University of California
Santa Barbara

Experimental and Computational Interrogation of the Hippocampal Formation

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Dynamical Neuroscience

by

William Thomas Redman

Committee in charge:

Professor Michael J. Goard, Chair
Professor Emily G. Jacobs
Professor Ikuko T. Smith
Professor Caleb T. Kamere

June 2023

The Dissertation of William Thomas Redman is approved.

Professor Emily G. Jacobs

Professor Ikuko T. Smith

Professor Caleb T. Kamere

Professor Michael J. Goard, Committee Chair

May 2023

Experimental and Computational Interrogation of the Hippocampal Formation

by

William Thomas Redman

To the mighty, mighty Redmans

Ode to Santa Ynez Valley

If Los Olivos means the olives

Why did I see grapes?

If Solvang is a Danish town

Why did I see crepes?

But maybe I'm too narrow minded

Not seeing the beau-ty

Wise words the coffee shop owner did tell me

"Redecorate, for su-shi"

Goard Lab Sea Shanty[1]

There once was a lab

That looked on the sea

Writing MATLAB code

And using the 2P

They fought pirates

For buried booty

Eating free pizza

Was their sworn duty

Hey!

[1]To be performed in rugged, handsome Scottish.

Acknowledgements

"We mistake induction for generation"

Richard Fariña (1966)

I have read the acknowledgements section of many theses (often, that is as far as I get), and I have found a general trend for a short set of acknowledgements. While I respect the succinctness with which they have been written, the separation of public and private spheres, and the lack of dramatics that suggest a PhD is the greatest honor one can achieve in life, mine will not be brief. This is for two reasons.

First, it is a unique opportunity to get to acknowledge all the people who have supported you, encouraged you, given you opportunities, taught you, fought for you, etc., and this, to my mind, makes it worthwhile to take full advantage of. And second, Richard Fariña's quote is highly relevant to the work you will find in the following chapters. To pretend otherwise, and not acknowledge in full that I have been thoroughly indebted to the process of induction, far more so than generation, would be a crime. This is equally true beyond the lab, where many of my life's role models reside. And, as I will hopefully make clear in the next few pages, I have had the fortune to interact with a great number of incredible, generous, brilliant humans.

Science was not on my radar, especially in the context of being something I could be good enough at to pursue as a career, until my freshman year of high school when I took Mrs. Lucci's biology class. At the year's end, she chose me to present an award to. It is hard to overstate the surprise I experienced receiving this, being a middle tier student who felt unqualified compared to some of the other students in the class. Mom called Mrs. Lucci that afternoon, to ask if she had really given the award to the right

kid. Mrs. Lucci's response is something I have cherished and opened up a new world of opportunities for me.

My senior year of high school, I was introduced to research through the Immersion Science Program at Fox Chase Cancer Center, led by Prof. Alana O'Reilly and Dr. Dara Ruiz-Whalen. Alana and Dara encouraged my curiosity, and have continued to support my career in science. They have always served as inspirations for non-traditional ways in which scientific inquiry can be pursued, including ways that give back to the larger community.

Through Alana, I was connected with Prof. Thomas Gregor, whose biophysics lab at Princeton University I was a part of from 2014-2015. I am greatly indebted to Thomas taking a chance on a high school kid, and the Gregor Lab and collaborators, particularly Prof. Madhav Mani and Dr. Danny Wells, greatly shaped my interest and independence in research.

My time at NYU was marked by a number of remarkable friends and professors, and was fostered by community driven departments. Profs. Antoine Cerfon, Joel Spencer, Sylvain Cappell, Edo Kussell, and Alexander Grosberg all taught particularly exciting and engaging courses that challenged me. Antoine and Prof. Claude Desplan were willing to listen to my research ideas and provide some structured projects. Profs. David McLaughlin, David Cai, and Aurel Lazar taught computational neuroscience classes that first got me interested in neuroscience, and working with Dr. Eliott Levy and Prof. André Fenton deepened my enthusiasm and made me consider that I might have something to offer. In particular, Eliott's patience and guidance is forming my project, and André's passion in lab meetings, gave me my first taste of the kinds of questions I would end up thinking about while at UCSB.

The NYU Society of Physics Students was an incredibly close knit group that became a home, an experience that I have learned to be unique. Dr. George Wong, Eric Kleinart,

Iraj Eshghi, Chris Ick, and Nick Verga all included me in the group and nurtured the community that became so integral to my time at NYU. Shiloh Pitt, Hark Kanwal, Sanchit Chaturvedi, and Kevin Zhuang all made me laugh and made many great memories. Kelly Wurtz was a very close friend and challenged me to become who I am today. She remains an inspiration for living as herself. Abby Jawahar was a friend when I needed it most. Dean Huang was an especially close friend that I have always felt at home with. Akshunna Dogra was my first friend at NYU and became a constant example for what putting myself out there could achieve. Our scientific collaborations during my time at UCSB has been a highlight and I look forward to all that we pursue together in the future.

The move to Santa Barbara was daunting, but one that I undertook for two reasons. First, the interdisciplinary nature and flexibility afforded by DYNS was incredibly appealing. I am proud to have taken full advantage of what DYNS has to offer while at UCSB, and as a result I got to meet some great faculty, colleagues, and friends. Profs. Julie Simpson and Andres Ludwigs' courses both stand out due to their well-thought out design and engaging material. Prof. Igor Mezić's courses opened up the world of Koopman operator theory to me. Igor's support and encouragement, as well as his integrity and devotion to doing meaningful, high impact science, has fundamentally impacted my research career. My work in the Koopman world has brought me in contact with other great researchers, such as Prof. Yannis Kevrekidis, Michael Mahoney, Omri Azencot, Ben Erichson, Enoch Yeung, Pietro Lio, Drs. Maria Fonoberova, Ryan Mohr, Boumediene Hamzi, and Paris Flood and Daniel Yeh. Additionally, the Koopman community's embrace of my work has led me to travel to Cancun, Mexico, Toronto, Canada, and Portland, Oregon, all of which have made my graduate school experience more rich.

The independence of DYNS additionally encouraged me to work on projects beyond purely systems neuroscience. Pursuing this, I have gotten a chance to work with Prof.

Atlas Wang, whose limitless energy for new questions and feedback, has been an inspiration. Like Igor, Atlas has proved a formidable mentor who has provided opportunities to engage with the research community in many different facets. My collaboration with Atlas also brought me to work with Dr. Tianlong Chen, whose output as a scientist is unrivalled, and who made sure to include me during ICML, as well as Prof. Sebastian Goldt and Dr. Alessandro Ingrosso, who blend physics, neuroscience, and machine learning in ways that inspire me.

Profs. Spencer Smith and Jeff Moehlis provided great direction to DYNS as chairs. The older DYNS graduate students always made me feel welcome, especially Drs. Arturo Deza and Josh Mueller. My cohort mates, Angela Bontempo, Sudhanshu Srivastava, and Justin Smith, have always made me feel not alone. Filip Tomasaka's initiative in building community in DYNS, as well as our fast friendship and cycling adventures, have been the source of many cherished memories. I was included into the PBS cohort as well, where I have found great friendship (and rivalry) with Jordan Garrett, who has put up with many of my pranks, but never ceases to hype up his friends and stay humble about his own accomplishments. I also became friends with Courtney and Ian Durdle, whose hospitality and expertise in organizing and including people has been a mainstay in my time at UCSB.

The second reason I made the decision to come to UCSB was because of my advisor Prof. Michael Goard. Everyone I talked to when I interviewed was excited to tell me that Michael was a smart and nice guy. After five years of working closely with Michael, I can say that these descriptors are true and more, and Michael's patience, guidance, and support allowed me to get the hands-on systems neuroscience experience I wanted out of graduate school. He has also inspired me as a father and husband, who makes his work life balance look easy. In addition to his role as my academic advisor, Michael has shaped my time at UCSB by prioritizing lab dynamic and maintaining an incredibly

fun and healthy lab culture. The Goard Lab defined my time in Santa Barbara, and my love for Dr. Kevin Sit, Nora Wolcott, Tyler Marks, Luis Franco, Fiona McBride, Santiago Acosta, Jasmine Yeo, and Dylan Martins is immense. There are far too many memories to share, but the weekly potlucks we had during the pandemic were a source of great comfort and deepend our friendship. Additionally, the members of our sister lab (Jacobs Lab), Laura Pritschet, Dr. Shuying Yu, Hannah Grotzinger, and Elle Murata joint potlucks that much more fun.

Daniel Godinez, while not a member of UCSB, brought me under his wing when I first moved to Santa Barbara, and our weekly BBQs were often the highlights of my week.

From my family, I have been given so much, including the opportunity to come to UCSB and spend the past five years living a dream life, becoming the person I never knew I always wanted to be. My grandparents have always encouraged my passions and made me feel supported. Mom has always been there for me, and has taught me to ask questions and do the little things that make other peoples lives better. Dad has been my greatest role model, and from whom my love of learning, humor, and strength comes from. And Emma, no matter what, has always been in my corner and my biggest fan, while never failing to amaze me at the funny, strong, smart person she has become.

Dr. Alyssa Lawson, who I remember first noticing when we interviewed together, puts the "partner" in partner by being supportive, caring, and loving, and making our relationship over the past 4+ years so incredibly easy and fun. Alyssa reminds me that there is more to life than work every time we laugh together.

Curriculum Vitæ

William Thomas Redman

Education

2023	Ph.D. in Dynamical Neuroscience (Expected), University of California, Santa Barbara.
2018	B.A. in Physics, New York University.
2018	B.A. in Mathematics, New York University.

Publications

2022	**W. T. Redman**, et al., "Long-term Transverse Imaging of the Hippocampus with Glass Microperiscopes". eLife.
2022	**W. T. Redman**, M. Fonoberova, R. Mohr, I.G. Kevrekidis, and I. Mezić, "Algorithmic (Semi-)Conjugacy via Koopman Operator Theory". IEEE Conference on Control and Decision (CDC).
2022	**W. T. Redman**, T. Chen, Z. Wang, and A. S. Dogra, "Universality of Winning Tickets: A Renormalization Group Perspective". International Conference on Machine Learning (ICML).
2022	**W. T. Redman**, M. Fonoberova, R. Mohr, I.G. Kevrekidis, and I. Mezić, "An Operator Theoretic View on Pruning Deep Neural Networks". International Conference on Learning Representations (ICLR).
2021	**W. T. Redman**, "On Koopman Mode Decomposition and Tensor Component Analysis". Chaos.
2020	A. S. Dogra* and **W. T. Redman***, "Optimizing Neural Networks via Koopman Operator Theory". Advances in Neural Information Processing Systems (NeurIPS). (*equal contribution)
2020	**W. T. Redman**, "Renormalization group as a Koopman operator". Physical Review E.
2019	**W. Redman**, "An *O(n)* method of calculating Kendall correlations of spike trains". PLOS ONE.

Publications in progress

2023	**W. T. Redman**, S. A. M. Acosta, X. X. Wei, and M. J. Goard, "Robust Variability of Grid Cell Orientation and Spacing Within Single Grid Module Enhances the Coding Capacity of Local Space". *In preparation.*

2023 **W. T. Redman**, D. Huang, M. Fonoberova, R. Mohr, and I. Mezić, "Deja Vu Prediction: Time Series Pattern Matching for Forecasting via Extended Dynamic Mode Decomposition with Memory". *In preparation.*

2023 K. F. Yei, P. D. L. Flood, **W. T. Redman**, and P. Lio, "Learning Linear Embeddings for Non-Linear Network Dynamics with Koopman Message Passing". *In preparation.*

2023 **W. T. Redman**, Z. Wang, A. Ingrosso, and S. Goldt, "Sparsity Enhances Non-Gaussian Data Statistics During Local Receptive Field Formation". *Under review.*

2023 **W. T. Redman**, J. M. Bello-Rivas, M. Fonoberova, R. Mohr, I. G. Kevrekidis, and I. Mezić, "Identifying Equivalent Machine Learning Algorithms". *Under review.*

2022 E. R. J. Levy, S. Carrillo-Segura, E. H. Park, **W. T. Redman**, J. Hurtado, A. A. Fenton, "A neuronal code for space in hippocampal coactivity dynamics independent of place fields". *Under review*

Abstract

Experimental and Computational Interrogation of the Hippocampal Formation

by

William Thomas Redman

The hippocampal formation plays an important role in spatial navigation and episodic memory. At the core of the neuroscience community's understanding of this role, are classes of functional cell types, such as place cells and grid cells. Despite being the subject of intense study for over half a century, how the complex recurrent circuitry of the hippocampal formation generates these functional classes, and how they interact with each other to support the critical cognitive processes the hippocampal formation underlies, is an area of open research. In this thesis, we present experimental and computational work aimed at addressing some aspects of these questions.

First, we develop a novel technique to optically access the transverse plane of the hippocampus *in vivo*, allowing – for the first time – the neural activity of multiple subregions of the hippocampus to be simultaneously recorded. We utilize this tool to characterize functional and structural properties across the hippocampal circuit, finding stable and unstable populations of dendritic spines on the apical dendrites of CA1 pyramidal neurons and heterogeneity in the amount of spatial information encoded by place cells along the CA1-CA3 axis. Second, we test a long held assumption on properties of grid cells in medial entorhinal cortex, analyzing recent cutting-edge experimental data. We find evidence for the distribution of grid spacings and orientations in individual modules to be non-uniform, having small, but significant variation. Computational modeling leads us to conjecture that a grid code with such heterogeneity in its properties enables robust encoding of local spatial information. And third, we study the emergence of localized

responses in artificial neural network models. We find that, as these local features form, the statistics of the internal representations of the network become increasingly non-Gaussian, in-line with theoretical work that has suggested this to be a general mechanism for driving localized responses. Sparsifying the network, via pruning or regularization, amplifies the non-Gaussian statistics, emphasizing the role of sparsification on internal representations. Each of these experimental and computational approaches motivates exciting avenues of future direction to shed light on the computations performed by the hippocampal formation.

Contents

Chapter 1

Introduction

"Little bit of this / Little bit of that"

Vince Staples (2016)

1.1 The Hippocampal Formation and Spatial Navigation

All organisms exist within their environment. This environment has affordances (i.e., what the environment offers an organism [1]), which can be accessed by performing specific actions, at specific locations in space. Examples include scavenging for food, courting a mate, and hiding from predators. In each case, the ability to efficiently navigate the environment is of great importance. This spatial navigation includes simple behaviors, such as the "run and tumble" strategy bacteria perform to explore new parts of their environment [2], as well as highly complex behaviors, such as migration patterns of bats that involves traveling thousands of miles to specific caves [3].

Given the range of organisms that spatially navigate, it is perhaps not surprising that there are many different strategies that nature has discovered for performing spatial navigation. One such approach is to maintain and store internal representations of physical space. These representations can be with respect to oneself, known as *egocentric* representations (e.g., the garage is to my left when I face my front door), or with respect to a fixed external reference frame, known as *allocentric* representations (e.g., my house is south of Los Padres mountains).

A major discovery in the past century of neuroscience, has been that there exist neural correlates that support both egocentric and allocentric representations [4, 5, 6, 7, 8, 9, 10], and in certain cases, conjunctively contain egocentric and allocentric information [11, 12, 13]. While allocentric encoding is transmitted broadly across the brain, being found in primary visual cortex [14, 15] and somatosensory cortex [16], it is believed to be generated by, and maintained in, the hippocampal formation. This is a brain region comprised of the hippocampus (HPC – itself partitioned into the dentate gyrus (DG) and cornu Ammonis (CA) subfields: CA1, CA2, and CA3), the subiculum, and the entorhinal cortex (EC). The size of the hippocampal formation has been found to vary in some animals in response to spatial navigation demands [17, 18, 19], and contains several different functional cell types that encode space, and features in space, in an allocentric manner. Below, we discuss some of these types of functional classes, as well as their properties, in more detail.

1.1.1 Place Cells

The first functional class of neurons discovered in the hippocampal formation was place cells (PCs) [20, 21]. These are neurons whose activity is local in space, and (largely) insensitive to the way in which an animal is facing (Fig. 1.1A). The region in which an

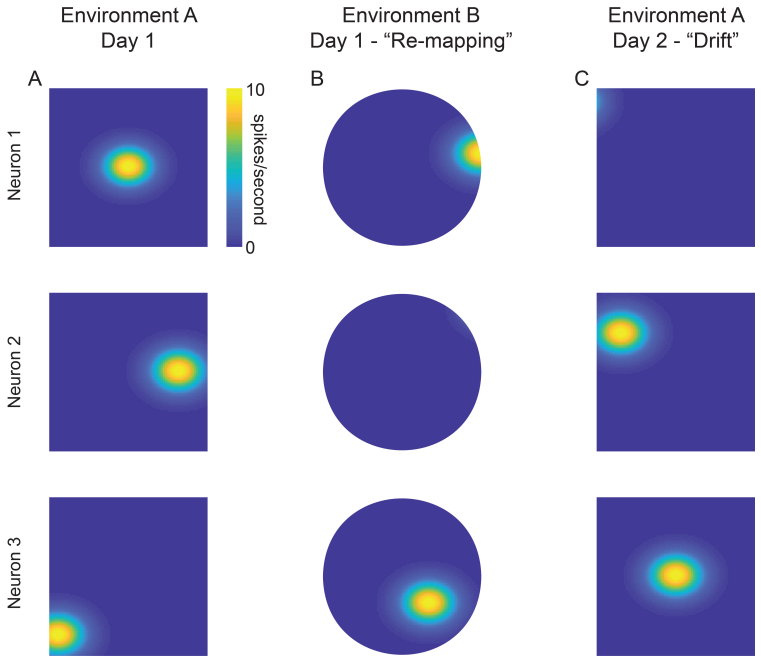

Figure 1.1: **Illustration of key properties of place cells.** (A) The activity of three "idealized" PCs in a rectangular environment (Environment A). Yellow color denotes 10 spikes per second, and blue color denotes 0 spikes per second. Note that the activity is localized, with respect to allocentric space. (B) The same three PCs' activity in a circular environment (Environment B). Note that the place fields move (or are absent, in the case of Neuron 2), illustrating "re-mapping". (C) The same three PCs' activity in Environment A, one day later. Note that the place fields have moved, illustrating representational drift.

individual PC fires action potentials is referred to as its place field. While PCs were historically characterized by having a single place fields, recording from the HPC in larger, more realistically sized environments has revealed that they can have multiple place fields [22, 23, 24, 25], each of which can have very different sizes [26], although the population distribution contains statistical structure [27].

The amount of spatial information an individual PC contains (a quantity that can be formally defined [28]) depends on several factors, including the size of the place fields and the maximum firing rate (relative to baseline) of a neuron. In general, the more local a place fields is, the more information it encodes about a single specific location, and the less information it encodes about the rest of space. Indeed, place cells are not

the most efficient way to encode space (a single neuron that has a different firing rate for all positions in space – at a given resolution – would suffice), however they are robust to noise (providing a faithful representation) and can be used to more easily associate affordances (such as rewards) with individual locations. The locality and specificity of PCs, which were not *a priori* guaranteed to be used by the brain, provided the first explicit evidence that allocentric information was represented at the level of individual neurons, and inspired the dominating hypothesis that the HPC generates and maintains a cognitive map [4] that can be used for spatial navigation.

PCs have been discovered in all major subregions of the HPC. While there do appear to be differences in the place fields properties among PCs in CA1, CA2, CA3, and DG [29, 30, 31], the fact that PCs are present throughout the HPC suggests that they are a fundamental "currency" of HPC computations.

Further evidence for map-like representations in the HPC came from the discovery that in distinct environments, with different visual stimuli, the encoding by PCs changes. In particular, PCs have place fields in different locations [32, 33], relative to the same environmental boundaries, a phenomenon termed "re-mapping" (Fig. 1.1B). The extent to which re-mapping occurs (i.e., what percentage of PCs change their place fields) has been extensively studied, with different subregions in the HPC having different sensitivities to alterations of the stimuli and geometric landmarks that make up a given environment [34, 35, 36]. These results have suggested distinct roles for each subregion in enabling navigation in similar environments.

Additional support for the hippocampus as a cognitive map came from the fact that early experiments to record from PCs in CA1 found their place fields to be stably present across time [37, 38]. Subsequent work found that the stability of place fields was controlled by NMDA receptors [39], and that attention could increase stability [40]. However, experiments over the past five years with modern technology have cast uncertainty on

the stability of place fields in general environments. Indeed, many studies describe instability in place fields [41, 42, 43, 44, 45, 46], although the exact nature of this instability differs (Fig. 1.1C) [41, 42, 44, 43, 45]. Due to the fact that this "representational drift" appears to be present in sensory cortices [47, 48, 49], as well as higher order areas, like retrosplenial and parietal cortex [50, 51], there has been considerable interest in studying how information encoded by local and specific neural responses can be stably extracted in the face of unstable tuning [52, 53, 54].

1.1.2 Grid Cells

While PCs in the HPC encode a different kind of information than neurons in sensory cortices, such as visually responsive cells in primary visual cortex, they do so in similar ways (i.e., a single localized receptive field in visual or physical space). It was then of major surprise when a new functional class was found in medial entorhinal cortex (mEC), which responded to multiple locations in physical space [6]. Because these neurons responded to positions located on an allocentric triangular lattice (and thus, coined "grid cells" – GCs) (Fig. 1.2A), they contain considerable geometric structure, making them unlike the PCs with multiple place fields discovered later [22, 23, 24, 25, 26].

While GCs have multiple firing fields (or grid fields – grid fields), their responses can be described by three different quantities (Fig. 1.2B): 1) the distance between individual grid peaks, called grid spacing (λ); 2) the angular orientation of the grid, called grid orientation (θ); and 3) the displacement of the grid relative to the center of the arena, called the grid phase (ϕ). While grid cells of all phases have been found in mEC, grid spacing and orientation have been found to take discrete sets of values, suggesting a modular organization [55] (Fig. 1.2C). These modules are roughly arranged along the dorsal-ventral axis, with grid spacing becoming increasingly large (from under half a

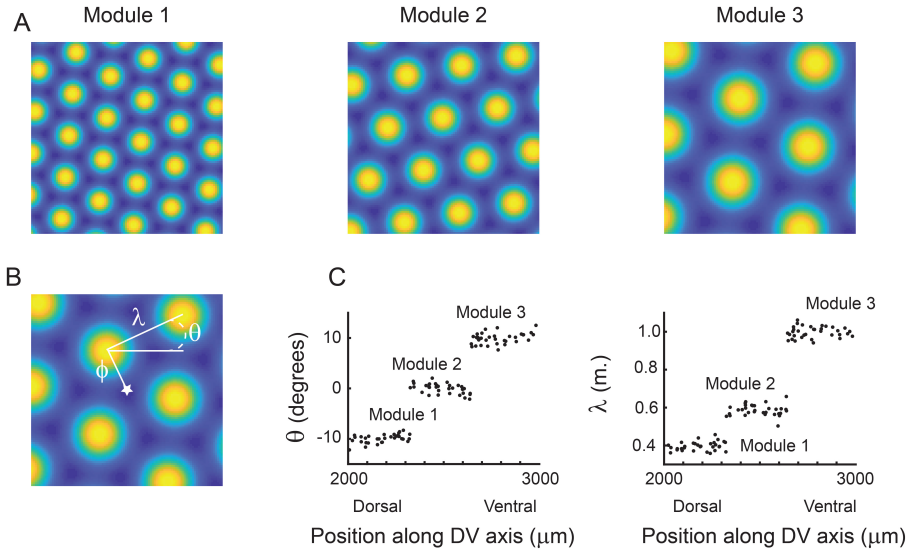

Figure 1.2: **Illustration of key properties of grid cells.** (A) Three "idealized" grid cell firing fields, each of which is from a different module. (B) Overview of the three quantities that describe grid cells: grid spacing (λ); grid orientation (θ); and grid phase (ϕ). Star denotes center of arena. (C) Illustration of grid properties taking on discrete values as a function of dorsal ventral position.

meter to over a meter) as the position becomes increasingly ventral. Theoretical work has shown that, while individual GCs has a degeneracy in their encoding of space, integration of information across multiple grid modules enables accurate and efficient representation of local space [56, 57, 58, 59, 60].

The triangular lattice structure of the grid fields has inspired considerable theoretical interest, and many models have been put forth on how this property might be implemented in neural circuits, such as those that make use of PCs [61] and finer resolution theta fields [62]. More recent work has focused on pattern formation mechanisms [63, 64, 65]. Beyond the formation of GCs, there has also been significant attention paid to how GCs representations can be maintained. These have primarily put relied on continuous attractor dynamics [66, 67, 68, 69], which has found experimental support [70, 71, 72].

Given the strong connectivity between mEC and HPC (in both directions), work has

been performed in understanding how representations, especially PCs and GCs, interact and whether one emerges from the other. In particular, when GCs were first discovered, a simple linear model, with inhibition, was shown to suffice to transform GCs, across multiple modules, to PCs [73, 74]. This is aligned with experimental results that have found the scale of grid fields impact the scale and stability of place fields [75, 76], as well as findings that GCs make up the largest percent of inputs to PCs [77]. However, PCs have been found to emerge before GCs during development [78, 79] and mEC lesioning does not entirely destroy the place code of HPC [80], implying that GCs cannot be the sole generator of PCs. There have also been proposals that PCs and GCs are learned in the same manner using spike-timing dependent plasticity (STDP), with GCs or PCs emerging depending on the smoothness properties of excitatory and inhibitory inputs [63].

Beyond PCs, GCs have been suggested to play an important role in path integration [67] (i.e., the process of maintaining and updating one's own direction and distance travelled along a path). This is because the triplet (λ, θ, ϕ) describing each GC can be turned into a measure of distance and angle. Further support for this theory has come from the discovery that artificial recurrent neural networks (RNNs), trained to perform path integration, can generate internal representations that include grid cells [81, 82, 64, 83]. This suggests that the need for path integration is a strong inductive bias on the neural circuits that are tasked with performing it.

1.1.3 Other Functional Allocentric Cell Types

While PCs and GCs are the primary functional cell types studied in the hippocampal formation, there exist several other classes of allocentric responses. These include border cells, band cells, and spatially informative non-place cells.

Border cells. An important determinant of the affordances, as well as the "dis-affordances", an environment provides are its boundaries. Cells responsive to borders of an environment have been found in mEC [7, 84, 85], among other brain regions. These border cells have been found to be present earlier in development than GCs [86], suggesting a primacy in their representations for spatial navigation.

Band cells. In many environments, there is a symmetry along the x-axis or the y-axis, with respect to certain affordances. For example, along a hallway, the doors are spaced along the y-axis. In such settings, maintaining a representation of position along the x-axis (to keep track of how close you are to the doors) and the y-axis (to keep track of what door you are near) is useful. In-line with this intuition, cells in the mEC that fire symmetrically along an axis have been discovered [8]. These so called band cells have been subject to little study, so beyond this little is known about them.

Spatially informative non-place/grid cells. As noted earlier, while coding space via PCs is interpretable (a single subset of cells active for a single location) and robust to noise, it is not the only way to encode space. Indeed, cells that reliably respond differently at different parts of allocentric space can, in theory, be used for decoding space by themselves. Recordings in mEC found that a majority of cells are not classifiable as border cells or GCs, but do contain significant and robust encoding of space [87]. Similarly, in HPC, it was found to be possible to achieve similar performance on approximating the position of an animal from a decoder built using the responses of neurons that had the most PC like responses and the least PC like responses [88]. Additionally, considering the activity of non-PCs was found to contain information about the responses of PCs, including allowing for prediction of when a PC would not fire in its place fields [89]. Collectively, this suggests that allocentric spatial encoding in mEC and HPC is a broader phenonomenon and not restricted so purely PCs, GCs, border cells, or band cells.

1.2 The Hippocampal Formation and Episodic Memory

Just as all organisms exist within their environment, all episodic memories must similarly exist within the context of an environment. Thus, many of the memories most necessary for survival (e.g., the location of food, the location of home, the location of mates) inextricably involve a notion of space. That the hippocampal formation, which encodes spatial information, similarly has been found to play an important role in episodic memory [90, 91], makes this point only more salient.

Research implicating the role of HPC in episodic memory has taken two main approaches. The first is to remove or inactivate the HPC, and examine the effects on tasks that require episodic memory (early rodent literature on this topic is extensively reviewed in [4]). In rare cases, this lesion occurs in humans, either by elective experimental surgery or disease. Cases that particularly captured the imagination of the scientific community, as well as the general public, were patients H.M. [92, 93] and E.P. [94]. In both cases, the bilateral loss of the HPC led to an irrecoverable and nearly complete loss of the ability to make and store new memories. Memories occurring before the loss, such as those in childhood, were left intact.

The second approach, made possible through the development of systems neuroscience tools, has been to create "false" memories in the HPC. In particular, the dominant theory of how neural circuits are involved in memory holds that ensembles, or engrams [95, 91], of cells are responsible for the encoding of individual memories. A corollary of this is that direct re-activation of the correct ensemble of cells should lead to a recollection of the memory. Revolutionary work did just that, re-activating neurons in the DG that were active during fear conditioning, eliciting a freezing response in rodents in a neutral environment [96, 97, 98, 99].

1.2.1 Place and Grid Cells Beyond Space

Just as the HPC and mEC are used in both spatial and episodic memory, there is evidence that the same functional cell classes are used as well. That is, that PCs and GCs encode information along feature dimensions beyond space.

First, just as locations in space are at certain distances from each other, so too are events in time. For instance, yesterday is a lot "closer" to today than three years ago. This temporal metric plays an important role in episodic memory, as it provides contextual information, such as the ordering of events, which makes their causes and consequences more interpretable. Recordings from HPC and mEC have found that neurons that are PCs and GCs can also encode the timing between events [100, 101, 102, 103, 104].

In addition to pure time, foundational work has shown that PCs and GCs also encode task relevant, "abstract" information [105]. This was demonstrated by training rats to pull a lever, while a pure tone underwent a frequency modulation, when the tone was within a specific frequency range. The researchers found that PCs and GCs encoded one, or multiple, tones in the sweep, and slowing the sweep widened the responses of these tone responsive cells, demonstrating that they were not encoding time.

Another property of PCs, and GCs to a lesser extent, that has been interpreted as a signature of the hippocampal formation playing a more general role in memory is the phenomena of re-mapping. Indeed, the initial interpretation of the discovered change of the place fields in different environments was that it was the HPC circuit recognizing a change in context [32]. Such contextual cuing is essential for correct memory recall. This has been recently formalized in a computational theory of hidden state inference [106]. Additionally, recent work has shown that remapping can be thought of more generally as a change in the correlation structure of neurons in HPC, with different environments causing changes in the way in which pairs of neurons were correlated [107].

These results have led to the hypothesis that PCs might be an emergent property of a computational system that compresses incoming sensory information in such a way that maximizes the distinguishability of it with other, previously stored, memories [108]. Various model implementations of such a system led to place-like tuning when the memories being encoded were linked to spatial positions. Finally, another interpretation of many disparate results from the PC and GC literature has been that of the predictive map [109]. In this framework, representations of space becomes directly connected to known and/or remembered affordances, as the HPC is used to encode the expected future rewards at a given location. This inductive bias was found to be sufficient to lead to the formation of PCs, and other PC properties.

1.3 Learning Localized Responses from Data

The growing appreciation for the multifaceted role PCs and GCs play in cognition, beyond pure spatial navigation, has led to an increased appreciation for the computational advantages a (sub-)population of neurons, with responses localized over the possible values a given abstract variable can take, provides. In particular, PCs and GCs are a "useful" representation, not only because they enable a simple GPS-like read-out of allocentric space, but also because they can be used to learn temporal orderings [110, 111], predict future outcomes [109], and form hierarchies [27]. In addition, such codes are sparse (in that only a subset of neurons are active at a given time), and therefore, are more efficient [112].

One of the first neural codes to be discovered as having with localized responses was in primary visual cortex (V1), where a given visually driven neuron responds only when a stimulus is present in a specific part of visual space [113]. This region is called the receptive field (RF) of the neuron. While there are many differences between PCs, GCs,

and V1 neurons with RFs, in each case, the neurons respond only in a local neighborhood of all the possible visual or allocentric positions. Note that this differs from the representation present in spatially informative non-place/grid cells [87, 88], where the responses are non-localized.

While RFs in V1 are largely due to hard-wired connectivity patterns that emerge during development, the slower time-scale over which PCs and GCs emerge [78, 79], as well as the experience dependent nature of PC formation [114], suggest that the localized responses in the hippocampal formation are learned and adapted. How exactly this occurs, is not well understood.

Beyond elucidating fundamental properties of neural circuits, especially those present in HPC and mEC, an understanding of how localized responses can emerge will lead to better machine learning (ML) methods. Indeed, convolutional neural networks (CNNs) [115, 116, 117], whose development, along with increasing computational resources and curated data sets, led to the first wave of breakthroughs in deep learning (DL), include in their architectural design a form of local RFs [118, 119]. This encourages the representations of CNNs to be invariant or equivariant to translations [120, 121, 122], thereby respecting a fundamental symmetry present in the images they are trained on. This makes CNNs more sample- and parameter-efficient than fully-connected networks (FCNs), which do not take this symmetry into account explicitly [123, 124].

As deep learning moves into ever more diverse applications in science and beyond, it is confronted with data sets whose symmetries are not necessarily known *a priori*. It would thus be desirable to "learn" the localized responses directly from data. Below, we discuss some approaches that have been able to learn localized responses in both biologically plausible, as well as artificial, neural networks, from data.

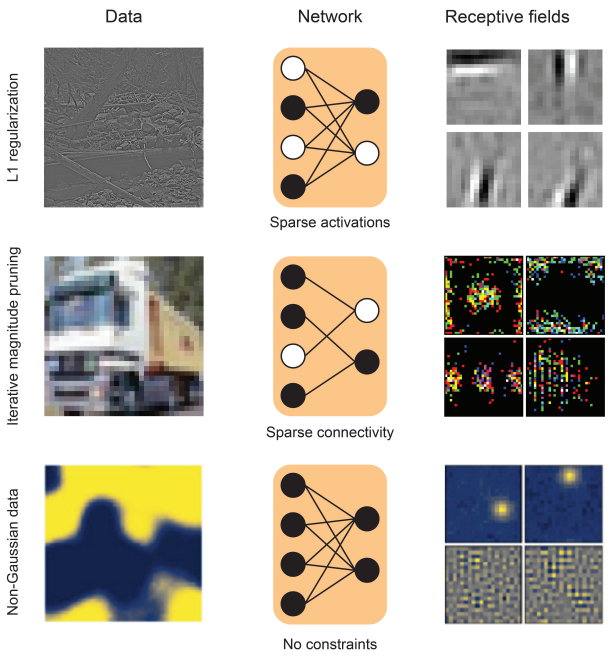

Figure 1.3: **Illustration of approaches that have learned local receptive fields from data.** (Top row) Learning localized RFs via L1 regularization on the activations [125]. (Middle row) Learning localized RFs via iterative magnitude pruning [127]. (Bottom row) Learning localized RFs via non-Gaussian data statistics [128]. (Left column) Data each method was trained on. (Middle column) Schematic of network implementation, with constraints listed below. (Right column) Example receptive fields. Data and receptive fields of bottom row taken from [128]. Other figures reproduced from code.

1.3.1 Learning Local Receptive Fields

Foundational work showing that a FCN, with no architectural constraints, could learn (in an unsupervised manner) local RFs when a penalty (i.e. regularizer) enforcing sparse activity was added [125] (Fig. 1.3 top row). That a global regularization on network activity could lead to local responses led to the hypothesis that metabolically efficient coding [126] is an inductive bias strong enough to shape the local neural responses found in the brain [112].

Despite the early success, learning local RFs from "scratch" on larger FCNs has proven difficult with stochastic gradient descent (SGD) on image classification tasks

[124, 129]. It has only been within the past few years that new approaches have been developed, including a more complex regularization technique [130] and iterative weight sparsification (or "pruning" – Fig. 1.3, middle row) [127], that have enable (supervised) learning of RFs.

Unlike past approaches, which have relied on modifications to the learning process (via regularization or pruning), local RFs can emerge in FCNs when trained in an online manner on synthetic images that are translation-invariant and contain sharp edges [128] (Fig. 1.3, bottom row). The signal in the data that triggers the formation of the localized receptive field is carried by higher-order (i.e., non-Gaussian) statistics of the data, and the representations present in the activation of the units of the neural network become distributed in an increasingly non-Gaussian manner.

1.3.2 Learning Place and Grid Cells

As discussed in Chapter 1.1.2, there exist many models that can generate PCs and GCs, to various degrees of biological plausibility [131, 132, 133, 66, 73, 134, 135, **?**, 67, 136, 137, 138, 69, 139, 140, 141, 61, 142, 143, 63, 81, 82, 64, 107, 83, 65]. While the GC models have been, to some extent, unified under the lens of pattern formation [64, 83, 65], the connection between various different PC models is less understood. Additionally, while some of the models have been shown to be capable of learning weights, via synaptic plasticity rules, such that exhibit PC tuning, other models rely on hand-tuned choice of weights. Alternatively, some models assume all the weights to take the same value. Such models are inherently limited in their biological plausibility.

1.4 Summary and Motivation

The hippocampal formation has been appreciated by the neuroscience community as playing an integral role in supporting episodic and spatial memory for over half a century. Decades of neurophysiology have identified neural correlates to these cognitive processes the hippocampal formation is supporting: individual neurons encode information about spatial locations, as well as abstract variables that are related by some metric. However, how this information, which is distributed across the mEC and HPC, as well as other brain regions, is integrated at the circuit level is still largely unknown. Here, we use a wide range of experimental and computational techniques to study some of the facets of this open question. First, we developed a novel technique for optically imaging the transverse plane of the HPC *in vivo*, enabling (for the *first time*) the simultaneous recording of all three primary subregions of the HPC. Next, we re-analyze state-of-the-art neural recordings of mEC, finding evidence for greater heterogeneity in the responses properties of GCs, which we show via computational models, enables significantly more information about local space than previously thought possible. Finally, we examine computational models that give rise to localized activity (like that found in PCs), showing how sparsity interacts with the statistics of the internal representations to give local RFs. These diverse studies, provide new ways in which to interrogate and understand the complex ways in which the hippocampal formation plays an key role in episodic and spatial memory.

1.5 Permissions and Attributions

1. The contents of Chapter 2 have been published in the journal *eLife* [144] and are the result of collaboration with Prof. Michael Goard, Prof. Spencer Smith, Nora Wolcott, Luca Montelisciani, Gabriel Luna, Tyler Marks, Dr. Kevin Sit, and Dr.

Che-Hang Yu.

2. The contents of Chapter 3 are in preparation for submission and are the result of collaboration with Prof. Michael Goard, Prof. Xuexin Wei, and Santiago Acosta.

3. The contents of Chapter 4 are under review at NeurIPS and are the result of collaboration with Prof. Sebastian Goldt, Dr. Alesandro Ingrosso, and Prof. Zhangyang "Atlas" Wang.

Chapter 2

Long-term Transverse Imaging of the Hippocampus with Glass Microperiscopes

Now here's something we hope you'll really like

Rocket J. Squirrel (1964)

2.1 Introduction

The hippocampus is critical for episodic and spatial memory formation [4, 90, 91], but the neural computations underlying these functions are not well understood. The trisynaptic circuit linking entorhinal cortex (EC) to dentate gyrus (DG), DG to CA3, and CA3 to CA1 is believed to endow the hippocampus with its functional capabilities. Since the circuit was first described in the anatomical studies of Ramon y Cajal over a century ago [145], considerable work has focused on each of the major hippocampal subfields

(CA1-3 and DG) to identify their roles in hippocampal processing. The resulting body of literature has indicated that the subfields have related, but distinct roles in pattern separation and completion [146, 147, 34, 35, 36, 148, 149, 150, 151, 152], response to novelty [153, 154, 155, 156, 157] and the encoding of social variables [158, 159, 160]. Additionally, there appear to be differences between the subfields in the stability of their place fields [157, 41, 44]. Although this work has increased our understanding of each of the subfields individually, it is not yet clear how neuronal activity is coordinated across the hippocampus.

This lack of knowledge comes, in part, from the technological limitations that prevent the recording of neuronal populations across hippocampal subfields in the same animal. Historically, electrophysiology has been the principal tool used to study the hippocampus. Electrophysiological recordings have the advantage of high temporal resolution and can directly measure spiking, but they are typically limited to small numbers of neurons in particular subfields. Additionally, localization of recorded neurons within the hippocampus is approximate, requiring post-hoc histological analysis to estimate the position of the electrode tracks, with the distances between the recorded neurons and the electrode sites being poorly defined. In recent years, calcium imaging approaches (i.e. single photon mini-endoscopes and two-photon microscopy) have been used to record hippocampal activity, allowing for the simultaneous measurement of large numbers of neurons with known spatial relationships [44, 161, 162, 42, 163, 164]. However, these approaches require aspiration of the overlying neocortex and are generally restricted to a single subfield for each animal.

Taken together, current experimental techniques are limited in their ability to: (1) record response dynamics and coordination across the hippocampus, (2) identify and distinguish between different neural subtypes, (3) allow for the chronic recording of cells across subfields, and (4) resolve key cellular structures, such as apical dendrites.

To address these challenges, we have developed a procedure for transverse imaging of the trisynaptic hippocampal circuit using chronically implanted glass microperiscopes. As has been found with previous studies using implanted microprisms in cortex [165, 166, 167], the neural tissue remained intact and healthy for prolonged periods of time (up to 10 months), and both dendritic structure and calcium activity could be repeatedly measured in awake mice. Optical modeling and point spread function measurements showed that axial resolution is decreased compared to traditional cranial windows, but is sufficient to image individual apical dendritic spines in hippocampal neurons several millimeters below the pial surface. Using this approach, we quantified spine turnover in CA1 apical dendrites across days. We then measured functional responses from CA1, CA3, and DG in head-fixed mice as they explored a floating carbon fiber arena [168, 169]. We found neurons in all regions whose activity met criteria to be considered place cells and speed cells. Further, we found a non-uniform distribution of spatial information across the extent of the DG-to-CA1 axis, supporting findings from earlier electrophysiological studies that found similar heterogeneity [29, 30, 31]. Taken together, this approach adds to the existing neurophysiological toolkit and will enable chronic structural and functional measurements across the entire transverse hippocampal circuit.

2.2 Methods

2.2.1 Animals

For dendritic morphology experiments, Thy1-GFP-M (Jax Stock #007788) transgenic mice (n = 7) were used for sparse expression of GFP throughout the forebrain. For forebrain-wide calcium indicator expression, Emx1-IRES-Cre (Jax Stock #005628) × ROSA-LNL-tTA (Jax Stock #011008) x TITL-GCaMP6s (Jax Stock #024104) triple

transgenic mice (n = 2) or Slc17a7-IRES2-Cre (Jax Stock #023527) × TITL2-GC6s-ICL-TTA2 (Jax Stock #031562) double transgenic mice (n = 6) were bred to express GCaMP6s in excitatory neurons. For imaging experiments, 8-51 week old (median 17 weeks) mice of both sexes (6 males and 9 females) were implanted with a head plate and cranial window and imaged starting 2 weeks after recovery from surgical procedures and up to 10 months after microperiscope implantation. The animals were housed on a 12 hr light/dark cycle in cages of up to 5 animals before the implants, and individually after the implants. All animal procedures were approved by the Institutional Animal Care and Use Committee at University of California, Santa Barbara.

2.2.2 Surgical Procedures

All surgeries were conducted under isoflurane anesthesia (3.5% induction, 1.5 - 2.5% maintenance). Prior to incision, the scalp was infiltrated with lidocaine (5 mg kg-1, subcutaneous) for analgesia and meloxicam (2 mg kg-1, subcutaneous) was administered preoperatively to reduce inflammation. Once anesthetized, the scalp overlying the dorsal skull was sanitized and removed. The periosteum was removed with a scalpel and the skull was abraded with a drill burr to improve adhesion of dental acrylic.

For hippocampal imaging, we used two types of custom-designed glass microperiscope (Tower Optical). The first ($v1_{CA1}$), for imaging the upper part of the hippocampus (CA1/CA2) consisted of a 1 × 1 × 1 mm square base and a 1 mm right angle prism, for a total length of 2 mm on the longest side (Fig. 2.1B, left). The second ($v2_{HPC}$), for imaging the entire transverse circuit (CA1-3, DG) had a 1.5 × 1.5 × 1.0 mm (L × W × H) square base and a 1.5 mm right angle prism, for a total length of 2.5 mm on the longest side (Fig. 2.1B, right). The hypotenuse of the right angle prisms were coated with enhanced aluminum for internal reflectance. The microperiscope was attached to a 5 mm

diameter coverglass (Warner Instruments) with a UV-cured optical adhesive (Norland, NOA61). Prior to implantation, the skull was soaked in sterile saline and the cortical vasculature was inspected to ensure that no major blood vessels crossed the incision site. If the cortical vasculature was suitable, a 3-4 mm craniotomy was made over the implantation site (centered at 2.2 mm posterior, 1.2-1.7 mm lateral to Bregma). For the smaller microperiscope ($v1_{CA1}$), a 1 mm length anterior-to-posterior incision centered at -2.1 mm posterior, 1.2 mm lateral to Bregma was then made through the dura, cortex, and mediodorsal tip of the hippocampus to a depth of 2.2 mm from the pial surface with a sterilized diamond micro knife (Fine Science Tools, #10100-30) mounted on a manipulator (Fig. 2.2A). For the larger microperiscope ($v2_{HPC}$), two overlapping 1.0 mm length anterior-to-posterior incisions were made centered at -1.8 mm posterior / 1.7 mm lateral and -2.4 mm posterior / 1.7 mm lateral to Bregma to a depth of 2.7 mm, with a total anterior-to-posterior incision length of 1.6 mm. Note that placements in the regions shown in Fig. 2.2B required incision coordinates slightly posterior to those indicated on the atlas. Care was taken not to sever any major cortical blood vessels. Gelfoam (VWR) soaked in sterile saline was used to remove any blood from the incision site. Once the incision site had no bleeding, the craniotomy was submerged in cold sterile saline, and the microperiscope was lowered into the cortex using a manipulator, with the imaging face of the microperiscope facing lateral (Fig. 2.2B). Once the microperiscope assembly was completely lowered through the incision until the coverglass was flush with the skull, the edges of the window were sealed with silicon elastomer (Kwik-Sil, World Precision Instruments), then with dental acrylic (C&B-Metabond, Parkell) mixed with black ink (Fig. 2.2C). Care was taken that the dental cement did not protrude over the window, as it could potentially scratch the objective lens surface. Given the working distance of the objective used in this study (3 mm), the smaller microperiscope ($v1_{CA1}$) implant enabled imaging from 2250 - 2600 μm below the coverglass surface, corresponding to

approximately 150 - 500 μm into the lateral hippocampus (the 150 μm of tissue nearest to the microperiscope face was not used for imaging). The larger microperiscope ($v2_{HPC}$) implant enabled imaging from 2650 - 2850 μm below the coverglass surface, corresponding to approximately 150 - 350 μm into the lateral hippocampus (the 150 μm of tissue nearest to the microperiscope face was not used for imaging). The microperiscope implantations were stable for up to ten months following the surgery (Fig. 2.8E).

After microperiscope implantation, a custom-designed stainless steel head plate (eMachineShop.com) was affixed using dental acrylic (C&B-Metabond, Parkell) mixed with black ink. After surgery, mice were administered carprofen (5 - 10 mg kg-1, oral) every 24 hr for 3 days to reduce inflammation. Microperiscope designs and head fixation hardware are available on our institutional lab web site (`https://goard.mcdb.ucsb.edu/resources`).

2.2.3 Point Spread Function Measurements

To measure empirical point spread functions, fluorescent microspheres (0.2 μm yellow-green fluorescent microspheres; ThermoFisher F8811) were embedded 1:2000 in 0.5% agar and placed under the cranial window or on the face of the microperiscope. Image stacks were taken through the microspheres (0.07 μm per pixel in XY; 0.5 μm per plane in Z) located 50-100 μm from the microperiscope face (except for the microsphere-microperiscope distance comparisons in Fig. 2.4J-K, which ranged from 25-500 μm from the microperiscope face). Candidate microspheres were initially isolated using the FindCircles function in the Image Segmenter app (MATLAB image processing toolbox). Only microspheres that were >25 pixels (1.7 μm) away from nearest neighbor microspheres and were completely contained within the Z-stack were used for further analysis. We registered isolated microspheres at their centroids and measured the full width at half

maximum (FWHM) of the average XY and XZ profiles to determine the lateral and axial resolution, respectively.

Since the geometry of the microperiscope limits the angle of the focusing light cone through the microperiscope, it predominately determines the functional numerical aperture at the imaging plane. Based on the microperiscope geometry, we calculated the effective NA of the $\mathrm{v1_{CA1}}$ microperiscope and $\mathrm{v2_{HPC}}$, and used it to calculate the theoretical lateral and axial resolution according to the following formulae [**?**]:

$$\omega_{XY} = \begin{cases} \frac{0.320\lambda}{\sqrt{2}NA} & NA < 0.7 \\ \frac{0.325\lambda}{\sqrt{2}NA^{0.91}} & NA > 0.7 \end{cases} \tag{2.1}$$

$$\omega_Z = \frac{0.532\lambda}{\sqrt{2}} \left[\frac{1}{n - \sqrt{n^2 - NA^2}} \right] \tag{2.2}$$

where ω_{XY} and ω_Z are the theoretical $1/e$ widths of the lateral and axial point spread function, λ is the wavelength, and NA is the numerical aperture. Note that the FWHM was calculated by multiplying the $1/e$ width (ω) by $2\sqrt{\ln(2)}$.

To perform aberration correction with adaptive optics, a deformable mirror (Multi-3.5, Boston Micromachines Corporation) was set at a plane conjugate to the raster scanning mirrors and the back aperture of the objective lens in the two-photon imaging system. Fluorescent microspheres (0.2 μm) were imaged, and the standard deviation of the image brightness was maximized under different configurations of the deformable mirror. Twelve selected Zernike modes are applied and modulated sequentially over a total of three rounds. The 12 zernike modes are: 1) oblique astigmatism, 2) vertical astigmatism, 3) vertical trefoil, 4) vertical coma, 5) horizontal coma, 6) oblique trefoil, 7) oblique quadrafoil, 8) oblique secondary astigmatism, 9) primary spherical, 10) vertical secondary astigmatism, 11) vertical quadrafoil, 12) secondary spherical. For each

Zernike mode, 21 steps of amplitudes were scanned through, and images were acquired for each step. The amplitude that resulted in the largest standard deviation was saved and set as the starting point of the DM configuration for the scanning of the next Zernike mode. The brightness, the lateral resolution, and the axial resolution are compared with and without the application of the deformable mirror correction.

2.2.4 Air-Floated Chamber

For measurement of spatial responses, mice were head-fixed in a floating carbon fiber chamber [168] (Mobile Homecage, NeuroTar, Ltd). The chamber base was embedded with magnets to allow continual tracking of the position and angular displacement of the chamber. Behavioral data was collected via the Mobile HomeCage motion tracking software (NeuroTar, versions 2.2.0.9, 2.2.014, and 2.2.1.0 beta 1). During imaging experiments, image acquisition was triggered using a TTL pulse from the behavioral software to synchronize the timestamps from the 2P imaging and chamber tracking.

A custom carbon fiber arena (250 mm diameter) was lined with four distinct visual patterns (5.7 cm tall, 18.1 cm wide) printed on 7 mil waterproof paper (TerraSlate) with black rectangles (5.7 cm tall and 1.5 cm wide) placed in between the four patterns. A circular track (Fig. 2.11A; Video S4) was made by adding a removable inner circle (14 cm in diameter and 4.2 cm tall) with visual cues that were matched to the outer wall printed on 7 mil waterproof paper. The resulting circumference, along the middle of the circular track, is 61.26 cm. Transparent tactile stickers (Dragon Grips) were placed on the arena floors to give differential tactile stimuli along the track. In between each recording and or behavioral session, the arena walls and floors were thoroughly cleaned.

Mice were acclimated to the arena by the following steps: 1) On the first day the mice were placed into the chamber and allowed to freely explore without head fixation

for 15 - 20 minutes. A piece of plexiglass with holes drilled through was placed on top of the arena to keep the mice from climbing out. 2) On the second day, the mice were head-fixed to a crossbar extending over the floating chamber (Fig. 2.11A) and allowed to freely explore the floating chamber freely for 15 minutes. Air flow (3 - 6.5 psi) was adjusted to maximize steady walking/running. On subsequent days, the head fixation time was increased by increments of 5 minutes, as long as the mice showed increased distance walked and percent time moving. This was continued until the mice would explore for 30 - 40 minutes and run for greater than 15% of the time. 3) Mice were head-fixed in the floated chamber for 20 minutes with a custom light blocker attached to their headplate. 4) Mice were head-fixed and placed on the 2P microscope to allow habituation to the microscope noise. 5) After mice were fully habituated, 20 - 40 minute duration recording sessions on the 2P microscope were performed.

If at any point during the above acclimation protocol the mouse significantly decreased distance traveled or percentage of time moving, then the mouse was moved back to the previous step.

Custom software was written to process the behavioral data output by the Mobile HomeCage motion tracking software. Because the Mobile HomeCage motion tracking software sampling rate was faster than the frame rate of our 2P imaging, all behavioral variables (speed, location, polar coordinates, and heading) that were captured within the acquisition of a single 2P frame were grouped together and their median value was used in future analysis. For the polar angle (which we used as the location of the mouse in 1D track), the median was computed using an open source circular statistics toolbox (CircStat 2012a) written for MATLAB [170]. We removed any time points when the mouse was not moving, as is standard for measurement of place fields [161]. This helps separate processes that are related to navigation from those that are related to resting state. To do this, we smoothed the measured instantaneous speed and kept time periods

>1 s that had speeds greater than 20 mm/s (adding an additional 0.5 s buffer on either side of each time period).

2.2.5 Two-Photon Imaging

After recovery from surgery and behavioral acclimation, GFP or GCaMP6s fluorescence was imaged using a Prairie Investigator 2P microscopy system with a resonant galvo scanning module (Bruker). For fluorescence excitation, we used a Ti:Sapphire laser (Mai-Tai eHP, Newport) with dispersion compensation (Deep See, Newport) tuned to λ = 920 nm. Laser power ranged from 40-80 mW at the sample depending on GCaMP6s expression levels. Photobleaching was minimal (<1% min^{-1}) for all laser powers used. For collection, we used GaAsP photomultiplier tubes (H10770PA-40, Hamamatsu). A custom stainless-steel light blocker (eMachineShop.com) was mounted to the head plate and interlocked with a tube around the objective to prevent light from the environment from reaching the photomultiplier tubes. For imaging, we used a $16\times/0.8$ NA microscope objective (Nikon) to collect 760 $\times$ 760 pixel frames with field sizes of 829 $\times$ 829 μm or 415 x 415 μm. Images were collected at 20 Hz and stored at 10 Hz, averaging two scans for each image to reduce shot noise.

For imaging spines across days, imaging fields on a given recording session were aligned based on the average projection from a reference session, guided by stable structural landmarks such as specific neurons and dendrites. Physical controls were used to ensure precise placement of the head plate, and data acquisition settings were kept consistent across sessions. Images were collected once every day for 5 - 10 days.

2.2.6 Two-Photon Post-Processing

Images were acquired using PrairieView acquisition software (Bruker) and converted into multi-page TIF files.

For spine imaging, registration and averaging was performed for each z-plane spanning the axial width of the dendrite to ensure all spines were captured across z-planes. The resulting projections were weighted according to a Gaussian distribution across planes. Non-rigid registration was used to align dendritic segments across consecutive recording sessions. The registered images underwent high-pass filtering to extract low amplitude spine features using code adapted from Suite2P's enhanced mean image function [171] (Fig. 2.5-Fig. 2.2A). The resulting ROIs were binarized using Otsu's global threshold method for spine classification (Fig. 2.5-Fig. 2.2A). In most cases, the global threshold successfully isolated the single most prominent dendrite. In fields with higher background dendrites that were not desired, these extraneous dendrites were manually excluded. To identify spines that fall below the global threshold, the user manually specifies incrementally lower thresholds from which to select spines that were excluded in the initial binarization. (Fig. 2.5-Fig. 2.2A). Spines above the global threshold with an area of > 1 μm^2 were included in our analysis. To classify each spine as one of the four major morphological classes, we performed the following steps. First, we found the base of the spine by identifying the region closest to the dendritic shaft. Second, we calculated the length of the spine by taking the Euclidean distance between the midpoint of the spine base and the most distant pixel. Third, this vector was divided evenly into three segments to find the spine head, neck, and base areas respectively. Finally, spines were classified in the four categories, considering the following threshold parameters (Fig. 2.5-Fig. 2.2B): stubby (neck length < 0.2 μm and aspect ratio < 1.3), thin (neck length > 0.2 μm, spine length < 0.7 μm, head circularity < 0.8 μm), mushroom (neck length > 0.2 μm, head

circularity > 0.8 μm), and filopodium (neck length > 0.2 μm, spine length < 0.8 μm, aspect ratio > 1.3).

For calcium imaging sessions, the TIF files were processed using the Python implementation of Suite2P [171]. We briefly summarize their pipeline here. First, TIFs in the image stack undergo rigid registration using regularized phase correlations. This involves spatial whitening and then cross-correlating frames. Next, regions of interest (ROIs) are extracted by clustering correlated pixels, where a low-dimensional decomposition is used to reduce the size of the data. The number of ROIs is set automatically from a threshold set on the pixel correlations. We manually checked assigned ROIs based on location, morphology and DF/F traces.

Since the hippocampal pyramidal cells are densely packed and the microperiscope reduces the axial resolution, we perform local neuropil subtraction using custom code (`https://github.com/ucsb-goard-lab/two-photon-calcium-post-processing`) to avoid neuropil contamination. The corrected fluorescence was estimated according to

$$F_{\text{corrected}}(n) = F_{\text{soma}}(n) - \alpha(F_{\text{neuropil}}(n) - \bar{F}_{\text{neuropil}}), \tag{2.3}$$

where F_{neuropil} was defined as the fluorescence in the region < 30 μm from the ROI border (excluding other ROIs) for frame n. $\bar{F}_{\text{neuropil}}$ is F_{neuropil} averaged over all frames. α was chosen from $[0, 1]$ to minimize the Pearson's correlation coefficient between $F_{\text{corrected}}$ and F_{neuropil}. The $\Delta F/F$ for each neuron was then calculated as,

$$\frac{\Delta F}{F} = \frac{F_n - F_0}{F_0} \tag{2.4}$$

where F_n is the corrected fluorescence ($F_{\text{corrected}}$) for frame n and F_0 is defined as the first mode of the corrected fluorescence density distribution across the entire time series.

We deconvolved this neuropil subtracted $\Delta F/F$ to obtain an estimate for the instantaneous spike rate, which we used (only) for the computation of neurons' spatial information (see below). This inferred spike rate was obtained via a MATLAB implementation of a sparse, nonnegative deconvolution algorithm (OASIS) used for Ca^{2+} recordings [172] We used an auto-regressive model of order 2 for the convolution kernel. Any cells that had spike rate $>$ 10 spikes/sec. or $<$1 spikes/sec. were manually checked and were removed from consideration if their traces appeared too noisy or sparse. Such cells were not considered in future analysis and were not included in the total number of cells recorded from.

2.2.7 Spine Imaging Data Analysis

After nonrigid registration, high pass filtering, and binarization of the dendritic segment, individual spines were extracted based on standard morphological criteria [173]. Spines projecting laterally from the dendritic segment were extracted and analyzed as individual objects, as described previously (Fig. 2.5-Fig. 2.2). The sum of the members of each spine class, as well as the total number of all spines, was recorded for each session. Spine totals (S_{total}) were then broken down into 10 μm sections of the dendritic segment (S_{section}) using the following calculation:

$$S_{\text{section}}(n) = \frac{S_{\text{total}}(n)}{D_{\text{length}} \times \frac{F_{\mu\text{m}}}{F_{\text{pixels}}}} \times 10 \tag{2.5}$$

where length of the dendritic segment, D_{length}, was determined by skeletonizing the dendritic shaft to 1 pixel in diameter, then taking the area of the pixels. F_{pixels} is the FOV in pixels, which here was 760 $\times$ 760 at 16$\times$ magnification, and $F_{\mu\text{m}}$ is the FOV in microns, which was 52 $\times$ 52 μm.

Turnover was estimated at 24 h increments; turnover here is defined as the net change

in spines per day for each morphological class (Fig. 2.5F). To determine which specific spines were involved in turnover across days, segments recorded 24 h apart were aligned and overlaid using a custom MATLAB interface, which allowed the user to manually select new or removed spines. Percent addition/subtraction $S_{a/s}$ was calculated as:

$$S_{a/s} = \frac{N_{a/s}(t)}{N(t)} \times 100, \tag{2.6}$$

where $N_{a/s}(t)$ is spines that have been added or subtracted and $N(t)$ is the total average number of spines. To account for variance in spine classification across days, turnover of specific classes of spines was normalized to total cumulative turnover per day.

To calculate the survival fraction curve $S(t)$, we determined which spines were present at time tn that were not present at time t_0 [174, 175, **?**]. The dendritic segment from t_0 was transparently overlaid with segments from t_n , and replacement spines that were present in t_0 but not tn were manually identified. Survival fraction was quantified as

$$S(t) = \frac{N_r(t_n)}{N(t_0)} \times 100, \tag{2.7}$$

where $N_r(t_n)$ are the total spines at tn that were also found in t_0, and $N(t_0)$ are the total number of spines that were present in t_0. Survival fraction, as well as % addition and subtraction, was calculated in 10 μm sections to control for segment length.

Dendrite registration (Fig. 2.5-Fig. supplement 2) was performed using built-in registration functions in the MATLAB image processing toolbox (rigid registration: imregister, non-rigid registration: imregdemons for displacement field estimation and warping).

2.2.8 Calcium Imaging Data Analysis

For calcium imaging experiments during exploration of the air-floated chamber, processed and synchronized behavioral data and 2P imaging data were used to identify place and speed cells as follows.

First, the 1D track was divided into 72 equal bins (each ∼0.85 cm in length). Activity as a function of position (we refer to these as spatial tuning curves) was computed for each lap, with activity divided by occupancy of each binned location. We observed that in certain cases, the mice traversed the track at high speeds. To avoid misattribution of slow calcium signals to spatial bins (which were relatively small due to our small track), any lap where the average instantaneous speed was greater than 180 mm/s was removed and not considered for further analysis (an average of 7% of laps were removed). To assess the consistency of spatial coding of each cell, we randomly split the laps into two groups and computed the correlation coefficient between the averaged spatial tuning curves. We then did the same for shuffled data in which each lap's spatial tuning curve was circularly permuted by a random number of bins. Note that this was done for each lap, to avoid trivial effects that might emerge from circularly permuting data that was stereotyped along the track. This was performed 500 times, and the distribution of actual correlation coefficient values was compared to the distribution of circularly shuffled values using a two-sample Kolmogorov-Smirnov test ($\alpha = 0.01$). The distribution also had to pass a Cohen's D analysis, having a score of greater than 0.5. A cell that passed these tests was considered a "consistent" cell.

To identify a neuron as a place cell, the neuron had to pass the consistency test, in addition to being well fit by a Gaussian function, $R_{\mathrm{DF/F}} = A_0 + Ae^{\frac{(X-B)^2}{C^2}}$, with FWHM $= 2C\sqrt{\ln 2}$. Note that in this convention, $C^2 = 2\sigma^2$. Specifically, we required that: 1) the adjusted $R^2 > 0.375$; 2) 2.5cm < FWHM < 30.6 cm (50% of track length – results

in Fig. 2.11H did not significantly depend on this threshold); 3) A > 0; 4) A/A$_0 > 0.50$. Cells that met these conditions were characterized as place cells; with place fields at the location of maximal activity and width defined as the FWHM. Note that these criteria are somewhat strict compared to traditional place field criteria. When tested with data in which individual laps were time shuffled, the approach yielded a false positive rate of 0%.

Speed cells were identified using a standard process developed for identification of speed cells in medial entorhinal cortex and hippocampus [**?**, 176, 177]. We computed the Pearson correlation of each cells' DF/F trace with the mouse's speed across the experiment. This value is considered as a "speed score". We then circularly shuffled the DF/F 100 times (making sure that the amount shuffled was greater than 10 frames to ensure that the shuffled distribution did not have artificially high correlations). Cells whose speed score was greater than 99%, or less than 1%, of the shuffled distribution were considered speed cells.

To compute the spatial information [28] of cell j (SI$_j$), we used the following formula,

$$SI_j = \frac{1}{\bar{a}_j} \sum_{k=1}^{72} p(k) a_j(k) \log_2 \left[\frac{a_j(k)}{\bar{a}_j} \right], \tag{2.8}$$

where $\bar{a}_j$ is the mean inferred spike rate of cell j, $a_j(k)$ is the mean inferred spike rate of cell j at position bin k, and $p(k)$ is the probability of being at position bin k. We divide by $\bar{a}_j$ to have SI in units of bits/inferred spike. To align recordings where we recorded along the CA1-DG axis, we found the inflection point of the axis and then computed the distance of each cell to that point. To do this, we performed the following steps. 1) We extracted the position of each identified cell using Suite2P's centroid output. 2) We then fit a function of the form $a(x-b)^2 + c$ to the cell positions by rotating the field-of-view from 0 to 180 degrees and finding the rotation that maximized the R^2 value of the fit.

3) We determined the inflection point as the peak of the curve and de-rotated the fit to determine the inflection point and curve in the original coordinates. The distance of each cell to the inflection point was found by finding the point along the fit curve that had the minimal distance to the cell's centroid.

2.2.9 Immunohistochemistry

Samples were perfusion fixed using 4% paraformaldehyde in 0.1M sodium cacodylate buffer (pH = 7.4) for 10 mins, and then immersion fixed overnight at 4°C. Next, sections were rinsed in cold PBS 5 × 5 mins and 1 × 1hr. Whole brains were then embedded in 10% low-melting agarose. Subsequently, 100 μm coronal sections were cut using a vibratome (Leica, Lumberton, NJ). Sections were then blocked overnight in normal donkey serum (Jackson ImmunoResearch; West Grove, PA) diluted 1:20 in PBS containing 0.5% bovine serum albumin, 0.1% Triton-X 100, and 0.1% sodium azide, hereafter, PBTA at 4°C. Next, primary antibodies anti-GFAP (1:500; abcam; ab53554), anti-S100β (1:1000; DAKO; Z0311) were diluted in PBTA and incubated overnight at 4°C. Then, sections were rinsed 5 × 5 mins and 1 × 1hr before corresponding secondary antibodies along with the nuclear stain Hoechst33342 (1:5000; Molecular Probes; H-3570) were incubated overnight at 4°C. Lastly, secondary antibodies were rinsed, and sections mounted using Vectashield (Vector laboratories Inc; H-1200) and sealed under #0 coverslips.

High resolution wide-field mosaics of brain samples were then imaged with a 20× oil immersion lens and an Olympus Fluoview 1000 laser scanning confocal microscope (Center Valley, PA) at a pixel array of 800 × 800 and then registered using the bio-image software Imago (Santa Barbara CA).

We then calculated glial cell density as a function of distance from the microperiscope face. First, each mosaic was rotated so that the medial-lateral axis of the brain sample

was aligned to be parallel with the horizontal axis of the image. Then each mosaic was cropped to remove extraneous pixels outside of the imaged brain slice. Next, a line denoting the face of the microperiscope was manually drawn parallel to the dorsal-ventral axis aligned with the location of the microperiscope face. We then used a custom cell-counting algorithm that identified potential regions of interest (ROIs). We limited the ROIs to be within the hippocampal formation in the brain slices. The Euclidean distance between the closest point on the defined microperiscope face and each ROI's centroid was calculated. Afterward, a similar procedure was performed on the contralateral side of the brain slice, with a mock "microperiscope face" defined at symmetric coordinates to the true microperiscope face, to serve as a control. These steps were repeated for each channel of the mosaic.

After extracting each ROIs distance from the microperiscope face, we counted the number of cells in each 50 μm distance bin. To account for basal glial cell density, we calculated the percent change of glial cell density on the microperiscope side with respect to the control side. This procedure was repeated 1000 times using randomly sampled distances, with replacement, to bootstrap the sample variance.

2.2.10 Statistical Information

Violin plots were made using an open-source MATLAB package [178]. Statistical tests for spine morphological types were calculated using a one-way ANOVA. Reliability across laps was tested with a two-sample Kolmogorov-Smirnov test. Comparisons between model fits for spatial distribution of spatial information and place field width used a General Linear F-test.

2.3 Results

2.3.1 Optical Access to the Transverse Hippocampus Using Implanted Microperiscopes

To image the transverse hippocampal circuit using two-photon (2P) imaging, we developed a surgical procedure for chronically implanting a glass microperiscope into the septal (dorsomedial) end of the mouse hippocampus (Fig. 2.1A; Fig. 2.2; see Sec. 2.2.2). For imaging CA1 only, we used a 1 mm × 1 mm × 2 mm microperiscope ($v1_{CA1}$; Fig. 2.1B, left), and for imaging the entire transverse hippocampus (CA1-CA3, DG), we used a 1.5 mm × 1.5 mm × 2.5 mm microperiscope ($v2_{HPC}$; Fig. 2.1B, right). The microperiscope hypotenuse was coated with enhanced aluminum in order to reflect the imaging plane orthogonally onto the transverse plane (Fig. 2.1C, D). To insert the microperiscope, we made an incision through the dura and tissue (Fig. 2.2A), then lowered the tip of the microperiscope into the incision (Fig. 2.2B), pushing the cortical tissue medially (Fig. 2.2C). Although this approach eliminated the need for the aspiration of cortical tissue typically performed prior to hippocampal imaging [161], it nonetheless results in severed connections and compressed tissue medial to the implant. Since the septal end of the hippocampus was affected by the implant, we used immunohistochemistry to quantify the effect of microperiscope implantation on astrocyte proliferation, as a function of distance from the imaging face of the microperiscope (Fig. 2.1E; Fig. 2.3). Similar to previous research using microprism implants [166], we found an increase in the prevalence of astrocytes close to the microperiscope face, but the density decreased with distance and was indistinguishable from the control hemisphere beyond 300 μm from the microperiscope face (Fig. 2.1F).

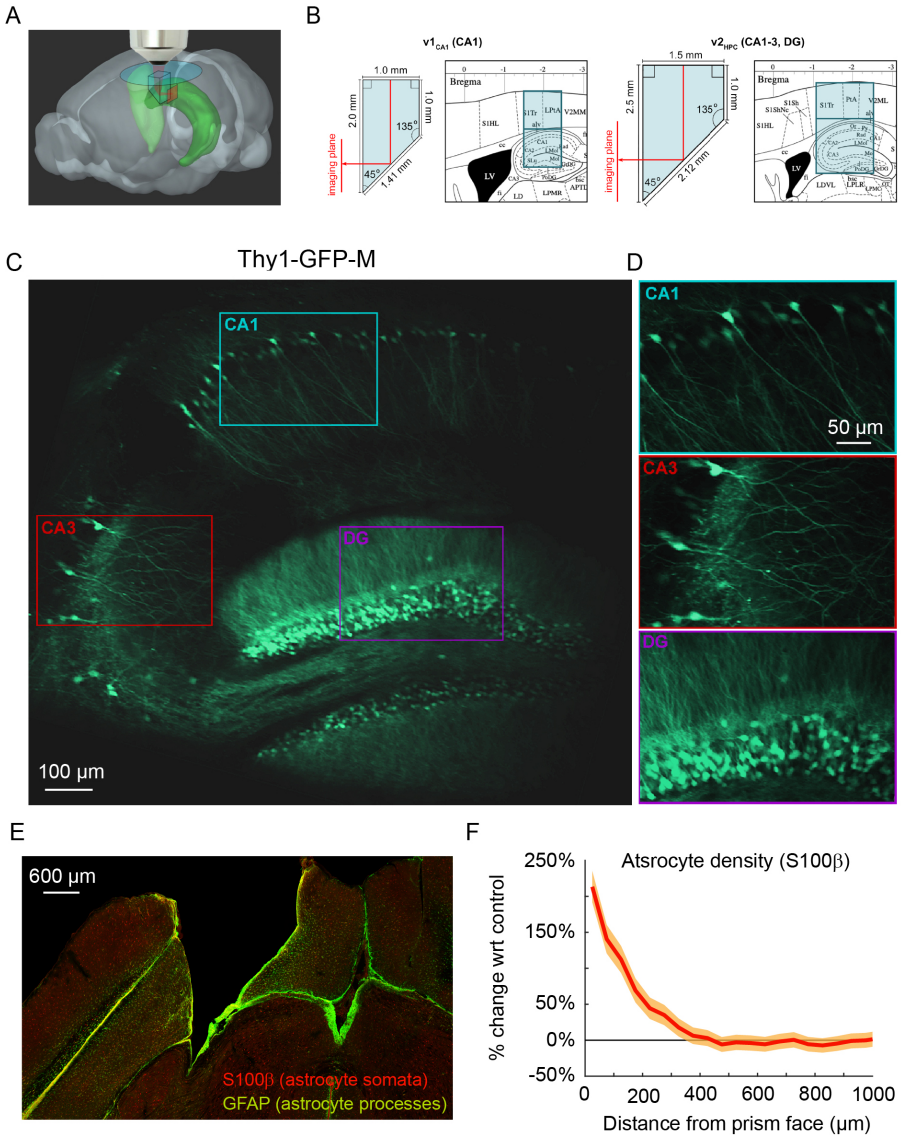

Figure 2.1: **Implanted microperiscopes allow imaging of the hippocampal transverse plane.** (A) Three-dimensional schematic [179] illustrating microperiscope implantation and light path for hippocampal imaging. (B) Schematics [180] showing the imaging plane location of $v1_{CA1}$ (1 mm imaging plane, 2 mm total length) and $v2_{HPC}$ (1.5 mm imaging plane, 2.5 mm total length) microperiscopes. (C) Tiled average projection of the transverse imaging plane using the $v2_{HPC}$ microperiscope implant in a Thy1-GFP-M transgenic mouse. Scale bar = 100 μm. Enlarged images of hippocampal subfields (CA1, CA3, DG) corresponding to the rectangles in (C). Scale bar = 50 μm. (E) Example histological section stained for astrocyte cell bodies (S100β, red) and processes (GFAP, green). Scale bar = 600 μm. Quantification of astrocyte density as a function of distance from the microperiscope face, normalized to the density in the unimplanted contralateral hemisphere (n = 2 mice; 288 and 316 days post-implant; mean ± bootstrapped s.e.m.).

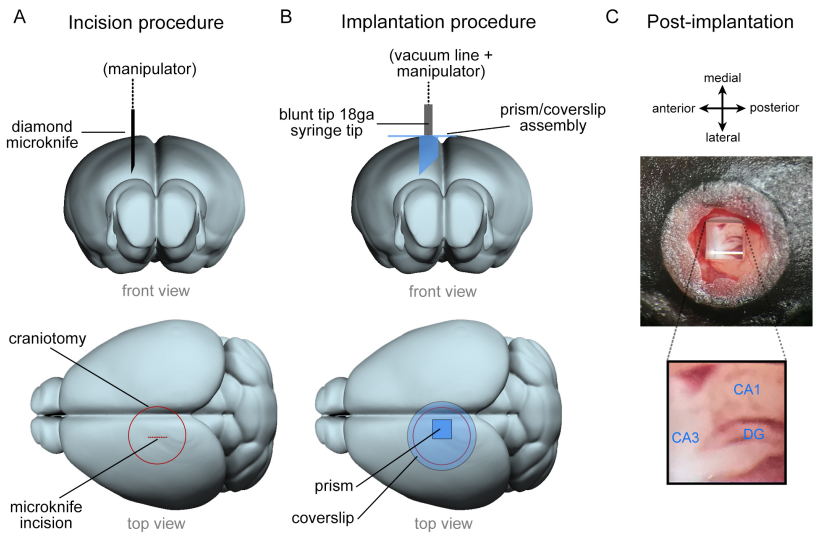

Figure 2.2: **Overview of microperiscope implantation surgery.** (A) Schematic view from front (top row) and top (bottom row) of mouse brain indicating placement of the incision. The incision is made using a diamond microknife prior to the implantation of microperiscope. The microknife bevel faces medial to leave a flat surface for the microperiscope face. (B) Schematic view from front (top row) and top (bottom row) of the brain during microperiscope implantation. The microperiscope is attached to the coverglass prior to implantation using optical glue, and the assembly is lowered into the incision using a blunt 18 gauge syringe mounted on a manipulator and held using a vacuum line. Once the assembly has been cemented to the skull, the vacuum is released, and the manipulator removed. The microperiscope faces lateral for imaging of the transverse plane of the dorsomedial hippocampus. (C) Post implantation view of microperiscope assembly, with identified hippocampal subregions.

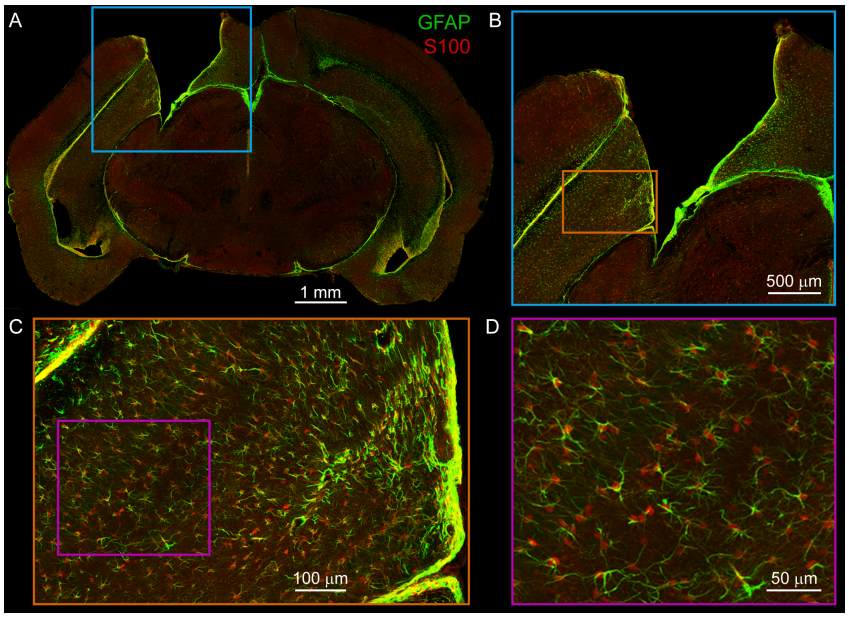

Figure 2.3: **Additional visualization of astrocyte immunohistochemistry.** (A) Mosaic confocal image of entire slice from mouse implanted with $v2_{HPC}$ microperiscope. Note that tissue shrinkage around implant margins increases apparent implant size. Green: GFAP (astrocyte processes); Red: S100β (astrocyte cell bodies); scale bar = 1mm. (B) Higher zoom image from cyan box in (A). Scale bar = 500 μm. (C) Higher zoom image from orange box in (B). Scale bar = 100 μm. (D) Higher zoom image from cyan box in (C). Note co-staining of individual astrocytes with GFAP and S100β; S100β used for all quantification (Fig. 2.1F) due to improved localization of cell bodies. Scale bar = 50 μm.

2.3.2 Characterizing Optical Properties of the Microperiscope

Use of the microperiscope requires imaging through several millimeters of glass, which could cause beam clipping or optical aberrations, potentially resulting in decreased signal intensity and optical resolution. However, the extent to which imaging through glass prisms affects optical signal quality is not well understood. To determine how our approach affects 2P image properties, we modeled the expected optical properties and compared it to the experimentally-determined signal intensity and point spread function measurements using fluorescent microspheres (Fig. 2.4A). To characterize signal intensity, we measured the minimum laser power necessary for saturation at the center of the microsphere, holding other imaging parameters constant (Fig. 2.4B). We found that signal intensity imaged through the microperiscope was reduced compared to a standard coverslip ($v1_{CA1}$ microperiscope: 59.0% ± 6.6%; $v2_{HPC}$ microperiscope: 67.2% ± 8.5%; mean ± s.d.; Fig. 2.4C). To measure resolution, we imaged small diameter (0.2 μm) fluorescent microspheres and determined the point spread function (PSF), measured as the full width at half maximum (FWHM) of the lateral and axial microsphere profiles. Compared to a standard cranial window using a 0.15 mm coverslip. We found that the FWHM was higher for both lateral (coverslip: 0.6 μm; $v1_{CA1}$ microperiscope: 1.0 μm; $v2_{HPC}$ microperiscope: 0.8 μm; Fig. 2.4D-G) and axial (coverslip: 3.4 μm; $v1_{CA1}$ microperiscope: 11.7 μm; $v2_{HPC}$ microperiscope: 9.5 μm; Fig. 2.4D-G) dimensions. Optical modeling indicated that the decrease in resolution is predominantly due to clipping of the excitation beam, resulting in a reduction of the functional numerical aperture of the imaging system (theoretical axial PSF FWHM of $v1_{CA1}$ microperiscope: 10.9 μm; $v2_{HPC}$ microperiscope: 7.7 μm; see Sec. 2.2.3) rather than optical aberrations. As a result, use of adaptive optics did not significantly improve the axial resolution, though adaptive optics did improve the signal intensity by 40-80% (data not shown). We next mea-

sured whether the optical properties changed across the microperiscope imaging plane, measuring the lateral and axial microsphere FWHM across the horizontal extent of the microperiscope (Fig. 2.4H). We found that the resolution was mostly uniform, with no significant differences observed in lateral or axial resolution as a function of horizontal position (Fig. 2.4I; lateral: $F(14, 593) = 0.53$, $p = 0.91$; axial: $F(14, 593) = 0.91$, $p = 0.55$; one-way ANOVA). Finally, we measured resolution as a function of distance from the face of the microperiscope (Fig. 2.4J). The lateral FWHM did not change as a function of distance from the microperiscope face (Fig. 2.4K; $F(19, 109) = 0.53$, $p = 0.26$; one-way ANOVA). The axial FWHM slightly increased with greater distances from the microperiscope face (Fig. 2.4K; $F(19, 109) = 2.51$, $p = 0.002$; one-way ANOVA), though none of the individual positions were significantly different when corrected for multiple comparisons ($p > 0.05$ for all positions, Tukey-Kramer post hoc test).

Taken together, these results indicate that imaging through the microperiscope lowers signal intensity and axial resolution. Despite these effects, the quality is still sufficient to image individual HPC neurons (Fig. 2.1C) and sub-micron morphological structures (Fig. 2.5) with laser power well below the photodamage threshold.

2.3.3 Imaging Spines on the Apical Dendrites of CA1

Dendritic spines are highly dynamic and motile structures that serve as the postsynaptic sites of excitatory synapses in the hippocampus [181, 182]. Previous in vitro studies suggest a role for dendritic spines in structural and functional plasticity, but the transient and dynamic nature of these structures make them ideally suited for being studied *in vivo* [183, 184, 174]. Although existing techniques allow imaging of spines on the basal CA1 dendrites near the surface of the hippocampus [174, 175, **?**], as well as transverse offshoots of apical dendrites [185, 186], imaging hippocampal neurons along the major

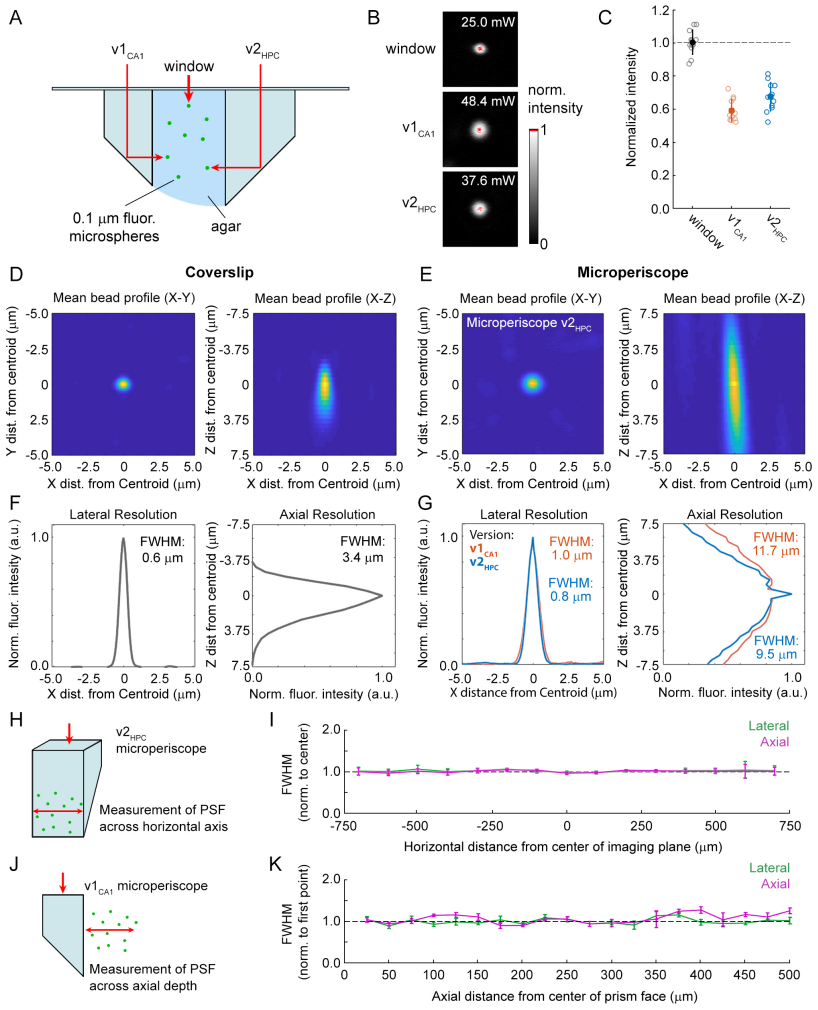

Figure 2.4: **Optical characterization of cranial window and microperiscopes.** (A) Schematic of signal yield and resolution characterization experiments to allow comparison of microperiscopes ($v1_{CA1}$ and $v2_{HPC}$) and standard coverslip window (0.15 mm thickness) using 0.2 μm fluorescent microspheres imaged with a 16×/0.8 NA objective. (B) Example microspheres imaged through the window, $v1_{CA1}$ microperiscope, and $v2_{HPC}$ microperiscope. Laser power was gradually increased to the minimum level necessary to saturate the centroid of the microsphere with all other imaging parameters held constant. (C) Distribution of signal intensity values (n = 12 microspheres for each condition), measured as the reciprocal of the minimum laser intensity for saturation, and normalized to the mean intensity through the window (v1CA1: 59.0% ± 6.6%; v2HPC: 67.2% ± 8.5%, mean ± s.d.). (D) Average X-Y profile (left) and X-Z profile (right) of fluorescent microspheres (n = 58 microspheres) imaged through the window. (E) Average X-Y profile (left) and X-Z profile (right) of fluorescent microspheres (n = 48 microspheres) imaged through the $v2_{HPC}$ microperiscope (2.5 mm path length through glass). *Caption continued on next page.*

Figure 2.4: (F) Plot of normalized fluorescence intensity profile of X dimension (lateral resolution; FWHM = 0.6 μm) and Z dimension (axial resolution; FWHM = 3.4 μm) through the centroid of the microsphere (n = 58 microspheres) imaged through the window. (G) Plot of normalized fluorescence intensity profile of X dimension (lateral resolution; $v1_{CA1}$ FWHM = 1.0 μm, orange; $v2_{HPC}$ FWHM = 0.8 μm, blue) and Z dimension (axial resolution; $v1_{CA1}$ FWHM = 11.7 μm; $v2_{HPC}$ FWHM = 9.5 μm) through the centroid of the microsphere (n = 28 microspheres for $v1_{CA1}$, n = 48 microspheres for $v2_{HPC}$). (H) Schematic of the experiment used to characterize lateral and axial resolution across the horizontal axis of the microperiscope. Version $v2_{HPC}$ used for larger horizontal range. (I) Lateral (green) and axial (magenta) FWHM (mean ± s.e.m.) of average microsphere profile as a function of horizontal distance from the center of the microperiscope. FWHM values normalized to the mean at the center of the imaging axis (-100 to +100 μm). There was no effect of horizontal position on lateral FWHM ($F(14, 593) = 0.53$, $p = 0.91$) or axial FWHM ($F(14, 593) = 0.91$, $p = 0.55$; one-way ANOVA). (J) Schematic of the experiment used to measure resolution as a function of distance from the microperiscope face. Version $v1_{CA1}$ used for larger working distance. (K) Lateral (green) and axial (magenta) FWHM (mean ± s.e.m.) of average microsphere profile as a function of distance from the face of the microperiscope, with FWHM values normalized to the closest position (25 μm). There was no effect of distance of imaging depth on lateral FWHM ($F(19, 109) = 0.53$, $p = 0.26$). There was a difference for axial FWHM ($F(19, 109) = 2.51$, $p = 0.002$; one-way ANOVA), though no individual position was significantly different from the first position when corrected for multiple comparisons (Tukey-Kramer post hoc test).

somatodendritic axis has not previously been possible. Using the microperiscope, we were able to track apical dendritic spine dynamics along the dendrite for CA1-3 neurons in awake mice, though we focused on CA1 dendrites in this study.

In order to visualize apical dendrites in CA1 neurons, we implanted a cohort of Thy1-GFP-M mice, sparsely expressing GFP in a subset of pyramidal neurons [187], with $v1_{CA1}$ microperiscopes (Fig. 2.5A; n = 7 mice). We focused on CA1 apical dendrites, but both apical and basal dendrites could be imaged in CA1-3 neurons. Although it is theoretically possible to image DG dendritic structures using the microperiscope, the Thy1-GFP-M mouse line has dense expression throughout DG (Fig. 2.1C, D), which prevented clear identification of distinct processes. To resolve individual spines along the dendrite, high-resolution images were taken from several axial planes spanning the segment, and a

composite image was generated using a weighted average of individual planes (Fig. 2.5B; see Methods). We reduced noise by filtering and binarizing, and isolated dendrites of interest for tracking across days (Fig. 2.5C, D; Fig. 2.6A; see Methods).

Previous studies have shown that dendritic spines fall into four major morphological subtypes: filopodium, thin, mushroom, and stubby [188, 189, 190]. We found that our resolution was sufficient to classify dendritic spines into their relative subtypes and evaluate density and turnover based on these parameters (Fig. 2.5E, F; Fig. 2.6B). Consistent with previous studies, we found a non-uniform distribution of dendritic spines: 30.4% thin, 41.0% stubby, 26.2% mushroom, and 2.3% filopodium ($F(3,100) = 51.47$, $p < 0.0001$, one-way ANOVA; Fig. 2.5E). The low proportion of filopodium found in this and previous 2P imaging and histological studies [191, 192] (2-3%), as compared to electron microscopy studies [193, 194] ($\sim 7\%$), may result from the narrow width of these structures causing them to fall below the detection threshold [174, **?**, 185, 195]. Consistent with previous work [196, 197, 173], we found that particular classes, such as filopodium, had a high turnover rate, while other classes were more stable across sessions ($F(3,100) = 7.17$, $p < 0.001$, one-way ANOVA; Fig. 2.5F).

We found that total turnover dynamics reflect 15.0% $\pm$ 2.0% spine addition and 13.0% $\pm$ 1.9% spine subtraction across consecutive days, representing a high degree of instability (Fig. 2.5H, I). We computed the survival fraction, a measure indicating the fraction of spines still present from day one [184, 174, **?**]. Although daily spine addition and subtraction was approximately 15.0% (Fig. 2.5H, I), the survival fraction curve yields a more conservative estimate of turnover dynamics (Fig. 2.5G). Across 10 days, we found a 23.5% net loss in original spines, indicating that most spine turnover takes place within an isolated population of transient spines. Both our cumulative turnover and survival fraction results were generally similar to previous findings from basal dendrites in CA1 [174, 175, **?**], indicating that apical and basal dendrites exhibit similar spine dynamics.

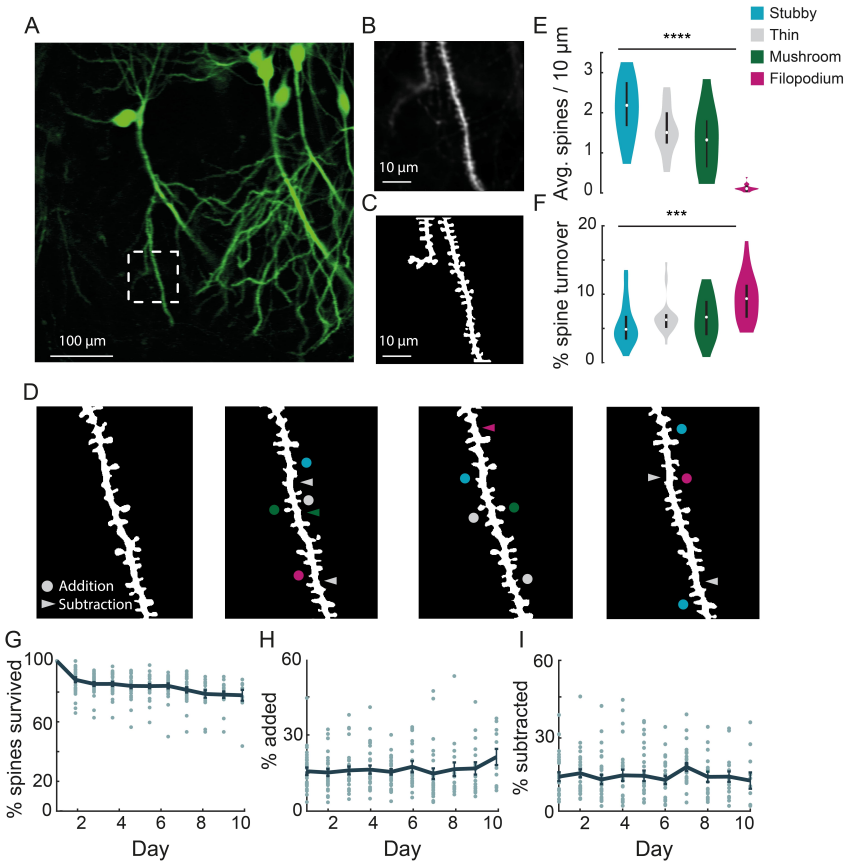

Figure 2.5: **Chronic imaging of spine morphology in CA1 apical dendrites.** (A) Average projection of CA1 neurons sparsely expressing a GFP reporter (Thy1-GFP-M) imaged through the $v1_{CA1}$ microperiscope. Scale bar = 100 μm. (B) Weighted projection (see Methods) of the apical dendrites shown in the dashed box of (A). Scale bar = 10 μm. (C) Filtered and binarized image (Figure 3-figure supplement 1A; see Methods) of the dendrites in (B) to allow for the identification and classification of individual dendritic spines. Scale bar = 10 μm.(D) Tracking CA1 dendritic spines over consecutive days on a single apical dendrite. Arrowheads indicate subtracted spines and circles indicate added spines. Colors indicate spine type of added and subtracted spines: filopodium (magenta), thin (grey), stubby (blue), and mushroom (green). (E) Average number of spines per 10 μm section of dendrite, for each of the four classes of spine (n = 26 dendrites from 7 mice); one-way ANOVA, F(3,100) = 51.47, ****p $<$ 0.0001. Error bars indicate the interquartile range (75th percentile minus 25th percentile) and circle is median. (F) Percent spine turnover across days in each spine type (n = 26 dendrites from 7 mice); one-way ANOVA, F(3,100) = 7.17, ***p $<$ 0.001. Error bars indicate the interquartile range (75th percentile minus 25th percentile) and circle is median. Spine survival fraction across processes (n = 26 dendrites from 7 mice) recorded over 10 consecutive days. (H) Percent of spines added between days over 10 days of consecutive imaging. (I) Percent of spines subtracted between days over 10 days of consecutive imaging.

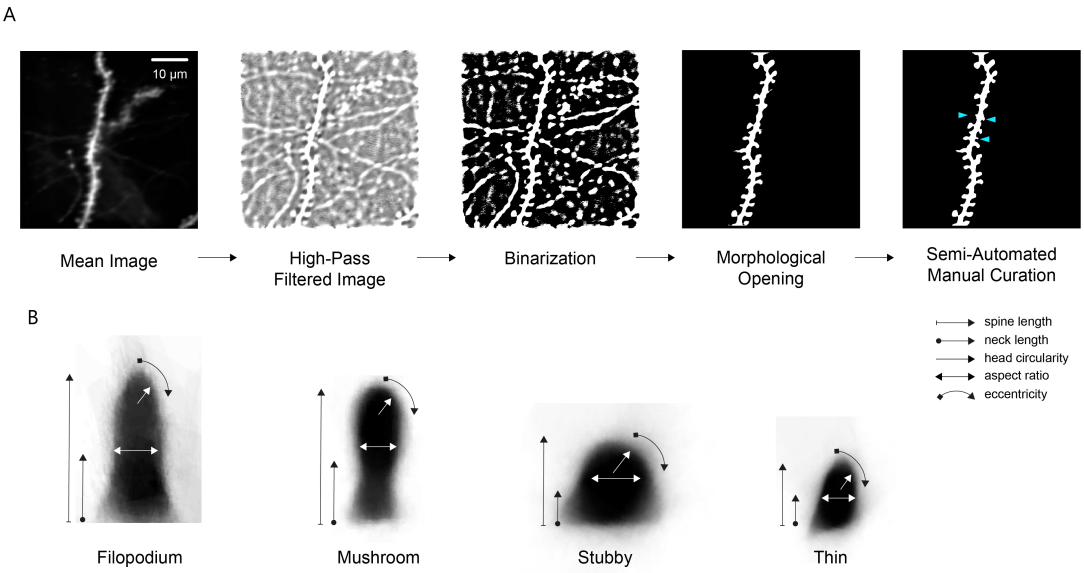

Figure 2.6: **Dendritic morphology image processing pipeline and spine classification.** (A) Steps in dendritic morphology image processing pipeline (see Sec. 2.2.7). Weighted mean image is obtained by acquiring average projections from 4 planes spaced by 3 μm and weighting the images around the brightest part of the dendrite. Next, the mean image is high-pass filtered to reveal fine structures. The image is then binarized using a global threshold, to filter out less prominent dendrites. Morphological opening is applied to remove any elements of the high-pass filtered image that survive the binarization process that are too small to be the principal dendrite. A semi-automated manual curation process is then used to add back any individual spines that were lost during binarization and remove any unwanted dendritic segments. Cyan arrowheads indicate manually added spines. To isolate the spines, the image is skeletonized and structures that protrude laterally from the dendritic shaft and have a total area of >1 μm^2 are identified as spines. Scale bar = 10 μm. (B) Mean profile of all filopodium, mushroom, stubby, and thin spines (n = 47, 431, 695, and 495, respectively), computed by averaging across the registered spines of each morphological classification. Spines were classified according to spine length, neck length, neck width, head circularity, and eccentricity (see Sec. 2.2.7).

It should be noted that some previous studies used super-resolution techniques to detect smaller spines and reduce optical merging. Thus, our results may suggest an inflated degree of stability due to resolution limitations that prevent capture of filipodia and other small spine structures [174, **?**, 185, 195]. Despite these limitations, we found that we could identify the same dendritic processes over long time periods (up to 150 days; Fig. 2.7), allowing for longitudinal experiments tracking isolated dendritic structures over long-time intervals.

2.3.4 Recording Place Cells and Speed Cells in CA1, CA3, and DG

Much of the experimental work testing the hypothesized roles of CA1, CA3, and DG neurons has come from place cell (PC) recordings [34, 35, 36, 151, 198, 88]. While the results of these studies have been instrumental, it has not been possible to measure activity throughout the transverse hippocampal circuit in the same animals. We therefore investigated the ability of our microperiscope to record from PCs in each of the hippocampal subfields during exploration of a spatial environment.

To measure functional responses, we implanted $v2_{HPC}$ microperiscopes in transgenic mice expressing GCaMP6s in glutamatergic neurons [199, 200] (see Methods). As with the Thy1-GFP-M mice, we were able to image neurons from CA1-CA3, and DG in the same animal (Fig. 2.8A, B). In some cases, depending on microperiscope placement, we were able to record from all three areas simultaneously (Fig. 2.9). In all HPC subfields, we found normal calcium dynamics with clear transients (Fig. 2.8C, D). The imaging fields remained stable, and the same field could be imaged over 100 days later (Fig. 2.8E). In addition, microperiscope implantation allowed measurement of neural responses from mossy cells in the dentate gyrus (Figs. 2.9, 2.10A) and, depending on microperiscope

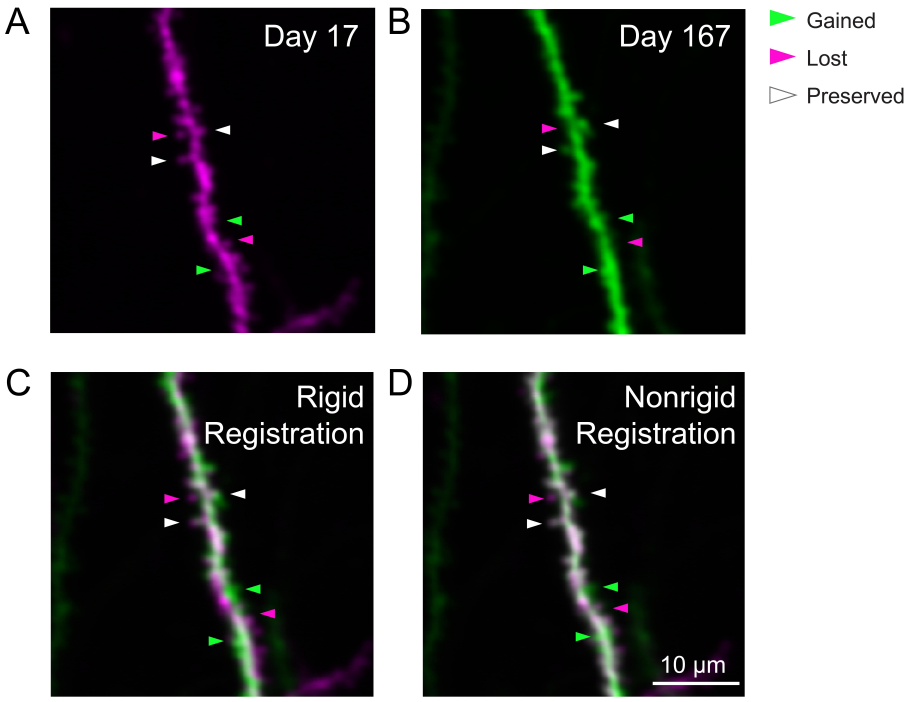

Figure 2.7: **Long-term imaging of the same dendrite.** (A) Example weighted average projection of a dendritic process imaged 17 days after microperiscope implantation. Same as (A), imaged 167 days after microperiscope implantation (150 days later). Arrowheads mark examples of spines that are present in locations along the dendrite at day 17, but not day 167 ("lost", magenta), present in locations at day 167, but not 17 ("gained", green), and present in locations at both days 17 and 167 ("preserved", white). Note that the dendrite was not imaged continuously, so we cannot be certain the preserved spines were there during the entire duration. (B) Images from (A) and (B) overlaid, using rigid registration. (C) Images from (A) and (B) overlaid, using nonrigid registration (see Sec. 2.2.7). Scale bar = 10 μm.

placement, simultaneous imaging of deep-layer cortical neurons in parietal cortex (2.10B).

As 2P microscopy generally requires the animal to be head-fixed, the behavioral assays used to probe PC activity are limited. Previous work has made use of virtual reality (VR) [44, 161], however it remains unclear how similar rodent hippocampal activity in real world environments is to that in VR [201, 202]. To study PC activity in a physical environment, our head-fixed mice explored a carbon fiber arena that was floated using an air table [168] (Fig. 2.11A; see Methods). The mice were thus able to navigate the physical chamber by controlling their movement relative to the floor. Although this approach lacks the vestibular information present in real world navigation, it captures somatosensory and proprioceptive information missing from virtual environments [169, 202, 203]. Moreover, recent work has found that place field width and single cell spatial information using this approach is comparable to the responses of free foraging animals [169]. For measurement of place coding, we allowed mice to navigate a curvilinear track over the course of 20-40 minutes (Video S4). As found previously [169], using a curvilinear track allowed for robust sampling of the spatial environment and improved place field localization, though they could also be measured in the open field. In order to measure spatial properties of the hippocampal neurons independent of reward, we relied on exploration rather than active reward administration for sampling of the environment.

To characterize place fields, we recorded from neurons in CA1, CA3, and DG (CA1: n = 1001; CA3: n = 832; and DG: n = 463) in transgenic mice with panexcitatory expression of GCaMP6s (Fig. 2.11B; n = 8 mice). We found PCs by assessing lap-by-lap reliability and goodness-of-fit to a Gaussian in all three subfields (Fig. 2.11C; false positive rate in shuffled data = 0%; see Methods). The distribution was in general agreement with previous 2P imaging experiments in mice [44, 157] (Fig. 2.11D, G; CA1: 26.0%; CA3: 18.9%; DG: 14.0%), with fields that spanned the entirety of the track (Fig. 2.11F). We found that place field widths were comparable across the three regions (Fig.

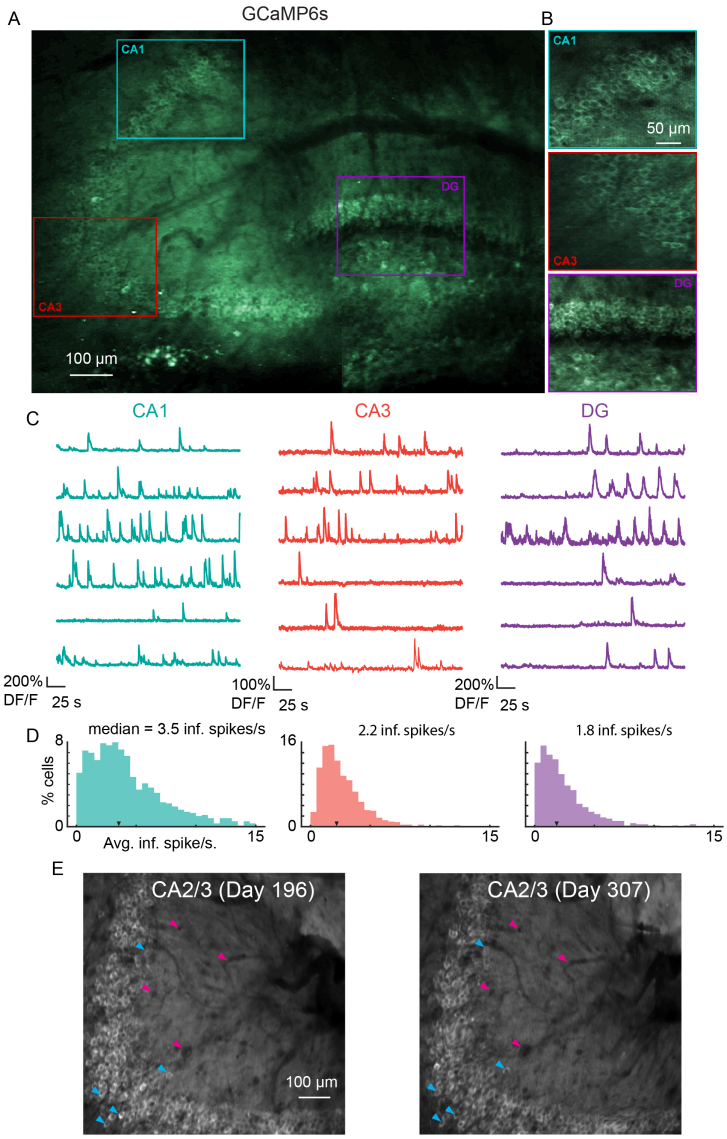

Figure 2.8: **Microperiscope imaging of calcium dynamics across subfields in awake mice.** (A) Example tiled image (as in Fig. 2.1C) for a single panexcitatory transgenic GCaMP6s mouse (Slc17a7-GCaMP6s). Scale bar = 100 μm. (B) Enlarged images of each hippocampal subfield (CA1, CA3, DG) corresponding to the rectangles in (A). Scale bar = 50 μm. (C) Example GCaMP6s normalized fluorescence time courses (% DF/F) for identified cells in each subfield. (D) Distribution of average inferred spiking rate during running for each hippocampal subfield (CA1, CA3, DG; see Methods). Median is marked by black arrowhead. Example average projection of CA2/3 imaging plane 196 days post implantation (left) and 111 days later. Image was aligned using non-rigid registration (see Methods) to account for small tissue movements. Magenta arrowheads mark example vasculature and blue arrowheads mark example neurons that are visible in both images. Scale bar = 100 μm.

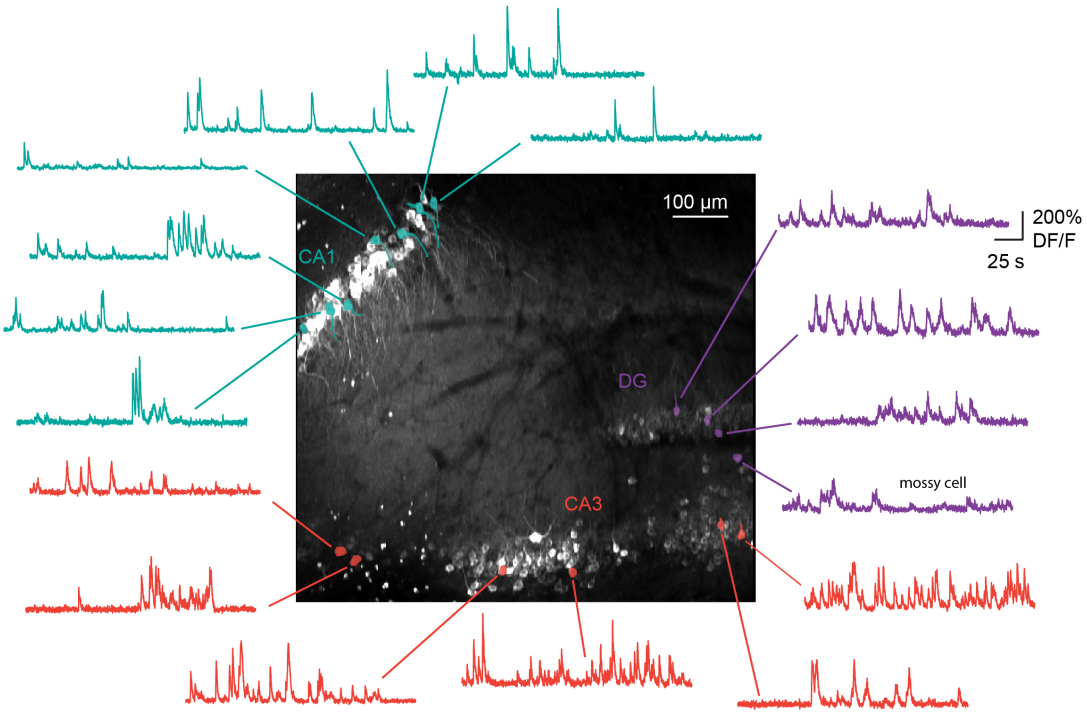

Figure 2.9: **Simultaneous imaging of all three hippocampal subfields.** Maximum projection, and example GCaMP6s fluorescence time courses for identified cells, from a recording in which CA1 (n = 55 cells), CA3 (n = 158), and DG (n = 28) were simultaneously recorded from the same image plane. Scale bar = 100 μm.

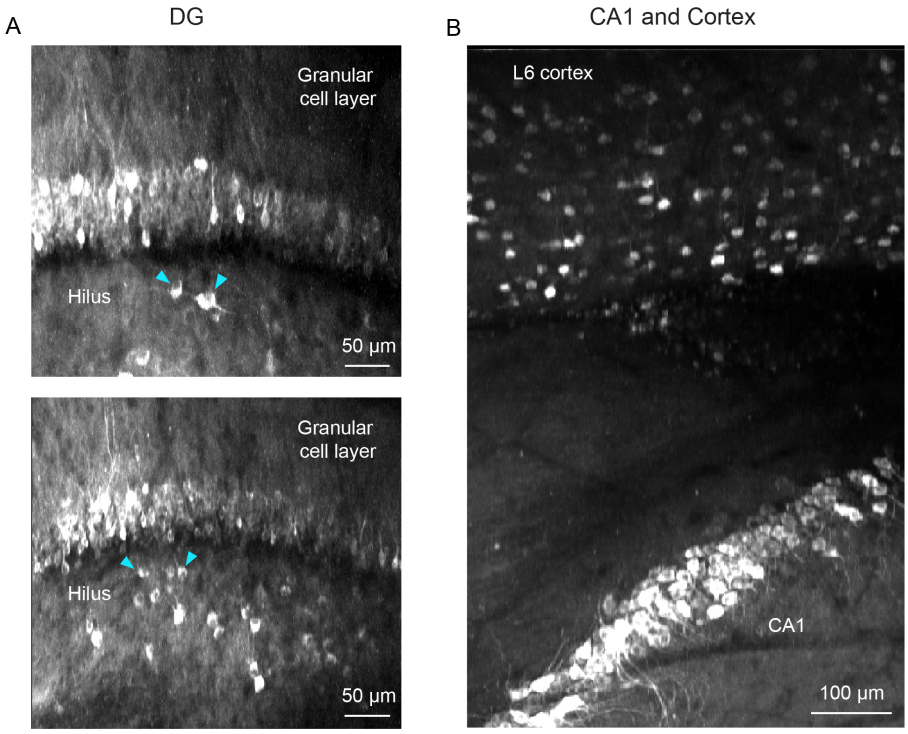

Figure 2.10: **Additional applications for microperiscope imaging.** (A) Two maximum projections of imaging planes in which both the granular cell layer and the hilus in the DG could be imaged simultaneously. Example putative mossy cells indicated with cyan arrowheads. Scale bar = 50 μm. (B) Maximum projection from an imaging session in which both CA1 and L6 of the neocortex are visible, allowing simultaneous imaging of neurons in CA1 and deep layers of the parietal cortex. Scale bar = 100 μm.

2.11H; CA1: mean (median) ± s.e.m = 18.8 (19.4) ± 0.4 cm; CA3: 18.6 (18.8) ± 0.5 cm; DG: 18.9 (19.5) ± 0.8 cm) and that spatial information was highest in CA1, followed by CA3, then DG (Fig. 2.11I; CA1: mean (median) ± s.e.m. = 0.95 (0.87) ± 0.03 bits/inferred spike; CA3: 0.89 (0.68) ± 0.05 bits/inferred spike; DG: 0.79 (0.61) ± 0.08 bits/inferred spike). The place field width and spatial information of the CA1 place cells were similar to those found in a previous study using the same floating chamber design with a traditional hippocampal window implant [169].

To determine if neurons close to the microperiscope face showed unusual response properties, we imaged functional responses as a function of distance. Place cell width did not vary significantly as a function of depth of the imaging plane from the face of the microperiscope (Fig. 2.9A; $F(4, 241) = 1.6$, $p = 0.17$, one-way ANOVA). Spatial information did vary as a function of depth ($F(4, 1872) = 15.9$, $p = 8.6 \times 10^{-13}$, one-way ANOVA), but there was no systematic effect (Fig. 2.12B). Decay constant for the fitted transients [172] did have a significant relationship with imaging depth ($F(4, 1872) = 29.8$, $p = 4.5 \times 10^{-24}$), which may be due to aberrant activity in a small subset of neurons close ($\leq$130 μm) to the microperiscope face (Fig. 2.12C).

As cells responsive to speed have recently been found in the medial EC[**?**, 176] and CA1 [176, 177], we also identified neurons as speed cells (SCs) if their activity was significantly related to running speed [**?**] (see Methods; average speed per recording: CA1, 71.3 ± 33.7 mm/s; CA3, 76.2 ± 32.1; DG, 101.1 ± 49.0; Two-sample Kolmogorov-Smirnov test: CA1-CA3, $p = 0.90$; CA3-DG, $p = 0.53$; CA1-DG, $p = 0.32$). We found SCs in CA1, CA3, and DG, with all areas having cells that showed both increased and decreased activity with higher running speeds (Fig. 2.11E). Speed cells were most abundant in DG (Fig. 2.11G; CA1: 9.4%; CA3: 13.5%; DG: 30.2%), consistent with recent work that found it was possible to decode the speed of freely moving animals from the activity of DG, but not CA1 [88].

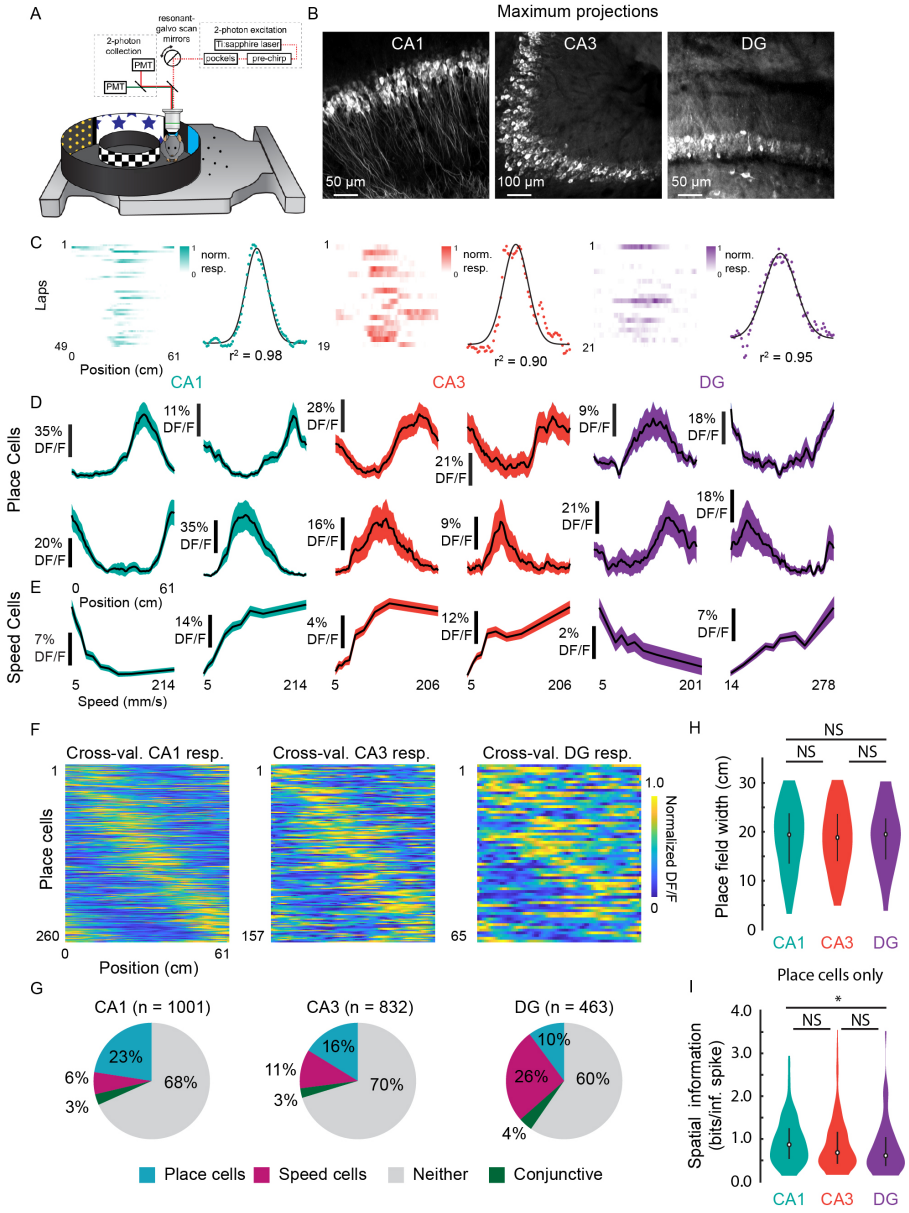

Figure 2.11: **Prevalence of place and speed cells across hippocampal subfields.** (A) Schematic of air-lifted carbon fiber circular track (250 mm outer diameter) that the mice explored during imaging. Four sections of matched visual cues lined the inner and outer walls. (B) Example maximum projections of GCaMP6s-expressing neurons in each subfield. Scale bar = 50, 100, 50 μm respectively. (C) Lap-by-lap activity and Gaussian fits, with r^2 value reported, of example place cells for each hippocampal subregion to illustrate our process of identifying place cells (see Methods). (D) Examples of mean normalized calcium response (% DF/F) versus position along the circular track for four identified place cells (PCs) in each subregion. Shaded area is s.e.m.

Figure 2.11: (E) Examples of mean calcium response (% DF/F) versus speed along the circular track for two identified speed cells (SCs) in each subregion. Shaded area is s.e.m. (F) Cross-validated average responses (normalized % DF/F) of all PCs found in each subfield, sorted by the location of their maximum activity. Responses are plotted for even trials based on peak position determined on odd trials to avoid spurious alignment. (G) Distribution of cells that were identified as PCs, SCs, conjunctive PC + SCs, and non-coding in each subregion. (H) Distribution of place field width for all place cells in each subregion. Error bars indicate the interquartile range (75th percentile minus 25th percentile) and circle is median. Two-sample Kolmogorov-Smirnov test: CA1-CA3, $p = 0.67$; CA3-DG, $p = 0.70$; CA1-DG, $p = 0.51$. NS, not significant ($p > 0.05$). (I) Distribution of spatial information (bits per inferred spike) for all place cells in each subregion. Error bars and circle are the same as in (H). Two-sample Kolmogorov-Smirnov test: CA1-CA3, $p = 0.05$; CA3-DG, $p = 0.36$; CA1-DG, $p = 0.01$. $^{*}p < 0.05$.

2.3.5 Distribution of Place Cell Properties Along the DG-CA1 Axis

Recent work has suggested that PC properties are heterogeneously distributed along the extent of the DG-to-CA1 axis. In particular, place field width and spatial information have been found to vary among different subregions of CA3, with distal CA3 (dCA3) and CA2 having lower spatial information and larger place fields than medial CA3 (mCA3) [29, 30, 31], and proximal CA3 (pCA3) having values most similar to DG [151]. Such distributions could be supported by known anatomical gradients in connectivity of CA3 [204, 205, 206, 207, 208]. However, given that these studies required separate animals for the recording of each location along the DG-to-CA1 axis, and that electrophysiology has limited spatial resolution along the transverse axis, we used our microperiscopes to measure these properties throughout the DG-to-CA1 axis in the same mice with high spatial resolution.

Using the $v2_{HPC}$ microperiscope, we simultaneously imaged from several hundred cells (mean $\pm$ s.d.; 215 ± 74 neurons per recording; total: 1075 neurons) extending from pCA3 to pCA1 (Fig. 2.13A). Recordings from distinct imaging planes in different mice ($n = 5$)

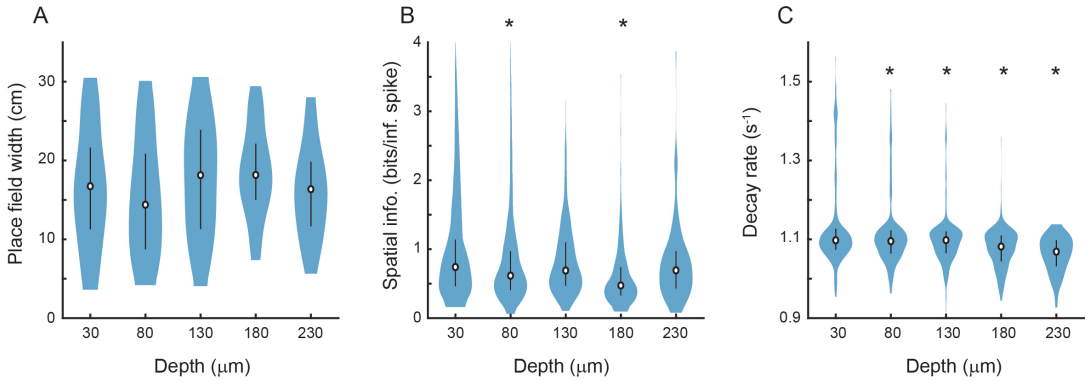

Figure 2.12: **Place cell properties as a function of distance from the face of the microperiscope.** (A) Place field width as a function of imaging depth relative to the microperiscope face. Error bars indicate the interquartile range (75th percentile minus 25th percentile) and circle is median ($F(4, 241) = 1.6$, $p = 0.17$, one-way ANOVA). (B) Same as (A), but for spatial information of place cells ($F(4, 1872) = 15.9$, $p = 8.6\times10^{-13}$, one-way ANOVA; Tukey-Kramer post hoc test for depth vs. 30μm: 80μm, $p = 1.1\times10^{-3}$; 130μm, $p = 0.34$; 180μm, $p = 9.92\times10^{-9}$; 230μm, $p = 0.15$). *$p < 0.05$ Tukey-Kramer post hoc test. (C) Same as (A), but for decay rate [172] of place cells ($F(4, 1872) = 29.8$, $p = 4.5\times10^{-24}$; Tukey-Kramer post hoc test for depth vs. 30μm: 80μm, $p = 6.70\times10^{-3}$; 130μm, $p = 9.38\times10^{-7}$; 180μm, $p = 9.92\times10^{-9}$; 230μm, p = = 9.92×10^{-9}). *p <0.05 Tukey-Kramer post hoc test.

were compared by calculating the distance of individual cells from the inflection point of the DG-to-CA1 transverse axis (Fig. 2.13A; see Sec. 2.2.8). In agreement with previous studies [29, 30, 31], we found a non-uniform distribution of spatial information along this axis (Fig. 2.13B, C; $F(4, 1047) = 5.10$, $p = 4.6 \times 10^{-4}$; General Linear F-test against a flat distribution with the same mean). In particular, we found that pCA3 cells had spatial information that was closer in value to those in DG than mCA3 (Fig. 2.13B, C; Fig. 2.11I), and that mCA3 had spatial information that was greater than dCA3/CA2 (Fig. 2.13B, C). We found that place field widths were narrowest in mCA3, although we failed to find a statistically significant non-uniform distribution with respect to place field width across the extent of the DG-to-CA1 axis (Fig. 2.14; $F(4, 206) = 1.63$, $p = 0.17$, General Linear F-test against a flat distribution with the same mean).

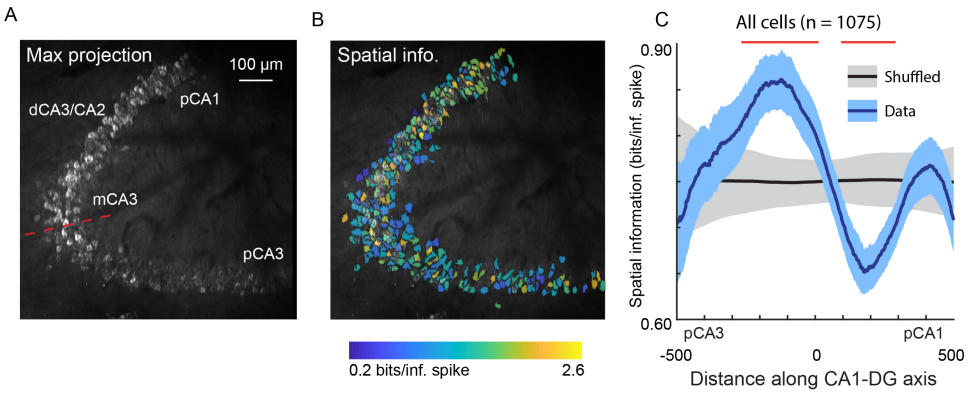

Figure 2.13: **Spatial information of neurons varies along the DG-to-CA1 axis.** (A) Maximum projection of an example DG-to-CA1 axis recording. Approximate locations of CA3 and CA1 subfields are labeled. Inflection point labeled with red line. Scale bar = 100 μm. (B) Spatial information (bits/inferred spike), pseudo–colored on a logarithmic scale, for each neuron, overlaid on the maximum projection in (A). (C) Spatial information, as a function of distance along the DG-to-CA1 axis (pCA3 to dCA1), calculated with a sliding window; mean ± bootstrapped s.e.m.; real data (blue), shuffled control (black). Red lines indicate values that are outside the shuffled distribution ($p < 0.05$). A general Linear F-test against a flat distribution with the same mean revealed significant non-uniformity: $F(4, 1047) = 5.10$, $p = 4.6 \times 10^{-4}$ (cells with distance greater than 600 or less than -600 were not included in the statistical analysis).

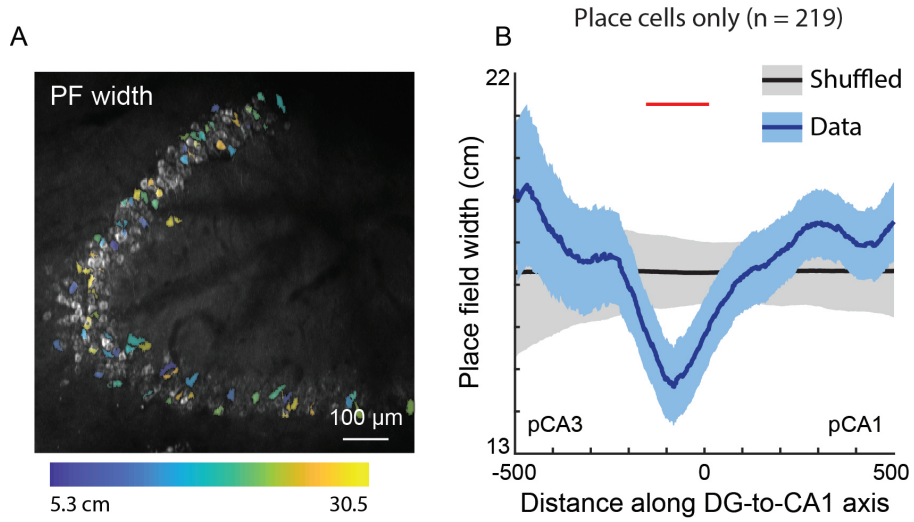

Figure 2.14: **Place field width does not significantly vary along the DG–to-CA1 axis.** (A) Place field width, pseudo-colored for each place cell, overlaid on the maximum projection from Fig. 2.13 Any cell that was not a place cell is not pseudo-colored. Scale bar = 100 μm. (B) Place field width, as a function of distance along the DG-to-CA1 axis (pCA3 to pCA1), of real data (blue) vs. shuffled control (black). Shaded area is s.e.m. Red line indicates individual positions with place field values that were outside of the shuffled distribution ($p < 0.05$). A general Linear F-test, against a flat distribution with the same mean, revealed no significant non-uniformity: $F(4, 206) = 1.63$, $p = 0.17$ (cells with distance greater than 600 or less than -600 were not included in the statistical analysis).

2.4 Discussion

The microperiscope hippocampal imaging procedure we developed allows researchers, for the first time, to chronically image neuronal structure and functional activity throughout the transverse hippocampal circuit in awake, behaving mice. This approach builds on microprism procedures developed for imaging cortical columns [165, 166, 167], allowing multiple hippocampal subfields in the same animal to be accessed optically. Using the microperiscope, we were able to resolve spines on the apical dendrites of CA1 pyramidal cells and track them across time. Additionally, we were able to characterize place cells (PCs) and speed cells (SCs) in all three hippocampal subfields and investigate their anatomical distribution across subfields.

2.4.1 Comparison to Other Methods

Over the past five decades, electrophysiology has been the principal tool used to study the hippocampus. Electrophysiological recordings have much higher temporal resolution than calcium imaging and can directly measure spikes, but have limited spatial resolution. Our approach allows large scale imaging of neurons across multiple hippocampal subfields, with known spatial and morphological relationships. In addition, our approach allows genetically-controlled labeling of particular cell types, imaging of cellular structures such as dendrites and spines, and unequivocal tracking of functional and structural properties of the same cells across time, none of which are possible with existing electrophysiological approaches.

Several approaches making use of optical imaging have been developed for use in the hippocampus. These include gradient index (GRIN) lenses [209, 210], and cannulas that can be combined with both one-photon head-mounted microendoscopes [162, 42, 163] and 2P imaging [44, 161, 164]. These methods have been limited to horizontal imaging

planes, making it difficult to image CA3 and DG, and intractable to image all three subfields in the same animal. While all of these methods cause damage to the brain, and some require the aspiration of the overlying cortex, the damage is largely restricted to superficial hippocampus. This contrasts with the implantation of our microperiscope, which is inserted into the septal end of the hippocampus and necessarily causes some damage to the structure. Despite this, we find normal response properties (Fig. 2.8, Fig. 2.11-Fig. 2.2), including selectivity for location and speed, in CA1-CA3 and DG (Fig. 2.11). Damage in the direction orthogonal to the transverse axis caused by the implantation of the microperiscope is similar to that caused by microprisms in cortex [166], as glial markers were found to decay to baseline levels with increased distance from the face of the microperiscope (Fig. 2.1E, F). We showed that, following successful implantation, the imaging fields were stable, and the same cells could be imaged up to 3 months later (Fig. 2.8E). However, we emphasize that tissue damage caused by the microperiscope assembly to the hippocampus should be taken into consideration when planning experiments and interpreting results, and that imaging planes should be at least 150 μm away from microperiscope face to avoid aberrant responses.

We found that imaging through the microperiscope had some effect on signal intensity and axial resolution. The effect depended not only on the path length through the glass, but also on the geometry of the microperiscope. For example, the larger $v2_{HPC}$ microperiscope had slightly better optical properties than the smaller $v1_{CA1}$ microperiscope (Fig. 2.4), likely due to reduced beam clipping. The optical properties should be considered along with the potential tissue damage during the planning of experiments. Moreover, further customizing the microperiscope geometry for particular applications may help optimize signal quality and minimize damage to neural structures.

To assess the ability of our optical and behavioral experimental methods to capture spatial coding, we chose to measure place fields in the absence of reward, which is known

to modulate hippocampal activity [211]. Recent work has found that place cells are significantly less stable in the absence of reward [212, 213], which may explain the lower lap-wise reliability in our experiments compared to similar experiments using the same behavioral apparatus [169]. However, we found similar percentages of place fields as other optical imaging studies, despite using a strict criteria for place cell inclusion (see Methods), and these properties did not change systematically as a function of distance from the face of the periscope (Fig. 2.11-Fig. 2.2A, B). This provides further support that the functional properties of the hippocampal circuitry remained intact when using the microperiscope. However, we note that it is possible that the place cell characteristics we measured might be altered in the presence of reward, and that this decision should be taken into account when interpreting the results.

2.4.2 Structural and Functional Properties Along the Transverse Hippocampal Circuit

We utilized the microperiscope in two experiments that would not have been possible with existing methods: (1) we tracked the spines along apical CA1 dendrites *in vivo* (Fig. 2.5); (2) we simultaneously recorded from PCs along the extent of the DG-to-CA1 transverse axis (Fig. 2.13).

Several studies have tracked the spines of basal CA1 dendrites *in vivo* by imaging the dorsal surface of the hippocampus [174, 175, **?**, 186, 195]. However, imaging spines along the major dendritic axis has not been accomplished *in vivo*. Given that these spines make up the majority of the input to CA1 pyramidal neurons, there is a significant need for understanding their dynamics, and how the dynamics differ as a function of position along the somatodendritic axis. Using the microperiscope, we tracked isolated apical dendrites for up to 10 consecutive days (Fig. 2.5D). We found considerable addition

and subtraction across days, indicating dynamic turnover in apical dendrites (Fig 3H, I). However, consistent with previous studies in basal spines [174, 175, **?**], we found the majority of spines (76.5%) survived throughout the imaging period (Fig. 2.5G). This high survival fraction suggests that the daily turnover rates we observe are reflective of a distinct pool of unstable transient spines, while the majority of spines are moderately stable, and remain present over a longer timescale. It should be noted that some prior results were obtained using super-resolution microscopy techniques that are more suited to capturing the dynamics of filipodia and small spines, and thus our results may reflect an inflated degree of stability [174, **?**, 186]. The microperiscope is in principle compatible with super-resolution microscopy, and future imaging studies that combine the two could characterize apical dendritic spine dynamics with improved accuracy [174, **?**, 186]. Taken together, our results add to the increasingly appreciated idea that, even in the absence of salient learning and reward signals, dendritic spines are dynamic and unstable structures [174, 214, 215] Understanding the nature and timescales of these dynamics has significant implications for the reported instability of PCs [41, 44, 42, 40, 43, 46].

Previous studies have found gradients of connectivity in CA3 [204, 205, 206, 207, 208], suggesting that there may be functional gradients as well. Electrophysiological recordings have indeed found that spatial information and place field width vary as a function of distance along the DG-to-CA1 axis [29, 30, 31], and that the most proximal part of CA3 is functionally more similar to DG than to the rest of CA3 [151]. This has led to a more nuanced understanding of the hippocampal circuit, with coarse anatomical subdivisions having finer functional subdivisions. However, given that these previous studies relied on recordings in different animals for each location along the DG-to-CA1 axis, and given that electrodes have poor spatial resolution in the transverse axis, the results must be cautiously interpreted. Using the microperiscope, we imaged several hundred neurons in multiple mice along the extent of the DG-to-CA1 axis (Fig. 2.13A). The location

of each neuron relative to the CA3 inflection point could be easily identified, allowing for unequivocal characterization of spatial information and place field width along the DG-to-CA1 axis across animals. Similar to the previous studies [29, 30, 31], we found a non-uniform distribution of spatial information (Fig. 2.13B, C). Place field width also appeared to vary along the DG-to-CA1 axis, though we failed to find statistically significant non-uniformity with respect to place field width (Fig. 2.13-Fig. 2.2B). Indeed, we found that spatial information, but not place field width, differed significantly between CA1, CA3, and DG (Fig. 2.11H, I). Given the high spatial resolution and ability to simultaneously record from cells across the DG-to-CA1 axis using this approach, our results strengthen the hypothesis of distributed spatial coding across, and within, hippocampal subfields.

2.4.3 Future applications

This paper explored a few possible uses for the microperiscope in interrogating the hippocampal circuit. However, there are a number of candidate applications we did not pursue that could reveal novel insight into hippocampal function. Here, we describe several of these, chosen to highlight the utility of the microperiscope in addressing questions that are challenging or intractable with existing approaches: (1) The microperiscope allows for the recording from multiple hippocampal subfields simultaneously (Fig 6; Fig. 2.8-Fig. 2.2), allowing investigation of interactions between neurons in different subfields during behavior. This could also be useful for determining the effect of neuron- or subfield-specific optogenetic manipulations on downstream subfields. (2) The microperiscope enables the investigation of local circuits by allowing morphological and genetic identification of different cell types. This includes identifying genetically-distinct hippocampal neurons (e.g. CA2 neurons, specific interneuron subtypes), as well was iden-

tifying particular cell types by position or morphological characteristics (e.g. mossy cells in DG; Fig. 2.8-Fig. supplement 2A). As these distinct cell types play important roles in hippocampal function [158, 114, 216, 217, 218, 215], having access to them will enable a greater understanding of the hippocampal circuit. (3) The microperiscope can be combined with retrograde-transported viruses to allow projection-based cell identification, making it possible to identify neurons that project to specific downstream targets. Since hippocampal neurons that project to different brain regions have been found to exhibit distinct functional properties (e.g. neurons in ventral CA1 that project to the nucleus accumbens shell have been implicated in social memory [159]), the union of these tools will be a powerful means for understanding hippocampal outputs. (4) Finally, the microperiscope provides access to the entire dendritic tree of pyramidal neurons (Fig.s 1C, 3A, 5B; Video S2), giving optical access to dendritic signaling over a much larger spatial extent than has been previously possible [219, 220, 221]. By sparsely expressing calcium or glutamate sensors in hippocampal pyramidal neurons, spines throughout the somatodendritic axis could be imaged, allowing determination of how place field responses arise from the responses of individual spines, similar to experiments in visual cortex investigating the cellular origin of orientation tuning from synaptic inputs [222, 223].

Combined with existing electrophysiological and imaging approaches, imaging of the transverse hippocampal circuit with microperiscopes will be a powerful tool for investigating hippocampal circuitry, structural dynamics, and function.

Chapter 3

Robust Variability of Grid Cell Properties Within Module Enhances the Coding Capacity of Single Grid Modules

"More is different."

Paul W. Anderson (1972)

3.1 Introduction

The discovery of grid cells (GCs) in medial entorhinal cortex (mEC) [6], characterized by multiple firing fields centered along a triangular lattice, has led to considerable experimental and computational work aimed at identifying their origin [68, 66, 62, 69, 81, 82, 63, 64, 65, 83] and their function [224, 73, 74, 67, 225, 226, 227, 75, 76].

The organization of GCs into discrete modules [6, 55], with grid properties (e.g.,

grid spacing and grid orientation) being conserved within module, but not between, has shaped the way in which these questions of origin and function have been studied. For instance, the increasing size and spacing of grid modules along the dorsal-ventral axis of mEC, by a near constant ratio, has been argued to be optimal for encoding local spatial information [59, 60], when GC activity across all modules is integrated together [56, 57, 58, 59, 60]. This is despite the fact that the hippocampus, downstream target of the mEC, recieves inputs from only a portion of the dorsal-ventral axis [228]. Additionally, the modularity of the grid system has been proposed to simplify the wiring necessary for continuous attractor networks [66, 68, 67, 69], a computational mechanism for supporting grid cells that enjoys considerable experimental support [70, 71, 72].

This prior work has made the additional corollary assumption that GC properties are "identical within module". However, the distributions of measured spacing and orientation show non-zero variability [6, 55, 72]. This could be due to finite recording times, neurophysiological noise, and/or sensitivity of numerical methods used to fit grid properties. Alternatively, this variability could reflect an inhomogeneity, at a fine scale, within module. Despite the fundamental way in which GC function has been informed by their modular organization, to the best of our knowledge, no characterization of the degree and stability of variability in grid properties within individual modules has been performed. If robust variability of grid properties does exist, then it is possible that this heterogeneity could be exploited to encode additional information about local space, reducing the requirement of integration across multiple grid modules.

In this paper, we perform detailed analysis of state-of-the-art mEC electrophysiological recordings [72]. We characterize the degree of variability of grid properties within individual modules, and find evidence for small, but robust variability of grid spacing and orientation. This variability is present whether a single value of grid spacing and orientation is assigned to each GC, or whether grid distortion [229, 230] is taken into

account by considering the three grid spacings and orientations independently. To assess the computational implications of this heterogeneity, we perform numerical experiments on synthetic grid cell populations. We find that the variability in GC spacing and orientation, at a similar degree as present in the data we analyzed, leads to near zero decoding error of local space, when using the activity of a *single* module. To our knowledge, this is the first time this has been demonstrated. This leads us to hypothesize that the small degree of inhomogeneity enhances the capability of individual modules, while at the same time maintaining the distinction between modules and not breaking the continuous attractor network.

Taken together, our results challenge a decades long held assumption and support a growing appreciation of the local spatial information grid populations can encode [87, 231, 232, 233]. Additionally, they encourage the field to think more broadly about what benefits, beyond the encoding of local space, multiple modules endows the grid code with.

3.2 Methods

3.2.1 Electrophysiology Recordings

The data analyzed in this chapter comes from a publicly available data set, which is described in detail in [72]. Here we give a briefly summarize the methodology and the experimental paradigms used during the recordings.

Three male rats (Long Evans – Rats Q, R, and S) were implanted with Neuropixels silicon probes [234, 235]. These probes were targeted at the mEC-parasubiculum region and the surgery was performed as described in [236, 235]. After three hours of recovery, recordings were performed.

In the recordings analyzed in this chapter, the animals foraged for randomly dispersed

corn puffs in a 1.5×1.5m square open field arena, with walls of height 50cm. The rats were familiar with the environment and task, having trained 10–20 times prior to the implantation. The rats were fiid restricted to motivate their foraging, being kept at a minimum of 90% of their original body weight (300–500 grams).

All procedures were approved by the Norwegian Food and Safety Authority and done in accordance with the Norwegian Animal Welfare Act and the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes.

3.2.2 Electrophysiology Post-Processing

The data utilized in this chapter had been previously post-processed. We describe, in brief, the post-processing performed by [72], as well as the post-processing we performed on the downloaded data.

Spike sorting was done on the data from the electrode arrays using KiloSort 2.5 [235]. Individual units were deemed putative cells if their average spike rate was in the range of 0.5–10Hz, and 99% of their interspike intervals were greater than 2ms.

For each putative cell, ratemaps were constructed by averaging the activity at binned spatial positions in the open field arena. This raw ratemap was smoothed, using a Gaussian kernel. The autocorrelation of these ratemaps were computed, and a grid score calculated, as describe previously [12]. To determine which putative cells were GCs, and to which module they belonged, autocorrelograms of raw ratemaps were constructed, with some of the bins corresponding to the closest and furthest distances removed. These "cropped" ratemaps were then clustered using UMAP [237] to project the data onto a two-dimensional subspace. All the cells in the non-largest cluster were found to have similar grid spacing and orientation, as well as high grid scores. They were therefore

deemed putative GCs. All cells in these clusters had their spike trains placed public repository from which we downloaded the data.

From the downloaded spike trains, we similarly constructed ratemaps and autocorrelograms in a similar manner, using code made publicly available by [82]. To ensure that the effects we saw were not driven by cells whose GC quality was low, we included in our analysis only GCs grid score greater than 0.8.

3.2.3 Electrophysiology Data Analysis

To compute the grid spacing and orientation, we used the following procedure. First, we computed the autocorrelogram of each GC from their smoothed ratemaps. We then computed the sum of the autocorrelation along a circle of growing radius (centered at the center of the autocorrelation). Interpolation methods were used to judge the autocorrelation value at arbitrary points. The radius corresponding to the first non-trivial peak of this quantity, λ_0, is used as a fist pass of the grid spacing. The values of the autocorrelogram along the circle with radius λ_0 is unwrapped and each peak is identified using `scipy.stats.find_peaks`. If there were at least five peaks (we allowed for one less than six, in the case that the last peak is at the edge of the autocorrelegram) at a minimum distance of 45 degrees from each other, the orientations of the first three peaks (which, by symmetry of the autocorrelogram, are the same as the last three peaks, up to a rotation of 180 degrees) were defined as the three orientations of the grid $(\theta_1, \theta_2, \theta_3)$. With these orientations, we then computed the values of the autocorrelation along lines that emanated from the center of the autocorrelogram and went the length of the autocorrelogram, at each of three orientations. The distances along the lines where the first non-trivial peak was located were defined to be the grid spacings $(\lambda_1, \lambda_2, \lambda_3)$. For some analyses, we considered a single grid spacing (λ) and orientation (θ) for the entire grid,

defining those to be the mean of the three individual quantities.

To characterize the within and between GC variability of grid spacing and orientation, we employed the following approach. First, we split the data into time bins of fixed length (both 1 second and 5 second bins were considered). From this, we randomly assigned each interval to one of two blocks (denoted as blocks a and b), with exactly half the total number of intervals in each block. For each GC, we computed the grid spacing ($\lambda_a^{(i)}$ and $\lambda_b^{(i)}$) and orientation ($\theta_a^{(i)}$ and $\theta_b^{(i)}$), from the data in each block. The within cell differences of grid spacing and orientation are determined as $\Delta\theta^{(i)} = \theta_a^{(i)} - \theta_b^{(i)}$ and $\Delta\lambda^{(i)} = \lambda_a^{(i)} - \lambda_b^{(i)}$. Each GC's properties were also compared to another GC, with the match being made using random sampling without replacement. These comparisons were determined as $\Delta\theta_{\text{shuffle}}^{(i,j)} = \theta_a^{(i)} - \theta_b^{(j)}$ and $\Delta\lambda_{\text{shuffle}}^{(i,j)} = \lambda_a^{(i)} - \lambda_b^{(j)}$. This process was repeated 100 times, per recording. The resulting distributions of $\Delta\theta$, $\Delta\lambda$, $\Delta\theta_{\text{shuffle}}$, and $\Delta\lambda_{\text{shuffle}}$, across GCs and splits of the data were then compared. If the variability in grid spacing and orientation for each GC was smaller than that between GCs, we concluded that the variability of GC properties was a robust feature of the data and not due to noise.

3.2.4 Synthetic grid cells

To study how variability in GC properties might endow the grid code with computational advantages, we generated synthetic GC ratemaps, so that specific manipulations could be performed. These synthetic GC ratemaps were constructed as follows.

First, the lengths of the "arena", within which the simulated GCs exist, along each dimension (L_x and L_y), were set. Then, for $N \in \mathbb{N}$ GCs, the grid spacing and orientation were sampled via $\lambda^{(i)} \sim \mathcal{N}(\mu_\lambda, \sigma_\lambda)$ and $\theta^{(i)} \sim \mathcal{N}(\mu_\theta, \sigma_\theta)$ where $\mathcal{N}(\mu, \sigma)$ is a normal distribution with mean μ and variance σ^2. Grid phase was sampled as $\phi^{(i)} \sim \mathcal{U}([0, L_x] \times [0, L_y])$, where ϕ is a two dimensional vector, with first component uniformly sampled from $[0, L_x]$

and second component uniformly sampled from $[0, L_y]$. To construct a population with no variability in grid properties (to use as a control), we set $\sigma_\lambda = 0$ m. and $\sigma_\theta = 0°$.

For each GC, we generated idealized grid responses by summing three two-dimensional sinusoids [73], such that the activity at $\mathbf{x} = (x, y) \in [-L_x/2, L_x/2] \times [-L_y/2, L_y/2]$ is given by

$$X^{(i)}(\mathbf{x}) = X^{\max}\frac{2}{3}\left(\frac{1}{3}\sum_{j=1}^{3}\cos\left[R(\theta_i)\mathbf{k}_j(\mathbf{x}+\phi_i)\pi/\lambda_i\right]\right), \tag{3.1}$$

where $X^{\max}$ is the maximal firing rate, $R(\theta_i)$ is the two-dimensional rotation matrix, with rotation θ_i, and $\mathbf{k}_j$ are the wave vectors with 60° and 120° angular differences [i.e., $\mathbf{k}_1 = (1, 0)^T, \mathbf{k}_2 = (1/2, \sqrt{3}/2)^T, \mathbf{k}_3 = (1/2, -\sqrt{3}/2)^T$].

3.2.5 Linear decoding

To determine the extent to which local spatial information can be decoded from the activity of populations of GCs with different degrees of variability in their grid properties, we performed the following analysis.

We generated noisy synthetic spike rates of N GCs by assuming a Poisson process and sampling using the idealized ratemaps found as described in Sec. 3.2.4. More concretely, the activity of GC i at position (x, y) was assumed to be a random variable with a Poisson distribution, whose mean was $X^{(i)}(x, y)$,

$$\tilde{X}^{(i)}(x, y) \sim \mathcal{P}\left(X^{(i)}(x, y)\right). \tag{3.2}$$

For a given resolution of the arena, we generated $\tilde{X}_j^{(i)}(x, y)$, where $j = 1, ..., 10$. That is, we constructed 10 noisy ratemaps. We performed cross-validated decoding by averaging across 9 of the 10 ratemaps, to get an average ratemap $\hat{X}^{(i)}(x, y)$. For sake of

simplicity, consider $\hat{X}^{(i)}(x,y) = 1/9 \sum_{j=1}^{9} \tilde{X}_j^{(i)}(x,y)$. To decode local position from the held out noisy ratemap [e.g. $\tilde{X}_{10}^{(i)}(x,y)$], we multiply the activity at each position by all the positions in the average ratemap, taking the sum and assigning the decoded position as that with the largest value. The Euclidean distance between the decoded position and the true position is considered the error.

3.3 Results

3.3.1 Consistent Difference in Grid Orientation Across Grid Modules

To determine the extent to which variability of grid properties in individual modules exists, and to what extent this variability is due to noise, as opposed to being a property of the underlying grid code, we analyzed previously published data [72] (see Sec. 3.2.1 for more details). We focused on the six recordings with the most GCs with high grid scores (see Sec. 3.2.3), for a total of 316 GCs, across four modules in three rats.

A visualization of the distribution of GC properties for an example recording (Rat R Day 1 Module 2 – 58 GCs with grid score > 1.0) is presented in Fig. 3.1A, B. While many of the computed values of grid spacing (λ) and orientation (θ) are near the module mean, both of these distributions have a sizeable range ($\lambda_{\max} - \lambda_{\min} \approx 0.3$m and $\theta_{\max} - \theta_{\min} \approx 8^{\circ}$).

To assess the robustness of this variability, and to determine whether it may be due to noise or a property of the underlying grid code, we performed a shuffle analysis, splitting data into two equal length, non-overlapping halves (see Sec. 3.2.3 for details). For ease of notation, we refer to one half as "even" and the other as "odd", although we do not actually split the data into odd and even intervals, but instead randomly sample. For

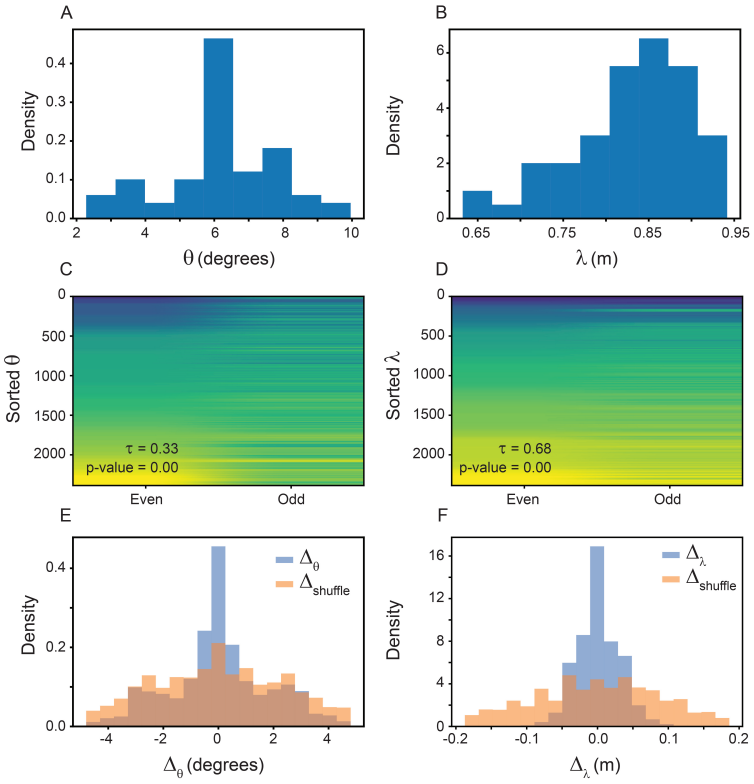

Figure 3.1: **Grid spacing and orientation is variable across an individual grid module.** (A)-(B) Distribution of θ and λ across individual grid module [72] (58 GCs total). (C)-(D) θ and λ computed across 50 separate splits of the spiking data, sorted according to the ordering in the "even" half. (E)-(F) Difference in θ and λ, across each split of the data (Δ_λ and Δ_θ), compared to a shuffled distribution, where the difference was computed for a randomly selected cell. (C)-(F) were computed from data first binned into 1 second, non-overlapping intervals.

both grid spacing and orientation, we find that, across 50 random shuffles of the data, the ordering of θ and λ is fairly conserved (Fig. 3.1 C, D – Kendall correlation: $\tau_\lambda = 0.68$, p-value < 0.05, $\tau_\theta = 0.33$, p-value < 0.05). We compared the difference of each cell's computed λ and θ for odd and even trials ("within" GC – Δ_λ and Δ_θ) and between the odd and even trials of randomly sampled GCs ("between" GCs – Δ_{shuffle}). We find that the distribution of Δ_{shuffle} is much broader, and has less of a peak at 0, than the respective Δ_{shuffle} distributions (Fig. 3.1E, F).

We performed the same analysis on the other five recordings, finding similar results.

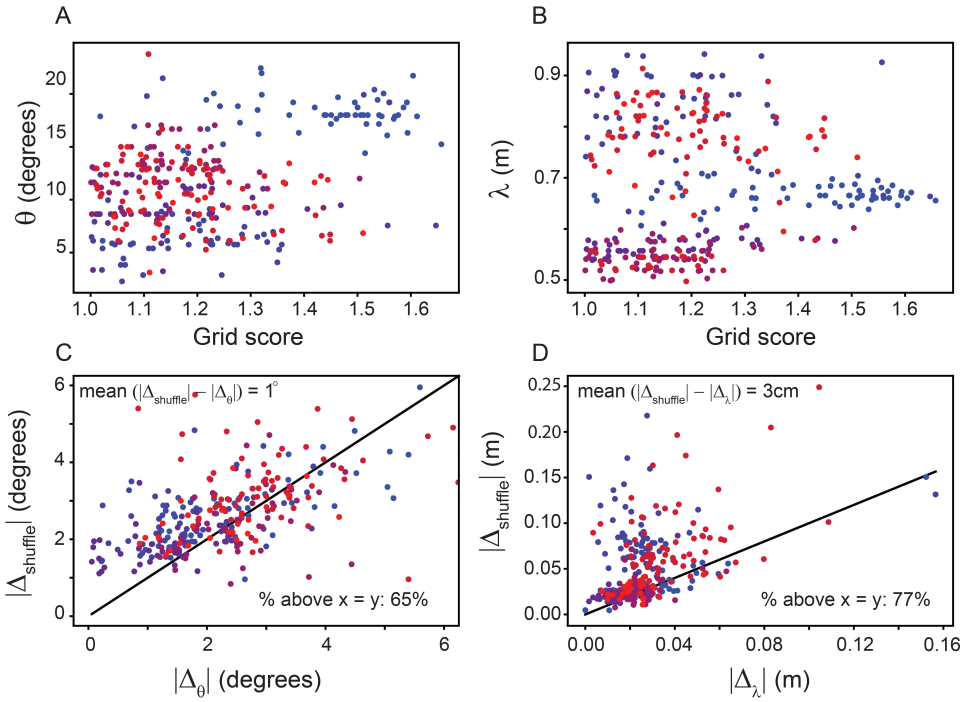

Figure 3.2: **Variability of grid spacing and orientation is more consistent than expected from noise.** (A)-(B) Grid orientation and spacing for all six recordings considered (3 rats, 4 modules, 316 GCs), as a function of grid score. Dots colored by recording each GC belongs to. (C)-(D) Mean Δ_λ and Δ_θ versus their respective shuffle values, Δ_{shuff}, computed across 50 randomized splits of the data. Percentage of cells with the within cell measure greater than the between cell measure (denoted as "% above x = y") and mean difference for those cells, are reported. (C)-(D) computed from data first binned into 1 second, non-overlapping intervals.

To quantify this across the entire data set, we computed the average value of Δ_λ and Δ_θ, as well as their respective shuffled distributions Δ_{shuffle}, across all 50 randomizations of the data, for each GC. We find that the majority of the GCs have greater average Δ_{shuffle}, than Δ_λ and Δ_θ, respectively (Fig. 3.1 C, D – 65% of GCs satisfy $|\Delta_\theta| < |\Delta_{\text{shuffle}}|$, 77% of GCs satisfy $|\Delta_\lambda| < |\Delta_{\text{shuffle}}|$). For those GC with smaller degree of within cell variation, than between cell variation, we find that the mean $|\Delta_{\text{shuffle}}| - |\Delta_\theta| \approx 1^\circ$ and the mean $|\Delta_{\text{shuffle}}| - |\Delta_\lambda| \approx 3.0$cm.

These results do not depend on the bin size used to divide the data into intervals to be sampled from during the shuffling procedure (Figs. 3.3 and 3.4).

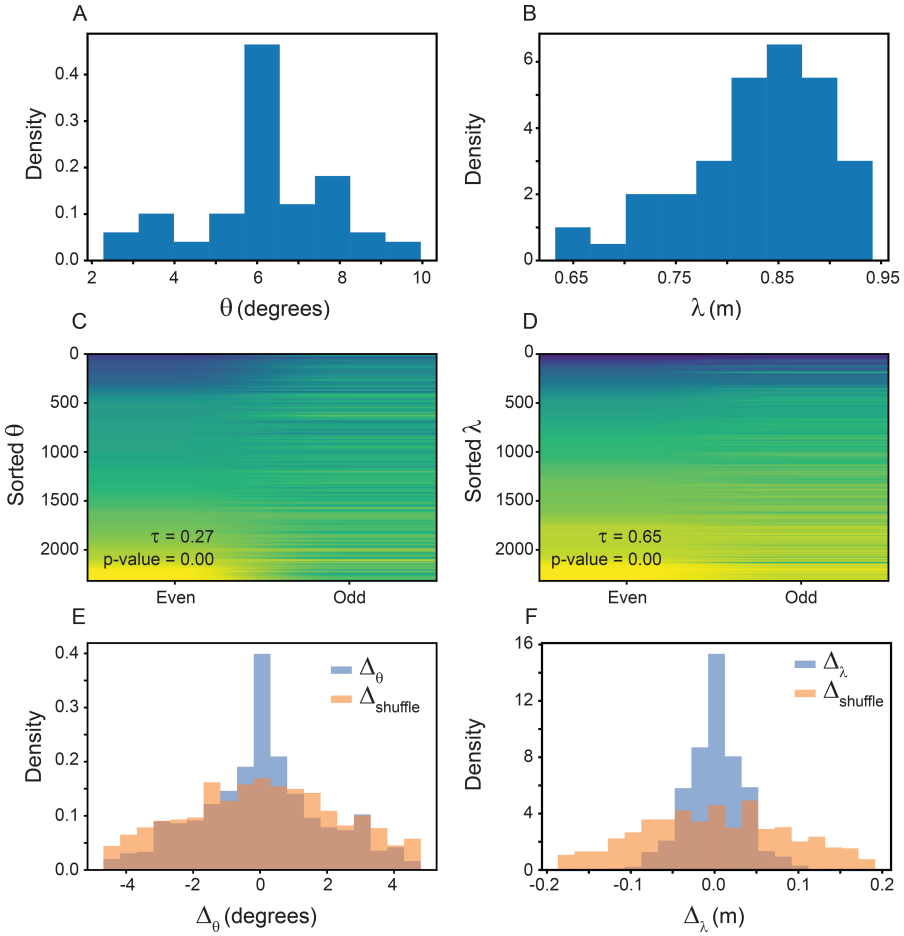

Figure 3.3: **Grid spacing and orientation is variable across individual grid module, for increased choice of time bins.** Same as in Fig. 3.1, but for using 5 second, non-overlapping intervals.

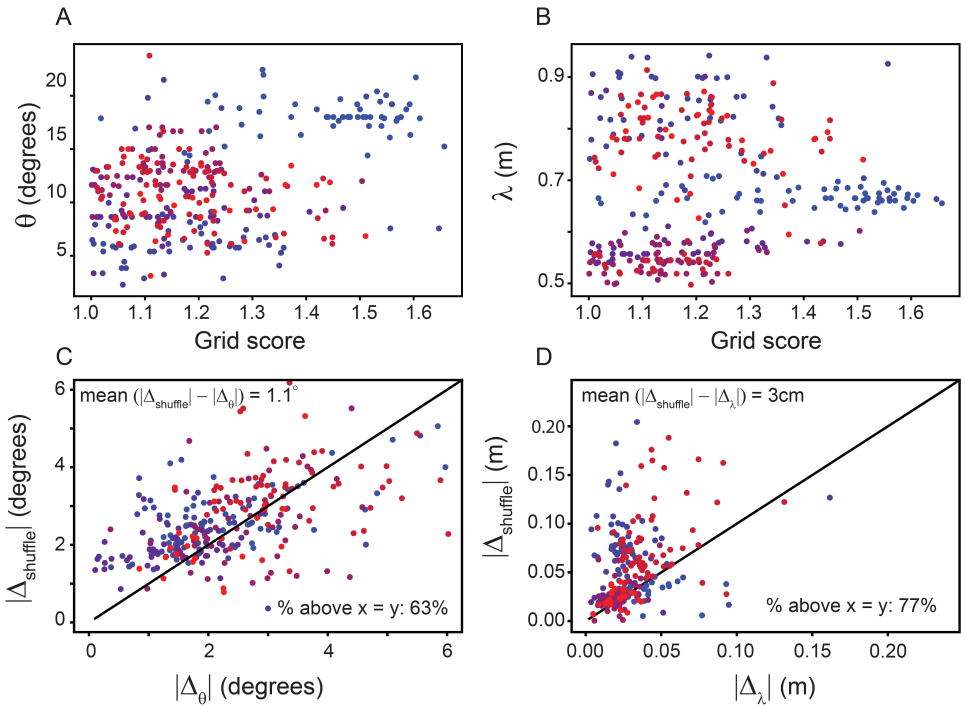

Figure 3.4: **Variability of grid spacing and orientation is more consistent than expected from noise, for increased choice of time bins.** Same as in Fig. 3.2, but for using 5 second, non-overlapping intervals.

Because previous work has found that grid fields can become distorted, shearing in response to the presence of objects and boundaries placed in arenas, as well as having small deviations from true hexagonal grids in open field arenas [229, 230], we examined to what extent the variability of grid properties changed when measured for each of the three unique peaks in the autocorrelogram separately. Note that there are three, and not six, unique peaks because the autocorrelogram is symmetric. We denote the grid spacing and grid orientation $\lambda_1, \lambda_2, \lambda_3$ and $\theta_1, \theta_2, \theta_3$, respectively.

Analyzing the same grid module as before (Fig. 3.1), we find that, in addition to the individual grid fields having different medians of their respective distributions (median $\lambda_1 = 0.86$m, median $\lambda_2 = 0.84$m, median $\lambda_3 = 0.82$m; median $\theta_1 = 5.6°$, median $\theta_2 = 5.9°$, $8.5°$), suggesting that there are indeed some distortions, the individual grid fields also show significant range in their spacing and orientation ($\lambda_{1,\max} - \lambda_{1,\min} \approx 0.3$m, $\lambda_{2,\max} - \lambda_{2,\min} \approx 0.5$m, $\lambda_{3,\max} - \lambda_{3,\min} \approx 0.4$m; $\theta_{1,\max} - \theta_{1,\min} \approx 8°$, $\theta_{2,\max} - \theta_{2,\min} \approx 14°$, $\theta_{3,\max} - \theta_{3,\min} \approx 14°$) (Fig. 3.5A, B).

As before, splitting the data into non-overlapping halves and computing the grid properties (this time, on the individual fields) leads to similarly ordered values (Fig. 3.5C, D – Kendall correlation: $\tau_\lambda = 0.66$, p-value < 0.05; $\tau_\theta = 0.47$, p-value < 0.05). Performing the shuffle analysis similarly identifies that the variability of grid spacing and orientation within GC is much smaller than the variability of grid spacing and orientation across GCs (Fig. 3.5E, F). This result holds when considering all recordings (Fig. 3.6C, D). For those GC fields that had smaller within variability in their computed for their computed grid properties, as compared to the shuffled distribution, $|\Delta_{\text{shuffle}}| - |\Delta_\lambda| \approx 3.4$cm and $|\Delta_{\text{shuffle}}| - |\Delta_\theta| \approx 1.9°$.

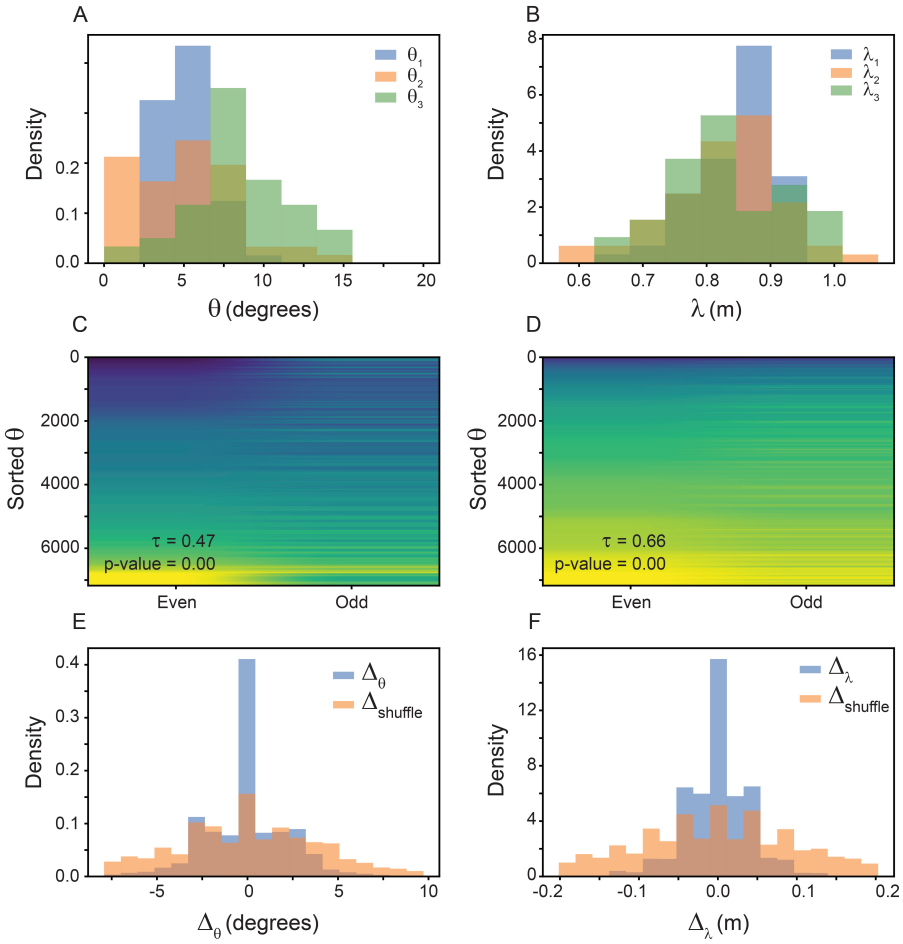

Figure 3.5: **Grid spacing and orientation is variable across individual grid fields.** Same as Fig. 3.2, except all individual grid peaks' λ and θ are reported, as opposed to averaged.

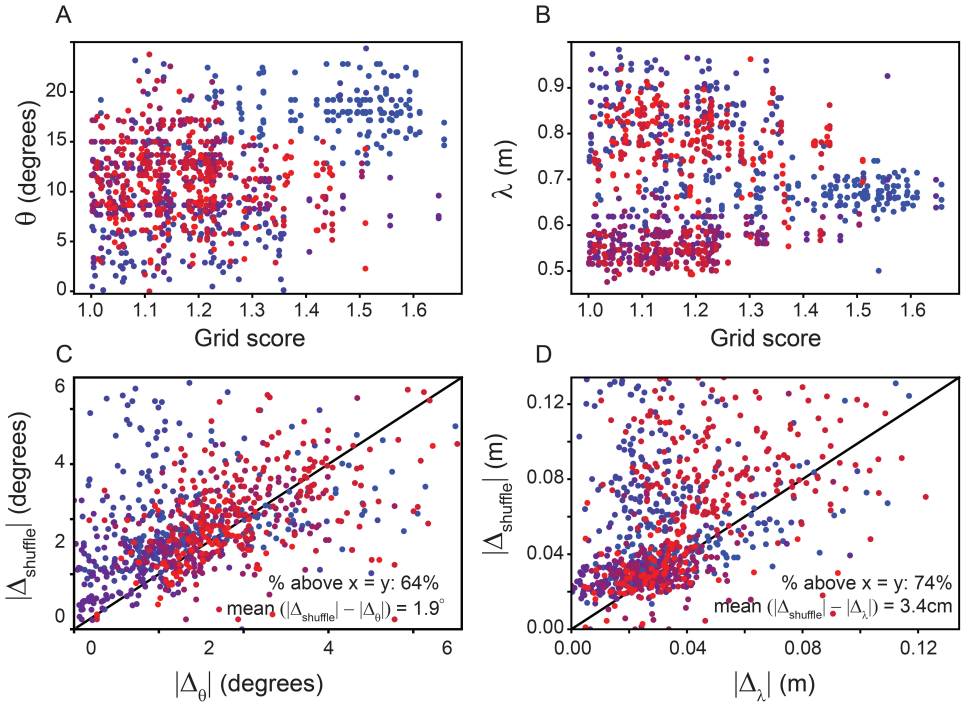

Figure 3.6: **Variability of grid spacing and orientation is more consistent than expected from noise, when treating each grid field separately.** Same as Fig. 3.2, except all individual grid peaks' λ and θ are reported, as opposed to averaged.

3.3.2 Variability in Grid Orientation Across Single Module Improves Spatial Decoding

The variability in GC properties identified via the data analysis performed in Sec. 3.3.1, while statistically significant, is small. Can a computational benefit be gained from such inhomogeneity?

To test this at a level where we have complete control over the variability of GC responses (σ_λ and σ_θ), we generated populations of synthetic GCs (see Sec. 3.2.4 for more details). By sampling from a Poisson process, we were able to create noisy "GC-like" ratemaps (Fig. 3.7A) for populations of varying sizes. Leveraging a simple linear decoder (see Sec. 3.3.2, we were able use these populations of noisy GCs to generate likelihood densities of where a given population vector came from. When the population of synthetic GCs being considered had no variability in λ and θ ($\sigma_\lambda = 0$m and $\sigma_\theta = 0°$), the likelihood densities were highly grid like, leading to large decoding errors, as the decoded position was essentially randomly chosen among the peaks of the grid (Fig. 3.7B, bottom row). However, when the simulated GC population had variability like that in the data analyzed ($\sigma_\lambda = 0.03$m and $\sigma_\theta = 1°$), the "gridiness" of the likelihood densities was decreased, and the decoded position was much closer to the true position (Fig. 3.7B, top row). This difference in decoding error, between the fixed and variable simulated GC populations, was present across the entirety of the simulated arena, except for the center (Fig. 3.7 C).

Despite the observed variability being small, we find that the decoding error scales like a powerlaw, with respect to GC population size. That is, the decoding error for a population of 2^N simulated GCs, $e(2^N)$, decreases as

$$e(2^N) \propto \left(2^N\right)^{-\beta} \tag{3.3}$$

for some exponent β (Fig. 3.7D, green line). Plotting this on a $\log_2$ scale gives us that the error scales linearly, which is what we observe (Fig. 3.7D). This is in contrast with the nearly static decoding performance of a population of GCs with fixed grid properties (Fig. 3.7D, black line). By 512 neurons, the mean decoding error of the variable population is $\approx 6\times$ smaller than that of the fixed population. This is *despite* the fact that the σ_θ is a single degree, and σ_λ is on the order of a centimeter.

To determine the extent to which the increase in decoding performance depends on the exact variability, we performed a grid search over 36 pairs of $(\sigma_\lambda, \sigma_\theta)$ values in $[0°, 5°] \times [0.00\text{m.}, 0.05\text{m.}]$. As expected, when the variability of both grid properties was large, we found a large decrease in decoding error (Fig. 3.7E). However, even when one $\sigma_\lambda = 0$ or $\sigma_\theta = 0$, increasing σ_θ or σ_λ, respectively, even over a small range, led to a large decrease in decoding error. Finally, we find that, even when $\sigma_\lambda = 0.01$m and $\sigma_\theta = 1°$, the decoding error is still decreased by approximately half of what it was for the fixed population.

Similar results were found when using GCs of different grid spacing ($\lambda = 0.75$ m. shown in Fig. 3.7, $\lambda = 0.5$ m. and $\lambda = 1.0$ m., shown in Fig. 3.8). However, a greater improvement was found when λ was smaller, presumably due to the fact that the number of grid fields was greater when λ was smaller, as well as the fact that $\sigma_\lambda = 0.03$ m. is a greater percentage of λ when λ is small.

Taken together, these results suggest that *individual* grid modules can gain a significant computational benefit, in terms of encoding local space, by containing heterogeneity, even to a small degree, in their grid properties.

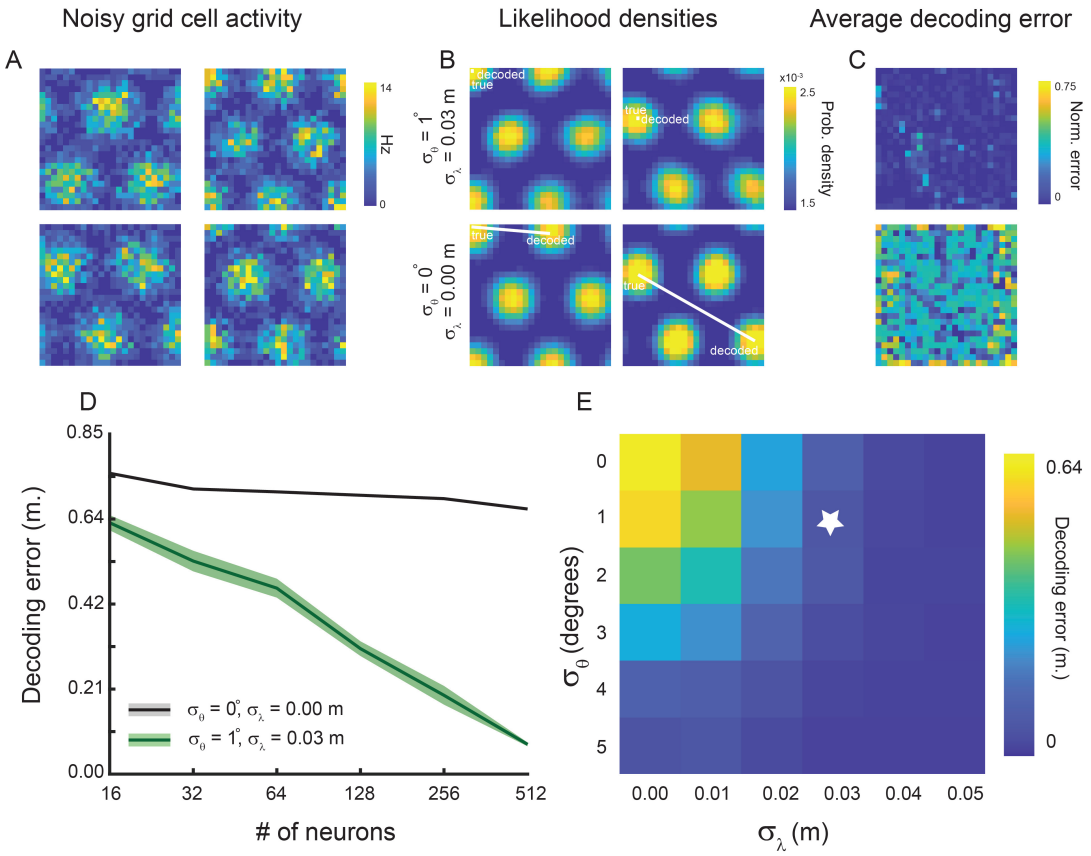

Figure 3.7: **Variability in GC properties enables increased decoding from activity from single grid module.** (A) Example noisy GC ratemaps sampled from Poisson process. Size of square arena taken to be 1.5m × 1.5 m [72]. (B) Likelihood estimates from linear decoder applied to the population vectors of 512 noisy GC inputs, for two example positions. Top row, variable GC population (with values for σ_θ and σ_λ taken from experimental analysis – Fig. 3.2). Distance between true and decoded position denoted by white line. (C) Mean decoding error, across 10 samples of GC input, for each position, across the entirety of the arena. Top and bottom rows same as in (B). (D) Decoding error, as a function of population size. Solid line is mean across 5 independent GC populations. Shaded area is ± S.E.M. of the 5 indendepnt populations. (E) Decoding error, over a grid of σ_λ and σ_θ values. White star denotes value used in plots (B) - (D).

3.4 Discussion

While comprising a minor portion of the neurons in mEC [87], GCs have been subject of major investigation. The organization of their multiple firing fields along a triangular lattice, a unique functional feature, has been historically interpreted as a limiting feature of their encoding of local space, but a powerful feature enabling the encoding of path integration [67]. Particularly influential in shaping this view has been the discovery of the distribution of GCs into distinct modules, with GC spacing (λ) and orientation (θ) preserved within, but not across, modules [6, 55]. While evidence for discontinuity in the GC properties across modules is strong, the corollary assumption, that within module the values of λ and θ are, up to noise, identical, has not been studied.

Dissecting recently performed, state-of-the-art mEC recordings led to the identification of variability in the grid properties within individual modules (Fig. 3.1). Despite considering only GCs that were unequivocally GCs (grid score > 1.0), we found this variability was large, with GC in the same module having up to a 0.3 m. difference in grid spacing and 8° difference in grid orientation. Statistical analysis of this variability shows that it is more robust, for the majority of GCs, than expected from noise (Fig. 3.2). Importantly, our analysis is robust to choices of parameters (Fig. SI2) and whether we treat each grid field independently, or take the average across grid fields (Fig. 3.6).

We speculated that this variability can be used to increase the fidelity at which *individual* grid modules can encode local space. This is in-line with a large literature showing that heterogeneity in the responses of populations of neurons increases the robustness of encoding [238, 239, 240, 241]. We indeed find, in noisy models of GCs, that the observed level of variability in grid properties leads to highly accurate encoding of spatial position (Fig. 3.7). This benefit is increased with increasing number of neurons by approximately a power-law (Fig. 3.7D), is general to even smaller amounts of variability (Fig. 3.7E),

and was most pronounced in modules with smallest λ (Fig. 3.8), although even the module with $\lambda = 1.0$ m saw a considerable decrease in decoding error.

We note that these results extend recent work which has argued that heterogeneity in maximum firing rates across individual grid fields [87, 231, 232] and grid sheering [229, 230, 233] lead to the ability to encode local spatial information. That is, even in the absence of any of these types of perturbations, individual grid modules can contain information about spatial position. This suggests that the encoding of local spatial information is a fundamental property of the grid code.

Importantly, these results motivate the field to ask "why are multiple grid modules present in the mammalian brain"? We hypothesize that the use of GCs beyond spatial navigation [242, 243, 105, 244, 245, 246], where hierarchical representations are important [247, 111], may be a sufficient implicit bias for the formation of multiple modules, extending upon the recent work that has shown that grid responses emerge in recurrent neural networks (RNNs) tasked with path integration [82, 81, 64, 83]. This is aligned with calls for asking different questions of GCs beyond purely path integration [248].

Finally, given that previous attempts to add variability in maximum firing rates of grid fields to the continuous attractor neural network models [66, 68, 67, 69] that have enjoyed broad experimental support [70, 71, 236, 72] has been challenging [232], we wonder whether and how the heterogeneity of grid properties discovered in this work can be naturally extended to continuous attractor neural network. Future work can explore whether this variability is an emergent phenomenon of RNN models [82, 81, 64, 83], as well as other mechanistic models that generate GCs [65].

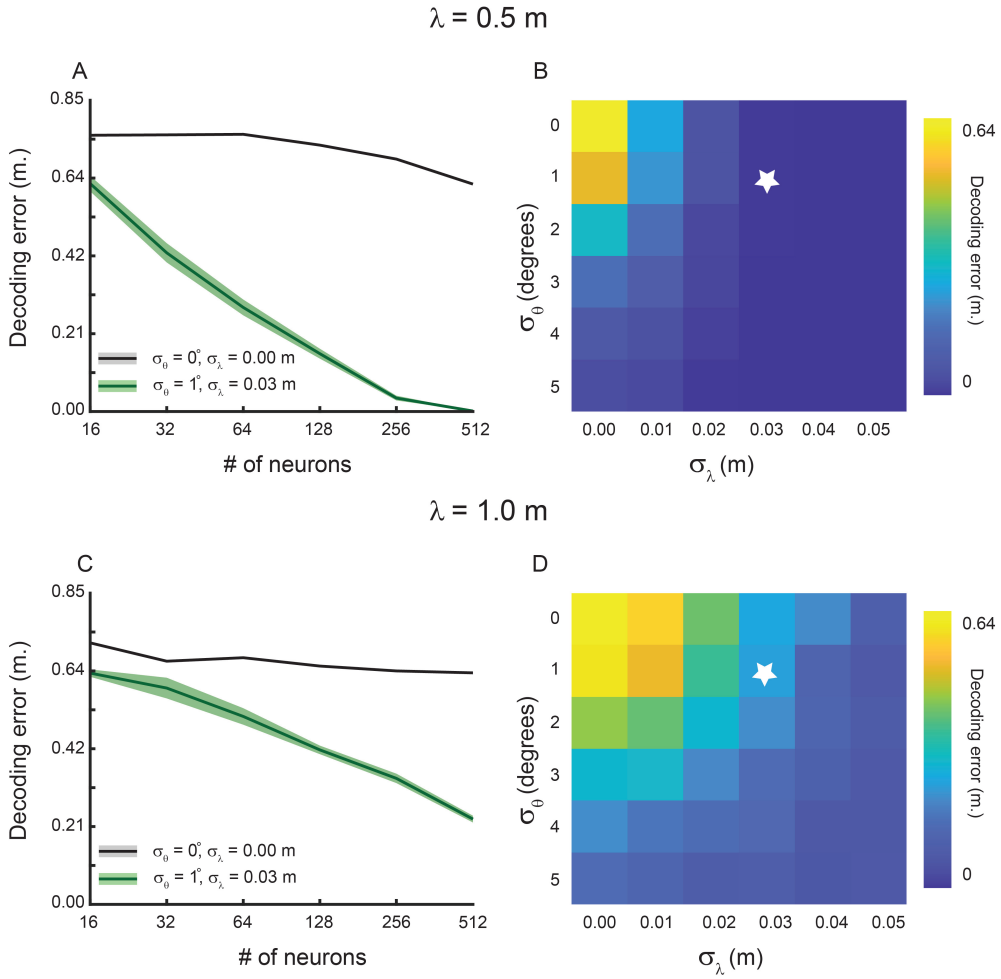

Figure 3.8: **Variability in GC properties enables improved decoding of local space from the activity of GCs from individual grid modules with different spacing.** (A), (C) Decoding error, as a function of population size, for $\lambda = 0.5$ m. and $\lambda = 1.0$ m., respectively. Solid line is mean across 5 independent GC populations. Shaded area is $\pm$ S.E.M. of the 5 independent populations. (B), (D) Decoding error, over a grid of σ_λ and σ_θ values, for $\lambda = 0.5$ m. and $\lambda = 1.0$ m., respectively. White stars denotes value used in plots (A), (C).

Chapter 4

Sparsity Enhances Non-Gaussian Data Statistics During Local Receptive Field Formation

"I asked her for water/ She brought me gasoline"

Howlin' Wolf (1959)

4.1 Introduction

The first wave of deep learning breakthroughs came in computer vision, when deep convolutional neural networks (CNNs) [115, 116, 117], armed with the necessary compute and (large) amounts of data [249], achieved huge improvements on the ImageNet benchmark. This success has been attributed to the interplay of network architecture and symmetries in the data. The architecture of deep CNNs [118, 119] encourages representations to be invariant or equivariant to translations, depending on whether pooling

layers are included or not [120, 121, 122]. The learned representations thus respect the same fundamental symmetry as the images they are trained on, making CNNs more sample- and parameter-efficient than fully-connected neural networks (FCNs), which do not explicitly take this symmetry into account [123, 124].

As deep learning moves into ever more diverse applications, it is confronted with data sets whose symmetries are not necessarily known *a priori*. It would thus be desirable to "learn" the right architecture directly from the data. For image classification however, learning convolutions from scratch has proven difficult with vanilla stochastic gradient descent (SGD) on image classification tasks [124, 129], and prior successful work has relied on elaborate regularisation mechanisms [130].

While the combination of vanilla SGD on image classification tasks does not yield receptive fields by itself, two recent works observed receptive field formation in slightly augmented settings. Ingrosso & Goldt (2022) [128] showed that two-layer fully-connected networks trained on synthetic images that are translation-invariant and contain sharp edges learn localised receptive fields from scratch, without any explicit regularisation. In particular, they showed that the signal in the training data that triggers the formation of the localised RFs was carried by the higher-order, i.e. non-Gaussian, input statistics, and in particular the fourth-order cumulant related to the kurtosis. While this work offered a novel statistical perspective on receptive field formation and how such structure can be learnt from "scratch", computer vision data sets have non-Gaussian properties, such as sharp luminosity changes, and have therefore non-trivial higher-order cumulants. So why do FCNs trained on such data sets not contain local RFs?

One possible explanation is that this structure is **embedded** within trained FCNs. Indeed, when Pellegrini & Biroli (2022) [127] analysed the architecture resulting from iterative magnitude pruning (IMP) [250, 251] of simple, fully-connected neural networks, they found that the surviving node connectivity is local in input space, and organized in

patterns reminiscent of the ones found in convolutional networks. In other words, SGD on image classification tasks imprints a localised structure on the weights, which can be recovered with the help of IMP, but it is not clear where the signal for the localisation of the receptive fields comes from.

These results raise two questions: first, do embedded local RFs form in FCN trained on real images in the same way as suggested by Ingrosso and Goldt (2022), namely via higher-order cumulants? Second, given that use of IMP leads to stronger localisation, does IMP amplify the non-Gaussian statistics of the learnt representations of the network?

In this work, we take the first step to address these questions. Our main contributions are as follows:

- We find that the increasing localisation of RFs, embedded in FCNs during training, is accompanied by neural representations becoming increasingly non-Gaussian (Sec. 4.4.1). Changing the task and hyper-parameters associated with training leads to changes in pre-activation statistics that are concomitant with those found in localisation (Sec. 4.4.2).

- We analyze the impact of pruning dynamics on the statistics of network internal representations, finding that each round of IMP makes the statistics of representations increasingly non-Gaussian, as measured by their kurtosis (Sec. 4.4.3). To the best of our knowledge, this is the first time that the statistics of internal representations during IMP have been characterised.

- We study two forms of dynamic sparse machine learning, L1 weight regularisation and RigL [252], a form of dynamic sparse training (DST). While L1 regularisation leads to localisation and a concomitant increase in pre-activaton kurtosis, RigL does not (Sec. 4.4.4). We hypothesize that starting from a dense FCN architecture

(as pruning and L1 regularisation do) may enable more efficient discovery of local RFs.

- We show evidence that increases in non-Gaussian preactivation statistics accompanying local RF formation goes beyond simple FCNs trained on supervised computer vision tasks. In particular, we find increases accompanying localisation of RFs during an unsupervised learning task [125] (Sec. 4.5).

4.1.1 Further related work

Learning the right architectures for various data modalities. A large amount of work has recently proposed architectures that are tailored to different invariances and geometries in datasets, such as social or gene regulatory networks [253, 254, 255, 256, 257, 258, 259, 260, 261]. Deep scattering networks [262, 263, 264] have been proposed as architectures that are invariant to a rich class of transformations. The vision transformer [265, 266, 267] is such a general-purpose architecture that learns approximately localised attention maps when pre-trained on large amounts of images. Here, we focus instead on training the simplest general architecture, an FCN, and trying to learn convolutions directly from data. Neybashur (2020) recently proposed a lasso-type regularisation technique, called β-LASSO, which encourages localised RFs, at the price of explicitly encouraging sparsity, which may not be the appropriate constraint for general data modalities [130]. Additionally, this approach relied on extensive data augmentation, whose role in driving local RF formation is not known.

What is learnt from pruning? After the discovery of "winning lottery tickets" in the initialisation of deep neural networks [251], a considerable body of work has focused on extending the original idea to larger data sets and architectures, e.g. [268]. While

a detailed understanding the inductive biases that come with IMP is lacking, it has been found that winning tickets have properties that make them transferable across tasks [269, 270, 271, 272]. Additional attention has been focused on understanding the lottery ticket hypothesis through a loss landscape perspective [268, 273, 274]. Recent work has studied the interaction between data and masks found via IMP [275, 276]. To the best of our knowledge, our work is the first to analyse the effect of IMP on the statistics of the internal presentations in neural networks.

4.2 Methods

4.3 Methods

4.3.1 Measuring Non-Gaussian Statistics

We denote a neural network by $f(\theta; x)$, where θ are its N parameters and x are data samples used for training, which come from an underlying distribution, $\mathcal{X}$.

The activations of units in the first hidden layer of the neural network are given by

$$a_i^{(1)} = h\left[\sum_j w_{ij}^{(1)} x_j\right] + b_i^{(1)}, \tag{4.1}$$

where $h(\cdot)$ is a nonlinear function (e.g., ReLU, sigmoid), $w_{ij}^{(1)}$ are the weights going from the data to the first hidden layer, and $b_i^{(1)}$ is the bias of unit i in the first hidden layer.

Note that, given Eq. 4.1, the input to the neural network is not x, but the product of x with the $W^{(1)}$, the weight matrix describing all $w_{ij}^{(1)}$. This quantity is referred to as

the "preactivation", and is computed as

$$\lambda_i = \sum_j w_{ij}^{(1)} x_j = W_i^{(1)} x. \tag{4.2}$$

In some cases, the data is transformed using batch normalization [277], which applies the following transformation

$$\tilde{x}_j = \gamma_j x_j + \beta_j, \tag{4.3}$$

where γ_j and β_j are free parameters that are learned during the course of neural network training. In this case, Eq. 4.2 is modified such that $\lambda_i = W_i^{(1)} \tilde{x}$.

One way to measure the statistics of the input to the neural network, that is, to measure the statistics of the preactivations, is to consider how Gaussian it is. To quantify this, the kurtosis can be computed

$$\text{kurt}[\lambda_i] = \frac{\mathbb{E}\left(\lambda_i - \mathbb{E}\lambda_i\right)^4}{\left[\left(\lambda_i - \mathbb{E}\lambda_i\right)^2\right]^2}, \tag{4.4}$$

where $\mathbb{E}$ is the expectation (mean) computed over some sample of x's. If λ_i is distributed according to a Gaussian distribution, then $\text{kurt}[\lambda_i] = 3$. If $\text{kurt}[\lambda_i] > 3$, then the distribution in more peaked than a Gaussian. If $\text{kurt}[\lambda_i] < 3$, the distribution is more broad than a Gaussian.

4.3.2 Measuring Localisation

To quantify how local the remaining weights of a pruned FCN are, we compute the distribution of pixels at a given distance that are connected to the same input unit in the first layer of the FCN, $S(\mathbf{d})$, as per Pellegrini and Biroli (2022). Distance, in this case, is defined as $\mathbf{d}_{xx'} = \mathbf{d}[(x, y), (x', y')] = (x - x', y - y')$, where (x, y) and (x', y') are the spatial

coordinates of the two pixels. $S(\mathbf{d})$ is defined as, $S(\mathbf{d}) = \sum_{(x,y)} \sum_{(x',y')|\mathbf{d}_{xx'}=\mathbf{d}} m_{x,y} m_{x',y'}$. The sum can be restricted to subsets of the mask that come from pixels in the same color channel, $S^{(\mathrm{sc})}(\mathbf{d})$, or different channels, $S^{(\mathrm{dc})}(\mathbf{d})$. We quantify *receptive field locality* by fitting two-dimensional Gaussians on $S(\mathbf{d})$, giving a measure of width in the x and y directions, σ_x and σ_y respectively.

4.3.3 FCN Training

For our experiments, we follow the approach of Pellegrini and Biroli (2020) [127]. We used the official implementation publicly available at `https://github.com/phiandark/SiftingFeatures`, with several modifications. First, we scaled the FCN down to have two hidden layers, instead of three. Second, we decreased the batch size from 1000 to 40. Lastly, we decreased the total number of training iterations from $100,000$ to $40,000$. For more details, see Table 4.1. Similar choices were made when training our FCNs on CIFAR-10 and CIFAR-10 GP (Table 4.2).

Table 4.1: Default FCN parameters for ImageNet32 experiments (unless otherwise noted).

Parameters	
Architecture	3072:1024:1024:1000
Activation function	ReLU
Batch size	40
Learning rate	0.1
Optimizer	SGD
Training iterations	40,000
Validation size	7,000
Rewind iteration	1,000

Table 4.2: Default FCN parameters for CIFAR-10 and CIFAR-10 GP experiments (unless otherwise noted).

Parameters	
Architecture	3072:1024:1024:10
Activation function	ReLU
Batch size	40
Learning rate	0.1
Optimizer	SGD
Training iterations	40,000
Validation size	1,000
Rewind iteration	1,000

4.3.4 Iterative Magnitude Pruning

A foundational result in the field of deep learning is that deep neural networks (DNNs) can (generally) have a large percentage of their parameters removed (i.e., set to 0) after training, with little to no effect on the DNN's performance. The removal of DNN parameters is called "pruning", and substantial work has been focused on determining which parameters are optimal, with respect to some metric, to remove [278, 279, 280, 281, 282, 250, 283]. These include using the magnitude of a parameter (i.e., its absolute value), the gradient of a parameter, and the Hessian of the parameter to determine. While these tend to be static measures, there exist methods that take into account dynamics [284], which, in certain cases, is equivalent to the static measures.

While DNN pruning was known to be possible, and possible to a large extent (i.e., DNNs admitted large sparsification using some of the pruning strategies), for several decades, training sparse DNNs, at initialization, was found to be challenging, yielding poor performance. This led to the paradigm that overparametrization, or dense connectivity, was necessary in order for standard DNN training techniques to find a well performing DNN state.

However, work on the Lottery Ticket Hypothesis (LTH) [251] hypothesized that sparse DNNs, that were trainable to the same performance as the dense model, from initialization (or near initialization) exist. To support this conjecture, they leveraged a simple technique, called iterative magnitude pruning (IMP) [250], to find such sparse subnetworks.

An explosion of work has been performed on the LTH, and it has become clear that the subnetworks discovered by IMP are intimately linked to the underlying geometry of the loss landscape [268, 274], are transferable to other tasks [269, 285, 270, 272, 286] and architectures [287].

IMP is performed according to the following steps. First, the DNN $f(\theta; x)$ is initialized with parameters θ_0. The DNN is trained, using the optimizer $\mathcal{T}$ (e.g., stochastic gradient descent) for T training steps. The parameters of the DNN are scored according to their magnitude $s(\omega_i) = |\omega_i|$. The parameters with score in the p^{th} percentile are set to 0, by setting an auxiliary variable $m \in \{0, 1\}^N$, called the "mask", to 0. That is, if $s(\omega_i) < \tau_p$, where τ_p is a threshold determined by the p^{th} percentile, then $m_i = 0$. The DNN is then evaluated as $f(\omega \odot m; x)$, where $\odot$ is the elementwise product. The thresholding of the parameters, τ_p, can be done using the p^{th} percentile, across the distribution of all the parameters in the entire DNN, or using the p^{th} percentile, computed for each layer individually. While the latter was performed initially with respect to the LTH [251], subsequent work has found that the former (a global threshold) leads to better accuracy [268, 283].

The remaining parameters (i.e., those parameters such that $m_i = 1$) are "rewound" to a previous value that occurred during training θ_{rewind}. While the original LTH found it possible to set $\theta_{\text{rewind}} = \theta_0$ [251], subsequent work found that, in order for IMP to find good performing sparse subnetwork, $\theta_{\text{rewind}} = \theta_t$, for some $t > 0$, must be chosen [268, 288]. The previous steps are then performed, with the p^{th} percentile of the remaining

parameters used as the threshold for pruning.

This entire process is iterated n times, inducing an trajectory in the space of DNN parameters, via $\theta_{\text{rewind}} \rightarrow \theta_{\text{rewind}} \odot m^{(1)} \rightarrow \theta_{\text{rewind}} \odot m^{(2)} \rightarrow ... \rightarrow \theta_{\text{rewind}} \odot m^{(n)}$. This iterative sparsificiation changes the way in which the units of the network are coupled, leading to a coarse-graining that has common characteristics for among different tasks [289].

4.3.5 L1 Weight Regularisation

In the main text, we consider adding L1 regularisation to the loss function. The loss function then becomes

$$\mathcal{L} = [\text{cross entropy}] + \lambda \sum_{i,j} |w_{ij}^{(1)}|, \tag{4.5}$$

where the sum is over all the weights between the input and the first hidden layer and λ is the weight of the regularisation. In our experiments, we set $\lambda = 5 \cdot 10^{-5}$. After training the same FCN detailed in Table 4.1, using the loss function described by Eq. , we oneshot pruned the FCN parameters.

4.3.6 L1 Sparse, Unsupervised Learning Model

The discovery of local RFs in the mammalian visual system [113], much like the discovery of place cells and grid cells in the hippocampal formation, inspired a large body of computational and theoretical work that aimed to understand how this functional property might be utilized in computations the brain performed. By the mid 1990's it was appreciated that local RFs, along with the other properties found in visually responsive neurons, such as orientation selectivity, enabled an efficient coding of the sensory data that the visual system deals with [290, 291, 292, 293, 294, 295]. In particular, the statistics of natural images, which have been found to contain unique properties

[294, 296]. Remarkably, efficient coding, from the standpoint of minimizing the activity, was then found to be a strong enough inductive bias to lead to the formation of local RFs [125], much like path integration has been found to be a strong enough inductive bias to lead to the formation of grid cells [82, 81, 64].

The framework for studying how enforcing sparse activations impacts local RF formation is the following [125]. A given image $I(x, y)$, with x and y indexing the n_x and n_y pixels in the x and y dimension of the image, respectively, is assumed to admit a linear decomposition into a set of N basis functions $\phi_i(x, y)$. That is,

$$I(x, y) = \sum_{i=1}^{N} a_i \phi_i(x, y), \tag{4.6}$$

where a_i are coefficients specific to each image. In the context of the visual system, the ϕ_i can be thought of as the tuning profiles of visual neurons, and the a_i as the activity of the neurons. Important, the ϕ_i are assumed to not only be a set of basis functions that describes the image $I(x, y)$, but all images that come from the same underlying distribution, $\mathcal{I}$.

To learn the basis functions ϕ_i, from data, an optimization problem is cast, using the following cost function,

$$E = \sum_{x,y} \left[I(x, y) - \sum_i a_i \phi_i(x, y) \right]^2 + \lambda \sum_i S\left(\frac{a_i}{\sigma}\right). \tag{4.7}$$

Here $S(\cdot)$ is a nonlinear function, σ is a scaling function, and λ is a positive real number, whose size controls how much the cost function depends on the activations. Note that $\sum_i a_i \phi_i(x, y)$ is the reconstructed image using the estimated basis functions and their associated coefficients. This reconstruction is denoted as $\hat{I}(x, y)$. For sake of computational resources, the training occurs not over individual images, but patch subsampled

on natural images (in Olshausen and Fields (1996), these images were 512×512 grayscale images taken in the woods [125]; in our work we use CIFAR-10 and CIFAR-10 GP).

To minimize Eq. 4.7, gradient descent is performed on the ϕ_i, where the update rule is performed as follows,

$$\Delta\phi_i(x_m, y_n) = \eta\mathbb{E}\left(a_i\left[I(x_m, y_n) - \hat{I}(x_m, y_n)\right]\right), \tag{4.8}$$

where η is the learning rate and the expectation, $\mathbb{E}(\cdot)$ is taken over a batch of images. For a given image, the value of a_i is estimated by taking the inner product of the image and the associated basis function, $a_i = \eta I\phi_i^T$.The a_i are then fed through the nonlinearity, and their values optimised using gradient descent, with learning rate η, such that the norm of vector A, composed of all N of the a_i, is sufficiently small. Because this occurs on the activations, after the nonlinearity has been performed, we measure the kurtosis of $a_i = I\phi_i^T$.

We implemented this approach using publicly available code `https://github.com/takyamamoto/SparseCoding-OlshausenField-Model`. We used all the off-the-shelf parameters, except for the following modifications: 1) we trained on grayscale CIFAR-10 or CIFAR-10 GP, instead of the grayscale natural images trained on by Olshausen and Fields (1996) [125]; 2) we used ReLU as our nonlinearity (Eq. 4.7). When training on CIFAR-10 GP, we create the Gaussian model for each patch that is sampled. For more details, see Table 4.3

4.3.7 Dynamic Sparse Training

While IMP, and many other pruning methods, transform a dense DNN (dense, in terms of connectivity) into a sparse DNN, other approaches, called "dynamic sparse training" (DST), start with a sparse DNN, and optimize which parameters are kept,

Table 4.3: Default parameters of L1 activation regularisation.

Parameters	
N	10
n_x	16
n_y	16
Batch size	250
η	0.01
Training iterations	500

while maintaining sparsity [297, 298, 299, 300]. DST methods have been found to be scalable to very large DNNs [301], have been cast in the framework of the LTH [301], and appear to gain some of their capability from being "dense in-time", by sampling a large proportion of the total possible connectivity in the DNN [301].

We focus on one approach of the many that DST encompasses, called RigL ("Rigging the Lottery") [252]. This was chosen due to its connection to the LTH. We made use of the publicly available repository of a PyTorch implementation of RigL, found at `https://github.com/nollied/rigl-torch`. For the choices used in our implementation, see Table 4.4.

This was chosen due to its connection to the LTH. RigL has four main steps. First, a sparse DNN is initialized with a given distribution. This can be a uniform distribution (each layer initialized with $(1-p)\%$ of its weights), or one that keeps more weights in some layers and less in others (e.g. Eros-Renyi distribution [298]). Here we use a uniform distribution, as it is consistent with the IMP work used to discover local RFs [127]. Second, a schedule to update the topology of the DNN (i.e., the identity of which weights kept and which are removed) is chosen. This has several free parameters, including the number of iterations between updates (ΔT), the iteration at which the updates stop and the topology is fixed for the rest of training (T_{end}), the initial fraction of the weights updated (α), and a decaying function that changes what percentage of

weights at each step are updated (e..g cosine annealing). To make our analysis simple, we fixed T_{end} to be the total number of iterations and use a constant decay function (that is, no decay, but a constant fraction of weights that are updated at each step). Third, a criterion for dropping weights must be selected. In RigL, and this work, parameter magnitude is used (i.e., $s(\theta) = |\theta|$). And fourth, a criteria for growing weights must be chosen. In RigL, this is taken to be the magnitude of the gradient of the parameters, (i.e., $|\nabla_\theta \mathcal{L}|$, for some choice of loss function, $\mathcal{L}$). Because prior work has found that this choice can lead to less exploration of the underlying parameter space, we use random growth, scoring all possible weights to be grown via a random number.

Table 4.4: Default FCN parameters for RigL experiments.

Parameters	
Data set	CIFAR-10
Architecture	3072:1024:1024:10
Activation function	ReLU
Batch size	10
Learning rate	0.1
Optimizer	SGD
Epochs	50
p	0.95
ΔT	2500 training steps
T_{end}	50 epochs
α	0.95
Decay function	Constant
Sparsity distribution	Uniform
Growth criteria	Gradient
Dropping criteria	Magnitude

4.4 Results

We began by replicating the results of Pellegrini and Biroli (2022), training an FCN (with a few modifications – see Sec. 4.3.3 for details) on ImageNet32. In line with their results, we find that the masks found from one-shot pruning after training are localised (Fig. 4.1A, C), illustrating that SGD brings the FCN towards a local minima that contains local RFs, albeit in an *embedded* manner, where pruning is required to extract this feature. Additionally, we find that performing multiple rounds of IMP leads to stronger localisation than is found using one-shot pruning (Fig. 4.1B, D), illustrating that the iterative nature of IMP affects the solution to which the FCN converges to.

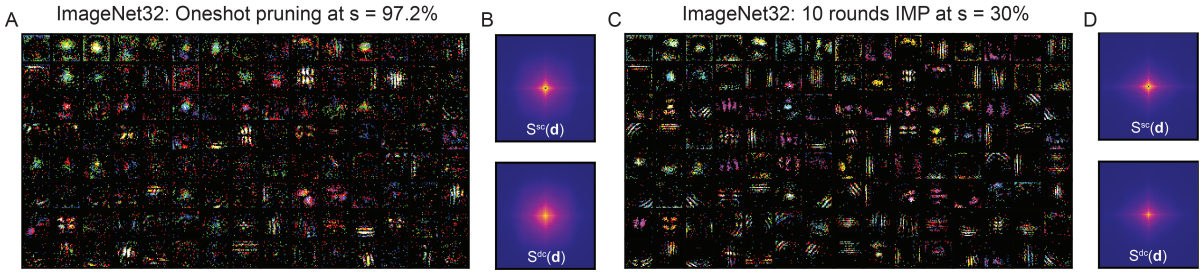

Figure 4.1: **One-shot pruning and IMP discover local RFs embedded in FCNs trained on ImageNet32.** (A)-(B) Local RFs present in the masks found after pruning (oneshot and IMP, respectively) FCNs trained on ImageNet32 to 97.2% sparsity (equivalent to 10 rounds of IMP, with $s = 0.3$). The masks shown are those with the greatest number of weights remaining [127]. (C)-(D) Local distance matrices, $S(\mathbf{d})$, for same and different channel, of masks in (A)-(B).

Motivated by these results, and in light of the work of Ingrosso and Goldt (2022) [128], we ask the following four questions, which we answer in turn below:

- **Q1**: Is the emergence of embedded local RFs in FCNs during training accompanied by an increase in the non-Gaussianity of the pre-activation statistics?
- **Q2**: How does changing the dataset and the hyper-parameters used for training affect the localisation of one-shot pruning masks and statistics of the pre-activations?

- **Q3**: What effect does IMP have on the statistics of the internal representations?
- **Q4**: How does IMP compare to other methods for sparsifying FCNs such as L1 and RigL in terms of localisation and preactivation statistics.

4.4.1 Emergence of embedded local RFs in FCNs trained on ImageNet32 is accompanied by increase in non-Gaussian preactivation statistics.

To understand the nature of the preactivation statistics, as local RFs emerge, we made copies of the FCN parameters at intervals during training. From these copies, we computed the preactivations (Eq. 4.2), using the ImageNet32 test set ($50,000$ images, 50 images per class) as inputs. The kurtosis of the preactivations (Eq. 4.4), per class, was then evaluated, with the average being taken across all classes for each unit. The saved parameters were also used to construct oneshot pruning masks, from which the locality of the RFs (σ_x and σ_y) were estimated. This allows us to estimate the extent to which any embedded local structures has emerged.

We find that, early in training ($5,000$ training iterations), the oneshot pruning masks are highly diffuse, showing little to no localisation (Fig. 4.2A). By the mid-part of training ($15,000$ training iterations), the oneshot pruning masks of some units have begun to show localisation (Fig. 4.2B). In-line with this, we find that the mean σ_x^{dc} steadily decreases with training time. Interestingly, as the oneshot masks are becoming increasingly localised, we find that the mean kurtosis of the preactivations increases, implying that the preactivatons are becoming increasingly non-Gaussian with training time. We therefore able to answer our first question (**Q1**) in the affirmative.

A1: Non-Gaussian preactivation statistics *do* emerge as localised structure, embedded within FCNs, emerges.

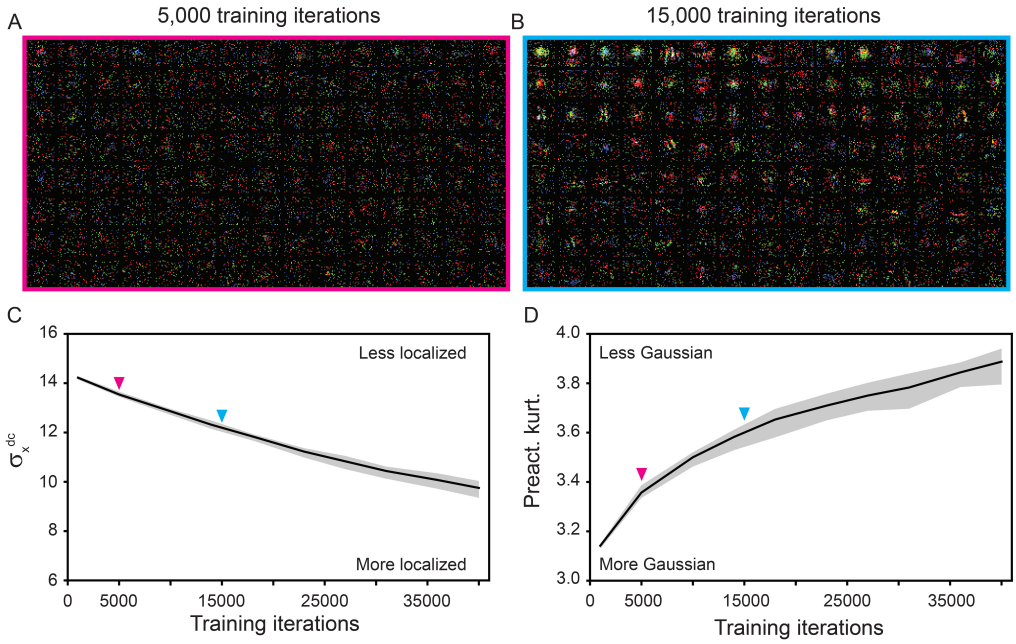

Figure 4.2: **Increased localisation of RFs occurs during an increase in non–Gaussian preactivation statistics.** (A)-(B) Masks found after pruning to 97.2% after 5,000 and 15,000 training steps, respectively. (C) Median σ_x^{dc}, as a function of training time. (D) Median preactivation kurtosis, computed per class, as a function of training time. In (A)-(B), black line denotes mean and gray shaded area denotes minimum and maximum across three independent seeds. The corresponding preactivation kurtosis and σ_x^{dc} values, at the training times shown in (A)-(B), is indicated with pink and blue arrows, respectively.

4.4.2 Training data and hyper-parameters affect localisation of RFs and statistics of preactivations.

Given that the emergence of embedded local RFs is a dynamic phenomenon, requiring training to increase localisation and non-Gaussian statistics of preactivations, we wondered how the data used to train the FCN, as well as the hyper-parameters associated with training, affected these results.

To examine this, we first trained our FCN on CIFAR-10 (see Appendix 4.3.3 for more details). When we used the same hyper-parameters as used for ImageNet32 (batch size, $b = 40$, and learning rate, $\eta = 0.1$), localisation was less strong (Fig. 4.3A-B, pink) than

Table 4.5: Performance of FCNs trained with different choices of training data, hyper-parameter, and sparsification strategies. Minimum and maximum of three independent seeds are presented.

Data set	Sparsification	Hyper-parameters	Best val. accuracy
ImageNet32	SGD training	$b = 40, \eta = 0.1$	[0.066, 0.067]
ImageNet32	L1 weight reg.	$b = 40, \eta = 0.1$	[0.053, 0.055]
ImageNet32	10 rounds IMP	$b = 40, \eta = 0.1$	[0.095, 0.095]
CIFAR-10	SGD training	$b = 10, \eta = 0.1$	[0.533, 0.537]
CIFAR-10	RigL	$b = 10, \eta = 0.1$	[0.491, 0.500]
CIFAR-10	SGD training	$b = 40, \eta = 0.1$	[0.561, 0.567]
CIFAR-10	SGD training	$b = 400, \eta = 0.1$	[0.570, 0.572]
CIFAR-10 GP	SGD training	$b = 10, \eta = 0.1$	[0.781, 0.785]
CIFAR-10 GP	SGD training	$b = 10, \eta = 0.1$	[0.722, 0.727]

it had been. By decreasing the batch size (and thus, increasing the learning rate to batch size ratio, η/b, which has been proposed to lead to convergence to flat minima with good generalization properties [302], as well as to lead to learning of sparse features [303]) we were able to increase localisation (Fig. 4.3A-B, orange). Decreasing η/b, by increasing b, led to nearly a complete loss in localisation (Fig. 4.3A-B, red). This difference in localisation is not explainable by the larger batch sizes leading to worse accuracy (see Table 4.5).

Concomitant with these changes in localisation of oneshot pruning masks, were changes to the preactivation kurtosis, measured on the CIFAR-10 test set ($10,000$ images, 100 per class). In particular, we found that the more localised the RFs were, the larger the preactivation kurtosis was (Fig. 4.3C). Note that, because we trained our FCNs with batch normalisation [277], and because the exact value the kurtosis takes can be sensitive

to scaling of variance, we considered the normalised preactivation kurtosis, dividing by the preactivation kurtosis of the FCN after 1,000 training steps, when comparing across experiment conditions.

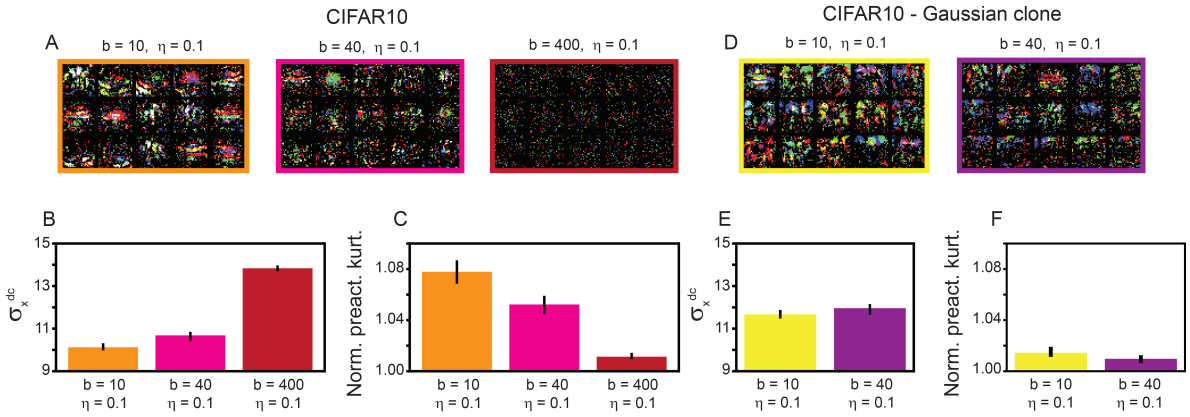

Figure 4.3: **Training FCNs on CIFAR-10 leads to less localisation of embedded RFs, with batch size and data statistics playing an important role.** (A) Masks found pruning FCNs trained on CIFAR-10 at 97.2% sparsity, for three choices of batch size (b) and learning rate (η). (B) Mean preactivation kurtosis, normalized to the preactivation kurtosis of the units after 1000 training iterations, for all three choices of b and η. (C) Mean σ_x^{dc}, for all three choices of b and η. (D)-(F) same as (A)-(C), but for FCNs trained on Gaussian clone of CIFAR-10. In (B)-(C) and (E)-(F), bar plot shows mean value, and error bars show minimum and maximum, across three separate seeds.

Sharp luminosity changes, which feature contours and edges in images, have been recognised as hallmarks of natural images [304, 295, 125, 305]. From a statistical perspective, these sudden changes translate into large higher-order cumulants. A generative model which samples images from a Gaussian distribution, where all cumulants after the second (i.e., the covariance) are zero, thus cannot reproduce this feature of realistic images. While we have found that local RFs emerge with *increasing* non-Gaussian statistics of local representations, we also verified the importance of having a non-Gaussian contribution to the training data itself.

To test this, we trained FCNs on a "Gaussian clone" of CIFAR-10, for which we re-sampled the test set from a mixture of 3×10 Gaussians, one for each colour channel

of each class of inputs [306]. The mean and covariance of each Gaussian are fitted to the images in that class. Training on this clone allows us to test the impact that higher-order cumulants of the CIFAR-10 images have on RF formation. We find that masks found when oneshot pruning after training on the Gaussian clones are less local for $b = 10$ and $b = 40$ (compare Fig. 4.3B and E). Additionally, the kurtosis hardly changes with training (Fig. 4.3F), highlighting that a necessary ingredient of the localisation is in the higher-order cumulants of the training images, even though, by themselves, they are not enough to trigger localisation. Note that sufficiency in this case has not been shown, as training on noise with non-Gaussian statistics, but no connected sharp edges, is not expected to drive local RF formation.

A2: The data used to train FCNs plays a large role on the localisation that emerges. Increasing the decreasing the batch size increases localisation and non-Gaussian statistics of the internal representations, but only when the data itself has non-Gaussian features.

4.4.3 IMP increases non-Gaussian preactivation statistics.

We next turn our attention to understanding how IMP is able to improve upon the masks found by oneshot pruning and increase localisation (Fig. 4.1, Fig. 4.4A). Importantly, this ability of IMP does not come from an increase in training time, because of the rewinding step in the IMP algorithm. Thus, while many rounds of IMP have to be performed (in our case, 10 rounds), the network parameters at the rewind state (θ_{rewind}) themselves are always trained the same amount of iterations as they are in the case of vanilla SGD. The only change is which weights are included in the training.

Because the iterative pruning of IMP affects the FCN parameters, it also necessarily affects the preactivations (Eq. 4.2). We hypothesised that IMP may be able to increase the localisation of its masks by removing parameters that made the internal representa-

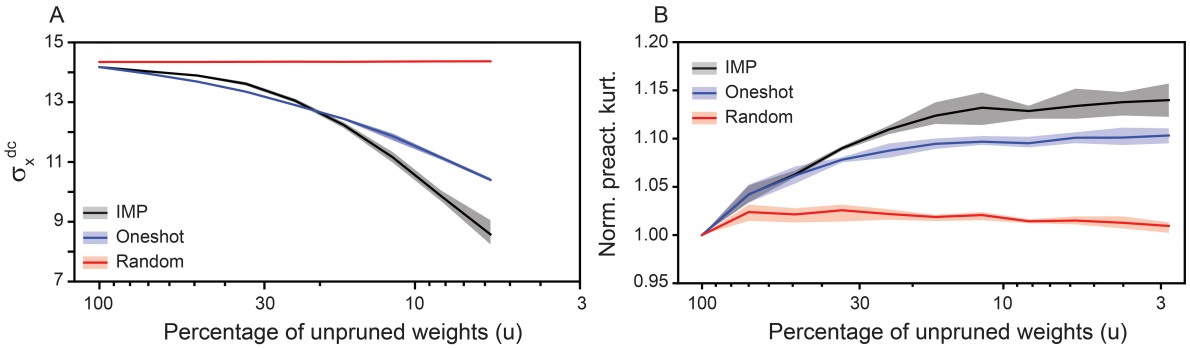

Figure 4.4: **IMP increases preactivation kurtosis and localisation on FCNs trained on ImageNet32, to a greater extent than oneshot pruning.** (A) Mean σ_x^{dc}, as a function of sparsity induced by IMP, oneshot, and random pruning. (B) Mean normalized kurtosis of preactivations, per class, as a function of sparsity induced by IMP, oneshot, and random pruning. In (A)-(B), solid line is mean and shaded area is minimum and maximum of three independent seeds.

tions more Gaussian. We indeed find that the kurtosis of the preactivations at the rewind state, normalized to its original value pre-pruning, increases with rounds of IMP (Fig. 4.4B, black curve). This increase is significantly greater than when random pruning is used, implying that it is not merely an artifact of sparsity (Fig. 4.4B, red curve). Additionally, the increase in non-Gaussian statistics of the preactivations is greater when using IMP than when using the masks found using oneshot pruning, after a single round of training (Fig. 4.4B, blue curve).

A3: IMP selectively enhances the non-Gaussian statistics of the preactivations at the rewind point, thus amplifying the degree to which the inputs to the FCN are non-Gaussian.

4.4.4 Local RFs and non-Gaussian preactivation statistics when training with other sparse supervised learning methods.

Given the large amount of development in sparse machine learning, we wondered whether other sparsification strategies, applied to our FCNs, might find localised RFs,

with a concomitant increase in non-Gaussian statistics of internal representations.

We first tried adding L1 regularisation of the weights in the first hidden layer to our loss function. Unlike IMP, this sparsifier is applied at every step of training (see Sec. **??** for more details). However, like IMP, it induces overall sparsity in the weights (although, in terms of the L_1, and not L_0, norm). We find strongly localised RFs when using L1 regularisation (Fig. 4.5A), with an increase in preactivation kurtosis across training (Fig. 4.5B). Interestingly, while the localisation is increased, with respect to vanilla SGD, the preactivation kurtosis does not see as large an increase (Fig. 4.6). This may be due to the fact that the added L1 regularisation allows for an additional mechanism by which localisation can occur, making each increase in non-Gaussian statistics more effective at driving local RF formation. Note that, aside from finding the correct scale of L1 penalty, this approach is much less costly than IMP, in terms of wall clock time and compute. We therefore expect that L1 regularisation may be optimal in some trade-off between cost and localisation.

We then tried using RigL, a dynamic sparse training (DST) method [252] (see Sec. 4.3.7 for more details). Due to computational constraints, we were not able to run our experiments on ImageNet32. We focused instead of CIFAR-10, with batch size $b = 10$ and learning rate $\eta = 0.1$. Such a setting was found to lead to some, albeit less strong, localisation (Fig. 4.3A-B, orange). Despite training with the same underlying FCN backbone, the masks found with RigL did not show any localisation (Fig. 4.5C). Consistent with this, we found no increase in normalized preactivation kurtosis during the course of training (Fig. 4.5D). We hypothesise that this may be due to the fact that CIFAR-10 does not explicitly require localisation for the FCN to perform well (indeed, we find similar accuracy between FCNs trained using vanilla SGD, before sparsification, and RigL – Table 4.5), and RigL is not biased towards finding a localised solution. While there exist many of DST approaches that future work should explore, it is possible that the

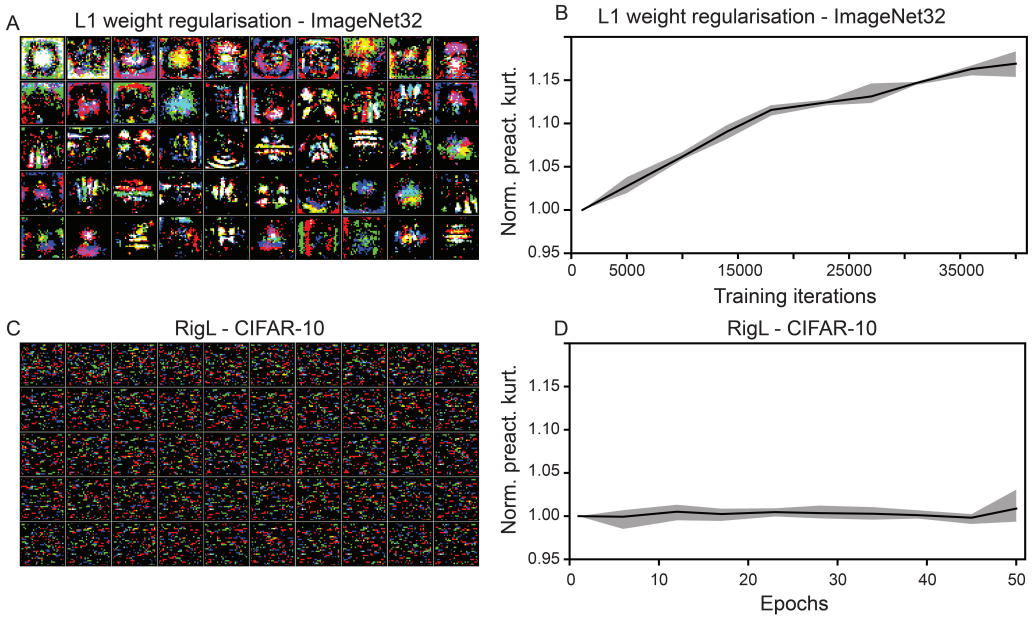

Figure 4.5: **Masks and statistics of internal representations using L1 regularisation of weights and RigL.** (A) Example masks found oneshot pruning after training on ImageNet32 using SGD with L1 regularisation. (B) Mean normalised kurtosis of preactivations, per class, as a function of training iterations, when using SGD with L1 regularisation. (C)-(D), same as (A)-(B), but for RigL trained on CIFAR-10. In (B) and (D), solid line is mean and shaded area is minimum and maximum of three independent seeds.

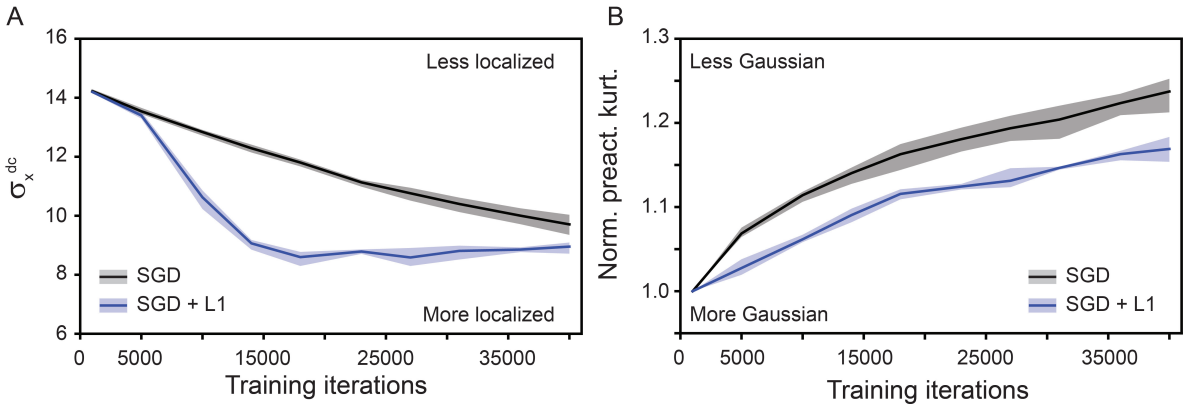

Figure 4.6: **Comparing SGD with and without L1 weight regularisation.** (A) Mean σ_x^{dc}, as a function of sparsity induced by oneshot pruning, for training via SGD with and without L1 weight regularisation. (B) Mean normalized kurtosis of preactivations per class, as a function of sparsity induced by oneshot pruning, for training via SGD with and without L1 weight regularisation. (A)-(B), solid line is mean and shaded area is minimum and maximum of three independent seeds.

discovery of local RFs requires (or is much enhanced by) starting from the full dense architecture (as one-shot pruning, IMP, and L1 weight regularisation all are).

A4: L1 regularisation is able to find increased localised RFs and non-Gaussian preactivation statistics, like IMP, but RigL is not.

4.5 Discussion

In this work, we took the first step to addressing the question of how the statistics of the preactivations evolve as FCNs, trained on standard computer vision benchmark data sets, develop embedded local RFs. As predicted from the theory developed by Ingross and Goldt (2022), we find increasingly non-Gaussian statistics of the preactivations accompanying increases in embedded locality (Fig. 4.2). This signal is further strengthened by the use of IMP (Fig. 4.4), whereby recursively rewinding and pruning leads to the removal of weights that support Gaussian statistics of the preactivations. Importantly, this improvement comes not just from sparsity or increased training, but from information that is gained by identifying how the FCN weights are coupled together.

We additionally expanded on the work done by Pellegrini and Biroli (2022), studying the dependence on the localisation using data beyond ImageNet32 (Fig. 4.3) and sparsification beyond IMP, SNIP [307], and SynFlow [308] (Fig. 4.5). We find that L1 weight regularisation leads to increased localisation, via an increase in preactivation kurtosis, as well as via mechanisms related to the continuous sparsification (Fig. 4.6 of Sec. 4.3.5). RigL [252], while a powerful framework to train sparse DNNs, is not able to discover local RFs, and sees no increase in preactivation kurtosis.

Limitations. This work is the first to study the statistics of internal representations during local RF formation embedded within general FCNs trained on computer vision

tasks, as well as the first to study how IMP affects these representations. While the results are consistent with the theory developed by Ingrosso and Goldt (2022), which proved that the local RF formation was *driven* by the non-Gaussian distribution of preactivations, in this extended setting, no such proof of causality exists. We view this as an important direction for the future. We additionally recognize that our definition of locality, namely the σ_x and σ_y, fitted on $S(\mathbf{d})$, is not the only possible measure. Finally, while an explicit goal of ours was to connect the works of Pellegrini and Biroli (2022) and Ingrosso and Goldt (2022), both of which studied FCNs on computer vision tasks, this necessarily limited our scope. We believe future exploration of how the structure that emerges in other network architectures, on other tasks, depends on the internal representations, is an exciting direction motivated by this work (we discuss this more below). Finally, our failure to find local RFs using RigL could be due to the fact that there are a number of hyper-parameters that must be tuned, of which we explored only part of the combinatorial space. Additionally, even if RigL is not able to find local RFs, there a number of other DST methods that exist, whose ability to find local RFs should be tested (we discuss this more below).

Beyond FCNs. While the work has been focused exclusively on FCNs trained on supervised learning computer vision tasks, there exist other settings where localised RFs emerge. These include unsuperivsed learning [125], multi-layer perceptrons (MLP) mixer [**?**], and vision transformers [266]. A natural question to ask is whether the framework developed in Ingrosso and Goldt (2022) [128], and extended here, holds in such cases. That is, are increasing deviations from Gaussian distributions in the statistics of the internal representations, a necessary ingredient in localisation? Evidence of the answer being an affirmative, especially in contexts beyond computer vision, would greatly improve the field's understanding of how to learn useful structure in general tasks.

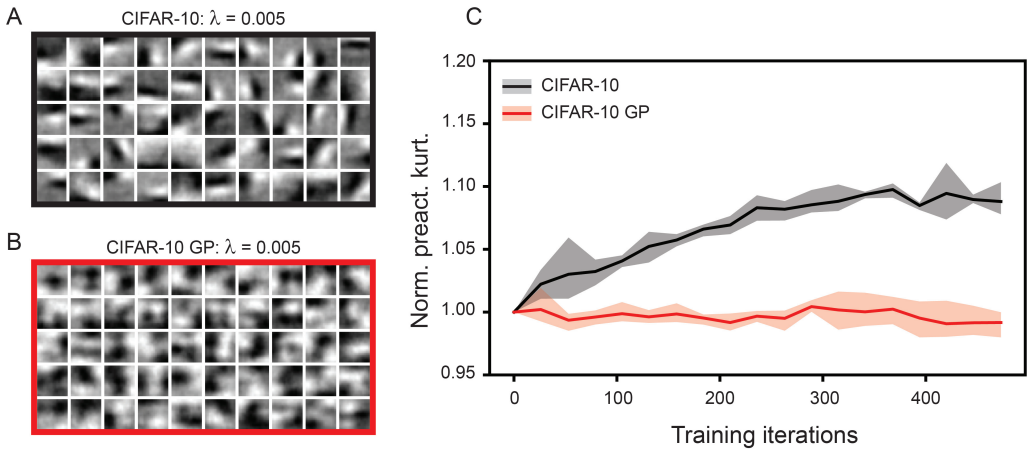

Figure 4.7: **Localisation of RFs and increase in non-Gaussian statistics of preactivations both rely on statistics of data in unsupervised learning task.** (A)-(B) RFs learned when applying unsupervised learning on CIFAR-10 and GP CIFAR-10, respectively, with L1 penalty $\lambda = 0.005$. (C) Median normalized preactivation kurtosis, as a function of training time. Solid line denotes mean, and shaded area denotes minimum and maximum, across three independent seeds.

While studying this problem is beyond the scope of this paper, we demonstrate that some of the key observations of this paper hold in the case of unsupervised learning [125] (Fig. 4.7). In particular, we find that, while L1 regularisation (on the activations, not the weights) enables the discovery of Gabor-esque filters from CIFAR-10 data, it is not sufficient when the data is the Gaussian clone of CIFAR-10 (Fig. 4.7A, B). This emphasizes the point made earlier (Fig. 4.3) that the data must be itself sufficiently non-Gaussian for local RFs to emerge. Additionally, we find that the non-Gaussanity of the preactivations increases only in the case where local RFs are found (Fig. 4.3C).

Lessons for sparse machine learning. We end by noting that our work suggests several fruitful directions for the sparse machine learning community. First, our failure to find local RFs when using RigL motivates the question, when (if at all) do IMP and dynamic sparse training methods learn the same structure? Secondly, our result showing that the degree of locality discovered by IMP depended on the data set (Fig. 4.3),

as well as Pellegrini and Biroli's (2022) result that locality was reduced when training on a smaller percentage of ImageNet32 [127], make us question whether attempts to discover winning tickets with less data [309, 275] may lead to less sparse subnetworks with less structure. Finally, identifying whether structure in the IMP masks found on large, complex data sets is what enables for transfer on downstream, smaller data sets [269, 270, 272] could provide a mechanistic explanation for the "universality" of some winning tickets. Addressing these questions will lead to additional research avenues and increase what the community has learned from studying sparse machine learning [310].

Chapter 5

Conclusion

"This looks like the way to
Zanesville, Ohio"

Aunt Denise (1972)

5.1 Summary

5.1.1 Optically Accessing the Transverse Plane of the Hippocampus

The circuit level mechanisms that enable the hippocampus (HPC) to support critical cognitive functions, such as episodic [90, 91] and spatial [21] memory has remained unclear, despite decades of study. A major contributor to this is the complex anatomy of the HPC, which distributes the principal subregions (CA1, CA3, and DG) across a curved volume of tissue that extends both dorsoventrally and medialaterally (see green region in Fig. 2.1A). This has challenged experimental paradigms, making it difficult (or impossible) to record neural from multiple subregions simultaneously, to identify structural

components of the circuit (e.g., dendritic spines on the apical dendrites), and to unequivocally identify distinct cell types (e.g., mossy and granular DG neurons, pyramidal and inter-neurons).

By developing a novel approach, via the implantation of microperiscopes that respect the geometry of the mouse HPC, we were able to circumvent these limitations. In particular, we were able to resolve and track spines on the apical dendrites of CA1 neurons (Figs. 2.4, 2.6, and 2.7) and record neural activity, simultaneously, across the HPC circuit 2.8, 2.9, and 2.13).

Leveraging our microperiscopes, we were able to show that there are populations of stable and unstable spines on apical dendrites, in similar proportion to what has been observed on basal dendrites [174], and that there is inhomogeneity in spatial encoding along the CA1-DG axis, in-line with electrophysiological recordings in distinct animals [29, 30, 31]. These results demonstrate how the microperiscope is a powerful addition to the existing neurophysiological toolkit used to study HPC function, and we proposed several ways in which it could be further used for chronic structural and functional measurements across the entire transverse HPC circuit.

5.1.2 Identifying Variability in Grid Spacing and Orientation in Individual Grid Modules

The discovery that grid cells (GCs) along the dorsoventral axis in medial entorhinal cortex (mEC) are organized into modules, with each module having GCs of similar grid spacing and orientation [6, 55], has led to the assumption that activity from GCs present in multiple modules must be integrated across in order for local spatial information to be decoded [73, 56, 57, 58, 59, 60]. However, the extent to which GCs in an individual module contain the same grid properties, has not been quantified.

By analyzing data from recent state-of-the-art mEC recordings [72], we found that the distribution of grid spacing and orientation covered a range of values (Fig. 3.1). evidence for variability in grid spacing and orientation that was more robust than what would be expected by from noise (Figs. 3.2). This property was consistent across choices made in the analysis (Fig. 3.4) and when considering each grid field independently 3.6).

While the observed variability was small (on the order of centimeters and degrees), modeling analysis showed a large increase in the ability to encode local spatial information, as compared to populations of GCs with fixed spacing and orientation (Fig. 3.7). These results challenge the field to answer the question of why are there multiple grid modules?

5.1.3 Studying the Statistics of Neural Network Representations During the Formation of Local RFs

Neurons that belong to functional classes characterized by activity that is local (with respect to some set of stimuli – e.g., physical position, retinal position), such as PCs, GCs, and visually tuned neurons in visual cortex, contain several advantages, from a coding perspective. First, they respect certain key symmetries, such as translational invariance. Second, they are robust to noise, and enable error correction mechanisms. And third, they enable hierarchical association. While local receptive fields (RFs) appear to be hard-wired in the early stages of the visual system, PCs and GCs develop over the course of development [78, 79], and the emergence of individual PCs depend on experience, as well as underlying physiological properties [114, 163]. Many models exist that lead to the learning of PCs and GCs [131, 132, 133, 66, 73, 134, 135, **?**, 67, 136, 137, 138, 69, 139, 140, 141, 61, 142, 143, 63, 81, 82, 64, 107, 83, 65], but what common mechanisms exist between them all, remains unclear.

Recent work has suggested that a generic driver of local RF formation could be non-Gaussian statistics present in data [128]. Examples of this include sharp luminosity changes, which feature contours and edges, in the context of images. While this work was performed when studying artificial neural networks, trained on synthetic data, given that most real-world data is distributed in ways that contain non-Gaussian features, such a mechanism could apply more generally. The first step in validating that this could be the case, was to address why, in the case of training FCNs on natural images, instead of synthetic data, no local RFs emerge.

One explanation is that local RFs emerge only in an embedded form [127]. That is, pruning must be performed in order for this structure to be resolved. We show that during the formation of the embedded local RFs, the "non-Gaussianity" of the data indeed plays an important role (Fig. 4.2), just as it did in the case of the synthetic data [128]. Further, we find that the data set used to train, as well as the training's associated hyper-parameters, play an important role in shaping the amount of non-Gaussanity that can be extracted during training (Fig. 4.3). The use of iterative magnitude pruning (IMP) [250, 251] amplifies the non-Gaussian structure of the data (Fig. 4.4), as does L1 regularisation of the FCN weights (Fig. 4.5). Collectively this suggests that sparsity can play a powerful role in the statistics of the network internal representations. We find evidence that this framework holds even in the context of unsupervised learning [125] (Fig. 4.7).

5.2 Future Directions

5.2.1 Circuit Level Computations in HPC

While considerable experimental and theoretical work has established distinct functional roles for each of the major subregions of the HPC, these roles have not always found support in more recent experiments, most particularly those in CA3 and DG [311, 198]. To better establish what computations each of these subregions are performing, for a given task, a necessary step is to be able to simultaneously record neural activity across the HPC circuit.

Our developed microperiscope (Chapter 2) enables such an experimental approach. Combining opto- and chemo-genetical with the microperiscope will further enable high levels of control of the HPC. A particularly exciting avenue of future work would be to optogenetically manipulate DG spiking activity, which has been found to control HPC function [312], and monitor the effects on neural activity in CA3 and CA1.

5.2.2 The Role of Multiple Grid Modules

By showing that individual grid modules, via the heterogeneity present in their grid properties, can encode considerable information about local space (Chapter 3), the question of why there exist multiple grid modules must be answered. One possibility is that GCs are utilized more generally than for just path integration [105, 111]. In such a case, a desirable property of a generalized coding scheme would be the ability to form hierarchical representations, which provide computational benefits [313, 314].

Such an organization has recently been found in the activity of PCs of CA1 in large environments [27]. Given that grid fields increase in their size across grid modules, this is a natural feature of a grid code with multiple modules. Additionally, given the inter-

connectivity between HPC and mEC, multiple modules may be useful for driving PCs with multiple place fields in large environments [22, 23, 24, 25, 26]. Whether heterogeneity in GC properties, within single grid modules, can enhance this, is an interesting question to address in future work.

5.2.3 Continuous Attractor Networks that Can Support Heterogeneity in Grid Properties

A major goal of work on the grid cell circuit has been the identification of the circuit level mechanism that allows for stable, multi-field representations. To this end, the success of continuous attractor neural network (CANN) models to show the capacity for path integration [67, 315], as well as extensive experimental support [70, 71, 236, 72] has led this model as the canonical model.

However, such CANN models require specific wiring, and the reported variability in grid field firing [87, 231] has proven challenging to adapt to CANNs (although see [**?**]). How our additional discovery of heterogeneity in the distribution of grid properties in individual grid modules (Chapter 3) can be supported in CANNs, without highly restrictive wiring, is unclear. A straightforward next step would be to examine whether recurrent neural network (RNN) models that have developed GC properties [82, 81, 64, 315] also see variability in grid properties. Examination for variability in another recently proposed mechanism for the establishment of GCs [65] should also be performed.

We view this general research program as synergistic with other work that has sought to understand why and how non-GC spatially modulated cells emerge [316].

5.2.4 A Causal Link between Non-Gaussian Preactivations and Local RFs

While our results on studying the statistics of the internal representations of FCNs during the formation of local RFs shows correlation between locality and the presence/degree to which non-Gaussian statistics emerge (Chapter 4), we have yet to establish causality. In addition to extending the theory developed in the case of synthetic images where local RFs emerge without the need of pruning [128], we can probe this causality by performing experiments where the weights that lead the most to non-Gaussian internal representations are removed at each iteration. Should this lead to strong local RFs, this would be evidence for a causal link. Additionally, control experiments, where the weights that contribute the most to Gaussian internal representations are removed, should be performed.

Additionally, identifying what elements of the training dynamics encourage or suppress local RF formation would be an interesting avenue of future work. This could be achieved by utilizing a dynamical systems theoretic framework, such as Koopman operator theory [317, 318, 319, 320, 321, 322, 323], whose spectral objects and linear representation have enabled the acceleration, optimization, and analysis of deep neural networks [324, 325, 326, 327, 328, 284, 329, 330, 331, 332].

5.2.5 Non-Gaussian Inputs, Excitatory and Inhibitory Plasticity, and the Formation of place fields

Our results supporting the role of non-Gaussian statistics for local RF formation in FCNs trained on computer vision motivate the question, "how far can this be extended?" Of particular interest is whether biologically plausible neural network models for the formation of PCs exhibit a similar increase in the statistics of their internal representations.

We note that some of these recent models have placed importance on inhibitory, as well as excitatory, plasticity [63, 107]. Weber and Spreckler (2018) specifically showed that the relative smoothness between the excitatory and inhibitory inputs, in a spiking neural network with spike-time dependent plasticity, determined what functional responses the downstream neuron would develop. These included PCs, GCs, and head direction cells. We speculate that inhibitory inputs, modulated via plasticity mechanisms, may determine the overall statistics of inputs into real neurons [333]. If the inputs are sufficiently non-Gaussian, then representations such as place fields may emerge. We note that, if the resulting distribution is broader than a Gaussian (i.e., its kurtosis is less than 3, as is the case of a uniform distribution), the resulting phenomenon of over-dispersion [334].

5.3 Conclusion

The brains remarkable ability to store and manipulate information has fascinated scientists and scholars for millennia. A brain region integral to this capability is the hippocampal formation, whose composite structure, diverse morphology, and loop network architecture, endow it with a high computational capacity to support such critical cognitive processes. Despite decades of work revealing aspects of the neurophysiology of the hippocampal formation, the circuit level mechanisms used are poorly understood. In this thesis, we present new tools and perspectives for understanding the important open question, at the heart of neuroscience.

Bibliography

[1] J. J. Gibson, *The ecological approach to visual perception: classic edition.* Psychology press, 2014.

[2] H. C. Berg, *Bacterial behaviour*, *Nature* **254** (1975), no. 5499 389–392.

[3] D. Genzel, Y. Yovel, and M. M. Yartsev, *Neuroethology of bat navigation*, *Current Biology* **28** (2018), no. 17 R997–R1004.

[4] J. O'Keefe and L. Nadel, *The Hippocampus as a Cognitive Map.* Oxford University Press, 1978.

[5] *Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis*, *Journal of Neuroscience* **10** (1990), no. 2 420–435.

[6] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, *Microstructure of a spatial map in the entorhinal cortex*, *Nature* **436** (2005), no. 7052 801–806.

[7] T. Solstad, C. N. Boccara, E. Kropff, M.-B. Moser, and E. I. Moser, *Representation of geometric borders in the entorhinal cortex*, *Science* **322** (2008), no. 5909 1865–1868.

[8] J. Krupic, N. Burgess, and J. O'Keefe, *Neural representations of location composed of spatially periodic bands*, *Science* **337** (2012), no. 6096 853–857.

[9] Ø. A. Høydal, E. R. Skytøen, S. O. Andersson, M.-B. Moser, and E. I. Moser, *Object-vector coding in the medial entorhinal cortex*, *Nature* **568** (2019), no. 7752 400–404.

[10] A. S. Alexander, L. C. Carstensen, J. R. Hinman, F. Raudies, G. W. Chapman, and M. E. Hasselmo, *Egocentric boundary vector tuning of the retrosplenial cortex*, *Science advances* **6** (2020), no. 8 eaaz2322.

[11] L. M. Frank, E. N. Brown, and M. Wilson, *Trajectory encoding in the hippocampus and entorhinal cortex*, *Neuron* **27** (2000), no. 1 169–178.

[12] F. Sargolini, M. Fyhn, T. Hafting, B. L. McNaughton, M. P. Witter, M.-B. Moser, and E. I. Moser, *Conjunctive representation of position, direction, and velocity in entorhinal cortex*, *Science* **312** (2006), no. 5774 758–762.

[13] A. S. Alexander and D. A. Nitz, *Retrosplenial cortex maps the conjunction of internal and external spaces*, *Nature neuroscience* **18** (2015), no. 8 1143–1151.

[14] A. B. Saleem, A. Ayaz, K. J. Jeffery, K. D. Harris, and M. Carandini, *Integration of visual motion and locomotion in mouse visual cortex*, *Nature neuroscience* **16** (2013), no. 12 1864–1869.

[15] E. M. Diamanti, C. B. Reddy, S. Schröder, T. Muzzu, K. D. Harris, A. B. Saleem, and M. Carandini, *Spatial modulation of visual responses arises in cortex with active navigation*, *Elife* **10** (2021) e63705.

[16] X. Long and S.-J. Zhang, *A novel somatosensory spatial navigation system outside the hippocampal formation*, *Cell research* **31** (2021), no. 6 649–663.

[17] D. F. Sherry, A. L. Vaccarino, K. Buckenham, and R. S. Herz, *The hippocampal complex of food-storing birds*, *Brain, behavior and evolution* **34** (1989), no. 5 308–317.

[18] A. Barnea and F. Nottebohm, *Seasonal recruitment of hippocampal neurons in adult free-ranging black-capped chickadees.*, *Proceedings of the National Academy of Sciences* **91** (1994), no. 23 11217–11221.

[19] E. A. Maguire, D. G. Gadian, I. S. Johnsrude, C. D. Good, J. Ashburner, R. S. Frackowiak, and C. D. Frith, *Navigation-related structural change in the hippocampi of taxi drivers*, *Proceedings of the National Academy of Sciences* **97** (2000), no. 8 4398–4403.

[20] J. O'Keefe and J. Dostrovsky, *The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat.*, *Brain research* (1971).

[21] J. O'Keefe, *Place units in the hippocampus of the freely moving rat*, *Experimental neurology* **51** (1976), no. 1 78–109.

[22] A. A. Fenton, H.-Y. Kao, S. A. Neymotin, A. Olypher, Y. Vayntrub, W. W. Lytton, and N. Ludvig, *Unmasking the ca1 ensemble place code by exposures to small and large environments: more place cells and multiple, irregularly arranged, and expanded place fields in the larger space*, *Journal of Neuroscience* **28** (2008), no. 44 11250–11262.

[23] K. B. Kjelstrup, T. Solstad, V. H. Brun, T. Hafting, S. Leutgeb, M. P. Witter, E. I. Moser, and M.-B. Moser, *Finite scale of spatial representation in the hippocampus*, *Science* **321** (2008), no. 5885 140–143.

[24] E. Park, D. Dvorak, and A. A. Fenton, *Ensemble place codes in hippocampus: Ca1, ca3, and dentate gyrus place cells have multiple place fields in large environments*, *PloS one* **6** (2011), no. 7 e22349.

[25] P. D. Rich, H.-P. Liaw, and A. K. Lee, *Large environments reveal the statistical structure governing hippocampal representations*, *Science* **345** (2014), no. 6198 814–817.

[26] B. Harland, M. Contreras, M. Souder, and J.-M. Fellous, *Dorsal ca1 hippocampal place cells form a multi-scale representation of megaspace*, *Current Biology* **31** (2021), no. 10 2178–2190.

[27] H. Zhang, P. D. Rich, A. K. Lee, and T. O. Sharpee, *Hippocampal spatial representations exhibit a hyperbolic geometry that expands with experience*, *Nature Neuroscience* (2022) 1–9.

[28] W. Skaggs, B. McNaughton, and K. Gothard, *An information-theoretic approach to deciphering the hippocampal code*, *Advances in Neural Information Processing Systems* **5** (1993) 1030–1037.

[29] H. Lee, C. Wang, S. S. Deshmukh, and J. J. Knierim, *Neural population evidence of functional heterogeneity along the ca3 transverse axis: Pattern completion versus pattern separation*, *Neuron* **87** (9, 2015) 1093–1105.

[30] L. Lu, K. M. Igarashi, M. P. Witter, E. I. Moser, and M. B. Moser, *Topography of place maps along the ca3-to-ca2 axis of the hippocampus*, *Neuron* **87** (9, 2015) 1078–1092.

[31] E. A. Mankin, G. W. Diehl, F. T. Sparks, S. Leutgeb, and J. K. Leutgeb, *Hippocampal ca2 activity patterns change over time to a larger extent than between spatial contexts*, *Neuron* **85** (1, 2015) 190–201.

[32] R. U. Muller and J. L. Kubie, *The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells*, *Journal of Neuroscience* **7** (1987), no. 7 1951–1968.

[33] C. B. Alme, C. Miao, K. Jezek, A. Treves, E. I. Moser, and M.-B. Moser, *Place cells in the hippocampus: eleven maps for eleven rooms*, *Proceedings of the National Academy of Sciences* **111** (2014), no. 52 18428–18435.

[34] S. Leutgeb, J. K. Leutgeb, A. Treves, M. B. Moser, and E. I. Moser, *Distinct ensemble codes in hippocampal areas ca3 and ca1*, *Science* **305** (8, 2004) 1295–1298.

[35] J. K. Leutgeb, S. Leutgeb, A. Treves, R. Meyer, C. A. Barnes, B. L. McNaughton, M. B. Moser, and E. I. Moser, *Progressive transformation of hippocampal neuronal representations in "morphed" environments*, *Neuron* **48** (10, 2005) 345–358.

[36] J. K. Leutgeb, S. Leutgeb, M. B. Moser, and E. I. Moser, *Pattern separation in the dentate gyrus and ca3 of the hippocampus*, *Science* **315** (2, 2007) 961–966.

[37] R. U. Muller, J. L. Kubie, and J. B. Ranck, *Spatial firing patterns of hippocampal complex-spike cells in a fixed environment*, *Journal of Neuroscience* **7** (1987), no. 7 1935–1950.

[38] L. Thompson and P. Best, *Long-term stability of the place-field activity of single units recorded from the dorsal hippocampus of freely behaving rats*, *Brain research* **509** (1990), no. 2 299–308.

[39] C. Kentros, E. Hargreaves, R. D. Hawkins, E. R. Kandel, M. Shapiro, and R. V. Muller, *Abolition of long-term stability of new hippocampal place cell maps by nmda receptor blockade*, *Science* **280** (1998), no. 5372 2121–2126.

[40] C. G. Kentros, N. T. Agnihotri, S. Streater, R. D. Hawkins, and E. R. Kandel, *Increased attention to spatial context increases both place field stability and spatial memory*, *Neuron* **42** (2004), no. 2 283–295.

[41] E. A. Mankin, F. T. Sparks, B. Slayyeh, R. J. Sutherland, S. Leutgeb, and J. K. Leutgeb, *Neuronal code for extended time in the hippocampus*, *Proceedings of the National Academy of Sciences of the United States of America* **109** (11, 2012) 19462–19467.

[42] Y. Ziv, L. D. Burns, E. D. Cocker, E. O. Hamel, K. K. Ghosh, L. J. Kitch, A. E. Gamal, and M. J. Schnitzer, *Long-term dynamics of ca1 hippocampal place codes*, *Nature Neuroscience* **16** (3, 2013) 264–266.

[43] N. R. Kinsky, D. W. Sullivan, W. Mau, M. E. Hasselmo, and H. B. Eichenbaum, *Hippocampal place fields maintain a coherent and flexible map across long timescales*, *Current Biology* **28** (11, 2018) 3578–3588.e6.

[44] T. Hainmueller and M. Bartos, *Parallel emergence of stable and dynamic memory engrams in the hippocampus*, *Nature* **558** (6, 2018) 292–296.

[45] W. G. Gonzalez, H. Zhang, A. Harutyunyan, and C. Lois, *Persistence of neuronal representations through time and damage in the hippocampus*, *Science* **365** (2019), no. 6455 821–825.

[46] W. Mau, M. E. Hasselmo, and D. J. Cai, *The brain in motion: How ensemble fluidity drives memory-updating and flexibility*, *eLife* **9** (12, 2020) 1–24.

[47] T. D. Marks and M. J. Goard, *Stimulus-dependent representational drift in primary visual cortex*, *Nature communications* **12** (2021), no. 1 5169.

[48] D. Deitch, A. Rubin, and Y. Ziv, *Representational drift in the mouse visual cortex*, *Current biology* **31** (2021), no. 19 4327–4339.

[49] C. E. Schoonover, S. N. Ohashi, R. Axel, and A. J. Fink, *Representational drift in primary olfactory cortex*, *Nature* **594** (2021), no. 7864 541–546.

[50] L. N. Driscoll, N. L. Pettit, M. Minderer, S. N. Chettih, and C. D. Harvey, *Dynamic reorganization of neuronal activity patterns in parietal cortex*, *Cell* **170** (2017), no. 5 986–999.

[51] L. M. Franco and M. J. Goard, *A distributed circuit for associating environmental context with motor choice in retrosplenial cortex*, *Science advances* **7** (2021), no. 35 eabf9815.

[52] M. E. Rule, A. R. Loback, D. V. Raman, L. N. Driscoll, C. D. Harvey, and T. O'Leary, *Stable task information from an unstable neural population*, *Elife* **9** (2020) e51121.

[53] J. Xia, T. D. Marks, M. J. Goard, and R. Wessel, *Stable representation of a naturalistic movie emerges from episodic activity with gain variability*, *Nature communications* **12** (2021), no. 1 5170.

[54] M. E. Rule and T. O'Leary, *Self-healing codes: How stable neural populations can track continually reconfiguring neural representations*, *Proceedings of the National Academy of Sciences* **119** (2022), no. 7 e2106692119.

[55] H. Stensola, T. Stensola, T. Solstad, K. Frøland, M.-B. Moser, and E. I. Moser, *The entorhinal grid map is discretized*, *Nature* **492** (2012), no. 7427 72–78.

[56] I. R. Fiete, Y. Burak, and T. Brookings, *What grid cells convey about rat location*, *Journal of Neuroscience* **28** (2008), no. 27 6858–6871.

[57] S. Sreenivasan and I. Fiete, *Grid cells generate an analog error-correcting code for singularly precise neural computation*, *Nature neuroscience* **14** (2011), no. 10 1330–1337.

[58] A. Mathis, A. V. Herz, and M. Stemmler, *Optimal population codes for space: grid cells outperform place cells*, *Neural computation* **24** (2012), no. 9 2280–2317.

[59] X.-X. Wei, J. Prentice, and V. Balasubramanian, *A principle of economy predicts the functional architecture of grid cells*, *Elife* **4** (2015) e08362.

[60] M. Stemmler, A. Mathis, and A. V. Herz, *Connecting multiple spatial scales to decode the population activity of grid cells*, *Science Advances* **1** (2015), no. 11 e1500816.

[61] Y. Dordek, D. Soudry, R. Meir, and D. Derdikman, *Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis*, *Elife* **5** (2016) e10094.

[62] H. T. Blair, A. C. Welday, and K. Zhang, *Scale-invariant memory representations emerge from moire interference between grid fields that produce theta oscillations: a computational model*, *Journal of Neuroscience* **27** (2007), no. 12 3211–3229.

[63] S. N. Weber and H. Sprekeler, *Learning place cells, grid cells and invariances with excitatory and inhibitory plasticity*, *Elife* **7** (2018) e34560.

[64] B. Sorscher, G. Mel, S. Ganguli, and S. Ocko, *A unified theory for the origin of grid cells through the lens of pattern formation*, *Advances in neural information processing systems* **32** (2019).

[65] M. Khona, S. Chandra, and I. R. Fiete, *From smooth cortical gradients to discrete modules: spontaneous and topologically robust emergence of modularity in grid cells*, *bioRxiv* (2022) 2021–10.

[66] M. C. Fuhs and D. S. Touretzky, *A spin glass model of path integration in rat medial entorhinal cortex*, *Journal of Neuroscience* **26** (2006), no. 16 4266–4276.

[67] Y. Burak and I. R. Fiete, *Accurate path integration in continuous attractor network models of grid cells*, *PLOS Computational Biology* **5** (2, 2009) e1000291.

[68] A. Guanella and P. F. Verschure, *A model of grid cells based on a path integration mechanism*, in *Artificial Neural Networks–ICANN 2006: 16th International Conference, Athens, Greece, September 10-14, 2006. Proceedings, Part I 16*, pp. 740–749, Springer, 2006.

[69] J. J. Couey, A. Witoelar, S.-J. Zhang, K. Zheng, J. Ye, B. Dunn, R. Czajkowski, M.-B. Moser, E. I. Moser, Y. Roudi, *et. al.*, *Recurrent inhibitory circuitry as a mechanism for grid formation*, *Nature neuroscience* **16** (2013), no. 3 318–324.

[70] K. Yoon, M. A. Buice, C. Barry, R. Hayman, N. Burgess, and I. R. Fiete, *Specific evidence of low-dimensional continuous attractor dynamics in grid cells*, *Nature neuroscience* **16** (2013), no. 8 1077–1084.

[71] B. Dunn, M. Mørreaunet, and Y. Roudi, *Correlations and functional connections in a population of grid cells*, *PLoS computational biology* **11** (2015), no. 2 e1004052.

[72] R. J. Gardner, E. Hermansen, M. Pachitariu, Y. Burak, N. A. Baas, B. A. Dunn, M.-B. Moser, and E. I. Moser, *Toroidal topology of population activity in grid cells*, *Nature* **602** (2022), no. 7895 123–128.

[73] T. Solstad, E. I. Moser, and G. T. Einevoll, *From grid cells to place cells: a mathematical model*, *Hippocampus* **16** (2006), no. 12 1026–1031.

[74] E. T. Rolls, S. M. Stringer, and T. Elliot, *Entorhinal cortex grid cells can map to hippocampal place cells by competitive learning*, *Network: Computation in Neural Systems* **17** (2006), no. 4 447–465.

[75] J. Ormond and B. L. McNaughton, *Place field expansion after focal mec inactivations is consistent with loss of fourier components and path integrator gain reduction*, *Proceedings of the National Academy of Sciences* **112** (2015), no. 13 4116–4121.

[76] C. S. Mallory, K. Hardcastle, J. S. Bant, and L. M. Giocomo, *Grid scale drives the scale and long-term stability of place maps*, *Nature neuroscience* **21** (2018), no. 2 270–282.

[77] S.-J. Zhang, J. Ye, C. Miao, A. Tsao, I. Cerniauskas, D. Ledergerber, M.-B. Moser, and E. I. Moser, *Optogenetic dissection of entorhinal-hippocampal functional connectivity*, *Science* **340** (2013), no. 6128 1232627.

[78] R. F. Langston, J. A. Ainge, J. J. Couey, C. B. Canto, T. L. Bjerknes, M. P. Witter, E. I. Moser, and M.-B. Moser, *Development of the spatial representation system in the rat*, *Science* **328** (2010), no. 5985 1576–1580.

[79] T. J. Wills, F. Cacucci, N. Burgess, and J. O'Keefe, *Development of the hippocampal cognitive map in preweanling rats*, *science* **328** (2010), no. 5985 1573–1576.

[80] J. B. Hales, M. I. Schlesiger, J. K. Leutgeb, L. R. Squire, S. Leutgeb, and R. E. Clark, *Medial entorhinal cortex lesions only partially disrupt hippocampal place cells and hippocampus-dependent place memory*, *Cell reports* **9** (2014), no. 3 893–901.

[81] C. J. Cueva and X.-X. Wei, *Emergence of grid-like representations by training recurrent neural networks to perform spatial localization*, in *International Conference on Learning Representations*, 2018.

[82] A. Banino, C. Barry, B. Uria, C. Blundell, T. Lillicrap, P. Mirowski, A. Pritzel, M. J. Chadwick, T. Degris, J. Modayil, *et. al.*, *Vector-based navigation using grid-like representations in artificial agents*, *Nature* **557** (2018), no. 7705 429–433.

[83] B. Sorscher, G. C. Mel, S. A. Ocko, L. M. Giocomo, and S. Ganguli, *A unified theory for the computational and mechanistic origins of grid cells*, *Neuron* (2022).

[84] F. Savelli, D. Yoganarasimha, and J. J. Knierim, *Influence of boundary removal on the spatial representations of the medial entorhinal cortex*, *Hippocampus* **18** (2008), no. 12 1270–1282.

[85] C. Lever, S. Burton, A. Jeewajee, J. O'Keefe, and N. Burgess, *Boundary vector cells in the subiculum of the hippocampal formation*, *Journal of Neuroscience* **29** (2009), no. 31 9771–9777.

[86] T. L. Bjerknes, E. I. Moser, and M.-B. Moser, *Representation of geometric borders in the developing rat*, *Neuron* **82** (2014), no. 1 71–78.

[87] G. W. Diehl, O. J. Hon, S. Leutgeb, and J. K. Leutgeb, *Grid and nongrid cells in medial entorhinal cortex represent spatial location and environmental features with complementary coding schemes*, *Neuron* **94** (2017), no. 1 83–92.

[88] F. Stefanini, L. Kushnir, J. C. Jimenez, J. H. Jennings, N. I. Woods, G. D. Stuber, M. A. Kheirbek, R. Hen, and S. Fusi, *A distributed neural code in the dentate gyrus and in ca1*, *Neuron* **107** (8, 2020) 703–716.e4.

[89] L. Meshulam, J. L. Gauthier, C. D. Brody, D. W. Tank, and W. Bialek, *Collective behavior of place and non-place neurons in the hippocampal network*, *Neuron* **96** (2017), no. 5 1178–1191.

[90] L. R. Squire, *Memory and the hippocampuss: A synthesis from findings with rats, monkeys, and humans*, *Psychological Review* **99** (1992) 195–231.

[91] S. Tonegawa, X. Liu, S. Ramirez, and R. Redondo, *Memory engram cells have come of age*, *Neuron* **87** (9, 2015) 918–931.

[92] W. B. Scoville and B. Milner, *Loss of recent memory after bilateral hippocampal lesions*, *Journal of neurology, neurosurgery, and psychiatry* **20** (1957), no. 1 11.

[93] B. Milner, S. Corkin, and H.-L. Teuber, *Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of hm*, *Neuropsychologia* **6** (1968), no. 3 215–234.

[94] L. Stefanacci, E. A. Buffalo, H. Schmolck, and L. R. Squire, *Profound amnesia after damage to the medial temporal lobe: A neuroanatomical and neuropsychological profile of patient ep*, *Journal of Neuroscience* **20** (2000), no. 18 7024–7036.

[95] R. W. Semon, *Die Mneme als erhaltendes Prinzip im Wechsel des organischen Geschehens*, vol. 7101. Wilhelm Engelmann, 1911.

[96] X. Liu, S. Ramirez, P. T. Pang, C. B. Puryear, A. Govindarajan, K. Deisseroth, and S. Tonegawa, *Optogenetic stimulation of a hippocampal engram activates fear memory recall*, *Nature* **484** (2012), no. 7394 381–385.

[97] A. R. Garner, D. C. Rowland, S. Y. Hwang, K. Baumgaertel, B. L. Roth, C. Kentros, and M. Mayford, *Generation of a synthetic memory trace*, *Science* **335** (2012), no. 6075 1513–1516.

[98] S. Ramirez, X. Liu, P.-A. Lin, J. Suh, M. Pignatelli, R. L. Redondo, T. J. Ryan, and S. Tonegawa, *Creating a false memory in the hippocampus*, *Science* **341** (2013), no. 6144 387–391.

[99] T. Kitamura, S. K. Ogawa, D. S. Roy, T. Okuyama, M. D. Morrissey, L. M. Smith, R. L. Redondo, and S. Tonegawa, *Engrams and circuits crucial for systems consolidation of a memory*, *Science* **356** (2017), no. 6333 73–78.

[100] C. J. MacDonald, K. Q. Lepage, U. T. Eden, and H. Eichenbaum, *Hippocampal "time cells" bridge the gap in memory for discontiguous events*, *Neuron* **71** (2011), no. 4 737–749.

[101] C. J. MacDonald, S. Carrow, R. Place, and H. Eichenbaum, *Distinct hippocampal time cell sequences represent odor memories in immobilized rats*, *Journal of Neuroscience* **33** (2013), no. 36 14607–14616.

[102] M. N. Modi, A. K. Dhawale, and U. S. Bhalla, *Ca1 cell activity sequences emerge after reorganization of network correlation structure during associative learning*, *Elife* **3** (2014) e01982.

[103] B. J. Kraus, M. P. Brandon, R. J. Robinson II, M. A. Connerney, M. E. Hasselmo, and H. Eichenbaum, *During running in place, grid cells integrate elapsed time and distance run*, *Neuron* **88** (2015), no. 3 578–589.

[104] M. W. Howard and H. Eichenbaum, *Time and space in the hippocampus*, *Brain research* **1621** (2015) 345–354.

[105] D. Aronov, R. Nevers, and D. W. Tank, *Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit*, *Nature* **543** (2017), no. 7647 719–722.

[106] H. Sanders, M. A. Wilson, and S. J. Gershman, *Hippocampal remapping as hidden state inference*, *Elife* **9** (2020) e51140.

[107] E. R. Levy, S. Carrillo-Segura, E. H. Park, W. T. Redman, J. Hurtado, and A. A. Fenton, *A neuronal code for space in hippocampal coactivity dynamics independent of place fields*, *bioRxiv* (2021) 2021–07.

[108] M. K. Benna and S. Fusi, *Place cells may simply be memory cells: Memory compression leads to spatial tuning and history dependence*, *Proceedings of the National Academy of Sciences* **118** (2021), no. 51.

[109] K. L. Stachenfeld, M. M. Botvinick, and S. J. Gershman, *The hippocampus as a predictive map*, *Nature neuroscience* **20** (2017), no. 11 1643–1653.

[110] H. Eichenbaum, *Time cells in the hippocampus: a new dimension for mapping memories*, *Nature Reviews Neuroscience* **15** (2014), no. 11 732–744.

[111] J. W. Rueckemann, M. Sosa, L. M. Giocomo, and E. A. Buffalo, *The grid code for ordered experience*, *Nature Reviews Neuroscience* **22** (2021), no. 10 637–649.

[112] B. A. Olshausen and D. J. Field, *Sparse coding of sensory inputs*, *Current opinion in neurobiology* **14** (2004), no. 4 481–487.

[113] D. H. Hubel and T. N. Wiesel, *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex*, *The Journal of physiology* **160** (1962), no. 1 106.

[114] M. A. Wilson and B. L. McNaughton, *Dynamics of the hippocampal ensemble code for space*, *Science* **261** (1993), no. 5124 1055–1058.

[115] A. Krizhevsky, I. Sutskever, and G. Hinton, *Imagenet classification with deep convolutional neural networks*, in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[116] K. Simonyan and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, in *International Conference on Learning Representations*, 2015.

[117] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[118] K. Fukushima, *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position.*, *Biol. Cybernetics* **36** (1980) 193—-202.

[119] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, *Handwritten digit recognition with a back-propagation network*, in *Advances in neural information processing systems*, pp. 396–404, 1990.

[120] D. Scherer, A. Müller, and S. Behnke, *Evaluation of pooling operations in convolutional architectures for object recognition*, in *International conference on artificial neural networks*, pp. 92–101, Springer, 2010.

[121] J. Schmidhuber, *Deep learning in neural networks: An overview*, *Neural Networks* **61** (2015) 85–117.

[122] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT Press, 2016.

[123] E. P. Simoncelli and B. A. Olshausen, *Natural image statistics and neural representation*, *Annual review of neuroscience* **24** (2001), no. 1 1193–1216.

[124] G. Urban, K. J. Geras, S. E. Kahou, O. Aslan, S. Wang, A. Mohamed, M. Philipose, M. Richardson, and R. Caruana, *Do deep convolutional nets really need to be deep and convolutional?*, in *International Conference on Learning Representations*, 2017.

[125] B. A. Olshausen and D. J. Field, *Emergence of simple-cell receptive field properties by learning a sparse code for natural images*, *Nature* **381** (1996), no. 6583 607–609.

[126] P. Sterling and S. Laughlin, *Principles of neural design*. MIT press, 2015.

[127] F. Pellegrini and G. Biroli, *Neural network pruning denoises the features and makes local connectivity emerge in visual tasks*, in *International Conference on Machine Learning*, pp. 17601–17626, PMLR, 2022.

[128] A. Ingrosso and S. Goldt, *Data-driven emergence of convolutional structure in neural networks*, *arXiv preprint arXiv:2202.00565* (2022).

[129] S. d'Ascoli, L. Sagun, G. Biroli, and J. Bruna, *Finding the needle in the haystack with convolutions: on the benefits of architectural bias*, *Advances in Neural Information Processing Systems* **32** (2019).

[130] B. Neyshabur, *Towards learning convolutions from scratch*, *Advances in Neural Information Processing Systems* **33** (2020) 8078–8088.

[131] M. Tsodyks and T. Sejnowski, *Associative memory and hippocampal place cells*, *International journal of neural systems* **6** (1995) 81–86.

[132] F. P. Battaglia and A. Treves, *Attractor neural networks storing multiple space representations: a model for hippocampal place fields*, *Physical Review E* **58** (1998), no. 6 7738.

[133] A. Arleo and W. Gerstner, *Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity*, *Biological cybernetics* **83** (2000), no. 3 287–299.

[134] M. Franzius, R. Vollgraf, and L. Wiskott, *From grids to places*, *Journal of computational neuroscience* **22** (2007) 297–299.

[135] M. Franzius, H. Sprekeler, and L. Wiskott, *Slowness and sparseness lead to place, head-direction, and spatial-view cells*, *PLoS computational biology* **3** (2007), no. 8 e166.

[136] L. M. Giocomo, M.-B. Moser, and E. I. Moser, *Computational models of grid cells*, *Neuron* **71** (2011), no. 4 589–603.

[137] N. Burgess and J. O'Keefe, *Models of place and grid cell firing and theta rhythmicity*, *Current opinion in neurobiology* **21** (2011), no. 5 734–744.

[138] E. A. Zilli, *Models of grid cell spatial firing published 2005–2011*, *Frontiers in neural circuits* **6** (2012) 16.

[139] D. Bush and N. Burgess, *A hybrid oscillatory interference/continuous attractor network model of grid cell firing*, *Journal of Neuroscience* **34** (2014), no. 14 5065–5079.

[140] L. Castro and P. Aguiar, *A feedforward model for the formation of a grid field where spatial information is provided solely from place cells*, *Biological cybernetics* **108** (2014) 133–143.

[141] A. Stepanyuk, *Self-organization of grid fields under supervision of place cells in a neuron model with associative plasticity*, *Biologically Inspired Cognitive Architectures* **13** (2015) 48–62.

[142] T. D'Albis and R. Kempter, *A single-cell spiking model for the origin of grid-cell patterns*, *PLoS computational biology* **13** (2017), no. 10 e1005782.

[143] M. M. Monsalve-Mercado and C. Leibold, *Hippocampal spike-timing correlations lead to hexagonal grid fields*, *Physical review letters* **119** (2017), no. 3 038101.

[144] W. T. Redman, N. S. Wolcott, L. Montelisciani, G. Luna, T. D. Marks, K. K. Sit, C.-H. Yu, S. Smith, and M. J. Goard, *Long-term transverse imaging of the hippocampus with glass microperiscopes*, *Elife* **11** (2022) e75391.

[145] S. R. y Cajal, *Histologie du système nerveux de l'homme et des vertèbres*. Maloine, volume 2 ed., 1911.

[146] P. E. Gilbert, R. P. Kesner, and I. Lee, *Dissociating hippocampal subregions: A double dissociation between dentate gyrus and ca1*, *Hippocampus* **11** (2001) 626–636.

[147] I. Lee, D. Yoganarasimha, G. Rao, and J. J. Knierim, *Comparison of population coherence of place cells in hippocampal subfields ca1 and ca3*, *Nature* **430** (7, 2004) 456–459.

[148] T. J. McHugh, M. W. Jones, J. J. Quinn, N. Balthasar, R. Coppari, J. K. Elmquist, B. B. Lowell, M. S. Fanselow, M. A. Wilson, and S. Tonegawa, *Dentate gyrus nmda receptors mediate rapid pattern separation in the hippocampal network*, *Science* **317** (7, 2007) 94–99.

[149] T. Nakashiba, J. D. Cushman, K. A. Pelkey, S. Renaudineau, D. L. Buhl, T. J. McHugh, V. R. Barrera, R. Chittajallu, K. S. Iwamoto, C. J. McBain, M. S. Fanselow, and S. Tonegawa, *Young dentate granule cells mediate pattern separation, whereas old granule cells facilitate pattern completion*, *Cell* **149** (3, 2012) 188–201.

[150] M. A. Kheirbek, L. J. Drew, N. S. Burghardt, D. O. Costantini, L. Tannenholz, S. E. Ahmari, H. Zeng, A. A. Fenton, and R. Henl, *Differential control of learning and anxiety along the dorsoventral axis of the dentate gyrus*, *Neuron* **77** (2013) 955–968.

[151] J. P. Neunuebel and J. J. Knierim, *Ca3 retrieves coherent representations from degraded input: Direct evidence for ca3 pattern completion and dentate gyrus pattern separation*, *Neuron* **81** (1, 2014) 416–427.

[152] C. Rennó-Costa, J. E. Lisman, and P. F. Verschure, *A signature of attractor dynamics in the ca3 region of the hippocampus*, *PLoS Computational Biology* **10** (2014) e1003641.

[153] L. M. Frank, G. B. Stanley, and E. N. Brown, *Hippocampal plasticity across multiple days of exposure to novel environments*, *Journal of Neuroscience* **24** (9, 2004) 7681–7689.

[154] M. P. Karlsson and L. M. Frank, *Network dynamics underlying the formation of sparse, informative representations in the hippocampus*, *Journal of Neuroscience* **28** (12, 2008) 14271–14281.

[155] C. Kemere, M. F. Carr, M. P. Karlsson, and L. M. Frank, *Rapid and continuous modulation of hippocampal network state during exploration of new places.*, *PloS one* **8** (2013) e73114.

[156] M. C. Larkin, C. Lykken, L. D. Tye, J. G. Wickelgren, and L. M. Frank, *Hippocampal output area ca1 broadcasts a generalized novelty signal during an object-place recognition task*, *Hippocampus* **24** (2014) 773–783.

[157] C. Dong, A. D. Madar, and M. E. Sheffield, *Distinct place cell dynamics in ca1 and ca3 encode experience in new environments*, *Nature Communications* **12** (5, 2021) 1–13.

[158] F. L. Hitti and S. A. Siegelbaum, *The hippocampal ca2 region is essential for social memory*, *Nature* **508** (2, 2014) 88–92.

[159] T. Okuyama, T. Kitamura, D. S. Roy, S. Itohara, and S. Tonegawa, *Ventral ca1 neurons store social memory*, *Science* **353** (9, 2016) 1536–1541.

[160] T. Meira, F. Leroy, E. W. Buss, A. Oliva, J. Park, and S. A. Siegelbaum, *A hippocampal circuit linking dorsal ca2 to ventral ca1 critical for social memory dynamics*, *Nature Communications* **9** (12, 2018) 1–14.

[161] D. A. Dombeck, C. D. Harvey, L. Tian, L. L. Looger, and D. W. Tank, *Functional imaging of hippocampal place cells at cellular resolution during virtual navigation*, *Nature Neuroscience* **13** (11, 2010) 1433–1440.

[162] K. K. Ghosh, L. D. Burns, E. D. Cocker, A. Nimmerjahn, Y. Ziv, A. E. Gamal, and M. J. Schnitzer, *Miniaturized integration of a fluorescence microscope*, *Nature Methods* **8** (2011) 871–878.

[163] D. J. Cai, D. Aharoni, T. Shuman, J. Shobe, J. Biane, W. Song, B. Wei, M. Veshkini, M. La-Vu, J. Lou, *et. al.*, *A shared neural ensemble links distinct contextual memories encoded close in time*, *Nature* **534** (2016), no. 7605 115–118.

[164] L. Sheintuch, A. Rubin, N. Brande-Eilat, N. Geva, N. Sadeh, O. Pinchasof, and Y. Ziv, *Tracking the same neurons across multiple days in ca2+ imaging data*, *Cell Reports* **21** (10, 2017) 1102–1115.

[165] T. H. Chia and M. J. Levene, *Microprisms for in vivo multilayer cortical imaging*, *Journal of Neurophysiology* **102** (2009) 1310–1314.

[166] M. L. Andermann, N. B. Gilfoy, G. J. Goldey, R. N. Sachdev, M. Wölfel, D. A. McCormick, R. C. Reid, and M. J. Levene, *Chronic cellular imaging of entire cortical columns in awake mice using microprisms*, *Neuron* **80** (11, 2013) 900–913.

[167] R. J. Low, Y. Gu, and D. W. Tank, *Cellular resolution optical access to brain regions in fissures: Imaging medial prefrontal cortex and grid cells in entorhinal cortex*, *Proceedings of the National Academy of Sciences of the United States of America* **111** (12, 2014) 18739–18744.

[168] M. Kislin, E. Mugantseva, D. Molotkov, N. Kulesskaya, S. Khirug, I. Kirilkin, E. Pryazhnikov, J. Kolikova, D. Toptunov, M. Yuryev, R. Giniatullin, V. Voikar, C. Rivera, H. Rauvala, and L. Khiroug, *Flat-floored air-lifted platform: A new method for combining behavior with microscopy or electrophysiology on awake freely moving rodents*, *Journal of Visualized Experiments* (6, 2014).

[169] M. A. Go, J. Rogers, G. P. Gava, C. E. Davey, S. Prado, Y. Liu, and S. R. Schultz, *Place cells in head-fixed mice navigating a floating real-world environment*, *Frontiers in Cellular Neuroscience* **15** (2, 2021) 618658.

[170] P. Berens, *Circstat : A matlab toolbox for circular statistics*, *Journal of Statistical Software* **31** (9, 2009) 1–21.

[171] M. Pachitariu, C. Stringer, M. Dipoppa, S. Schröder, L. F. Rossi, H. Dalgleish, M. Carandini, and K. Harris, *Suite2p: beyond 10,000 neurons with standard two-photon microscopy*, *bioRxiv* (7, 2016) 061507.

[172] J. Friedrich, P. Zhou, and L. Paninski, *Fast online deconvolution of calcium imaging data*, *PLoS Computational Biology* **13** (3, 2017).

[173] A. J. Holtmaat, J. T. Trachtenberg, L. Wilbrecht, G. M. Shepherd, X. Zhang, G. W. Knott, and K. Svoboda, *Transient and persistent dendritic spines in the neocortex in vivo*, *Neuron* **45** (1, 2005) 279–291.

[174] A. Attardo, J. E. Fitzgerald, and M. J. Schnitzer, *Impermanence of dendritic spines in live adult ca1 hippocampus*, *Nature* **523** (7, 2015) 592–596.

[175] A. Mizrahi, J. C. Crowley, E. Shtoyerman, and L. C. Katz, *High-resolution in vivo imaging of hippocampal dendrites and spines*, *Journal of Neuroscience* **24** (3, 2004) 3147–3151.

[176] M. Iwase, T. Kitanishi, and K. Mizuseki, *Cell type, sub-region, and layer-specific speed representation in the hippocampal–entorhinal circuit*, *Scientific Reports* **10** (12, 2020) 1–23.

[177] Z. H. T. Góis and A. B. Tort, *Characterizing speed cells in the rat hippocampus*, *Cell Reports* **25** (11, 2018) 1872–1884.e4.

[178] B. Bechtold, *Violin plots for matlab, github project*, 2016.

[179] Q. Wang, S. L. Ding, Y. Li, J. Royall, D. Feng, P. Lesnar, N. Graddis, M. Naeemi, B. Facer, A. Ho, T. Dolbeare, B. Blanchard, N. Dee, W. Wakeman, K. E. Hirokawa, A. Szafer, S. M. Sunkin, S. W. Oh, A. Bernard, J. W. Phillips, M. Hawrylycz, C. Koch, H. Zeng, J. A. Harris, and L. Ng, *The allen mouse brain common coordinate framework: A 3d reference atlas*, *Cell* **181** (5, 2020) 936–953.e20.

[180] G. Paxinos and K. B. J. Franklin, *The Mouse Brain in Stereotaxic Coordinates*. Academic Press, 2nd ed., 2001.

[181] F. Engert and T. Bonhoeffer, *Dendritic spine changes associated with hippocampal long-term synaptic plasticity*, *Nature* **399** (5, 1999) 66–70.

[182] K. M. Harris, *Structure, development, and plasticity of dendritic spines*, *Current Opinion in Neurobiology* **9** (1999) 343–348.

[183] R. Yuste and T. Bonhoeffer, *Morphological changes in dendritic spines associated with long-term synaptic plasticity*, *Annual Review of Neuroscience* **24** (2001) 1071–1089.

[184] G. Yang, F. Pan, and W. B. Gan, *Stably maintained dendritic spines are associated with lifelong memories*, *Nature* **462** (12, 2009) 920–924.

[185] L. Gu, S. Kleiber, L. Schmid, F. Nebeling, M. Chamoun, J. Steffen, J. Wagner, and M. Fuhrmann, *Long-term in vivo imaging of dendritic spines in the hippocampus reveals structural plasticity*, *Journal of Neuroscience* **34** (10, 2014) 13948–13953.

[186] A. F. Ulivi, T. P. Castello-Waldow, G. Weston, L. Yan, R. Yasuda, A. Chen, and A. Attardo, *Longitudinal two-photon imaging of dorsal hippocampal ca1 in live mice*, *JoVE (Journal of Visualized Experiments)* **2019** (6, 2019) e59598.

[187] G. Feng, R. H. Mellor, M. Bernstein, C. Keller-Peck, Q. T. Nguyen, M. Wallace, J. M. Nerbonne, J. W. Lichtman, and J. R. Sanes, *Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of gfp*, *Neuron* **28** (10, 2000) 41–51.

[188] F. L. F. Chang and W. T. Greenough, *Transient and enduring morphological correlates of synaptic activity and efficacy change in the rat hippocampal slice*, *Brain Research* **309** (8, 1984) 35–46.

[189] A. Rodriguez, D. B. Ehlenberger, D. L. Dickstein, P. R. Hof, and S. L. Wearne, *Automated three-dimensional detection and shape classification of dendritic spines from fluorescence microscopy images*, *PLoS ONE* **3** (4, 2008) 1997.

[190] J. Son, S. Song, S. Lee, S. Chang, and M. Kim, *Morphological change tracking of dendritic spines based on structural features*, *Journal of Microscopy* **241** (3, 2011) 261–272.

[191] B. Lendvai, E. A. Stern, B. Chen, and K. Svoboda, *Experience-dependent plasticity of dendritic spines in the developing rat barrel cortex in vivo*, *Nature* **404** (4, 2000) 876–881.

[192] W. C. Risher, T. Ustunkaya, J. S. Alvarado, and C. Eroglu, *Rapid golgi analysis method for efficient and unbiased classification of dendritic spines*, *PLoS ONE* **9** (9, 2014) e107591.

[193] J. C. Fiala, M. Feinberg, V. Popov, and K. M. Harris, *Synaptogenesis via dendritic filopodia in developing hippocampal area ca1*, *Journal of Neuroscience* **18** (1998) 8900–8911.

[194] M. G. Stewart, N. I. Medvedev, V. I. Popov, R. Schoepfer, H. A. Davies, K. Murphy, G. M. Dallérac, I. V. Kraev, and J. J. Rodríguez, *Chemically induced long-term potentiation increases the number of perforated and complex postsynaptic densities but does not alter dendritic spine volume in ca1 of adult*

mouse hippocampal slices, *European Journal of Neuroscience* **21** (6, 2005) 3368–3378.

[195] T. P. Castello-Waldow, G. Weston, A. F. Ulivi, A. Chenani, Y. Loewenstein, A. Chen, and A. Attardo, *Hippocampal neurons with stable excitatory connectivity become part of neuronal representations*, *PLOS Biology* **18** (11, 2020) e3000928.

[196] J. Grutzendler, N. Kasthuri, and W. B. Gan, *Long-term dendritic spine stability in the adult cortex*, *Nature* **420** (12, 2002) 812–816.

[197] J. T. Trachtenberg, B. E. Chen, G. W. Knott, G. Feng, J. R. Sanes, E. Welker, and K. Svoboda, *Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex*, *Nature* **420** (12, 2002) 788–794.

[198] M. T. van Dijk and A. A. Fenton, *On how the dentate gyrus contributes to memory discrimination*, *Neuron* **98** (2018) 832–845.e5.

[199] T. W. Chen, T. J. Wardill, Y. Sun, S. R. Pulver, S. L. Renninger, A. Baohan, E. R. Schreiter, R. A. Kerr, M. B. Orger, V. Jayaraman, L. L. Looger, K. Svoboda, and D. S. Kim, *Ultrasensitive fluorescent proteins for imaging neuronal activity*, *Nature* **499** (7, 2013) 295–300.

[200] T. L. Daigle, L. Madisen, T. A. Hage, M. T. Valley, U. Knoblich, R. S. Larsen, M. M. Takeno, L. Huang, H. Gu, R. Larsen, M. Mills, A. Bosma-Moody, L. A. Siverts, M. Walker, L. T. Graybuck, Z. Yao, O. Fong, T. N. Nguyen, E. Garren, G. H. Lenz, M. Chavarha, J. Pendergraft, J. Harrington, K. E. Hirokawa, J. A. Harris, P. R. Nicovich, M. J. McGraw, D. R. Ollerenshaw, K. A. Smith, C. A. Baker, J. T. Ting, S. M. Sunkin, J. Lecoq, M. Z. Lin, E. S. Boyden, G. J. Murphy, N. M. da Costa, J. Waters, L. Li, B. Tasic, and H. Zeng, *A suite of transgenic driver and reporter mouse lines with enhanced brain-cell-type targeting and functionality*, *Cell* **174** (7, 2018) 465–480.e22.

[201] D. Aronov and D. W. Tank, *Engagement of neural circuits underlying 2d spatial navigation in a rodent virtual reality system*, *Neuron* **84** (10, 2014) 442–456.

[202] Z. M. Aghajan, L. Acharya, J. J. Moore, J. D. Cushman, C. Vuong, and M. R. Mehta, *Impaired spatial selectivity and intact phase precession in two-dimensional virtual reality*, *Nature Neuroscience* **18** (1, 2015) 121–128.

[203] P. Ravassard, A. Kees, B. Willers, D. Ho, D. Aharoni, J. Cushman, Z. M. Aghajan, and M. R. Mehta, *Multisensory control of hippocampal spatiotemporal selectivity*, *Science* **340** (6, 2013) 1342–1346.

[204] B. J. Claiborne, D. G. Amaral, and W. M. Cowan, *A light and electron microscopic analysis of the mossy fibers of the rat dentate gyrus*, *Journal of Comparative Neurology* **246** (4, 1986) 435–458.

[205] N. Ishizuka, J. Weber, and D. G. Amaral, *Organization of intrahippocampal projections originating from ca3 pyramidal cells in the rat*, *Journal of Comparative Neurology* **295** (5, 1990) 580–623.

[206] X. i, P. Somogyi, A. Ylinen, and G. Buzsáki, *The hippocampal ca3 network: An in vivo intracellular labeling study*, *Journal of Comparative Neurology* **339** (1, 1994) 181–208.

[207] N. Ishizuka, W. M. Cowan, and D. G. Amaral, *A quantitative analysis of the dendritic organization of pyramidal cells in the rat hippocampus*, *Journal of Comparative Neurology* **362** (11, 1995) 17–45.

[208] M. P. Witter, *Intrinsic and extrinsic wiring of ca3: Indications for connectional heterogeneity*, *Learning and Memory* **14** (11, 2007) 705–713.

[209] M. J. Levene, D. A. Dombeck, K. A. Kasischke, R. P. Molloy, and W. W. Webb, *In vivo multiphoton microscopy of deep brain tissue*, *Journal of Neurophysiology* **91** (4, 2004) 1908–1912.

[210] R. P. Barretto, T. H. Ko, J. C. Jung, T. J. Wang, G. Capps, A. C. Waters, Y. Ziv, A. Attardo, L. Recht, and M. J. Schnitzer, *Time-lapse imaging of disease progression in deep brain areas using fluorescence microendoscopy*, *Nature Medicine* **17** (2, 2011) 223–229.

[211] J. L. Gauthier and D. W. Tank, *A dedicated population for reward coding in the hippocampus*, *Neuron* **99** (7, 2018) 179–193.e7.

[212] S. Krishnan, C. Cherian, and M. E. J. Sheffield, *Changing reward expectation transforms spatial encoding and retrieval in the hippocampus*, *bioRxiv* (2020).

[213] N. L. Pettit, X. C. Yuan, and C. D. Harvey, *Hippocampal place codes are gated by behavioral engagement*, *Nature Neuroscience 2022 25:5* **25** (4, 2022) 561–566.

[214] A. R. Chambers and S. Rumpel, *A stable brain from unstable components: Emerging concepts and implications for neural computation*, *Neuroscience* **357** (8, 2017) 172–184.

[215] G. Mongillo, S. Rumpel, and Y. Loewenstein, *Inhibitory connectivity defines the realm of excitatory plasticity*, *Nature Neuroscience* **21** (10, 2018) 1463–1470.

[216] N. B. Danielson, G. F. Turi, M. Ladow, S. Chavlis, P. C. Petrantonakis, P. Poirazi, and A. Losonczy, *In vivo imaging of dentate gyrus mossy cells in behaving mice*, *Neuron* **93** (2, 2017) 552–559.e4.

[217] D. GoodSmith, X. Chen, C. Wang, S. H. Kim, H. Song, A. Burgalossi, K. M. Christian, and J. J. Knierim, *Spatial representations of granule cells and mossy cells of the dentate gyrus*, *Neuron* **93** (2, 2017) 677–690.e5.

[218] Y. Senzai and G. Buzsáki, *Physiological properties and behavioral correlates of hippocampal granule cells and mossy cells*, *Neuron* **93** (2, 2017) 691–704.e5.

[219] M. E. Sheffield and D. A. Dombeck, *Calcium transient prevalence across the dendritic arbour predicts place field properties*, *Nature* **517** (1, 2015) 200–204.

[220] M. E. Sheffield, M. D. Adoff, and D. A. Dombeck, *Increased prevalence of calcium transients across the dendritic arbor during place field formation*, *Neuron* **96** (10, 2017) 490–504.e5.

[221] M. D. Adoff, J. R. Climer, H. Davoudi, J. S. Marvin, L. L. Looger, and D. A. Dombeck, *The functional organization of excitatory synaptic input to place cells*, *Nature Communications* **12** (12, 2021).

[222] H. Jia, N. L. Rochefort, X. Chen, and A. Konnerth, *Dendritic organization of sensory input to cortical neurons in vivo*, *Nature* **464** (4, 2010) 1307–1312.

[223] D. E. Wilson, D. E. Whitney, B. Scholl, and D. Fitzpatrick, *Orientation selectivity and the functional clustering of synaptic inputs in primary visual cortex*, *Nature Neuroscience* **19** (8, 2016) 1003–1009.

[224] B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, and M.-B. Moser, *Path integration and the neural basis of the'cognitive map'*, *Nature Reviews Neuroscience* **7** (2006), no. 8 663–678.

[225] L. de Almeida, M. Idiart, and J. E. Lisman, *The input–output transformation of the hippocampal granule cells: from grid cells to place fields*, *Journal of Neuroscience* **29** (2009), no. 23 7504–7512.

[226] J. L. Kubie and S. E. Fox, *Do the spatial frequencies of grid cells mold the firing fields of place cells?*, *Proceedings of the National Academy of Sciences* **112** (2015), no. 13 3860–3861.

[227] D. Bush, C. Barry, D. Manson, and N. Burgess, *Using grid cells for navigation*, *Neuron* **87** (2015), no. 3 507–520.

[228] N. Van Strien, N. Cappaert, and M. Witter, *The anatomy of memory: an interactive overview of the parahippocampal–hippocampal network*, *Nature reviews neuroscience* **10** (2009), no. 4 272–282.

[229] D. Derdikman, J. R. Whitlock, A. Tsao, M. Fyhn, T. Hafting, M.-B. Moser, and E. I. Moser, *Fragmentation of grid cell maps in a multicompartment environment*, *Nature neuroscience* **12** (2009), no. 10 1325–1332.

[230] J. Krupic, M. Bauza, S. Burton, C. Barry, and J. O'Keefe, *Grid cell symmetry is shaped by environmental geometry*, *Nature* **518** (2015), no. 7538 232–235.

[231] R. Ismakov, O. Barak, K. Jeffery, and D. Derdikman, *Grid cells encode local positional information*, *Current Biology* **27** (2017), no. 15 2337–2343.

[232] B. Dunn, D. Wennberg, Z. Huang, and Y. Roudi, *Grid cells show field-to-field variability and this explains the aperiodic response of inhibitory interneurons*, *arXiv preprint arXiv:1701.04893* (2017).

[233] G. Ginosar, J. Aljadeff, L. Las, D. Derdikman, and N. Ulanovsky, *Are grid cells used for navigation? on local metrics, subjective spaces, and black holes*, *Neuron* (2023).

[234] J. J. Jun, N. A. Steinmetz, J. H. Siegle, D. J. Denman, M. Bauza, B. Barbarits, A. K. Lee, C. A. Anastassiou, A. Andrei, Ç. Aydın, *et. al.*, *Fully integrated silicon probes for high-density recording of neural activity*, *Nature* **551** (2017), no. 7679 232–236.

[235] N. A. Steinmetz, C. Aydin, A. Lebedeva, M. Okun, M. Pachitariu, M. Bauza, M. Beau, J. Bhagat, C. Böhm, M. Broux, *et. al.*, *Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings*, *Science* **372** (2021), no. 6539 eabf4588.

[236] R. J. Gardner, L. Lu, T. Wernle, M.-B. Moser, and E. I. Moser, *Correlation structure of grid cells is preserved during sleep*, *Nature neuroscience* **22** (2019), no. 4 598–608.

[237] L. McInnes, J. Healy, and J. Melville, *Umap: Uniform manifold approximation and projection for dimension reduction*, *arXiv preprint arXiv:1802.03426* (2018).

[238] M. Shamir and H. Sompolinsky, *Implications of neuronal diversity on population coding*, *Neural computation* **18** (2006), no. 8 1951–1986.

[239] M. I. Chelaru and V. Dragoi, *Efficient coding in heterogeneous neuronal populations*, *Proceedings of the National Academy of Sciences* **105** (2008), no. 42 16344–16349.

[240] J. Gjorgjieva, G. Drion, and E. Marder, *Computational implications of biophysical diversity and multiple timescales in neurons and synapses for circuit performance*, *Current opinion in neurobiology* **37** (2016) 44–52.

[241] N. Perez-Nieves, V. C. Leung, P. L. Dragotti, and D. F. Goodman, *Neural heterogeneity promotes robust learning*, *Nature communications* **12** (2021), no. 1 5791.

[242] N. J. Killian, M. J. Jutras, and E. A. Buffalo, *A map of visual space in the primate entorhinal cortex*, *Nature* **491** (2012), no. 7426 761–764.

[243] A. O. Constantinescu, J. X. O'Reilly, and T. E. Behrens, *Organizing conceptual knowledge in humans with a gridlike code*, *Science* **352** (2016), no. 6292 1464–1468.

[244] J. B. Julian, A. T. Keinath, G. Frazzetta, and R. A. Epstein, *Human entorhinal cortex represents visual space using a boundary-anchored grid*, *Nature neuroscience* **21** (2018), no. 2 191–194.

[245] M. Nau, T. Navarro Schröder, J. L. Bellmund, and C. F. Doeller, *Hexadirectional coding of visual space in human entorhinal cortex*, *Nature neuroscience* **21** (2018), no. 2 188–190.

[246] N. Wilming, P. König, S. König, and E. A. Buffalo, *Entorhinal cortex receptive fields are modulated by spatial attention, even without movement*, *Elife* **7** (2018) e31745.

[247] M. Klukas, M. Lewis, and I. Fiete, *Efficient and flexible representation of higher-dimensional cognitive variables with grid cells*, *PLoS computational biology* **16** (2020), no. 4 e1007796.

[248] M. Frey, M. W. Mathis, and A. Mathis, *Neuroai: If grid cells are the answer, is path integration the question?*, *Current Biology* **33** (2023), no. 5 R190–R192.

[249] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, *Imagenet: A large-scale hierarchical image database*, in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[250] S. Han, J. Pool, J. Tran, and W. Dally, *Learning both weights and connections for efficient neural network*, in *Advances in Neural Information Processing Systems* (C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, eds.), vol. 28, Curran Associates, Inc., 2015.

[251] J. Frankle and M. Carbin, *The lottery ticket hypothesis: Finding sparse, trainable neural networks*, in *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, OpenReview.net, 2019.

[252] U. Evci, T. Gale, J. Menick, P. S. Castro, and E. Elsen, *Rigging the lottery: Making all tickets winners*, in *International Conference on Machine Learning*, pp. 2943–2952, PMLR, 2020.

[253] T. Cohen and M. Welling, *Group equivariant convolutional networks*, in *Proceedings of The 33rd International Conference on Machine Learning* (M. F. Balcan and K. Q. Weinberger, eds.), vol. 48 of *Proceedings of Machine Learning Research*, (New York, New York, USA), pp. 2990–2999, PMLR, 2016.

[254] T. Cohen, M. Weiler, B. Kicanaoglu, and M. Welling, *Gauge equivariant convolutional networks and the icosahedral CNN*, in *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.), vol. 97 of *Proceedings of Machine Learning Research*, pp. 1321–1330, PMLR, 2019.

[255] R. Kondor and S. Trivedi, *On the generalization of equivariance and convolution in neural networks to the action of compact groups*, in *Proceedings of the 35th International Conference on Machine Learning* (J. Dy and A. Krause, eds.), vol. 80 of *Proceedings of Machine Learning Research*, pp. 2747–2755, PMLR, 2018.

[256] D. J. Rezende, S. Racanière, I. Higgins, and P. Toth, *Equivariant hamiltonian flows*, 2019.

[257] T. S. Cohen, M. Geiger, J. Köhler, and M. Welling, *Spherical CNNs*, in *International Conference on Learning Representations*, 2018.

[258] F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, and M. M. Bronstein, *Geometric deep learning on graphs and manifolds using mixture model cnns*, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[259] M. Weiler, M. Geiger, M. Welling, W. Boomsma, and T. S. Cohen, *3d steerable cnns: Learning rotationally equivariant features in volumetric data*, in *Advances in Neural Information Processing Systems* (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, eds.), vol. 31, Curran Associates, Inc., 2018.

[260] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, *Geometric deep learning: Going beyond euclidean data*, *IEEE Signal Processing Magazine* **34** (2017), no. 4 18–42.

[261] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, *Geometric deep learning: Grids, groups, graphs, geodesics, and gauges*, 2021.

[262] S. Mallat, *Group invariant scattering*, *Communications on Pure and Applied Mathematics* **65** (2012), no. 10 1331–1398.

[263] J. Bruna and S. Mallat, *Invariant scattering convolution networks*, *IEEE transactions on pattern analysis and machine intelligence* **35** (2013), no. 8 1872–1886.

[264] S. Mallat, *Understanding deep convolutional networks*, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **374** (2016), no. 2065 20150203.

[265] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, *Stand-alone self-attention in vision models*, in *Advances in Neural Information Processing Systems* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds.), vol. 32, Curran Associates, Inc., 2019.

[266] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, *An image is worth 16x16 words: Transformers for image recognition at scale*, in *International Conference on Learning Representations*, 2021.

[267] S. D'Ascoli, H. Touvron, M. L. Leavitt, A. S. Morcos, G. Biroli, and L. Sagun, *Convit: Improving vision transformers with soft convolutional inductive biases*, in *Proceedings of the 38th International Conference on Machine Learning* (M. Meila and T. Zhang, eds.), vol. 139 of *Proceedings of Machine Learning Research*, pp. 2286–2296, PMLR, 2021.

[268] J. Frankle, G. K. Dziugaite, D. Roy, and M. Carbin, *Linear mode connectivity and the lottery ticket hypothesis*, in *Proceedings of the 37th International Conference on Machine Learning* (H. D. III and A. Singh, eds.), vol. 119 of *Proceedings of Machine Learning Research*, pp. 3259–3269, PMLR, 13–18 Jul, 2020.

[269] A. Morcos, H. Yu, M. Paganini, and Y. Tian, *One ticket to win them all: generalizing lottery ticket initializations across datasets and optimizers*, in *Advances in Neural Information Processing Systems* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds.), vol. 32, Curran Associates, Inc., 2019.

[270] T. Chen, J. Frankle, S. Chang, S. Liu, Y. Zhang, Z. Wang, and M. Carbin, *The lottery ticket hypothesis for pre-trained bert networks*, in *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, eds.), vol. 33, pp. 15834–15846, Curran Associates, Inc., 2020.

[271] H. Yu, S. Edunov, Y. Tian, and A. S. Morcos, *Playing the lottery with rewards and multiple languages: lottery tickets in rl and nlp*, in *International Conference on Learning Representations*, 2020.

[272] T. Chen, J. Frankle, S. Chang, S. Liu, Y. Zhang, M. Carbin, and Z. Wang, *The lottery tickets hypothesis for supervised and self-supervised pre-training in computer vision models*, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16306–16316, June, 2021.

[273] M. Paul, F. Chen, B. W. Larsen, J. Frankle, S. Ganguli, and G. K. Dziugaite, *Unmasking the lottery ticket hypothesis: What's encoded in a winning ticket's mask?*, *arXiv preprint arXiv:2210.03044* (2022).

[274] U. Evci, Y. Ioannou, C. Keskin, and Y. Dauphin, *Gradient flow in sparse neural networks and how lottery tickets win*, in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 6577–6586, 2022.

[275] M. Paul, B. W. Larsen, S. Ganguli, J. Frankle, and G. K. Dziugaite, *Lottery tickets on a data diet: Finding initializations with sparse trainable networks*, *arXiv preprint arXiv:2206.01278* (2022).

[276] Z. Ankner, A. Renda, G. K. Dziugaite, J. Frankle, and T. Jin, *The effect of data dimensionality on neural network prunability*, *arXiv preprint arXiv:2212.00291* (2022).

[277] S. Ioffe and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, in *International conference on machine learning*, pp. 448–456, pmlr, 2015.

[278] S. A. Janowsky, *Pruning versus clipping in neural networks*, *Phys. Rev. A* **39** (1989) 6600–6603.

[279] M. C. Mozer and P. Smolensky, *Skeletonization: A Technique for Trimming the Fat from a Network via Relevance Assessment*, p. 107–115. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1989.

[280] M. C. Mozer and P. Smolensky, *Using relevance to reduce network size automatically*, *Connection Science* **1** (1989), no. 1 3–16.

[281] Y. LeCun, J. S. Denker, and S. A. Solla, *Optimal brain damage*, in *Advances in neural information processing systems*, pp. 598–605, 1990.

[282] E. D. Karnin, *A simple procedure for pruning back-propagation trained neural networks*, *IEEE Transactions on Neural Networks* **1** (1990), no. 2 239–242.

[283] D. Blalock, J. J. Gonzalez Ortiz, J. Frankle, and J. Guttag, *What is the state of neural network pruning?*, *Proceedings of machine learning and systems* **2** (2020) 129–146.

[284] W. T. Redman, M. Fonoberova, R. Mohr, Y. Kevrekidis, and I. Mezic, *An operator theoretic view on pruning deep neural networks*, in *International Conference on Learning Representations*, 2021.

[285] H. Yu, S. Edunov, Y. Tian, and A. S. Morcos, *Playing the lottery with rewards and multiple languages: lottery tickets in rl and nlp*, *arXiv preprint arXiv:1906.02768* (2019).

[286] T. Chen, Z. Zhang, S. Liu, S. Chang, and Z. Wang, *Long live the lottery: The existence of winning tickets in lifelong learning*, in *International Conference on Learning Representations*, 2021.

[287] X. Chen, Y. Cheng, S. Wang, Z. Gan, J. Liu, and Z. Wang, *The elastic lottery ticket hypothesis*, *Advances in Neural Information Processing Systems* **34** (2021) 26609–26621.

[288] A. Renda, J. Frankle, and M. Carbin, *Comparing rewinding and fine-tuning in neural network pruning*, in *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*, OpenReview.net, 2020.

[289] W. T. Redman, T. Chen, Z. Wang, and A. S. Dogra, *Universality of winning tickets: A renormalization group perspective*, in *International Conference on Machine Learning*, pp. 18483–18498, PMLR, 2022.

[290] M. V. Srinivasan, S. B. Laughlin, and A. Dubs, *Predictive coding: a fresh view of inhibition in the retina*, *Proceedings of the Royal Society of London. Series B. Biological Sciences* **216** (1982), no. 1205 427–459.

[291] D. J. Field, *Relations between the statistics of natural images and the response properties of cortical cells*, *Josa a* **4** (1987), no. 12 2379–2394.

[292] J. J. Atick, *Could information theory provide an ecological theory of sensory processing?*, *Network: Computation in neural systems* **3** (1992), no. 2 213–251.

[293] *Real and optimal neural images in early vision*, *Nature* **360** (1992), no. 6399 68–70.

[294] D. L. Ruderman and W. Bialek, *Statistics of natural images: Scaling in the woods*, *Physical review letters* **73** (1994), no. 6 814.

[295] D. J. Field, *What is the goal of sensory coding?*, *Neural computation* **6** (1994), no. 4 559–601.

[296] S. Saremi and T. J. Sejnowski, *Hierarchical model of natural images and the origin of scale invariance*, *Proceedings of the National Academy of Sciences* **110** (2013), no. 8 3071–3076, [https://www.pnas.org/content/110/8/3071.full.pdf].

[297] G. Bellec, D. Kappel, W. Maass, and R. Legenstein, *Deep rewiring: Training very sparse deep networks*, *arXiv preprint arXiv:1711.05136* (2017).

[298] D. C. Mocanu, E. Mocanu, P. Stone, P. H. Nguyen, M. Gibescu, and A. Liotta, *Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science*, *Nature communications* **9** (2018), no. 1 2383.

[299] X. Dai, H. Yin, and N. K. Jha, *Nest: A neural network synthesis tool based on a grow-and-prune paradigm*, *IEEE Transactions on Computers* **68** (2019), no. 10 1487–1497.

[300] H. Mostafa and X. Wang, *Parameter efficient training of deep convolutional neural networks by dynamic sparse reparameterization*, in *International Conference on Machine Learning*, pp. 4646–4655, PMLR, 2019.

[301] S. Liu, D. C. Mocanu, A. R. R. Matavalam, Y. Pei, and M. Pechenizkiy, *Sparse evolutionary deep learning with over one million artificial neurons on commodity hardware*, *Neural Computing and Applications* **33** (2021) 2589–2604.

[302] S. Jastrzkebski, Z. Kenton, D. Arpit, N. Ballas, A. Fischer, Y. Bengio, and A. Storkey, *Three factors influencing minima in sgd*, *arXiv preprint arXiv:1711.04623* (2017).

[303] M. Andriushchenko, A. Varre, L. Pillaud-Vivien, and N. Flammarion, *Sgd with large step sizes learns sparse features*, *arXiv preprint arXiv:2210.05337* (2022).

[304] H. B. Barlow, *Unsupervised learning*, *Neural computation* **1** (1989), no. 3 295–311.

[305] A. Bell and T. J. Sejnowski, *Edges are the 'independent components'of natural scenes.*, in *Advances in Neural Information Processing Systems* (M. Mozer, M. Jordan, and T. Petsche, eds.), vol. 9, MIT Press, 1996.

[306] M. Refinetti, A. Ingrosso, and S. Goldt, *Neural networks trained with sgd learn distributions of increasing complexity*, *arXiv preprint arXiv:2211.11567* (2022).

[307] N. Lee, T. Ajanthan, and P. Torr, *SNIP: SINGLE-SHOT NETWORK PRUNING BASED ON CONNECTION SENSITIVITY*, in *International Conference on Learning Representations*, 2019.

[308] H. Tanaka, D. Kunin, D. L. Yamins, and S. Ganguli, *Pruning neural networks without any data by iteratively conserving synaptic flow*, *Advances in neural information processing systems* **33** (2020) 6377–6389.

[309] Z. Zhang, X. Chen, T. Chen, and Z. Wang, *Efficient lottery ticket finding: Less data is more*, in *International Conference on Machine Learning*, pp. 12380–12390, PMLR, 2021.

[310] S. Liu and Z. Wang, *Ten lessons we have learned in the new" sparseland": A short handbook for sparse neural network researchers*, *arXiv preprint arXiv:2302.02596* (2023).

[311] T. Hainmueller and M. Bartos, *Dentate gyrus circuits for encoding, retrieval and discrimination of episodic memories*, *Nature Reviews Neuroscience* **21** (2020), no. 3 153–168.

[312] D. Dvorak, A. Chung, E. H. Park, and A. A. Fenton, *Dentate spikes and external control of hippocampal function*, *Cell reports* **36** (2021), no. 5 109497.

[313] A. Mathis, A. V. Herz, and M. B. Stemmler, *Resolution of nested neuronal representations can be exponential in the number of neurons*, *Physical review letters* **109** (2012), no. 1 018103.

[314] T. O. Sharpee, *An argument for hyperbolic geometry in neural circuits*, *Current opinion in neurobiology* **58** (2019) 101–104.

[315] B. Sorscher, R. Geirhos, S. Shekhar, S. Ganguli, and A. S. Morcos, *Beyond neural scaling laws: beating power law scaling via data pruning*, *arXiv preprint arXiv:2206.14486* (2022).

[316] A. Nayebi, A. Attinger, M. Campbell, K. Hardcastle, I. Low, C. S. Mallory, G. Mel, B. Sorscher, A. H. Williams, S. Ganguli, *et. al.*, *Explaining heterogeneity in medial entorhinal cortex with task-driven neural networks*, *Advances in Neural Information Processing Systems* **34** (2021) 12167–12179.

[317] B. O. Koopman, *Hamiltonian systems and transformation in hilbert space*, *Proceedings of the National Academy of Sciences* **17** (1931), no. 5 315–318, [https://www.pnas.org/content/17/5/315.full.pdf].

[318] B. O. Koopman and J. v. Neumann, *Dynamical systems of continuous spectra*, *Proceedings of the National Academy of Sciences* **18** (1932), no. 3 255–263, [https://www.pnas.org/content/18/3/255.full.pdf].

[319] I. Mezić, *Spectral properties of dynamical systems, model reduction and decompositions*, *Nonlinear Dynamics* **41** (2005) 309–325.

[320] M. Budišić, R. Mohr, and I. Mezić, *Applied koopmanism*, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **22** (2020/10/24, 2012) 047510.

[321] I. Mezić, *Spectrum of the koopman operator, spectral expansions in functional spaces, and state-space geometry*, *Journal of Nonlinear Science* **30** (Oct, 2020) 2091–2145.

[322] I. Mezić, *Koopman operator, geometry, and learning of dynamical systems*, *Not. Am. Math. Soc* **68** (2021), no. 7 1087–1105.

[323] S. L. Brunton, M. Budišić, E. Kaiser, and J. N. Kutz, *Modern koopman theory for dynamical systems*, *SIAM Review* **64** (2022), no. 2 229–340.

[324] A. S. Dogra and W. T. Redman, *Optimizing neural networks via koopman operator theory*, in *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, eds.), vol. 33, pp. 2087–2097, Curran Associates, Inc., 2020.

[325] M. E. Tano, G. D. Portwood, and J. C. Ragusa, *Accelerating training in artificial neural networks with dynamic mode decomposition*, *arXiv preprint arXiv:2006.14371* (2020).

[326] R. Mohr, M. Fonoberova, I. Manojlovic, A. Andrejcuk, Z. Drmac, Y. Kevrekidis, and I. Mezic, *Applications of koopman mode analysis to neural networks.*, in *AAAI Spring Symposium: MLPS*, 2021.

[327] R. Mohr, M. Fonoberova, Z. Drmač, I. Manojlović, and I. Mezić, *Predicting the critical number of layers for hierarchical support vector regression*, *Entropy* **23** (2021), no. 1.

[328] I. Naiman and O. Azencot, *A koopman approach to understanding sequence neural models*, *arXiv preprint arXiv:2102.07824* (2021).

[329] D. Luo, J. Shen, R. Dangovski, and M. Soljačić, *Koopman operator learning for accelerating quantum optimization and machine learning*, *arXiv preprint arXiv:2211.01365* (2022).

[330] Z. Liang, C. Zhao, W. Liu, B. Xue, W. Yang, and Z. Pang, *Credit assignment for trained neural networks based on koopman operator theory*, *arXiv preprint arXiv:2212.00998* (2022).

[331] P. Šimánek, D. Vašata, and P. Kordík, *Learning to optimize with dynamic mode decomposition*, in *2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2022.

[332] W. T. Redman, J. M. Bello-Rivas, M. Fonoberova, R. Mohr, I. G. Kevrekidis, and I. Mezić, *On equivalent optimization of machine learning methods*, *arXiv preprint arXiv:2302.09160* (2023).

[333] W.-M. M. Hsu, D. B. Kastner, S. A. Baccus, and T. O. Sharpee, *How inhibitory neurons increase information transmission under threshold modulation*, *Cell reports* **35** (2021), no. 8 109158.

[334] A. A. Fenton and R. U. Muller, *Place cell discharge is extremely variable during individual passes of the rat through the firing field*, *Proceedings of the National Academy of Sciences* **95** (1998), no. 6 3182–3187.