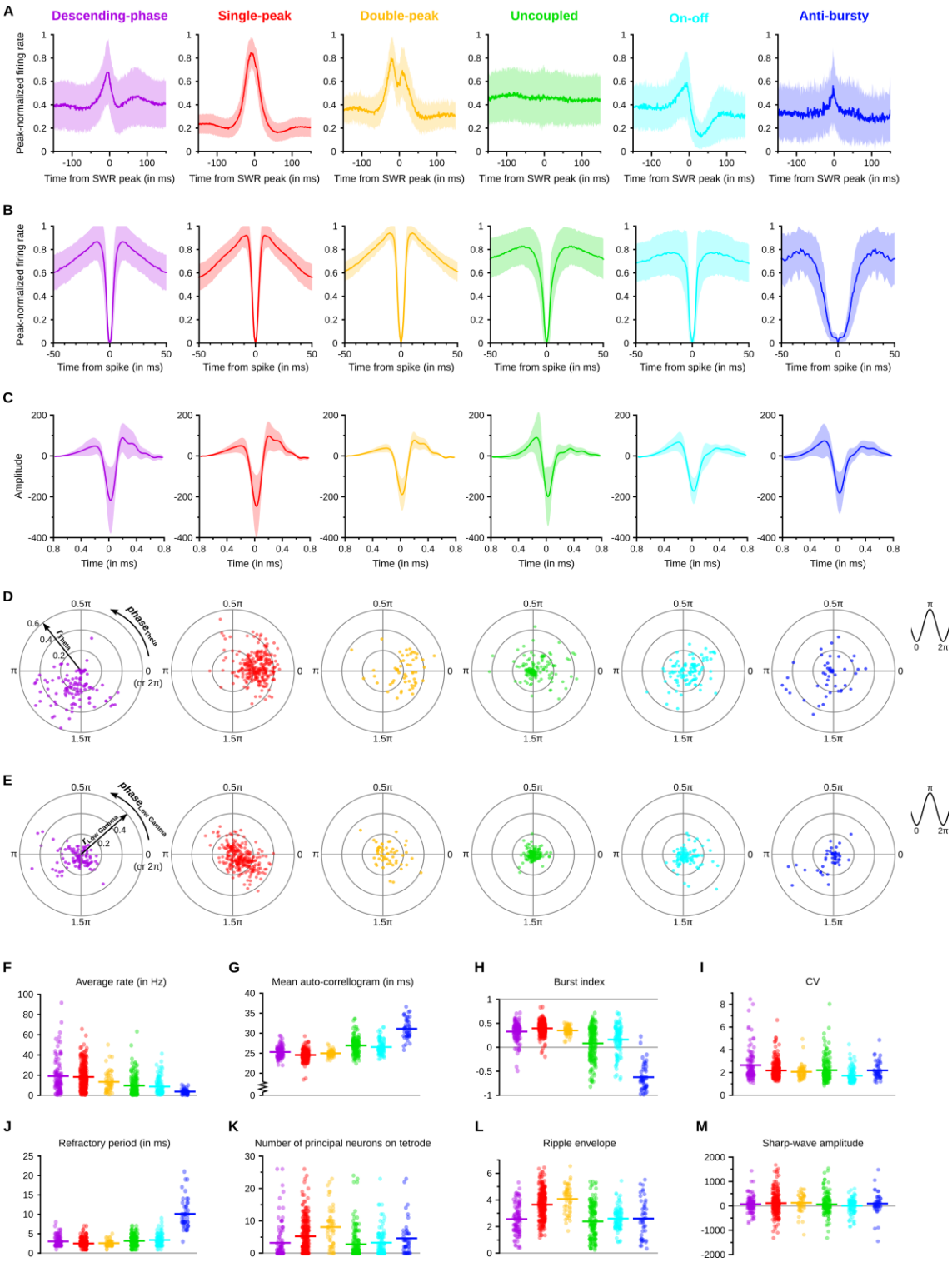

Figure 5.14 *Interneuron clusters identified based on characteristics extracted from extracellular recordings.*

Using *k*-means clustering on 16 manually crafted features (listed in the text), 757 putative interneurons are divided into six clusters. The clusters are ordered according to their average firing rate. The different interneuron clusters can be characterized by a theta coupling to the early descending phase (purple, "descending-phase"; n = 107 neurons), a single SWR peak and coupling to the trough of theta (red, "single-peak"; n = 280), a double SWR peak and coupling to the trough of theta (orange, "double-peak"; n = 55), no coupling to SWR or theta (green, "uncoupled"; n = 163), an increase in rate during SWRs

followed by a decrease (cyan, "on-off"; $n = 105$) and a low rate with high refractory period and no burstiness (blue, "anti-bursty"; $n = 47$). Shown are each cluster's (A) peak-normalized SWR-response (mean ± SD); (B) peak-normalized auto-corrollogram (mean ± SD); (C) spike waveform (mean ± SD); (D) coupling to theta phase; (E) coupling to low gamma phase; (F) average firing rate; (G) mean of the auto-correllogram; (H) burst-index; (I) coefficient of variation (CV); (J) refractory period; (K) number of principal neurons recorded on the same tetrode; (L) ripple envelope; (M) sharp-wave amplitude.
*[As in **Figure 5.9**, the 23 putative interneurons recorded from the animal that was recorded on the system with a different amplification are left out in the plot of the spike waveforms (C) and in the plot of the sharp-wave amplitudes (M).]*

through of theta cycles makes that both the single-peak and the double-peak interneuron class likely correspond to parvalbumin basket cells and/or to bistratified cells (Lapray et al., 2012; Varga et al., 2012, 2014; Katona et al., 2014). An open question is whether one of these two classes corresponds to the parvalbumin basket cells and the other to the bistratified cells, or whether both classes contain a mixture of interneurons from either type. Surprisingly, I could not find any description of a "double peak" in the SWR-response of interneurons in any of the studies using juxtacellular labelling and recording. The only report I could find that describes something similar is a paper from Csicsvari and colleagues (1999b). The characteristic theta phase coupling of the descending-phase interneuron class, in combination with their not very pronounced SWR-reponse, suggests that (some of) these interneurons are axo-axonic cells (Viney et al., 2013; Varga et al., 2014). The remaining three interneuron classes identified here are harder to match to anatomically defined interneuron types.

5.6) Interneuron coupling to principal cell assembly patterns

Having identified interneuron classes based on extracellular recordings, it is now possible to return to the initial question raised in this chapter's introduction (see **Figure 5.1**). Are there differences in how different types of interneurons couple to principal cell assembly patterns? Such differences could be reflected in that some types of interneurons might couple more strongly to assembly patterns in general, but also in that some interneuron types might differentiate more between different assembly patterns (i.e., having strong coupling to one

pattern, but not to another). Finally, it is possible that interneuron types differ in the extent to which such differential coupling with regard to assembly patterns is maintained over time (i.e.,

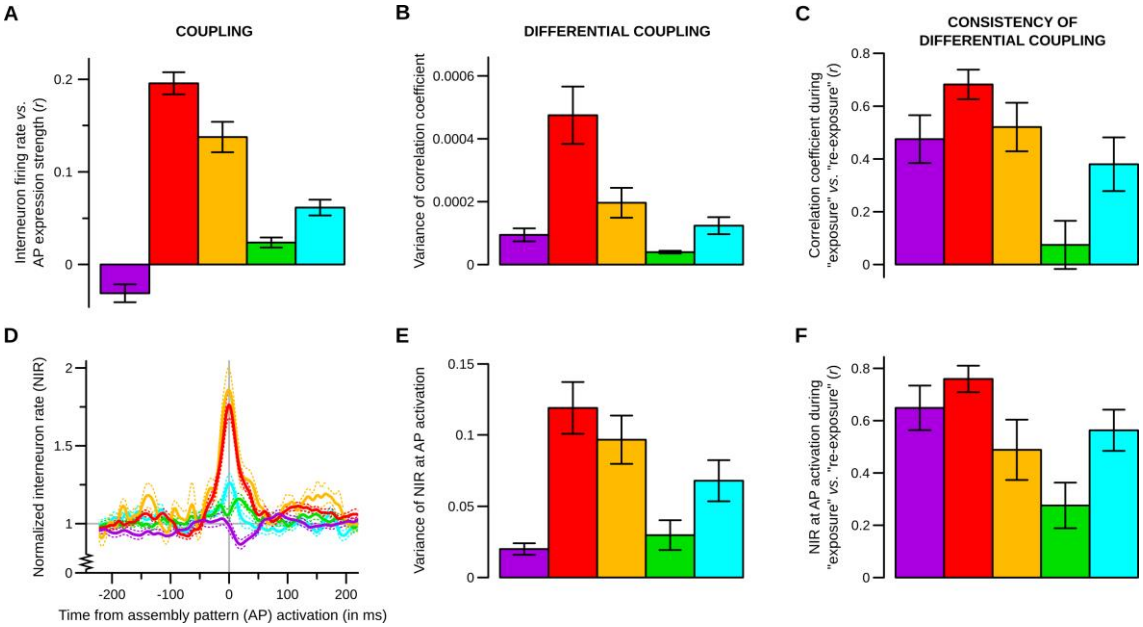

Figure 5.15 *Coupling of interneurons to principal cell assembly patterns.*

Using two different measures, this figure shows for interneurons of each class how they couple to principal cell assembly patterns ("COUPLING"), how much they differentiate between different assembly patterns ("DIFFERENTIAL COUPLING"), and to what extent such differential coupling is preserved between a first exposure to an enclosure and later re-exposure ("CONSISTENCY OF DIFFERENTIAL COUPLING").
The first measure is based on the correlation between an interneuron's firing rate and the expression strength of an assembly pattern calculated in 25 ms time bins. Depicted are (A) the mean (± SEM) of these correlation coefficients over all interneuron-assembly pattern pairs, (B) the variance—per interneuron and per "exposure" session (mean ± SEM)—of these correlation coefficients over the in that session expressed assembly patterns, and (C) the correlation—per interneuron and per "exposure"-"re-exposure" session-pair (mean ± SEM)—of the correlation coefficients with the assembly patterns expressed during "exposure" versus the correlation coefficients with the expression of those same assembly patterns during "re-exposure".
The second measure is based on the normalized interneuron firing rate at the time of assembly pattern activations. Depicted are (D) the normalized interneuron firing rate (mean ± SEM) in a window [-220 ms, 220 ms] around the assembly pattern activations, (E) the variance—per interneuron and per "exposure" session (mean ± SEM)—of the normalized interneuron firing rate at the time of assembly pattern activation over the in that session expressed patterns, and (F) the correlation—per interneuron and per "exposure"-"re-exposure" session-pair (mean ± SEM)—of the normalized interneuron firing rate at the time of assembly pattern activations of the patterns expressed during "exposure" versus the normalized interneuron firing rate at the time of activations of those same patterns during "re-exposure".
This figure is based on the same data set used in chapter 4, and includes 13 "descending-phase" interneurons (purple), 26 "single-peak" interneurons (red), 8 "double-peak" interneurons (orange), 19 "uncoupled" interneurons (green) and 19 "on-off" interneurons (cyan). There were no "anti-bursty" interneurons in this data set. For (A) and (D), each interneuron-assembly pattern pair during an "exposure" session provides one data point (descending phase: n = 183; single-peak: n = 348; double-peak: n = 119; uncoupled: n = 219; on-off: n = 200). For (B), (C), (E) and (F), each interneuron provides one data point per "exposure"-"re-exposure" session-pair with at least three assembly patterns identified during the exposure session (descending phase: n = 29; single-peak: n = 56; double-peak: n = 21; uncoupled: n = 38; on-off: n = 37).

if an interneuron during a particular session couples strongly to pattern A and weakly to pattern B, to what extent is that still the case during a later session?).

To investigate these questions in a systematic way is not straight forward. Here, using the interneuron classification scheme of the previous section and the data set described in the previous three chapters, a tentative first attempt is made to address these questions (see **Figure 5.15**). It is found that the single-peak and double-peak interneuron classes couple most strongly to principal cell assembly patterns in general (**Figure 5.15A,D**). This likely relates to that these interneuron classes are the ones that couple most strongly to SWRs. Of special note is that the double-peak interneuron class does not show a double peak in their coupling with assembly pattern activations (**Figure 5.15D**, orange trace). The single-peak and double-peak interneuron classes are also found to have the highest differential coupling between assembly patterns (**Figure 5.15B,E**). This result should however be interpreted with caution. As the differential coupling is measured by the variance of the coupling-strength over simultaneously expressed assembly patterns, this result might simply reflect that the average coupling-strength of these two classes is the highest. Finally, for the consistency of the differential coupling from the exposure to the re-exposure session, less clear differences between the interneuron classes are found (**Figure 5.15C,F**). Only noteworthy result is that the "uncoupled" interneuron class is less consistent in its differential coupling than the other interneuron classes, further suggesting that this class of interneurons is unrelated to the rest of the hippocampal circuit.

5.7) Discussion

In this chapter, a framework for the characterization of extracellular recorded neurons has been described, which I applied to the full database in the lab of recordings from the dorsal CA1 region of the mouse hippocampus. Although I did not find clear support for the possibility to identify discrete types of interneurons solely based on their extracellular recordings, I did find structure in this data set indicative of clusters of interneurons with partially overlapping firing

properties. To be more specific, I found that the shape of an interneuron's SWR-response predicts properties such as its firing rate, its firing pattern and its theta phase coupling. I then performed *k*-means clustering on manually crafted features representing these firing properties to identify six interneuron classes, some of which could cautiously be matched to anatomically defined interneuron types. A tentative attempt was made to analyse the coupling of these different interneuron classes to principal cell assembly patterns. I suggest that the here presented framework for an unsupervised interneuron clustering solely based on extracellular recordings, although not absolute, nevertheless provides a useful way of classifying hippocampal interneurons that could contribute to further the understanding of their diverse roles in network dynamics and behaviour.

Automated feature extraction

In this thesis, all features used for the clustering were manually crafted from the neurons' SWR-response, auto-correllogram, ISI distribution, spike waveform and phase coupling. Besides that this was a time-consuming task, the selected features are subjective and might not be optimal. A potential refinement of the here presented interneuron classification framework would therefore be to replace the manual crafting of features by automated feature extraction.

One way to automatically extract features from high-dimensional data would be to use linear dimensionality reduction techniques such as principal component analysis or independent component analysis. However, disadvantage of these methods is that they do not take into account the temporally ordered structure of, for example, the neurons' SWR-response, auto-correllogram or spike waveform. An alternative way to automatically extract features, which takes into account such temporal structure, would be to use a wavelet transform in combination with a criterion deciding which wavelet coefficients to select as features (e.g., Xu and Kezunovic, 2002; Quiroga et al., 2004).

Silencing parvalbumin-expressing interneurons

As a side-project during my DPhil studies, together with Claire Bratley, I performed tetrode recordings from the dorsal CA1 hippocampus of freely-behaving mice while transiently silencing (both SWR-triggered and random) parvalbumin(PV)-expressing neurons using optogenetics (22 recording days, 3 PV-Cre mice bilaterally injected with a CAG-FLEX-ArchT-GFP viral vector in dorsal CA1). The aim of this project was two-fold: (1) "optotagging" of PV-expressing interneurons to assist in mapping the above described interneuron classes to anatomically defined interneuron types and (2) characterizing the effect of temporarily de-activating PV-expressing interneurons to elucidate the role of this broad group of interneurons (containing multiple anatomically defined types, see for example Somogyi et al., 2014).

Unfortunately, the cell yield in these animals was low. From the 9 recording days that were worth pre-processing, 7 putative interneurons were identified (out of 107 recorded CA1 neurons), all of which had an average firing rate below 4 Hz. One of these interneurons was assigned to the "descending-phase" cluster, two to the "uncoupled" cluster and the remaining four to the "on-off" cluster. None of the in these animals recorded neurons was found to be directly silenced by the optogenetic feedback. That the optogenetic silencing was nevertheless working—at least to some extent, was indicated by the observation of a negative deflection in the recorded LFP in response to light pulses (**Figure 5.16A**) and by a small increase in the

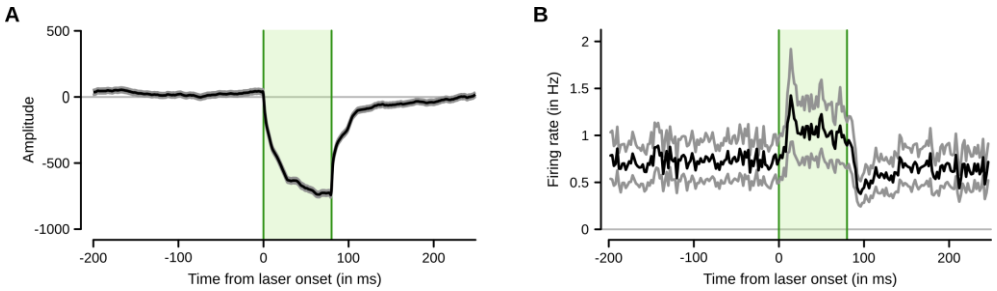

Figure 5.16 *Silencing PV-expressing interneurons.*

(A) Example of a CA1 tetrode's average (± 2 SEM) LFP triggered by randomly delivered 80 ms light pulses (n = 1999) aimed at silencing PV-expressing interneurons.
(B) Average (± 2 SEM) firing rate of putative CA1 principal neurons (n = 84) in response to these randomly delivered optogenetic pulses.

average firing rate of the principal neurons (**Figure 5.16B**). Of the interneurons, only two—both of which were assigned to the "on-off" cluster—had a clear response of their firing rate to optogenetic silencing, with both transiently increasing their rate upon randomly delivered light pulses.

Computational modelling

Another complementary approach to furthering the understanding of the diverse roles of interneurons is computational modelling. I explored this approach during a summer school on computational neuroscience at the Okinawa Institute of Science and Technology (Japan). I adopted a detailed, biophysically motivated computational model of the hippocampal CA1 microcircuit containing pyramidal neurons and four types of inhibitory interneurons (basket, axo-axonic, bistratified and oriens lacumosum-moleculare cells), which was developed by Cutsuridis and colleagues (2010; **Figure 5.17A**). During the summer school, I wrote a Python-wrapper for the freely available source code of this model, which was written for the NEURON simulation environment (Hines and Carnevale, 1997). This allowed me to simulate the spiking activity of a population of CA1 pyramidal neurons while feeding different sets of external stimuli to the network (**Figure 5.17B**). I then detected and tracked cell assembly patterns in the resulting CA1 pyramidal neuron spike trains (**Figure 5.17C**). The detected assembly patterns closely correspond to the different external inputs, but rarely any assembly pattern expression could be detected during periods without external patterned input. This absence of self-organised assembly dynamics indicates that this computational model in its current form might not be useful for studying reactivation.

I had also hoped to be able to use this model to study the role of hippocampal interneurons by in- or excluding particular types from the model and observing how the behaviour of the network changes. However, I found that this model was too small or simple to identify functions of interneurons other than those that were purposefully programmed into

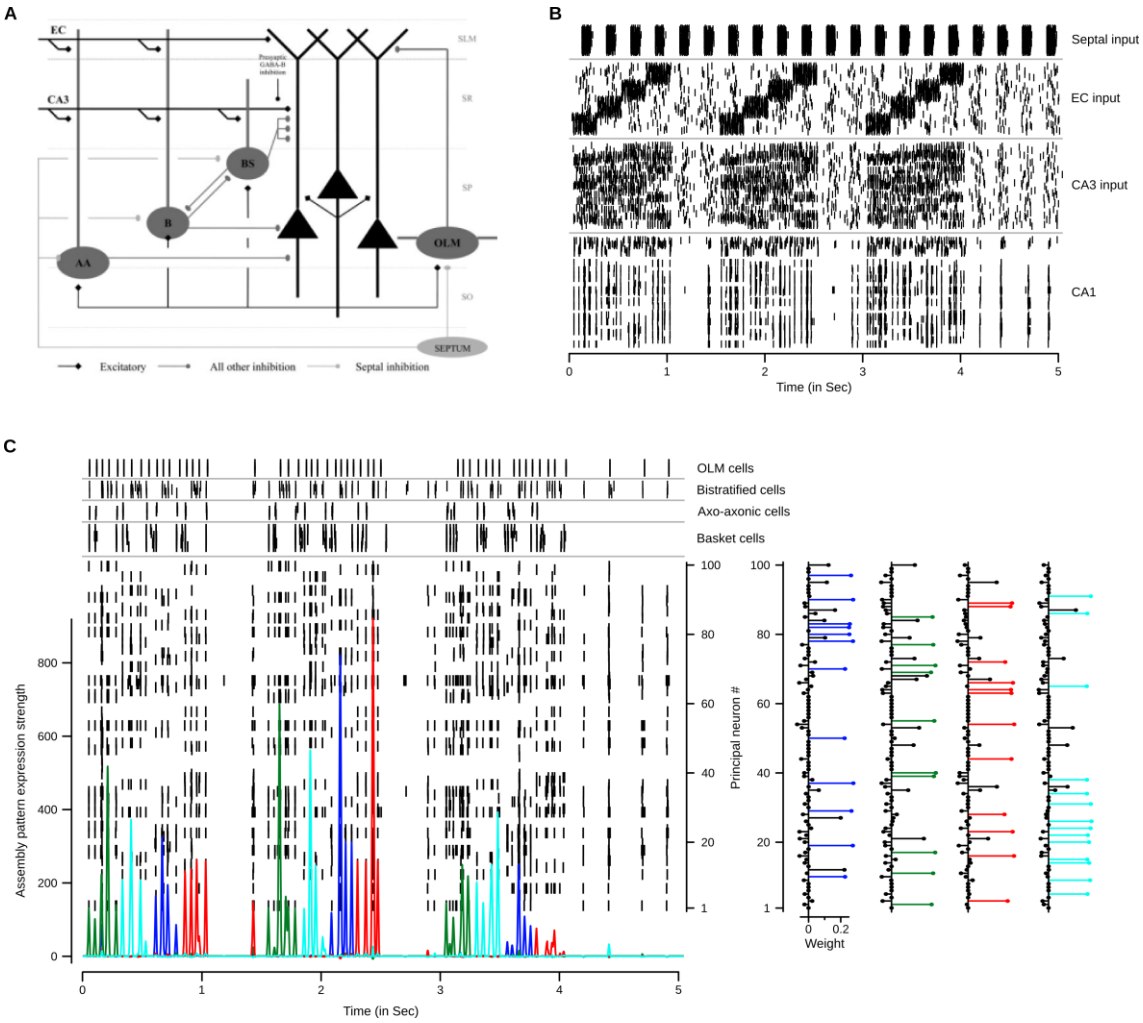

Figure 5.17 *Simulated hippocampal network activity with various types of interneurons.*

(A) Schematic of the cell types and their connectivity in the computational model of the hippocampal CA1 microcircuit. Reproduced from Cutsuridis et al. (2010).
(B) Results of a simulation with 115 CA1 neurons [see (C)] and 210 input units [30 inhibitory septal inputs, 80 excitatory entorhinal cortex (EC) inputs and 100 excitatory CA3 inputs], whereby four different input patterns are each repeated three times.
(C) Close-up of the spiking activity of the CA1 neurons [100 principal neurons, 6 basket cells (B), 3 axo-axonic cells (AA), 3 bistratified cells (BS) and 3 oriens lacunosum-moleculare cells (OLM)], together with the weight vectors (right) and temporal expression strength (bottom) of the detected principal cell assembly patterns.

Abbreviations: SLM, *stratum lacunosum moleculare*; SR, *stratum radiatum*; SP, *stratum pyramidale*; SO, *stratum oriens*.

them. I suggest that instead of this "bottom-up" approach of computational neuroscience, a more fruitful tactic for gaining insights into the roles of different types of interneurons might be a "top-down" approach to computational neuroscience (see section 6.4).

CHAPTER 6: GENERAL DISCUSSION

The main aim of this DPhil thesis has been to investigate network-level processes in the brain underlying memory consolidation. By using a mouse model of declarative memory, it was possible to combine *in vivo* multi-unit extracellular recordings from the dorsal CA1 region with closed-loop optogenetic feedback. I used these techniques to test and confirm the long-standing hypothesis in the field of memory consolidation that the stability of newly formed, hippocampal cell assembly patterns depends on their offline reactivation. As a by-product of this process, I validated a recently proposed statistical assembly detection method by demonstrating that the patterns it detects in recordings of hippocampal principal neurons indeed group together co-active neurons and that these patterns are "memory-representing" in the sense that their expression is spatially selective and environment specific. This result opens up the possibility for this promising analysis tool to be used with data from other brain regions and/or species.

Another interesting, somewhat surprising finding was the functional diversity in spatially selective cell assembly patterns expressed in the mouse hippocampus. Unexpectedly, offline reactivation seemed to be only important for the stability of a subset of the patterns expressed in a novel environment. On the one hand there were "Hebbian-like" assembly patterns whose strength gradually increased throughout their first expression and whose later stability was related to their offline reactivation; on the other hand there were "reactivation independent" assembly patterns whose expression rapidly stabilized upon their first expression and whose later stability appeared to be unrelated to their offline reactivation.

Finally, I started a project aimed at classifying hippocampal interneurons based on extracellular recordings. Although no evidence was found for the existence of discrete classes of interneurons based on their firing properties alone, the resulting unsupervised clustering framework nevertheless promises to be a useful analytical tool.

6.1) Does optogenetic silencing disrupt reactivation selectively?

The interpretation of the results presented in this thesis as direct evidence for a causal role of offline reactivation in the stabilization of hippocampal cell assembly patterns, critically depends on the assumption that reactivation was *selectively* disrupted by optogenetic SWR silencing. The main argument for this assumption is that reactivation peaks during SWRs (see **Figure 4.2C**; Wilson and McNaughton, 1994; O'Neill et al., 2008), and indeed also around the on-the-fly detections of SWRs that were used to trigger the optogenetic light pulses (see ***Appendix A***). Note that this same rationale was previously used for a similar interpretation of the results of studies using electrical SWR disruption (Girardeau et al., 2009; Ego-Stengel and Wilson, 2010). However, as discussed in section 4.5, the electrical stimulation used by these studies might have disrupted more than just the reactivation. Indeed, an important implicit assumption made so far in this thesis is that optogenetic silencing is a "clean" way to transiently inhibit neuronal activity.

Although this attitude does not seem to be uncommon in the Neuroscience literature (Adamantidis et al., 2015), my observations indicating a potential "network response" to first-time application of repeated optogenetic silencing (see **Figure 2.2**) shed some doubt on this assumption. These observations suggest that repeated optogenetic silencing—besides transiently inhibiting neuronal firing—might have additional, longer lasting effects on the hippocampal network, which, interestingly, might be more pronounced the first time optogenetic pulses are delivered. Unfortunately, my experiments were not designed to investigate the effect of first-time repeated optogenetic silencing (e.g., some mice received a varying number of undocumented "test" pulses before their first experiment), which means that more research is needed to formally test for this phenomenon. If first-time repeated optogenetic silencing indeed has such a marked effect on the network, this would have profound consequences for the design and interpretation of optogenetic experiments. Moreover, such a finding could provide a basis

for gaining new insights into the plastic properties of brain networks (i.e., how does the network re-arrange so that subsequent repeated optogenetic silencing no longer triggers the response?).

Taking this into account, let me get back to the interpretation of the SWR silencing results. Firstly, regardless of whether or not optogenetic silencing is a "cleaner" way to disrupt hippocampal reactivation compared to electrical stimulation, the side effects of both approaches are likely different. Already for this reason, the in this thesis presented results are an important incremental contribution to the growing evidence for a causal role of offline reactivation in memory consolidation. Moreover, the main concern with the previous electrical SWR disruption studies (Girardeau et al., 2009; Ego-Stengel and Wilson, 2010) was that the by them demonstrated memory impairments could have been caused by generic effects (e.g., induced plasticity) of their stimulation specifically coupled to SWRs. In this thesis, the familiar environment condition controls for such an interaction between SWRs and the disruptive intervention. That SWR-triggered silencing impairs assembly pattern reinstatement for a novel, but not a familiar, enclosure, convincingly demonstrates that optogenetic SWR silencing disrupts something that specifically contributes to the stabilization of new cell assemblies. Although I believe that offline reactivation is the most likely candidate for this "something"—a believe that is further endorsed by the in this thesis demonstrated correlation between assembly pattern reactivation and subsequent reinstatement; I admit that sceptics could argue that this remains open to some speculation.

Additional speculation about a causal role of SWR reactivation in memory consolidation was provided by a recent study that failed to find impaired stability of newly formed hippocampal spatial maps following optogenetic SWR silencing (Kovács et al., 2016). This result was particularly surprising at first, because the experimental protocol—mice exploring a novel open field enclosure, followed by rest in a sleep box and re-exposure to the same enclosure—was very similar to the one used in this thesis. A major difference, however, is in the size of the data

sets: while the study of Kovács and colleagues was based on 233 principal neurons, the results on SWR silencing reported in this thesis included a total of 4109 principal neurons. Although their small data set is likely the main reason for this study's null result, there are a couple of other differences that could have contributed. Firstly, Kovács and colleagues only used single-neuron and pairwise measures to assess the effect of SWR silencing on the stability of the firing patterns of hippocampal principal neurons. It is possible that these simple measures are less sensitive, and perhaps more noisy, than the cell assembly pattern measures used for the central analyses of this thesis (e.g., compare **Figure 4.6B,C** with **Figure 4.5B**). Secondly, the optogenetic silencing used by Kovács and colleagues was considerably less effective than the one used for this thesis. As indicated by their Figure 4, their optogenetic pulses reduced the firing rate of principal neurons only by about 50%, while the optogenetic protocol employed for this thesis caused a near complete cessation of principal neuron spiking (see **Figure 4.4D,F**). This difference in optogenetic efficacy is also highlighted by the far less pronounced LFP deflections observed in their study (compare their Figure 2 with **Figures 4.4E** and **4.7A**). This difference could relate to that they used Archaerhodopsin instead of the more recently discovered ArchT, or to that that they used a double-transgenic mouse line to express this optogenetic actuator as opposed to targeted injection of a viral construct in a single-transgenic Cre-mouse line. Thirdly, Kovács and colleagues only compared the effect of silencing during SWRs with the effect of silencing after a fixed delay following on-the-fly detected SWRs. This could have contributed to their null result since it is possible that optogenetic silencing not coupled to SWRs also impairs the stability of new hippocampal maps, potentially because—as speculated above—repeated optogenetic silencing could destabilize newly formed cell assemblies by means other than disrupting reactivation. Somewhat consistent with this, I also found a smaller and less pronounced—albeit still significant—effect of SWR silencing when compared with random silencing instead of with no silencing (see **Figure 4.7B**). Note however that the difference in assembly pattern reinstatement after random silencing compared to after

no silencing was not significant, and that a small difference would be expected regardless because optogenetic silencing not coupled to SWRs inevitably disrupts some reactivation.

An experiment to control for effects of optogenetic silencing on the stabilization of new cell assemblies other than through cancelling reactivation, would be to selectively disrupt SWRs only when a particular cell assembly pattern is reactivated during that SWR. The hypothesis is that such an intervention would impair the stability of the selected assembly pattern (and the memory represented by that pattern), but not—or at least substantially less—the stability of other assembly patterns (and their corresponding memories). The main technical challenge with this experiment is to detect on-the-fly which assembly patterns are active during a SWR. Moreover, this detection should ideally be fast enough for the disruption to be applied before the expression of the assembly pattern is finished. (Note that in this thesis the expression of cell assembly patterns is assumed to be within 25 ms time windows!) However, if it is not feasible to have a detection this fast, it might still be possible to disrupt the downstream effect of a pattern's reactivation—i.e., its presumed "transfer" to the cortex. Although in that case the intervention would probably not result in an impaired stability of the hippocampal pattern (but it might, see discussion in the next section), it would be hypothesized to still result in a behavioural memory impairment relating to the mental concept represented by the selected pattern (e.g., a particular reward location). As another approximation to the ideal experiment, it might be more feasible to—instead of a cell assembly pattern—select a single, well isolated place cell and perform optogenetic silencing whenever a spike of this neuron is detected during the sleep following exploration. Because place cells are predominantly reactivated together with peer neurons that have overlapping place fields (Wilson and McNaughton, 1994), such an intervention would be expected to mostly impair the stability of assemblies of place cells with fields around the place field of the selected neuron.

6.2) Dual role for reactivation: local *vs*. systems consolidation

Sometimes a distinction is made between two presumed roles of hippocampal reactivation (e.g., Csicsvari and Dupret, 2014). On the one hand, tracing back to Donald Hebb's proposal of "reverberatory activity" (Hebb, 1949), reactivation is thought to stabilize newly formed cell assemblies within the hippocampus, by providing additional pairings of pre- and post-synaptic activity. On the other hand, with roots in the theoretical work of David Marr (1970, 1971), hippocampal SWR reactivation is thought to have a role in the "transfer"—or perhaps the "transform" (McClelland et al., 1995; Winocur and Moscovitch, 2011)—of memories from the hippocampus to the cortex. These different roles are also reflected in another often made distinction between *local consolidation*, referring to the stabilization of memory traces within a given brain structure, and *systems consolidation*, referring to the reorganization of memory traces between brain structures (e.g., Dudai, 2004; Frankland and Bontempi, 2005). However, while it is indeed typically hypothesized that reactivation—in particular during SWRs—plays an important role in systems consolidation, this appears to be a less common hypothesis for local consolidation. Local consolidation, which is more often called *cellular* or *synaptic* consolidation, is usually thought to involve cellular and molecular processes that strengthen synaptic connections that are involved in the storage of new memories. The possibility that a "network-level" process as offline reactivation could reinforce these local synaptic processes is not often acknowledged.

The finding presented in this thesis—that disrupting reactivation by optogenetic SWR silencing impairs the stability of local hippocampal cell assembly patterns—does however support a role of reactivation in local consolidation. Other recent evidence indicative of such a local role for reactivation was provided by a study reporting that most replay events in the medial entorhinal cortex, which is densely connected with the hippocampus, are unrelated to SWRs or reactivation in the CA1 area (O'Neill et al., 2017). This suggests that many reactivation events are only local. It has to be said, however, that there is ample evidence that—at least statistically—

reactivation in the hippocampus is coordinated with reactivation in various cortical areas (e.g. Ji and Wilson, 2007; Lansink et al., 2009; Ólafsdóttir et al., 2016; Rothschild et al., 2017). Moreover, although the results presented in this thesis support a local role, they of course do not exclude the possibility that reactivation also has more global roles in systems consolidation.

In fact, it is possible that the in this thesis observed local stabilization could have been a "by-product" of processes primarily aimed at systems consolidation. For example, the enhanced stability of local patterns due to SWR reactivation might be mediated by reciprocal connections with the cortex and actually reflect stabilized cortical traces. Another possibility relates to a recent hypothesis (Harris, 2008) stating that the stabilization of changes in input synapses might depend on retroaxonal signals which a neuron receives if it succeeds to excite changes in its output synapses. That is, the changes in downstream target-structures excited by the global role of a pattern's SWR reactivation might in turn assist the stabilization of previous changes in its input-synapses, thereby ensuring the stable reinstatement of the pattern locally. It is thus possible that the local and system-wide functions of reactivation are intertwined.

6.3) Awake SWR replay

In this thesis I have focussed on offline reactivation that takes place during sleep or rest, but it should be mentioned that in the last decade there has been a growing interest in "awake replay"—temporally structured reactivation that takes place during active behavioural states (for reviews see Carr et al., 2011; Roumis and Frank, 2015). Although most prevalent during sleep, SWRs have also been shown to occur during brief periods of inactivity or even during waking exploration (Kudrimoti et al., 1999; O'Neill et al., 2006). During these awake SWRs, temporally compressed sequences of place cells have been detected reflecting paths taken by the animal in both the forward and backward direction (Foster and Wilson, 2006; Diba and Buzsáki, 2007). As these sequences often begin at the animal's current location (Csicsvari et al., 2007) and frequently occur at points were the animal needs to make a choice (Karlsson and Frank,

2009), it has been proposed that awake replay could reflect cued memory retrieval for the purpose of planning and decision making (Carr et al., 2011). In support of this hypothesis, electrical disruption of awake SWRs during a spatial alteration task specifically impaired performance on the working memory component of that task (Jadhav et al., 2012) and awake replay sequences were found to predict upcoming choices (Pfeiffer and Foster, 2013; Singer et al., 2013).

On the other hand, because there are also many similarities between SWRs in the awake and in the sleep state (e.g., high neuronal synchronization, place cell activity at compressed temporal scale, ripple frequency that appears ideal for synaptic plasticity), awake SWRs are also—similarly to sleep SWRs—hypothesized to contribute to memory consolidation by stabilizing recently formed neuronal memory traces (O'Neill et al., 2010; Carr et al., 2011). In fact, because awake SWRs occur closer in time to the events to be remembered and because current sensory experience has been shown to bias the content of awake SWRs (O'Neill et al., 2006), it has been suggested that awake SWRs might be even better suited than sleep SWRs to start consolidating new experience-dependent cell assemblies (Csicsvari and Dupret, 2014). An on-going project led by Lisa Roux from the lab of György Buzsáki is currently testing for such a role by probing the stability of the hippocampal spatial map following disruption of awake SWRs. For their experiment, mice are first trained to learn every day a new set of three rewarded goal locations on a mutli-well "cheeseboard" maze (Dupret et al., 2010). They then use closed-loop optogenetic feedback to silence a subset of the recorded CA1 principal neurons during on-the-fly detected awake SWRs. The preliminary results of this study that have been presented so far (poster B073 at FENS Forum 2016; Eichenbaum et al., 2016; pp.1242-1243) indicate that the place fields of SWR-silenced neurons are less stable than the place fields of control neurons (i.e., neurons that were not silenced or neurons that were silenced after a delay upon awake SWR detection). Besides providing compelling evidence for a causal role of awake SWRs in local consolidation, such a result would strengthen the confidence in the main

conclusion of this thesis by confirming the importance of SWR-associated neuronal activity (either in the awake or sleep/rest state) in the stabilization of newly formed place cell assemblies.

If awake SWRs indeed contribute to the stabilization of new cell assemblies as well, this raises the possibility that the "early stabilized" patterns of section 4.4 might have profited from these awake SWRs for their rapid stabilization. An interesting experiment would therefore be to test whether disrupting awake SWRs would also impair the stability of these early stabilized patterns, whose stability was not impaired by disrupting sleep SWRs. If the stability of these patterns would indeed depend on awake SWRs, it would indicate that the early stabilized patterns are not completely "reactivation independent", but merely that they need less of it.

6.4) Offline reactivation in artificial intelligent agents

The research presented in this thesis is a clear example of "basic science" aimed at improving our understanding of the neural mechanisms in the brain underlying the encoding, consolidation and retrieval of declarative memories. In the long run, it is hoped that knowledge produced by this type of research will inform the development or refinement of therapeutic interventions, in the case of this thesis for example for treating epilepsy or for combatting memory loss in dementia. It should be stressed, however, that these goals are at best long-term perspectives and that the main aim of the research for this thesis has been gaining "basic" knowledge.

There is, however, a growing pressure for such basic research—especially within neuroscience—to have tangible translational applications. Traditionally, it is often attempted to suggest that long-term perspectives, such as those sketched above, are closer or more concrete than they might actually be. Instead, here I would like to highlight another, often overlooked yet important translational direction which I hope my research will contribute towards.

With roots tracing back to the cell assembly theory of Donald Hebb, a class of machine learning models called deep neural networks (DNNs) have a central role in the recent surge in the capabilities of artificial intelligent agents (LeCun et al., 2015). Although originally inspired by connectionist models of the brain, recent progress in DNN models has largely been due to advances in engineering and optimization, achieved in isolation from neuroscience and often involving biologically implausible mechanisms. However, a recent major breakthrough was inspired by contemporary advances in our understanding of how the brain works. By incorporating offline reactivation into state-of-the-art DNN models, researchers at the start-up company DeepMind managed to gain substantial improvements in the performance of these models (Mnih et al., 2015).

This success raises the intriguing possibility that other empirically demonstrated properties of, or computations performed by, the brain could be incorporated into state-of-the-art machine learning models as well. For example, the representation of concepts by co-activation patterns, oscillatory activity and/or different types of interneurons could all be candidates for such an endeavour. Besides to gains in performance of artificial intelligence agents, such an approach might also lead to a better understanding of the functions of these empirical properties of the brain.

APPENDIX A

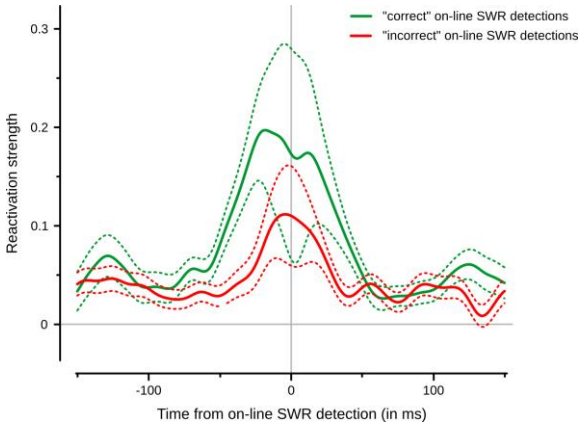

Figure A.1 *On-line SWR detections signal periods of increased assembly pattern reactivation, whether they correspond to offline SWR identifications or not.*

Assembly pattern reactivation is emphasised around on-line SWR detections that correspond to offline identified SWRs (green), but also around on-line SWR detections that do not have a corresponding offline identified SWR (red). Assembly patterns detected in the familiar enclosure (n = 108 assembly patterns) and those detected in a novel enclosure (n = 139) were pooled together for this figure. Solid lines represent mean difference in expression strength between sleep/rest after and sleep/rest before, dashed lines ±1 SEM. Both the mean- and SEM-traces are smoothed with a Gaussian kernel with SD = 4 ms.

APPENDIX B

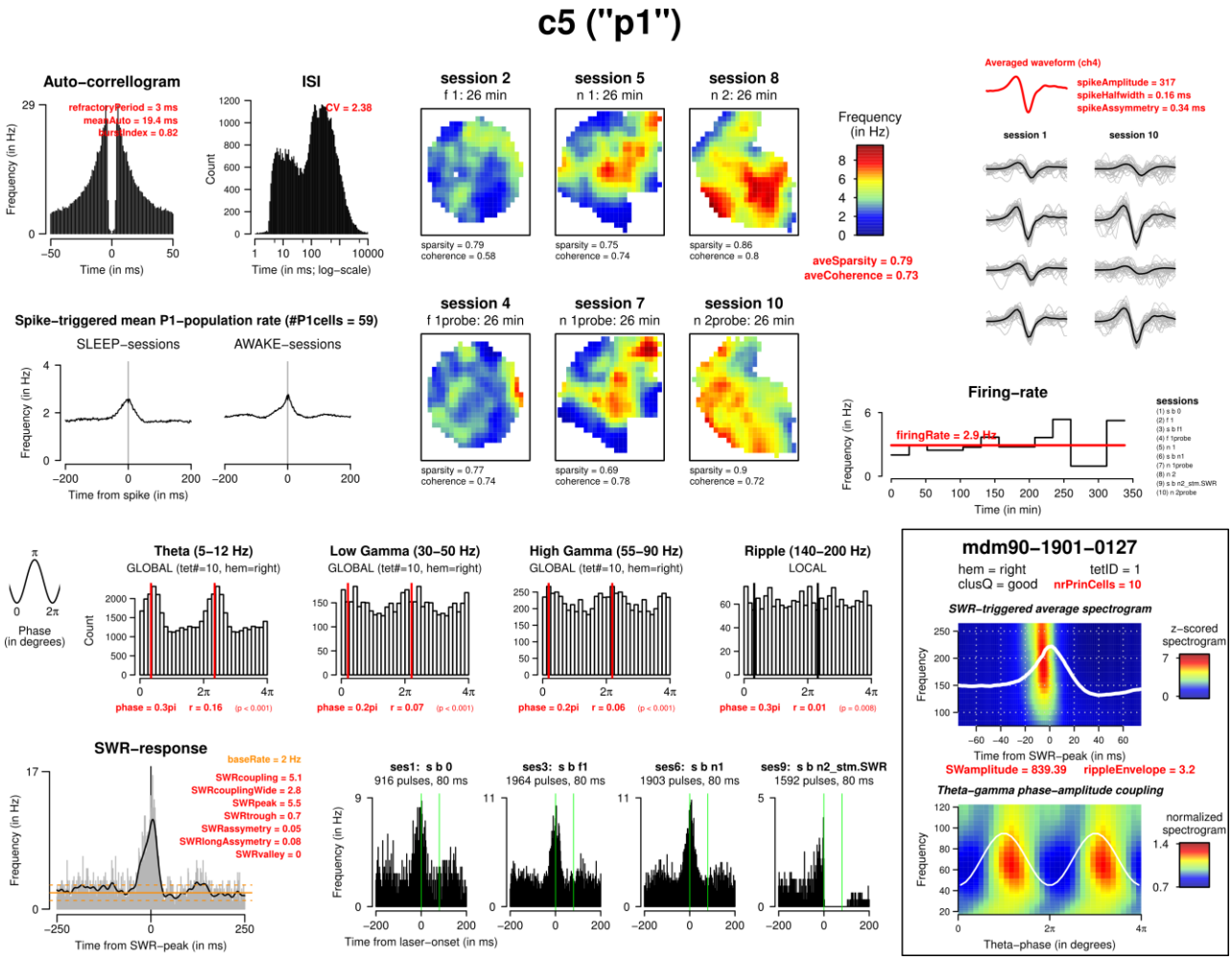

Figure B.1 *Illustration of the output of the computer program "neuronOverview".*

Besides the measures and plots described in section 5.2, this program also generates for each neuron two plots illustrating its population coupling (during sleep and during exploration, see Okun et al., 2015), plots illustrating its firing rate response to light pulses (one for each session with light pulses, either without the laser powered [in this case the first three plots] or with the laser powered [last plot]) and a plot to further characterize the tetrode from which that neuron was recorded (theta-gamma phase-amplitude coupling, see Schomburg et al., 2014).

REFERENCES

Ackerknecht, E.H., and Vallois, H.V. (1956). Franz Joseph Gall, Inventor of Phrenology and His Collection (Department of History of Medicine, University of Wisconsin Medical School).

Adamantidis, A., Arber, S., Bains, J.S., Bamberg, E., Bonci, A., Buzsáki, G., Cardin, J.A., Costa, R.M., Dan, Y., Goda, Y., et al. (2015). Optogenetics: 10 years after ChR2 in neurons--views from the community. Nat. Neurosci. *18*, 1202–1212.

Agostinelli, C., and Lund, U. (2013). R package "circular": Circular Statistics (version (0.4-7).

Agranoff, B.W., Davis, R.E., and Brink, J.J. (1966). Chemical studies on memory fixation in goldfish. Brain Res. *1*, 303–309.

Aiba, A., and Nakao, H. (2007). Conditional mutant mice using tetracycline-controlled gene expression system in the brain. Neurosci. Res. *58*, 113–117.

Akerboom, J., Chen, T.-W., Wardill, T.J., Tian, L., Marvin, J.S., Mutlu, S., Calderón, N.C., Esposti, F., Borghuis, B.G., and Sun, X.R. (2012). Optimization of a GCaMP calcium indicator for neural activity imaging. J. Neurosci. *32*, 13819–13840.

Almeida-Filho, D.G., Lopes-dos-Santos, V., Vasconcelos, N. a P., Miranda, J.G.V., Tort, A.B.L., and Ribeiro, S. (2014). An investigation of Hebbian phase sequences as assembly graphs. Front. Neural Circuits *8*, 34.

Ascher, P., and Nowak, L. (1988). The role of divalent cations in the N-methyl-D-aspartate responses of mouse central neurones in culture. J. Physiol. *399*, 247–266.

Barnes, C.A., Suster, M.S., Shen, J., and McNaughton, B.L. (1997). Multistability of cognitive maps in the hippocampus of old rats. Nature *388*, 272–275.

Barondes, S.H., and Cohen, H.D. (1966). Puromycin effect on successive phases of memory storage. Science *151*, 594–595.

Bartlett, F.C. (1960). Karl Spencer Lashley. 1890-1958. Biogr. Mem. Fellows R. Soc. *5*, 107–118.

Belkin, M., and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. *15*, 1373–1396.

Bliss, T.V.P., and Gardner-Medwin, A.R. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetized rabbit following stimulation of the perforant path. J. Physiol. *232*, 357–374.

Bliss, T.V.P., and Lømo, 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.

Burke, R.E. (2007). Sir Charles Sherrington's The integrative action of the nervous system: a centenary appreciation. Brain *130*, 887–894.

Buzsáki, G. (1986). Hippocampal Sharp Waves : Their Origin and Significance. Brain Res. *398*, 242–252.

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

Buzsáki, G. (2010). Neural syntax: cell assemblies, synapsembles, and readers. Neuron *68*, 362–85.

Buzsáki, G. (2015a). Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. Hippocampus *25*, 1073–1188.

Buzsáki, G. (2015b). Our skewed sense of space. Science *347*, 612–613.

Buzsáki, G., and Moser, E.I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. Nat. Neurosci. *16*, 130–138.

Buzsáki, G., Leung, L., and Vanderwolf, C. (1983). Cellular bases of hippocampal EEG in the behaving rat. Brain Res. *287*, 139–171.

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

Carr, M.F., Jadhav, S.P., and Frank, L.M. (2011). Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. Nat. Neurosci. *14*, 147–153.

Cembrowski, M.S., Wang, L., Sugino, K., Shields, B.C., and Spruston, N. (2016). Hipposeq: a comprehensive RNA-seq database of gene expression in hippocampal principal neurons. eLife *5*, e14997.

Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. J. Stat. Softw. *61*, 1–36.

Chen, T.-W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., and Jayaraman, V. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature *499*, 295–300.

Clark, R.E., West, A.N., Zola, S.M., and Squire, L.R. (2001). Rats with lesions of the hippocampus are impaired on the delayed nonmatching-to-sample task. Hippocampus *11*, 176–186.

Cohen, N.J., and Squire, L.R. (1981). Retrograde amnesia and remote memory impairment. Neuropsychologia *19*, 337–356.

Cole, A.J., Saffen, D.W., Baraban, J.M., and Worley, P.F. (1989). Rapid increase of an immediate early gene messenger RNA in hippocampal neurons by synaptic NMDA receptor activation. Nature *340*, 474–476.

Colgin, L.L., Moser, E.I., and Moser, M.-B. (2008). Understanding memory through hippocampal remapping. Trends Neurosci. *31*, 469–477.

Collingridge, G., Kehl, S., and McLennan, H. t (1983). Excitatory amino acids in synaptic transmission in the Schaffer collateral-commissural pathway of the rat hippocampus. J. Physiol. *334*, 33–46.

Csicsvari, J., and Dupret, D. (2014). Sharp wave/ripple network oscillations and learning-associated hippocampal maps. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *369*, 20120528.

Csicsvari, J., Hirase, H., Czurko, A., and Buzsáki, G. (1998). Reliability and State Dependence of Pyramidal Cell–Interneuron Synapses in the Hippocampus. Neuron *21*, 179–189.

Csicsvari, J., Hirase, H., Czurko, A., Mamiya, A., and Buzsáki, G. (1999a). Fast network oscillations in the hippocampal CA1 region of the behaving rat. J. Neurosci. *19*, 1–4.

Csicsvari, J., Hirase, H., Czurko, A., Mamiya, A., and Buzsáki, G. (1999b). Oscillatory Coupling of Hippocampal Pyramidal Cells and Interneurons in the Behaving Rat. J. Neurosci. *19*, 274–287.

Csicsvari, J., Hirase, H., Mamiya, A., and Buzsáki, G. (2000). Ensemble Patterns of Hippocampal CA3-CA1 Neurons during Sharp Wave–Associated Population Events. Neuron *28*, 585–594.

Csicsvari, J., O'Neill, J., Allen, K., and 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*, 704–16.

Cutsuridis, V., Cobb, S., and Graham, B.P. (2010). Encoding and retrieval in a model of the hippocampal CA1 microcircuit. Hippocampus *20*, 423–446.

DeFelipe, J. (2006). Brain plasticity and mental processes: Cajal again. Nat Rev Neurosci *7*, 811–817.

Denny, C.A., Kheirbek, M.A., Alba, E.L., Tanaka, K.F., Brachman, R.A., Laughman, K.B., Tomm, N.K., Turi, G.F., Losonczy, A., and Hen, R. (2014). Hippocampal memory traces are differentially modulated by experience, time, and adult neurogenesis. Neuron *83*, 189–201.

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

Douglas, R.M. (1977). Long lasting synaptic potentiation in the rat dentate gyrus following brief high frequency stimulation. Brain Res. *126*, 361–365.

Douglas, R.M., and Goddard, G.V. (1975). Long-term potentiation of the perforant path-granule cell synapse in the rat hippocampus. Brain Res. *86*, 205–215.

Dragoi, G., Harris, K.D., and Buzsáki, G. (2003). Place Representation within Hippocampal Networks Is Modified by Long-Term Potentiation. Neuron *39*, 843–853.

Dudai, Y. (2004). The Neurobiology of Consolidations, Or, How Stable is the Engram? Annu. Rev. Psychol. *55*, 51–86.

Dudai, Y. (2012). The restless engram: consolidations never end. Annu. Rev. Neurosci. *35*, 227–247.

Duncan, C.P. (1949). The retroactive effect of electroshock on learning. J. Comp. Physiol. Psychol. *42*, 32–44.

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

Dupret, D., O'Neill, J., and Csicsvari, J. (2013). Dynamic reconfiguration of hippocampal interneuron circuits during spatial learning. Neuron *78*, 166–80.

Ego-Stengel, V., and Wilson, M.A. (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., Amaral, D.G., Buffalo, E.A., Buzsáki, G., Cohen, N., Davachi, L., Frank, L., Heckers, S., Morris, R.G., and Moser, E.I. (2016). Hippocampus at 25. Hippocampus *26*, 1238–1249.

Ekstrom, A.D., Kahana, M.J., Caplan, J.B., Fields, T.A., Isham, E.A., Newman, E.L., and Fried, I. (2003). Cellular networks underlying human spatial navigation. Nature *425*, 184–188.

Eschenko, O., Ramadan, W., Mölle, M., Born, J., and Sara, S.J. (2008). Sustained increase in hippocampal sharp-wave ripple activity during slow-wave sleep after learning. Learn. Mem. *15*, 222–228.

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex *1*, 1–47.

Forwood, S.E., Winters, B., and Bussey, T. (2005). Hippocampal lesions that abolish spatial maze performance spare object recognition memory at delays of up to 48 hours. Hippocampus *15*, 347–355.

Foster, D.J., and Wilson, M.A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature *440*, 680–683.

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

Franklin, K.B.J., and Paxinos, G. (2007). Mouse Brain in Stereotaxic Coordinates (Academic Press).

Friedman, J., Hastie, T., and Tibshirani, R. (2001). The elements of statistical learning (Springer series in statistics Springer, Berlin).

Fujii, T., Moscovitch, M., and Nadel, L. (2000). Memory consolidation, retrograde amnesia, and the temporal lobe. In Handbook of Neuropsychology: Memory and Its Disorders, L.S. Cermak, ed. (Amsterdam: Elsevier Science Publishers B.V.), pp. 223–250.

Gall, F.J. (1822). On the functions of the brain and of each of its parts: with observations on the possibility of determining the instincts, propensities, and talents, or the moral and intellectual dispositions of men and animals, by the configuration of the brain and head. (Boston, Marsh, Capen & Lyon).

Gerrard, J.L., Burke, S.N., McNaughton, B.L., and Barnes, C.A. (2008). Sequence Reactivation in the Hippocampus Is Impaired in Aged Rats. J. Neurosci. *28*, 7883–7890.

Giese, K.P., Fedorov, N.B., Filipkowski, R.K., and Silva, A.J. (1998). Autophosphorylation at Thr286 of the α calcium-calmodulin kinase II in LTP and learning. Science *279*, 870–873.

Gilboa, A., Winocur, G., Grady, C.L., Hevenor, S.J., and Moscovitch, M. (2004). Remembering our past: functional neuroanatomy of recollection of recent and very remote personal events. Cereb. Cortex *14*, 1214–1225.

Girardeau, G., Benchenane, K., Wiener, S.I., Buzsáki, G., and Zugaro, M.B. (2009). Selective suppression of hippocampal ripples impairs spatial memory. Nat. Neurosci. *12*, 1222–3.

Götze, F., Tikhomirov, A., and others (2004). Rate of convergence in probability to the Marchenko-Pastur law. Bernoulli *10*, 503–548.

Grün, S., Diesmann, M., and Aertsen, A. (2002a). Unitary events in multiple single-neuron spiking activity: I. Detection and significance. Neural Comput. *14*, 43–80.

Grün, S., Diesmann, M., and Aertsen, A. (2002b). Unitary events in multiple single-neuron spiking activity: II. Nonstationary data. Neural Comput. *14*, 81–119.

Guzowski, J.F., Timlin, J. a, Roysam, B., McNaughton, B.L., Worley, P.F., and Barnes, C.A. (2005). Mapping behaviorally relevant neural circuits with immediate-early gene expression. Curr. Opin. Neurobiol. *15*, 599–606.

Han, X., Chow, B.Y., Zhou, H., Klapoetke, N.C., Chuong, A., Rajimehr, R., Yang, A., Baratta, M.V., Winkle, J., Desimone, R., et al. (2011). A high-light sensitivity optical neural silencer: development and application to optogenetic control of non-human primate cortex. Front. Syst. Neurosci. *5*, 18.

Harris, K.D. (2008). Stability of the fittest: organizing learning through retroaxonal signals. Trends Neurosci. *31*, 130–136.

Harris, E.W., Ganong, A.H., and Cotman, C.W. (1984). Long-term potentiation in the hippocampus involves activation of N-methyl-D-aspartate receptors. Brain Res. *323*, 132–137.

Harris, K.D., Henze, D.A., Csicsvari, J., Hirase, H., and Buzsáki, G. (2000). Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. J. Neurophysiol. *84*, 401–414.

Harris, K.D., Csicsvari, J., Hirase, H., Dragoi, G., and Buzsáki, G. (2003). Organization of cell assemblies in the hippocampus. Nature *424*, 552–556.

Harris, K.D., Quiroga, R.Q., Freeman, J., and Smith, S.L. (2016). Improving data quality in neuronal population recordings. Nat. Neurosci. *19*, 1165–1174.

Hebb, D.O. (1949). The Organization of Behavior: A Neuropsychological Theory (New York: Wiley).

Hines, M.L., and Carnevale, N.T. (1997). The NEURON simulation environment. Neural Comput. *9*, 1179–1209.

Holt, G.R., Softky, W.R., Koch, C., and Douglas, R.J. (1996). Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. J. Neurophysiol. *75*, 1806–1814.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. *10*, 626–634.

Jackson, J., and Redish, A.D. (2007). Network dynamics of hippocampal cell-assemblies resemble multiple spatial maps within single tasks. Hippocampus *17*, 1209–29.

Jackson, J.C., Johnson, A., and Redish, A.D. (2006). Hippocampal Sharp Waves and Reactivation during Awake States Depend on Repeated Sequential Experience. J. Neurosci. *26*, 12415–12426.

Jadhav, S.P., Kemere, C., German, P.W., and Frank, L.M. (2012). Awake hippocampal sharp-wave ripples support spatial memory. Science *336*, 1454–8.

James, W. (1890). The Principles of Psychology (London: Macmillan and co.).

Ji, D., and Wilson, M.A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. Nat. Neurosci. *10*, 100–7.

Jones, E.G. (1994). The neuron doctrine 1891. J. Hist. Neurosci. *3*, 3–20.

Josselyn, S.A., Köhler, S., and Frankland, P.W. (2015). Finding the engram. Nat. Rev. Neurosci. *16*, 521–534.

Karlsson, M.P., and Frank, L.M. (2009). Awake replay of remote experiences in the hippocampus. Nat. Neurosci. *12*, 913–918.

Katona, L., Lapray, D., Viney, T.J., Oulhaj, A., Borhegyi, Z., Micklem, B.R., Klausberger, T., and Somogyi, P. (2014). Sleep and movement differentiates actions of two types of somatostatin-expressing GABAergic interneuron in rat hippocampus. Neuron *82*, 872–86.

Kentros, C., Hargreaves, E., Hawkins, R.D., Kandel, E.R., Shapiro, M., and Muller, R.V. (1998). Abolition of Long-Term Stability of New Hippocampal Place Cell Maps by NMDA Receptor Blockade. Science *280*, 2121–2126.

Kentros, C.G., Agnihotri, N.T., Streater, S., Hawkins, R.D., and Kandel, E.R. (2004). Increased attention to spatial context increases both place field stability and spatial memory. Neuron *42*, 283–295.

King, C., Henze, D.A., Leinekugel, X., and Buzsáki, G. (1999). Hebbian modification of a hippocampal population pattern in the rat. J. Physiol. *521*, 159–167.

Klausberger, T., and Somogyi, P. (2008). Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. Science *321*, 53–57.

Klausberger, T., Magill, P.J., Cobden, P.M., Somogyi, P., Márton, L.F., Roberts, J.D.B., and Buzsáki, G. (2003). Brain-state- and cell-type-specific firing of hippocampal interneurons in vivo. Nature *421*, 844–8.

Kobayashi, T., Nishijo, H., Fukuda, M., Bures, J., and Ono, T. (1997). Task-Dependent Representations in Rat Hippocampal Place Neurons. J. Neurophysiol. *78*, 597–613.

Koch, C., Rapp, M., and Segev, I. (1996). A brief history of time (constants). Cereb. Cortex *6*, 93–101.

Köhler, W. (1940). Dynamics in Psychology (London: Faber and Faber Limited).

Kovács, K.A., O'Neill, J., Schoenenberger, P., Penttonen, M., Guerrero, D.K.R., and Csicsvari, J. (2016). Optogenetically Blocking Sharp Wave Ripple Events in Sleep Does Not Interfere with the Formation of Stable Spatial Representation in the CA1 Area of the Hippocampus. PloS One *11*, e0164675.

Kruskal, J.B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika *29*, 1–27.

Kruskal, P.B., Stanis, J.J., McNaughton, B.L., and Thomas, P.J. (2007). A binless correlation measure reduces the variability of memory reactivation estimates. Stat. Med. *26*, 3997–4008.

Kudrimoti, H.S., Barnes, C.A., and McNaughton, B.L. (1999). Reactivation of hippocampal cell assemblies: effects of behavioral state, experience, and EEG dynamics. J. Neurosci. *19*, 4090–4101.

Kumaran, D., Hassabis, D., and McClelland, J.L. (2016). What learning systems do intelligent agents need? complementary learning systems theory updated. Trends Cogn. Sci. *20*, 512–534.

Lansink, C.S., Goltstein, P.M., Lankelma, J.V., McNaughton, B.L., and Pennartz, C.M. (2009). Hippocampus leads ventral striatum in replay of place-reward information. PLoS Biol *7*, e1000173.

Lapray, D., Lasztoczi, B., Lagler, M., Viney, T.J., Katona, L., Valenti, O., Hartwich, K., Borhegyi, Z., Somogyi, P., and Klausberger, T. (2012). Behavior-dependent specialization of identified hippocampal interneurons. Nat. Neurosci. *15*, 1265–71.

Lashley, K.S. (1929). Brain mechanisms and intelligence: A quantitative study of injuries to the brain. (Chicago, IL: University of Chicago Press).

Lashley, K.S. (1942). The problem of cerebral organization in vision. In Visual Mechanisms, H. Klüver, ed. (Lancaster, PA: The science press printing company), pp. 301–322.

Lashley, K.S. (1950). In search of the engram. Soc. Exp. Biol. Symp. *4*, 454–482.

Lashley, K.S. (1951). The problem of serial order in behavior. In Cerebral Mechanisms in Behavior, (New York, NY: Wiley), pp. 112–136.

Laubach, M., Shuler, M., and Nicolelis, M.A.L. (1999). Independent component analyses for quantifying neuronal ensemble interactions. J. Neurosci. Methods *94*, 141–154.

de Lavilléon, G., Lacroix, M.M., Rondi-Reig, L., and Benchenane, K. (2015). Explicit memory creation during sleep demonstrates a causal role of place cells in navigation. Nat. Neurosci. *18*, 493–495.

Lechner, H.A., Squire, L.R., and Byrne, J.H. (1999). 100 years of consolidation—remembering Müller and Pilzecker. Learn. Mem. *6*, 77–87.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444.

Lee, A.K., and Wilson, M.A. (2002). Memory of Sequential Experience in the Hippocampus during Slow Wave Sleep. Neuron *36*, 1183–1194.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature *445*, 168–176.

Levy, W.B., and Steward, O. (1979). Synapses as associative memory elements in the hippocampal formation. Brain Res. *175*, 233–245.

Liu, X., Ramirez, S., Pang, P.T., Puryear, C.B., Govindarajan, A., Deisseroth, K., and Tonegawa, S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. Nature *484*, 381–5.

Lloyd, S. (1982). Least squares quantization in PCM. IEEE Trans. Inf. Theory *28*, 129–137.

Lømo, T. (1966). Frequency potentiation of excitatory synaptic activity in dentate area of hippocampal formation. In Acta Physiologica Scandinavica, p. 128.

Lopes-dos-Santos, V., Conde-Ocazionez, S., Nicolelis, M.A.L., Ribeiro, S.T., and Tort, A.B.L. (2011). Neuronal assembly detection and cell membership specification by principal component analysis. PloS One *6*, e20996.

Lopes-dos-Santos, V., Ribeiro, S., and Tort, A.B.L. (2013). Detecting cell assemblies in large neuronal populations. J. Neurosci. Methods *220*, 149–166.

Lorente de Nó, R. (1939). Transmission of impulses through cranial motor nuclei. J. Neurophysiol. *2*, 402–464.

Lorento de Nó, R. (1938). Analysis of the activity of the chains of internuncial neurons. J. Neurophysiol. *1*, 207–244.

Louie, K., and Wilson, M.A. (2001). Temporally Structured Replay of Awake Hippocampal Ensemble Activity during Rapid Eye Movement Sleep. Neuron *29*, 145–156.

Löwel, S., and Singer, W. (1992). Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. Science *255*, 209–212.

Luczak, A., McNaughton, B.L., and Harris, K.D. (2015). Packet-based communication in the cortex. Nat. Rev. Neurosci. *16*, 745–755.

Maaten, L. van der, and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

MacDermott, A.B., Mayer, M.L., Westbrook, G.L., Smith, S.J., and Barker, J.L. (1986). NMDA-receptor activation increases cytoplasmic calcium concentration in cultured spinal cord neurones. Nature *321*, 519–522.

Maingret, N., Girardeau, G., Todorova, R., Goutierre, M., and Zugaro, M. (2016). Hippocampo-cortical coupling mediates memory consolidation during sleep. Nat. Neurosci. *19*, 959–964.

Malvache, A., Reichinnek, S., Villette, V., Haimerl, C., and Cossart, R. (2016). Awake hippocampal reactivations project onto orthogonal neuronal assemblies. Science *353*, 1280–1283.

Marčenko, V.A., and Pastur, L.A. (1967). Distribution of eigenvalues for some sets of random matrices. Sb. Math. *1*, 457–483.

Marchini, J.L., Heaton, C., and Ripley, B.D. (2013). fastICA: FastICA Algorithms to perform ICA and Projection Pursuit.

Mariño, J., Schummers, J., Lyon, D.C., Schwabe, L., Beck, O., Wiesing, P., Obermayer, K., and Sur, M. (2005). Invariant computations in local cortical networks with balanced excitation and inhibition. Nat. Neurosci. *8*, 194–201.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004). Interneurons of the neocortical inhibitory system. Nat. Rev. Neurosci. *5*, 793–807.

Marr, D. (1970). A Theory for Cerebral Neocortex. Proc. R. Soc. Lond. B Biol. Sci. *176*, 161–234.

Marr, D. (1971). Simple Memory: A Theory for Archicortex. Philos. Trans. R. Soc. Lond. B Biol. Sci. *262*, 23–81.

Mayer, M.L., Westbrook, G.L., and Guthrie, P.B. (1984). Voltage-dependent block by Mg2+ of NMDA responses in spinal cord neurones. Nature *309*, 261–263.

McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol. Rev. *102*, 419–457.

McCloskey, M., and Cohen, N.J. (1989). Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. In Psychology of Learning and Motivation, G.H. Bower, ed. (Academic Press), pp. 109–165.

McGaugh, J.L. (1966). Time-Dependent Processes in Memory Storage. Science *153*, 1351–1358.

McGaugh, J.L. (2007). Searching for memory in the brain: Confronting the collusion of cells and systems. In Neural Plasticity and Memory: From Genes to Brain Imaging, F. Bermudez-Rattoni, ed. (CRC Press), pp. 1–13.

McHugh, T.J., Blum, K.I., Tsien, J.Z., Tonegawa, S., and Wilson, M.A. (1996). Impaired hippocampal representation of space in CA1-specific NMDAR1 knockout mice. Cell *87*, 1339–1349.

McNamara, C.G., Tejero-Cantero, Á., Trouche, S., Campo-Urriza, N., and Dupret, D. (2014). Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. Nat. Neurosci. *17*, 1658–1660.

McNaughton, B.L. (2003). Long-term potentiation, cooperativity and Hebb's cell assemblies: a personal history. Philos. Trans. R. Soc. Lond. B Biol. Sci. *358*, 629–634.

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

McNaughton, B.L., Douglas, R., and Goddard, G.V. (1978). Synaptic enhancement in fascia dentata: cooperativity among coactive afferents. Brain Res. *157*, 277–293.

Miller, S.G., and Kennedy, M.B. (1986). Regulation of brain Type II Ca2+ calmodulin-dependent protein kinase by autophosphorylation: A Ca2+-triggered molecular switch. Cell *44*, 861–870.

Milner, B. (1962). Les troubles de la mémoire accompagnant les lésions hippocampiques bilatérales. In Physiologie de l'Hippocampe, (Paris, C.N.R.S.: Colloques Internationaux No. 107), pp. 257–272.

Milner, P.M. (1957). The cell assembly: Mark II. Psychol. Rev. *64*, 242–252.

Milner, P.M. (1993). The Mind and Donald O. Hebb. Sci. Am. *268*, 124–129.

Milner, B., Corkin, S., and Teuber, H.-L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of H.M. Neuropsychologia *6*, 215–234.

Milner, B., Squire, L.R., and Kandel, E.R. (1998). Cognitive neuroscience and the study of memory. Neuron *20*, 445–468.

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

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., and Ostrovski, G. (2015). Human-level control through deep reinforcement learning. Nature *518*, 529–533.

Morris, R.G. (1981). Spatial localization does not require the presence of local cues. Learn. Motiv. *12*, 239–260.

Morris, R., Garrud, P., Rawlins, J., and O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. Nature *297*, 681–683.

Morris, R., Anderson, E., Lynch, G. a, and Baudry, M. (1986). Selective impairment of learning and blockade of long-term potentiation by an N-methyl-D-aspartate receptor antagonist, AP 5. Nature *319*, 774–776.

Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A., and Rosenbaum, R.S. (2006). The cognitive neuroscience of remote episodic, semantic and spatial memory. Curr. Opin. Neurobiol. *16*, 179–190.

Mugavin, M.E. (2008). Multidimensional scaling: a brief overview. Nurs. Res. *57*, 64–68.

Muller, R.U. (1996). A quarter of a century of place cells. Neuron *17*, 813–822.

Müller, G.E., and Pilzecker, A. (1900). Experimentelle beiträge zur lehre vom gedächtniss (Z. Psychol. Ergänzungsband).

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

Muller, R.U., and Kubie, J.L. (1989). The firing of hippocampal place cells predicts the future position of freely moving rats. J. Neurosci. *9*, 4101–4110.

Muller, R.U., Kubie, J.L., and Ranck, J.B. (1987). Spatial Firing Patterns of Hippocampal Complex-Spike Cells in a Fixed Environment. J. Neurosci. *7*, 1935–1950.

Nadasdy, Z., Hirase, H., Czurko, A., Csicsvari, J., and Buzsáki, G. (1999). Replay and Time Compression of Recurring Spike Sequences in the Hippocampus. J. Neurosci. *19*, 9497–9507.

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

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

Nicoll, R.A. (2017). A Brief History of Long-Term Potentiation. Neuron *93*, 281–290.

Nowak, L., Bregestovski, P., Ascher, P., Herbet, A., and Prochiantz, A. (1984). Magnesium gates glutamate-activated channels in mouse central neurones. Nature *307*, 462–465.

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

O'Keefe, J., and Nadel, L. (1978). The hippocampus as a cognitive map (Clarendon Press Oxford).

Okun, M., Steinmetz, N.A., Cossell, L., Iacaruso, M.F., Ko, H., Barthó, P., Moore, T., Hofer, S.B., Mrsic-Flogel, T.D., Carandini, M., et al. (2015). Diverse coupling of neurons to populations in sensory cortex. Nature *521*, 511–515.

Ólafsdóttir, H.F., Carpenter, F., and Barry, C. (2016). Coordinated grid and place cell replay during rest. Nat. Neurosci. *19*, 792–794.

O'Neill, J., Senior, T., and Csicsvari, J. (2006). Place-selective firing of CA1 pyramidal cells during sharp wave/ripple network patterns in exploratory behavior. Neuron *49*, 143–55.

O'Neill, J., Senior, T.J., Allen, K., Huxter, J.R., and Csicsvari, J. (2008). Reactivation of experience-dependent cell assembly patterns in the hippocampus. Nat. Neurosci. *11*, 209–15.

O'Neill, J., Pleydell-Bouverie, B., Dupret, D., and Csicsvari, J. (2010). Play it again: reactivation of waking experience and memory. Trends Neurosci. *33*, 220–9.

O'Neill, J., Boccara, C., Stella, F., Schoenenberger, P., and Csicsvari, J. (2017). Superficial layers of the medial entorhinal cortex replay independently of the hippocampus. Science *355*, 184–188.

Ono, T., Nakamura, K., Nishijo, H., and Eifuku, S. (1993). Monkey hippocampal neurons related to spatial and nonspatial functions. J. Neurophysiol. *70*, 1516–1529.

Otto, T., and Eichenbaum, H. (1992). Complementary roles of the orbital prefrontal cortex and the perirhinal-entorhinal cortices in an odor-guided delayed-nonmatching-to-sample task. Behav. Neurosci. *106*, 762–775.

Pavlides, C., and Winson, J. (1989). Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. J. Neurosci. *9*, 2907–2918.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

Penfield, W., and Milner, B. (1958). Memory Deficit Produced by Bilateral Lesions in the Hippocampal Zone. AMA Arch. Neurol. Psychiatry *79*, 475–497.

Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S.I., and Battaglia, F.P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. Nat. Neurosci. *12*, 919–26.

Peyrache, A., Benchenane, K., Khamassi, M., Wiener, S.I., and Battaglia, F.P. (2010). Principal component analysis of ensemble recordings reveals cell assemblies at high temporal resolution. J. Comput. Neurosci. *29*, 309–25.

Pfeiffer, B.E., and Foster, D.J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. Nature *497*, 74–79.

Qin, Y.-L., McNaughton, B.L., Skaggs, W.E., and Barnes, C.A. (1997). Memory reprocessing in corticocortical and hippocampocortical neuronal ensembles. Philos. Trans. R. Soc. B Biol. Sci. *352*, 1525–1533.

Quintilian (1856/2006). Institutes of oratory (Iowa State).

Quiroga, R.Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering. Neural Comput. *16*, 1661–1687.

Ramadan, W., Eschenko, O., and Sara, S.J. (2009). Hippocampal sharp wave/ripples during sleep for consolidation of associative memory. PloS One *4*, e6697.

Ramirez, S., Liu, X., Lin, P.-A., Suh, J., Pignatelli, M., Redondo, R.L., Ryan, T.J., and Tonegawa, S. (2013). Creating a false memory in the hippocampus. Science *341*, 387–391.

Ramón y Cajal, S. (1894). The Croonian Lecture: La Fine Structure des Centres Nerveux. Proc. R. Soc. Lond. *55*, 444–468.

Reijmers, L., and Mayford, M. (2009). Genetic control of active neural circuits. Front. Mol. Neurosci. *2*, 1–8.

Reijmers, L.G., Perkins, B.L., Matsuo, N., and Mayford, M. (2007). Localization of a stable neural correlate of associative memory. Science *317*, 1230–3.

Ribot, T. (1882). Diseases of Memory: an essay in the positive psychology (London: Kegan Paul, Trench & Co.).

de la Rocha, J., Doiron, B., Shea-Brown, E., Josić, K., and Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. Nature *448*, 802–806.

Rochester, N., Holland, J., Haibt, L., and Duda, W. (1956). Tests on a cell assembly theory of the action of the brain, using a large digital computer. IRE Trans. Inf. Theory *2*, 80–93.

Rothschild, G., Eban, E., and Frank, L.M. (2017). A cortical-hippocampal-cortical loop of information processing during memory consolidation. Nat. Neurosci. *20*, 251–259.

Roumis, D.K., and Frank, L.M. (2015). Hippocampal sharp-wave ripples in waking and sleeping states. Curr. Opin. Neurobiol. *35*, 6–12.

Rousseeuw, P.J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. *20*, 53–65.

Royer, S., Zemelman, B.V., Losonczy, A., Kim, J., Chance, F., Magee, J.C., and Buzsáki, G. (2012). Control of timing, rate and bursts of hippocampal place cells by dendritic and somatic inhibition. Nat. Neurosci. *15*, 769–775.

Russo, E., and Durstewitz, D. (2017). Cell assemblies at multiple time scales with arbitrary lag constellations. eLife *6*, e19428.

Ryan, T.J., Roy, D.S., Pignatelli, M., Arons, A., and Tonegawa, S. (2015). Engram cells retain memory under retrograde amnesia. Science *348*, 1007–1013.

Sadowski, J.H.L.P., Jones, M.W., and Mellor, J.R. (2016). Sharp-Wave Ripples Orchestrate the Induction of Synaptic Plasticity during Reactivation of Place Cell Firing Patterns in the Hippocampus. Cell Rep. *14*, 1916–1929.

Schiller, D., Eichenbaum, H., Buffalo, E.A., Davachi, L., Foster, D.J., Leutgeb, S., and Ranganath, C. (2015). Memory and Space: Towards an Understanding of the Cognitive Map. J. Neurosci. *35*, 13904–13911.

Schomburg, E.W., Fernández-Ruiz, A., Mizuseki, K., Berényi, A., Anastassiou, C.A., Koch, C., and Buzsáki, G. (2014). Theta Phase Segregation of Input-Specific Gamma Patterns in Entorhinal-Hippocampal Networks. Neuron *84*, 470–485.

Scoville, W.B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. J. Neurol. Neurosurg. Psychiatry *20*, 11–21.

Sherrington, C.S. (1892). Note on the knee-jerk and the correlation of action of antagonistic muscles. Proc. R. Soc. Lond. *52*, 556–564.

Sherrington, C.S. (1906). The integrative action of the nervous system (New Haven, CT: Yale University Press).

Sherry, D.F., and Schacter, D.L. (1987). The evolution of multiple memory systems. Psychol. Rev. *94*, 439–454.

Singer, A.C., and Frank, L.M. (2009). Rewarded outcomes enhance reactivation of experience in the hippocampus. Neuron *64*, 910–21.

Singer, A.C., Carr, M.F., Karlsson, M.P., and Frank, L.M. (2013). Hippocampal SWR activity predicts correct decisions during the initial learning of an alternation task. Neuron *77*, 1163–1173.

Skaggs, W.E., and McNaughton, B.L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. Science *271*, 1870–1873.

Skaggs, W.E., McNaughton, B.L., Wilson, M.A., and Barnes, C.A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. Hippocampus *6*, 149–172.

Smeyne, R.J., Schilling, K., Robertson, L., Luk, D., Oberdick, J., Curran, T., and Morgan, J.I. (1992). Fos-IacZ transgenic mice: Mapping sites of gene induction in the central nervous system. Neuron *8*, 13–23.

Somogyi, P., Katona, L., Klausberger, T., Lasztóczi, B., and Viney, T.J. (2014). Temporal redistribution of inhibition over neuronal subcellular domains underlies state-dependent rhythmic change of excitability in the hippocampus. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *369*, 20120518.

Spruston, N., and Johnston, D. (1992). Perforated patch-clamp analysis of the passive membrane properties of three classes of hippocampal neurons. J. Neurophysiol. *67*, 508–529.

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

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

Squire, L.R., and Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. Curr. Opin. Neurobiol. *5*, 169–177.

Squire, L.R., and Zola-Morgan, S. (1991). The medial temporal lobe memory system. Science *253*, 1380–1387.

Squire, L.R., Clark, R.E., and Knowlton, B.J. (2001). Retrograde amnesia. Hippocampus *11*, 50–55.

Squire, L.R., Genzel, L., Wixted, J.T., and Morris, R.G. (2015). Memory consolidation. Cold Spring Harb. Perspect. Biol. *7*, a021766.

Staresina, B.P., Bergmann, T.O., Bonnefond, M., Van Der Meij, R., Jensen, O., Deuker, L., Elger, C.E., Axmacher, N., and Fell, J. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. Nat. Neurosci. *18*, 1679–1686.

Stark, E., Roux, L., Eichler, R., Senzai, Y., Royer, S., and Buzsáki, G. (2014). Pyramidal cell-interneuron interactions underlie hippocampal ripple oscillations. Neuron *83*, 467–480.

Sullivan, D., Csicsvari, J., Mizuseki, K., Montgomery, S., Diba, K., and Buzsáki, G. (2011). Relationships between hippocampal sharp waves, ripples, and fast gamma oscillation: influence of dentate and entorhinal cortical activity. J. Neurosci. *31*, 8605–16.

Sutherland, G., and McNaughton, B.L. (2000). Memory trace reactivation in hippocampal and neocortical neuronal ensembles. Curr. Opin. Neurobiol. *10*, 180–186.

Tanaka, K.Z., Pevzner, A., Hamidi, A.B., Nakazawa, Y., Graham, J., and Wiltgen, B.J. (2014). Cortical Representations Are Reinstated by the Hippocampus during Memory Retrieval. Neuron *84*, 347–354.

Tayler, K.K., Tanaka, K.Z., Reijmers, L.G., and Wiltgen, B.J. (2013). Reactivation of neural ensembles during the retrieval of recent and remote memory. Curr. Biol. *23*, 99–106.

Tenenbaum, J.B., De Silva, V., and Langford, J.C. (2000). A global geometric framework for nonlinear dimensionality reduction. Science *290*, 2319–2323.

Theis, L., Berens, P., Froudarakis, E., Reimer, J., Rosón, M.R., Baden, T., Euler, T., Tolias, A.S., and Bethge, M. (2016). Benchmarking spike rate inference in population calcium imaging. Neuron *90*, 471–482.

Thompson, L.T., and Best, P.J. (1990). Long-term stability of the place-field activity of single units recorded from the dorsal hippocampus of freely behaving rats. Brain Res. *509*, 299–308.

Thorndike, R.L. (1953). Who belongs in the family? Psychometrika *18*, 267–276.

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. Ser. B Stat. Methodol. *63*, 411–423.

Tonegawa, S., Liu, X., Ramirez, S., and Redondo, R. (2015). Memory engram cells have come of age. Neuron *87*, 918–931.

Treves, A., and Rolls, E.T. (1994). Computational analysis of the role of the hippocampus in memory. Hippocampus *4*, 374–91.

Trouche, S., Perestenko, P.V., van de Ven, G.M., Bratley, C.T., McNamara, C.G., Campo-Urriza, N., Black, S.L., Reijmers, L.G., and Dupret, D. (2016). Recoding a cocaine-place memory engram to a neutral engram in the hippocampus. Nat. Neurosci. *19*, 564–567.

Tsien, J.Z., Huerta, P.T., and Tonegawa, S. (1996a). The essential role of hippocampal CA1 NMDA receptor–dependent synaptic plasticity in spatial memory. Cell *87*, 1327–1338.

Tsien, J.Z., Chen, D.F., Gerber, D., Tom, C., Mercer, E.H., Anderson, D.J., Mayford, M., Kandel, E.R., and Tonegawa, S. (1996b). Subregion- and Cell Type–Restricted Gene Knockout in Mouse Brain. Cell *87*, 1317–1326.

Ulanovsky, N., and Moss, C.F. (2007). Hippocampal cellular and network activity in freely moving echolocating bats. Nat. Neurosci. *10*, 224–233.

Valero, M., Cid, E., Averkin, R.G., Aguilar, J., Sanchez-Aguilera, A., Viney, T.J., Gomez-Dominguez, D., Bellistri, E., and de la Prida, L.M. (2015). Determinants of different deep and superficial CA1 pyramidal cell dynamics during sharp-wave ripples. Nat. Neurosci. *18*, 1281–1290.

Van Strien, N.M., Cappaert, N., and Witter, M.P. (2009). The anatomy of memory: an interactive overview of the parahippocampal–hippocampal network. Nat. Rev. Neurosci. *10*, 272–282.

Varga, C., Golshani, P., and Soltesz, I. (2012). Frequency-invariant temporal ordering of interneuronal discharges during hippocampal oscillations in awake mice. Proc. Natl. Acad. Sci. *109*, E2726–E2734.

Varga, C., Oijala, M., Lish, J., Szabo, G.G., Bezaire, M., Marchionni, I., Golshani, P., and Soltesz, I. (2014). Functional fission of parvalbumin interneuron classes during fast network events. Elife *3*, e04006.

Villette, V., Malvache, A., Tressard, T., Dupuy, N., and Cossart, R. (2015). Internally recurring hippocampal sequences as a population template of spatiotemporal information. Neuron *88*, 357–366.

Viney, T.J., Lasztoczi, B., Katona, L., Crump, M.G., Tukker, J.J., Klausberger, T., and Somogyi, P. (2013). Network state-dependent inhibition of identified hippocampal CA3 axo-axonic cells in vivo. Nat. Neurosci. *16*, 1802–11.

Vyazovskiy, V.V., Faraguna, U., Cirelli, C., and Tononi, G. (2009). Triggering slow waves during NREM sleep in the rat by intracortical electrical stimulation: effects of sleep/wake history and background activity. J. Neurophysiol. *101*, 1921–1931.

von Waldeyer-Harz, W. (1891). Über einige neuere Forschungen im Gebiete der Anatomie des Centralnervensystems. Dtsch. Med Wschr *17*, 1213–1218, 1244–1246, 1267–1269, 1287–1289, 1331–1332, 1352–1356.

Weir, K., Blanquie, O., Kilb, W., Luhmann, H.J., and Sinning, A. (2014). Comparison of spike parameters from optically identified GABAergic and glutamatergic neurons in sparse cortical cultures. Front. Cell. Neurosci. *8*, 460.

Wikenheiser, A.M., and Redish, A.D. (2015). Decoding the cognitive map: ensemble hippocampal sequences and decision making. Curr. Opin. Neurobiol. *32*, 8–15.

Wilson, M.A., and McNaughton, B.L. (1993). Dynamics of the hippocampal ensemble code for space. Science *261*, 1055–1058.

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

Winocur, G., and Moscovitch, M. (2011). Memory transformation and systems consolidation. J. Int. Neuropsychol. Soc. *17*, 766–780.

Winocur, G., Moscovitch, M., and 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.

Winocur, G., Sekeres, M.J., Binns, M.A., and Moscovitch, M. (2013). Hippocampal lesions produce both nongraded and temporally graded retrograde amnesia in the same rat. Hippocampus *23*, 330–341.

Xu, X., and Kezunovic, M. (2002). Automated feature extraction from power system transients using wavelet transform. (IEEE), pp. 1994–1998.

Yizhar, O., Fenno, L.E., Prigge, M., Schneider, F., Davidson, T.J., O'Shea, D.J., Sohal, V.S., Goshen, I., Finkelstein, J., and Paz, J.T. (2011). Neocortical excitation/inhibition balance in information processing and social dysfunction. Nature *477*, 171–178.

Zar, J.H. (1999). Biostatistical Analysis (Upper Saddle River, New Jerset, USA: Prentice Hall International, Inc.).

Zola-Morgan, S. (1995). Localization of Brain Function: The Legacy of Franz Joseph Gall (1758-1828). Annu. Rev. Neurosci. *18*, 359–383.

Zola-Morgan, S.M., and Squire, L.R. (1990). The primate hippocampal formation: Evidence for a time-limited role in memory storage. Science *250*, 288–290.

Zugaro, M.B., Monconduit, L., and Buzsáki, G. (2005). Spike phase precession persists after transient intrahippocampal perturbation. Nat. Neurosci. *8*, 67–71.