The Dissertation Committee for Ki Jung Yoon
certifies that this is the approved version of the following dissertation:

Unraveling the dynamics and structure of grid cells as a spatial map in the brain

Committee:

Ila Fiete, Supervisor

Sriram Vishwanath, Co-Supervisor

Alan Bovik

Peter Stone

Jonathan Pillow

Unraveling the dynamics and structure of grid cells as a spatial map in the brain

by

Ki Jung Yoon, B.S.; M.S.E.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2015

Dedicated to my family

Jisun, Taehee, Hwaram

&

to my parents

Acknowledgments

This thesis would not have been possible without the support of a huge number of people. I am grateful for all the guidance I have received during the past five years by my advisors, Drs. Ila Fiete and Sriram Vishwanath. Ila's insight, focus, and remarkable clarity of thought helped me mature in ways I could have the confidence to develop into the researcher I am today. I am very lucky to have Sriram as my ECE advisor, and am grateful for his openness that allowed me to explore and focus on new directions. To the rest of my commitee: Drs. Jonathan Pillow, Alan Bovik, and Peter Stone, thank you for the time you devoted to my thesis, the advice, and direction you gave me whenever I needed it. I would also like to thank Dr. Haris Vikalo for his role as an advisor of CommNetS, which provided me with valuable comments on my research.

I am fortunate to have had several fantastic collaborators: many thanks go to Sam Lewallen, Amina Kinkhabwala, Cristina Domnisoru, and Yi Gu for exciting and warmhearted conversations and for introducing me to many interesting questions on the navigation in virtual reality. I won't forget the many light moments we shared over meals at Princeton and Austin. I thank David Tank, Caswell Barry, and Neil Burgess: lots of discussions and the opportunity to get access to unique grid cell data made my research influential

from the beginning.

I wish to thank John Widloski, Yongseok Yoo, Daniel Robles-Llana, Michael Buice, Sameet Sreenivasan, Birgit Kriener, Rishi Chaudhuri, and Ingmar Kanitscheider for all the help and great discussions in the Fiete lab. Thanks to Il Memming Park for his mentorship. And my fellow students in ECE and CLM – the list would go on much longer, but a special thanks to Duehee Lee, Juhun Lee, Mijung Park, Larkkwon Choi, Yujin Kim, Kwangwoo Ko, and Chungsub Kim for being there when the going got tough.

And lastly, I would not have even started on this path without unwavering support, love, and encouragement of my parents and family. I am deeply grateful to Jisun who has selflessly and lovingly given her all to adorable boys, Taehee & Hwaram.

Unraveling the dynamics and structure of grid cells as a spatial map in the brain

Publication No. _____

Ki Jung Yoon, Ph.D.
The University of Texas at Austin, 2015

Supervisor: Ila Fiete
Co-Supervisor: Sriram Vishwanath

Grid cells, defined by their strikingly periodic spatial responses in open fields, have spurred widespread theoretical interest, and numerous models have been proposed to explain how grids are formed, how they are differentiated from the others, and how they might use idiothetic (self motion) information to path integrate. This dissertation leverages unique grid cell data together with computational and mathematical approaches to unravel grid cell dynamics and structure during navigation in general. First, we analyze several extensive datasets of grid cells recorded in 2-dimensional (2D) environments under a number of experimental manipulations, and show that the multi-dimensional network activity of grid cells is embedded into a two-dimensional continuous attractor manifold. Second, we analyze grid cell responses on linear 1-dimensional (1D) tracks to extract an underlying 2D grid structure. Combining Fourier analytical methods and numerical refinements, we show that

the system remains in the same dynamical regime during navigation in 2D and 1D environments. Finally, we introduce a state-space point process filter to track the temporal evolution of spatial tuning curves and examine the error accumulation of grid cell system. We show that we can accurately infer the drift of the internal estimate of positions subsumed in the grid cell system as a path integrator.

Table of Contents

List of Figures

Chapter 1

Introduction

When we stand in a room, we can tell where we are by visual stimuli, but if we close our eyes we would still have an awareness of our location in the room. For example, most people probably experience in an unfamiliar hotel room that they have to get up and go to the bathroom in the middle of the night during complete darkness, and that depends on one's ability to keep track of one's location independent of any other sensory stimuli. This is a function called *path integration* or *dead reckoning* – an integrative process that requires computing ongoing location estimates through the integration of self-motion cues. Long before neuroscientists began investigating the cortex to understand the function, Darwin hypothesized over a century ago that humans and animals may have some parts of the brain specialized for this function that does not solely depend on the external world [24]:

With regard to the question of the means by which
animals find their way home from a long distance,...
Yet no one will suppose that they possessed
any special sense which is quite absent in us.
We must bear in mind that neither a compass,
nor the north star, nor any other such sign,
suffices to guide a man to a particular spot,...

unless the deviations are allowed for,
or a sort of "dead reckoning" is kept.
All men are able to do this
in a greater or less degree,...
though probably in an unconscious manner.

(Darwin, 1873)

The study of the neural basis for spatial navigation and representation started more recently by discovering potential neural correlates of path integration from the hippocampus [73] and the entorhinal cortex [41], located at the top of the cortical hierarchy and farthest from the primary sensory cortices. The findings opened a new era of experimental and computational studies of a spatial map and more broadly the high-level cortical functions in the brain because the firing of cells were generally known to become dissociated from any experimentally defined stimulus features as we go away from the low-level sensory cortices [37, 26]. Contrary to the widely-accepted theory, the discovery was an obvious exception to the apparent decoupling from the external world. Many cells in this part of the brain emitted action potentials only when the animal was in a specific set of locations in its local environment. The preferred locations differed from cell to cell, such that as a population, a sequence of locations (or movement trajectory) was coded by a sequence of firings of a unique cell population in the environment. Because of these unique activity combinations, people believed that the cells would effectively serve as a map of the animal's position.

1.1 Neural basis of spatial representation in the brain

The study of the neural basis of space began in 1971 by O'Keefe and Dostrovsky [73] with investigating activity from single neurons in the hippocampal subregion (CA1) of freely moving rats. Many of their cells with location-dependent firing were named "place cells" because the cell fired at a high rate when the rat was in the cell's "place field" while the firing activity decreased and remained low as the animal left the field until the next time the animal came to the same field. Their successive study also discovered that most hippocampal cells had place fields and that the exact firing locations of one cell differed from the other [71]. The fact that each location in the environment was associated with a unique combination of active place cells suggested that the place cells of the hippocampus manifested the *cognitive map* of Tolman [70], who proposed that behavior was guided by a map-like neural representation of stimulus relationships in the environment, rather than by chains of stimulus-response associations [92]. The strong relationship between neural activity and an external property – the animal's location – was unique among all the recordings that had been made in high-level cortices by that time.

The discovery of place cells was followed by three decades of studies uncovering that place cells may have functions that extend beyond a specific role in the mapping of the physical space. This idea was supported by the report that different maps could be activated by small environmental changes [67, 59, 97, 57, 58], implying that if the brain has a universal map for distances

and orientations that are independent of what the environment looks like, then such a map should not be located in the hippocampus but elsewhere, possibly the entorhinal cortex that is upstream of the hippocampus.

Pursuing the question, Moser group discovered another type of cells in the medial entorhinal cortex in 2005 [41]. Individual cells had discrete firing fields, like place cells in the hippocampus, but each cell had multiple fields, and the fields collectively formed a hexagonal array, with the equilateral triangles as a unit spanning the entirety of the animal's environment. Because of their grid-like periodic firing pattern of these cells, they were named "grid cells". The grid structure was similar for all grid cells, but the spacing of the fields, the orientation of the grids, and the x-y coordinate of the grid vertices varied. Grid cells were initially observed in rats and mice, and more recently they have been reported also in bats [100], monkeys [52], and humans [49], suggesting that they are present widely across all the mammals.

A notable property of the grid cells was the persistence of the firing pattern even in darkness and in the presence of changes in the animal's speed and direction [41]. Moreover, when two grid cells were recorded at the same time, the relationship between their grid fields tended to be preserved from one environment to the next [33, 102]. If the grid fields of two simultaneously recorded cells were out of phase in one task, they would generally be in the next too. The stability of this relationship was quite different from the behavior of place cells in the hippocampus, which were known to have the capacity to switch between completely uncorrelated firing patterns (*global remapping*).

The coherent responses of simultaneously recorded grid cells suggested that the grid cells may provide a universal metric for space in the brain.

Further recordings revealed that grid cells are organized into a small number of discrete modules along the dorsal-ventral axis of the entorhinal cortex [87]. Different grid maps varied on a number of parameters, including the spacing and the orientation of the grid axes. At the dorsal end, most grid cells had tightly packed grid fields and the grid spacing expanded in a discrete manner – not continuously – as we moved away from the dorsal border toward the ventral part. The most significant finding was the fact that different modules could respond independently to the geometric changes of the environment [87].

One of the explanations for such an organization of the brain's space map is grounded on the way the output of the entorhinal cortex is used downstream in the hippocampus. The hippocampus is known to be critical for certain types of memory, often referred to as *episodic memory*. The general statement to understand the episodic memory is "What did you do at a certain time in a certain place?" [93]. Spatial memory is a fundamental subset of this memory and thus the hippocampus needs to find a way of constructing, maintaining, differentiating, and retrieving lots of internal maps in the brain for processing thousands of episodic memories every day. This is where multiple grid modules could contribute to be helpful. If two modules respond independently to a change in the environment, their coactivity will vary. The change in coactivity will trigger a new subset of hippocampal neurons, which in turn

with a small number of grid modules the entorhinal cortex can be linked to a huge amount of hippocampal activity patterns – putative memories. An analogy can be made with a clock storing 43,200 codes (12hrs $\times$ 60mins $\times$ 60secs) with only 3 hands. Input from a few independent grid modules may enable the hippocampus to generate a large number of distinct representations linked to specific places and experiences.

1.2 Theoretical perspectives on grid cells

The universality of grid cell responses across environments in terms of the persistently periodic firing pattern regardless of the color, shape, location, or size of the enclosure suggests that they encode relative displacements rather than absolute location, and thus grid cells may be the neural loci of path integration.

The grid cell code has been analyzed theoretically and found to be a highly efficient code for encoding large displacements at a high spatial resolution [32, 29, 85, 94]. This is an essential characteristic for velocity-to-position integrators, because a coarse representation of displacement would accrue errors during path integration that would severely constrain the practicality of the computation. Disparate computational models of the possible mechanisms underlying the observed response of grid cells also rely on the conversion of velocity inputs into location signals [64, 32, 21, 38, 44, 20, 10, 43, 18, 95]. That is, devising mechanisms that might underlie the instantaneous responses of grid cells usually involves assigning grid cells the function of path integration.

What are the main ingredients for path integration? It requires computing ongoing location estimates through the integration of self-motion cues, where short-term memory is essential because data sources (current location estimates) should not disappear before consequent decisions or actions are taken. The neural substrate for short-term memory is believed to be a stable pattern of persistent neural activity, and such persistent activity is thought to possess an underlying continuous attractor structure [8, 103, 83, 18]: the instantaneous value of a continuous variable is represented by setting the system to a point on a continuous manifold of stable fixed points, and the stability of the fixed points generates a far longer time constant for memory than cellular time-constants. This property enables the system to track continuous time-varying stimuli such as the direction of heading during navigation, the location of the eye, or the orientation of a visual stimulus. This is the reason continuous attractor dynamics have been widely hypothesized to underlie the system of neural substrates (grid cells, place cells, head-direction cells) for spatial navigation.

1.3 Chapter overview

Chapter 2: What are the implications of grid cells being generated by the attractor network model? One of the major criticisms that are often raised against continuous attractor models of grid cells and against other attractor network models in many different brain areas is that they do not easily generate testable predictions. This work published in *Nature Neuroscience*

(2013) has taken this challenge head-on. The main predicted signature of an attractor mechanism in the neural context is that the responses of cells, when plotted against an external represented variable, may change substantially, but pairwise relationships between cells should remain stable. It is a feature that arises naturally in the attractor model because the mapping from the attractor states to the external variable might vary due to changing associations between the network and the external world (for example, rotation of the mapping induced by rotation of the world or more elaborate changes involving stretching or squeezing of tuning curves), but the cell-cell relationships are determined by fixed recurrent connectivity in the network.

Based on this theory, we analyzed several extensive datasets of grid cells (from Moser group & Burgess group), and showed that the spatial phase relationship between pairs of grid cells persists over time and across environmental manipulations, despite substantial changes in the sensory input. The finding that cell-cell relationships are better preserved across environments than responses of single cells suggests that the multi-dimensional network activity of grid cells is embedded into a two-dimensional (2D) manifold. Pursuing further the question of whether the 2D manifold is attractive (stable), we found evidence that external perturbations, in the form of velocity inputs, drive the grid-cell system away from the 2D manifold. Moreover, the off manifold (perturbed) components of the network state were shown to relax back to the 2D manifold on a time scale similar to or faster than the few second time scale on which the perturbing input changed. Our population analysis showing that

each grid-cell network is localized to a 2D manifold and that the manifold is attractive constitutes specific and direct evidence in support of continuous attractor network models of grid-cell activity.

Chapter 3: Does grid cell system remain in the same dynamical and computational regime during 2D and 1D navigation? Grid cells are defined by their striking, periodic spatial firing patterns in 2D enclosures. However, their responses in 1D environments (linear/circular track, T-/radial-/hairpin maze) appear to be considerably more irregular and complex. To date, many studies record grid cell responses in 1D (to study spatial tuning over large distances; to generate imaging records in virtual environments; etc.), but there is no systematic study or understanding of the relationship between the 1D and 2D response of grid cells. The key question is whether grid cells in 1D really continue to operate like grid cells in 2D, or whether they transition into a distinct computational and dynamical regime.

This work under review in *Neuron (2015)* presents an analytical and numerical study of the same grid cells, recorded first in 2D environments then on 1D linear virtual tracks. We consider the hypothesis that 1D grid cell responses are generated by slicing an underlying periodic 2D lattice. We first establish analytically that slices through regular lattices must have very specific structures in their Fourier spectra, and proceed to demonstrate that the data exhibit these structures. We show that lattice slices generate excellent fits to the 1D responses. Next, we further show that group properties of simulta-

neously recorded cells in the 2D response are predicted by parallel slice fits to the 1D responses, and therefore the system remains in the same fundamental dynamical and computational regime during navigation along stereotyped 1D routes as during free foraging in open arenas.

Chapter 4: How to infer integrative properties of grid cell system? To understand the network dynamics and fully investigate the predictions of a theoretical model for grid cell system, it is essential to look into fine temporal dynamics at the resolution of the same order of magnitude as the neural spiking. Grid cells are thought to perform path integration, an idea that dominates the current thinking about the functional role of the medial entorhinal cortex. The additive nature of continuous integration of self-motion cues results in an accumulation of error because an error in the current position estimate will be propagated to all the following estimates without any correction mechanism. In such models, a natural question is whether there exists any evidence of error accumulation [42]. Experimentally recorded grid cell data, however, appear relatively stable over a few tens of minutes of recording session, implying that either there is no accumulation of error, there is an ongoing error-correction mechanism, or there is an accumulation of error that cannot easily be seen by the naked eye – the current histogram-based approach – but be captured in the microscopic scale.

To study fine temporal dynamics and examine the error accumulation (and possibly correction) of grid cells, we apply a linear state-space model

observed through a point process [28] for characterizing the temporal evolution of spatial firing patterns, or *spatial tuning curves*. The point process adaptive filtering framework aims not only to estimate the system parameters at the same temporal resolution of the spiking ($\sim$ 1-10 msec), but also to infer how diffusive the grid cell system would be. Here we show that the mismatch between the actual and the internal estimate of an animal's positions over time can be well kept track of by this modeling framework given simulated data from a continuous attractor network model of grid cells. We find that the model allows us to accurately infer to what extent the spatial tuning of grid cells undergoes diffusion over time.

Chapter 2

Specific evidence of low-dimensional continuous attractor dynamics in grid cells

We examined simultaneously recorded spikes from multiple rat grid cells, to explain mechanisms underlying their activity. Among grid cells with similar spatial periods, the population activity was confined to lie close to a two-dimensional (2D) manifold: grid cells differed only along two dimensions of their responses and otherwise were nearly identical. Relationships between cell pairs were conserved despite extensive deformations of single-neuron responses. Results from novel environments suggest such structure is not inherited from hippocampal or external sensory inputs. Across conditions, cell-cell relationships are better conserved than responses of single cells. Finally, the system is continually subject to perturbations that, were the 2D manifold not attractive, would drive the system to inhabit a different region of state space than observed. These findings have strong implications for theories of grid-cell activity and substantiate the general hypothesis that the brain computes using low-dimensional continuous attractors.

2.1 Introduction

A set of N uncoupled spiking neurons, each with dynamic range Q, supply a vast representational space (volume $\sim Q^N$), (**Figure 2.1a**, top). However, the representation has poor resistance to noise: each state is independent and if changed to another, there is no restoring dynamics to correct the state. Even in the absence of noise, the states persist only for the time-constant of single neurons.

Coupling between neurons generally disallows many states, shrinking the representational space (**Figure 2.1a**, top and bottom). An advantage of coupling is that it can, in special cases, produce stable fixed points (attractors) of the network dynamics that allow the network to hold a state after inputs are removed, for far longer than the single-neuron time-constant. Moreover, if noise is present in the system, it may perturb the system off the attractor, but the perturbations are transient and automatically corrected as the system rapidly flows back toward the attractor (**Figure 2.1a**, top). Discrete or point attractors, as in Hopfield networks, may be used to represent discrete items [48]. In many cases, the brain must represent continuous variables. In these cases, the value of the variable could be represented as a point on a continuous manifold of stable fixed points, of the same dimensionality D as the variable [8, 83, 103, 84]. This manifold is called a low-dimensional continuous attractor, if its dimensionality is much smaller than the number of neurons in the network ($D \ll N$). In these ways, attractors enable robust representation and memory, albeit at the cost of diminished representational space.

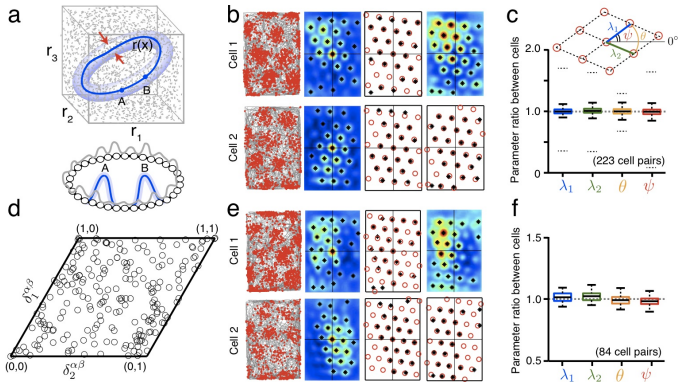


Figure 2.1: **(a)** Top (state space): The state of N independent neurons, each described by a firing rate r_i in $[0, r_{max}]$, may lie anywhere in an N-dimensional cube of side length r_{max} (shown for $N = 3$ neurons). Appropriate coupling between the neurons can shrink the allowed states to a low-dimensional attractor (dark blue). All other states are transient, rapidly decaying back to the attractor, and are thus rarely seen. States very close to the attractor (light blue), through transient, may be observed if perturbations frequently drive the system into those states. Bottom: An example network of N neurons (small circles) with 1-d continuous attractor dynamics. Local excitatory and global inhibitory connections (not shown) between all neurons stabilize population states that are local activity bumps (e.g. blue bump A or B; gray: transient/unstable activity profiles). An activity bump is a single point on the continuous attractor (top) of all possible translations of the bump. If points on the attractor are identified with values of some circular variable, then all neural tuning curves for that variable will be identical, except for a phase shift (translation). **(b)** Column one: Recorded spikes (red dots) of two simultaneously recorded cells as a function of space (rat trajectory: gray lines). Column two: Autocorrelograms of the smoothed spatial response (peaks identified by black asterisks). Column three: A template lattice (red circles) is fit to all the peaks of the autocorrelogram. Parameters of the template (see **c**, inset) include the two primary axis lengths (λ_1, λ_2) and two angles (θ, ψ). Column four: Crosscorrelogram between the two cells (top), and the corresponding template fit (bottom). **(c)** Box plot of the ratio of each lattice parameter across 223 cell pairs (e.g., θ(cell i)/θ(cell j) where $i > j$) (median ratio: center line in box; interquartile ranges: box; lowest and highest values within $1.5\times$ of interquartile range: outer horizontal lines; 95% confidence interval based on 223 randomly chosen pairs not recorded simultaneously: dotted outer horizontal lines). **(d)** The distribution of relative phases (black circles) between all cell pairs, plotted within a canonical unit cell of the grid lattice. **(e)** Discharge maps (as in **b**) of the same cell pair, recorded again after an interval of > 60 minutes. **(f)** Box plot of parameter ratios (as in **c**) from this later trial, for the subset of cell pairs from **c** that were also recorded in this trial ($N = 84$ cell pairs).

Low-dimensional continuous attractor dynamics have been widely hypothesized to underlie the stable tuning curves of population codes [8, 56], motor control [83, 22], neural integration [83, 103, 9, 80, 91, 1, 32, 17, 38, 18], and parametric working memory [79, 65]. The predicted signatures of low-dimensional continuous attractors in the neural context are systematic differences in neural responses along the attractor manifold (e.g. preferred angles that vary along a continuum for a 1-d ring attractor, **Figure 2.1a**, bottom) but conformity and stability otherwise (e.g. tuning curves generated by pulling a rigid activity bump across the 1-d ring attractor). Critically, the mapping of states on the attractor to specific values of the represented variable may vary based on changing associations between the network and the external world (e.g. rotation of the network states relative to the world), but relationships between cells (e.g., whether they are in phase or counter-phase or quadrature-phase in their relative activity patterns), must remain absolutely stable [83, 103, 32, 18].

In other words, the responses of cells – when plotted against the external represented quantity – may change substantially, but pairwise relationships between cells should not. Despite these predictions, and beautiful empirical results [91, 1, 78, 2], definitive validation of the low-dimensional continuous attractor hypothesis has been somewhat elusive: In most cases, partly because of the difficulty of inducing sufficient change or perturbation in the neural responses and partly because quantitative analyses on simultaneously recorded neural pairs have not been conducted, it has been unclear whether the dynam-

ics are truly low-dimensional, what the dimension is, or where the dynamics originates [91, 78, 2].

Mammalian grid cells [41], each of which fires at the vertices of a regular spatial grid as the animal moves through its environment, are hypothesized [32, 17, 38, 18, 41, 72, 64, 21] to compute ongoing location estimates through integration of self-motion cues, based on the theoretical argument that their responses constitute a relatively context-independent code for spatial displacements. Across different familiar environments, the firing field locations in a grid cell change only through global phase shifts and rotations [41, 33], in contrast with place cells, which change in more elaborate ways, by gains or losses and shifts in subsets of their place fields [67, 77, 12, 61, 99]. The notable regularity and stability of a grid cell's response hints that the population might be well-described by relatively few parameters, a signature of an underlying low-dimensional dynamical system.

On the other hand, spatially periodic firing in grid cells is neither necessary nor sufficient for inferring low-dimensional population dynamics: It is theoretically possible for the population dynamics to be low-dimensional and periodic *without* spatially regular firing in individual cells, because of poor velocity integration [18]. Conversely, if the N cells in a single population have periodic spatial responses, but each displays independent shifts (relative to the other cells) of its spatial phase across environments, the dimensionality of the population response would be high, or $\sim N$. Finally, experiments that involve resizing of a familiar environment, or exploration in novel environments, reveal

that grid cell spatial responses stretch along one or both dimensions [7, 6, 87], a malleable response that is possibly consistent with higher-dimensional dynamics.

Here we examine spikes from simultaneously recorded grid cell pairs, in experiments where the single-cell responses undergo significant change, and where external inputs do not provide reliable spatial cues, to rigorously determine across conditions the dimensionality of the population response for each grid cell *network* or *module* (diiscrete networks or modules, consisting of local groups of cells with a common grid period and orientation, were predicted to exist through modeling [32, 18, 64, 29] and experimentally validated [7, 87]), and thus probe for evidence of low-dimensional continuous attractor dynamics in the brain. We relate the empirical findings to dynamical models of grid cells, to generate constraints on the mechanisms that underlie grid cell response.

2.2 Results

We examine several datasets of grid cell recordings in their entirety. The results reported below include all simultaneously recorded cell pairs from these datasets, in which both cells of the pair share a common spatial period and meet a modified gridness score that includes cells with regular triangular grids, even if the triangles are not equilateral (see Online Methods).

2.2.1 Identical spatial responses up to 2-d translation

We examine spikes from neurons recorded simultaneously from the same or nearby tetrodes. The activity peaks of a sample pair (**Figure 2.1b**) are arranged in the spatially periodic firing patterns characteristic of grid cells. Our definition of the spatial responses of grid cells, here and in the rest of this work, is the set of locations of the firing peaks. Six parameters are sufficient to characterize any periodic tiling in 2-d, regardless of the shape of the tiles [4]. Thus, the spatial response of an individual cell in a particular environment is well-described by four parameters for the angles and lengths of two primary lattice vectors (**Figure 2.1c**, inset), with two additional parameters that specify the 2-d spatial phase of the lattice, relative to some reference phase or location.

We find that cell pairs from the same or nearby tetrodes have extremely similar values for the first four parameters (**Figure 2.1c**, $N = 223$ cell pairs: 24 from [41]; 97 from [81]; 12 from [7]; 90 from [6]). This is the case even though the cells have very different spatial phases (**Figure 2.1d**), i.e., even when the cells are active in complementary parts of the environment. The *relative* phase between cell pairs, defined as the difference in their spatial phases, appears to be uniformly distributed ($N = 223$ cell pairs) over the unit cell of the lattice (**Figure 2.1d**; consistent with similar result from [41]).

2.2.2 Cell-cell relationships more stable than single cells

We next examine the stability over time of each cell's response and of cell-cell response relationships. Without any detailed analysis, the fact that a clear grid pattern is visible in the responses of individual neurons over a 20-minute recording session means that the individual spatial phases remain essentially constant over the session; if the phase shifted over time, the cell would fire at these different phases, and the grid response would be washed out. It follows directly that the relative phase between cells of the same spatial period and orientation will also be constant over that interval.

In this analysis, we probe the grid cell responses over longer time intervals: cells are recorded in an environment, then following an interval in which the animal is tested under varying conditions and environments, the cells are recorded again in the original environment (**Figure 2.1e**). The elapsed time between recordings in the original environment is > 60 minutes. We find that cells that shared essentially identical values of the first four grid parameters in the first measurement (**Figure 2.1b,c**) continue to share essentially identical parameter values in the subsequent measurement in the original environment (**Figure 2.1f**, $N = 84$ cell pairs from [7, 6]).

Moreover, the relative phases between cell pairs remain essentially constant over this interval (**Figure 2.2a**, $N = 84$ cell pairs), consistent with continuous attractor dynamics that stabilizes cell-cell relationships. However, this stability of relative phase between cells may be attributable to the stability of the phases of individual cells across visits to the original environment.

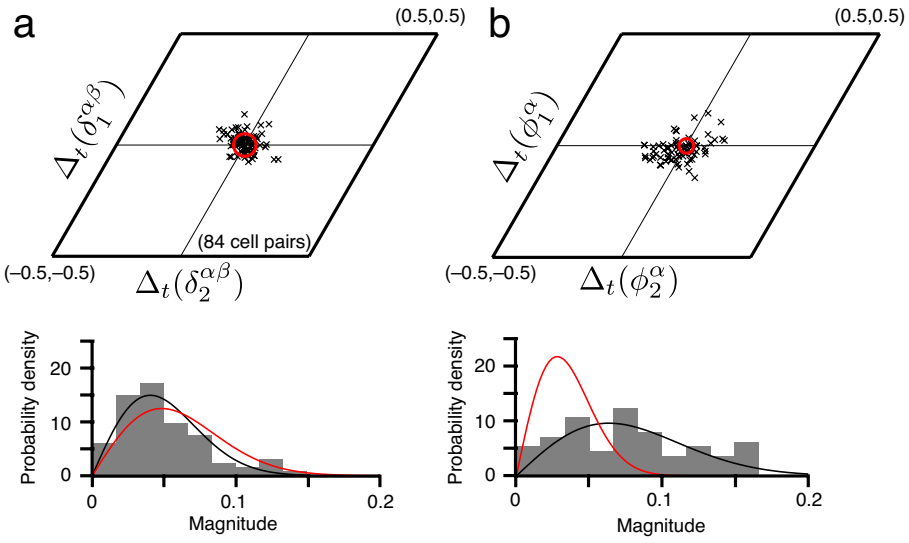


Figure 2.2: **(a)** Top: The difference across time (trials separated by > 60 mins) in the relative phase between cell pairs is clustered near zero (black x's, see Online Methods). Red circle: uncertainty in estimating relative phase differences (see Online Methods for error analysis). Bottom: Normalized histogram of the magnitudes of these relative phase differences (gray), with the *null* distribution (red), in which phase differences are not significantly different from zero and drawn independently from a Gaussian with standard deviation equal to the uncertainty in phase estimation. The null distribution of magnitudes is Rayleigh. Black: best-fit Rayleigh distribution to the data. **(b)** Difference across time (i.e., trials) of the phase of single cells (top), and the normalized histogram of magnitudes (bottom). Black, red defined similarly as in **a**. The data in **a** are not significantly different from the null hypothesis, while those in **b** are (**a**: $P = 0.58 \gg 0.05$, **b**: $P \ll 10^{-4}$ under the F-test for whether the data and the null distribution come from a distribution of the same variance). Finally, $P < 0.001$ under the F-test for whether the data in **a, b** (bottom) come from a distribution of the same variance.

To differentiate between the two possibilities, we therefore compared the stability in the spatial phase of single cells with the stability of relative phase between cells. We find that, notably, the relative phase between cells is more tightly preserved than the phase of each cell (**Figure 2.2b**), strongly suggesting a low-dimensional internal dynamics that yokes together the responses of different cells in the network, rather than spatially informative external cues.

2.2.3 Cell-cell relationships stable despite grid distortions

Next, we test pairwise relationships between grid cells that undergo a non-uniform rescaling of their individual grid responses when a familiar enclosure is suddenly resized [7] (**Figure 2.3a,b**). This rescaling constitutes a major change in the responses and grid parameters of individual cells (**Figure 2.3c**). But despite rescaling, the ratios of the first four grid parameters between cell pairs remain fixed very close to 1, indicating that these parameters change in tandem across the population (**Figure 2.3d**). This result holds for all cell pairs (**Figure 2.3e**, 7 cell pairs from [7]).

The relative phases between cells remain constant (**Figure 2.3f**) despite the changes in grid cell responses, and the constancy of relative phases is independent of whether the cells have similar or very different phases relative to each other. Thus we see that relative phase is strongly conserved even when the responses of single cells – influenced by changing external cues – have changed significantly, again strongly suggesting a 2-d state space and simultaneously suggesting that internal dynamics rather than external cues

are responsible for the low-dimensional response.

2.2.4 Constancy of cell-cell relationships without place-cell stability

The preceding results are consistent with a population response that is confined to a 2-d manifold: Given the spatial response of one cell, the responses of the others are always predictable, differing from the single cell only by a fixed 2-d phase shift, which remains invariant across conditions. However, it remains possible that, rather than arising from attractor dynamics generated within the grid cell network, the low-dimensional response of grid cells is externally imposed: by sensory cues from the familiar room, that somehow are flexible enough to permit rescaling of individual responses yet rigidly force fixed relative phases; or more plausibly, by inputs from the hippocampus.

In an attempt to address these possibilities, we analyze grid cell responses from animals' first exposures to novel environments. Grid cells are recorded first in a familiar environment, then in a novel environment, then again in several subsequent sessions in the novel environment as it becomes gradually more familiar [6]. In the first exposure to a novel environment, the spatial periods of grid cells expand suddenly, and the responses become less regular (**Figure 2.4a-c**, 24 cell pairs from [6]). As the novel environment becomes more familiar, significant changes occur: the responses become more grid-like (**Figure 2.4b**), and the grid periods contract steadily (**Figure 2.4c**). Indeed, the four parameters of the grid response all change in the novel en-

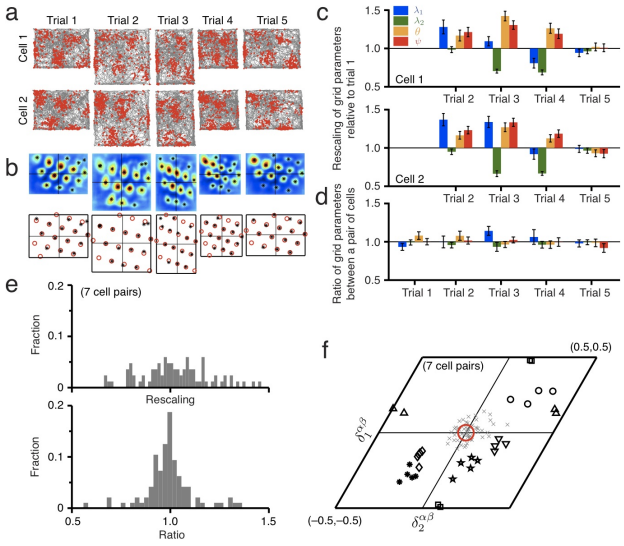


Figure 2.3: **(a)** Firing fields of two simultaneously recorded cells in a familiar environment (trials 1, 5) and resized versions of the familiar environments (trials 2, 3, and 4). **(b)** Spatial crosscorrelograms for the cell pair (top) and the best-fit template lattices (bottom). Asterisks denote local peaks in the crosscorrelogram. **(c)** Each grid parameter for cell 1 (top) or cell 2 (bottom) normalized by the value from trial 1. The parameters are substantially rescaled across trials 2-4. Error bars indicate ±1 s.d. (Online Methods). **(d)** The ratio, between cells 1 and 2, of each grid parameter, for each trial. The ratios are statistically very close to one, despite the significant rescaling in each cell, seen in **c**. Error bars indicate ±1 s.d. (Online Methods). **(e)** Top: Histogram of all grid parameters for the 11 cells in the resizing experiments from trials 2,3,4 normalized to the corresponding value from trial 1. Bottom: Histogram of the ratios of all grid parameter values between cells 1 and 2 for all 7 cell pairs from trials 1-4. This distribution is strongly peaked at 1 and different from the distribution at top. The *Kolmogorov-Smirnov* test for whether the two data samples come from the same distribution produces $P < 0.001$. The F-test for whether the two data samples come from a distribution of the same variance produces $P < 0.001$. **(f)** The relative phases for the 7 cell pairs span the unit cell (each black symbol represents a different cell pair; each marker for a given symbol represents a different trial). Gray x's: relative phase differences, computed across all cell pairs and trials. Red circle: uncertainty in the relative phase difference magnitude (Online Methods). The relative phase differences are not significantly different from zero ($P \approx 0.6 \gg 0.05$ for the same null hypothesis as in **Figure 2.2**).

vironment then relax back over days to values seen in familiar environments (**Figure 2.4d**), while the environmental sensory input remains unchanged. This suggests that the response of grid cells and their relationships are not determined, and by extension, not stabilized, by the environmental sensory input during this contraction period.

With this in mind, we next examine the relationships of grid parameters between cells. We find that the ratios of each grid parameter between cell pairs remain close to unity in the novel environment, as in familiar environments, starting with the very first exposure in the novel environment (**Figure 2.4e**), and continuing throughout the period of contraction of the grid period. Again, crucially, we find that the relative phase between cells remains essentially unchanged between the familiar environment and the very first exposure to a novel environment (**Figure 2.5**, 24 cell pairs from [6]), despite the large changes in the grid responses of neurons, and thus in their absolute phases. The relative phase between cells continues to remain fixed as the grid response shrinks over repeated exposures to the novel environment (**Figure 2.5**).

Because the animal in a novel space has not yet learned to associate external sensory cues with location [66, 86], and because external cues remain fixed while the grid responses shrink over several days, it follows that the relationship between grid cells and the external world is less stable than is the relationship between grids cells. It is therefore unlikely that external sensory cues are stabilizing cell-cell relationships across novel and familiar environments.

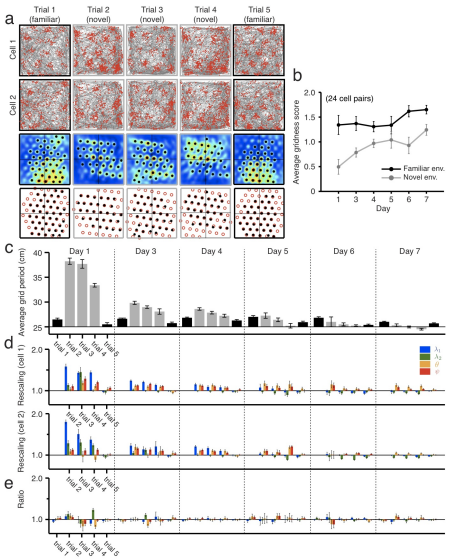


Figure 2.4: **(a)** Firing fields of pairs of simultaneously recorded cells in a familiar environment (black squares) and novel ones (gray squares) across five consecutive trials on one day, and the corresponding crosscorrelograms and best-fit template lattices. Note that on different days, the recordings involve different cells from the same tetrodes in the same area in the animal (**Supplementary Figure 2.4** for all cell pairs). **(b)** Development of average modified gridness in novel environments (gray) across seven days. The grid score gradually approaches that measured in familiar environments (black) (24 cell pairs, from [6]; means ± s.e.m.). **(c)** Change, across trials and days in the novel environment (gray), of the average grid period. Average grid period is the mean of the first two grid parameters across all cells in a trial (means ± s.e.m. 24 cell pairs total: 1, 6, 10, 3, 1, and 3 on days 1, 3, 4, 5, 6, and 7, respectively; no cells in day 2 passed the gridness criterion). The grid period significantly rescales in a novel environment, compared to when measured in the familiar environment (black), then gradually relaxes to its original value over seven days. **(d)** Grid parameters of one typical cell pair from each day (all cell pairs shown in **Supplementary Figure 2.4**), normalized by the corresponding parameter values from the first trial (familiar environment) of the day. Clusters of four narrow bars represent the four parameters, in the same ordering and color scheme as in **Figure 2.1** and **2.3**. Error bars indicate ±1 s.d. (Online Methods). **(e)** Grid parameter ratios for the two cells, across trials and days. Almost all these ratios are statistically indistinguishable from 1 (for all cell pairs, see **Supplementary Figure 2.4** and **2.5**). Error bars indicate ±1 s.d. (Online Methods).

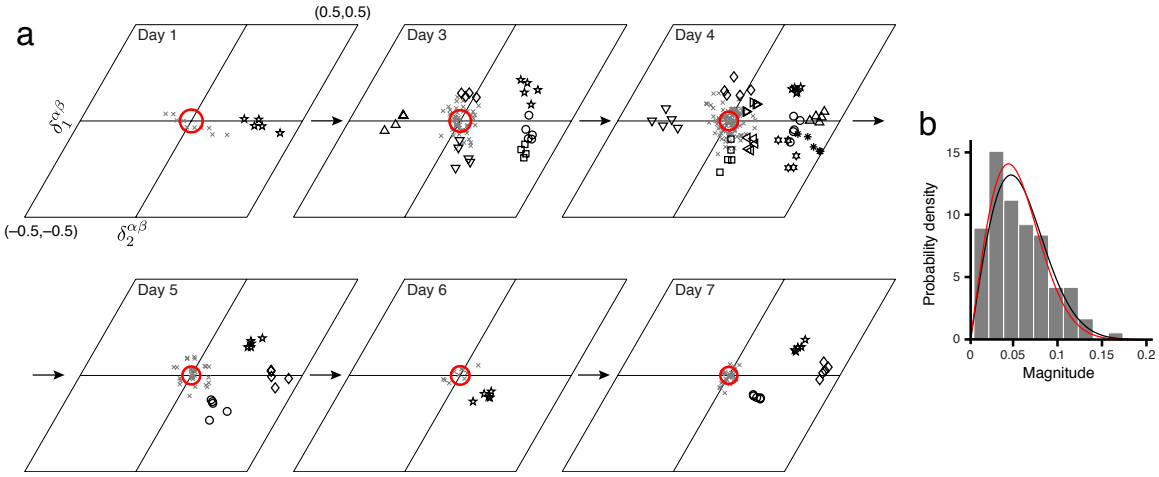


Figure 2.5: **(a)** The relative phases of the cell pairs (each distinct symbol represents a pair), across different trials and days ($N = 24$ cell pairs from [6], across all 7 days). Gray x's: the relative phase difference for every trial and all pairs. Red circle: uncertainty in the magnitude of relative phase differences (Online Methods). **(b)** The relative phase differences are not significantly different from zero for the same null hypothesis as in **Figure 2.2** ($P = 0.38 \gg 0.05$ under the F-test for whether the data and the null distribution come from a distribution of the same variance).

In contrast to grid cell responses in the novel environment, simultaneously recorded place cells underwent complete remapping [6] – defined as the loss of some of their firing fields and gain of others, with little preservation of spatial correlations between place cell firing fields. Like grid cells, place cell firing fields also expanded, then shrank, but to a much lesser extent than the grids, and for a shorter time. The remapped responses were not immediately stable, taking hours to stabilize [6] (see also [67, 30, 98, 50] for similar results on place cell remapping). Thus, hippocampal representations, and by extension, place cell-grid cell relationships, are in flux while cell-cell relationships between grid cells remain stable, suggesting that hippocampal input is not stabilizing relative phases between grid cells in the transition from familiar to novel environments, and within the novel environment. Taken together, these findings suggest that the hippocampus cannot be generating and feeding forward the 2-d stable responses observed in grid cells.

The above result is not inconsistent with the finding that hippocampal inputs seem required for grid cell activity [11]. A reduction of excitatory drive from the hippocampus and diminished activation of grid cells can, if the recurrent connections between grid cells are dominantly inhibitory [75, 23], result in a failure of the recurrent connections to induce population patterning and low-dimensional dynamics, as in the models of [18, 11]. Hippocampal inputs might also correct path integration errors by selecting the appropriate population state for a given location from a set of stable population patterns [21, 96, 85], thus enabling accurately patterned spatial grid cell responses over

a trajectory [18]. In either of these cases, the 2-dimensionality of the grid cell population response is intrinsic to the entorhinal cortex, but abolishing hippocampal drive may abolish spatial patterning.

We may conclude, first, that the grid cell population response is restricted to the same 2-d manifold at the first exposure to novel environments as in familiar environments, and second, that this restriction to the 2-d manifold cannot easily be ascribed to external sensory cues or hippocampal inputs, because relative phases and parameter ratios are stable even when these inputs are not.

2.2.5 Smoothness (continuity) of 2-d manifold

We next more closely examined the 2-d manifold of stable grid parameters and relative phases, to determine whether it exhibits a granular or "lumpy" structure, in which cell-cell stability is dependent on cell-cell similarity. A scenario in which cells with similar spatial phases conserved their relationships with each other more strongly, would be consistent with distinct subnetworks of cells with similar spatial response patterns stabilizing each other and not others with more dissimilar responses. Thus, we reexamined the results from familiar, rescaled, and novel environments, plotting parameter ratios as a function of relative phase between cell pairs. Parameter ratios were consistently close to one, independent of the distance in phase between cells (**Figure 2.6a**). Importantly, the *stability* in cell-cell relationships across rescaled and novel enclosure trials, as measured by stability of parameter ratios (**Figure 2.6b**) and

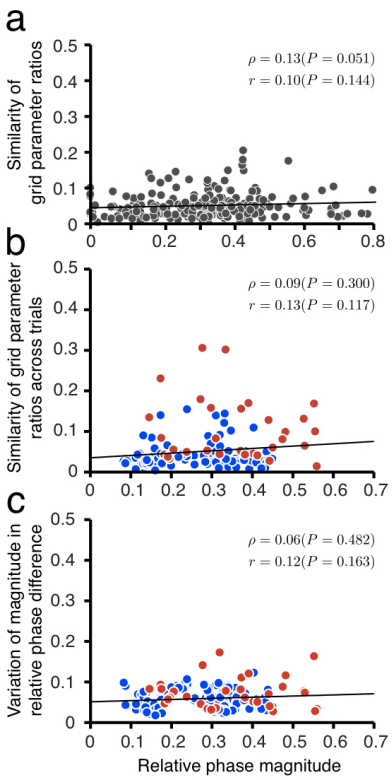


Figure 2.6: **(a)** Parameters between cell pairs (223 cell pairs from **Figure 2.1c,d**) in the same network are very similar (as reported in **Figure 2.1**), and moreover, the degree of similarity does not vary with the difference in spatial phase (i.e. magnitude of relative phase) between cells. (Parameter similarity is defined as the square-root of the squared deviation of parameter ratios from 1, averaged over all parameters per pair.) Each dot represents one trial from one cell pair. Black: linear regression; ρ: Spearman's rank correlation; r: Pearson's product-moment correlation. **(b)** The stability of parameter ratios between cell pairs across rescaling trials (red dots, 7 pairs from **Figure 2.3f**) and novel enclosure trials (blue dots, 24 pairs from **Figure 2.4** and **2.5**) is independent of the pair's relative phase. **(c)** The stability of relative phase (mean of magnitude of relative phase *differences*) across rescaling and novel enclosure trials is independent of relative phase between cells in a pair (same dataset and color-coding as in **b**).

stability of relative phases (**Figure 2.6c**), did not vary with relative phase magnitude between cell pairs. Therefore, up to experimental resolution, the 2-d manifold of states is continuous.

2.2.6 Stability (attractiveness) of the 2-d manifold of states

We have established that the grid cell population response is localized to the neighborhood of a 2-d manifold, over extended periods of time and across varying external conditions that induce significant changes in the grid cell responses. Further, this localization is very likely due to internal recurrent dynamics, not a low-dimensional external input. What can we say about the dynamical stability of the 2-d manifold?

Real-world dynamical systems are localized to their stable or attractive states. They are seldom (with vanishing probability) found at or near unstable fixed points, precisely because such points are unstable and the smallest perturbation will drive the system away. Conversely, the dimensionality of state space occupied by the system is indicative of the dimensionality of the attractive states of the system. If a low-dimensional manifold is stable but is part of a higher-dimensional manifold of stable fixed points (**Figure 2.7a**), then high-dimensional noise, however small, will cause the system to random-walk through the larger manifold [19]. The system will consequently be found to inhabit any of the states across all dimensions of the stable higher-dimensional manifold. Grid cells are likely subject to high-dimensional *internal* noise: For example, stochastic vesicle release in synapses [89] causes independent pertur-

bations in every postsynaptic neuron (noise dimension $\sim N$, where N is the number of neurons). An analysis of spiking variability as a source of noise in grid cells is given in the SI (**Supplementary Figure 2.3**). The fact that grid cells nevertheless primarily occupy the neighborhood of a 2-d manifold of states, in which cell-cell relationships are tightly conserved, suggests that the 2-d manifold is *attractive* (**Figure 2.7b**) and that locally, the attractive manifold has a dimension neither greater than nor smaller than 2.

Besides the deduction, above, that the grid cell system is subject to perturbations off the attractor through *internal* stochastic dynamics, we find evidence that *external* perturbations, in the form of velocity inputs, drive the grid cell system away from the 2-d manifold. To see this, we select all the spikes emitted by a cell during parts of the trajectory when the animal is headed 'northward' (north ± 45 deg), to form a *North* spatial activity map. We then compute the relative phase between a cell pair only for the North maps of each. The difference between this relative phase and the relative phase computed over the full trajectory is the perturbation off the attractor that northward motion induces in the state. All grid cells in our analysis have negligible directional tuning (**Supplementary Figure 2.2**), thus changes in relative phase cannot be attributed to a direct velocity modulation. We do the same for the South, East, and West conditions. The resulting directional shifts in relative phase are statistically significant, compared to controls computed from similarly sized trajectory fragments unsorted by heading direction (**Figure 2.7c,d**: data from the same cells in Figure 2.1c,d). The shifts in relative phase

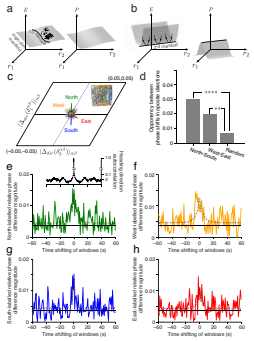


Figure 2.7: **(a-b)** Schematic energy landscape (left) and occupation probability (right) plots of a dynamical system. The independent variables on the plots depicting energy (left) and probability density (right) as heights are the firing rates of the neurons in the network. If the energy landscape has a flat plateau (**a**, left) of dimension $D > 2$, in which the 2-d manifold is embedded (depicted as a line), the system will likely be found off the 2-d manifold because there is no specific restoring drive back to it. (**a**, right). When there is a 2-d valley in energy (**b**, left), the system state will be localized to the 2-d manifold (**b**, right) even in the presence of noise. **(c)** Inset: animal trajectory from one trial color-coded (green, blue, red, yellow) by the instantaneous movement direction (North, South, East, West quadrants, respectively). Main plot, green vector: the difference in relative phase between cell pairs, computed as the relative phase obtained from spikes obtained only during Northward trajectory fragments minus the relative phase obtained from all spikes in the trajectory, averaged across all cell pairs (223 pairs). Gray vectors: Samples from the null hypothesis of randomly segmenting the full trajectory into four sets of fragments of the same average lengths as the directional fragments, without directional specificity. Black circle: one standard deviation of the null hypothesis distribution (Online Methods). p-values for the directional shifts in relative phase under null hypothesis: North ($P = 0.0004$), West ($P = 0.0444$), South ($P = 0.0003$), East ($P = 0.0201$). **(d)** Opposing shifts in relative phase from opposing trajectory directions: North-South ($****P < 10^{-4}$), West-East ($** P = 0.0014$) (Online Methods). **(e-h)** Green: Directional shifts induced in relative phase decay as the spike selection windows spanning Northward trajectory fragments are shifted in time, 1 second at a time, away from the centers of those directional segments, to include spikes emitted just before or after Northward movements. Black solid line: Radius of black circle from c. Gray dotted curve: Best-fit Laplace distribution with zero mean. Inset: Autocorrelation of movement direction in trajectories.

between cells are consistent with the respective heading directions (**Figure 2.7c**), suggesting that these shifts are indeed due to directional velocity input to the system. Thus, ongoing movements push the network away from the attractor by causing small deformations of the population pattern, in the form of a slight stretching of the population pattern along the direction of motion.

Finally, we examine the dynamics of perturbation by sliding the spike-selection windows in time relative to the centers of e.g. the Northward trajectory fragments (see Online Methods). We see that the shift in relative phase decays as the window is slid by a few seconds, and the decay time-constant is very similar to the autocorrelation time for directional motion in the animal trajectories. Thus, the off-manifold (perturbed) components of the network state relax back to the 2-d manifold on a time-scale similar to or faster than the few-second time-scale on which the perturbing input changes. The quick return of the system back to the 2-d manifold of states after perturbation from the external velocity input is direct evidence that the 2-d manifold is attractive.

2.3 Discussion

We have shown that, over short times and in familiar enclosures, the spatial responses of individual grid cells are well-characterized by a low-dimensional set of six parameters, with essentially the only difference between cells in the same network given by a 2-dimensional phase representing a rigid translation of the same basic response pattern. Over time and across environmental

manipulations, the responses of individual grid cells change, and thus the parameters that describe their responses vary. Therefore, the responses of an individual grid cell are not described by a single set of six parameters.

Crucially, however, the dimensionality of the population response remains invariant. The responses of different grid cells are yoked tightly together: Over time or with experimental manipulation, when the spatial response patterns of the individual neurons change significantly – not only through rigid rotations but anisotropic and isotropic deformations of the grid pattern through stretching – the grid parameter ratios and relative phases between simultaneously recorded neurons in each network or module remain essentially constant (Cells with distinct grid periods – i.e., cells from different networks – could never share a stable spatial phase relationship even if all single-neuron grid parameters were perfectly stable, because the relative spatial phase between two perfectly periodic patterns of different frequency will necessarily precess relative to each other across cycles). In familiar enclosures, where stability may be attributed to external cues, we show that the cell-cell relationships are *more* stable than single-neuron responses, which argues against the external cue hypothesis. Cell-cell relationships persist with the same fidelity immediately upon entering novel environments, even though landmark cues suddenly change and remain unassociated with specific locations, while place cell responses continue to change, arguing strongly for stabilizing constraints within the grid cell system.

Thus, population activity is confined to the immediate vicinity of a 2-d

manifold, across time and across conditions in different environments. Confinement of the system to a 2-d manifold despite stochastic internal dynamics and external velocity-driven perturbations off the manifold suggest that the 2-d manifold of states is an attractor.

2.3.1 Relationship to past work and implications for models.

A glimpse that the responses of different grid cells are yoked together was afforded by the data of [33] (not included in our present analysis): the spatial phases of $\sim 5-9$ simultaneously recorded grid cells shifted when the animal was moved from one environment to another, and the shifts appeared to be of a similar magnitude and direction across cells. However, both environments were familiar, so that hippocampal representations were stable (albeit different), and external cues could be used to provide locational information. Thus, without a comparative analysis of variability of phase within and between cells in a single environment, or an experiment involving destabilized hippocampal representations and external cues, the qualitatively different possibility of feedforward stabilization by hippocampal inputs or external cues could not have been ruled out.

Intracellular recordings in head-fixed animals navigating through virtual environments [27, 82] show that grid cell membrane potentials exhibit a substantial DC depolarization at the firing field locations. This is consistent with excitatory (or disinhibitory) network drive that depolarizes neurons for the duration of a field. However, given that the intracellular recordings can-

not distinguish between feedforward and feedback network inputs, and do not examine the structure and dimensionality of population activity, they do not provide specific evidence for continuous attractor models.

We have shown through population analysis that each grid cell network is localized to a 2-d manifold and that the manifold is attractive, which constitutes specific and direct evidence in support of continuous attractor network models of grid cell activity [32, 17, 38, 18, 64].

To be consistent with our findings, models of populations of grid cells need to include recurrent interactions that constrain the system to lie on a 2-d continuous attractor. The translation-invariant recurrent connectivity patterns of [8, 103, 32, 38, 18, 64] are examples of such an interaction. Within the constraint of 2-d continuous attractor dynamics, however, models of grid cell networks may be quite different: Some consist of a single recurrent network with a 2-d attractor, in which the grid cells integrate velocity inputs [32, 38, 18]. In others, the 2-d dynamics of grid cells arises from the feedforward summation of inputs, from two 1-d ring attractors, each of which integrates one component of animal velocity [95]. In future work, it will be interesting to distinguish experimentally between such alternatives.

To conclude, our analysis contributes strong new evidence (see also [91, 1, 2]) supporting the idea that the brain uses low-dimensional continuous attractor dynamics in its integration and memory functions.

2.4 Methods

2.4.1 Binning and rate maps

Cell-sorted spikes of putative grid cells from foraging rats were assigned to (1cm $\times$ 1cm) spatial bins derived from position samples taken at 50 Hz. The number of spikes assigned to each bin was divided by the animal's total dwell time in that bin, to remove the effects of inhomogeneous spatial exploration on estimating the probability of spiking at each location. This defines the *rate map*. Smoothed rate maps were generated by convolving the rate maps with a two-dimensional Gaussian kernel ($\sigma = 4$ bins).

2.4.2 Autocorrelations and crosscorrelations

To characterize the spatial response patterns of grid cells, we computed spatial autocorrelations from the smoothed rate maps of individual cells. To compare pairs of cells, we computed spatial crosscorrelations from the smoothed rate maps of simultaneously recorded cell pairs. If the smoothed rate maps are R_1 and R_2, both spatial correlations are generated as follows:

$$\gamma(u,v) = \frac{\sum_{x,y\in\Gamma}\left[R_1(x,y) - \overline{R}_{1,\Gamma}\right]\left[R_2(x-u,y-v) - \overline{R}_{2,\Gamma}\right]}{\sqrt{\sum_{x,y\in\Gamma}\left[R_1(x,y) - \overline{R}_{1,\Gamma}\right]^2 \sum_{x,y\in\Gamma}\left[R_2(x-u,y-v) - \overline{R}_{2,\Gamma}\right]^2}}$$

where $\gamma(u,v)$ is the correlation coefficient at the bin (u,v), Γ is region of spatial overlap between R_1 and R_2, and $\overline{R}_{i,\Gamma}$ is the mean of $R_i(x,y)$ within the region Γ. For autocorrelations, R_1 replaces R_2.

2.4.3 Modified gridness score

The standard gridness score [81, 55] penalizes any regular grid pattern if it is not of an equilateral triangular pattern **(Supplementary Figure 2.1a,b)**. A novel scoring procedure was proposed [13], with the aim of assigning high scores despite elliptical distortions in a hexagonal grid. However, it was limited by the fact that either the major or minor axis of the circumscribed ellipse was always assumed to pass through one of six nearest peaks. We used a modified gridness score that more generally allows both isotropic (equilateral triangle) and anisotropic (squeezed or stretched) grids to get a high score. For maximal sample size, and to test in principle the dimensionality of the grid cell response, it is important to include cells with anisotropic triangular grids: it is critical to determine whether deformed grid responses still lie on a low-dimensional manifold.

The modified gridness score is defined on the autocorrelogram of a cell. We first apply a transform on the correlogram that maps the central lattice cell (given by the six peaks nearest the center) into a regular hexagon **(Supplementary Figure 2.1c Left)**. This transform is determined by mapping the ellipse that circumscribes the central cell into a circle via a combined rotation, rescaling, and rotation transformation. If the number of nearest peaks (defined as the six or fewer peaks whose distance from the origin is less than 2 times the shortest distance from the origin to a neighboring peak) is less than or equal to four, the transformation is not applied.

Given the transformed autocorrelogram, we define an annular region

with inner radius R_i and outer radius R_o. We rotate the autocorrelation map in steps of 6° and compute the Pearson correlation between the rotated map and the original map with each confined to the annular region. The gridness for a given annular region is defined by the minimum difference between crests and troughs in rotated correlations:

$$\text{Gridness}(R_i, R_o) = \min\{\rho_{i,o}(60^\circ), \rho_{i,o}(120^\circ)\} - \max\{\rho_{i,o}(30^\circ), \rho_{i,o}(90^\circ), \rho_{i,o}(150^\circ)\}$$

where $\rho_{i,o}(\phi)$ is the correlation value when one map is rotated by angle ϕ relative to another over the annular region defined by R_i, R_o.

We do this for various values of R_i and R_o, letting R_i change from $0.5r$ to r and R_o change from $R_i + 1$cm to $1.5r$ (or to the maximum allowed value based on the autocorrelogram), each independently and in steps of 1cm. r is the mean distance to the nearest six peaks from the center in the transformed autocorrelogram. The modified gridness of the cell is then defined as the maximum gridness score over these various annular regions.

Our results are not qualitatively changed if we use the former gridness scoring technique [13]; they also do not qualitatively change if we use our technique with a higher threshold (=0.5).

2.4.4 Cell selection

In this paper, we analyzed data sets from four different sources [41, 81, 7, 6] (data [41, 81] available at http://www.ntnu.no/cbm/gridcell). A modified gridness score was computed for each grid cell, and cells with a gridness

score less than zero were rejected **(Supplementary Figure 2.2)**. (When we restricted our cell sample based on the gridness score more commonly used in the past, the sample become smaller but the qualitative results remained unchanged. Indeed, because the error in grid parameter estimation drops for more clean grid responses, the cell-cell relationships and ratios become slightly tighter.)

For pairwise analyses, we used all possible cell pairs that were simultaneously recorded in the same individual animal and shared a common grid period whose maximum difference between cells was smaller than 30% in the familiar enclosure (provided both members were above threshold on the modified gridness score). Cells above threshold that had no simultaneously recorded cells to pair with that also met the threshold for cell-cell comparisons, were also rejected from further analysis. Data from 223 cell pairs are shown in **Figure 2.1c,d**, from numerous experiments (24 from [41]; 97 from [81]; 12 from [7]; 90 from [6]). Data from 84 cell pairs (75 cells) are shown in **Figure 1f** and **2a,b**, from both resizing [7] and novel enclosure experiments [6], recorded in both trials 1 and 5. Data from the same cells in **Figure 2.1c** are shown in **Figure 2.7c-h**. Note that all grid cells from all datasets that met our gridness criterion happened to have extremely weak directional sensitivity (Watson U2 score < 10, by the scoring method of [81, 13]; **Supplementary Figure 2.2**). In **Figures 2.3-5**, cell pairs that have lower gridness than the fixed threshold in both familiar enclosures or in more than two intermediate trials were discarded, which results in 7 cell pairs in **Figure 2.3** and 24 cell

pairs in **Figure 4** and **5**.

2.4.5 Template matching algorithm to estimate grid parameters

We identified local maxima in the autocorrelogram and noted the coordinates and heights of the peaks. The local maxima whose heights were lower than 1% of the height of the global maximum were not considered as "local peaks". We then generated a 2-dimensional template lattice. Any 2-dimensional periodic lattice centered at the origin is fully specified by the magnitudes (λ_1, λ_2) and orientations $(\psi_1, \psi_2) = (\psi, \theta - \psi)$ of two basis vectors (**Figure 2.1c**). The angles are measured from the x axis. The template lattice is generated by populating the explored spatial environment by vertices whose locations are determined by the basis vectors and their translations. The lattice parameters are determined by finding values that minimize a cost function that quantifies the fit between the template and the data. The cost function is given by the sum of the squared distances from every data peak to the nearest vertex in the template lattice, weighted by the autocorrelation amplitude at that data peak:

$$C = \sum_{i=1}^{n} w_i \|p_i - v_i\|_2$$

where p_i is the (x, y) location of the data peak, v_i is the vector for the point in the lattice nearest to the ith data peak, w_i is the correlation coefficient at p_i, and n is the total number of peaks in the autocorrelogram.

The central peak of a crosscorrelogram is typically not at the origin, but

shifted by some displacement vector $\vec{d}$ from the origin. Thus, 2 additional parameters $\vec{d} = (d_x, d_y)$ were estimated simultaneously with (λ_1, λ_2) and (ψ_1, ψ_2) by minimizing the same cost function in order to find the best-fit template lattice of the crosscorrelogram.

2.4.6 Relative phase, relative phase difference, and phase magnitude

Let $\vec{\phi}^{\alpha}$ represent the phase of cell α, where the component ϕ_i^{α} is the phase along the ith lattice basis vector. The relative phase between cells α and β is then

$$\vec{\delta}^{\alpha\beta} \equiv (\vec{\phi}^{\alpha} - \vec{\phi}^{\beta}) \mod 1$$

where "mod 1" is understood to apply to each component. The relative phase between a pair of cells is closely related to the shift $\vec{d}$ in the peak of their spatial crosscorrelogram. If the two cells have precisely the same lattice parameters, then $\vec{d}$ will equal the relative shift of the two cells' grid patterns. The oblique projection of $\vec{d}$ onto the two primary lattice vectors $\vec{e}_1$ and $\vec{e}_2$ produces the components (d_1^{proj}, d_2^{proj}). When the components are normalized by λ_1 and λ_2, respectively, and considered modulo 1, we get the relative phase:

$$\begin{aligned} (\delta_1^{\alpha\beta}, \delta_2^{\alpha\beta}) &= (\phi_1^{\alpha} - \phi_1^{\beta}, \phi_2^{\alpha} - \phi_2^{\beta}) \mod 1 \\ &= (d_1^{proj}/\lambda_1, d_2^{proj}/\lambda_2) \mod 1 \end{aligned}$$

If the relative phase between the same two cells is measured in two different conditions (such as in distinct trials), we can define the relative phase

difference as the difference in $\vec{\delta}^{\alpha\beta}$ between the two conditions. We will denote this as $\Delta_C(\vec{\delta}^{\alpha\beta})$, where C labels the condition. Relative phase differences are equivalent modulo 1 and representing them on the unit lattice cell with components in the interval $[0, 1)$ has the consequence that values close to zero will appear in the four corners of the unit cell **(Supplementary Figure 2.6a-b)**. To avoid this and map values close to zero together, we remap the unit lattice cell to the equivalent unit cell with components in the interval $[-0.5, 0.5)$. Thus, the relative phase difference at distinct times t_1, t_2 is given by:

$$\begin{aligned} \Delta_t(\vec{\delta}^{\alpha\beta}) &= f\left[(\vec{\delta}^{\alpha\beta}_{t_1} - \vec{\delta}^{\alpha\beta}_{t_2}) \mod 1\right] \\ f(\vec{x}) &= \begin{cases} (x_1 - 1, x_2 - 1) & \text{if } x_1 \geq 0.5 \text{ and } x_2 \geq 0.5 \\ (x_1 - 1, x_2) & \text{if } x_1 \geq 0.5 \text{ and } x_2 < 0.5 \\ (x_1, x_2 - 1) & \text{if } x_1 < 0.5 \text{ and } x_2 \geq 0.5 \\ (x_1, x_2) & \text{if } x_1 < 0.5 \text{ and } x_2 < 0.5 \end{cases} \end{aligned}$$

where $\vec{\delta}^{\alpha\beta}_{t_1} = (\delta_1^{\alpha\beta}, \delta_2^{\alpha\beta})_{t_1}$ and $\vec{\delta}^{\alpha\beta}_{t_2} = (\delta_1^{\alpha\beta}, \delta_2^{\alpha\beta})_{t_2}$ are relative phases between cells α, β at time t_1 and t_2. The function $f(\vec{x})$ maps (x_1, x_2) within the interval $[0, 1)$ into $[-0.5, 0.5)$ by subtracting 0.5 from each component, if that component is greater than or equal to 0.5 **(Supplementary Figure 2.6c)**.

For relative phase, we take the relative *phase magnitude* to be of the form $||\vec{\delta}^{\alpha\beta}||^2 = (\delta_1^{\alpha\beta})^2 + (\delta_2^{\alpha\beta})^2 + \delta_1^{\alpha\beta}\delta_2^{\alpha\beta}$ and similarly for relative phase difference magnitudes.

2.4.7 Error analysis

The error bar (ϵ) in **Figures 2.3c,d** and **2.4d,e** for the ratio of two uncertain quantities λ_1/λ_2 is given by the standard method of error propagation given a covariance matrix Σ for the uncertainties in λ_1 and λ_2. Σ is estimated via bootstrap resampling: given an original spike (discharge) map of M total spike locations, we create 100 new spike maps of M total spikes each, by picking spike locations from the original map one at a time, at random, and with replacement, from the original map. Next, we use these spike maps to generate rate maps using the same procedure as for the original, with the original trajectory data (i.e. with a normalization given by the same visitation frequency as the original spike map) and estimate grid parameters from the spatial autocorrelogram with the template matching algorithm. This procedure generates 100 samples of the grid parameters, from which we compute the covariance matrix as an estimate for Σ. The error for the ratio $\frac{\lambda_1}{\lambda_2}$ is given by:

$$\frac{\lambda_1}{\lambda_2} \pm \epsilon = \frac{\lambda_1}{\lambda_2} \pm \frac{\lambda_1}{\lambda_2}\sqrt{\frac{\Sigma_{11}}{\lambda_1^2} + \frac{\Sigma_{22}}{\lambda_2^2} - 2\frac{\Sigma_{12}}{\lambda_1\lambda_2}}$$

The radius (r) of the red circles in **Figures 2.2a,b, 2.3f,** and **2.5**, signifies the measurement error of the differences across trials, in phase per cell (**Figure 2.2b**) or in relative phase per cell pair (**Figure 2.2a, 2.3f,** and **2.5**). This measurement error is estimated by applying the same bootstrapping technique to every cell (pair), computing now the (relative) phase difference as defined above for each bootstrap sample, and then subtracting the mean of

sampled (relative) phase differences for each cell (pair). The error in (relative) phase difference across (pairs of) cells is given by collapsing all the bootstrap samples of zero-mean (relative) phase differences and fitting a 2d Gaussian with a multiple of the identity matrix as a covariance to those samples (by expectation-maximization algorithm). The radius in the Figures is the square root of the estimated covariance.

2.4.8 Analysis of velocity-driven perturbation

Define cardinal direction labels as North ($45^\circ \sim 135^\circ$), West ($135^\circ \sim 225^\circ$), South ($225^\circ \sim 315^\circ$), or East ($315^\circ \sim 45^\circ$). Each time point t in the animal's trajectory is labeled by the animal's velocity at that time (given by the vector difference quotient between the position at $t + \Delta t$ and t). Spikes that occur at time t are labeled by the trajectory direction label at that time. This produces four sets of trajectories and for each cell, four corresponding sets of spike maps, labeled by North, South, East, and West. For each direction, we generate rate maps and relative phases (as we did earlier for the full trajectory and full spike maps). For each cell pair, we thus obtain four "direction-labeled relative phases", given by $\vec{\delta}^{\alpha\beta}_{dir} = (\vec{\phi}^{\alpha}_{dir} - \vec{\phi}^{\beta}_{dir})$ where $dir \in \{North, South, East, West\}$ and α, β refer to the cells. We denote the relative phase for the full spike maps, obtained earlier, as $(\vec{\delta}^{\alpha\beta}_{full})$. The "direction-labeled relative phase differences" for each cell pair are defined as the differences between the direction-labeled relative phase and the full relative phase (i.e. $\Delta_{dir}(\vec{\delta}^{\alpha\beta}) = \vec{\delta}^{\alpha\beta}_{dir} - \vec{\delta}^{\alpha\beta}_{full}$). The mean value of the direction-labeled rel-

ative phase difference, with the average taken over all simultaneously recorded cell pairs, is shown in **Figure 2.7c**. It is written as $\langle \Delta_{dir}(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta}$.

To assess whether the shifts in relative phase as a function of motion direction are meaningful, we create a null hypothesis distribution by segmenting the full trajectory into continuous pieces of a length consistent with the lengths of the continuous pieces generated in the direction-labeled trajectory segmentation described above (the fragment length was set equal to the correlation time of the animal's heading direction; a representative value of the heading direction correlation time across experiments is approximately 0.65 sec, and we chose 1.6 seconds to provide a window of at least two time constants). However, the segmentation did not correlate with movement along a specific direction. The segments resulting from this process were divided, randomly, into 4 sets of equal size. Consider one such set of directionally mixed or random segments, and label it "R", in contrast to the (North, South, East, West) labels of the directional trajectory segments. The relative phase difference for this one set of directionally mixed trajectory segments is denoted $\Delta_R(\vec{\delta}^{\alpha\beta}) = \vec{\delta}_R^{\alpha\beta} - \vec{\delta}_{full}^{\alpha\beta}$. Averaging this relative phase difference for one set of trajectory segments, across all simultaneously recorded cell pairs, gives one sample of $\langle \Delta_R(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta}$, which can be seen as one gray vector in **Figure 2.7c**. Repeating this procedure 400 times produce the 400 different gray vectors in **Figure 2.7c**. The lengths of these vectors represent the expected magnitude of deviation from the full phase simply due to subsampling errors from subdividing the full trajectory into 4 sets, independent of directional effects.

Thus, these vectors provide the null hypothesis for no directional motion effects on relative phase. We approximate the sampled distribution of vectors $\langle \Delta_R(\vec{\delta}^{\alpha\beta})\rangle_{\alpha\beta}$ by a symmetric 2-d Gaussian with standard deviation σ^2. Using this distribution, the p-value of the mean direction-labeled relative phase differences, $l = \langle \Delta_{dir}(\vec{\delta}^{\alpha\beta})\rangle_{\alpha\beta}$, is given by:

$$p = \int_{l^2}^{\infty} \exp(-\frac{r^2}{2\sigma^2})dr$$

In **Figure 2.7d**, we test the opponency of shifts in relative phase, for opposite movement directions, by computing the magnitude of the difference between opposing mean labeled relative phase differences, $\|\langle \Delta_{North}(\vec{\delta}^{\alpha\beta})\rangle_{\alpha\beta} - \langle \Delta_{South}(\vec{\delta}^{\alpha\beta})\rangle_{\alpha\beta}\|$ and $\|\langle \Delta_{West}(\vec{\delta}^{\alpha\beta})\rangle_{\alpha\beta} - \langle \Delta_{East}(\vec{\delta}^{\alpha\beta})\rangle_{\alpha\beta}\|$. The magnitude of opponency expected under the null hypothesis is given by the expected magnitude of the differences between the gray bars of **Figure 2.7c**, i.e., by averaging $\|\langle \Delta_R(\vec{\delta}^{\alpha\beta})\rangle^i_{\alpha\beta} - \langle \Delta_R(\vec{\delta}^{\alpha\beta})\rangle^j_{\alpha\beta}\|$, where i and j index values of $\langle \Delta_R(\vec{\delta}^{\alpha\beta})\rangle_{\alpha\beta}$ from the 400 sample shown in **Figure 2.7c**. This gives the height of the "random" bar in **Figure 2.7d**. The fraction of samples with $\|\langle \Delta_R(\vec{\delta}^{\alpha\beta})\rangle^i_{\alpha\beta} - \langle \Delta_R(\vec{\delta}^{\alpha\beta})\rangle^j_{\alpha\beta}\|$ that is larger in size than $\|\langle \Delta_{North}(\vec{\delta}^{\alpha\beta})\rangle_{\alpha\beta} - \langle \Delta_{South}(\vec{\delta}^{\alpha\beta})\rangle_{\alpha\beta}\|$, gives the p-value that the direction-labeled relative phase differences can be accounted for by the null hypothesis.

2.4.9 Analysis of relaxation from the perturbation

We tracked the mean shifts in relative phase for each cardinal direction as a function of time to examine the dynamics of perturbation off the attractor.

First, we used the Northward direction-labeled trajectory fragments to define a corresponding set of time-windows to select spikes for analyzing relative phase along that movement direction. Next, we slid the same set of time windows forward, so that instead of being centered in time at the Northward fragments, they were centered 1 second after the center (in time) of each Northward fragment, and so on, in steps of 1 second. We did the same in the opposite direction, sliding the windows back. For each position of the windows, we computed the relative phase for spikes obtained from those windows, and subtracted from it relative phase obtained from the full trajectory. For each time-shift, we averaged the result across all 223 cell pairs in our dataset.

2.4.10 Analysis of stochastic dynamics

To investigate the variance of spiking in grid cells, we compute the interspike interval (ISI) distribution using firing times $\{\tau_i\}$ that are rescaled from the original firing times $\{t_i\}$ via $\tau_i = \int_0^{t_i} \lambda(t)dt$, where $\lambda(t)$ is the time varying firing rate of the cell. $\lambda(t)$ was approximated by sliding a rectangular window function along the spike train. The coefficient of variation (CV) in **Supplementary Figure 2.3**, Left was derived from the rescaled ISI with $\Delta t = 0.5$ seconds. The CV as a function of window size for 4 representative cells is shown in **Supplementary Figure 2.3**, Right.

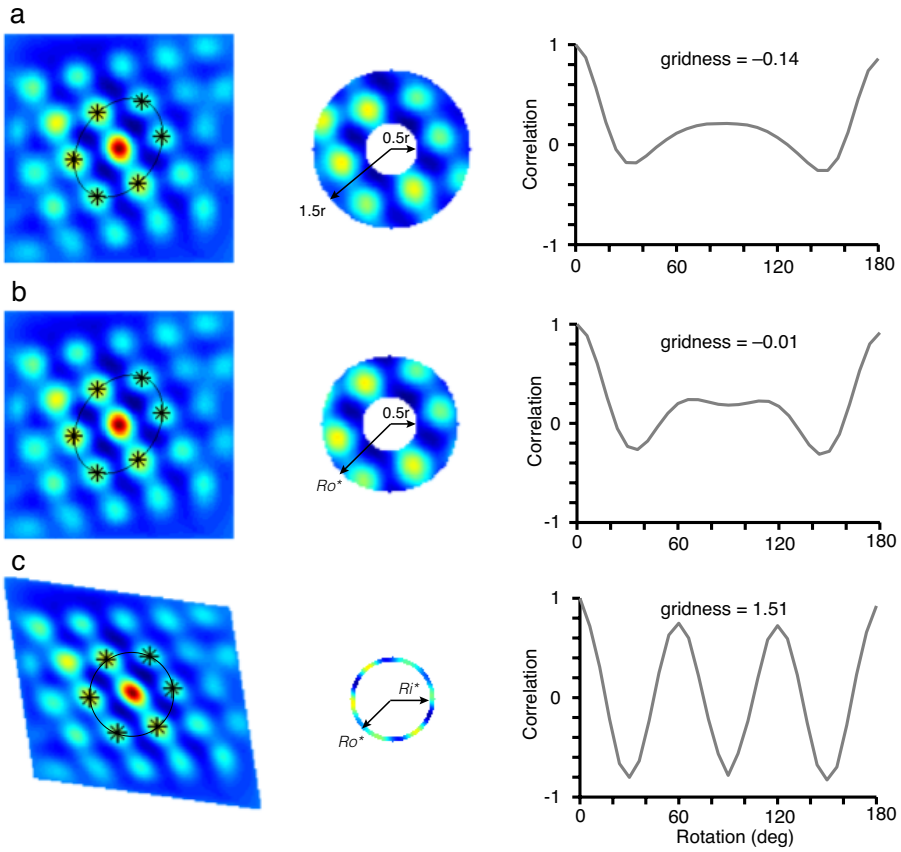
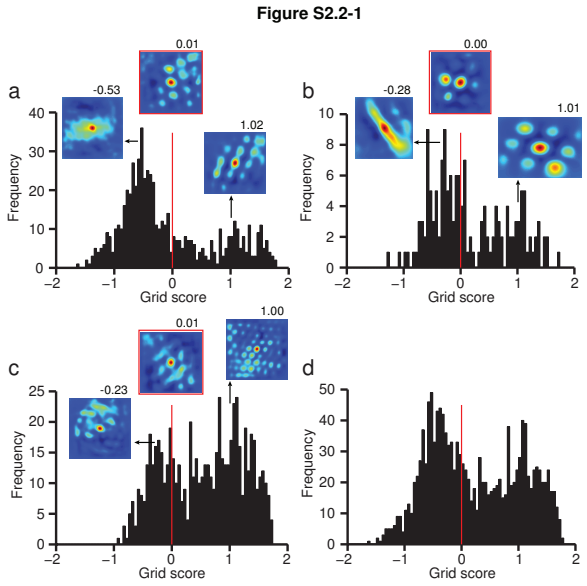


Figure S2.1: **Standard and modified (our method) gridness scores for anisotropic (non-equilateral) triangular grid cell responses. (a)** Left: A sample autocorrelogram is characterized by a non-equilateral triangular lattice. The nearest six local peaks (black asterisks) are circumscribed by the best-fit (black) ellipse. Middle: An annulus containing the nearest six local peaks ($R_i = 0.5r, R_o = 1.5r$; r is the average distance from the center to the six nearest peaks). Right: The Pearson correlation between the indicated annulus and the same annulus rotated by an angle in the interval $[0, \pi]$ as a function of angle. Gridness score is marked above the curve. **(b)** Left: the autocorrelogram (as in a). Middle: a choice of annulus with R_i fixed at $0.5r$ and where R_o is increased from $r + 1(cm)$ in steps of $1(cm)$. R_o^* is chosen as the value that produces the maximum gridness score ($= 0.01$). Right: As above. **(c)** Left: The transformed autocorrelogram with the six nearest peaks circumscribed by a circle. Middle: The annulus that resulted in the maximum gridness score ($=1.51$) over the set of annuli with $R_i \in [0.5r, r]$ and $R_o \in [r + 1, 1.5r]$. Right: As above.



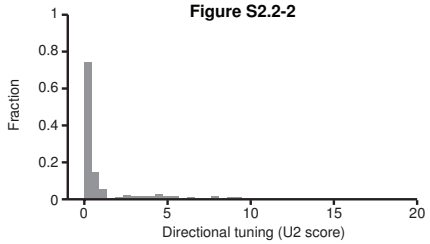


Figure S2.2-1: **Frequency distribution of modified gridness/U2 scores for the entire sample of MEC cells from the datasets analyzed in this paper.** **(a)** Cells from **Figs. 1, 2,** and **7**; **(b)** cells from **Figure 2.3**, and **(c)** cells from **Figs. 4** and **5**. The red line indicates our gridness threshold, which is equal to zero. Insets show spatial autocorrelograms of cells that represent non-grid cells (peak of the distribution on the left of the threshold), cells whose gridness score is close to the threshold, and cells above threshold (with a gridness score around 1.0). **(d)** The combined distribution of gridness scores for all previous samples. Note: Many cells in **a** were recorded from deep layers (III, V, VI) of MEC, and had grid scores lower than threshold.

Figure S2.2-2: Histogram of head-directional tuning for all cells in **Supplementary Figure 2.2-1a** that obtained higher grid score than the threshold. Directional tuning is expressed by Watson's U2 score.

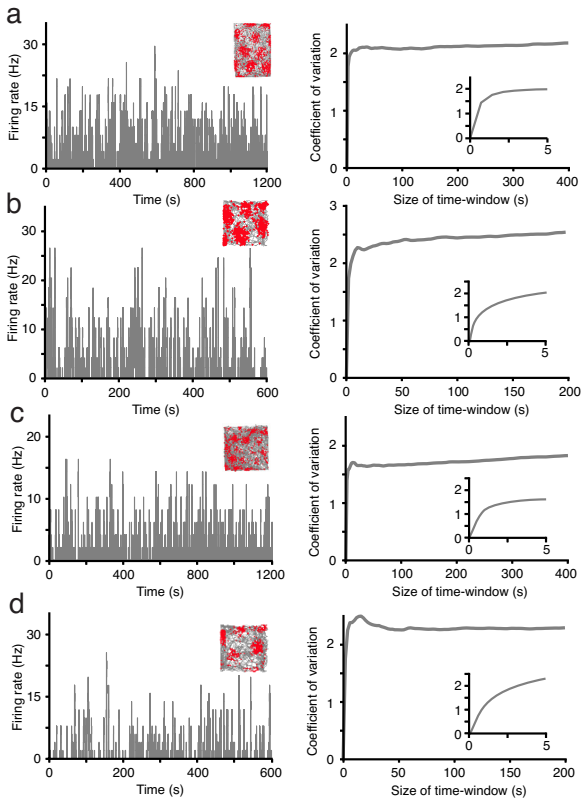


Figure S2.3: **Mean-corrected coefficient of variation (CV) analysis to assess the stochasticity of grid cell spike trains.** Left column: The time-varying firing rates of some example grid cells (N=4; inset), generated by sliding a window over each cell's spike train. The window is a boxcar (i.e., zero everwhere except for the interval $[t, t + \Delta t)$, where it is one; Δt is the window length and t is slid from $t = 0$ to $T - \Delta t$, where T is the length of the recording). Shown is the firing rate computed with $\Delta t = 0.5$ sec. Right column: Coefficient of variation (CV) for the interspike intervals, with spike times rescaled according to a windowed firing rate (as in Left), as a function of window size (see Online Methods). Inset: magnified representation of CV for windows lengths of 0 to 5 seconds. Note that the CV exceeds 1 for all cells for windows of length 600 ms and greater, and is considerable in size (exceeds 0.5) even for windows of length 300 ms. (ref 1: Softky, W. R. & Koch, C. The highly irregular firing of cortical cells in inconsistent with temporal integration of random EPSPs. *J. Neurosci.* **13**, 334350 (1993). ref 2: Shadlen, M. N. & Newsome, W. T. Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* **4**, 569-579 (1994)).

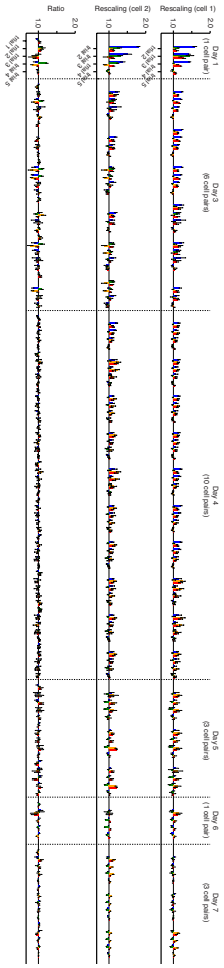


Figure S2.4: **A summary of all cell pairs in novel environments, analyzed as in Figure 2.4c-e.** Color scheme as in **Figs. 3c** and **4d,e**. Grid parameters for sets of cells (top: cell 1; middle: cell 2) normalized to the corresponding value on the first trial of the day (familiar environment), and ratio between the corresponding cells (aligned vertically) of each grid parameter for each trial (bottom) across all 24 cell pairs. Each day shows a set of parameters for cells recorded that day, similar to those in **Figure 2.3c,d** (e.g. day 3 has 6 cell pairs and six groups of parameters; the recorded cells on different days are different, but from the same area of one animal). A few trials are omitted due to the low gridness score of the cell's response. Note that "cell 1" and "cell 2" are not specific cells, i.e. each row does not represent the same cell across days; however, each column does represent the same cell *pair*.

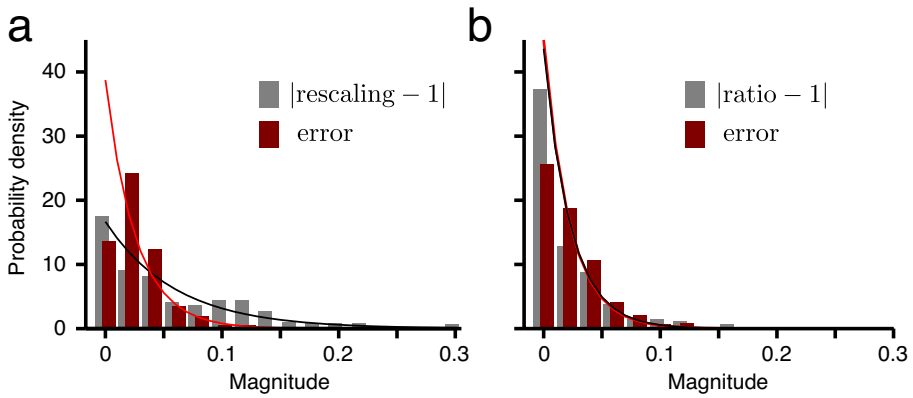


Figure S2.5: **Stable grid parameter ratios even when grids are rescaled. (a)** Histogram (normalized) of grid parameter rescalings (i.e., grid parameter ratios for each parameter and each cell, taken across time/trials) for all cells in **Figure 2.3f** (gray), with a normalized histogram of error magnitudes for each measurement (red). Parameter rescalings are plotted by subtracting one, and taking the absolute value, to compare against error magnitudes. Solid lines: best-fit exponential distributions ($\alpha \cdot e^{\alpha x}$). The width $1/\alpha_{data}$ of the parameter rescaling distribution is significantly larger than the width $1/\alpha_{null}$ of the error distribution ($\alpha_{data} < \alpha_{null}; P \ll 10^{-4}$, F-test for equality of variances of two samples). **(b)** Grid parameter ratios across time for all cell *pairs* (i.e., the ratio taken across time/trials of the ratio between cells of each parameter and each cell pair) from **Figure 2.3f**: here, the width of the two distributions is not significantly different: α_{data} is not significantly different from α_{null} ($P = 0.57 \gg 0.05$, F-test for equality of variances of two samples), showing that the grid parameter ratios between neighboring cell pairs remain significantly close to one.

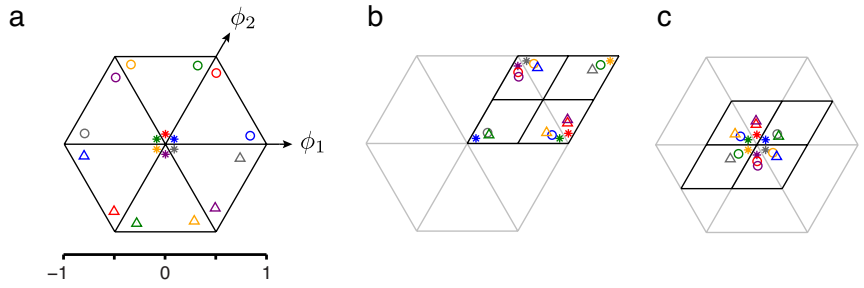


Figure S2.6: **Periodicity of the phase variable and how phase is computed.** Exemplar spatial phases adjacent to zero in **(a)** regular hexagonal unit cell, **(b)** rhomboidal unit cell mapped from **a** through mod 1, and **(c)** rhomboidal unit cell ranging from -0.5 to 0.5 by remapping phases in **b** to the origin. Each distinct symbol and color represents an identical spatial phase across transformed unit cells. Note that phases distant from the origin approximately by one grid period (circles and triangles) in **a** are mapped into the neighborhood of origin in **c**.

Chapter 3

Grid cell responses in 1D environments assessed as slices through a 2D lattice

Grid cells, defined by their striking periodic spatial responses in open 2D arenas, appear to respond differently on 1D tracks: The multiple response fields are not periodically arranged, peak amplitudes vary across fields, and the mean spacing between fields is larger than in 2D environments. We ask whether such 1D responses are consistent with the system's 2D dynamics, or whether they indicate that the system is in a different dynamical state. Combining analytical and numerical methods, we show that individual 1D responses are consistent with a linear slice through an internal 2D triangular lattice. Further, the 1D responses of comodular cells are well-described by *parallel* slices, and the offsets in the starting points of the 1D slices can predict the measured 2D relative spatial phase between the cells. These results show that the network remains in fundamentally the same dynamical regime, suggesting a common computation during both types of navigation behavior.

3.1 Introduction

Over the course of navigation in real environments, animals traverse open fields and run along paths defined by natural trails and boundaries such as underground burrow systems, streams, and rock faces. If grid cells demarcate spatial coordinates in 2D open fields, as seems possible, it is important to learn whether they perform a similar function during navigation along stereotyped 1D paths.

Arena experiments in the lab mimic open field exploration, while 1D track experiments approximate runs along trails. In arenas, individual grid cells of the mammalian entorhinal cortex exhibit spatially periodic tuning, in the form of increased firing at every vertex of a virtual triangular lattice overlaid on the floor of the enclosure [41] (Figure 3.1A-B). As a population, cells from nearby tetrodes have practically identical spatial firing patterns, or *spatial tuning curves*, up to a global phase shift (Figure 3.1C). The shift in phase of the spatial response between cell pairs, also known as the relative phase, remains stable across time and environmental conditions [33, 102].

Many grid cells also exhibit spatially specific responses when recorded on 1D tracks of various shapes [25, 101, 69, 39, 60]. On linear 1D tracks, the spatial tuning curves of grid cells consist of multiple firing fields with apparently unequal spacing, Figure 3.1D-E. The mean spacing between these firing fields is typically several times larger than in the same cell's 2D response. At present, there is little understanding of the detailed structure of these 1D responses. In particular, it is unclear whether there is a relationship between the

characteristic properties of 2D grid cell recordings – their hexagonal patterning, their similar tuning up to rigid shifts within a module – and the observed 1D responses. One wonders whether the cells are continuing to perform the same underlying computation when the animals navigate in environments of dimension different from two [51, 101, 69, 62, 45, 39, 35]. Here we seek to address this question and elucidate whether, and to what extent, the 1D grid response reflects an underlying pattern similar to that found in 2D.

In this paper, we consider the hypothesis that grid cell responses on linear 1D tracks are generated by slicing linearly through an underlying 2D triangular lattice [27, 51]. We begin by considering what this *slice hypothesis* predicts for the Fourier transform and power spectrum of the 1D spatial responses of cells. We present analytical methods and numerical refinements for extracting the parameters of the hypothetical slice of an underlying 2D grid given the power spectrum of the recorded 1D response, and show that the resulting slices can result in excellent fits to the 1D spatial responses. We show that groups of cells putatively from the same module [87] and recorded simultaneously in 1D have substantially similar power spectra, despite the apparent dissimilarity of their spatial responses, and are well-fit by parallel slices through a triangular lattice. The predicted 2D phase offsets between cells, based purely on the 1D responses, closely match the recorded 2D phase offsets for these cells. Finally, while the spatial tuning of the 1D response can gradually reshape over the course of an experiment, this remapping can be interpreted as a drift in the slice parameters, and slice drift appears to

occur in tandem across cells from the same putative network, preserving cell-cell relationships. We conclude that the data strongly support the hypothesis that 1D responses can be generated by slicing through a regular underlying 2D triangular lattice. Moreover, the group properties of cells from the same module in 1D environments are entirely consistent with the low-dimensional group properties of the same cells during 2D exploration, indicating that the system remains in the same dynamical regime during navigation in 2D and 1D environments.

3.2 Results

3.2.1 The Fourier power spectral density reveals underlying periodic structure

We find that the 1D tuning curves of simultaneously recorded grid cells – here recorded on 1D virtual tracks [27] – while not periodic and not necessarily related to each other by simple shifts (Figures 3.1A-F), nevertheless exhibit a shared "fingerprint" that hints of an underlying periodic structure. To see this, first consider the 2D responses of these cells, in which this fingerprint — the *power spectral density* (PSD) of the spatial response — is easy to understand and interpret (Figure 3.1G). The periodic 2D response of a grid cell generates six discrete peaks forming a hexagon around the origin in the PSD (these peaks form part of a triangular lattice in frequency space that is "dual" to the spatial triangular lattice, see Figure S3.1). The loss of phase information that results when squaring the Fourier spectrum to obtain the PSD

means that neighboring cells, which share common grid response parameters up to phase offsets [102], will exhibit the same spectra – all the spectral peak locations are matched (Figure 3.1G). Intriguingly, the PSDs of the 1D spatial responses of grid cells putatively from the same module also appear similar (Figure 3.1H). We next explore this phenomenon and its possible implications for how 1D grid cell responses are generated.

The PSD of a generic linear slice through an idealized 2D triangular lattice has exactly three dominant peaks of equal height (Figures 3.2A-2C). As the angle of the slice varies, the locations of these three peaks shift (Figures 3.2A-2C). These peak locations are independent of the origin (or spatial phase) of the linear slice for generic slices, thus parallel 1D lattice slices have nearly identical PSDs (Figures 3.2D-2F). Non-generic slices are those parallel to one of the primary lattice vectors (Figure 3.2B) or at the half-lattice angle of 30°; in these cases the PSD is predicted to exhibit only one or two major peaks, respectively (Figure 3.2C). In short, linear slices through a 2D lattice should exhibit 1, 2, or (most commonly) 3 peaks in their PSDs. We will quantify the degree to which the mass of a PSD is concentrated in 3 or fewer peaks by a *three-peakiness score* (Experimental Procedures).

This frequency-space characterization provides a concise description of the particular heights, spacing, and ordering of firing fields in a lattice slice, and of the constrained peak structure of the corresponding PSD. To illustrate the specific nature of these constraints, we compare lattice slices to two sets of statistically matched random controls: Given a 1D response, we generate

corresponding *gap-randomized controls* by randomly shuffling the order of the firing fields then assigning the fields new positions chosen randomly from a uniform distribution (Figure S3.2 and Experimental Procedures). We generate *gap-shuffled controls* by simply shuffling the gaps between fields (Figure S3.2 and Experimental Procedures). Examples of these two random controls and their PSDs can be seen in Figure 3.2B (reddish-gray and bluish-gray, respectively), with PSDs in Figure 3.2C. The tuning curves for both random controls can look very similar to slices, but generically their PSDs differ: they have variable numbers of PSD peaks, and usually more than three (Figures 3.2C and S3.3; three-peakiness score p_3=0.63 for the gap-randomized and p_3=0.69 for the gap-shuffled controls, compared to $p_3 = 1$ for the lattice slices, Experimental Procedures). We will return to these controls to statistically test our conclusions throughout this work.

It is useful to note that the 1D PSD of a 1D slice through a 2D lattice is equivalent to a 1D slice taken at the same angle starting at the origin of the 2D PSD of the 2D lattice (Figures 3.2G-2H and Experimental Procedures). This equivalence allows us to predict the PSD of a 1D slice at any angle by simply projecting the three closest-to-the-origin spectral peaks from the 2D PSD onto a line running through the origin at the desired angle (Figures 3.2G-2H). Conversely, given the locations of two major PSD peaks from the 1D response, we can solve analytically for the angle of the slice (Figure S3.4 and Experimental Procedures). The slice angle obtained in this way is exact if the 1D slice extends to infinity.

Using similar techniques (Figure S3.4 and Experimental Procedures), we can infer the *scale factor* of a cell, which is the ratio of the inferred 2D lattice period underlying the 1D response to the recorded period in 2D enclosures. From the Fourier transform of the 1D response, we can furthermore deduce the starting point or 2D spatial phase of the slice (Figure 3.2D) (Experimental Procedures). By taking the difference of the inferred 2D spatial phases for a pair of cells, we can predict their 2D relative phase. We can alternatively, and more directly, estimate the 2D relative phase for a cell pair through the Fourier transform of the cross-correlation of their 1D spatial responses (Experimental Procedures). These four parameters, slice angle (θ), scale factor (α), and 2D phase ($\vec{\phi}$), fully describe a *semi-infinite* linear slice that starts at the given phase within a unit cell in the canonical 2D grid and continues out to infinity at the given angle (like the schematic slices in Figure 3.2D). Once the scale factor has been determined, the length of the slice through the 2D grid is fully determined by the scale factor and the length of the track run by the animal.

Finite size effects induce uncertainty in the inferred slice parameters (Figures S5-S6), as does noise in the neural response (Figures S5 and S7). In the presence of such uncertainty, supplementing the analytical estimates with numerical optimization can result in improved parameter estimates (Figures S5-S8 and Experimental Procedures).

In summary, the mathematical duality between the spatial response domain and the Fourier spectral domain for slices through regular lattices allows us to determine, up to estimation uncertainty, whether the 1D spatial

responses are consistent with an underlying 2D lattice and if so, to infer the parameters of the slice relative to the lattice.

3.2.2 The distribution of gaps in a 1D response reveals underlying periodic structure

A second signature of lattice slices, complementary to the 1D PSD, is the distribution of the gaps between 1D firing fields. A linear slice through a 2D lattice cuts across several bumps, with a distinct sequence of gaps (spacing between bumps) depending on its angle (Figure 3.3A-B). The smallest gap occurs when the slice hits adjacent bumps in the 2D lattice; a larger gap occurs when the slice misses a 2D bump and hits the next-nearest bump, and so on. Thus, the set of possible gaps is restricted to the set of distances between a reference bump and all other bumps in the 2D lattice that can be reached by a straight line that does not cross any other bumps. By contrast, the gaps in gap-randomized controls (Figure 3.3C) can be of arbitrary size; thus, the distribution of gaps is unimodal and smooth (Figure 3.3D, bottom), rather than multimodal with a few discrete peaks as predicted for slices (Figure 3.3D, top). (The gap-shuffled controls preserve the full gap distribution of the original data by construction.)

Up to an overall scale factor to account for the lattice period, the gaps in lattice slices are entirely specified by the geometry of the 2D lattice: the relative lengths of lines connecting different vertices in the lattice. Thus, the ratios between gaps will exhibit relationships characteristic of an underlying

triangular lattice: $1/\sqrt{7} \approx 0.38; 1/\sqrt{3} \approx 0.58; \sqrt{3/7} \approx 0.65$; and so on (Figure 3.3E), and will be the same across modules with different spatial tuning periods and orientations, and different slice angles, so long as the underlying grid geometry is the same.

The distribution of gaps collected by pooling across 100 slices of an idealized triangular lattice with varying angle and starting position (phase) is shown in Figure 3.3E (bottom; the lattice period is $a = 20$ cm). The gaps (Figure 3.3D, top) from the two slices of Figures 3.3A-3B are simply samples from this gap distribution (Figure 3.3E), and a particular slice may sample only a subset of possible gaps from the lattice. To take an extreme example, a slice along one of the primary lattice vectors samples only the nearest-neighbor bump distances in the 2D lattice, thus the gap distribution consists of a single peak; Figure S3.9).

It is possible to extract gaps on a trial-by-trial basis from neural recordings then pool the gap samples across trials. Even if the 1D firing field locations drift from trial to trial (Figure S3.10), the gaps in every trial will be drawn from the same small set governed by the 2D lattice geometry, assuming that all trials are linear (but possibly changing) slices through that lattice. Moreover, because gap ratios are independent of grid orientation, grid period and scale factor, it should be possible to pool gap ratios across modules. By contrast, PSDs depend on slice angle and cannot be similarly pooled. On the other hand, the gap analysis focuses on adjacent bumps, while the PSD contains information about relationships between all bumps, not just adjacent ones.

Thus, gap analysis and PSDs provide complementary measures of a slice.

3.2.3 Grid cell responses on 1D virtual tracks are well-fit by linear slices through a triangular lattice

We applied the methods described above to the 1D virtual track responses of 25 neurons identified as well-isolated grid cells with stable 1D spatial responses from the experimental dataset (see Experimental Procedures). Examples of the 1D response are shown in Figure 3.4A (three cells from different experiments), with 30-trial firing rate averages in Figure 3.4B (black curves). The PSDs of the spatial tuning curves in Figure 3.4B (Figure 3.4C, black) exhibit three major peaks ($p_3 = 0.95, 0.99$, and 0.66, respectively), as expected for lattice slices.

Our Fourier-based slice analysis method (Experimental Procedures) returned slice parameters (inset values, Figures 3.4B) that produced the best fit (in the sense of Pearson's correlation coefficient) to the measured 1D spatial tuning curve. For each neuron, the slice angle was confined (without loss of generality) to the interval $[0°, 30°]$, the phase could fall anywhere within a predefined unit cell of the lattice, and the scale factor could take values between 1 and 8 (Experimental Procedures).

The 1D tuning curves predicted by the slice hypothesis (Figure 3.4B, green) provide excellent matches to the recorded 1D spatial tuning curves (average correlation value of 0.92 and average p-value of 0.04 for these three cells). The best-slice PSDs, shown superimposed on the data PSDs (Figure

3.4C, green), are also very close to the data PSDs. The gap histogram obtained from pooling gaps between adjacent bursts across individual trials (Figure 3.4D; see Experimental Procedures) displays a small set of distinct peaks. Moreover, most gaps and gap ratios (inset numerical values, Figure 3.4D) are strikingly similar to (within a few percent of) the values predicted for a regular triangular lattice (Figure 3.3E). By contrast, the predicted gap distribution for gap-randomized controls is a smooth, broad unimodal distribution (Figure 3.3D, bottom).

We augment our virtual track recordings with data obtained by Brun and colleagues on a long linear 1D track ([15], data shared by the Moser laboratory through www.ntnu.edu/kavli/research/grid-cell-data and personal exchange). The track is punctuated by discrete visuospatial landmarks like doorways, and the spatial response statistics of a majority of cells change visibly roughly halfway into the track and persist to the track's end in most of the data. For this reason, we analyze responses on the first half of the linear track (Figure 3.4E).

Applying the same analysis as in Figures 3.4B-D to two cells from [15] yields striking agreement between their track responses and the linear slice hypothesis (average slice correlation value of 0.89, and average p-value of 5×10^{-4} for the cell pair; the p-values, computed from the correlation coefficient and the number of above-threshold firing fields, are even more significant than for our virtual track data because of the greater track length and consequently larger number of fields), Figure 3.4F (green).

The longer track is also an opportunity to perform cross-validated prediction of 1D spatial tuning in grid cells: even after truncation of half the track, the remaining ≈ 8 meters is nevertheless roughly twice as long as the virtual tracks. When we fit the spatial response from the first four meters of the long track to a lattice slice, that slice provided a very good prediction of spatial tuning on the withheld four meters (Figure 3.4F, cyan). When the analytical slice prediction on the withheld data was good, then numerical local optimization of that solution from the first four meters of track produced even better prediction, indicating that the optimization step yields more accurate descriptions of the underlying slice.

3.2.4 Statistical analysis of slice hypothesis across virtual track data

As we have seen, the 1D responses of some grid cells are extremely well-fit by the slice hypothesis: the average fit correlation coefficients of the cells in Figure 3.4A are high ($\langle \rho \rangle = 0.92$), the PSDs are largely concentrated in three peaks ($\langle p_3 \rangle = 0.87$), and the gap distribution has peaks close to the predicted values for slices of triangular lattices.

To what extent is it generally true that the 1D responses of grid cells in our dataset can be well-described by slices? To answer this question, we focus closely on dataset-wide statistical analyses of cells recorded on the virtual track, because there we possess 2D recordings, from which we can extract ground-truth information on whether the cells are grid cells, on their 2D peri-

ods, and their co-modularity.

We generate the best-fit slice for each cell in the dataset (e.g., Figure 3.5A, top row, green; fit in black), together with large samples of corresponding gap-randomized and gap-shuffled controls (one sample of each shown in Figure 3.5A, middle row, reddish-gray and bluish-gray, respectively) and their best-fit slices (green). We can now compare metrics of the cell against the metrics of its two random controls, Figure 3.5A (bottom row). Typically, the correlation coefficients between the data and a slice are well to the right of those for the two matched control distributions; the same is true for the three-peakiness scores (Figure 3.5A, bottom row). We standardize the correlation score and PSD three-peakiness score of each cell (by subtracting the mean and dividing by the standard deviation of the distributions of these scores from matched controls, illustrated in Figure 3.5A, bottom row), so that the standardized scores reflect how many standard deviations away the data lies from the control distributions. Standardization allows us to collapse together results across all grid cells in the dataset. The resulting quality-of-fit distributions for the dataset are significantly to the right of both matched controls (Figure 3.5B; testing to see whether the data are drawn from a distribution with zero mean, as in the control distributions, yields $p = 3 \times 10^{-4}$; $t = 4.28$; $df = 24$ and 2×10^{-4}; $t = 4.49$; $df = 24$, respectively, on a one-sample t-test), as are the three-peakiness scores (Figure 3.5C; $p = 1 \times 10^{-3}$; $t = 3.65$; $df = 24$ and 8×10^{-4}; $t = 3.85$; $df = 24$ respectively; one-sample t-test). (Note that the controls are extremely conservative (not overly random), in that they preserve

many higher-order statistics of the data; this is particularly true for non-generic slice angles where the distribution of gaps is very narrow and thus the random controls closely resemble the original slice (see Figure S3.3); nevertheless, the slice hypothesis fits the data significantly better than it does the controls.)

Finally, the histogram of gaps, normalized by the inferred period (scale factor times 2D grid period) for each cell and pooled across all cells in the dataset, yields a distribution with a small set of distinct peaks (Figure 3.5D). The peaks in the gap distribution lie very close to the predicted values for an underlying regular triangular lattice (see Figure 3.3E). The slight, systematic leftward shift in all the peaks relative to the prediction can be attributed to the fact that each firing field has a finite width, and so the measured spacing from field to field will be reduced by an amount proportional to the field width.

Lacking access to 2D data for cells recorded on 1D real tracks [15], we instead identified putative grid cells from the 1D responses alone by applying a modified version of criteria used in [27] (Experimental Procedures). An analysis of all the cells that passed these criteria (40 out of 97 cells, or 41 percent) yields excellent agreement with results reported here: The 1D responses of putative grid cells are well-fit by the slice hypothesis (the average fit-quality on these responses is 0.72, close to the mean for the virtual track data in Figure 3.6) and are better-fit by it than are the statistically matched controls (Figure S3.11A-C); they exhibit a gap distribution consistent with slicing through a regular triangular lattice (Figure S3.11D); and the slice parameters tend to cluster into groups, indicating the presence of comodular cells with similar

slice parameters (Figure S3.11E).

3.2.5 Cell-cell relationships: pairs from the same putative module are well-fit by parallel slices

We next examine relationships in the 1D slice parameters of *comodular* cells, defined as simultaneously recorded cells deemed to be from the same module based on the similarity of their 2D responses. If the 1D responses of different cells from a module were generated by the coherent displacement (driven by animal motion) of an underlying 2D population pattern, we would expect their 1D tuning to correspond to parallel slices of the 2D population pattern with the same scale factor. As a consequence, their PSDs should also look similar (prediction in Figures 3.2D-F). Indeed, the scale factors, slice angles, and PSDs of comodular cells can be very similar, as we can see for an example cell pair on the virtual linear track and another pair on a long linear track (Figure 3.6A-D).

However, the best-fit slice parameters obtained individually for each cell in the dataset are not always very similar for comodular cells (Figure 3.6E, left). One reason for this discrepancy is the existence of a degeneracy in the solution for the best-fit slice of each cell. Because of response variability and the finite length of the track, there are sometimes two or three specific, discrete good slice solutions for a cell (see Experimental Procedures). These solutions are not near each other in parameter space because the relationship between fit quality and slice parameters is non-convex – they tend to take the

form of well-separated, similarly deep local optima in the solution landscape and correspond to a large step in slice angle and scale factor. Thus, there is an inherent uncertainty in estimating the best slice parameters and even intrinsically parallel slices may not generate similar parameter estimates if we choose different minima for each. In summary, there can be a many-to-one mapping from slice parameters to fit quality, and we may in some cases be picking the "wrong" slice for some cells when comparing across cell pairs.

The question therefore is whether there exist parallel slice solutions for comodular cells that also yield good fits. We refit cells with the same fitting procedure as above, but after a preliminary step that enforces consistency in which PSD peaks are selected across comodular cells to predict their slice parameters (Experimental Procedures and Figure S3.12). Not too surprisingly, the slice parameters for comodular cells are now well-clustered (Figure 3.6E, right). However, imposing constraints on a fit will generally worsen the quality of the fit, especially if the constraints are not consistent with structure in the data; in our case, we are forcing consistency across sets of cells instead of finding the globally best-fit solution for each cell as before, thus our fits are guaranteed to be worse. Indeed, there is a change in the mean of the quality-of-fit distribution when cell fits are individually optimized (Figure 3.6F, gray histogram) versus when they are fit consistently (Figure 3.6F, white histogram). Notably, however, the drop in quality of fit is only slight (the difference in means between the individually optimized and constrained fit-quality distributions is not significant, $p = 0.2$; t=1.30; df=48;

paired t-test) and imposing consistency still produces slice fits that are much better than in the controls ($p = 2 \times 10^{-3}$; $t = 3.72$; $df = 24$ for average fit quality after imposing consistency versus matched gap-randomized controls and $p = 1 \times 10^{-3}$; $t = 3.68$; $df = 24$ versus gap-shuffled controls; one-sample t-test). We conclude that the constraint is consistent with the underlying structure of the data: that is, comodular cells in 2D are well-described by parallel grid slices through a common 2D grid. We will see below that a parallel slice description further leads to accurate prediction from the 1D track data of the 2D spatial tuning phase relationships between comodular cells.

Finally, the angle selected by the grid networks for generating a 1D response is well-distributed relative to the underlying 2D lattice: there does not appear to be a preferred slice angle for generating 1D responses (Figure 3.6G).

3.2.6 Cell-cell relationships: 2D relative phases are predicted by 1D slice fits

If the 2D response of grid cells is generated by rigidly translating an underlying population activity pattern according to animal movements, then the phase offset in the 2D spatial tuning of a pair of comodular cells (called their relative phase) should equal their phase offset with respect to the underlying population pattern. If the 1D response is generated in the same way (but as a special case where the rigid translations occur along a straight line), the same should be true, and the 2D phase offset between a pair of predicted

1D slices should equal the measured 2D relative phase. A stringent test of the slice hypothesis, then, is to ask whether a predicted 2D relative phase from fitting slices to the 1D tuning curves of grid cells on tracks agrees with the 2D relative phase measured in box environments.

The predicted 2D relative phase from the slice analysis is the offset in the starting points of the two slices within a unit cell of the unit lattice (Figure 3.7A, inset; there is 12-fold degeneracy in the relative phase prediction intrinsic to the symmetry of the lattice, see Figure S3.13 and Experimental Procedures for how we obtain a unique solution and also apply similar procedures to generate fair controls). We generate 2D relative phase predictions from 1D track responses for all pairs of comodular cells (Experimental Procedures) and test them against the measured 2D relative phases obtained from 2D trajectories in boxes. We find very good agreement between the 2D relative phases predicted from the 1D slice analysis (under the consistency-imposed slice fits described above) and those measured in 2D (Figure 3.7B, left), up to the estimation uncertainty inherent to the prediction due to spiking variability (estimated by bootstrap sampling of spikes, Experimental Procedures): in other words, the predicted and measured 2D phases are not distinguishably different from one another ($N = 25$ total comodular pairs from the 3 comodular pairs, 2 comodular quadruples, and 1 comodular quintuple in the dataset).

The errors in predicted relative phase (predictions generated pairwise) are significantly smaller than if guessed at random ($p = 10^{-5}$; $\chi = 4.62$; $df = 24$ from a chi-square test; the errors in the random predictions are themselves

not large because each random pair is also given the benefit of a 12-fold degeneracy in solutions, as noted above; see Figure S3.13 and Experimental Procedures). If the relative phase predictions for cells from a comodular K-tuple are generated by selecting a common solution domain out of the 12-fold degenerate domains across all (K choose 2) pairs (rather than allowing a different choice for each pair as above), the 2D relative phase prediction error barely increases; importantly, the random predictions become worse because they are subject to the same constraint (see Figure S3.13 and Experimental Procedures), leading to a greater separation between the predicted relative phase errors and errors from random guessing ($p = 6 \times 10^{-9}$; $\chi = 2.26$; $df = 24$ from a chi-square test).

The 1D slice analysis prediction of the phase offset between cells yields substantially more information about the 2D relative phase vector from 1D track data than can be obtained from the Pearson cross-correlation of the 1D responses (Figure 3.7C). The latter is a scalar measure that provides no information about 2D relative phase beyond its magnitude. (On the other hand, if the goal is to predict only the *magnitude* of 2D relative phase from the 1D data, Pearson cross-correlation and slice analysis yield comparable results, Figure S3.14).

3.2.7 Grid expansion from 2D real environments to 1D real and virtual environments

The spatial tuning period of grid cells expands $\approx$ 4-fold in going from 2D real open fields to 1D virtual tracks (Figure 3.6E). What factors contribute to this expansion? One factor is the move from real to virtual environments: the same cell, recorded first in 2D real open fields then in 2D virtual reality, undergoes a $\approx$ 2.5-fold expansion in its spatial tuning period [3]. Nevertheless, nearly 40% of the expansion remains unexplained by this factor. We hypothesize that even within real environments, the move from 2D to 1D induces a change in period. This possibility was already noted in [15], but the 1D period estimation methods used, it was also observed, might overestimate the 1D period because of the propensity of grid cells to entirely "miss" fields, or fail to fire in a field. Thus, it remained unclear if there is an expansion in going from 2D to 1D.

The slice hypothesis is not sensitive to missed fields: the PSD method and especially the gap distribution reveal the underlying grid period even with missed fields: missed fields will reduce the height of gap distribution at the peak corresponding to the smallest gap, but so long as that peak is present, it provides a good estimate of period. Moreover, inferring 2D parameters from 1D tracks using the slice hypothesis provides more information than gained by 1D correlation analyses. Using the slice method, therefore, it is possible to more closely answer the question of 2D to 1D grid expansion. We estimate the 2D spatial tuning periods of cells based on their dorsoventral locations [15]

(Figure S3.15A) and compare these to the inferred 2D lattice periods underlying 1D track runs (Figure S3.15B). The 2D lattice period during navigation on real 1D tracks expands by a factor of ≈ 2 (Figure S3.15C) relative to that for familiar 2D enclosures. These two factors, from real to virtual and 2D to 1D, together explain the overall expansion (scale factor) we find between 2D real and 1D virtual environments. Later, we will discuss the implications of these expansions and other findings for grid cell mechanism.

3.2.8 Shifting 1D responses may be consistent with drifting slices

Occasionally, we observed systematic shifts across trials in the 1D spatial tuning of cells, Figure 3.8A. Averaging over trials in such cases yields a spatial tuning curve with broader firing bumps, extra bumps, and lower a signal-to-noise ratio; indeed, this pair of cells was responsible for some of the worse relative phase predictions and slice fits in previous figures.

The single trial responses during drift, however, do not appear to differ statistically from trials in which there is little drift. While it is difficult in general to perform statistical tests and the PSD analysis to extract slice parameters on rapidly drifting responses, on cells with moderate drift (Figure 3.8A) we may average small sets of adjacent trials and perform a more temporally resolved analysis. Our aim is to determine whether the drifting response can still be described by the slice hypothesis locally in time, so that each small group of trials is well-described as a slice, but more separated trials are described by slices with different parameters.

We first generate a gap histogram, Figure 3.8B, which we know should be invariant to changes in slice angle and phase, if the response is generated from a 2D lattice with invariant scale. Notice that the distribution has a few distinct peaks. The locations of the peaks (Figure 3.8B, green), scaled by the location of the third peak, closely coincide with the predicted ratios for a triangular lattice.

The PSDs of the two cells over small block of trials (5-trial averages), Figure 3.8C, are quite similar, even though the PSD of each cell changes considerably over trials. This indicates that the slice parameters of both cells are similar to each other locally in time, even as the slice parameters change over time. These PSDs tend to have more than three large peaks, an indication that, if the slice hypothesis is true, the slice has already drifted over the course of the 5-trial average (the PSD over a sum of drifting slices is the sum of the PSDs of those different slices).

Finally, we directly fit slices to 5-trial moving window average firing rates (with a 1-trial shift between windows) of the two cells, Figure 3.8D. We find that the best-fit slice solutions have a constant angle and scale-factor across trials (Figure 3.8D, rows 1-2). The main change across trials is in the 2D phases of the slices (Figure 3.8D, rows 3-4). Note that the quality of fit does not vary substantially over trials despite the drift in firing patterns (Figure 3.8E, dark red and blue), and remains reasonably high despite the smaller SNR in the 5-trial averaged firing rates (the SNR in the spatial responses from 30-trial averages over non-drifting responses is substantially higher), suggesting that

the slice hypothesis is equally valid locally in time in this set of drifting trials as it is for cells that exhibit little drift. By way of control, slices 5 degrees and 10 degrees away from the optimal slice produce much poorer fit quality (Figure 3.8E, pale curves).

For visualization, we plot the moving slices directly on the underlying 2D lattice (Figure 3.8F). Note that the drift in slice parameters is largely coherent for both cells (the slices drift roughly in parallel with each other, Figure 3.8D). The predicted relative phase from the 1D responses, generated in this time-resolved way from slice fits, remains unchanged across trials up to the uncertainty inherent in the noisy spiking responses (Figure 3.8G, different black triangles for the moving 5-trial window averages; red circle is the bootstrap estimation uncertainty from spikes), and all these estimates are close to the recorded relative phase from 2D environments (Figure 3.8G, black circle).

While the drift in slice parameters across trials for this pair is mostly confined to the slice phase, in other experiments the drift can be in angle or some combination of slice parameters (data not shown).

3.3 Discussion

3.3.1 Summary of findings

We characterized the responses of grid cells on 1D tracks based on the analytical properties of the Fourier transform of linear slices through regular lattices. In previous work, it was noted that spatially periodic 2D responses should consist of only a few Fourier components, and in particular a triangu-

lar lattice should consist of three Fourier components corresponding to plane waves at 60° angles [53]. We established that despite their irregular and aperiodic appearance, 1D slices through idealized 2D lattices will have a 1D Fourier power spectrum with no more than three peaks, and the responses will exhibit a characteristic gap distribution with gap ratios determined by the 2D lattice geometry. We further showed how to use the (real and imaginary) components of the Fourier components to analytically infer the parameters of the slice. Our theory is intimately related to that of quasicrystals [90] and the X-ray crystallography of lattices. Indeed, the 1D grid cell response, under the hypothesis that it originates from slicing a regular 2D lattice, is a quasicrystal.

Applying these theoretical methods and results to 1D grid cell responses produces excellent fits and cross-validated predictions of spatial tuning. Cells from the same module are well-fit by parallel slices through the same regular lattice, and moreover, it is possible to predict the 2D relative phase of cells from slice fits to their 1D track recordings. These successes strongly support the possibility that grid cell responses on 1D linear tracks are generated by linearly translating a population-level internal 2D triangular lattice.

The ability to generate accurate slice fits means that cells recorded in various experiments on linear 1D tracks [15, 46, 27, 40] can be more readily and accurately characterized as grid cells or not. The fact that we can generate such fits from a few runs down a linear track means it is possible to rapidly infer cell-cell relationships like comodularity and 2D relative phase between grid cells from short experiments. In principle, responses on longer tracks

provide more constraints on the parameters of a slice fit; at the same time, however, there may be a greater tendency toward remapping partway along longer tracks. There are instances from our analysis of data from [15] consistent with remapping-free linear slice responses along 8 m of a longer track. 1D tracks free of distinguishing spatial landmarks along this length could inhibit landmark-anchored remapping and allow for parameter estimation from < 40s of data (10 runs down an 8 meter track at a running speed of 0.2 m/s).

Finally, our methods can be used to characterize the structural characteristics of grid cell responses in three spatial dimensions [45, 34], by constructing 1D and 2D slices of the acquired 3D data and analyzing the resulting projections.

3.3.2 Mechanism: Implications for recurrent dynamics during 1D navigation

The observed responses of individual grid cells when animals run on virtual and real 1D tracks are consistent with sampling an underlying 2D lattice response along the confined 1D path of the animal, albeit with a global expansion of the grid period by a factor of ≈ 4 on virtual tracks and ≈ 2 real tracks. Moreover, the cell-cell relationships (2D relative phase) observed between comodular cells in 2D are preserved in their 1D responses. Finally, cells deemed to be comodular in 2D share the same slice angle and scale factor in their 1D responses, consistent with the picture of an internal 2D population lattice response whose phase is shifted in tandem with animal velocity as the

animal runs, regardless of whether it is running in nonlinear arcs through a 2D environment or in straight lines on a 1D track.

The source of a non-unity scale-factor in the spatial tuning of individual cells could in principle be caused by a global expansion of the population pattern or, as predicted in [18] under the continuous attractor hypothesis for grid cell generation, by a change in the gain with which one unit of animal velocity shifts the phase of the underlying population pattern: a decrease in gain would result in an expansion of the spatial tuning period, without a change in the period of the population pattern [18].

In a previous study [102], it was shown that changes in the spatial tuning of grid cells in rescaled 2D environments were inconsistent with the first possibility: a change in the population pattern would predict a change in the relative phases of spatial tuning between cells, which were strictly preserved. Similarly here, the preservation of cell-cell relationships rules out the possibility that grid period expansion is due to a change in the period of the underlying, recurrently-driven population pattern reflected in the responses of grid cells. Instead, we conclude that the recurrent dynamics underlying the neural responses must be preserved, with period expansion coming from a change in the efficacy with which actual animal motion drives changes in the population pattern phase. This change in efficacy could arise because of a reduced perception of speed in 1D and virtual environments or a reduced ability of the speed signal to shift the population phase [102].

Both these possibilities involve changes in the feedforward input to

the grid cell system in 2D and 1D, rather than a change in the recurrent dynamics of the system. In this sense, the grid cell system appears to remain in fundamentally the same dynamical and computational regime during 2D and 1D navigation.

3.3.3 Potential exceptions to the slice hypothesis and future work

There are limits to the ability of single linear lattice slices to describe the 1D responses of grid cells. First, as already noted, when animals run on long linear tracks with stable visual features, the 1D response characteristics (from [15]) can sometimes change partway down the track: a frequently firing cell may become more sparse, the field spacings may go from more apparently periodic to less, and so on. Similarly, in 1D tracks with hairpin turns that alternate with linear segments, the grid cell responses across all segments run in one direction are nearly identical to each other [25], while those run in the opposite direction are different and closely resemble each other. These kinds of effects might be explained within an expanded version of the slice hypothesis, according to which 1D track responses are piecewise compositions of different linear slice segments. Thus, a response remapping partway down a track would correspond to a resetting of the slice parameters from that point onward. This is a topic for future empirical and theoretical tests.

A closely related observation is that grid cells can exhibit different spatial tuning in the forward and return running directions on physical tracks [15], just as observed in place cells [63, 36]. This is the case even when the

spatial tuning in one run direction is consistent with the slice hypothesis (our analysis). An interesting question for future research is whether spatial tuning on the return journey is also well-fit by a slice, and how the two slices are related. If the slice hypothesis is a good description also of the return response, a shift in the slice parameters for different run directions or different run contexts [60] could explain the direction- and context- specificity of place cell tuning in similar experiments.

Second, spatial tuning curves obtained by averaging together blocks of trials tend to not be well-fit by slices if there is a rapid drift in the spatial responses across trials. In such cases, it is possible that the single trial responses are generated by slicing a lattice, but the slice parameters shift rapidly across trials. Data limitations arising from the length and duration of an individual trial in the existing data means that further experiments and analysis will be required to properly resolve this question as well as to better characterize what drives the shift in slice parameters.

Third, to understand how animals use 2D coordinate frames when navigating along 1D trails, it will be interesting to learn about the selection of slice angles: supplied with two constrained linear paths in a 2D arena starting the same point but diverging at an angular separation angle δ, are the slice angles chosen by the animal along those paths consistent with δ? The ability of animals to solve starburst mazes [74] suggests that animals can learn the geometrical relationships between paths, and a slice analysis of grid cells along the 1D arms may provide glimpses into the mechanisms underlying this ability.

Fourth, a broader question related to 1D navigation is whether responses along *nonlinear* 1D tracks resemble corresponding nonlinear slices of the 2D lattice. Mechanistically, if grid cell activity is generated by a velocity input that shifts an underlying 2D population pattern around in tandem with animal motion [18], we would expect the answer to be affirmative. Indeed, experiments on annular tracks [101, 69] suggest that the responses of grid cells on a circular 1D track is consistent with a circular slice through a 2D lattice, and that the remapping of these responses may be the result of a shift in the phase or orientation of the circular slice through the 2D lattice [68]. Analytical Fourier parameter extraction becomes substantially more complicated for circular 1D tracks, but there are ways in which it can be generalized (Yoon, Newman, Fiete, unpublished observations).

3.4 Methods

3.4.1 Binning and rate maps

Cell-sorted spikes from foraging mice were assigned to 1 cm $\times$ 1 cm spatial bins (2D open field) or 1 cm spatial bins (1D linear track). Positions were derived from samples taken at 30 Hz. The number of spikes assigned to a bin was divided by the rat's total dwell time in that bin, to remove the effects of inhomogeneous spatial exploration on estimating the probability of spiking at each location. This defined the rate map. Smoothed rate maps were generated by convolving the binned rate maps with a 2D or 1D Gaussian kernel, respectively ($\sigma = 4$ bins).

3.4.2 Cell selection

We analyzed units from [27], which were recorded in both 2D open fields and on virtual 1D linear tracks. Our starting sample consisted of 126 units (137 recorded, with 126 of these passing an interspike interval criterion that no more than 0.25% of spikes in each unit were emitted with interspike intervals (ISIs) shorter than 1 ms [27]).

We applied the intersection of three criteria on spatial tuning to select units individually, then excluded duplicate units by a separate criterion based on similarity of tuning. These methods are described below. The result was 25 grid units (6 singles, 3 pairs, 2 quadruples, and 1 quintuple) from the starting sample of 126.

Our selection procedure follows: We first computed three scores that reflect spatial response properties individually for each unit: 1) A standard gridness score [55] on the 2D spatial tuning, to determine whether the unit has sufficiently grid cell-like tuning. 2) A score on the overall stability of the response in 1D (see Trial selection, below) computed on the most-stable 30-trial block out of the variable number of trials recorded for each unit, to remove units that do not display a stable spatial response in 1D because we could not average trials to obtain a 1D spatial tuning curve. 3) The entropy of the 1D spatial tuning curves (with the spatial tuning curve obtained from averaging the most-stable 30 trials for each cell). High entropy units have high background firing, little spatial modulation, large drifts, or generally high noise in their responses. This entropy score also helps to catch units with variable

or drifting trial-to-trial responses that are not caught by the stability measure in (2) above.

Units with 2D gridness < 0.34 [27], 1D stability < 0.1, and 1D entropy > 8.7 were rejected, reducing the sample size from 126 to 26 units (Individually, 69 units fail the 2D gridness score, 39 fail the stability score and 64 fail the entropy score). Units had two kinds of stability problems in 1D: one was a slow, systematic drift of firing field locations across trials; the other was strongly fluctuating spike counts and other less-systematic effects from trial to trial. Both led to less-sharp 1D tuning curves with lower signal-to-noise ratio. The stability and entropy thresholds quoted here were set to eliminate excess drift and variability based on empirical observation of the 1D track spike rasters.

Finally, to avoid overcounting one cell as two units, we eliminated one unit out of each simultaneously recorded pair if they exhibited both very similar spatial tuning and a large amount of mutual spike contamination (Cell pairs with very similar tuning but small or no spike contamination are legitimate independent samples and should both be included – grid cells within a population are expected to have a spectrum of relative phases, including a relative phase of zero). For a cell pair with relative phase magnitude smaller than 10% of the maximal possible relative phase separation, and a contamination score (defined below) greater than 0.1, we kept only one cell of the pair (the one with the higher gridness). This process identified a single pair of units from the set of 26 (suggesting one double-counted cell), and led to the final dataset of 25 units.

3.4.3 Entropy

Entropy is given by the standard definition:

$$\text{Entropy} = -\sum_i P(x_i) \log P(x_i)$$

where $P(x_i)$ is the firing rate in the ith spatial bin along the linear track, normalized by the area of the full firing rate curve (firing rate per bin, summed over all bins).

3.4.4 Comodularity

Cells were defined as comodular if they were recorded simultaneously from the same animal and if the spatial periods (mean spacing of firing fields) of their recorded 2D responses differed by less than 10% (relative to the larger of the two estimated periods) and the 2D grid orientations differed by less than 10 degrees.

3.4.5 Spike cross-contamination for cell pairs

Following [47], we fit the pair of clusters C_i and C_j comprised of the waveforms of all spikes from two simultaneously recorded units i, j by a sum of (two) Gaussians. With this fit, we derive a contamination score as the average probability that a spike from C_i would come from C_j's distribution:

$$\text{contamination}(C_1, C_2) := \frac{1}{N_1} \sum_{v \in C_1} P(C = C_2 \mid V = v)$$

3.4.6 Additional cells from real 1D tracks

We additionally studied the 1d real-track responses of cells from [15]. As stated in the main manuscript, for left-to right traversals we used only the left (first) half of the track, and for right-to-left traversals, we used the right (first) half of the track. When we refer to track length L in the analysis below, we refer to these first half-tracks. The responses of a cell on the two different direction traversals were treated as two independent spatial responses. We possess no ground-truth data on which responses in [15] come from grid cells. Thus, we must identify responses as coming from putative grid cells based on characteristics of the spatial response. We do so as follows:

The selection criteria for identifying putative grid cells from [15] are adapted from Domnisoru et al. 2013, with adjustments to simplify the criteria slightly and take into account the very different track length. The modified criteria are as follows: A 1D response cannot be a putative grid cell response if (a) the number of transitions N_{trans} between an in-field and and out-of-field period for a track of length L is smaller than $L/(5\langle w\rangle)$, where $\langle w\rangle$ is the mean firing field width for that 1D response. (In-field and out-of-field periods are defined below.) (b) The widest field of the response has width greater than $5\langle w\rangle$. (c) Same as in [27], which is that 30% or fewer of the bins are assigned to either in-field or out-of-field periods. (d) the mean firing rate in-field divided by the mean firing rate out-of-field is smaller than 2.

The definitions of in-field and out-of-field periods are very similar to [27], but the numerical thresholds are changed and some of the steps are sim-

plified: out-of-field periods are defined as intervals of $\geq 8cm$ for which the firing rate was lower than the 20th percentile of the bootstrapped shuffle distribution for that bin (i.e., $1 - P_{value} \leq 0.2$). Firing fields are intervals of duration $\geq 12cm$ with firing rate higher than the 80th percentile of the bootstrapped shuffle distribution for that bin (i.e., $1 - P_{value} \geq 0.8$). There is no change in criteria for defining firing fields if they occurred at the ends of the track. Bins with intermediate firing rates remained unassigned. Steps omitted from [27] criteria for simplicity: we did not extend candidate firing fields by an adjacent bin on each side if their firing rates exceeded the 70th percentile of the bootstrapped shuffle distribution for that bin; we did not examine and discard fields based on whether they received spike contributions from only $\leq 20\%$ of all trials.

If the response of a cell in either traversal direction survived the selection process above, the cell was classified as a putative grid cell. We were left with 51 putative grid cells with 65 passing spatial responses, out of the full dataset of 97 total cells with 194 spatial responses. We applied slice fits to all 65 passing responses, provided they had ≥ 3 fields on the length-L half-track (this condition is necessary for generating meaningful slice fits: two 1D fields are mathematically not enough of a constraint to specify a slice or infer the underlying grid cell period, by any method). This leaves 40 putative grid cells and 53 responses; the excluded cells here tend to be from the most ventral end of the dorsolateral MEC.

3.4.7 Trial selection

Different cells were recorded for different numbers of trials, and some cells showed substantial drift in their spatial tuning across trials. To equalize the number of trials used per cell and obtain the most stable block of trials for each cell so that we could obtain reasonable trial-averaged rates, we adopted the following trial-selection procedure for Figures 3.1-8: Given a total of $N > 30$ trials, we computed a stability score, defined as the average of the pairwise Pearson's correlation coefficient across all pairs of single-trial rate responses (obtained by smoothing the single-trial spiking responses, as described in *Binning and rate maps* above) within the consecutive 30-trial block $[i, i + 29]$. We repeated this process for different starting positions i in steps of 1 trial, starting from $i = 1$ up to $i = N - 29$. From this analysis, we selected the 30-trial block with the largest stability score; for comodular cells, we selected a common 30-trial block with the largest stability score averaged across cells.

3.4.8 Fourier spectral analysis for inferring 2D lattice slice parameters from a 1D response: theory

Consider a firing rate r of unit height that varies as a function of the 2D variable $\mathbf{x} = (x_1, x_2)$, in the shape of an equilateral triangular lattice with period λ. This response can be written as the sum of three 2D sine waves along the directions $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$, where $\mathbf{b}_1 = \frac{1}{\lambda}(0, 2/\sqrt{3}), \mathbf{b}_2 = \frac{1}{\lambda}(1, -1/\sqrt{3}), \mathbf{b}_3 =$

$\frac{1}{\lambda}(1, 1/\sqrt{3})$ (Figure S3.1). Thus,

$$r(\mathbf{x}) = \frac{1}{3}\sum_{i=1}^{3}\cos\left(2\pi\,\mathbf{b}_i\,\mathbf{x}^T\right) \tag{3.1}$$

where the superscript T on a vector refers to the vector transpose, and $\mathbf{x}\mathbf{y}^T$ is the dot product between vectors $\mathbf{x}, \mathbf{y}$.

Next, consider a semi-infinite line $\mathbf{x}_\ell(t)$ parametrized by t in the 2D coordinate space above, with origin at $\mathbf{c} = \lambda(\phi_1, \phi_2)$ and angle θ relative to the x-axis. It is given by

$$\begin{aligned} \mathbf{x}_\ell(t) &= (\cos(\theta), \sin(\theta))\,t + \mathbf{c} && (3.2)\\ &\equiv \mathbf{u}(\theta)\,t + \mathbf{c} && (3.3)\end{aligned}$$

where we have defined the vector $\mathbf{u}(\theta) = (\cos(\theta), \sin(\theta))$.

The modulation of firing rate r along this line is given by restricting $r(\mathbf{x})$ to the coordinates $\mathbf{x}_\ell(t)$. Thus, along the line, the rate varies as:

$$\begin{aligned} r(\mathbf{x}_\ell(t)) &= \frac{1}{3}\sum_{i=1}^{3}\cos\left(2\pi\,\mathbf{b}_i\,\mathbf{x}_\ell(t)^T\right) && (3.4)\\ &= \frac{1}{3}\sum_{i=1}^{3}\cos\left(2\pi\,\mathbf{b}_i\,\mathbf{u}(\theta)^T t + 2\pi\,\mathbf{b}_i\,\mathbf{c}^T\right) && (3.5)\\ &= \frac{1}{3}\sum_{i=1}^{3}\cos\left(2\pi f_i(\lambda,\theta)t + 2\pi\delta_i(\mathbf{c})\right) && (3.6)\end{aligned}$$

where $f_i(\lambda, \theta) \equiv \mathbf{b}_i(\lambda)\,\mathbf{u}(\theta)^T$ and $\delta_i(\mathbf{c}) \equiv \mathbf{b}_i(\lambda)\,\mathbf{c}^T$ are all scalar quantities (we have made the dependence of the $\mathbf{b}_i$'s on λ explicit here because λ is a slice parameter we would like to infer from the following analysis, as is θ).

From Equation 3.6, it is clear that the Fourier transform of $r(\mathbf{x}_\ell(t))$ will have spectral peaks at the frequencies $f_1(\lambda,\theta)$, $f_2(\lambda,\theta)$, $f_3(\lambda,\theta)$ and phases $2\pi\delta_1(\mathbf{c}), 2\pi\delta_2(\mathbf{c}), 2\pi\delta_3(\mathbf{c})$ (Figures S4 A-D).

Given the Fourier transform of $r(\mathbf{x}_\ell(t))$ (more specifically, given the six scalars $\{f_i|i = 1,2,3\}$ and $\{\delta_i|i = 1,2,3\}$), our goal is to extract four quantities: the period and angle parameters (λ,θ) of the slice as well as the vector origin of the slice, recast as a phase with respect to a unit vector of the lattice, given by the vector $\boldsymbol{\phi} \equiv \mathbf{c}/\lambda$.

For any equilateral triangular lattice (characterized by the three vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ specified above) it is easy to see that $\mathbf{b}_1 + \mathbf{b}_2 - \mathbf{b}_3 = 0$. As a result, $f_1+f_2-f_3 = (\mathbf{b}_1+\mathbf{b}_2-\mathbf{b}_3)\,\mathbf{u}(\theta)^T = 0$ (Figure S3.4-F). Thus, only two of the f_i's provide independent information and a pair of f_i's is sufficient to analytically obtain λ,θ. The corresponding two δ_i's are sufficient to analytically determine the slice origin vector $\mathbf{c}$. This procedure will produce a reasonable solution if a recorded rate is indeed a slice.

Note that the scale factor (α) referred to in the main manuscript is defined as λ (the inferred period of the lattice that underlies the 1D response), divided by the measured period of the cell's response in 2D (λ_{2D}). Thus, $\alpha = \lambda/\lambda_{2D}$. The inferred 2D phase from the 1D slice analysis in the main manuscript is $\boldsymbol{\phi} = \mathbf{c}/\lambda$; we write it in terms of the oblique projection onto the two primary lattice vectors [102].

3.4.9 Direct inference of 2D relative phase from a pair of 1D responses

The relative 2D phase between two linear slices is simply the difference in the estimates of the 2D phases of each of the two slices ($\Delta\boldsymbol{\phi} \equiv \boldsymbol{\phi}^1 - \boldsymbol{\phi}^2$, where $\boldsymbol{\phi}^1, \boldsymbol{\phi}^2$ are the 2D phases of each linear slice).

However, it is also possible to directly infer the relative 2D phase from the two 1D responses, using the cross-correlation theorem:

$$\mathcal{F}\left[(g \star h)(t)\right] \;=\; G(f)\overline{H(f)} \tag{3.7}$$

where $\star$ denotes the correlation and G, H are the Fourier transforms of g, h, respectively. The Fourier transform of the cross-correlation of the two signals g, h (the left hand side of Equation 7) will have phase $2\pi\delta_i(\mathbf{c}_g - \mathbf{c}_h)$ (as in Equation 6) ascribed to the conjugated $H(f)$, which can be directly recast as a 2D relative phase, $\Delta\boldsymbol{\phi} = (\mathbf{c}_g - \mathbf{c}_h)/\lambda$.

3.4.10 Fourier spectral analysis for inferring 2D lattice slice parameters from a 1D response: practice

For each cell recorded in 1D (or each random control response), we first compute the PSD of its response, identify the two highest peaks in its PSD, and label the peak locations (i.e. the spatial frequencies at which the peaks occur) as q_1 and q_2 with $q_1 < q_2$. Given (q_1, q_2), there is a choice in whether (q_1, q_2) should be identified with (f_1, f_2), or (f_1, f_3), or (f_2, f_3) in the analytical definitions above. We consider the three possible solutions, in addition to the two solutions corresponding to the special cases of a 0 or 30 degree slice (for

which there are only two spectral peaks, see Figure 3.2). From this set of 5 discrete slice solutions, we pick the one with the best correlation with the recorded 1D response of the cell. The 2D period, obtained from 2D recordings for each cell, was was set to the average of lengths of two primary lattice vectors.

Note that different assignments of the two highest peaks to the f_i's (three of the five candidate solutions described above) can correspond to different local minima in the solution space; sometimes, these different local minima can provide roughly similar (in terms of correlation coefficient) slice fits to the measured 1D response. We explore this issue further in Figure 3.6 of the main manuscript.

We numerically refined the analytical solution by searching locally for values of the four slice parameters $(\theta, \alpha, \boldsymbol{\phi})$ that optimized the fit between the recorded 1D rate response and rate values predicted by the slice. The fit is quantified by Pearson's correlation coefficient (ρ). The slice parameters were initiated to the analytical solution $(\theta^*, \alpha^*, \vec{\phi}^*)$ then locally relaxed via a standard gradient ascent algorithm under the following constraints: $\max(0°, \theta^* - 3°) \leq \theta \leq \min(30°, \theta^* + 3°)$, $\alpha^* - 0.5 \leq \alpha \leq \alpha^* + 0.5$, and $[0, 0] \leq \vec{\phi} \leq [1, 1]$.

3.4.11 Imposing consistency in the slice fits across comodular grid cells

For a K-tuple of comodular cells, we imposed consistency in slice fits across the K cells as follows: First, compute the product of all K PSDs obtained from the spatial tuning curves of each cell. Select the spatial frequencies corresponding to the two highest peaks from the PSD product; call these q_1 and q_2. For each cell, we now pick the two peaks from their individual PSDs that lie closest to q_1, q_2. The rest of the procedure for finding the best slice for each cell is as described above. This procedure corresponds to selecting the same local minimum in the solution landscape across all cells in the K-tuple.

When reporting relative phases between a pair of cells (or doing a K-tuple analysis for K comodular cells), the 2D period was set from 2D recordings to the average value along the two lattice vectors, and averaged across cells in the pair (or the K-tuple).

3.4.12 Measure for "three-peakiness" in PSD

To quantify to what extent a given PSD has a three-peaked structure as predicted for lattice slices (Figures 3.2C and 3.2F), we define a "three-peakiness score" (p_3) : It is computed as the sum of the areas under the three largest peaks in the PSD, divided by the total area of the PSD, followed by a normalization that equalizes the score across different lattice slices. Normalization involves dividing the result by the same score for the best-fit linear slice (the normalization process leads to a score of $p_3 = 1$ for all ideal lattice

slices; in general, p_3 is high for a curve with three or fewer major peaks and little area under the rest of the curve).

3.4.13 Gap analysis

For all recorded cells, we collected gap statistics as follows: Per trial, we identified bursts of spikes (described next). A gap is the distance between the centers of mass of adjacent (consecutive) bursts. We compile a histogram of all such bursts across each trial and across all 30 selected trials (see above for Trial selection) for each cell.

To define a burst, we slide a small time-window of width w across a spike train. The window contains a burst if at least n_{spk} spikes occur in that window. We used $n_{spk} = 3$ and $w = 0.1$s. To avoid segmenting a burst artifactually into two, when building the gap histogram we ignored inter-burst distances smaller than the 2D grid period (as can happen when a single long burst within one field is punctuated by a chance pause of duration $> w_{burst}$).

For any 1D responses that contain no spiking records (this includes the random controls, which are rates), we define gaps based on rates. We compute the inter-peak distances between all adjacent firing rate peaks of height $> 25\%$ of the maximum firing rate (Figure 3.3C).

3.4.14 Generating matched random control data

Consider a cell with spatial tuning curve $\mathbf{r}(x)$, measured over the range $x \in [0, x_{max}]$. We generate two kinds of matched controls for this cell: gap-

randomized and gap-shuffled. For both types of control, we first process the data as follows.

For each spatial tuning curve, identify the locations of all peaks and troughs. Significant peaks are those whose heights exceed 25% of the maximal firing rate in the tuning curve. Label the locations of these peaks $\{p_i\}$ (with $i = 1, \cdots, Q$ and $p_1 < \cdots < p_Q$). We seek to identify firing fields, which may include more than one significant peak if the peaks are very closely spaced. Label a set of consecutive significant peaks as belonging to the same field if they are closer to each other than half the estimated field width for the cell (field width w is defined as the distance from the origin to the first trough in the autocorrelogram of $\mathbf{r}(x)$). In other words, K consecutive peaks indexed $i, i+1, \cdots, i+K-1$ belong to a single field if $p_{j+1} - p_j \leq w/2$ for each $j = i, \cdots, i+K-2$. (For a well-isolated peak separated by $> w/2$ from the rest, $K = 1$.) We now define a field interval as the spatial interval encompassing these K peaks, with the interval boundaries given by the closest troughs immediately flanking to the left and to the right the set of K peaks. Starting from the first significant peak in the spatial response, we identify field intervals as described above, and denote the αth field interval as $[x_s^\alpha, x_e^\alpha]$, where x_s^α, x_e^α are the start and end coordinates marked by the two troughs surrounding the field. The intervals $[x_e^\alpha, x_s^{\alpha+1}]$ between field intervals (and also the interval $[0, x_s^1]$), if non-empty, are designated non-field intervals.

Suppose the cumulative length of all L field intervals put together covers M bins, while the cumulative length of the full track covers $N = x_{max}/\Delta x$

bins, where Δx is the width of each spatial bin ($\Delta x = 1$cm; x_{max} and Δx have dimensions of centimeters, while L, M, and N are dimensionless). Then the cumulative length of non-field intervals covers $N - M$ bins.

To generate gap-randomized controls, we take the following steps: first generate a set of $N - M + L$ empty slots. Randomly and without replacement, assign each of the indices $\{1, \cdots, L\}$ to one empty slot. The slot with the index α now represents (contains) the firing rate segment $\mathbf{r}([x_s^\alpha, x_e^\alpha])$ (which may, and typically does, consist of multiple spatial bins). In the remaining $N - M$ empty slots, insert the $N - M$ spatial bins of non-field responses (we preserved the order of the non-field bins; this detail is unimportant because non-field responses are of low amplitude and randomizing or not randomizing bin ordering does not lead to quantitatively big effects). The result of this procedure is to exactly preserve the structure of each field, while randomizing the ordering of the fields and the lengths of the gaps between fields.

This defines one gap-randomized sample for a given cell. This procedure was used in Figure 3.5A (reddish-gray), to generate 100 random samples based on the cell in Figure 3.5A (top). For Figures 3.5B-C (reddish-gray), we generated 100 random samples for each cell in our virtual track dataset; this pooled dataset was our full random sample.

To generate gap-shuffled controls, we first note the center-of-mass location of each field (designate the center-of-mass of the αth field by x_c^α; the spatial bin index of this location in the original response is $x_c/\Delta x$), then compute the $L - 1$ gaps $g^\alpha \equiv x_c^{\alpha+1} - x_c^\alpha$ between adjacent pairs of the L identified

fields (the αth gap length in bins is $g^\alpha/\Delta x$). Next, we randomly permute the indices $\{1, \cdots, L\}$ to get a permuted vector of the indices, $\{c_1, \cdots, c_L\}$ (thus a random ordering of fields), and also randomly permute the indices $\{1, \cdots L-1\}$ to get another permuted vector $\{d_1, \cdots, d_{L-1}\}$ (thus a random ordering of gaps). We draw at random (uniformly distributed in the interval $[0, N - \sum_{\alpha=1}^{L-1} g^\alpha/\Delta x]$) a location to begin the c_1th field. We assign the gap g^{d_1} next by inserting the corresponding number of empty spatial bins after the first field. At the end of this gap, we place the c_2th field, then gap g^{d_2}, and so on, until all fields and gaps have been assigned. The empty gap bins are now filled in with non-field data, as in the gap-randomized controls above. The result of this procedure is to exactly preserve the structure of each field, and to preserve the set of all gap lengths from the data in each random control sample; however, a given field is typically adjacent to different gaps across different random samples.

As described above, in Figure 3.5A (bluish-gray) we generated 100 random samples based on the cell in Figure 3.5A (top). For Figure 3.5B-C (bluish-gray), we generated 100 random samples for each cell in our virtual track dataset.

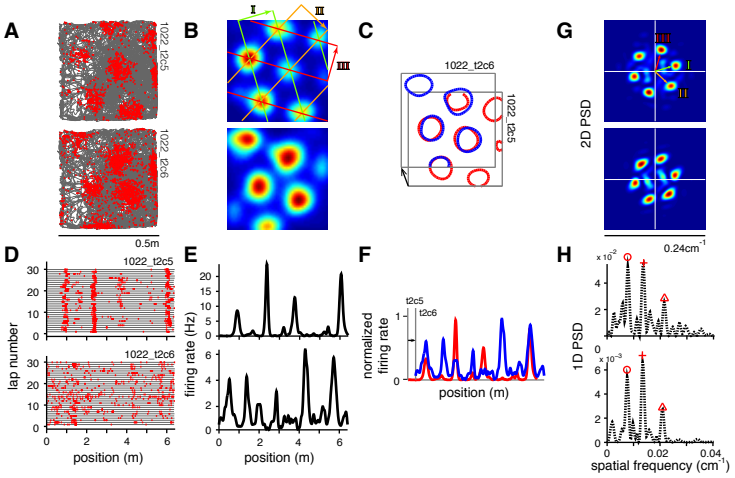


Figure 3.1: (A) Trajectory (gray) of an animal foraging in a square enclosure, with spike locations (red dots) for two simultaneously recorded cells (top and bottom, respectively). (B) Smoothed firing rate maps, with a decomposition of the triangular grid response into three 2D sinusoidal waves (three colors and the labels I, II, III). (C) Contour plot of the firing fields of cell 1 (red) overlaid on the rigidly translated fields of cell 2 (blue): the spatial tuning curves are simply shifted versions of each other. Black arrow: relative phase. (D-E) The same cells as in A), recorded on a virtual linear track (D: trajectory and spikes; E: firing rates). (F) The firing rate profile of cell 1 (red) overlaid on the rigidly translated fields of cell 2 (blue): the tuning curves are *not* simply shifted versions of each other. (G) The Fourier power spectral densities (PSDs) for the 2D responses of these cells. The peaks represent the three planar sinusoidal waves that make up the triangular lattice (I-III as in (B); see also Supplementary Fig 1). Both cells have identical PSDs, indicating their shared lattice structure up to phase shift. (H) The PSD of the 1D responses, with the highest three peaks marked by red symbols; the 1D PSDs are very similar for both cells.

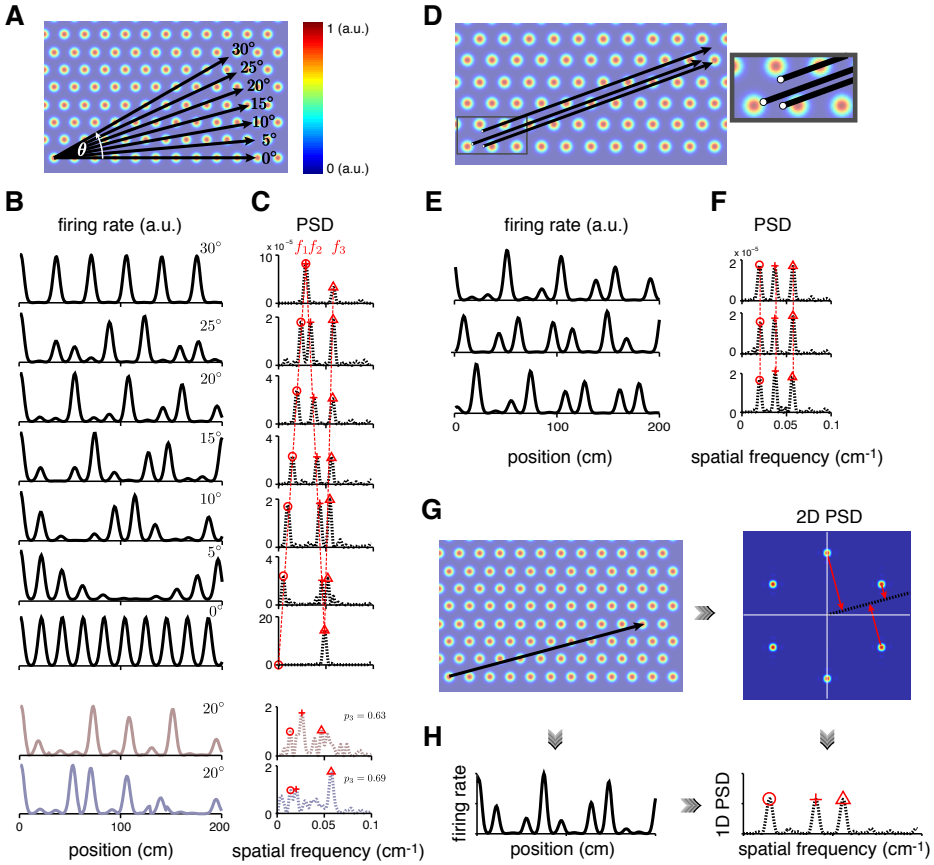


Figure 3.2: Slices of different orientations through a 2D triangular lattice (A, black) result in different 1D responses (B, black). (C) The PSDs (dashed black) also differ, yet have commonalities: They are characterized by three main peaks (red symbols) (but at two special angles, $\theta = 0°$ and 30°, two of the PSD peaks coalesce into one for a total of two peaks). (B-C), reddish gray and bluish gray: Two statistically matched random control tuning curves (B) generated from the 20° slice with their PSDs (C). (D) Parallel linear slices with different starting points or phases (inset: magnification of the starting points) result in different 1D responses (E), but identical PSDs (F). (G) Left: A linear slice through a 2D triangular lattice. Right: 2D Fourier transform of the 2D lattice, overlaid by a slice from the origin at the same angle (black). (H) Left: 1D response resulting from the linear slice in (G, left). Right: PSD of the 1D response at left. This 1D PSD is equal to the projection (red arrows in (G, right)) of the 2D PSD onto the same linear slice (also see Figure S3.4).

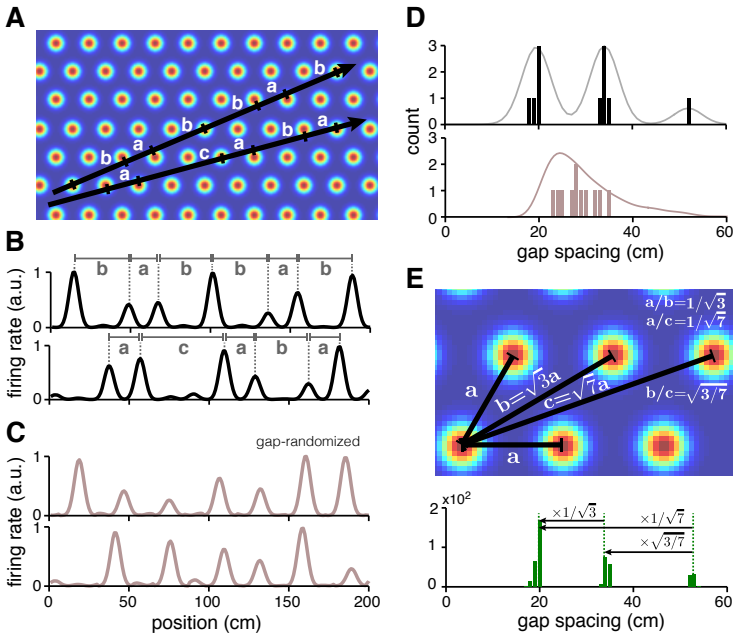


Figure 3.3: (A-B) Two linear slices (black) of different orientations and spatial phases generate different sequences of inter-bump distances or gaps (gaps delineated by vertical ticks in (A), and gaps of the same length marked by the same roman letter; only peaks with amplitudes above a threshold are considered in (B). (C) Matched random 1D control responses (gap-randomized controls) for the responses in (B). (D) Histogram of gaps for the slices in (B) (black) and the random controls in (C) (reddish-gray). Line: smoothed version of the histogram (gray) and a full gap distribution generated from 100 gap-randomized controls for each of the two slices in (A-B) (reddish-gray). (E) Top: the gaps have numeric values determined by the specific geometry of the 2D lattice (up to overall scaling by the lattice period), and gap ratios depend only on lattice geometry. Bottom: a histogram of gaps generated from pooling across 100 linear slices of the same length as in (B), with random angles and starting points.

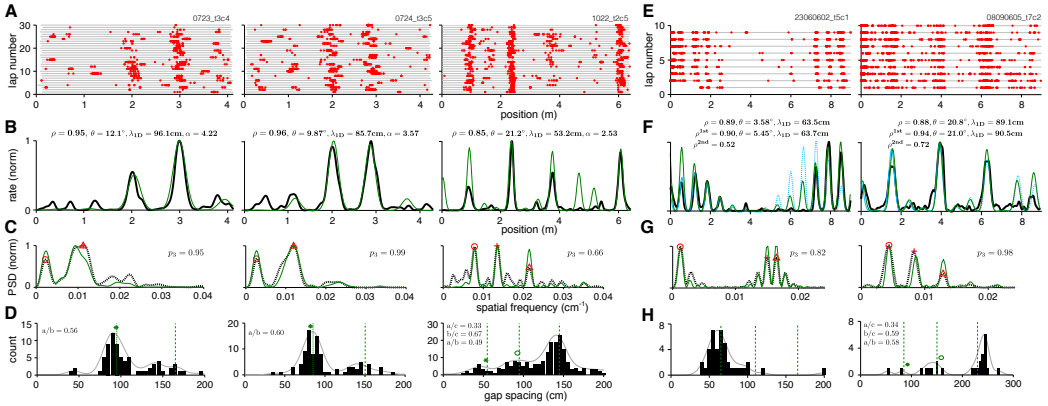


Figure 3.4: (A) Spike rasters of three cells recorded on a linear virtual track (right: same cell as in Figure 3.1D (top); center, left: cells recorded from different animals on different dates [27]). Smoothed trial-averaged rate (B, black), PSD (C, dashed black), and gap histogram (D, black bars; smoothed histogram indicated by solid gray line) for the same three cells. Gaps are computed from adjacent bursts per trial and pooled across trials. Red symbols in (C) indicate the PSD peaks used to generate the best-fit slice. The rate prediction from the slice is shown in (B, green), with fit quality (Pearson's correlation coefficient between lattice slice prediction and neural tuning curve) and slice parameters noted at top. PSD of the slice (C, green). Green dashed line in (D): predicted peaks locations for the gap distribution obtained by multiplying the ideal predicted gap ratios by the scale factor of the best-fit slice; green symbols: ideal gap ratios multiplied by the position of the actual rightmost peak in the gap distribution. (E-H) The same as (A-D), for two cells from animals running on a long linear track [15]. The cyan curve in (F) is the slice prediction from fitting only the first 4m of track. The data on the second 4m of track act as a cross-validated test of the predicted slice.

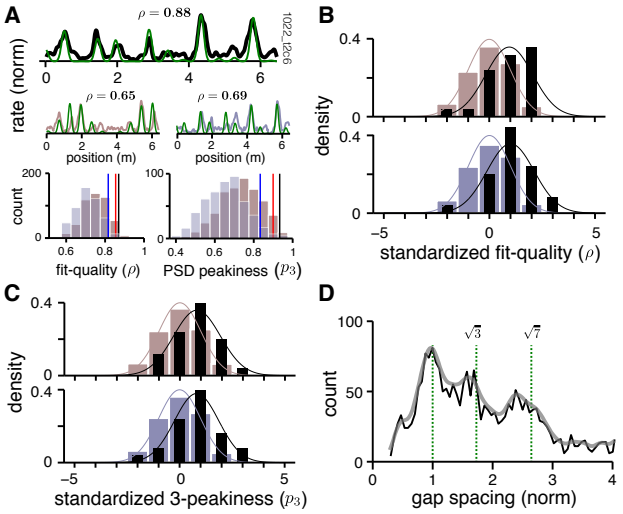


Figure 3.5: In all spatial tuning plots, the slice prediction is given in green. (A) Top: Cell firing rate (black). Middle: A gap-randomized (reddish-gray) and a gap-shuffled (bluish-gray) control sample of cell response above. Slice fit quality (ρ) noted above. Bottom left: Quality of slice fit histogram for 500 gap-randomized control samples (reddish-gray), and for 500 gap-shuffled controls (bluish-gray). Red/blue vertical lines: 95% percentile level of the random/shuffle control fit score. Black line: slice fit value of the response at Top. Bottom right: Distribution of three-peakiness scores for the PSDs of random/shuffle controls (reddish- and bluish-gray histograms, respectively), compared to the the score for the response at Top (black vertical line). Red/blue vertical lines: 95% percentile level of the random/shuffle distributions. (B) Top: Distribution of fit-quality scores for all cells in the dataset (black), with each cell's score standardized by the distribution for its matched gap-randomized controls (as in (A)); the randomized control distribution (reddish-gray) has zero mean and unit variance because of the standardization procedure. Bottom: The same as above, but the standardizing distributions are from gap-shuffled controls (bluish-gray); data in black as above. The difference in the means of the data and control histograms is highly significant (p-values noted in text, one-sample t-test). (C) Same as (B), but the distribution quantifies the statistics of three-peakiness scores in the data PSDs compared to PSDs from both types of control. (D) Aggregate gap distribution (black), pooled across all cells in the dataset, after the gaps in each cell's response are normalized by the estimated 1D lattice period (scale factor). Gray: smoothed version.

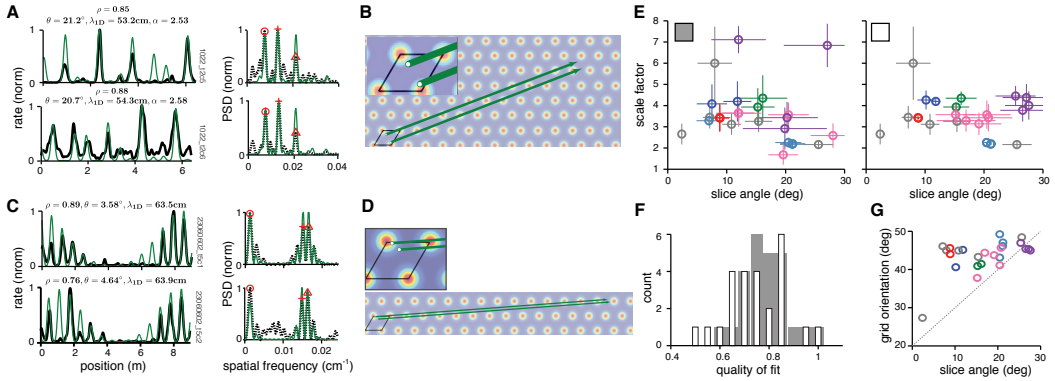


Figure 3.6: (A) Left: Firing rates of a pair of comodular cells, recorded on a virtual track (black). Best predicted firing rates from lattice slices, found separately for each cell (green). The slice angle, θ, and scale factor, α, are noted above. Right: PSD of the 1D response (black dashed) with the PSD of the predicted slice (green). (B) The best-fit linear slices for the cell pair from (A), with the starting points of the slices magnified in the inset. The starting point of a slice defines the 2D spatial phase of the cell (with respect to the rhomboidal unit cell, black), as inferred from the 1D response. (C-D) Same as (A-B), for a pair of putatively comodular cells recorded on a long linear track, data from [15]. (E) Left: the best-fit slice parameters for all cells in the dataset, obtained for each cell individually. Cells classified as comodular are plotted in a common color and single (non-comodular) cells are colored in gray (12 groups, 25 cells). Crosshair: error bars, $\pm$ 1 s.d. ($n = 100$ bootstrap samples). Right: same as left, but the PSD peaks selected for the slice fits are chosen consistently for comodular cells (see Experimental Procedures). Slice parameters for comodular cells now cluster together. (F) Histogram of fit-quality between measured 1D tuning and predicted slice fit, when fits are generated individually (as in (E), left; gray histograms) and when they are obtained by imposing consistency across comodular cells (as in (E), right; white histograms). Fit quality does not suffer in a way that is statistically significant when consistency is imposed (p-values noted in text, paired t-test). (G) Grid orientation measured in 2D versus slice angle estimated from 1D responses of the same cells (same color-codes as in (E)).

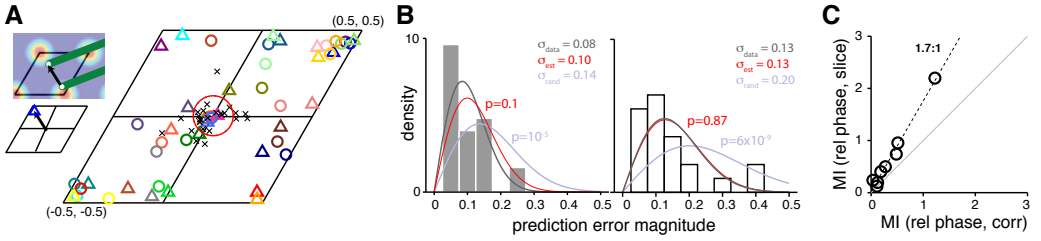


Figure 3.7: (A) Top inset: The offset in a pair of parallel slices, within a rhomboidal unit cell of the unit lattice, equals the predicted relative phase from the 1D slice analysis. Bottom inset: schematic of how the relative phase above is plotted as a symbol (blue triangle) within the unit cell. Main: Relative phases predicted from 1D responses (colored triangles) and actual relative phases measured from 2D recordings (circles in matched colors). Shown are the best predictions from across 12 degenerate solutions per pair. Black x's: differences between the predicted and measured 2D relative phase values. Red circle: relative phase estimation uncertainty (expected error) obtained by bootstrap (Experimental Procedures). (B) Left: Histogram of relative phase prediction error magnitudes for best pairwise slice fits (gray bars; Rayleigh distribution fit in black), inherent estimation uncertainty from bootstrap (Rayleigh distribution fit in red), and random prediction error magnitudes (Rayleigh fit in bluish-gray). Right: The same as left, but best fits constrained so that all pairs in a K-tuple of comodular cells were constrained a common solution domain from across the 12 possibilities (rather than a free choice per pair; Experimental Procedures and Figure S3.13). Left and right: p-values that the slice prediction errors are from the same distribution as the estimation uncertainty (red), or the same distribution as random (bluish-gray). (C) How much information (Shannon's mutual information, abbreviated MI) the correlation coefficients between cell pairs' 1D responses convey about their measured 2D relative phases (abscissa) versus the mutual information between measured 2D relative phases and the predicted values from slice analysis on the 1D responses (ordinate). The latter exceeds the former by a factor of 1.7.

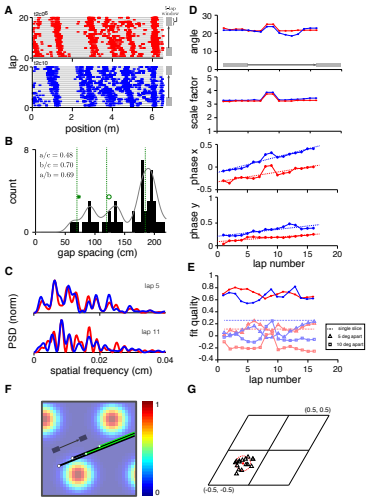


Figure 3.8: (A) Spike rasters for a pair of comodular cells show field drift, field splitting, and field merging across trials. (B) The gap histogram for the data in (A), pooled across both cells, contains three peaks. The three peaks are at the predicted values (green dashed lines; obtained from 200 randomly oriented and shifted slices through a triangular lattice), given the estimated slice period. (C) Top: the PSDs of the two cells computed from a 5-trial block centered on lap 5. Bottom: the same, centered on lap 11. The across-cell PSDs for a given block of trials is more similar than the PSD of a given cell across trials. (D) Slice parameter (slice angle, scale factor, and the two components of phase, respectively) evolution (estimated from 5-trial averages sliding one trial at a time). The best-fit slices have constant angle and scale factor, but the phase of the slice is drifting across trials. Since phase is a periodic variable, it is equivalent to plot it between [-0.5, 0.5] or equivalently, [0,1], as done in some plots here and in (E). (E) The quality-of-fit for the slice hypothesis (red, blue cells) over the drifting trials. Light red/light blue: the quality of fit between the data and slices with orientation 5 (triangle) or 10 (square) degrees from the best-fit slice. (F) Depiction, on the 2D lattice, of the inferred drifting best-fit slices from (D). Black, green: best-fit slices from near the beginning and end of the block of trials, respectively. (G) The estimated relative phase between cells across inferred slices over time (different black triangles) remains close to relative phase estimate obtained from averaging together all trials (red dot). The scatter in different time-resolved relative phase estimates is about equal to the inherent estimation uncertainty due to finite spike sampling (estimated by bootstrap), red dashed circle.

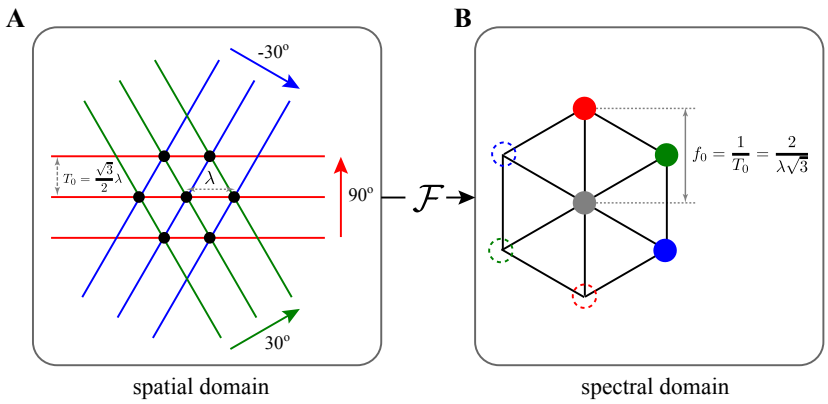


Figure S3.1: **Duality of the spatial and Fourier spectral domain characterizations of a hexagonal lattice.** (A) A regular triangular lattice of spatial period λ visually separated by two-dimensional sinusoidal waves of the period T_0 and three different orientations (red, green, and blue). Each line represents the crest of the waves. (B) Fourier transform of the hexagonal lattice represented by three spatial frequencies (same color-coded dots) that correspond to the frequency and directionality of three sinusoids in (A). Dashed circles: negative frequencies naturally come up from complex exponentials.

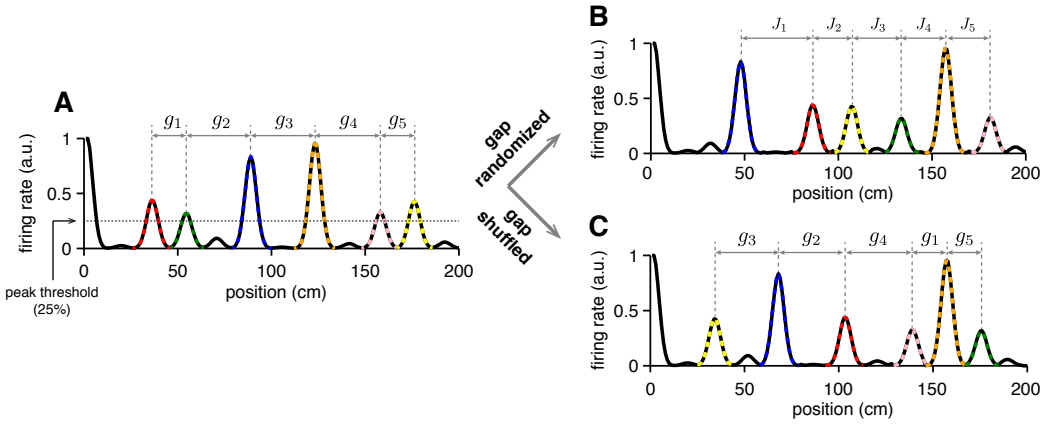


Figure S3.2: **Procedure for generating random control data.** (A) An ideal linear slice (black) through a 2D lattice. Fields are defined as the bump whose peak height is greater than a certain threshold (25% of the maximum firing rate) and in between two troughs: 6 fields color-coded by red, green, blue, orange, pink, and yellow. g_i is the gap (inter-peak distance). (B) Gap-randomized control by shuffling the order of firing fields and randomly placing the fields in between background signal (black). New gaps (J_i) does not preserve the original gaps (g_i). (C) Gap-shuffled control by shuffling the order of firing fields and placing them with keeping the original gaps.

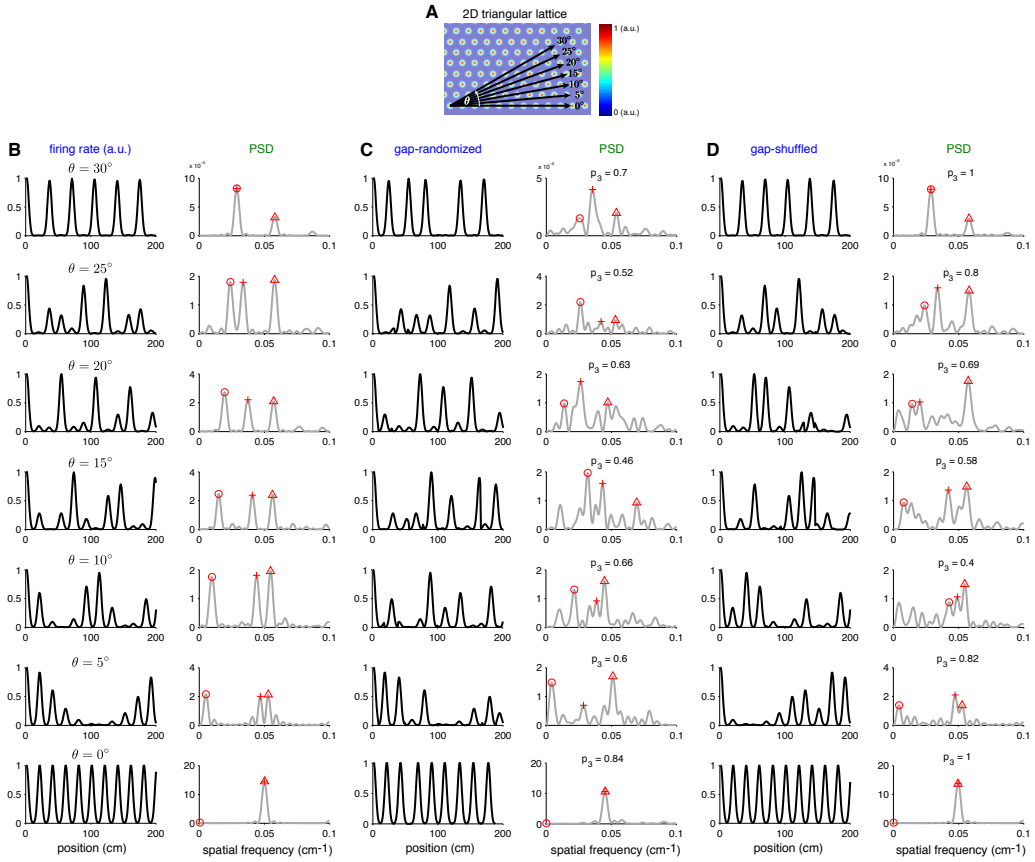


Figure S3.3: **Features of the PSD for random control data.** (A-B) Same as in Figures 3.2A-C. (C) Gap-randomized controls (left) of the data in (B) left and the PSDs (right) of the control data. Three peakiness score (p_3) is shown on the top. (D) Same as in (C) from gap-shuffled controls.

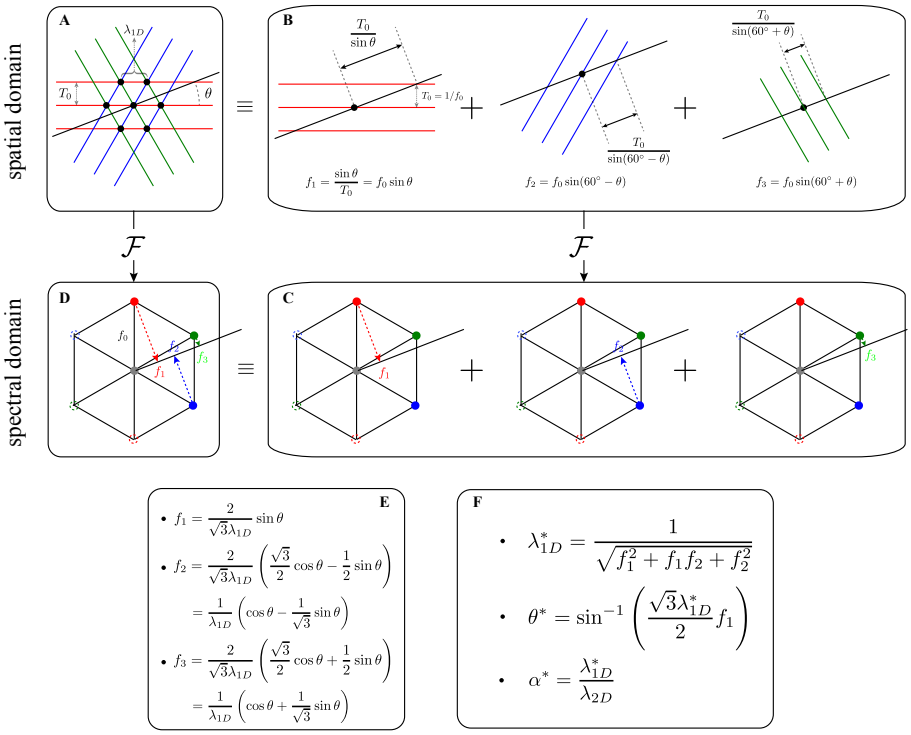


Figure S3.4: **Geometric interpretation of Fourier spectral analysis.** (A) The same triangular lattice as in Figure S3.1-A overlaid by a linear slice of orientation θ (black line). (B) The triangular lattice is the sum of three sinusoidal waves (red, blue, and green), and a linear slice through the lattice is equal to the sum of the linear slices through each sinusoid. The individual slice results in a periodic profile of the frequency $f_{i\in\{1,2,3\}}$. (C) The frequency f_i is geometrically equivalent to the projection of the corresponding spectral peak of i^{th} sinuosoidal waves in (B) onto a line of the angle θ. (D) Linearity of the Fourier transform induces that the Fourier transform of a linear slice through the lattice (A $\rightarrow$ D) is equal to projecting the three spectral peaks onto a line of that angle (A $\rightarrow$ B $\rightarrow$ C $\rightarrow$ D). (E) Locations of three spectral peaks (f_i) are represented by the underlying lattice spacing (λ_{1D}) and the slice angle (θ). (F) The analytical solution for slice parameters. Note that equations in (E) correspond to an overdetermined system, and the solutions in (F) are based on the first two spectral peaks. All pairs of significant spectral peaks are exploited to find the best-fit linear slice (see Experimental Procedures).

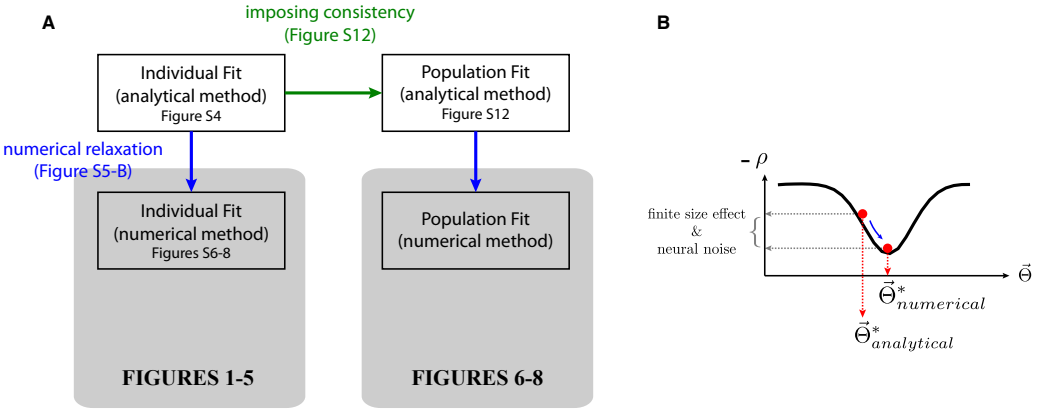


Figure S3.5: **Analytical and numerical approaches for fitting a linear slice.** (A) Fourier analytical method (Figure S3.4) is a simple and fast way of searching for best-fit linear slice of an individual cell. Nevertheless, the analytical method is limited by finite size effect (Figure S3.6) and neural noise (Figure S3.7). Thus, additional numerical optimization via a simple gradient descent rule (blue arrow; Figure S3.5-B) helps improve the quality of fit. Moreover, the dimensionality of underlying structure in PSD can be reduced by imposing consistency on peak locations of PSDs across population neurons (Figure S3.12), assuming that simultaneously recorded cells from the same putative network should share an identical structure (green arrow). (B) A schematic for how the numerical method can improve/minimize the fit-quality/negative correlation coefficient (ρ). Θ is the set of slice parameters.

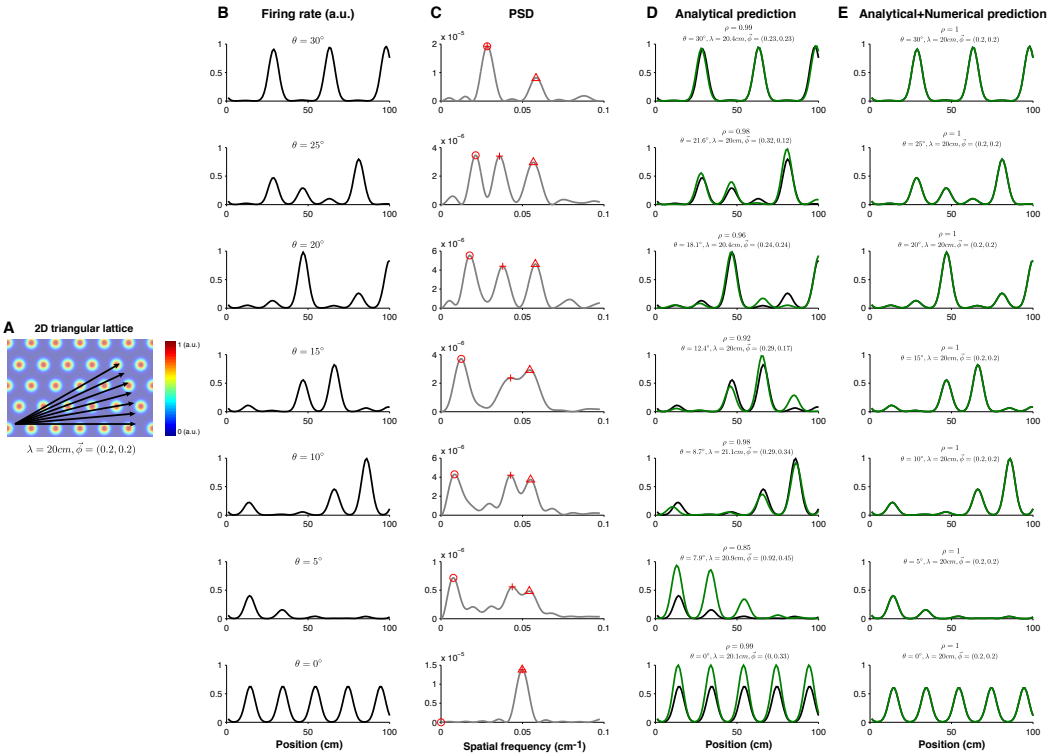


Figure S3.6: **Finite size effect on Fourier analytical method.** (A) The same set of linear slices as in Figure 3.2 but of half-length non-zero spatial phase. (B) 1D responses resulting from the linear slices. (C) The power spectral density (gray) of the 1D responses overlaid by three prominent peaks (red symbols). Note that the peaks are interfering with each other, and less separated than in Figure 3.2C. (D) Analytically predicted linear slice (green) from the spectral peak locations in (C). For comparison, the actual 1D response (black) is superimposed on the prediction, and their correlation coefficient is shown at the top together with predicted slice parameters. Note that analytical solutions have some estimation errors, especially for spatial phase $\vec{\phi}$. (E) Reconstructed responses (green) by analytical method followed by numerical optimization (see Experimental Procedures), which perfectly matches the actual 1D responses.

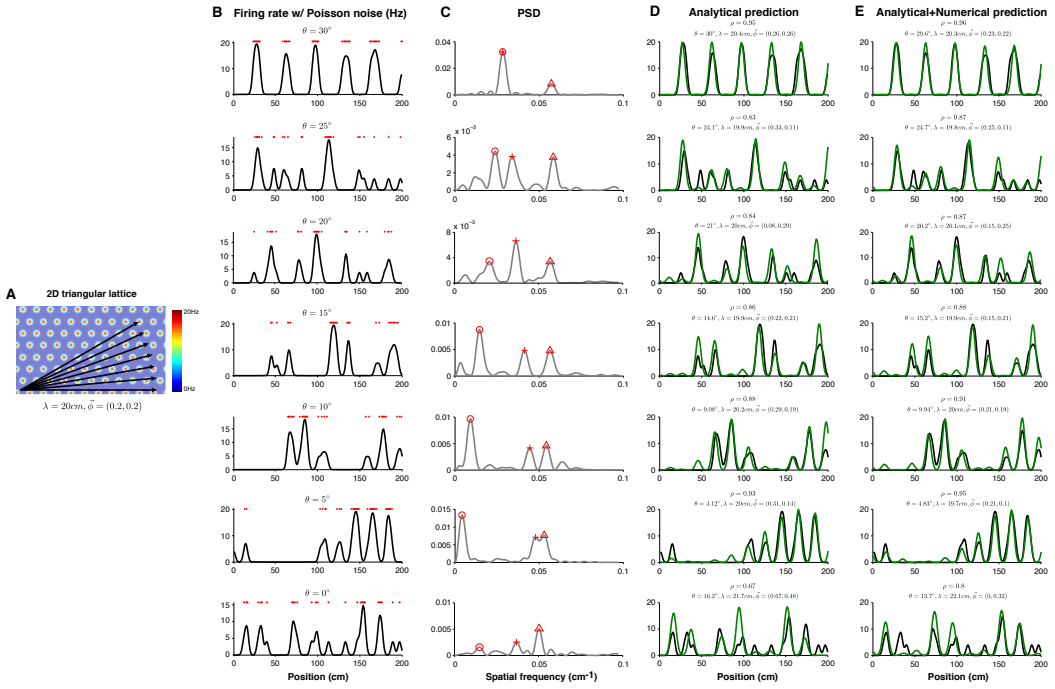


Figure S3.7: **The effect of neural variability on Fourier analytical method.** (A) The same set of linear slices as in Figure 3.2 but of a physical unit and non-zero spatial phase. (B) Noisy version of 1D responses (black) resulting from a Poisson spike train (red dots) sampled from the instantaneous firing rate (black arrow in (A)) in the unit of spikes/sec. Spatial bin size (1cm) is used as a time bin ($\Delta t = 1$ sec) to sample a spike train (red dots), and its smoothed firing rate (black) instantiates neural variability. (C) PSDs of the noisy 1D responses. (D-E) The same predictions as in Figures S6 D-E. Numerical optimization improves the predictability in terms of fit-quality as well as estimated slice parameters (see Figure S3.8 for statistical significance).

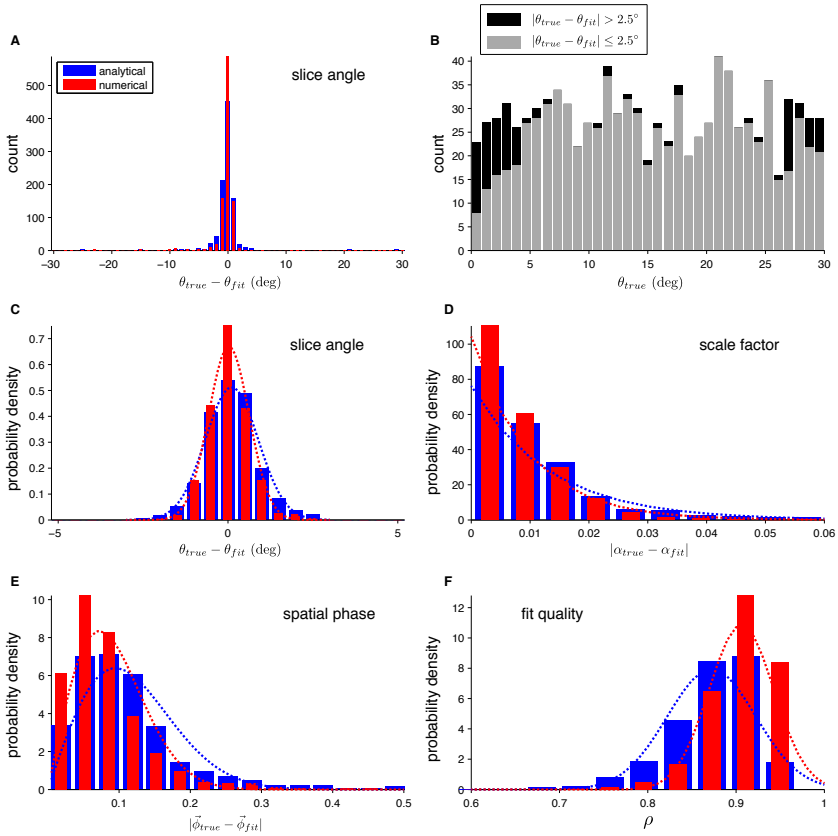


Figure S3.8: **Numerical method enhances the predictability of 1D linear slice.** (A) Histogram of estimation errors on slice orientation (blue: error by analytical method, red: error by analytical + numerical method) $N = 1000$ randomly generated linear slices given uniformly sampled slice orientations ($0° \leq \theta_{true} \leq 30°$), spatial phases ($[0, 0] \leq \vec{\phi}_{true} \leq [1, 1]$), and the rest of conditions same with Figure S3.7. (B) Distribution of the actual slice orientation with differentiating one that causes larger estimation errors ($> 2.5°$; black) from the other (gray). (C-F) Normalized histogram of estimation errors on slice orientation (C), scale factor (D), spatial phase (E), and quality of fit (F). The distributions are based on the samples that did not induce larger estimation errors on slice angle (gray bars in (B)). Numerical method lowers the estimation error on all slice parameters and improves the fit quality (C: P$= 2 \times 10^{-15}$, D: P$= 6 \times 10^{-7}$, E: P$= 5 \times 10^{-13}$, two-sample F-test, F: P$= 2 \times 10^{-20}$, two-sample t-test).

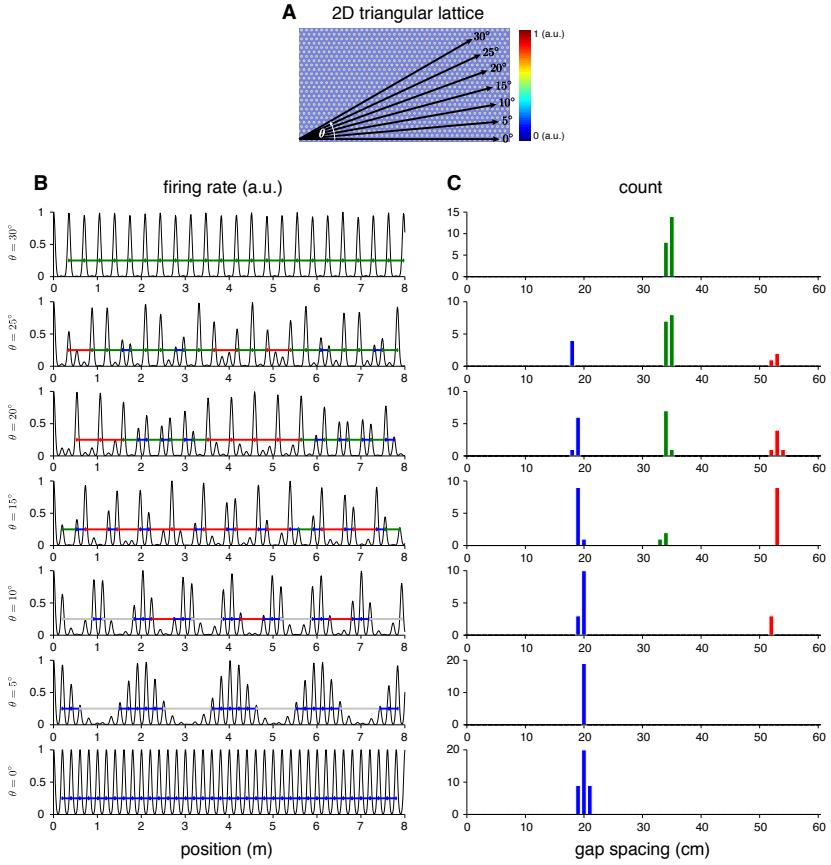


Figure S3.9: **Features of the gap distribution for linear slices of different angles.** (A) Slices of different orientations through a 2D triangular lattice as in Figure 3.2A, but four times longer than the slices in Figure 3.2A. (B) 1D responses resulting from the linear slices. Inter-peak distances are superimposed on the 1D response, color-coded by their magnitude (blue: $20cm$, green: $20\sqrt{3}cm$, red: $20\sqrt{7}cm$, and gray: the rest; peak threshold = 0.25). (C) Histogram of the gaps from 1D responses in (B). The same color code as in (B). Note that the inter-peak distances are clustered around the same positions without regard to the slice orientations, but the frequency of each cluster depends on the slice angle. Furthermore, the frequency can even change with a fixed slice orientation depending on the length and the portion of the linear slice along the orientation. For instance, when the orientation is 20 deg in (B), 1D response from 350cm to 550cm consists of the red gaps only, while the response from 160cm to 350cm consists of blue and green gaps.

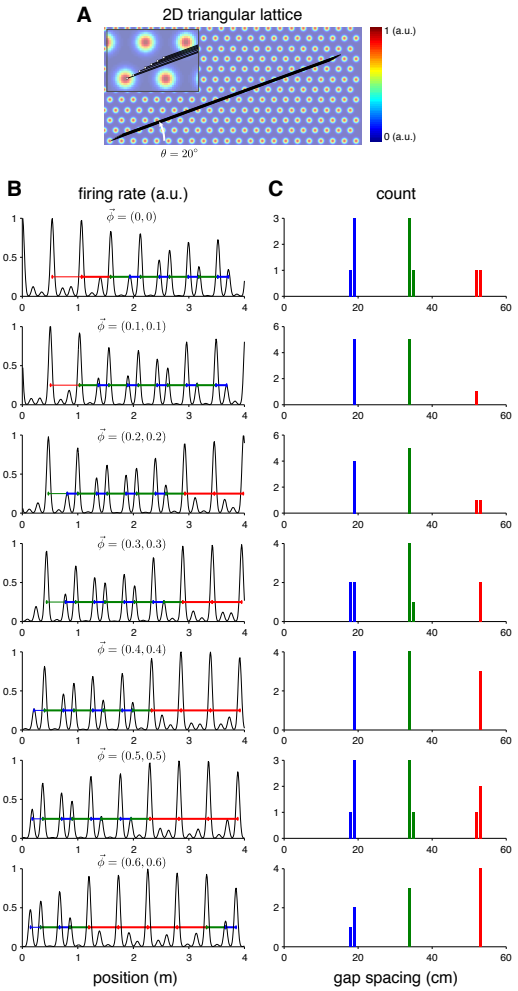


Figure S3.10: **Features of the gap distribution for parallel linear slices.** (A) Parallel slices of different spatial phases through a 2D triangular lattice as in Figure 3.2D. Inset: magnified view of the different starting points (spatial phases) of the parallel linear slices. (B) 1D responses resulting from the linear slices as in Figure S3.9-B. (C) Histogram of the gaps from 1D responses in (B). Similar to Figure S3.9-C, gaps are clustered around the same locations despite different spatial phases.

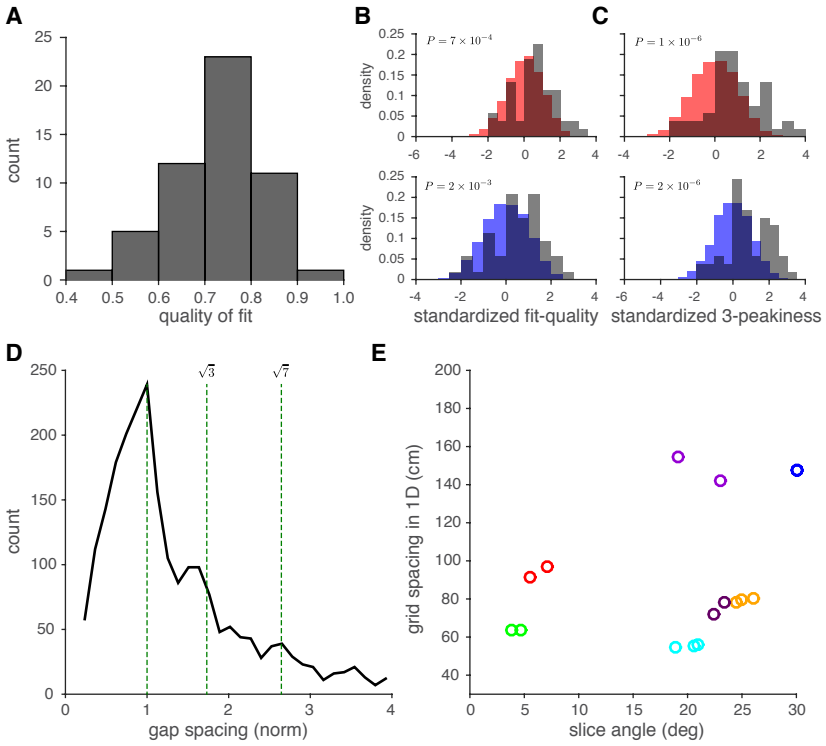


Figure S3.11: **Analysis of 1D real linear track data.** (A) Histogram of fit-quality between measured 1D tuning and predicted slice fit. (B-C) Same as in Figures 3.5B-C. p-values noted in each figure, one-sample t-test. (D) Aggregate gap distribution (black), pooled across all cells in the dataset, after the gaps in each cell's response are normalized by the estimated 1D lattice period. (E) The best-fit slice parameters for all comodular cells in the dataset, obtained for each cell individually. Cells classified as comodular are plotted in a common color.

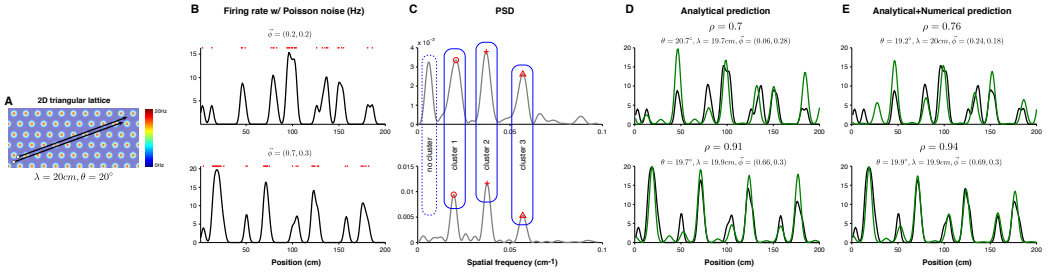


Figure S3.12: **Dimensionality reduction through imposing consistency on spectral peak identification.** (A) Parallel linear slices (of identical orientation) with different spatial phases. (B) Noisy version of 1D responses resulting from the linear slices and Poisson noise as in Figure S3.7. (C) The power spectral density of the 1D responses. Spectral peak locations (red symbols) are aligned (blue vertical bar; cluster) as in Figure 3.2F, but the first PSD (top) has an additional peak location that is missing in the second PSD (blue dashed vertical bar; no cluster). Presence of extra peaks in one but the other is due to the stochasticity of neural spike trains. Thus, imposing consistency in identifying spectral peaks from more than one instance is a good strategy to cope with neural spike variability and to search for best-fit (not strictly) parallel linear slices consistent with the assumption that the neural responses are generated by an underlying 2D pattern in the neural population (if neurons are appropriately rearranged) that shifts with animal motion, as in continuous attractor models of the grid cell system.

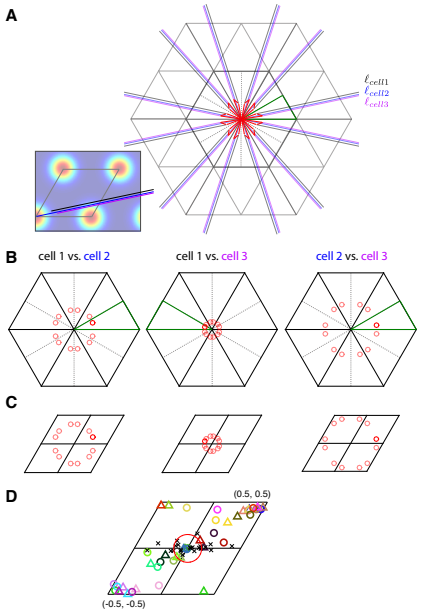


Figure S3.13: **Multiplicity of linear slices and ambiguity of the predicted 1D relative phase.** (A) Three parallel linear slices through a 2D lattice of orientation between 0° and 30° with spatial phases within a rhomboidal unit cell generate unique 1D response. The response, however, can be regenerated by reflecting the linear slice over the line of 30° and/or by rotating the linear slice through multiples of 60 degree around the origin (black/blue/purple: a triple of parallel linear slices populated up to 12 identical pairs). Nevertheless, the relative phases (red arrows) between parallel linear slices can be entirely different. (B) For each cell pair, there are 12 equivalent slice solutions, with 12 different relative phase values (red circles); these 12 values fall into 12 different domains of a hexagonal unit cell of the lattice (radial sections of the hexagon, marked by black and gray dashed lines). Allowing a choice of the best of the 12 solution domains in the pairwise relative phase prediction can produce a lower predicted total relative phase error than forcing a single common domain for all three pairwise predictions, because there is more freedom in the choice of solutions. (C) The same as (B), but for a rhomboid unit cell. (D) Same as in Figure 3.7A except that the best relative phase estimates for pairs in each comodular $K-$tuple in the dataset are not generated pairwise but by selecting a common solution domain out of the 12-fold degenerate domains for all (K choose 2) cell pairs. Histogram of relative phase prediction error magnitudes is Figure 3.7B right.

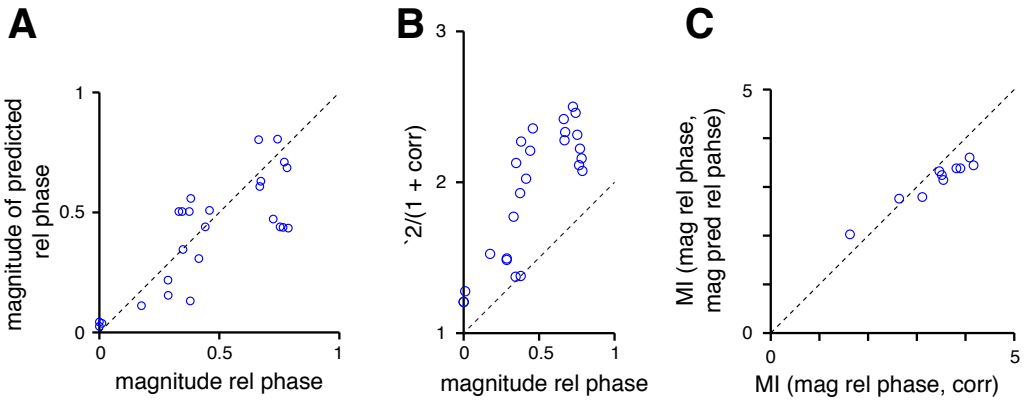


Figure S3.14: **Predictability of relative phase magnitude.** (A) Measured 2D relative phase magnitudes (abscissa) versus the magnitude of predicted values from slice analysis (ordinate). (B) Measured 2D relative phase magnitudes versus rescaled correlation coefficients between cell pairs' 1D responses. (C) Same as in Figure 3.7C, but based on data in (A-B). The magnitude of relative phase can be equally predicted by the slice analysis and correlation measure between cell pairs.

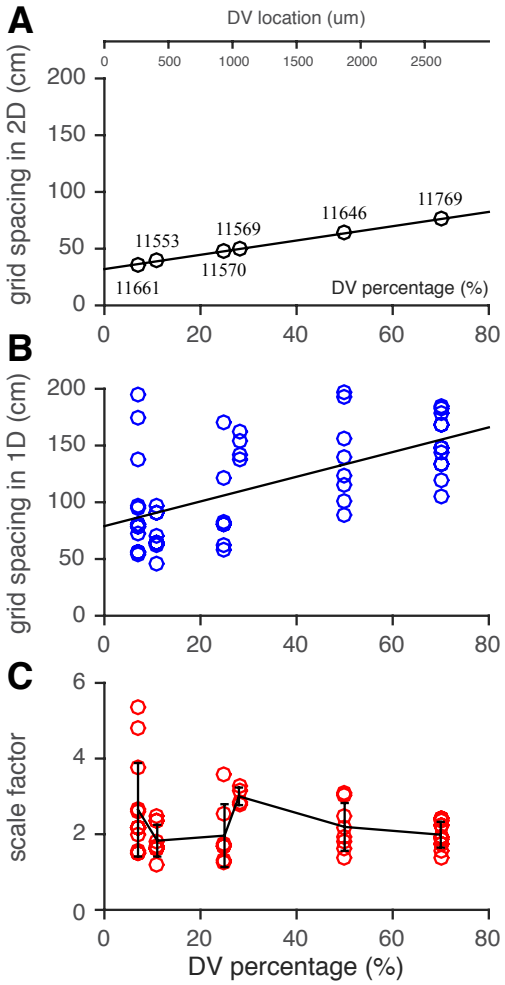


Figure S3.15: **Grid expansion from 2D to 1D real environments.** (A) Inferred grid spacing in 2D environment from the cell's dorsoventral location (Figure 1 in Brun et al. 2008) and the extrapolated linear relationship (Figure 2h in Hafting et al. 2005). Five digits are the animal number. (B) Predicted grid spacings in 1D from the 1D tuning curves (blue circles). A group of circles at the same DV location corresponds to all cells from the animal. (C) Scale factor (red circles) estimated from the measurements in (A-B).

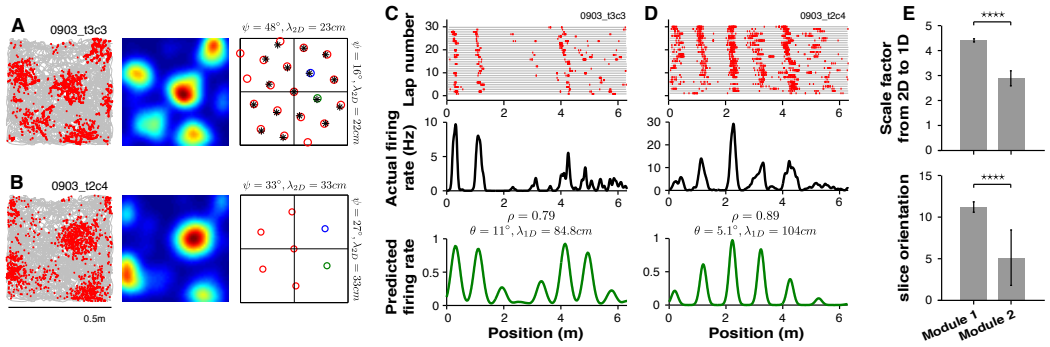


Figure S3.16: **Functional independence of grid modules.** (A-B) Two grid cells recorded simultaneously from different tetrodes in 2D open field. Left: Spike discharge map. Middle: Smoothed rate map. Right: Best-fit template lattice (red circles) to the autocorrelogram (top) and to the rate map (bottom). When grid spacing is large and firing fields are mostly cropped, we searched for the best-fit template lattice directly in rate map, which is the case of the bottom cell. Blue circle: End point of the first primary lattice vector (length and angle of the vector are shown on the top). Green circle: End point of the second lattice vector (length and angle are shown at the right). (C-D) Top: spike rasters of the same cells. Middle: Smoothed trial-average firing rate. Bottom: Rate prediction from a linear slice. (E) Both scale factor (top) and slice angle (bottom) do not change coherently during a complete environmental transition from 2D to 1D, when a cell pair is from different modules.

Chapter 4

Dynamic analysis of grid cell encoding

4.1 Introduction

Experience-dependent modification of grid cell responses has been reported in a number of experiments in both rats and mice. Boundary-induced change in shape or size of the environment can distort grid patterns [7, 88, 54], and the degree of environmental familiarity [6] or differently registered sensory cues [41, 16] can also provide notable changes in firing patterns without directly altering the environment. Such a global remapping may emerge from changing associations between the network and the external world, and is generally identified by summing the firing activity of a single neuron over a full trajectory ($\sim$ 20 min), which is an integrated measure over time of the activity of one cell. This histogram-based analysis retains the computational efficiency when we explore average spiking activity during a single entire recording session or compare two distinct trials, while it restricts studying the underlying dynamics of neurons in a temporally resolved way; there were a few attempts by parsing out a full trajectory within a short time window ($\sim$ 1-5 min) [102, 42].

To understand the network dynamics and fully investigate the predictions of a theoretical model for grid cell system, it is essential to look into

fine temporal dynamics at the resolution of the same order of magnitude as the neural spiking. To be specific, the maintenance of general periodic structure in grid cells across different environmental contexts and even in complete darkness suggest that grid cells may be the neural loci of path integration [41]. However, additive nature of continuous integration of self-motion cues results in an accumulation of error because an error in the current position estimate will be propagated to all the following estimates without any correction mechanism. In such models, a natural question is whether there exists any evidence of error accumulation [42]. Experimentally recorded grid cell data, however, appear relatively stable over a few tens of minutes of recording session, implying that either there is no accumulation of error, there is an ongoing error-correction mechanism, or there is an accumulation of error that cannot easily be seen by the naked eye – the current histogram-based approach – but be captured in the microscopic scale.

To study fine temporal dynamics and examine the error accumulation (and possibly correction) of grid cells, we apply a linear state-space model observed through a point process [28] for characterizing the temporal evolution of spatial firing patterns, or *spatial tuning curves*. The point process adaptive filtering framework aims not only to estimate the system parameters at the same temporal resolution of the spiking ($\sim$ 1-10 msec), but also to infer how diffusive the grid cell system would be. Here we show that the mismatch between the actual and the internal estimate of an animal's positions over time can be well kept track of by this modeling framework given simulated

data from a continuous attractor network model of grid cells. We find that the model allows us to accurately infer to what extent the spatial tuning of grid cells undergoes diffusion over time. Additionally, we pilot the method on grid cell data recorded in both familiar and novel environments, where in the first exposure to a novel environment, spatial firing patterns of grid cells expand in size and reduce in regularity [6]. Environmental novelty may be indicative of less frequent or less confident corrective inputs to the grid cell network. Thus, we show how the model can be used to infer and compare the amount of systematic drift between two environments.

4.2 Methods

4.2.1 State-space point-process model

In the context of attractor-based network model of grid cells, an internal estimate of an animal's position s_t is updated from the past estimate s_{t-1} through integrating the velocity input v_t at each time point. The stochastic nature of neural responses, along with heterogeneous synaptic weights and a finite number of neurons, introduces sources of error ϵ_d that can cause the diffusion in the positional estimates s_t [18]:

$$s_t = s_{t-1} + v_t \Delta_t + \epsilon_d \tag{4.1}$$

$$\epsilon_d \sim \mathcal{N}(0, \sigma_d^2) \tag{4.2}$$

where σ_d determines the amount of noise for diffusion process.

Individual neurons are modulated by the internal estimate of position

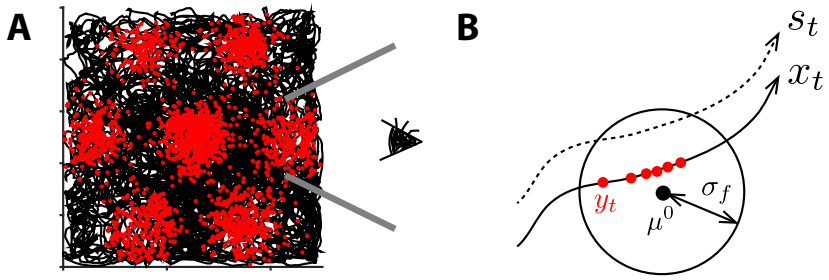


Figure 4.1: (A) Trajectory (black) of an animal foraging in a square enclosure, with spike locations (red dots) for a grid cell. (B) A schematic of the grid cell spiking (red dots, y_t) over a short time interval while the animal traverses one of the firing fields (black circle; field width σ_f; field width μ^0) along a movement trajectory (black curve; instantaneous position x_t). According to the continuous attractor model, grid cells are not modulated by the actual position x_t but by the internal estimate of the animal's position s_t, which could be away from each other through diffusion process.

s_t, and for the unobservable or latent stimulus s_t, grid cell spiking y_t is modeled as an inhomogeneous poisson process:

$$y_t \sim Poiss[\lambda_t \Delta_t] = Poiss[f(s_t;\theta)\Delta_t] \tag{4.3}$$

where the firing rate is

$$\begin{aligned} \lambda_t &= f(s_t;\theta) && (4.4)\\ &= \exp\left\{\alpha - \frac{1}{2\sigma_f^2}(s_t-\mu^i)^T(s_t-\mu^i)\right\} && (4.5) \end{aligned}$$

$\exp(\alpha)$ is the maximum firing rate, μ^i is the center of the i-th firing field most adjacent to s_t, σ_f is the field width, and $\theta = [\alpha\ \mu^0\ \sigma_f]^T$. $\mu^{i>0}$ is deterministic by treating them as the outcome of populating two primary lattice vectors of grid cells centered on μ^0 (Figure 4.1).

This latent variable model can be useful when there is enough number of neurons with sufficiently distinct spatial tuning phases to span the entire space of the latent variable. However, it is only possible to take simultaneous recordings from a limited number of grid cells; more importantly, how many neurons we need to accurately reconstruct s_t is a unanswered question. For that reason, we introduce an alternative state-space model that does not require to directly infer the latent variables s_t.

4.2.2 Point-process adaptive filter

We reformulate the previous model equations by replacing the latent stimulus s_t by the actual position of the animal x_t and transferring the source of stochasticity in s_t to the set of tuning curve parameters θ (Figure 4.2):

$$\theta_t = \theta_{t-1} + \epsilon_d \tag{4.6}$$

$$\epsilon_d \sim \mathcal{N}(0, Q) \tag{4.7}$$

$$y_t \sim Poiss[\lambda_t \Delta_t] = Poiss[f(x_t; \theta_t)\Delta_t] \tag{4.8}$$

where Q is the covariance matrix whose diagonal is the noise variance of each parameter in $\theta = [\alpha\ \mu^0\ \sigma_f]^T$, and the firing rate is

$$\lambda_t = f(x_t; \theta_t) \tag{4.9}$$

$$= \exp\left\{\alpha_t - \frac{1}{2(\sigma_t^f)^2}(x_t - \mu_t^i)^T(x_t - \mu_t^i)\right\} \tag{4.10}$$

This point process adaptive filter model was used in a detailed study of receptive field plasticity in the hippocampal CA1 or the entorhinal cortex

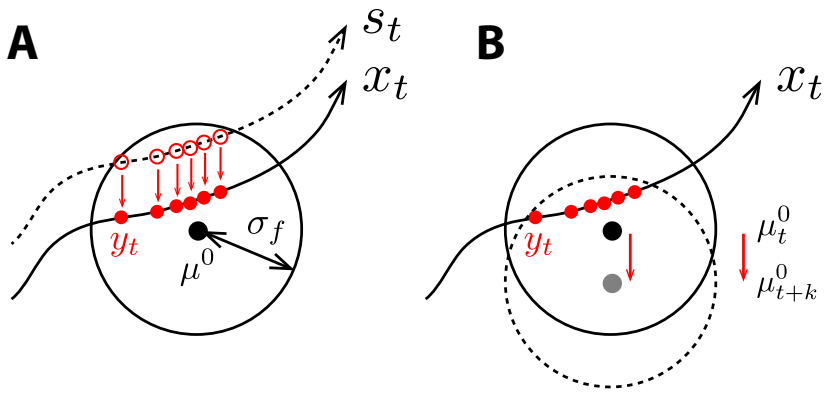


Figure 4.2: (A) A model for highlighting that grid cell responses are not activated by the actual position of an animal x_t but by the internal position estimate s_t, relative to the center of a firing field μ^0. The current spiking y_t (red dots) can be interpreted as the translocated activity (red circles) along the trajectory of internal position estimates. (B) Alternative model of the spiking events y_t. Grid cell activity is triggered by the firing field that shifts by the same amount (red arrow) through a random walk.

to track the evolution of these receptive fields on a millisecond timescale [14, 31, 28]. We aim to infer how s_t diverges from (or converges on) x_t over time by inferring the center of firing fields in a temporally resolved way, from which we examine whether the system is diffusive or not.

In this model, θ_t is a time-varying parameter vector to be estimated at each time point, and the estimates of θ_t will be based on its posterior density, conditioned on the set of past observations: $p(\theta_t|y_t, H_t)$ where $H_t = [\theta_{1:t-1}, x_{1:t}, y_{1:t-1}]$. This posterior density evolves with time and with each observation. Tracking this evolution allows for real-time estimation of the parameters of λ_t based on all of the spiking data observed up to the current

time. The recursive algorithm for updating the posterior mean and variance is given by [28]:

$$\theta_{t|t-1} = \theta_{t-1|t-1} \tag{4.11}$$

$$W_{t|t-1} = W_{t-1|t-1} + Q \tag{4.12}$$

$$(W_{t|t})^{-1} = (W_{t|t-1})^{-1} + \left[\left(\frac{\partial \log \lambda}{\partial \theta_t}\right)' (\lambda \Delta_t) \left(\frac{\partial \log \lambda}{\partial \theta_t}\right) - (y_t - \lambda\Delta_t)\left(\frac{\partial^2 \log \lambda}{\partial \theta_t \partial \theta_t'}\right)\right]_{\theta_{t|t-1}} \tag{4.13}$$

$$\theta_{t|t} = \theta_{t|t-1} + W_{t|t}\left[\left(\frac{\partial \log \lambda}{\partial \theta_t}\right)' (y_t - \lambda\Delta_t)\right]_{\theta_{t|t-1}} \tag{4.14}$$

where $\theta_{t|t-1}$ and $W_{t|t-1}$ are the mean and variance of the one-step prediction density, $\theta_{t|t}$ and $W_{t|t}$ are the posterior mean and variance, respectively. The first and second derivative terms in (4.13) and (4.14) are evaluated as:

$$\left.\frac{\partial \log \lambda}{\partial \theta}\right|_{\theta_{t|t-1}} = \begin{bmatrix} 1 \\ \sigma_{t|t-1}^{-2}(x_t - \mu_{t|t-1}^i) \\ \sigma_{t|t-1}^{-3}(x_t - \mu_{t|t-1}^i)^T(x_t - \mu_{t|t-1}^i) \end{bmatrix}$$

$$\left.\frac{\partial^2 \log \lambda}{\partial \theta \partial \theta'}\right|_{\theta_{t|t-1}} = \begin{bmatrix} 0 & \mathbf{0}^{1\times 2} & 0 \\ \mathbf{0}^{2\times 1} & -\sigma_{t|t-1}^{-2}\mathbf{I}^{2\times 2} & -2\sigma_{t|t-1}^{-3}(x_t - \mu_{t|t-1}^i) \\ 0 & -2\sigma_{t|t-1}^{-3}(x_t - \mu_{t|t-1}^i)^T & -3\sigma_{t|t-1}^{-4}(x_t - \mu_{t|t-1}^i)^T(x_t - \mu_{t|t-1}^i) \end{bmatrix}$$

4.3 Results

4.3.1 Model fitting in simulations

To examine the performance of our point-process adaptive filter, we fit the parameters to simulated spike trains from the same model with known parameters and from the continuous attractor model of grid cells. The first

simulated cells exhibited a canonical firing pattern of grid cells, with model parameters selected to reproduce the spatial tuning of grid cells. The second simulated cells qualitatively mimicked experimental grid cell data through sufficiently strong inhibitory interactions such that the connectivity produced a regular pattern of neural activity [18]. The stimulus consisted of an experimentally recorded two dimensional position estimates from a rat, binned at a millisecond resolution. We validated our point-process filter by examining error in the estimated values of parameters, or comparing the mismatch between the actual position x_t and the decoded estimate s_t to the change in field centers over time. With different conditions of model parameters, the parameter estimates reconstructed the true dynamics of spatial tuning for both simulated cells. Therefore, the filtering algorithm can reliably fit the generative model to grid cell spiking data, despite the fact that the model does not have a network dynamics in the standard mechanistic model of grid cells.

4.3.2 Inferring model parameters from spikes of a descriptive model

We fit our model (4.6-10) to spike trains ($\Delta_t = 20$ ms) obtained exactly from the same model with the model parameters fixed: grid period $\lambda = 20$cm, maximum firing rate $\exp(\alpha) = 20$Hz, and the field width $\sigma_f = \lambda/3$ cm (Figure 4.3A). However, the center of firing fields was allowed to vary over time through a Gaussian random walk by setting $\mu^0_{t=0} = [25, 25]$cm and $Q = \text{diag}[0, 0.05, 0.05, 0]$ (Figure 4.3B).

We tracked the center of a firing field μ^0_t over ~ 10 min recording session

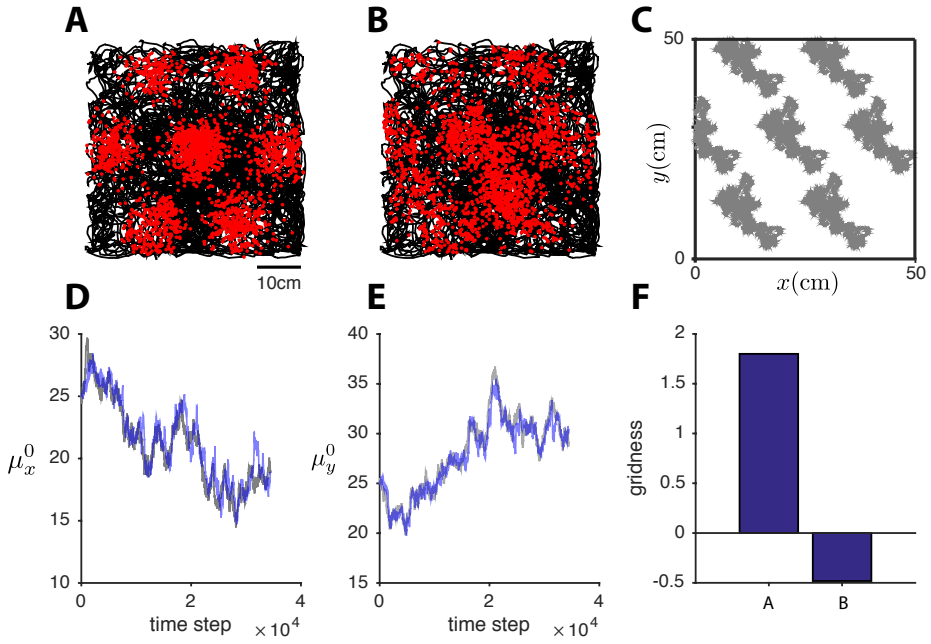


Figure 4.3: (A) Example trajectory (black) with simulated grid cell spikes in red. (B) Grid cell responses after adding a random walk in the center of firing fields. (C) The two-dimensional random walk of the firing field locations (gray) embedded in (B). (D-E) The trajectory of the random walk in each coordinate (gray) and the decoded field locations (blue). (F) Grid score measured in (A) and (B).

and compared with the estimates by the filtering algorithm. The point-process adaptive filter correctly predicted the diffusive nature of the firing field location simulated in the model (Figure 4.3C-E). To visualize the estimated field centers alongside the evolution of a simulated random walk, we decomposed the 2-dimensional coordinates of the field center into two scalars. The notable tracking even under the rapid evolution in some places was ascribed to the adaptive learning rate that was proportional to the one-step prediction vari-

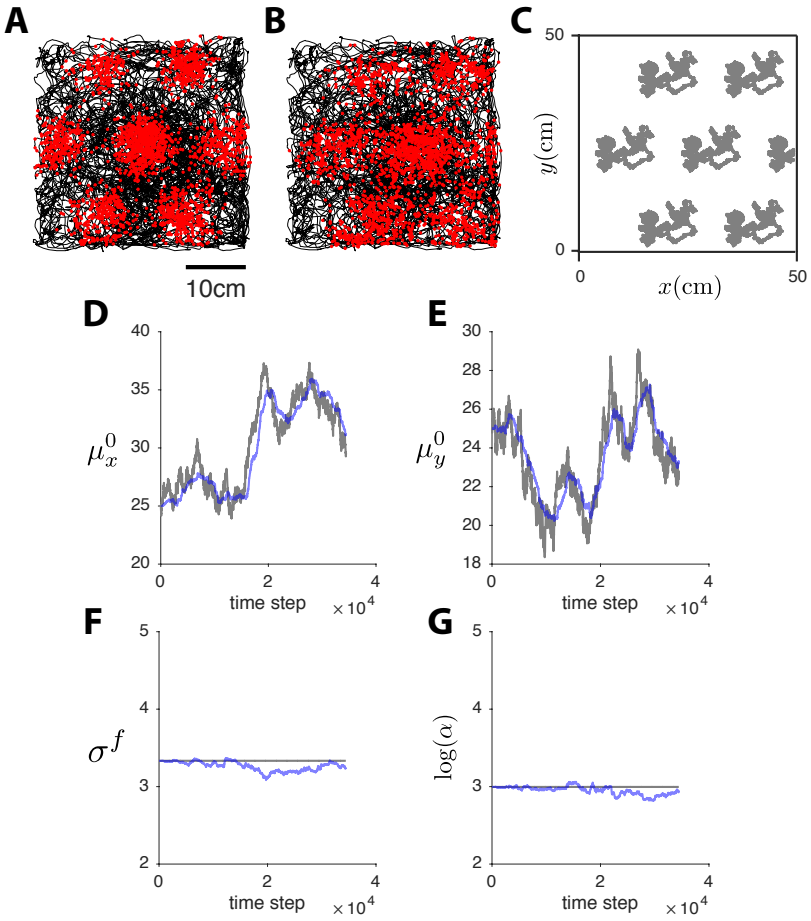


Figure 4.4: (A-E) Same as Figure 4.3 for another random walk in the center of firing fields. The inference is based on purposely approximated system noise variances. (F-G) Decoded field width and log of maximum firing rate (blue).

ance (4.13). The point-process filter scaled up its relatively low learning rates right after the jump, and then scaled them down rapidly to continue to track correctly the parameter by exploiting the error between the instantaneous firing rate estimate and the current spiking activity [28].

Decoding the parameters of a spatial tuning curve requires a covariance matrix Q *a priori* for updating the variance of the one-step prediction density (4.12). We applied exactly the same covariance Q as the one used in generating simulated spike trains in Figure 4.3A, but we now relaxed the constraint by setting $Q = \text{diag}[10^{-3}, 10^{-2}, 10^{-2}, 10^{-3}]$ to understand the dependence of the model on the noise variance. The reduced variance of μ_t^0 made the posterior mean change less swiftly, although the estimates of the field center were not qualitatively different from the ground truth (Figure 4.4D-E). The positive variance of σ_t^f and α_t constantly fed into (4.12) as opposed to zero produced a slight fluctuation in their posterior mean, but the field width and the maximum firing rate were fairly stable over time (Figure 4.4F-G). Nonetheless, a sizeable discrepancy between the system noise variance and our empirical initial value of Q could lead to inaccurate estimates (now shown), which is discussed in the next chapter.

4.3.3 Inferring model parameters from spikes of an attractor network model

We fit the point-process filter model to a population of grid cells simulated from a continuous attractor network model in response to velocity inputs. In this mechanistic model, grid cells are modeled as a network of neurons with asymmetrically centered inhibitory center-surround synaptic weight profiles. These neurons form a two-dimensional neural sheet where a grid activity pattern emerges in the presence of feedforward excitation. Individual grid cell responses are then generated when an animal's velocity is coupled to the

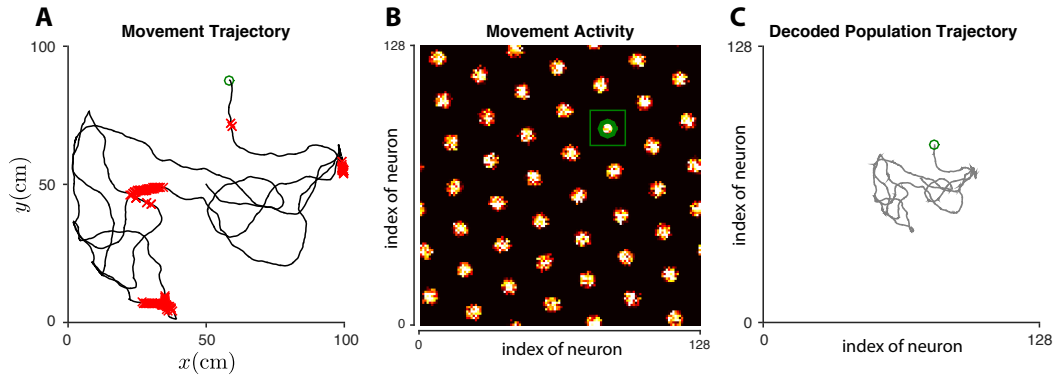


Figure 4.5: (A) Movement trajectory (black) of an animal (green circle) over a few seconds superimposed by spike locations (red) of a grid cell. (B) A snapshot of the population activity, when the network is driven by the animal's velocity input. Displacement of the population pattern is recorded by keeping track of a center of mass (green circle) from neurons within a 11×11 patch (green square). (C) Decoded population trajectory (gray).

movement of population activity pattern [18, 32].

We first performed a population decoding in the neural sheet to reconstruct the trajectory of internal estimates of the animal's position. An instantaneous displacement of an activity bump was decoded by computing the center of mass from the activity states and the absolute locations of 121 neurons within a 11×11 patch centered around the activity bump (Figure 4.5). The neural trajectory followed by an appropriate rescaling was compared to the actual movement trajectory of the animal to reconstruct the random walk subsumed in the velocity-to-position integrator (Figure 4.6A-C). We next examined the drift in the population state to assess the effect of noise on integration accuracy. Quantitatively, the mean squared error of the internal

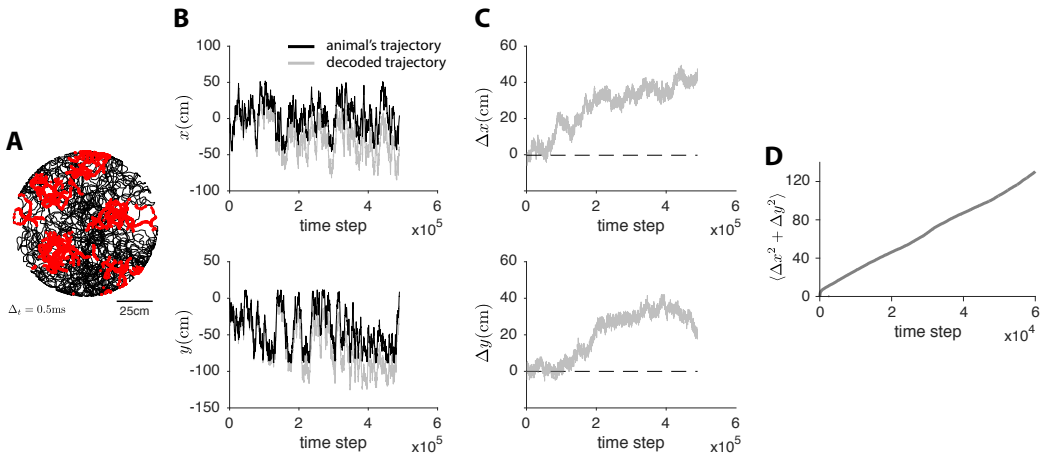


Figure 4.6: (A) Example spatial response of a grid cell simulated from a continuous attractor model. (B) Animal's movement trajectory (black) shown as x- (top) and y-components (bottom). Decoded trajectory of the population pattern (gray) is overlaid on the figure. (C) Difference between the animal's trajectory and the decoded trajectory (gray). (D) Mean squared error of the internal position estimate in a short time interval up to $\sim$ 1 minute.

position estimates in a time interval Δ_t was closely proportional to Δ_t (Figure 4.6D), which was consistent with the prediction of noisy spiking network [18].

Our adaptive filter model was fit to the $\sim$ 5 minutes of spike train recordings from the simulated grid cells. Two simultaneously recorded cells were randomly selected and their field centers were decoded independently but with enforcing consistency in the field width and the maximum firing rate. The latter constraint is consistent with the attractor hypothesis in that any cells in the same network share essentially identical values of the tuning curve parameters. Although the model was fit to each cell independently, the decoded center of the firing fields of two cells was well-matched to each other

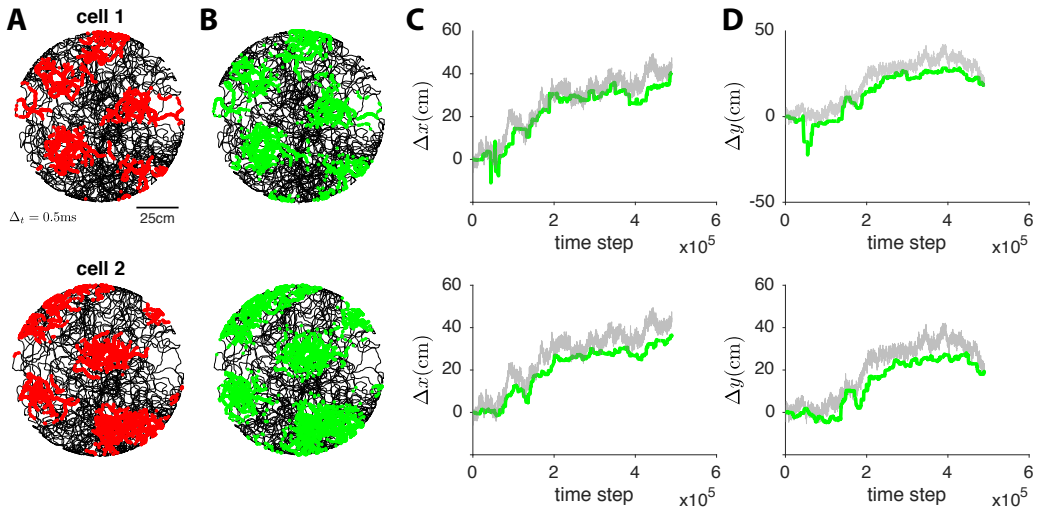


Figure 4.7: (A) Example spatial response of a grid cell simulated from a continuous attractor model. (B) Animal's movement trajectory (black) shown as x- (top) and y-components (bottom). Decoded trajectory of the population pattern (gray) is overlaid on the figure. (C) Difference between the animal's trajectory and the decoded trajectory (gray). (D) Mean squared error of the internal position estimate in a short time interval up to $\sim$ 1 minute.

(Figure 4.7). Furthermore, the estimated field center of an individual cell was not qualitatively different from the drift of the internal position estimate inferred from the population decoding (Figure 4.7), which was also consistent with the attractor hypothesis. The results of our analysis confirmed that the point-process filter model can infer the dynamic changes in spatial tuning curves of grid cell over time.

4.4 Conclusions

We have utilized a computationally tractable method for tracking the temporal evolution of tuning curves from spike train observations. The formulation involved modeling a probabilistic model of neural spiking along with a state-space model using the Chapman-Kolmogorov equation in a Bayesian framework to obtain a recursive expression for the posterior distribution of parameters to be tracked. We have applied this method to determine the dynamics underlying grid cell activity in the entorhinal cortex. According to the predictions of the attractor model, grid cells accumulate error by the nature of integration and stochastic neural responses. Thus, we should find the evidence of error accumulation over a short time, even under an error correction mechanism. Our results showed that we can infer the amount and behavior of the drifting error through the adaptive filtering algorithm. The following work with this model is to investigate the same problems with experimentally recorded grid cells. Specifically, the work includes analyzing grid cells recorded in both familiar and novel environments to examine how differently the error is accumulated. It aims to understand whether the familiarity of the external cues plays a role as a source of correction mechanism.

Future direction will be required to establish a solution to the dependence of the model on initial values of the parameter vector. We observed that the decoding performance can deteriorate when the covariance Q or the value of parameters at $t = 0$ is far from the ground-truth. Future work should aim to determine which model best describes actual neural data by applying

a classic Bayesian model comparison – to quantify the ratio of the marginal likelihoods of the data given two models, $p(\mathbf{y}|\mathcal{M}_1)/p(\mathbf{y}|\mathcal{M}_2)$.

Chapter 5

Conclusions

The computational and mathematical approaches presented in this dissertation aimed to unravel grid cell dynamics during navigation in general. The grid cell system offered the opportunity to study a cognitively meaningful neural computation at a mechanistic level. Here we leveraged unique datasets together with newly developed algorithms to examine integrative properties of grid cells in three aims.

In Chapter 2, the main implication was that, while grid cells may change their properties under the remapping such as when physically stretching the environment or under conditions of novelty, these changes must be made in tandem, such that the ratio between different grid parameters must remain stable between pairs of cells, before and after remapping. Our analyses strongly suggests that the multi-dimensional network activity of grid cells is embedded into a stable two-dimensional continuous attractor manifold. In addition, we showed that the animal's velocity could perturb the state of the network away from the stable state, in agreement with the important role that the velocity of the animal plays in most theoretical models of grid cells.

In Chapter 3, the responses of grid cells on 1D tracks were character-

ized based on the analytical properties of the Fourier transform of linear slices through regular lattices. We established that despite their irregular and aperiodic appearance, 1D slices through idealized 2D lattices will have a canonical 1D Fourier power spectrum, and the responses will exhibit a characteristic gap distribution with gap ratios determined by the 2D lattice geometry. We further showed that cells from the same module are well-fit by parallel slices through the same regular lattice, and moreover, it is possible to predict the 2D relative phase of cells from slice fits to their 1D track recordings. These successes strongly suggest that the network remains in fundamentally the same dynamical regime.

In Chapter 4, we piloted a state-space point-process model to infer the drift in a network state of population neurons. The statistical inference method could reconstruct the random walk subsumed in the grid cell system as a path-integrator, and the result was consistent with the "square root of time" rule for the errors of random walk. Finally, the random walks decoded from two grid cells recorded simultaneously showed coherent changes over time.

Bibliography

[1] E Aksay, G Gamkrelidze, H S Seung, R Baker, and D W Tank. In vivo intracellular recording and perturbation of persistent activity in a neural integrator. *Nature Neuroscience*, 4(2):184–193, 2001.

[2] E Aksay, G Major, M S Goldman, R Baker, H S Seung, and D W Tank. History dependence of rate covariation between neurons during persistent activity in an oculomotor integrator. *Cerebral Cortex*, 13(11):1173–1184, 2003.

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

[4] Michael Artin. *Algebra*. Prentice Hall, 1991 (first edition).

[5] Stephen A Baccus and Markus Meister. Fast and slow contrast adaptation in retinal circuitry. *Neuron*, 36(5):909–919, 2002.

[6] C. Barry, L.L. Ginzberg, J. O'Keefe, and N. Burgess. Grid cell firing patterns signal environmental novelty by expansion. *Proceedings of the National Academy of Sciences*, 109(43):17687–17692, 2012.

[7] Caswell Barry, Robin Hayman, Neil Burgess, and Kathryn J Jeffery. Experience-dependent rescaling of entorhinal grids. *Nature Neuroscience*, 10(6):682–684, 2007.

[8] R. Ben-Yishai, R.L. Bar-Or, and H. Sompolinsky. Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, 92(9):3844, 1995.

[9] H T Blair and P E Sharp. Anticipatory head direction signals in anterior thalamus: evidence for a thalamocortical circuit that integrates angular head motion to compute head direction. *J. Neurosci.*, 15(9):6260–6270, 1995.

[10] Hugh T Blair, Kishan Gupta, and Kechen Zhang. Conversion of a phase- to a rate-coded position signal by a three-stage model of theta cells, grid cells, and place cells. *Hippocampus*, 18(12):1239–55, 2008.

[11] Tora Bonnevie, Benjamin Dunn, Marianne Fyhn, Torkel Hafting, Dori Derdikman, John L Kubie, Yasser Roudi, Edvard I Moser, and May-Britt Moser. Grid cells require excitatory drive from the hippocampus. *Nature Neuroscience*, 16:309–317, 2013.

[12] E. Bostock, R.U. Muller, and J.L. Kubie. Experience-dependent modifications of hippocampal place cell firing. *Hippocampus*, 1(2):193–205, 1991.

[13] Mark P Brandon, Andrew R Bogaard, Christopher P Libby, Michael A Connerney, Kishan Gupta, and Michael E Hasselmo. Reduction of theta rhythm dissociates grid cell spatial periodicity from directional tuning. *Science*, 332(6029):595–9, Apr 2011.

[14] Emery N Brown, David P Nguyen, Loren M Frank, Matthew A Wilson, and Victor Solo. An analysis of neural receptive field plasticity by point process adaptive filtering. *Proceedings of the National Academy of Sciences*, 98(21):12261–12266, 2001.

[15] Vegard Heimly Brun, Stefan Leutgeb, Hui-Qiu Wu, Robert Schwarcz, Menno P Witter, Edvard I Moser, and May-Britt Moser. Impaired spatial representation in CA1 after lesion of direct input from entorhinal cortex. *Neuron*, 57(2):290–302, 2008.

[16] V.H. Brun, T. Solstad, K.B. Kjelstrup, M. Fyhn, M.P. Witter, E.I. Moser, and M-B Moser. Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. *Hippocampus*, 18(12):1200–1212, 2008.

[17] Yoram Burak and Ila Fiete. Do we understand the emergent dynamics of grid cell activity? *J. Neurosci.*, 26(37):9352–9354, 2006.

[18] Yoram Burak and Ila R Fiete. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput. Biol.*, 5(2):e1000291, 2009.

[19] Yoram Burak and Ila R Fiete. Fundamental limits on persistent activity in networks of noisy neurons. *Proceedings of the National Academy of Sciences*, 109(43):17645–17650, 2012.

[20] Neil Burgess. Grid cells and theta as oscillatory interference: theory and predictions. *Hippocampus*, 18(12):1157–1174, 2008.

[21] Neil Burgess, Caswell Barry, and John O'Keefe. An oscillatory interference model of grid cell firing. *Hippocampus*, 17(9):801–812, 2007.

[22] C. Cannon and D.A. Robinson. Loss of the neural integrator of the oculomotor system from brain stem lesions in monkey. *J. Neurophysiol.*, 57(5):1383–1409, May 1987.

[23] Jonathan J Couey, Aree Witoelar, Sheng-Jia Zhang, Kang Zheng, Jing Ye, Benjamin Dunn, Rafal Czajkowski, May-Britt Moser, Edvard I Moser, Yasser Roudi, et al. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nature Neuroscience*, 16:318–324, 2013.

[24] Charles Darwin. Origin of certain instincts. *Nature*, 7:417–418, 1873.

[25] Dori Derdikman, Jonathan R Whitlock, Albert Tsao, Marianne Fyhn, Torkel Hafting, May-Britt Moser, and Edvard I Moser. Fragmentation of grid cell maps in a multicompartment environment. *Nat Neurosci*, 12(10):1325–1332, 2009.

[26] Robert Desimone, Thomas D Albright, Charles G Gross, and Charles Bruce. Stimulus-selective properties of inferior temporal neurons in the macaque. *The Journal of Neuroscience*, 4(8):2051–2062, 1984.

[27] Cristina Domnisoru, Amina A Kinkhabwala, and David W Tank. Membrane potential dynamics of grid cells. *Nature*, 495:199–204, 2013.

[28] Uri T Eden, Loren M Frank, Riccardo Barbieri, Victor Solo, and Emery N Brown. Dynamic analysis of neural encoding by point process adaptive filtering. *Neural computation*, 16(5):971–998, 2004.

[29] Ila R Fiete, Yoram Burak, and Ted Brookings. What grid cells convey about rat location. *J. Neurosci.*, 28(27):6856–6871, 2008.

[30] L.M. Frank, G.B. Stanley, and E.N. Brown. Hippocampal plasticity across multiple days of exposure to novel environments. *J. Neurosci.*, 24(35):7681–7689, 2004.

[31] Loren M Frank, Uri T Eden, Victor Solo, Matthew A Wilson, and Emery N Brown. Contrasting patterns of receptive field plasticity in the hippocampus and the entorhinal cortex: an adaptive filtering approach. *The Journal of neuroscience*, 22(9):3817–3830, 2002.

[32] Mark C Fuhs and David S Touretzky. A spin glass model of path integration in rat medial entorhinal cortex. *J. Neurosci.*, 26(16):4266–4276, 2006.

[33] Marianne Fyhn, Torkel Hafting, Alessandro Treves, May-Britt Moser, and Edvard I Moser. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, 446(7132):190–194, 2007.

[34] G. Ginosar, A. Finkelstein, L. Las, and N. Ulanovsky. 3-d grid cells in flying bats. *Soc. Neurosci. Abstr.*, 94.22, 2014.

[35] G. Ginosar, A. Finkelstein, A. Rubin, L. Las, and N. Ulanovsky. 3d grid cells and border cells in flying bats. *Society for Neuroscience (SfN Abstracts)*, 2015.

[36] Katalin M Gothard, William E Skaggs, and Bruce L McNaughton. Dynamics of mismatch correction in the hippocampal ensemble code for space: interaction between path integration and environmental cues. *The Journal of neuroscience*, 16(24):8027–8040, 1996.

[37] Charles G Gross, CE Rocha-Miranda, and DB Bender. Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of neurophysiology*, 35(1):96–111, 1972.

[38] Alexis Guanella, Daniel Kiper, and Paul Verschure. A model of grid cells based on a twisted torus topology. *Int. J. Neural. Syst.*, 17(4):231–240, 2007.

[39] Kishan Gupta, Nathan J Beer, Lauren A Keller, and Michael E Hasselmo. Medial entorhinal grid cells and head direction cells rotate with

a t-maze more often during less recently experienced rotations. *Cerebral Cortex*, 24(6):1630–1644, 2014.

[40] T Hafting, M Fyhn, T Bonnevie, MB Moser, and EI Moser. Hippocampus-independent phase precession in entorhinal grid cells. *Nature*, 453:1248–1252, 2008.

[41] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E.I. Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, 2005.

[42] Kiah Hardcastle, Surya Ganguli, and Lisa M Giocomo. Environmental boundaries as an error correction mechanism for grid cells. *Neuron*, 86(3):827–839, 2015.

[43] Michael E Hasselmo. Grid cell mechanisms and function: contributions of entorhinal persistent spiking and phase resetting. *Hippocampus*, 18(12):1213–29, 2008.

[44] Michael E Hasselmo, Lisa M Giocomo, and Eric A Zilli. Grid cell firing may arise from interference of theta frequency membrane potential oscillations in single neurons. *Hippocampus*, 17(12):1252–1271, 2007.

[45] Robin Hayman, Madeleine A Verriotis, Aleksandar Jovalekic, André A Fenton, and Kathryn J Jeffery. Anisotropic encoding of three-dimensional space by place cells and grid cells. *Nature neuroscience*, 14(9):1182–1188, 2011.

[46] James G Heys, Krsna V Rangarajan, and Daniel A Dombeck. The functional micro-organization of grid cells revealed by cellular-resolution imaging. *Neuron*, 84(5):1079–1090, 2014.

[47] Daniel N Hill, Samar B Mehta, and David Kleinfeld. Quality metrics to accompany spike sorting of extracellular signals. *The Journal of Neuroscience*, 31(24):8699–8705, 2011.

[48] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554, 1982.

[49] Joshua Jacobs, Christoph T Weidemann, Jonathan F Miller, Alec Solway, John F Burke, Xue-Xin Wei, Nanthia Suthana, Michael R Sperling, Ashwini D Sharan, Itzhak Fried, et al. Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature Neuroscience*, 2013.

[50] Mattias P Karlsson and Loren M Frank. Network dynamics underlying the formation of sparse, informative representations in the hippocampus. *J. Neurosci.*, 28(52):14271–14281, 2008.

[51] Felix Maria Kempf, Alexander Mathis, Martin Stemmler, and Andreas V. M Herz. Neural representation of space: Relation between one- and two-dimensional environments. *Front. Comput. Neurosci. Conference Abstract*, (19), 2012.

[52] Nathaniel J Killian, Michael J Jutras, and Elizabeth A Buffalo. A map of visual space in the primate entorhinal cortex. *Nature*, 491(7426):761–764, 2012.

[53] J. Krupic, N. Burgess, and J. O'Keefe. Neural representations of location composed of spatially periodic bands. *Science*, 337(6096):853–857, 2012.

[54] Julija Krupic, Marius Bauza, Stephen Burton, Caswell Barry, and John O'Keefe. Grid cell symmetry is shaped by environmental geometry. *Nature*, 518(7538):232–235, 2015.

[55] Rosamund F Langston, James A Ainge, Jonathan J Couey, Cathrin B Canto, Tale L Bjerknes, Menno P Witter, Edvard I Moser, and May-Britt Moser. Development of the spatial representation system in the rat. *Science*, 328(5985):1576–1580, 2010.

[56] Peter E Latham, Sophie Deneve, and Alexandre Pouget. Optimal computation with attractor networks. *J. Physiol. Paris*, 97(4-6):683–694, 2003.

[57] Jill K Leutgeb, Stefan Leutgeb, May-Britt Moser, and Edvard I Moser. Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science*, 315(5814):961–966, 2007.

[58] Stefan Leutgeb and Jill K Leutgeb. Pattern separation, pattern completion, and new neuronal codes within a continuous ca3 map. *Learning*

& Memory, 14(11):745–757, 2007.

[59] Stefan Leutgeb, Jill K Leutgeb, Alessandro Treves, May-Britt Moser, and Edvard I Moser. Distinct ensemble codes in hippocampal areas ca3 and ca1. *Science*, 305(5688):1295–1298, 2004.

[60] Paul A Lipton, John A White, and Howard Eichenbaum. Disambiguation of overlapping experiences by neurons in the medial entorhinal cortex. *The Journal of neuroscience*, 27(21):5787–5795, 2007.

[61] E J Markus, Y L Qin, B Leonard, W E Skaggs, B L McNaughton, and C A Barnes. Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *J. Neurosci.*, 15(11):7079–7094, 1995.

[62] Alexander Mathis, Martin B Stemmler, and Andreas VM Herz. Probable nature of higher-dimensional symmetries underlying mammalian grid-cell activity patterns. *eLife*, page e05979, 2015.

[63] B L McNaughton, C A Barnes, and J O'Keefe. The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Exp Brain Res*, 52(1):41–49, 1983.

[64] Bruce L McNaughton, Francesco P Battaglia, Ole Jensen, Edvard I Moser, and May-Britt Moser. Path integration and the neural basis of the 'cognitive map'. *Nat. Rev. Neurosci.*, 7(8):663–678, 2006.

[65] Paul Miller, Carlos Brody, Ranulfo Romo, and Xiao-Jing Wang. A recurrent network model of somatosensory parametric working memory in the prefrontal cortex. *Cerebral Cortex*, 13(11):1208–1218, 2003.

[66] R.G.M. Morris. Spatial localization does not require the presence of local cues. *Learning and motivation*, 12(2):239–260, 1981.

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

[68] Joshua P Neunuebel, D Yoganarasimha, Geeta Rao, and James J Knierim. Conflicts between local and global spatial frameworks dissociate neural representations of the lateral and medial entorhinal cortex. *The Journal of Neuroscience*, 33(22):9246–9258, 2013.

[69] Ehren L Newman, Jason R Climer, and Michael E Hasselmo. Grid cell spatial tuning reduced following systemic muscarinic receptor blockade. *Hippocampus*, 24(6):643–655, 2014.

[70] J. O'Keefe and L. Nadel. *The hippocampus as a cognitive map*. Oxford University Press, Oxford, 1978.

[71] John O'Keefe. Place units in the hippocampus of the freely moving rat. *Experimental neurology*, 51(1):78–109, 1976.

[72] John O'Keefe and Neil Burgess. Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. *Hippocampus*, 15(7):853–866, 2005.

[73] John O'Keefe and Jonathan Dostrovsky. The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 34(1):171–175, 1971.

[74] David S Olton and Robert J Samuelson. Remembrance of places passed: spatial memory in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, 2(2):97, 1976.

[75] Hugh Pastoll, Lukas Solanka, Mark CW van Rossum, and Matthew F Nolan. Feedback inhibition enables theta-nested gamma oscillations and grid firing fields. *Neuron*, 77(1):141–154, 2013.

[76] Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, EJ Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

[77] G J Quirk, R U Muller, and J L Kubie. The firing of hippocampal place cells in the dark depends on the rat's recent experience. *J. Neurosci.*, 10(6):2008–2017, 1990.

[78] D L Ringach, M J Hawken, and R Shapley. Dynamics of orientation tuning in macaque primary visual cortex. *Nature*, 387(6630):281–284,

May 1997.

[79] R Romo, C D Brody, A Hernández, and L Lemus. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, 399(6735):470–3, Jun 1999.

[80] A Samsonovich and B L McNaughton. Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci.*, 17(15):5900–5920, 1997.

[81] Francesca Sargolini, Marianne Fyhn, Torkel Hafting, Bruce L McNaughton, Menno P Witter, May-Britt Moser, and Edvard I Moser. Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science*, 312(5774):758–762, 2006.

[82] Christoph Schmidt-Hieber and Michael Häusser. Cellular mechanisms of spatial navigation in the medial entorhinal cortex. *Nature Neuroscience*, 16(3):325–331, 2013.

[83] H S Seung. How the brain keeps the eyes still. *Proceedings of the National Academy of Sciences*, 93(23):13339–13344, 1996.

[84] H S Seung and D.D. Lee. The manifold ways of perception. *Science*, 290(5500):2268–2269, 2000.

[85] Sameet Sreenivasan and Ila R Fiete. Grid cells generate an analog error-correcting code for singularly precise neural computation. *Nature Neuroscience*, 14:1330–1337, 2011.

[86] RJ Steele and RGM Morris. Delay-dependent impairment of a matching-to-place task with chronic and intrahippocampal infusion of the NMDA-antagonist D-AP5. *Hippocampus*, 9(2):118–136, 1999.

[87] H. Stensola, T. Stensola, T. Solstad, K. Froland, M.-B. Moser, and E.I. Moser. The entorhinal grid map is discretized. *Nature*, 492:72–78, 2012.

[88] Tor Stensola, Hanne Stensola, May-Britt Moser, and Edvard I Moser. Shearing-induced asymmetry in entorhinal grid cells. *Nature*, 518(7538):207–212, 2015.

[89] C F Stevens and Y Wang. Changes in reliability of synaptic function as a mechanism for plasticity. *Nature*, 371(6499):704–7, Oct 1994.

[90] Jens-Boie Suck, Michael Schreiber, and Peter Häussler. *Quasicrystals: An introduction to structure, physical properties and applications*, volume 55. Springer Science & Business Media, 2013.

[91] Jeffrey S Taube, Robert U Muller, and James B Ranck. Head-direction cells recorded from the postsubiculum in freely moving rats. ii. effects of environmental manipulations. *J. Neurosci.*, 10(2):436–447, 1990.

[92] Edward C Tolman. Cognitive maps in rats and men. *Psychological review*, 55(4):189, 1948.

[93] Endel Tulving. Precis of elements of episodic memory. *Behavioral and Brain Sciences*, 7(02):223–238, 1984.

[94] Xue-xin Wei, Jason Prentice, and Vijay Balasubramanian. A principle of economy predicts the functional architecture of grid cells. *eLife*, page e08362, 2015.

[95] A.C. Welday, I.G. Shlifer, M.L. Bloom, K. Zhang, and H.T. Blair. Cosine directional tuning of theta cell burst frequencies: evidence for spatial coding by oscillatory interference. *J. Neurosci.*, 31(45):16157–16176, 2011.

[96] Peter E Welinder, Yoram Burak, and Ila R Fiete. Grid cells: the position code, neural network models of activity, and the problem of learning. *Hippocampus*, 18(12):1283–1300, 2008.

[97] T.J. Wills, C. Lever, F. Cacucci, N. Burgess, and J. O'Keefe. Attractor dynamics in the hippocampal representation of the local environment. *Science*, 308(5723):873–876, 2005.

[98] M A Wilson and B L McNaughton. Dynamics of the hippocampal ensemble code for space. *Science*, 261(5124):1055–1058, 1993.

[99] E.R. Wood, P.A. Dudchenko, R.J. Robitsek, and H. Eichenbaum. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron*, 27(3):623–633, 2000.

[100] Michael M Yartsev, Menno P Witter, and Nachum Ulanovsky. Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature*, 479(7371):103–7, Nov 2011.

[101] D Yoganarasimha, Geeta Rao, and James J Knierim. Lateral entorhinal neurons are not spatially selective in cue-rich environments. *Hippocampus*, 21(12):1363–1374, 2011.

[102] KiJung Yoon, Michael A Buice, Caswell Barry, Robin Hayman, Neil Burgess, and Ila R Fiete. Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature Neuroscience*, 16(8):1077–1084, 2013.

[103] K Zhang. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.*, 16(6):2112–2126, 1996.