The Dissertation Committee for John Eric Widloski
certifies that this is the approved version of the following dissertation:

Grid cell attractor networks: development and implications.

Committee:

Ila Fiete, Co-Supervisor

Michael Marder, Co-Supervisor

Vernita Gordon

Jonathan Pillow

Harry Swinney

Grid cell attractor networks: development and implications.

by

John Eric Widloski, B.S. PHY.

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 Denise Prince.

Acknowledgments

I wish to thank the multitudes of people who helped me over my years at the University of Texas at Austin. First and foremost, to my PI, Ila Fiete, who turned me on to grid cells and the hippocampus in general, and who encouraged a level of rigor and aptitude that has given me confidence to pursue further studies in academia. To Michael Marder, who heroically stepped in to be my physics mentor when no one else was willing, and who had the patience to understand the work I was doing, and offering critical insights in the process. To the Fiete lab, past and present, for always being open to sharing their knowledge and expertise, and for their helpful criticisms and suggestions regarding my work. To colleagues in the physics graduate program, for their intellectual support and friendship. To my family, especially my parents, to whom Im heavily indebted, for their encouragement and support, and for financing my adolescent extracurricular interests. And finally, to my partner during all these years in Austin, Denise Prince, for supporting me across all fronts, for letting me indulge my scientific curiosities at the expense of the relationsihp, and being a stabilizing force throughout. To everyone, I'm greatly indebted.

Grid cell attractor networks: development and implications.

Publication No. _____

John Eric Widloski, Ph.D.
The University of Texas at Austin, 2015

Co-Supervisors: Ila Fiete
Michael Marder

At the foundation of our ability to plan trajectories in complex terrain is a basic need to establish one's positional bearings in the environment, i.e., to self-localize. How does the brain perform self-localization? How does a network of neurons conspire to solve this task? How does it self-organize? Given that there might be multiple solutions to this problem, with what certainty can we say that any such model faithfully captures the neural structure and dynamics as it exists in the brain? This thesis presents a collection of three theoretical works aimed at addressing these problems, with a particular focus on biological plausibility and amenability to testing experimentally.

I first introduce the context within which the work in the thesis is situated. Chapter 1 provides a framework for understanding algorithmically how

the brain might solve the problem of self-localization and how a neural circuit could be organized to perform self-localization based on the integration of self-motion cues, an operation known as path integration. We also introduce the neurobiology that underlies self-localization, with special emphasis on the cell types found in and around the hippocampus. We discuss the case that a particular class of cells – grid cells – subserve path integration, because of their peculiar spatial response properties and their anatomical positioning as the recipients of self-motion information. Continuous attractor models are introduced as the favored description of the grid cell circuit. Key open questions are introduced as motivation for the subsequently described work.

I next focus on the question of how the grid cell circuit may have organized. In Chapter 2, it is demonstrated that an unstructured immature neural network, when subjected to biologically plausible inputs and learning rules, can learn to produce grid-like spatial responses and perform path integration. This model makes a number of predictions for experiment which are described at length.

In Chapter 3, I describe a theoretically motivated experimental probe of the organization and dynamics of the grid cell circuit. The proposed experiment relies on sparse neural recordings of grid cells together with global perturbations of the circuit (and is thus experimentally feasible). It promises to yield special insight into the hidden structure of the grid cell circuit.

Finally, in Chapter 4, I provide an analytical treatment of pattern formation dynamics in the grid cell circuit. This work focuses on nonlinear effects

to show how parameters of neurons and networks relate to the formed pattern in a model grid cell network beyond the results predicted by linear stability analysis.

Table of Contents

List of Figures

Chapter 1

Introduction

We take for granted that our brains are responsible for the complex workings of our inner lives. And yet the brain is really a biological entity, composed of neurons that are, arguably, individually unintelligent, yet synergistically capable of performing immensely complicated tasks like visual recognition, motor control, short and long-term memory, etc. How does the brain accomplish these processes? Given that the brain of an infant, while wondrously complex and worthy in its own right, is not like the fine-tuned apparatus of the adult, how does the brain self-organize? As computational neuroscientists, we address these questions by building models of the brain. However, this enterprise is generally fraught with difficulty, given the brain's dizzying multi-scale complexity, structurally and dynamically – it is often cited as the most complex object in the known universe. Even so, the field has made rapid advances in recent years, so that some brain processes have been successfully modeled down to the circuit-level description, e.g. [1, 19, 22, 30, 62, 63, 98].

This thesis focuses on a process that is often taken for granted: how the brain keeps track of location in an environment. We will consider one particular component of self-localization, which involves using self-motion cues

to update the instantaneous position estimate. This operation is called "path integration". Importantly for computational neuroscientists, there exists for this topic a relatively rich playground of experimental results [4, 5, 9, 10, 13, 14, 21, 34, 39, 43, 45, 48, 54, 71, 95, 97, 102, 111, 119], and much progress has been made toward understanding the putative circuit behind this process. However, key questions remain unanswered. This thesis presents work that addresses several of this key issues.

1.1 The brain as a computer: computational demands and neural hardware

From a computational perspective, the demands placed on the brain are three-fold: to *represent* (or *encode*) information to *compute* with/on that information, and to guide the agent to *act* on that information. Compared with the hardware of man-made computer systems serving the same computations for the purposes of, say, guiding an autonomous robot, neural hardware, while nonlinear, is particularly noisy and slow. However, it is massively parallel, with dense feedforward and feedback loops within and across its many computational modules (we will define a *module* to be a collection of neurons cooperating to perform some function, or subroutine). This organizational property lends the brain its capacity to rival, or even dwarf, the fastest man-made machines in computational power, as in the case of inference. Also exceptional is its capacity to learn, and to self-organize. In fact, this is a necessity: the 20,000 genes in the human genome seem to few to specify the 100

trillion synapses that connect the 100 billion neurons in the human brain.

This thesis is directed toward understanding the computations behind the brain's ability to self-localize based on internal cues related to the animal's motion (i.e., *path integrate*). In particular, how does the brain *represent* spatial location, and how does it *compute* it from its inputs? (The question of how this representation is used to guide the animal's behavior, though intriguing, will not be discussed further.) How are neurons connected to perform this function, and what guides the system's self-organization?

In the existing neuroscience literature, there is an important class of neural models that suppose that within the module subserving path integration, the underlying connectivity of the neurons is highly organized and feeds-back on itself: imagining that the neurons in the module are spread out on a sheet, the connectivity is such that neurons located close to one another excite (or disinhibit) one another, and neurons further apart inhibit one another (Mexican-hat connectivity), Figure 1.1. Consider providing the neurons with a constant external flux of energy, that activates all neurons uniformly across the network. Now consider adding a small, transient pulse of input to one of the neurons. If the strength of the feedback is sufficiently small, the pulse dissipates and the neural profile goes back to reflecting the uniform input. If, however, we increase the strength of the feedback, and again provide a small, transient pulse to one neuron, the pulse widens into a bump, as before (via the local excitatory coupling), but this time, because the feedback is larger, the long-range inhibition kicks in and shores up the bump, preventing

it from spreading. This region of silence then allows for neurons just beyond to spontaneously form, uninhibited, into their own activity clusters, and the process propagates until the hotspots of activity have organized themselves into a stable hexagonal array with a well-defined scale (period), orientation, and phase. The process is reminiscent of the Turing patterns that develop in activator-inhibitor reactions and is representative of pattern formation in biological systems in general, which occur when the external flux of energy coming into the system are balanced by the local fluxes of energy (governed by local self-enhancement and long-range inhibition between constituent elements of the system).

Because the coupling between neurons is translation invariant, all phases of this pattern are stable. This stability in phase endows the network naturally with a certain useful representational capacity. For the purposes of neural models responsible for keeping track of animal location, it is precisely this phase that is used to encode the animal's spatial displacement. However, the particular complication for such models is figuring out how to couple the pattern formation process to the motion of the animal, such that when the animal moves, the pattern also moves proportionally (path integration). Thus, pattern formation in this implementation is not static, but necessarily time-dependent. The coupling scheme that performs this function is a complex variation of the Mexican-hat connectivity described above (and is described in Chapter 2). The question of the self-organization of such a structure is not at all understood. One of the contributions of this thesis is to address the

question, using biologically plausible wiring rules and inputs. This biological realism allows, for the first time, detailed predictions concerning the development of the network connectivity and dynamics. This is the first part of the thesis. The second part of the thesis involves the development of a novel experimental probe for inferring certain structural and dynamical details normally hidden to experimentalists using conventional techniques. Specifically, the probe is able to infer the existence of Turing-like patterning in the putative circuit believed responsible for self-localization via path integration, and also the topology of the underlying network architecture. The probe relies on systematically perturbing the dynamics of the network in a controlled way and sparsely sampling the dynamics of only a handful of neurons, all well within the realm of experimental feasibility. The last part of the thesis involves developing, for the first time, an analytical understanding of the constraints that govern the patterning of the network dynamics.

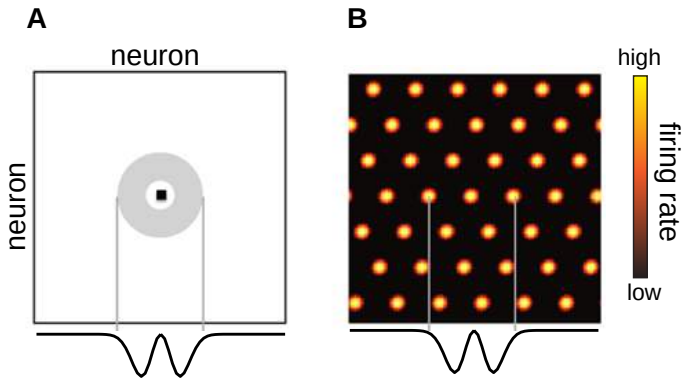

Figure 1.1: **Pattern formation in a neural network.** (A) 2D sheet of neurons in brain space, one of which is highlighted as the black square at center. The connectivity profile between this neuron and its neighbors is shown below as the black contour, and has the form of a Mexican-hat (or, difference-of-Gaussians), with local excitation (or, rather, in this specific case, disinhibition) and long-range inhibition. This connectivity profile is shared by all neurons in the sheet. (B) Steady-state pattern formation in the neural activity, given the connectivity in (A). The yellow blobs represent clusters of neurons with high firing rates, while the black surrounding regions indicate that the neurons are silent. Figure modified from [16]

1.2 What is self-localization and why we do it: A computational view based on Marr's levels of description

As an organizing principle for introducing the topics in this thesis, we borrow a descriptive framework from David Marr, the late, great computational neuroscientist from the middle of the 20th century. Marr was responsible for some of the first models linking high-level cognitive processes (e.g., memory recall) to plausible implementations at the neural circuit level. Marr's approach to understanding such processes was a description at three hierarchical levels: at a functional level, algorithmical level, and the level of a circuit implementation. In what follows, we will consider how the brain computes the animal's location, i.e., the process of self-localization. As we shall see, there are two algorithms for self-localization: a strategy akin to triangulation, and path integration. The focus of much of the work of this thesis is on how the brain learns and implements the latter algorithm.

Many animals, from insects to mammals, exhibit complex collections of spatial behaviors for survival, including foraging for food, remembering where home is, remembering safe routes between home and various known food sources, improvising new routes back home after an exploratory outbound path to a previously unvisited location, learning maps of new environments, and setting goal locations. At the core of these behaviors is a basic need to know where the animal is at any given time, i.e., to self-localize, which enables the animal to plan routes towards desirable places, and avoid undesirable places.

1.3 Algorithmical descriptions of self-localization: external vs. internal cues

Algorithmically, there are at least two well-known approaches to self-localization, based on two entirely different sources of information. These two strategies can be intuited in the context of navigation at sea where they are known as celestial navigation and dead-reckoning. Assuming that the vessel is well away from shore, the only available cues are the stars (and sun). We will define these as *external* cues. Celestial navigation involves the mapping of viewable constellations to coordinates on the globe. Importantly, the mapping between constellations and position is the culmination of many years of observation, and it is only through this lengthy process of familiarization with the night sky that a map is available to the sailor at all as a means to self-localize.

Distinguish this technique from dead-reckoning, or *path integration*, in which the local motion of the ship, i.e., an *internal cue*, is used to compute the ship's displacement with respect to some origin. Here, there is no need to reference landmarks. The strategy involves simple vector calculus: by holding course on a particular heading direction, over some known time period at fixed speed (the speed is measured, e.g., by throwing an object overboard and measuring how long it takes the object to move from bow to stern), the displacement vector can be computed and added to the previous estimate (computed from other legs of the journey) to keep track of the ship's relative position with respect to its starting point. In conditions that obscure the sky, dead-reckoning is all that is available to the sailor as a means of self-

localization.

Such strategies, while used overtly by sailors, are not unique to humans. Many species, from insects to mammals, are skilled at self-localization. For our purposes, we will restrict ourselves to mammals, and in particular, because of the wealth of behavioral experiments, to rodents. In familiar environments, in the presence of stable cues, rodents tend to use available landmarks as beacons to guide their immediate behavior. An important experiment, conducted by Edward Tolman in 1946, showed that they can do so by accessing an internal (i.e., in their head), "cognitive" map of the external environment [109]. After preliminary training rats in a spatially restricted, L-shaped enclosure, where they are rewarded to move from one end of the tract to the other, the rats were then given the choice of many novel radial arms (having blocked the original path) to move directly to the endpoint, i.e., along the hypotenuse of the original L-shape. Interestingly, many rats, above chance, were observed to take the shortest radial route corresponding to the beeline path between the start and end points. This suggested that rats could improvise paths and shortcuts through regions they had not previously traversed. Tolman reasoned that animals were capable of constructing mental representations of spatial locations in the environment and relations between them, independent of reward (because movement along the radial arms had never been experienced before, let alone reinforced). He called these representations "cognitive maps". By construction, cognitive maps could hypothetically provide route information between any two points in the environment and thus be used to navigate Tol-

man's sun-burst maze. Thus, like the celestial navigator, who relates perceived landmarks to particular coordinates on his map, the brain of the mammal relates proximal and distal landmarks to particular coordinates in its cognitive map.

In the absence of known cues, as in novel or light-deprived environments, animals, just like the sailors, must rely on their sense of self-motion to guide behavior. That animals are capable of navigation under such deprived circumstances was shown in a classic paper by Mittelstaedt and Mittelstaedt in 1980 [79]. Mother gerbils, in laboratory conditions in which their nest of pups was placed on a large platform, were observed, in complete darkness, to fetch a displaced pup and bring it back on a beeline path to the nest. Was she simply beaconing on the odor/noises of the nest, or was she computing a homing vector based on having maintained a representation of her position via path integration? To answer this, the authors carefully moved the nest after the mother had left to retrieve the pup. It was found that her return path was usually directed towards the old location rather than the current nest position, even if the nest location was relatively close by, showing that she was navigating using path integration.

As a side note, it is interesting to consider how maps are constructed in the first place. In the robotics literature, this is known as the SLAM problem (Simultaneous Localization and Mapping). If self-motion estimates are precisely integrated to determine location, then building a map of the environment involves simply visiting and attaching a coordinate to each landmark. Once

constructed, the map can supplant the path integrator as the main means by which the robot self-localizes. Typically, however, the sensory inputs (both internal and external) are noisy and unreliable and can lead to large inaccuracies and inconsistencies in the map, e.g., multiple coordinates mapping to the same landmark, or vice versa. Building a map that reflects the true arrangement of objects in the environment requires probabilistic sequential Bayesian methods (or approximations like Kalman filters) to optimally combine the expected sensory inputs with the current sensory inputs in order to correct both the self-motion estimate and to update the map.

These processes described above, i.e., map-based and self-motion-based self-localization, and the necessity of the latter for the development of the former, have striking neural correlates, to which we now turn.

1.4 Neurobiology of self-localization

1.4.1 Hippocampus as the neural substrate of the cognitive map

There is particular region of the brain that has been implicated in playing a crucial role in spatial navigation: the hippocampus. The first significant interest in the hippocampus came in the 1950's, when it was shown to be an area related to memory. Patient HM, whose hippocampus was bilaterally lesioned in order to quell the epileptic seizures that emanated from there, was left with the complete inability to form new memories. Importantly, his intellectual and perceptual abilities remained intact. Later, and seemingly at odds with this interpretation, was the discovery, by John O'Keefe in the early

1970's, of a special type of neuron recorded electrophysiologically within the hippocampus of a behaving rodent. This cell fired if and only if the animal was in the immediate neighborhood of a particular location (its place field) in a particular environment [85], Figure 1.2B. Many place cells possess place fields within any given environment. Because the recording environments (typically 0.5-1 m per dimension) tended to be covered by different place fields, place cells were hypothesized to form the basis for spatial mapping. This pointed to the hippocampus as the locus of a cognitive map for space, and, relating back to its role in memory, to the possibility that the encoding and retrieval of memories occurs within a spatial substrate.

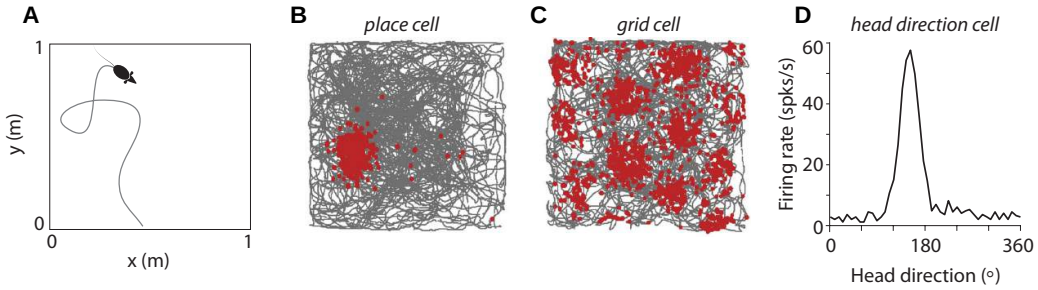

Figure 1.2: **Spatial-coding cell types found in hippocampal formation.** (A) Schematic of square environment, i.e., open field, in which rat is placed to measure responses of cells. (B) Response of a place cell in the hippocampus, from experiment [26]. The trajectory of the rat is marked by the gray line. Red dots show the locations of the animal when the cell fired a spike. (C) Response of a grid cell in the MEC, from experiment [26]. Gray line and red dots, same as in (B). (D) Response of a head-direction cell the in postsubiculum, from experiment [110].

The groundbreaking discovery of place cells prompted the renewal and further development of Tolman's cognitive map hypothesis from 20 years earlier. O'Keefe and his colleague Lynn Nadel extensively reviewed the psychological, behavioral, anatomical, and physiological evidence for the existence of abstract spatial maps in the brain, in their comprehensive and prescient book on the topic [86]. O'Keefe and Nadel provided a clear definition of a spatial map as an abstract representation of locations in an environment, the relationships between them, and the sensory inputs related to the locations. An important contribution to the cognitive map theory by O'Keefe and Nadel [86] was to elaborate on the problem of ongoing location identification. Reasoning that it was sufficiently difficult to estimate location purely from the observation of shifting angles of visible landmarks relative to the animal, they hypothesized that another system, sensitive to the movements of the animal, would be required. This second system was hypothesized to follow the self-motion of the animal through space, shifting the hippocampal place representation accordingly. The self-motion drive was suggested to supplement purely external sensory inputs, which provided cues originating from viewing the world from different locations and angles. Functionally, the "internal" system was seen as providing predictions about what to expect at a particular place, that were compared with the actual sensory input provided by the "external" system. Discrepancies between expectation and actual input might be conveyed via *misplace units* [84], whose hypothesized role was to signal mismatches between the two systems. In the theory, active misplace units would trigger

further exploration of the environment until enough information was acquired to fix the incongruities between the two inputs and silence the misplace units. Thus, in a way, each system was seen as providing partially accurate representations of the animal's location within the environment, with the interplay between the two suggested as leading to the formation of a consistent map.

At the time, there was no evidence for the "internal", path-integration-based module in the brain; the authors argued for its inclusion on theoretical grounds. However, there is now some reason to believe that the"internal" system is in the medial entorhinal cortex (MEC), which directly projects to the hippocampus.

1.4.2 The entorhinal cortex as gateway to the hippocampus

In addition to the hippocampus, the entorhinal cortex (EC) is a key brain area involved in spatial navigation. Lesion studies have implicated the EC in spatial computation [34, 87, 88, 101, 111]. The EC is the cortical gateway of inputs to the hippocampus, and divided into medial and lateral portions (MEC and LEC, respectively) [76, 121, 122]. Electrophysiological studies in the EC of the freely moving rat reveal a dissociation in the nature of the LEC and MEC representations [28]: LEC cells tend to respond to objects in the animal's immediate environment [28, 113, 126, 128], while cells in MEC ignore object locations and instead fire at multiple locations in the open field [2, 37, 41, 91, 117]. Thus, the LEC and MEC might form the two parallel streams postulated in the cognitive map hypothesis, carrying external sensory and

internal motion-based cues, respectively, to be synthesized in the hippocampus.

1.4.3 Grid cells as the hypothesized neural substrate of path integration

Less than a decade ago, the MEC of rats was found to contain a class of cells – grid cells – with astonishing spatial firing characteristics [48]: each cell fires at multiple locations in an environment, and the locations are arranged on the vertices of an essentially equilateral triangular grid, Figure 1.2C. This grid that tiles the space in which the animal moves (defined as the *movement space*) is parameterized by period, orientation, and phase. Grid cells have since been found in other mammals as well, including mice, bats, monkeys, and humans [31, 40, 48, 61, 66, 124].

A couple of definitions are in order. In what follows, we will make reference to three spaces: *movement space*, defined above; *brain space*, which refers to the physical arrangement of cells in the brain; and *topological space*, whose significance we will see later, defined as a special re-arrangement of the cells in brain space based on their connectedness (note that this latter space is only used as an aid for understanding the network structure). We define the cell's *spatial tuning curve* as follows: first, subdivide the environment into spatial bins of width Δx (which is usually on the order of 1 cm), with the ith bin centered at $\vec{x}_i$; then, count the number of spikes emitted by the cell in each bin and divide by how often the animal visits that spatial bin:

$$r(\vec{x}_i) = \frac{\#\text{of spikes fired in spatial bin centered at } \vec{x}_i}{\text{time spent by animal at spatial bin } \vec{x}_i}. \tag{1.1}$$

For grid cells, the spatial tuning curve has the shape of a grid. In contrast, we define the entire network's activity profile in brain/topological space as the *population activity*. Population activity is a snapshot of the instantaneous (on the order of a time bin $\approx$ 1ms in duration) dynamics of the network.

Below, we highlight six key properties of grid cells.

- *Atopography of cell arrangement in brain space.* Nearby cells in brain space share common grid periods and orientation, but do not change continuously in phase. This discontinuity in spatial tuning curves of nearby cells means that the arrangement of cell is atopographic. In other words, their arrangement in brain space has no bearing on the similarity of their spatial tuning curve phases (i.e., their arrangement in movement space). This is distinct from the sensory systems in the brain, most notably the visual system, in which nearby cells in brain space are tuned to nearby regions of visual space.

- *Modularity.* There is one other key organization principle, that applies along the longitudinal axis of the MEC, Figure 1.3. Along this axis, grid periods increase in size; however, the distribution of periods (and orientations) is discretized, or chunked [102]. Thus we define a grid cell network (GCN) as a collection of grid cells with a shared period and orientation, but with phases distributed uniformly, and randomly (i.e., atopographic organization of phases in brain space within each GCN). It is believed that there are approximately 4000-40000 neurons

in a GCN [16, 38, 77]. Moreover, evidence suggests that the different GCNs that span the longitudinal axis of the MEC function more or less independently [102]. A quick note on the apparent redundancy of this code: From the point of view of a downstream read-out, each GCN's encoding of position is only unique modulo the length of the GCN's characteristic spatial period; a unique encoding of position only comes from reading out from all GCN's simultaneously. The "why" question of this particular encoding of position (called a modulo code) is interesting in its own right but will not be discussed further in this thesis. Suffice it to say that this particular code is especially sensitive to noise, and thus can be used to correct error in the path-integrated estimate of location [100, 115].

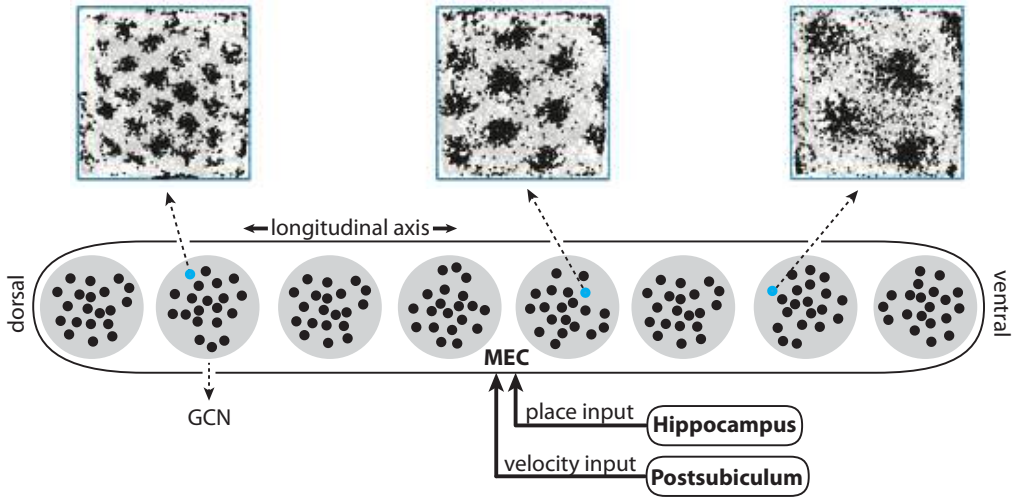

Figure 1.3: **Modular organization of grid cells along the longitudinal axis of the MEC.** Schematic organization of grid cells in MEC, showing the span of modules (solid gray circles) along the MEC's longitudinal axis, from the dorsal (left) to the ventral (right) end. A GCN module is defined as a collection of grid cells with common grid scale and orientation. GCN's are also reciprocally connected with the hippocampus and postsubiculum, which provide place and velocity information, respectively. Grid cell recordings shown above are from experiment [102].

- *Insensitivity to external cues: self-motion cues as primary drive of grid response.* The spatial tuning curves of grid cells (in movement space) can rotate when salient external cues are rotated [48], and periods of their spatial tuning can resize in response to a rescaling of a familiar environment [4], but other than simple modulations of the heights of the cell's firing fields (while preserving grid orientation, scale, and phase) [96], the spatial tuning curves of grid cells are relatively insensitive to the particulars of the environment. This is in contrast to the spatial responses of cells in the LEC and the hippocampus, which exhibit more detailed and complex changes to environmental manipulation [23, 28, 72, 73, 82, 113, 117, 126, 128]. The relative insensitivity of grid cells to external cues and the stability of their fields in cue-poor environments and darkness [48] suggests that self-motion is the primary determinant of grid cell firing. For these reasons, it is widely hypothesized that the grid cell system computes, or at least is responsive to, a path integrated estimate of the animal's position. However, direct evidence of the role of grid cells in path integration is lacking.

- *Conjunctivity: grid + head direction tuning.* Grid cells are most commonly found in the superficial layer (layer II) of MEC. The postsubiculum, a major source of input to the MEC, terminates in the deep layers [112]. The postsubiculum contains head direction cells, which fire when the animal's head points in a particular direction, usually with respect to some salient, stable cue in the environment, independent of the actual

location of the animal within the environment, Figure 1.2D. Because of these inputs, the deep layers (layers III-V) of the MEC contain cells responsive to the animal's head direction, either in the form of pure head direction tuning or combined head direction and grid-like tuning. The latter are known as "conjunctive" grid cells [95]. A fraction of these are modulated by the animal's speed as well [95], such that their firing rate increases linearly with speed. In contrast, grid cells in the superficial layer tend to be insensitive, or at least weakly tuned, to head direction ("pure" grid cells) [95].

- *Wiring constraints.* Grid cells are most probably excitatory cells, meaning that they excite the cells to which they project [32, 97]. In the deep layers of the MEC, grid cells can directly excite one another. In layer II, however, this is not the case; instead, communication between grid cells is likely mediated by inhibitory cells [25, 29, 89].

- *Grid cell inputs.* As mentioned earlier, the MEC is poised to receive information about the animal's location via the place cells of the hippocampus as well as information about the animal's velocity, via the speed-sensitive head direction cells of the postsubiculum [122], Figure 1.3.

1.5 Circuit implementation of path integration

1.5.1 Canonical model of neural activity

Given the evidence that suggests grid cells are the substrate on which a path-integrated estimate of the animal's location is computed and maintained, and given the fact that the different subnetworks of the MEC, i.e., GCNs, function more or less independently [102], we are faced with the task of figuring out how a collection of cells within a single GCN can both solve this task and elicit grid-like spatial responses (in movement space) in the process.

There are various levels at which to describe the dynamics of single neurons. Hodgkin and Huxley introduced the first model of neural spiking, which involved four coupled ode's, one of which governed the membrane potential of the cell, similar in form to the dynamics of the voltage in an RC circuit, the other three governing the dynamics of ionic currents responsible for the cell's action potential [57], or "spike". For sufficiently large firing rates (defined as the number of spikes per second), and for a sufficiently long membrane time constant, these four ode's can be reduced to a single ode governing the firing rate of the cell, embedded in a network of other neurons with identical dynamics [92]:

$$\frac{ds_i}{dt} + \frac{s_i}{\tau_s} = f\left(I_i\right), \tag{1.2}$$

where s_i is analogous to the firing rate of the ith cell, τ_s is the characteristic time scale of the cell's response (assumed to be on the order of tens of milleseconds), and I_i is the total input into the cell. The total input is the sum of two terms, derived from internal (within the GCN) and external (outside the

GCN) sources:

$$I_i = g_i^{rec} + g_i^{ext}, \tag{1.3}$$

where $g_i^{rec} = \sum_j W_{ij} s_i$ is the total recurrent input from all other cells in the GCN, W_{ij} is the (i,j)th element of the coupling matrix, describing the strength of the connection (or synapse) from cell j to cell i, and g_i^{ext} is the external input. We will assume that f is a simple rectification of the neuron's inputs, namely $f(I) = I$ if $I > 0$, otherwise, $f(I) = 0$. Note that the dynamics specified above is entirely deterministic, and that there is no explicit spiking (i.e., no action potentials). This model is known as the Linear-Nonlinear (LN) model of neural dynamics. There is evidence to believe that the cell's transformation of its input into output spiking events is a stochastic process with Poisson statistics. The LN model can be extended by sampling, in each time bin Δt, spiking events from the distribution $Poiss(n; f(I)\Delta t)$. This builds a spiking vector of 0's, 1's, 2's, etc., formalized in continuous time as the spike train variable:

$$\sigma(t) = \sum_k n_k \delta(t - t_k^{spk}), \tag{1.4}$$

where t_k^{spk} is the kth time bin in which n_k spiking events occur. This term then replaces the right hand side of Equation 1.2:

$$\frac{ds_i}{dt} + \frac{s_i}{\tau_s} = \sigma_i, \tag{1.5}$$

This extension is known as the Linear-Nonlinear-Poisson (LNP) model of neural spiking.

Given the canonical description of the dynamics of single neurons (either Equation 1.2 or Equation 1.5), the problem of circuit implementation is reduced to finding the appropriate coupling matrix, W, as well as the appropriate inputs, g_i^{ext}. For now, we will assume that g_i^{ext} is constant and uniform across the GCN. How, then, to specify W?

1.5.2 The challenge of modeling path integration with neurons: difference of time-scales

At the most basic level, path integration is spatial calculus. For any given time step, while the new displacement is computed, the previous estimate, or memory, is stored, to which the update is then added. Therefore, there are two components: a mechanism that stores memories, and a mechanism to update those memories by the appropriate amount related to animal motion.

At the level of the neural circuit, keeping track of slowly changing variables like animal position presents a problem: individual neurons have very fast time constants (see τ_s above, which is on the order of tens of milleseconds), orders-of-magnitude below what is needed. However, collectively, neurons can, when coupled appropriately, exhibit dynamics with very slow time constants, on the order of seconds. That is, a carefully constructed coupling matrix, W, endows the state space of the system (where each dimension represents the firing rate of a single neuron, and the total dimensionality is equal to the total number of neurons) with fixed points, or attractors. These attractors

are stable such that perturbations away from the attractor fall back in. These attractor states can have any arrangement; an arrangement we are particularly interested in is a 2D manifold of such attractor states, or a 2D continuous attractor. Thus, animal position, which is a 2D continuous variable, can be mapped onto the continuous manifold of stable network attractor states. As far as path integration is concerned, this 2D continuous attractor manifold will serve to store the current estimate of animal location. Secondly, we need a way to couple the network state, which rests somewhere on the manifold, to animal motion, so that when the animal moves the state is kicked along in correspondence with the animal's velocity.

What is the coupling matrix that solves the problem of path integration as laid out above, and also gives grid-like responses from its constituent neurons as a function of animal location in movement space? That particular problem has been solved before, and we outline its construction below.

1.5.3 Continuous attractor models of grid cells

Continuous attractor models of grid cells [15, 16, 38, 46, 77] posit that neuron-to-neuron connectivity in the form of short-range excitation and long-range inhibition destabilizes the uniform activity state in the neural population activity profile (in topological space, which is like brain space, except that the neurons have been reordered based on the connectedness to their neighbors) and stabilizes a state which displays regular triangular grid patterning of activity bumps, or hotspots.

Translation invariance of such connectivity stabilizes all translations of this pattern, and an asymmetric component of the connectivity allows external inputs signaling animal velocity to drive the pattern in direct proportion to the direction and speed of the animal's movements.

All cells in the continuous attractor network model share the same grid period and orientation, because their responses are generated by translations of the same pattern, and all spatial phases are exactly uniformly distributed, consistent with the data. Disjoint network copies (modules) are required to produce different grid periods, and because each network is large, leads to the prediction of a few, discrete grid periods within each animal [16, 38, 77]. This prediction was recently experimentally verified in [102]. The fundamental prediction of continuous attractor models is that the differences in preferred spatial activation phase between pairs of grid cells will remain stable over time and regardless of environmental manipulations that induce sizeable distortions in the grid fields, if the network architecture remains unchanged. Recent analysis of simultaneously recorded grid cells with similar period and orientation across experiments involving grid cell distortion (including the environmental stretching experiments of [4] that induce likewise stretching of the grid fields) establishes the stability of these predicted relationships and shows that the grid cell population responses within a GCN are confined to a 2-d manifold within the high-dimensional state space [125].

We have shown that continuous attractor models of grid cells provide a satisfying account of self-localization via path integration through grid cells,

thus fulfilling Marr's last level of description. However, there are several key questions left unaddressed in the literature.

1.6 Development of the grid cell circuit

It is poorly understood how the system might form. In rats, grid cell responses emerge at 3-4 weeks of age, after eye opening, [71, 119], suggesting that the system is not hardwired at birth. However, there are no existing models of grid cell development consistent with the above observations. Thus questions regarding the nature of the synaptic learning rules (see below) that lead to a fully functioning grid cell network, or whether the emergence of the network requires extensive experience of the animal exploring the environment, are wholly unknown.

1.6.1 Spike time-dependent plasticity

A famous postulate by Donald Hebb, in 1948, stated without experimental support, said the following: "Whenever an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased."

The implied temporal specificity of this learning rule has since received experimental support, Figure 1.4 [6]. This has led to the following mathematical formalization of the the dynamics of the synapse connecting cell j (the presynaptic cell) to cell i (the postsynaptic cell), which depends on the cells'

relative spiking times (termed STDP, for Spike Time-Dependent Plasticity):

$$\frac{1}{\eta}\frac{dW_{ij}}{dt} = \int_0^\infty \sigma_i(t)\sigma_j(t-t')k_+(t')dt' + \int_0^\infty \sigma_j(t)\sigma_i(t-t')k_-(t')dt', \quad (1.6)$$

where $\sigma(t)$ denotes the cell's spike vector (which consists of 0's and 1's in discretized time, as described above in Equation 1.4), k_+ and k_- represent the learning kernels relating how changes in the synapse are related to the difference in times of the pre- and post-synaptic spiking events (k_+ (k_-) is used when presynaptic spike temporally precedes (follows) the postsynaptic spike), and η is the synaptic learning rate.

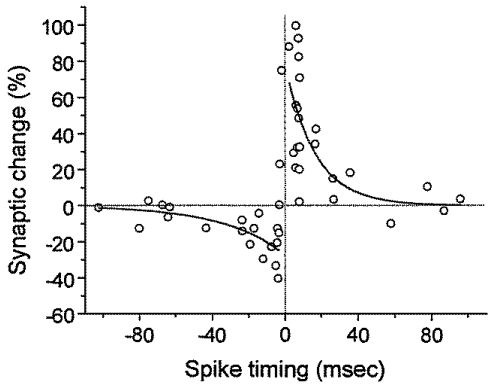

Figure 1.4: **Spike time-dependent plasticity in the hippocampus.** Modification of the synapse connecting two cells as a function of the relative time between pre- and post-synaptic spikes (i.e., "spike timing"), measured in experiment [6]. The right (left) hand side of graph shows that when the presynaptic cell fires before (after) the postsynaptic cell with a latency of less than 40 ms, the synaptic weight increases (decreases). The solid lines are exponential fits to the data, and would correspond to the functions k_+ and k_- in Equation 1.6.

1.6.2 Why the question of grid cell development is hard

The question of grid cell development poses several conceptual challenges. First, the wiring needed to perform path integration is sophisticated because of the necessary requirement that movements of the animal in movement space relate to movements of the population activity pattern in brain space. Secondly, the known atopographic arrangement (in brain space) of grid cells precludes the usage of mechanisms for building the topographic maps common to areas in the sensory cortices, like the visual cortex [120, 123]. Thirdly, it is common to model the emergence of tuning curves by assuming that a supervisory input imposes the desired tuning on the network, and that rules governing the development of neuron-to-neuron coupling (i.e., plasticity rules, as described above) consolidate the imposed patterns [49, 105]. But it is implausible that fully functional supervisory grid-patterned inputs exist before grid cell maturation.

1.6.3 What we show

In Chapter 2, we describe a neural network model for the emergence of grid cells based on spatial experience. Despite its relative simplicity, the model overcomes the conceptual hurdles described above. The inputs into the model are based on the available anatomical and electrophysiological data (outline above), which suggest that the MEC has access to both animal location information (via place cells that are themselves assumed to be driven by landmarks in an enriched environment) as well as animal velocity information (via speed-

sensitive head direction cells in the post-subiculum). The stochastic dynamics of individual neurons is modeled according to Equation 1.2. The initially random cell-to-cell coupling matrix is developed according to Equation 1.6, where the learning kernels are assumed generically to be exponentials. The network is then allowed to develop, driven by inputs derived from a quasi-random exploration of a 1D environment by a virtual rat. After development, the mature network exhibits grid cell-like activity patterns, and is capable of path integration, even in the absence of the external landmark-derived place information (i.e., darkness). It is consistent with the existing (but limited) data on the organization of grid cells in layer II of the MEC, including the atopography of the grid cell layer (see above), the use of conjunctive grid cells (see above), and the exhibition of approximate 2D continuous attractor dynamics [125].

Beyond being the first developmental model of grid cells that exhibits continuous attractor dynamics, the model is unique in that it is constructed using stochastically spiking neurons, and thus is able to implement what is widely considered to be the most biologically realistic form of synaptic learning, STDP. In doing so, the model provides a unique opportunity to make predictions regarding the developing, *immature* network as well, and thus to connect with a host of experiments regarding grid cells in development [71, 118, 119].

Lastly, there is an important takeaway that ties into the next chapter of the thesis; namely, that the topology of movement space constrains the topology of brain/topological space (recall that "topological" space refers to a specific reordering of the neurons in brain space based on connectivity,

such that strongly connected neurons are placed next to one another). Here, because movement space is necessarily planar, the topology of the network structure is also planar. This is an important prediction of the developmental model, and naturally leads to the next part of the thesis.

1.7 Inferring network topology and local connectivity structure from experiment

The fact that the population activity of grid cells is localized around a continuous low-dimensional manifold [39, 125] provides strong evidence for strongly recurrent networks at the heart of the grid cell response. However, such a result does not sufficiently discriminate between available continuous attractor models. For one, it does not specify whether the computations that give rise to the attractor dynamics are performed within [15, 16, 38, 46, 77] or upstream [18, 68, 78, 114] of the grid cell layer. Further, amongst models in the former category, there is ambiguity in the nature of the network connectivity: some models posit that the topology of the network connectivity is toroidal [16, 46], while others suggest that it is planar [16, 38, 116]. As discussed at the end of the last section, this structural difference between models has qualitative ramifications for how the circuit could have developed. Another unknown is the spatial extent of the center-surround connectivity. This is directly related to the nature of the population activity patterning in the network's dynamics, both in how large and how many activity bumps make up the population activity pattern, as well as the permissible range of population activity pattern

periods.

1.7.1 Why circuit discrimination is hard

Despite their differences, the models are difficult to distinguish on the basis of existing data, because all of them produce grid-patterned outputs and exhibit approximate 2D continuous attractor dynamics. Worse, neither complete cataloguing of the activity of every cell in the network or the strength of every synapse in the network would be sufficient to distinguish between proposed mechanisms.

1.7.2 What we show

We show how it is possible to gain surprisingly detailed information about the network structure and dynamics of the grid cell circuit from a feasible experimental strategy that depends on a circuit perturbation and sparse neural recording. The proposed strategy can allow the experimenter to discriminate between various distinct candidate mechanisms that are currently undifferentiated by experiment. We show how, in particular, both the topology of the network (planar vs. toroidal), and, under certain restrictions, how many bumps there are in the network population activity (i.e., how spatially restricted are the network connections) can be deduced through a feasible experimental manipulation of the circuit.

1.8 Analytical derivation of constraints related to pattern formation of population activity

The last portion of the thesis is concerned with beginning to understand what controls the period of the population activity pattern. In particular, we want to understand a particular result of the previous two chapters, namely that the period of the pattern changes with certain parameters of the network dynamics (e.g., changing the neural time constant, τ_s). In this chapter, we seek understand this dependency.

1.8.1 Why this is challenging

According to linear stability analysis, the period of the pattern formation in the population activity should depend on the characteristic spatial scale of the neural coupling. Changing dynamical parameters (like τ_s) is only expected to modulate the threshold of the instability that leads to pattern formation in the dynamics, but have no effect on the critical wavelength. Thus, understanding why this is apparently not the case requires going beyond simple linear stability analysis to consider the specific nonlinearities of the system.

1.8.2 What we show

We derive constraints between the dynamical/structural parameters of the network, and relate them to the expected pattern parameters, including the period of the pattern and the size of the individual hotspots that comprise the pattern. These equations constrain the solution phenotype to a continu-

ous, low-dimensional manifold. We confirm these predictions with numerical simulations, and investigate, also with simulations, how the network boundary conditions (i.e., topology) affect solution convergence.

Chapter 2

A model of grid cell development through spatial exploration and spike time-dependent plasticity

2.1 Introduction

The unusual tuning properties of mammalian grid cells [31, 40, 48, 66, 124] have spurred a number of theoretical and experimental efforts to dissect their function and mechanisms. The periodic response of grid cells to two-dimensional (2D) animal location is relatively independent of non-spatial variables. Thus, grid cells are widely conjectured to be responsible for computing spatial displacements, by integrating self-motion cues [16, 35, 38, 48].

On the question of mechanism, grid cells from a module (defined as the set of all cells with similar spatial period and orientation [102]) appear to collectively exhibit 2D continuous attractor dynamics [125]. Consistent with this finding, network input in the form of slow depolarizing current ramps drives the spatial firing rate patterns of grid cells [32, 97]. Several hardwired models demonstrate how grid cell-like activity can arise based on recurrent circuits that exhibit low-dimensional continuous attractor dynamics [8, 16, 38, 46, 77, 78, 89]. Other models of individual grid cells are based on interfering temporal oscillations [17, 52] that are mapped into spatially periodic responses.

Nevertheless, key mechanistic questions remain unanswered. It is poorly understood how the system might form. In rats, grid cell responses emerge at 3-4 weeks of age, after eye opening, [71, 119], suggesting that the system is not hardwired at birth. Here we describe a neural network model for the emergence of grid cells based on spatial experience.

2.2 Results

2.2.1 Model ingredients: initial architecture and learning rules

Our model relies on active exploration of a spatially cue-rich environment. This work focuses on the assembly of an individual grid cell network (GCN) whose model neurons correspond to grid cells in one experimentally observed module. The GCN consists of excitatory (E) and inhibitory (I) neurons in a 5 : 1 ratio, each modeled as LNP neurons (as described in Chapter 1; for a more complete description of the simulations, see the Methods section at the end of this chapter):

$$\frac{ds_i}{dt} + \frac{s_i}{\tau_s} = \sigma_i \tag{2.1}$$

where $\sigma_i(t) = \sum_k n_k \delta(t - t_k^{spk})$ is the cell's spike train, based on Poisson samples of its instantaneous firing rate: $Poiss(n_k, f(I_i)\,\Delta t)$ (n_k is the number of spike events sampled within the kth time window, of size Δt; $f(I_i)$ is the same rectification nonlinearity as described in Chapter 1). We will now describe the input, I_i, into each cell.

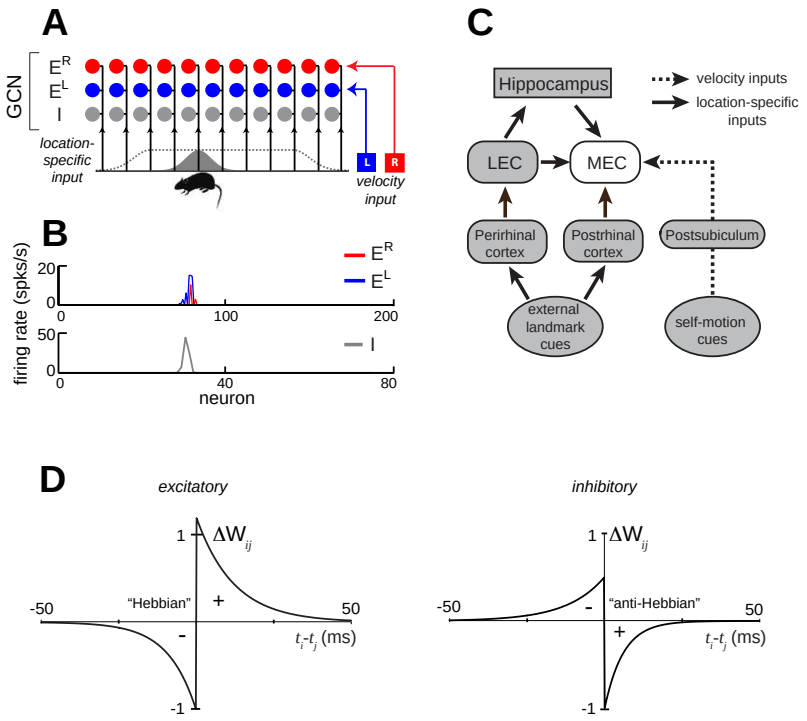

Figure 2.1: **Initial architecture and learning rule.** (A) The 1D GCN consists of inhibitory (I; gray circles) and velocity-sensitive excitatory cells (E^L and E^R and blue and red circles, respectively). All cells are assigned location-specific inputs (gray bell-shaped curve: schematic of a location-specific input; dotted gray envelope suppresses location-specific inputs near the environment boundaries). The colored boxes labelled 'L' and 'R' are the two pools of speed-modulated head-direction cells (not modeled explicitly) that supply velocity input to the GCN and project only to the E cells. (B) Snapshot of population activity derived from location-specific inputs during the learning. (C) Possible pathways for velocity and location-specific inputs to grid cells in MEC. (D) The STDP windows (kernels) used in this work, for excitatory (left) and inhibitory (right) synapses. The "+" and "-" indicate the positive and negative lobes of the STDP windows, respectively.

During exploration, the E and I populations receive inputs related to the position and self-motion of a virtual animal moving in 1D, Figure 2.1A, (see Figure A.1 for the statistics of the trajectory used for training). The former inputs, termed *location-specific*, could hypothetically derive from any number of sources connected to the MEC, including the hippocampus, Figure 2.1B, and are such that each neuron in the GCN randomly inherits a *location preference* in movement space somewhere in the environment to which it is driven to fire maximally (as we will see, this randomness in location-preference leads to apparent randomness in how cells develop connections with each other in brain space, and thus for the *mature* GCN to exhibit an atopographical organization; in order to visualize, and therefore understand, the development of patterning in the network connectivity, we will resort to displaying neurons in "topological" space, i.e., an ordering that respects their underlying connectedness). As we will see, once the GCN is mature, the location-specific inputs become ineffective in driving neural activation and setting the network state.

The other input, velocity, is multiplicative (see below) and divides the E cells into two populations: one population (E^R) receives biased speed-dependent excitation when the animal moves rightward, the other (E^L) when the animal moves leftward, Figure 2.1A. The I cells receive no velocity input. In sum, the input into each cell is given by

$$I_i = \alpha_i^{vel} \left(g_i^{rec} + g_i^{ext} \right) \tag{2.2}$$

where the recurrent inputs is $g_i^{rec} = \sum_j W_{ij} s_j$ (this term implicitly includes inputs from all cells across all populations), and the external input comprises

of two terms, $g_i^{ext} = g_i^{loc} + g^0$, where $g_i^{loc} = \mathcal{N}(x - x_i, \sigma_{loc})$ is the location-specific input ($\mathcal{N}$ is a Normal distribution, x is the location of the animal, x_i is the cell's location preference, and σ_{loc} is the width of the location-specific input into the GCN, all in units of meters in movement space) and g^0 is the constant uniform input, and $\alpha_i^{vel} = 1 + \beta(\hat{e}_i \cdot \vec{v})$ is the multiplicative velocity input ($\vec{v}$ is the animal's velocity, β is the gain of the velocity input, and $\hat{e}_i$ is the cell's direction preference equal to $(0,1), (0,-1), (0,0)$ for the E^R, E^L, and I populations respectively). We note that the velocity input, α^{vec}, is also technically derived from outside the network.

The synaptic weights (specifically E^L-to-I, E^R-to-I, I-to-E^L, I-to-E^R, and I-to-I, which will be refered in shorthand as E-to-I, I-to-E and I-to-I, respectively; we have excluded E-to-E connection to be consistent with experimental data [5, 25, 29, 89]) are initially random and weak, and are subject to change according to spike time-dependent plasticity (STDP) rules, as described in Chapter 1 in Equation 1.6. For example, the form of the STDP for excitatory synapses (i.e., E-to-I) is the conventional Hebbian form (as shown in Figure 1.4), where the synapse is strengthened (weakened) if the presynaptic cell fires before (after) the postsynaptic cell [33, 36, 60, 75], Figure 2.1D, left. For the inhibitory synapses, the window is reversed (anti-Hebbian [58]), Figure 2.1D, right. As a note, simultaneously reversing the time-axes of all STDP windows used here, which would correspond to anti-Hebbian STDP in excitatory synapses [36, 47] and Hebbian STDP in inhibitory synapses, results in quantitatively the same outcome (up to a sign-flip of an arbitrary sign during

integration, see Figure A.1).

During plasticity, neural activity is driven only by the location (g^{loc}) and velocity-specific (α^{vel}) inputs (g^{rec} and g^0 are suppressed), Figure 2.1B. We define this phase as the *plasticity phase*, and distinguish it from the *activation phase* in which all inputs are active, but learning is suppressed ($\eta = 0$ in Equation 1.6). This division of phases is common practice for learning in recurrent networks. We will return to the plausibility of this phase segregation in the Discussion section at the end of the Chapter.

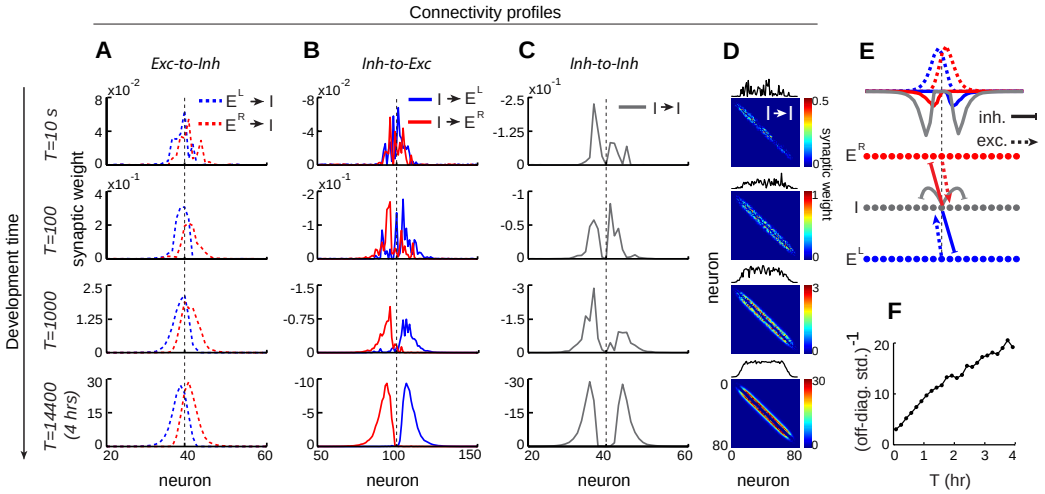

Figure 2.2: **Evolution of network connectivity.** (A-C) Profiles of synaptic weights across development: (A) E-to-I weights (E^L-to-I weight profile in solid blue; E^R-to-I in solid red); (B) I-to-E (I-to-E^L in dotted blue; I-to-E^R in dotted red); (C) I-to-I weights (in gray). Each curve represents the output projections of a single cell (located at the black dashed line) to its postsynaptic targets. Note the change in scale across rows. (D) Development of the full I-to-I synaptic weight matrix, with the off-diagonal plotted above, showing the relative translational invariance of the connectivity. (E) Top: Weight profiles in the mature GCN (from (A-C), row 4). Bottom: Schematic of mature connectivity between populations. (F) Measure of translation invariance (inverse standard deviation of the off-diagonal shown in D) as a function of time during development.

2.2.2 Development of connectivity

Emergence of "local" connectivity and weight asymmetries. Driven by the localized inputs shown in Figure 2.1B, cells with similar location preferences are activated within a short latency of each other. These short-latency activations rapidly result in stronger synaptic weights, both excitatory and inhibitory, that is "local" in the functional sense of location preference, i.e., neurons with similar location preferences are wired up together (Figure 2.2A-C, row 1).

GCN neurons also spontaneously develop weight asymmetries based on direction preference (Figure 2.2A-B). Consider the excitatory synapses, in particular the synapses from the E^R population to the I population. During rightward traversals, the positive STDP lobe (Figure 2.1D, left) strengthens weights from the E^R population of cells to I population of cells with slightly more rightward location preferences, which fire a short time later. The weight gain is not fully cancelled by the opposite lobe of the same STDP kernel during leftward trajectories, because during such runs, the net activation of the E^R neurons is lower (through of the modulatory term, α^{vel}). Thus, E^R cells project to I cells with slightly right-shifted location preferences (Figure 2.2A, row 4). The opposite happens for the E^L population.

By a similar argument, I-to-E projections acquire slight shifts in the opposite direction (Figure 2.2B, rows 3-4), because the STDP learning window is flipped when the presynaptic cell is inhibitory (Figure 2.1D, right). The I-to-

I connections remain symmetric and unbiased because I cells are not modulated by velocity (Figure 2.2C-D, rows 3-4). The symmetries and asymmetries of these network projections are summarized in Figure 2.2E (full weights are shown in Figure A.2).

Learning time. The basic architecture of the GCN is established within the first few minutes of exploration and plasticity (Figure 2.2A-C, rows 1-3). However, two key developments unfold over hours. The first is a strengthening of the weight profiles: until the weights reach a threshold in size, they cannot drive pattern formation in the population activity in brain space (next subsection). The second is a progressive increase in translation invariance of the synaptic weight profiles across cells. Translation invariance, which is important for the formation of a continuum of fixed points for analog integration and memory [127], is quantified by the increase in smoothness along the off-diagonals of the weight matrices (Figure 2.2D, F). There is a tradeoff between rapidity of weight growth and translation invariance in the mature GCN, so that major changes in learning rate either cause learning to be too slow or cause the mature network to exhibit too little translation invariance (Figure A.2). Thus, the estimate of learning time is not susceptible to order-of-magnitude changes through corresponding adjustments of the rate parameter.

We have not explored the dependence of learning time on neuron number. However, larger networks allow for more averaging, so we expect that larger networks may reach a comparable level of performance sooner. Finally, note that the accrual of $\sim$ 4 hours of plasticity time, as taken for matura-

tion by the GCN in Figure 2.2, may span half a day or several days to weeks in animals, if plasticity occurs only part of the time during exploration and exploration itself occurs only in small bouts over the day.

2.2.3 Emergence of patterned activity

Next, we examine how the growing synaptic weights shape the population activity (in brain space) and single-cell spatial tuning curves (in movement space). In all that follows, neural responses are probed in the activation phase, meaning that all the inputs active, but plasticity is turned off.

Population activity and path integration. Early in development, population activity is roughly uniform across cells (Figure 2.3A-B, rows 1-2). In the presence of the uniform excitatory input, the relative contribution of the location-specific input is small, and the largest source of activity modulation is the velocity input (Figure 2.3A-B, rows 1-2).

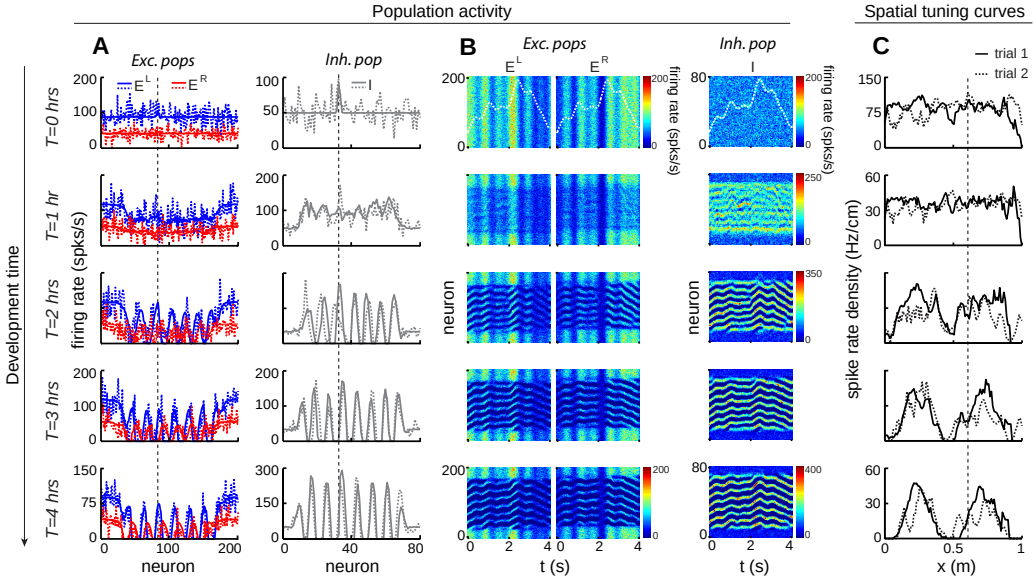

Figure 2.3: **Evolution of population activity, path integration, and spatial tuning.** (A) Snapshot of population activities in the activation phase during a leftward run ($v_{rat} = 0.4$ m/s), across development (different rows). (A) E^R (red) E^L (blue) populations; (B) I population (black). Solid lines: mean rate used to generate Poisson spikes; dashed lines: filtered spiking activity (spike trains convolved with Gaussian, $\sigma = 15$ ms). The location-specific bump in the population activity (location marked by the vertical dashed line) is obscured by the larger-amplitude spike variability. (B) Population activities over a 4-second interval (spiking activity filtered as in (A)). The white dashed line (top row) tracks the location-specific input. Note the difference in movement gain between the white line and the population phase. (C) Spatial tuning curves (as defined in Chapter 1; also, see Appendix A) of a cell in one of the excitatory populations across two trials, measured across development (the two trials consist of different 10-second variable-velocity runs that span the enclosure, beginning at the same initial population pattern phase and similar initial location of the animal within some tolerance, see Appendix A). The location-specific input of this cell drives (insignificant) activation at the dashed vertical line.

The growing synaptic weights eventually destabilize uniform activity states within the GCN, and force a periodic patterning of the population activity as viewed in topological space (Figure 2.3A-B, rows 3-5). Cells in the E^R (E^L) population drive the activity pattern rightward (leftward), by exciting I cells with a rightward (leftward) bias, which in turn inhibit the left (right) flanks of the corresponding E bumps (Figure 2.2E). When the animal is still, the two E populations exert an and equal and opposite influence, and the pattern remains stationary. When the animal moves, one of the E populations receives biased excitation (Figure 2.3A, row 5) and succeeds in driving the activity pattern along its preferred direction. Thus, the GCN behaves as a path integrator by flowing its population activity pattern (the patterns of the three populations are yoked together via the connectivity) in an amount corresponding to the movement of the animal. The integration accuracy of the GCN improves with the number of neurons, and decreases with the variability of the spiking process (i.e., by using a model of neural spiking that is sub-Poisson), Figure A.3.

(Note: In this model, both E and I populations become patterned. This result is a consequence of prohibiting direct E-to-E coupling. When E-to-E coupling is permitted, it is possible for E cells to be patterned while I cells remain largely unpatterned and exhibit minimal spatial tuning, Figure A.3.)

Spatial tuning curves of individual cells. During early development, GCN neurons are not spatially tuned (Figure 2.3C, rows 1-2), as evident by their lack of well-defined spatial firing fields in movement space, consistent

with the uniform, unpatterned population activities in brain space at the same stage (Figure 2.3A-B, rows 1-2). However, once the population activity begins to become patterned, and the GCN begins to be able to path integrate, stable spatial tuning curves appear (and at a relative lag – compare the emergence of the population activity patterns in rows 2-3 of Figure 2.3C with the spatial tuning curves of a single cell in rows 3-4; quantification to follow).

By the time the GCN has matured (i.e., integrates its inputs to sufficient accuracy), the GCN's internal location estimate has decoupled from the location-specific input. This is because there is essentially a free parameter linking the speed of the internal estimate from the animal's speed. This gain factor depends on both the degree of asymmetries in the recurrent weights and the gain of the velocity input in the GCN. In fact, by the end of learning, the recurrent weights have become so dominant in determining network activity that the location-specific inputs even if present have become irrelevant.

2.2.4 Properties of the mature network

The connectivity structure of the mature GCN has aperiodic boundaries (cells at one edge of the neural sheet do not connect to neurons at the opposite, and neurons of a given preferred spatial phase are not connected with all others of the same phase). This is a consequence of the fact that movement space is itself aperiodic. As a consequence, cells that lie at the edge of the network (i.e., cells whose location preferences are at the boundary edges) display poor grid-like spatial tuning, but those in the bulk – a majority of the total – are

very grid-like, Figure 2.4A. We quantify how grid-like the spatial tuning curves are using gridness score, defined in Appendix A, which is essentially the power of the largest spatial frequency in the normalized spatial tuning curve.

(Note: If weak synaptic plasticity is permitted in the *activation* phase of the *mature* GCN, the same STDP rules will wire together all cells with common preferred phase (i.e., those cell's with identical spatial tuning curves), producing a network that is topologically equivalent to a single-bump patterned network with periodic boundary conditions [15, 16, 46], Figure A.4. However, plasticity in the activation phase is a strong positive feedback process, and generically leads to instability and bias in the GCN dynamics, Figure A.4. Therefore, it remains an open question whether there is a stable way to developmentally obtain a grid cell network with periodic boundaries.)

The mature GCN exhibits the key signatures of continuous attractor dynamics [15, 16, 38, 125]. For instance, neurons in the mature GCN exhibit very similar spatial tuning periods, Figure 2.4A, C. The distribution of spatial phases, ϕ^{α}, is uniform (Figure 2.4D), and relative spatial phases between cells, $\delta^{\alpha\beta}$, are stably preserved even when the spatial tuning of individual cells drifts over time (Figure 2.4E).

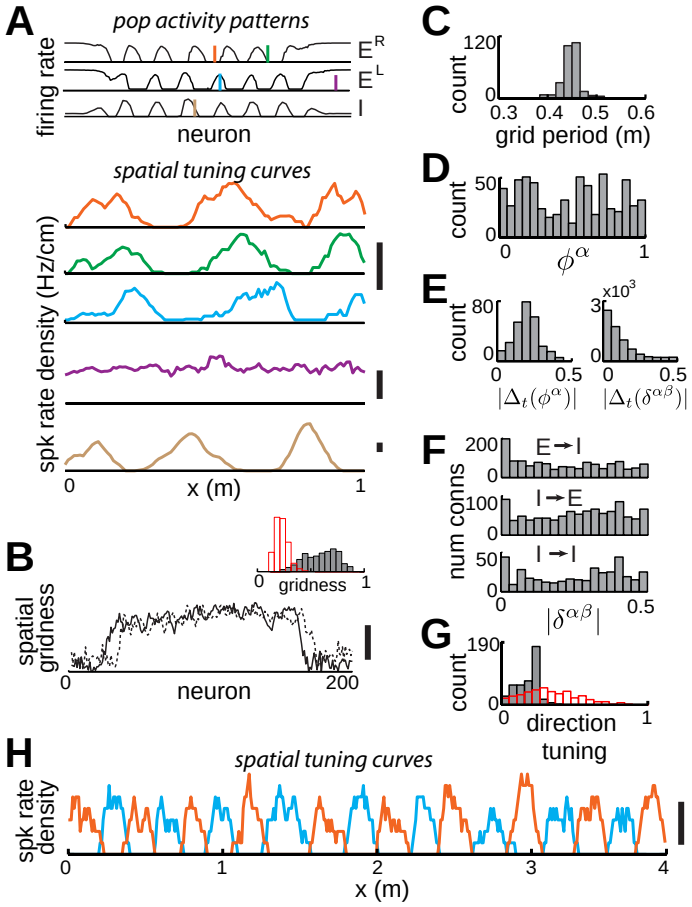

Figure 2.4: **Spatial tuning in the mature network.** (A) Top: Snapshot of the population activity pattern in all three populations. Vertical colored bars indicate which cells are probed for their spatial tuning curves (shown below). Bottom: Spatial tuning of a sample of cells from the mature GCN, obtained from a 10-second variable-speed run across the environment. Scale bars: Mean spike rate of 60 Hz (averaged over 1 cm). (B) Spatial gridness scores across the E^L (solid black) and E^R (dotted black) populations (scale bar: gridness of 0.5). Inset: Distribution of these scores (gray bars), as well as when the constant external inputs are "lesioned" (red bars; see Methods). For cells in the population whose spatial tuning curves have spatial gridness ≥ 0.5, grid periods are narrowly distributed (C), and spatial phases (ϕ^α) are

approximately uniformly distributed (D). (E) Left: Distribution of the change across trials in the spatial phase of individual cells (shown is the distribution of the magnitude of these changes, $|\Delta_t(\phi^\alpha)|$, assuming the same initial conditions in the two trials, pooled over all cells α that appear in (C-D)). Right: Change in the relative spatial phase between pairs of cells across trials, $\delta^{\alpha\beta}$ (shown is the distribution of the magnitude of changes in relative phase, $|\Delta_t(\delta^{\alpha\beta})|$). (F) The number of connections between cell pairs as a function of the spatial relative phase difference between the cells, $|\delta^{\alpha\beta}|$, for the E-to-I (top), I-to-E (middle), and I-to-I (bottom) weights. Any synapse whose strength exceeds 5% of the strongest synapse of that type counts as a connection for this plot. (G) Direction tuning scores in the mature GCN (gray bars), and after "lesion" (red bars) (see Appendix A; plotted for cells in (B-C)). (H) Spatial tuning curves in a 4-meter space (four times larger than the training environment) for two cells from (A). The trajectory is a single, constant-speed unidirectional sweep ($v_{rat} = 0.4$ m/s) across the space.

When the uniform excitation, g^0, is removed, the population pattern is lost and gridness scores drop precipitously (Figure 2.4B). The GCN neurons instead display strongly direction-tuned, head direction cell-like responses (Figure 2.4F), consistent with results from experimental studies in which inputs to MEC were lesioned [9, 10].

Cells in the mature GCN receive input from other cells with disparate spatial tuning curve phases, Figure 2.4G. This is the case despite the highly structured, "local" weight profile of the mature GCN.

Spontaneous periodic tuning in new and large environments. Cells in the mature GCN generate periodic spatial responses on the very first run-through in much larger environments than the training environment (Figure 2.4H). This is possible because, as described above, mature GCN neuron responses are generated by integration of velocity inputs, independent of external location-specific inputs.

2.2.5 The emergence of patterning is abrupt

To quantify how grid cell-like features emerge over development, we examine several metrics of patterning and stability as a function of time during development (Figure 2.5; Figure A.5).

As weights gradually reach and then exceed a threshold strength, patterning of the population activity in brain space emerges abruptly (around T_{crit} in Figure 2.5A), because of a weight-driven linear instability in the GCN dynamics. The population activity pattern is fully formed, in terms of reach-

ing a maximal gridness score (computed on the periodic population activity pattern just as it is for the periodic spatial tuning of individual cells, see Appendix A), by about the time labelled T_{patt}. To display spatial tuning, the GCN must be capable of sufficiently good path integration over the trajectory on which tuning is assessed, in addition to displaying patterned population activity. Thus, the emergence of grid-like spatial tuning in movement space consistently lags, and is smoother in onset, than population activity patterning in brain/topological space (Figure 2.5B). Spatial tuning gridness scores for shorter trajectories are sharper and better probes of population activity patterning than longer trajectories, because they involve less integration (Figure 2.5B, black versus gray). The evolution of across-trial spatial tuning stability (i.e., correlation between spatial tuning curves on consecutive trials with same initial conditions) and spatial coherence (how well spiking in a spatial bin in movement space is predicted by spiking in neighboring spatial bins, see Appendix A), Figure 2.5C-D, closely resembles the evolution of spatial tuning.

(Note: Early in development ($T < T_{crit}$), cells exhibit an artifactual spatial tuning over short trajectories. This is because of their strong velocity modulation, coupled with the fact that variations in velocity are not averaged over space in short trajectories. As recurrent weights develop, the artifactual spatial tuning and spatial coherence decline slightly before T_{crit} (black curves in Figure 2.5B, D). This effect is weaker in longer trajectories because of averaging (gray curves). However, longer trajectories obscure the emergence of grid-like tuning because of the accumulation of path integration errors. The

emergence of spatial tuning can also be obscured in datasets with fewer cells and trials, Figure A.5.)

Finally, the relative phase of the spatial tuning curves (defined as the phasic offset of one cell's spatial tuning curve with respect to the others) between cell pairs, Figure 2.5E, is a good measure of population activity patterning, because patterning in the relative phases is arguably more abrupt and possibly emerges sooner than spatial tuning gridness or spatial coherence (compare Figure 2.5E with Figure 2.5B,D). (Note: The relative phases between cells, though clearly themselves patterned (in the sorted population), are *not* constant over development (see the gradual expansion of the features in Figure 2.5E), because of a slight gradual expansion of the period of the population pattern after T_{patt}, Figure 2.5F. This expansion is partly responsible for the oscillatory variation in the spatial gridness and coherence measures late in development, Figure 2.5B,D.)

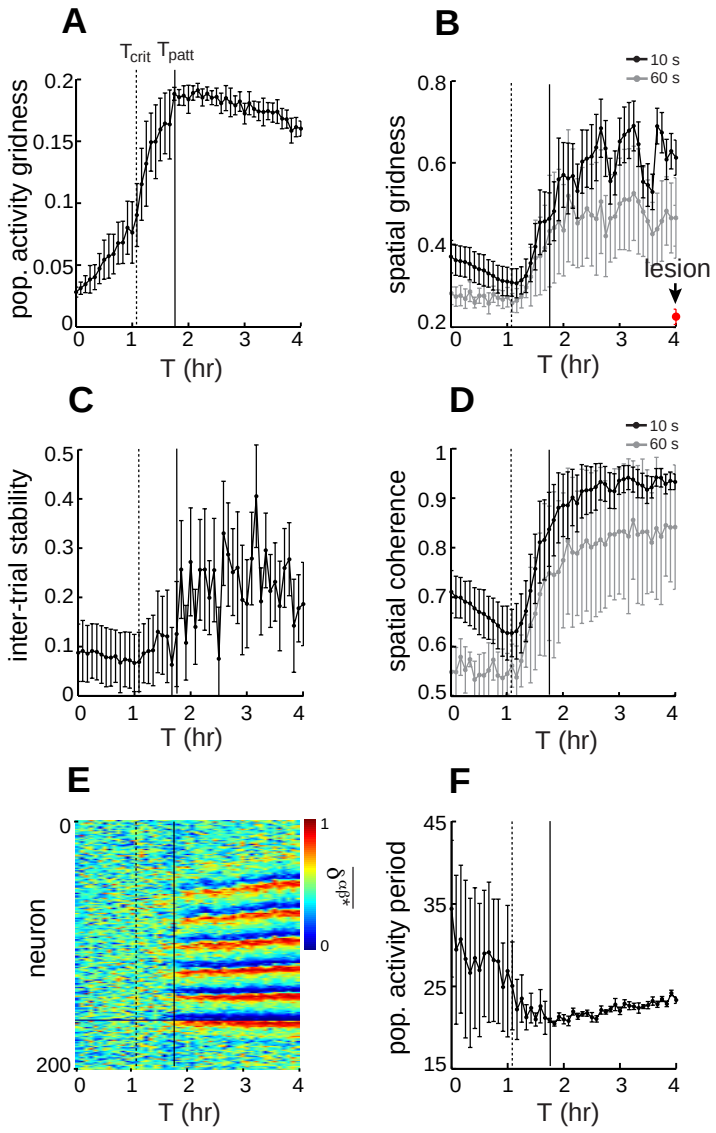

Figure 2.5: **Abrupt onset of patterning.** (A) Development of population activity gridness in the GCN (only E^L neurons are used in this analysis). At several points during development, the GCN is probed in the activation phase. At one such point, we collect a set of population activity snapshots (one every 5 ms, for one second); from this set, an average gridness score (error bars are $\pm$ one standard deviation) is obtained. Long vertical lines in (A-F) designate

$T_{crit} \approx 1.08$ hours (dotted) and $T_{patt} \approx 1.75$ hours (solid), defined as the times at which the population pattern gridness score reaches its half-maximum and maximum value, respectively. (B-D) Development of gridness (B), inter-trial stability (C), and spatial coherence in the spatial tuning of cells (see Appendix A). Black dots (B-D) and gray dots (B,D) are average scores from a set of 10-second and 60-second trajectories (see Appendix A), respectively. (Average computed across n = 263 E cells and n = 10 trials. The same 10 trials are used at each point in development. Cells are included if their spatial tuning scores exceed 0.5 in the mature (at 4 hours) GCN). Red dot: average gridness of spatial tuning in the GCN with "lesioned" feedforward input. (E) Relative phases ($\delta^{\alpha\beta_*}$; see Appendix A) of E^L cells (labeled by α), with respect to one reference cell (β^*; dark horizontal line marks the reference cell), averaged across trials (same 10-second trajectories as in (B-D)). (F) Development of the period of the population activity pattern (in neurons, see Appendix A), estimated using the same data as in (A).

2.2.6 Signatures of development in cell-cell correlations, direction tuning, and speed tuning

By definition, population activity patterning involves the stable co-activation of some cells and counter-activation of others. Thus, hallmarks of population patterning should be visible in the emergence of stable cell-cell correlations (measured as the Pearson's correlation coefficient between the spike trains of the two cells).

The key prediction associated with the emergence of population activity patterning is that the pairwise correlation distribution should develop a *uniform component*. In simulated *in vitro* conditions, where the velocity inputs are absent (the animal is motionless), the immature GCN exhibits only very weak correlations (Figure 2.6A rows 1-2). With development of patterning, a uniform platform of correlations emerges (Figure 2.6A, row 3; especially see inset). (Note: The uniform component is not large and not fully flat because, in the absence of velocity inputs, the population activity pattern does not efficiently flow; as a result, distant pattern phases and thus the larger anticorrelations are simply not well-sampled, cf. Figure A.6). The predicted emergence of a uniform component in the *in vitro* correlation distribution is consistent with MEC slice data showing that (anti)correlations grow with maturation [71]. Over development, the standard deviation of the *in vitro* correlation distribution grows slowly (Figure 2.6B) even though population activity patterning is abrupt and the uniform component (which is small in amplitude) in the correlation distribution appears suddenly (insets, Figure 2.6A).

When correlations are measured *in vivo*, the velocity input induces strong (anti)correlations early in development, which can obscure the emergence of correlations based on pattern formation of the population activity, Figure A.6. However, *in vivo* data can provide a reasonable proxy for *in vitro* correlations, and show more clearly the abrupt emergence of a uniform component in the pairwise correlations, if the correlations are based only on segments of the trajectory when the animal is moving slowly (thus the velocity input is close to zero), Figure 2.6A-B, gray curves.

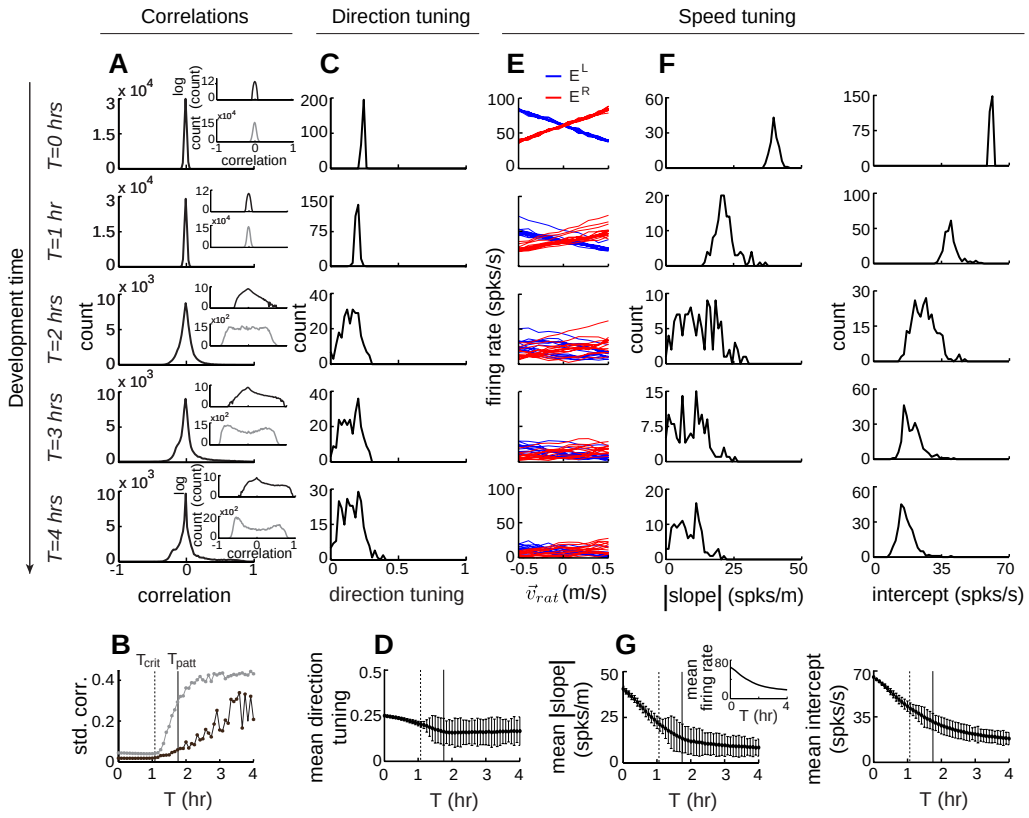

Figure 2.6: **Windows into development: Correlations, direction tuning, and speed tuning.** (A) Main plots: Development of *in vitro* pairwise activity correlations (Pearson's) between E cells. Pairs drawn from cells used in Figure 2.5B-D; correlations assessed over a 120 second trial (see Appendix A; same trial and same cells used for remaining panels). Insets: Top: Semi-log version of main plots.Bottom: *in vivo* correlations computed only for low-velocity ($v_{rat} < 0.1$ m/s) trajectory segments. (B) Finely sampled evolution of the standard deviation of the pairwise correlation distribution for the *in vitro* (black) and low-velocity *in vivo* (gray) cases. Vertical lines as in Figure 2.5A. (C) Histogram of direction tuning scores (which measure the strength of direction tuning and range between 0 and 1, see Appendix A) for cells in the E population. (D) Finely sampled evolution of mean direction tuning (error bars are $\pm$ one standard deviation, here and in rest of figure). (E) Examples of speed tuning in E cells; color-coded according to preferred direction. (F)

Absolute value of the slopes (left column) and intercepts (right column) of the regression lines used to fit speed tuning curves. (G) Finely sampled evolution of slopes and intercepts of speed tuning. Inset: Evolution of mean firing firing rate for cells in E population.

Over GCN development, the mean strength of direction tuning (how sensitive the cell is to the animal's head direction – in this case, left vs. right) decreases slightly, while the variance grows (Figure 2.6C-D). (Note: The inferred direction tuning (through the mean vector length statistic of a circular variable) depends on exploration speed: in a fixed network, faster speeds produce larger vectors (Figure A.6), so comparisons across development in experiment must be made carefully, with statistically matched trajectories.)

Speed tuning, the sensitivity of a cell's firing rate to the animal's speed, broken down by speed along its preferred and antipreferred directions (see Appendix A and Figure 2.6E, red and blue), decreases in strength over development, as quantified by the absolute value of the slopes and the intercepts of the firing rate vs. speed tuning curves (Figure 2.6F-G). This decrease can be attributed to the increasing influence of recurrent inputs on cell firing over development.

(Note: If speed tuning is instead computed by averaging together changes in firing rate as a function of speed without taking into account motion direction (and thus without taking the absolute values of the firing rate-input speed curves for different directions before averaging), the results are different (Figure A.6), but largely consistent with experimental results that use this definition of speed tuning [118].)

2.2.7 Multi-period GCNs

The present model allows us to parametrically explore possible mechanisms for the differences in grid period of GCNs along the dorsoventral (DV), or longitudinal, axis of the MEC.

If put in place over development, a smaller gain in velocity input, β, a smaller temporal width of the STDP windows, a longer intrinsic biophysical time-constant (τ_s) in neurons or synapses, or a larger learning rate for the inhibitory synapses (I-to-E and I-to-I), all result in larger periods in the spatial tuning curves(Figure 2.7A; Figure 2.7B, row 1, columns 1-4, respectively; Figure A.7).

Decreasing the velocity gain leaves the population activity period unchanged but reduces the ability of animal velocity to translate the population activity pattern, thus increasing the spatial tuning period (Figure 2.7B, column 1, rows 1-3). In general, changing developmental parameters can affect two or more distinct properties of the GCN that influence the spatial tuning period, sometimes in opposite directions. For instance, as the width of the STDP windows decreases, the period of the population activity pattern shrinks, but the GCN's velocity sensitivity increases (pattern translates faster for a given animal velocity). In total, the enhanced velocity sensitivity wins out over population activity pattern expansion, and grid periods actually increase (Figure 2.7B, column 2, rows 1-3).

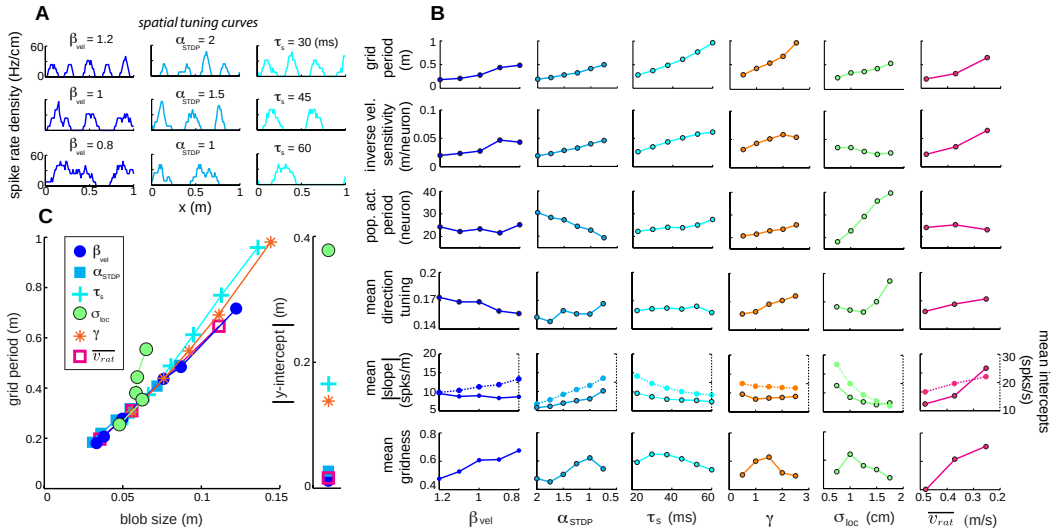

Figure 2.7: **Mechanisms for controlling spatial period across modules.** (A) Differences in the velocity gain, β_{vel}, the scaling of the STDP time-constant, α_{STDP} (see Methods), and biophysical time-constant, τ_s (columns 1-3, respectively; all other parameters held fixed) result in systematic variations in the spatial tuning period (assessed over a 10-second, constant-speed (0.4 m/s) sweep through the environment). (B) Metrics of the mature GCN (rows), and how they vary as a function of parameter settings over development. Note that many scales on the abscissa are inverted. GCN metrics (in order): average spatial tuning period, inverse velocity sensitivity (see Appendix A), population activity period, average strength of direction tuning, average strength (slope) of absolute values of speed tuning curves, and average intercepts of the speed tuning curves. Parameters (in addition to those mentioned in (A)): the scale of learning of inhibition (γ; see Appendix A), the width of the location-specific developmental input (σ_{loc}), and the mean speed of exploration during development ($\overline{v_{rat}}$). (C) Relationship between blob size (the widths of the activity bumps in the spatial tuning curves; see Appendix A) and grid period, for different spatial tunings that result from varying different single parameters. Each color marks the effects of variations in a single parameter.

Although variations in any of several parameters can in theory produce a range of spatial tuning periods (Figure 2.7B, row 1 and Figure A.7), they do so in different and experimentally distinguishable ways. A mechanism based on velocity gain alone would predict weaker direction tuning more ventrally (Figure 2.7B, column 1, rows 1-3), consistent with recent reports [44], and sufficiently explained by GCNs across the longitudinal axis to having identical connectivity. By contrast, increasing the spatial tuning period ventrally by increasing the STDP window width would predict narrower lateral inhibition (and a smaller population period), together with a reduction in the intercepts and slopes of speed tuning (Figure 2.7B, column 3). Increasing τ_s is predicted to decrease the intercepts and slopes of speed tuning more ventrally (Figure 2.7 B, column 3, rows 4-5), while not significantly affecting the strength of direction tuning; these effects are in contrast to the former two mechanisms. A 3-fold variation in τ_s can, in our simulations, explain most of the (roughly 10-fold) variation in grid period along the DV axis (data not shown), consistent with the experimentally estimated DV variation in the membrane and synaptic integration time-constant of layer II stellate cells [42].

These and other mechanisms for varying the spatial tuning period – scaling the learning rate (or equivalently, the overall strength) of lateral inhibition, scaling the width of the location-specific inputs, or differentially scaling the velocity input across GCNs – produce additional experimentally distinguishable predictions for DV variation across modules (Figure 2.7B, columns 4-6). A corollary of the predicted effect of exploration speed on grid period is

that if animals are forced to move more slowly during development, all spatial tuning periods would be larger than in control animals (Figure 2.7B, column 6).

Grid cells in MEC with different periods look like globally rescaled versions of one another; therefore, a plot of spatial tuning period against the mean width of the individual spatial firing fields (i.e., "blob size") should be linear, with zero intercept. Figure 2.7C illustrates which parameter scalings produce an appropriate relationship between spatial tuning period and blob size. We find that the velocity gain, STDP window width, and velocity standard deviation parameters produce the smallest offsets from a zero-intercept line.

We also plot the relationship between spatial tuning period agains the overall strength of the recurrent inhibition in the network connectivity, and find that stronger inhibition leads to larger spatial tuning periods, inconsistent with the data [5]. On the other hand, in other simplified non-developmental models of grid cells (e.g. [16], which uses neurons that both excite and inhibit other cells in the network, inconsistent with experimental findigs), leads to the correct dependence of spatial tuning period on inhibition. In more realistic, conductance-based neuron models, inhibitory inputs can effectively shorten the biophysical time-constant; if this effect is strong enough, more inhibition might result in smaller-period spatial tuning (Figure 2.7B, column 3, row 1). Alternatively, other variables explored above also influence grid period and in MEC, may dominate over the effects of inhibition strength.

We reiterate that the above discussion involved changing parameters in the developing network and then measuring certain properties of the network at the end of development. However, under certain experimental conditions, spatial responses can change instantly [3, 4]. Several of the parameters considered above are related to plasticity and, because of their slower time-scales, cannot be responsible for such rapid changes to the spatial responses. The parameters that are capable of driving fast change (e.g. through neuromodulation) include the velocity gain, the strength of recurrent inhibition (which developmentally was equated with the learning rate of inhibitory synapses), and the biophysical time-constant of neurons.

Analysis of grid cell data in such rapid rescaling experiments indicates that the underlying population activity pattern does not change [125]. However, according to our model, varying the strength of inhibition in the mature GCN changes the population pattern (Figure A.7), as does changing τ_s (Figure A.7). This lends support to the possibility that, regardless of the mechanisms underlying the gradient in spatial tuning periods along the longitudinal axis of the MEC, fast rescaling is driven by a change in the gain of the velocity inputs to the mature grid cell system (Figure 2.7B and Figure A.7) because it is the only parameter that does not change the period of the population activity pattern. The possibility that rapid rescaling is caused by a gain change in the velocity input can be tested by looking for variations in direction and speed tuning that are predicted to accompany such a change, as predicted by our model (Figure A.7).

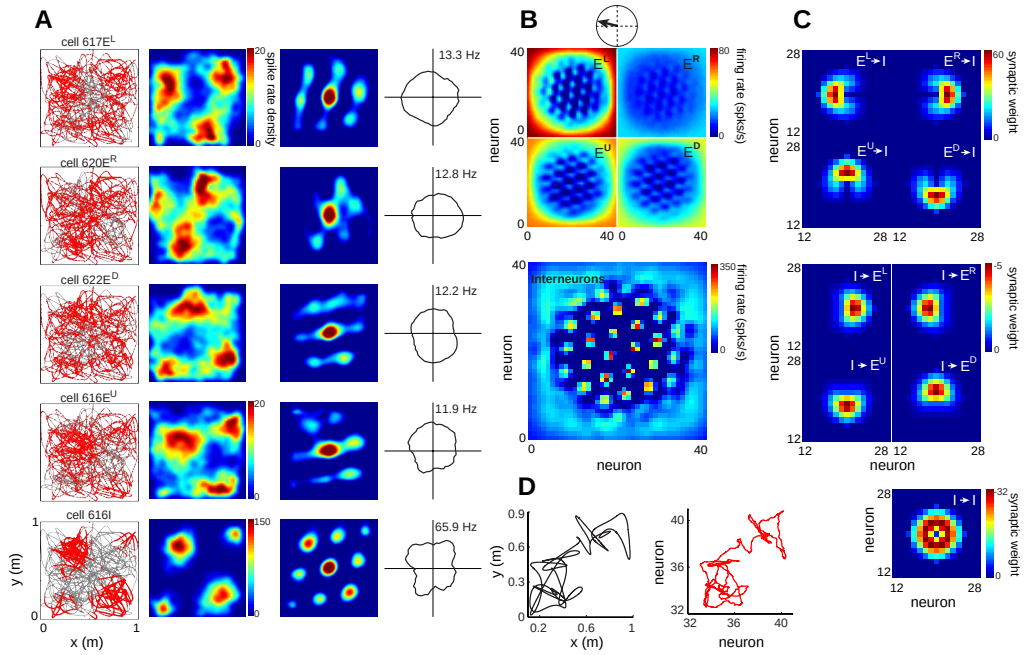

Figure 2.8: **Two-dimensional grid cell network.** (A) Spatial tuning of 5 cells in a mature 2D GCN, following development. The trajectory is a 5 min segment taken from actual animal motion recorded in a square environment. The location-specific input used during training was removed for this trajectory. Column 1: Spiking response of a cell (each red dot = 1 spike). Column 2: Spatially filtered response (bin = 2 cm; convolved with Gaussian, $\sigma = 3$ cm). Column 3: Autocorrelogram of column 2. Column 4: direction tuning curve, with peak firing rates indicated. (B) Snapshot of population activity when the animal is moving in the direction of the black vector shown above, with cells ordered according to their location-specific training inputs (E cells above; I cells at bottom). (C) Recurrent weights in the mature GCN. Each matrix shows output projections from one cell located at the GCN center. (D) Left: 30-second recorded animal trajectory. Right: Path-integrated estimate of the same trajectory by the GCN (GCN estimate is defined as the location of one of the population pattern bumps in the neural sheet (B) as the pattern flows with input velocity).

2.2.8 Two-dimensional network

The principles illustrated above for 1D GCN development carry directly over to 2D (Figure 2.8). There are now four sets of E cells, defined by whether the cells receive cosine-tuned inputs for north (E^U), south (E^D), east (E^L), or west (E^R) motion. Each E set contains $40\times40 = 1600$ cells, as does the I population, so that the E:I cell ratio is 4:1. The reason to choose four distinct sets of E cells with discrete direction tuning is to illustrate the systematic weight shifts of each population. We obtain qualitatively similar functionality if each E cell is independently and randomly assigned a preferred direction from a continuous and uniform distribution over all possible angles, Figure B.4. We already explored the effects of noise and variability (stochastic GCN neurons and random trajectories) on development in 1D; thus, for tractability, we employed deterministic cells (see Methods) and a simple trajectory to train the 2D system (see Methods). For testing, however, we reverted to fully stochastic dynamics, as in 1D. The STDP kernels are identical to those used in the 1D GCN, and other parameters are similar (Methods and Appendix A).

Cells in the mature 2D GCN display grid-like spatial tuning in response to velocity inputs (Figure 2.8A). (Note that the trajectory is 5 min long, with no corrective mechanisms from outside the GCN to reduce the accumulation of path integration errors over this long interval.) As expected, the population activity underlying spatial tuning is itself patterned (Figure 2.8B). The population pattern flows in proportion to and in the direction of animal displacement (Figure 2.8C); therefore, the GCN performs path integration.

The evolution of *in vitro* correlations, direction tuning, and speed tuning in the 2D GCN qualitatively track the trends from 1D development (Figure B.4; for a description and comparison of *in vivo* correlations, see Figure A.6). In particular, the *in vitro* cell-cell correlation distribution is initially narrow and centered around low correlation values, but evolves to display long tails that reflect strong pattern-related correlations (Figure B.4). Also consistent with 1D is the reduction in direction and speed tuning with development (Figure B.4).

2.3 Discussion

2.3.1 Summary

We have presented a model for the development of grid cell networks with continuous attractor dynamics, a recently substantiated property of grid cells in animals [125]. This experience-dependent model is a proof-of-principle demonstration of how grid tuning and path integration functionality might arise through synaptic plasticity, with no assumptions about topograpical order in the GCN. The mature cells in our model resemble layer II grid cells: Mature cells have strong grid-like spatial tuning and weak direction tuning [95], but lose their gridness and become strongly directional when the external uniform inputs are removed [9]; consistent with existing data on connectivity in MEC layer II [25, 29, 89], the principal (E) cells interact only through inhibitory interneurons; over development, gridness emerges abruptly and speed tuning decreases [118], and the *in vitro* correlation strengths increase [71].

2.3.2 Assumptions of model

If the role of grid cells is to estimate location during navigation, it may seem like we are putting the cart before the horse by assuming that location-specific inputs exist during GCN development. However, in our model, GCN development is a process of bootstrapping on spatially informative input derived from *external* cues in highly familiar cue-rich environments, so that the mature GCN becomes capable of *autonomous* spatial estimation in novel and cue-poor environments through integration of *internal* self-motion cues. This is an important gain in functionality, because, as noted in Chapter 1, a major computational challenge of navigation involves self-localization in novel spaces and across relatively featureless stretches of familiar environments.

Location-specific input to the GCN is assumed to derive from multisensory constellations of proximal and distal external cues during exploration around the familiar home nest area, possibly via other spatially tuned cell types of the hippocampal formation, including place cells [71, 119], border/boundary cells [7, 74, 99], and landmark-specific LEC cells [27], or via bottom-up pathways including the visuo-spatial stream through the postrhinal cortex [65]. We found that pattern formation can proceed with sparser or less uniform location-specific inputs, but the development of translation invariance and, thus, path integration, suffers.

The suppression of location-specific inputs at the boundaries of the environment during plasticity may be performed by inhibitory border cells, similar to those found in the subiculum [104]. The network "edge" is not its

topographic edge; it merely consists of cells that receive location-specific input from the boundaries of the training environment. These cells are edge cells, topologically speaking, because they are missing half of their potential lateral partners: the GCN contains no cells whose input preferred location is directly adjacent to the edge cells', if it is outside the environment. If border cells were to simply globally inhibit the GCN at the training environment boundary, the result would be a weaker activation of the then-active GCN cells (the edge cells). Attenuated activation of cells at the GCN edge during plasticity results in a mature GCN with less pinning and more accurate path integration (even though the GCN is tested in the activation phase without attenuated input at the borders). Boundary-evoked activity suppression might also be linked to the systematic orienting of grid fields in square environments (Figure B.4), as seen in experiment [103].

The self motion-based velocity inputs required in our model are likely derived from a combination of vestibular, optic flow, motor efference and proprioceptive cues [20, 24, 81, 107]. The main requirements for the velocity inputs are that the direction of movement be encoded by unimodal direction tuning curves, and that speed inputs (whether they arise from the same or a different pathway than the direction input) modulate the overall activity level of the GCN subpopulation corresponding to the present motion direction.

GCN dynamics are divided into plasticity and activity phases to avoid the deleterious effects of positive feedback associated with STDP, which tends to create discrete fixed points at the cost of translation invariance. Restricting

recurrent input during plasticity breaks the positive feedback loop; such procedures are widely espoused in the learning of continuous attractors or when learning the statistics of the external world [49, 50, 55, 56, 105]. *Activation* of recurrent inputs may occur during some runs in the home environment and when the animal explores test environments away from home [71, 119]. Or, the network may spend a fraction of each theta-cycle in an activation phase and another fraction in a plasticity phase, if modulators can control the alternation of recurrent synaptic transmission and plasticity on that time-scale [50]. In either case, the neural response collected in a given environment would sample from both plasticity and activity phases and thus would reflect the contribution of recurrent weights.

2.3.3 Questions for the future and relationship to existing work

There have been a number of papers examining development of continuous attractor networks. Below, we highlight a subset of such models, and also describe their particular shortcomings.

In [49], a continuous attractor network is organized by supervised learning – the desired population activity patterns are imposed on the network, and an error-driven rule makes these patterns permanent. In [105], the desired patterns are imposed on the network, and are consolidated by associative learning rules. However, to train a GCN with the desired patterns would require grid-like population pattern inputs that translate with animal motion, just like mature grid cells.

Plasticity in [105] is governed by firing rates not spikes. Modeling noisy spiking allows us to make an estimate of learning time; but other differences between spiking versus rate models are less significant because the key time-dependence of spikes in our model arises from the underlying time-varying rates, and the temporal asymmetries of our STDP rules play a functionally similar role to the asymmetric way in which pre- and post-synaptic neural firing rates drive plasticity in [105].

In [68] and [78], competitive learning rules act on location-specific inputs to produce stripe-like or grid-like spatial tuning. However, in these models, velocity inputs do not influence the network's spatial response, and the network is unable to path integrate; spatial tuning remains entirely dependent on the continued presence of location-specific inputs. In [38], associative plasticity rules acting on traveling activity waves in the neural sheet generate grid-like population patterning [38]. However, the resulting network is topographically organized, and there is no mechanism to associate activity patterns with animal location or velocity, so the model does not produce spatial representations.

The present model overcomes these hurdles to show that simple associative rules can result in the development of periodic, path integrating neural representations with the help of inputs that do not possess such features. However, fundamental questions about how the brain could form continuous attractors remain unanswered by our and all other works on the topic: Are translation invariant training inputs necessary for building a translation-

invariant network? Are infinitesimally small weight changes (and thus a very large dynamic range in synaptic weights) necessary to overcome noise and non-uniformity in the inputs? Is either the suppression of weights during plasticity or an alternation of learn-and-erase phases necessary to suppress positive feedback instability? Our next steps are to attempt to understand how the brain might solve these problems.

2.3.4 Predictions

Our model is robust – but therefore also non-specific – in the sense that various parameters may be varied substantially and yet produce a functional GCN. The model is insensitive to certain modifications of the STDP windows, for which there are many suitable combinations (for one example, see Figure A.1; various other examples: data not shown), although a temporal asymmetry is required. STDP windows for the three types of synapses between the E and I populations are under-constrained by the experimental data, even though there is support for the windows we used [33, 36, 58, 60, 75].

The GCN can be modified to model MEC layer III rather than layer II, by adding E-to-E connectivity and a larger feedforward velocity gain (thus increasing the strength of direction tuning; data not shown), and is robust to adding direct velocity inputs to the inhibitory cells. When velocity inputs drive I cells, these cells also develop asymmetries in their outgoing weight profiles and become direction selective.

We can modify the sign and shapes of the STDP windows to generate

anti-grid spatial tuning (constant background firing with inactivity at every vertex of a triangular lattice) in inhibitory neurons while the E population response remains grid-like [89]. For all these reasons, the key predictions of our model are not a specific set of STDP windows or weight profiles, but certain qualitative features and parametric trends, described next.

The model predicts that spatial experience is necessary for grid cell development. Restriction of spatial exploration during the normal period of grid cell development should delay or – if development occurs in a critical period that is not extended by experiential deprivation – prevent development. Changing the trajectory statistics (e.g., restricting the animal to a radial or linear track (Figure B.4) or changing the spatial metric of the environment (e.g., raising the animal on a non-Euclidean surface, as in [69]) is predicted to lead to distortions in network wiring and population patterning, and thus to qualitatively different single neuron spatial responses.

The prediction that velocity inputs are necessary for development is consistent with the experimental observation that head direction responses are stable before grid cells [71, 119]. It is unclear whether the predicted location-specific inputs arrive in MEC during development, though both place cells and border cells, which might supply such inputs, do display spatial tuning before grid cells [7, 71, 119]. Moreover, while the location-specific inputs implicitly carry all motion information (the time-derivative of location is velocity), such inputs would fail, in our model, to induce the requisite asymmetries in the weights needed for path integration. Thus, an *explicit* velocity input is neces-

sary for development. It follows that the replay of trajectories (e.g. in sleep), if unaccompanied by the proper velocity inputs [11], is itself insufficient for training the GCN.

The GCN weights are predicted to exhibit a simple and local connectivity profile (that can be inferred from the weight matrix even in the absence of information about neural ordering, Figure A.2). Despite the local connectivity, GCN cells are predicted to project with equal frequency to cells tuned to similar and orthogonal spatial phases (Figure 2.4E). All neurons with direct velocity input and synapses subject to asymmetric STDP are predicted to exhibit asymmetries in their outgoing weight profiles.

The model predicts that population patterning arises abruptly. The abruptness of patterning may be assessed by spatial tuning over short trajectories. The abruptness of patterning is likely to be masked or smoothed when assessed by spatial tuning over longer trajectories or when assessed based on a limited number of neurons and trials. Thus, grid cell responses on short linear tracks might more readily reveal pattern formation.

The mean direction tuning strength and the strength of speed tuning are predicted to decline (Figure 2.6F,G) because of the growth in recurrent weights over development. If pairwise neural correlations are computed in the absence of velocity inputs to the GCN, as when the animal is at rest or in slice preparations, the model predicts a considerable broadening of the correlation distribution over development, with the emergence of a uniform platform-like component. The *in vivo* system can, in the presence of velocity inputs, display

strong velocity-driven (anti)correlations in early development, that are difficult to tell apart from population pattern-induced (anti)correlations (Figure A.6).

The model also makes a number of parametric predictions, about how spatial tuning in the mature GCN depends on the biophysical time-constant of cells and synapses, the STDP window widths, the strength of the feedforward velocity input, the relative gain in excitatory and inhibitory synaptic strengths, and the statistics of spatial exploration. Our results show how to experimentally discriminate between possible mechanisms underlying the variation in grid period along the dorsoventral axis of the entorhinal cortex, based not only on spatial tuning period, but on other aspects of neural response, including direction and speed tuning, population pattern period, and the scaling of grid period to blob width in different modules.

2.4 Methods

Roman subscripts (e.g. i, j) refer to individual cells within population P. The population index P can take the values {I, E^R, E^L} in the 1D GCN and {I, E^R, E^L, E^U, E^D} in the 2D GCN. Integration in all simulations is by the Euler method with time-step dt.

In the 1D GCN, unless otherwise noted, the trajectories used during the plasticity phase and for probes of GCN development consist of random paths across a 1D environment (see Appendix A). In the 2D GCN, to reduce simulation time during development, the exploration trajectory consists of fixed-speed ($\vec{v}_{rat} = 1$ m/s) sweeps vertically and horizontally across the en-

vironment. Successive parallel sweeps are staggered (separated by a distance of 0.0125 meters) to provide approximately uniform coverage of the environment. The mature 2D GCN is evaluated with velocity inputs derived from rat trajectories recorded in the open field [4].

2.4.1 Neural and synaptic dynamics.

Given a summed current input $I_i^P(t)$ to the (P,i)th cell, the instantaneous firing rate of the cell is

$$r_i^P(t) = f(I_i^P(t)), \tag{2.3}$$

with the neural transfer function f given by

$$f(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0. \end{cases} \tag{2.4}$$

Based on this time-varying firing rate, neurons fire spikes according to an inhomogeneous (sub-Poisson) point process with a coefficient of variance of CV = 0.5 (see [16] and Appendix A for details on generating a sub-Poisson point process).

The activation $s_i^P(t)$ of synapses from the (P,i)th cell is given by

$$\frac{ds_i^P(t)}{dt} + \frac{s_i^P(t)}{\tau_s} = \sigma_i^P(t), \tag{2.5}$$

where

$$\sigma_i^P(t) = \begin{cases} r_i^P(t) & \text{(deterministic dynamics)} \\ \sum_k n_{i,k}\delta(t - t_{i,k}^P) & \text{(stochastic dynamics),} \end{cases} \tag{2.6}$$

where $n_{i,k}$ indicates the number of spike events of the ith cell at time $t^P_{i,k}$. For both the 1D and 2D GCN, the neural dynamics ars stochastic during testing in the activation phase. During the plasticity phase, dynamics are stochastic for the 1D GCN and deterministic for the 2D GCN.

The total synaptic current $I^P_i(t)$ into the (P,i)th cell is given by

$$I^P_i(t) = A^P_i \alpha^{P,vel}(v,t)(g^{P,rec}_i(t) + g^{P,loc}_i(t) + g^0) + A^P_i g^{0'}, \tag{2.7}$$

where g^0 and $g^{0'}$ are small, positive, constant bias terms common to all cells, $g^{P,rec}_i$ are the recurrent inputs, $g^{P,loc}_i$ are the location-specific inputs, $\alpha^{P,vel}$ are the velocity inputs, and A^P_i is the suppressive envelope function. The recurrent input is

$$g^{P,rec}_i(t) = \sum_{P'} \sum_{j=1}^{N} W^{PP'}_{ij} s^{P'}_j(t), \tag{2.8}$$

where $W^{PP'}_{ij}$ are the recurrent weights. The location-specific input is a Gaussian bump of height W^{loc} and width σ_{loc}:

$$g^{P,loc}_i(\vec{x}(t)) = W^{loc} \exp\left[\frac{-||\vec{x}(t) - \vec{x}^P_i||^2}{2\sigma^2_{loc}}\right], \tag{2.9}$$

where $\vec{x}(t)$ is the location variable and $\vec{x}^P_i$ is the preferred location of the input to the $(i,P)th$ cell. (In the 1D GCN, location and preferred location preference are scalars). The preferred locations are evenly distributed over the unit interval in $\mathbb{R}^1$ and $\mathbb{R}^2$ for the 1D and 2D GCNs, respectively.

All cells in the Pth population (with preferred direction given by the unit vector $\hat{e}^P$) receive a common velocity input:

$$\alpha^{P,vel}(\vec{v}(t)) = (1 + \beta^{vel}\vec{v}(t) \cdot \hat{e}^P), \tag{2.10}$$

where $\vec{v}(t)$ is velocity of the animal and β^{vel} sets the gain of the velocity input; $\hat{e}^P$ = (0,0), (0,1), (0,-1), (1,0), (-1,0) for the I, E^R, E^L, E^U, E^D populations, respectively. The multiplicative influence of velocity inputs on the rest of the inputs to each cell may be viewed as a shunting effect [53] from inputs that arrive simultaneously at different parts of a dendritic tree and combine nonlinearly.

The envelope function, which is only applied in the familiar home environment (and not during testing; although performance would improve during testing if the envelope were also applied then), is given by [16]:

$$A_i^P = \begin{cases} 1 & X_i^P < 1 - \Delta X \\ \exp\left[-a_0 \left(\frac{X_i^P - 1 + \Delta X}{\Delta X}\right)^2\right] & \text{otherwise} \end{cases} \tag{2.11}$$

where $X_i^P = ||\vec{x}_i^P - (0.5, 0.5)||$ (in 1D, $X_i^P = |x_i^P - 0.5|$), ΔX determines the range over which tapering occurs, and a_0 controls the steepness of the tapering.

2.4.2 Plasticity rule and development.

The recurrent weights $W_{ij}^{PP'}$ are drawn initially from a uniform distribution on the interval $[0, \pm w_0]$ ($-$ when the presynaptic cell is inhibitory, and $+$ when it is excitatory). We assume no direct E-to-E connections; thus, these weights are 0. During plasticity in the home environment, g^0, $g^{0'}$, and $g_i^{P,rec}$ are set to zero; neural activity is based only on the feedforward inputs (which are tapered at the edges according to the envelope function A_I^P). Weights are

incremented during exploration according to:

$$\frac{\Delta W_{ij}^{PP'}(t)}{\eta\gamma_{PP'}} = \int_0^\infty \sigma_i^P(t)\sigma_j^{P'}(t-t')k_+^{P'}(t')dt' + \int_0^\infty \sigma_j^{P'}(t)\sigma_i^P(t-t')k_-^{P'}(t')dt', \quad (2.12)$$

where η is the learning rate, $\gamma_{PP'}$ is a term of order 1 that controls the relative speed of learning for the different types of recurrent weights, and σ_i^P is given by (4). There are two learning kernels, k^E and k^I, depending on whether the presynaptic cell is excitatory or inhibitory, respectively (Figure 2.1). The causal (+) and acausal (−) sides of these two learning kernels are given by:

$$k_+^E(t) = Ae^{-t/2\alpha_{STDP}\tau_{STDP}} \quad (2.13)$$

$$k_-^E(t) = -e^{-t/1.5\alpha_{STDP}\tau_{STDP}}; \quad (2.14)$$

$$k_+^I(t) = Be^{-t/2\alpha_{STDP}\tau_{STDP}} \quad (2.15)$$

$$k_-^I(t) = -e^{-t/\alpha_{STDP}\tau_{STDP}}. \quad (2.16)$$

The coefficients A and B control the relative magnitudes of the two sides of the learning kernels, and the time constant τ_{STDP}, with a scale factor α_{STDP} of order 1, controls the widths.

In the testing condition (activation phase), η is set to zero and the envelope is removed; all figures probing GCN dynamics, Figures 2.3-2.8, are generated in the activation phase.

1D simulation parameters

$N_I = 80$; $N_E = 400$ (200 per E population); CV = 0.5; $dt = 0.5$ ms; $\tau_s = 30$ ms; $\sigma_{loc} = 1$ cm; $W_{loc}^I = 50$; $W_{loc}^E = 10$; $g^0 = 50$ (=1 for "lesioned" feedforward

input); $g^{0'} = 15$ (=0 for "lesioned" feedforward input); $\beta^{vel} = 0.9$; $w_0 = 0.001$; $\eta = 0.015$ s^{-1}; $\gamma_{II} = 7$; $\gamma_{EI} = 2$; $\gamma_{IE} = 1$; $\tau_{STDP} = 12$ ms; $\alpha_{STDP} = 1$; A = 1.2; B = 0.5; $\Delta r = 0.72$; $a_0 = 60$.

2D simulation parameters

$N_I = 1600$; $N_E = 6400$ ($40 \times 40 = 1600$ per E population); $\beta^{vel} = 2$; $w_0 = 0$; $\eta = 0.012$ s^{-1}; $\gamma_{II} = 5$; $\gamma_{EI} = 0.25$. $\gamma_{IE} = 16.7$; $\Delta r = 36$; $a_0 = 10$. All other parameters identical to 1D parameters.

Chapter 3

Cortical microcircuit determination through perturbation and sparse sampling in grid cells

3.1 Introduction

The unusual responses of grid cells present a challenge and simultaneously, an opportunity. The challenge is to understand how such a complex cognitive response is generated; the opportunity is the availability of versatile experimental tools and a rich set of relatively detailed models [16–18, 38, 46, 51, 52, 78, 83, 114] that are well-constrained by the very complexity of the grid cell response, to help meet the challenge.

The recent application of quantitative analyses to electrophysiological data reveals that the population dynamics of grid cells (within individual modules, or networks – we will refer to them as GCNs, for grid cell networks) is localized around a continuous low-dimensional (2D) manifold [39, 125], a finding that lends support to early models predicated on the idea of low-dimensional pattern formation through strong cell-cell (i.e., *recurrent*) coupling [15, 16, 38, 46, 77], as well as other models in which grid cells are the recipients of spatially tuned inputs that conspire to drive grid cell firing [18, 68, 78, 114].

These models are architecturally and mechanistically distinct in impor-

tant ways, both large and subtle: they differ in whether grid cells perform path integration, in whether pattern formation originates wholly or partly within grid cells, and in the structure of their connectivity. Some of the structural differences within recurrent models which seem subtle have qualitative ramifications for how the circuit could have developed. Despite their differences, the models are difficult to distinguish on the basis of existing data, because all of them produce grid-patterned outputs and exhibit approximate 2D continuous attractor dynamics. Worse, as we discuss at the end, neither having the complete neural activity records nor the complete single synapse-resolution weight matrices will be sufficient to distinguish between proposed mechanisms.

We show how it is nevertheless possible to gain surprisingly detailed information about the grid cell circuit from a feasible experimental strategy that depends on circuit perturbation, applied simultaneously to all cells in the GCN, and sparse neural recording. The proposed strategy can allow the experimenter to discriminate between various distinct candidate mechanisms that are currently undifferentiated by experiment.

3.2 Results

3.2.1 Experimentally undifferentiated grid cell models

Let us begin by considering recurrent pattern forming models, the same types of models we have been considering in Chapter 1 and 2, in which grid cells are assumed to integrate velocity inputs (path integrate) and output location-coded grid-like responses in movement space (the space in which the animal

moves) [16–18, 38, 46, 51, 52, 78, 83, 114]. These models can be distinguished primarily by the nature of their underlying connectivity along two dimensions: the *locality* of the connectivity, which specifies in topological space (an organization of cells based on their connectedness – see Chapter 1) the extent of the Mexican-hat connectivity with respect to the size of the network and can be changed continuously, and the *topology* of the connectivity, which is either toroidal (with periodic boundary conditions) or planar (with aperiodic boundary conditions). Dynamically, the locality knob controls the size and number of bumps in the population activity pattern, and therefore its wavelength. The topology knob imposes extra constraints on the pattern phenotype: For periodic boundary conditions, the number of bumps in the pattern must be an integer; for aperiodic boundary conditions, the situation is more complex and depends on whether the neural activity near the boundary edges are tapered or not (see Chapter 4 and reference [16]). If they are tapered, then the boundaries have no affect on the pattern phenotype, and so the wavelength is determined primarily by the locality – we will assume tapered boundaries for the remainder of this chapter.

These two knobs, locality and topology, naturally subdivide grid cell pattern-forming networks into three main types: (1) local connectivity (therefore supporting many activity bumps) with aperiodic boundaries, which we will call *aperiodic* [16, 116], Figure 3.1A, (2) global connectivity (supporting a singlebump) with periodic boundaries, called *singlebump periodic* [12, 15, 38, 46, 89, 116], Figure 3.1B, (3) local connectivity (multibump) with periodic

boundaries, called *multibump periodic* [16], Figure 3.1C. (Note that we have left out the fourth variation, i.e., global connectivity with aperiodic boundaries. In this case, in the limit when the population activity pattern expresses a single bump, the bump inevitably gets stuck at the network edge, preventing the network from performing path integration over large spaces as well as the ability to generate the multiple firing fields characteristic of grid cells.)

These aforementioned models, termed *recurrent* network models because they rely on interactions between grid cells, are distinct from the other major class of grid cell models in which grid cells are the result of the summation of inputs that are already spatially tuned, termed *feedforward* models [18, 68, 78, 114], Figure 3.1D-E. Whereas in recurrent models, in which path integration occurs within the grid cell layer, in feedforward models, path integration occurs upstream of grid cells. For example, in [78], individual networks upstream from grid cells, each encoding the animal's displacement along particular directions in movement space as a single bump of activity persisting on a ring of connected neurons (essentially a 1D singlebump periodic network), conspire to drive the hexagonal spatial responses of grid cells (Figure 3.1D).

We would like to experimentally discriminate amongst the aforementioned models, using an approach that is experimentally feasible. In theory, one should be able to deduce the underlying structure by either taking an activity snapshot of the entire population activity, or by measuring the full cell-to-cell connectivity (i.e., its *connectome*). However, the lack of topography in the grid cell network precludes learning much from snapshots of the

activity, because the population activity, in brain space, will appear salt-and-pepper-like, even if the population activity is patterned in topological space). On the other hand, trying to measure every connection in the network is entirely infeasible given the current techniques available. In the next section, we describe a novel method for discriminating these networks using available experimental techniques.

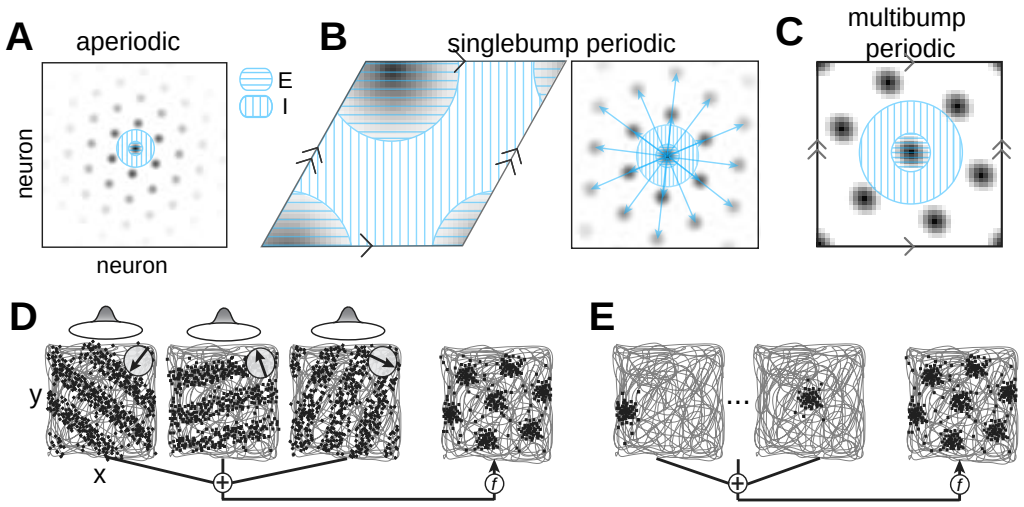

Figure 3.1: **Mechanistically distinct models not distinguished by existing data.** (A-C) Recurrent 2D pattern-forming models: Population activity (gray; darker indicates more activity) and the corresponding synaptic weights from a single representative cell (blue regions centered on the cell of origin; horizontal lines (E) indicate excitatory projections; vertical lines (I) indicate inhibitory projections), both in brain space. (A) Aperiodic network. (B) Singlebump periodic network. The boundaries of the network are sheared into a rhombus so that the spatial tuning curves of individual cells in the population are hexagonal grid-like. (C) Multibump periodic network. (D-E) Feedforward networks. (D) Grid cells are generated by selectively summing inputs from cells belonging to "ring attractor" networks, i.e., periodically-connected 1D network of neurons that have singlebump-like population activities, drawn as the ellipses in the figure [18, 78, 114]. Each network integrates animal velocity along a particular direction in movement space, so that cells exhibit stripe-like responses in movement space oriented orthogonal to this axis (see three boxes at left). (E) Grid cells are generated by selectively summing inputs from place cells (which fire at single locations in the environment, examples of which are shown in left two boxes) to make grid-like responses in movement space (right box) [68].

3.2.2 A perturbation-based probe of circuit architecture

The conceptual idea that we propose for differentiating between models of grid cells depends on multiple grid cell recordings (i.e., measuring their spatial tuning curves as the animal moves about space) before and after a sustained perturbation of the network. The idea derives from previous observations (Chapter 2) in which perturbing the dynamics of an aperiodic network, e.g. by increasing the strength of the inhibition in the network connectivity or increasing the neural time-constant, leads to increases in the period of population activity pattern (Figure 2.7B, third row, columns 3 and 4, Figure A.7B, third row, columns 2 and 3, and Figure B.2). This effect, not predicted by linear stability analysis, exists in simulations [16, 116]; an analytical understanding of such effects will be the goal of Chapter 4. Experimentally, modifying the strength of inhibition can be induced via the application of drugs, e.g., benzodiazepines that modulate the efficacy with which inhibition is communicated between two cells [94], while modifying the time constant, as we argue later, can be induced by cooling the tissue, and thus slowing down the temperature-dependent chemical reactions that subserve the neural dynamics [64, 80, 108].

What is the gain of such a perturbation-based approach? As we will see, perturbing the underlying dynamics of the system (in brain space) leads to predictable and characteristic shifts in the phase relationships between the spatial tuning curves (in movement space) of pairs of cells that reflect the "quantization" of the population activity pattern.

For the following, we assume a 1D network. However, the method we describe is completely generalizable to 2D. Let roman subscripts (e.g., i and j) refer to individual cells. If cells are arranged in topological space (i.e., based on their connectedness), then i refers to the ith cell's location (in neuron body-length units) within the population activity pattern. Recall that the population activity pattern is defined as a snapshot of the activity pattern of the network, in brain/topological space. In topological space, let the population activity pattern be periodic with period λ_{pop} (again in neuron body-length units), Figure 3.2A, blue curve.

We will define the network state before/after a perturbation as the *pre/post*-perturbation state. The size of the perturbation, or the perturbation *stretch factor*, is defined as $\alpha = |\lambda_{pop,post}/\lambda_{pop,pre} - 1|$, where $\lambda_{pop,pre}$ and $\lambda_{pop,post}$ are the wavelengths of the pre- and post-perturbation population activity patterns, respectively (Figure 3.2A). For example, as depicted in Figure 3.2, a perturbation that induces a stretching of the pattern wavelength from $\lambda_{pop,pre} = 250$ neurons to $\lambda_{pop,post} = 290$ neurons is characterized by the stretch factor $\alpha = 0.16$.

The ith cell's firing phase within the periodic population activity pattern, defined as the cell's *population phase*, is $\phi^i_{pop} = ((i-1) \bmod \lambda_{pop})/\lambda_{pop}$ (with the arbitrary choice, made without loss of generality, that neuron 1 has phase 0) (Figure 3.2B, blue curve).

After a perturbation, cells one wavelength apart in the population pattern (see the locations of circle, square, and triangle symbols in Figure 3.2A)

and thus firing at the same population phase (Figure 3.2B, blue), will experience a shift in their population phases (Figure 3.2A-B, compare red curve with blue curve). Defining this shift to be one quantum, ϵ (Figure 3.2B), then the shift in phase between cells originally separated by exactly K wavelengths will be K quanta, or $K\epsilon$, where $\epsilon = \frac{\alpha}{\alpha+1}$ (Figure 3.2B).

For each cell in the population, plotting the pre-perturbation phase against the post-perturbation phase (red vs. blue curves in Figure 3.2B) shows that the data is quantized and lies on a series of parallel manifolds, Figure 3.2C. This quantization is captured via the following transformation: First, the data is down shifted along the y-axis, so that these curves become extenuations of the other curves:

$$\Delta\phi^i_{pop} = \begin{cases} \phi^i_{pop,pre} - (1+\alpha)(\phi^i_{pop,post} - 1), & \text{if } \phi^i_{pop,pre} < (1+\alpha)\phi^i_{pop,post} \\ \phi^i_{pop,pre} - (1+\alpha)\phi^i_{pop,post}, & \text{otherwise,} \end{cases} \tag{3.1}$$

(We have assumed that the true stretch factor, α, is known – later, we will show how α can be inferred from the data.) The phase shifts are then mapped to the $[0, 1]$ interval via the modulo operation

$$\Delta\phi^i_{pop} = \Delta\phi^i_{pop} \bmod 1, \tag{3.2}$$

followed by reflecting about the midpoint of the interval

$$\Delta\phi^i_{pop} = \min\{\Delta\phi^i_{pop}, 1 - \Delta\phi^i_{pop}\}. \tag{3.3}$$

The distribution of these phase shift values, Figure 3.2D, has three special properties: (1) The distribution is quantized, due to the fact that population

activity pattern itself is quantized. (2) The number of peaks in the distribution is exactly equal to the number of bumps in the population activity pattern (this holds only for sufficiently small perturbations, such that $\alpha M < 0.5$, where M is the number of bumps in the pre-perturbation population activity pattern – see Figure B.3 for explanation). (3) The peak separation in the distribution is exactly equal to the stretch factor, α.

However, the construction of this distribution relies on experimentally difficult-to-access quantities, namely the population activity pattern phase for each cell. The utility of our proposed strategy arises because the distribution of shifts in the *population phase* across cells in topological space is mirrored in the distribution of shifts in the *relative phase of spatial tuning* (or *relative phase*) across cells in movement space (Figure 3.2E-F). Let d^{ij} represent the offset, in meters, of the peak closest to the origin in the cross-correlation of the two spatial tuning curves, and let λ be the spatial tuning period (in meters) of the two cells. The relative phase is defined as $\delta^{ij} = (d^{ij} \bmod \lambda)/\lambda$. The relative phase shift is computed as the population phase shift was, using equations Equations 3.1-3.3.

Cell pairs with zero relative phase in their spatial tuning pre-perturbation (because they have the same population phase) will exhibit post-perturbation shifts in relative phase that, like the shifts in the population phases, will be quantized, and for small changes in population period will be proportional to the number of bumps separating them, Figure 3.2F. We define this distribution to be the *distribution of relative phase shifts*, or DRPS.

In 2D, relative phase is a vector measured along the two principal axes of the hexagonal spatial tuning grid (see section "2D relative phase" in Appendix B). As in 1D, DRPS's for *each* component of δ^{ij} can be computed. The total number of bumps in the population activity pattern can be read out as the product of the number of peaks in the two single-component DRPS's (Figure B.4). For the rest of the chapter, statements that are made regarding the DRPS computed in 1D also hold for the component DRPS's computed in 2D.

The transformations described in Equations 3.1-3.3 require knowledge of the stretch factor, α, a quantity that is not directly observable. However, it can be inferred from the data, because the desired α value is the one that makes the distribution the most peak-y. This is equivalent to projecting the data onto its orthogonal axis, Figure 3.2C. Peak-y-ness is quantified as the Pearson's correlation coefficient between the DRPS and a comb-like function defined over the same interval. The comb function is a series of delta-functions laid out with a spacing equal to α. The desired α stretch factor is the one that maximizes this correlation.

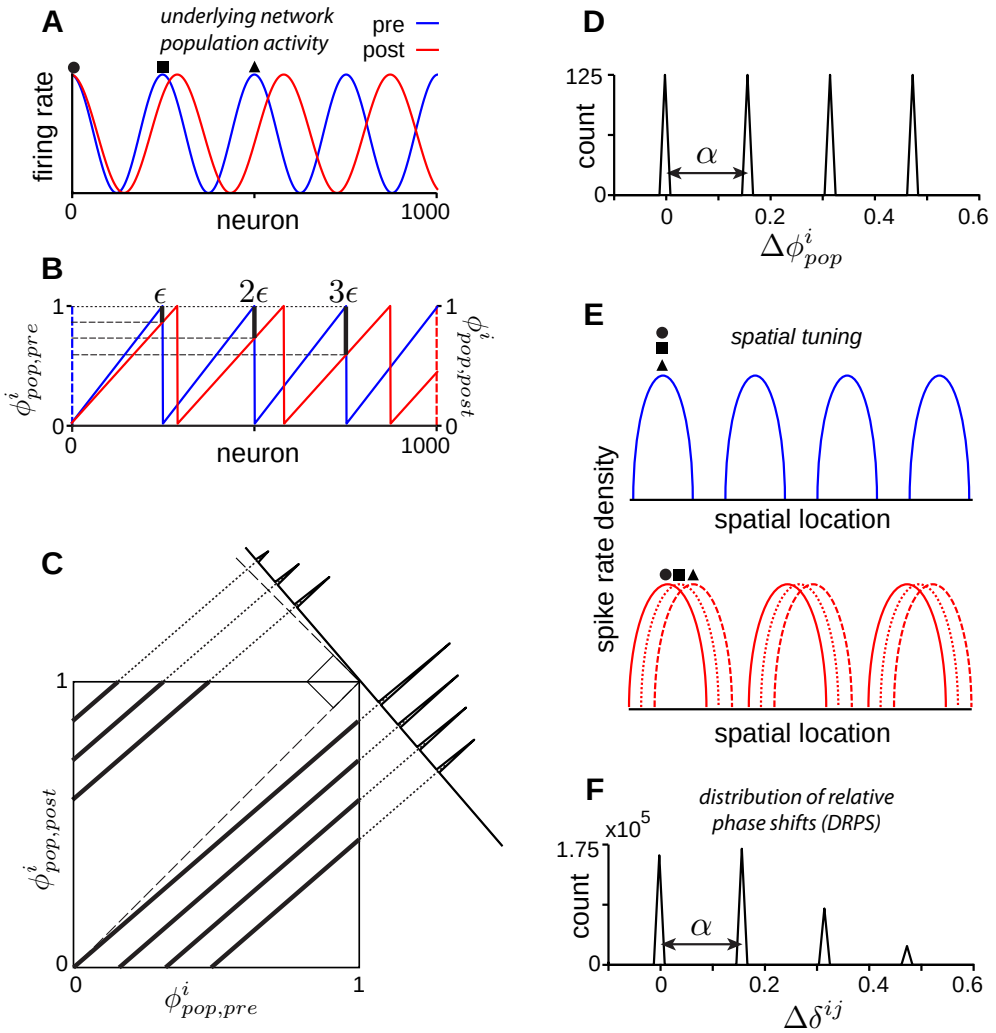

Figure 3.2: **Perturbation and phase shift analysis can reveal detailed features of population patterning.** (A) 1D population activity, pre- (blue) and post-perturbation, for a %16 increase the wavelength of the pattern ($\alpha = 0.16; \lambda_{pop,pre} = 250$ neurons), with pattern expansion is centered at the left network edge. Circle, square, and triangle: locations of cells that are separated by integer numbers of wavelengths. (B) Population phase, pre- (blue) and post- (red) perturbation. (C) Post-perturbation phase vs. pre-perturbation phase, with projection of data onto orthogonal axis shown at

upper right. (D) Distribution of shifts in population phase (n=1000). (E-F) Shift distributions for population phase (experimentally inaccessible) carry over to shift distributions for spatial tuning phase (experimentally observable). (E) The circle, square, and triangle cells, which original have identical spatial tuning (blue curves), now exhibit shifted spatial tuning curves (red curves). The shift in spatial phase for a pair is proportional to the number of activity bumps between them in the original population activity pattern. (F) Distribution of relative phase shifts (DRPS) (n = (1000 choose 2) relative phase samples because relative phase is computed pairwise).

3.2.3 Discriminating amongst recurrent architectures

Simulations of the grid cell models (in this case, their 1D equivalents) reveal that the effects of such a perturbation will differ across recurrent network architectures, with consequently different predictions for the DRPS.

In an aperiodic network, incremental perturbation results in incremental expansion of the population activity pattern (Figure 3.3A, red, and Figure B.2). Thus, the peak spacing of the DRPS will expand gradually and linearly with perturbation strength (Figure 3.3B-C, red and Figure B.2).

In a multibump periodic network (network with local connectivity and periodic boundary conditions, Figure 3.1B), the number of bumps in the population activity pattern is constrained to be an integer. Thus, incrementally increasing the perturbation strength should result first in no change to the population activity period, and then a sudden change when the network can accommodate an additional bump (or an additional row of bumps in 2D, assuming the pattern does not rotate as a result of the perturbation; see Discussion) (Figure 3.3A, purple). The fine structure of the DRPS will still be peak-y. However, counting peaks to estimate the number of bumps in the underlying population activity pattern will result in serious underestimation, as the effective stretch factor, α, will be much larger than $1/(2M)$ (Figure B.3 and e.g. Figure 3.3B, compare peaks in the solid and dashed lines for small and large perturbations, respectively)

In the singlebump periodic network (Figure 3.1C), the same perturba-

tions that alter the population activity pattern period in the other recurrent networks (Figure 3.1A-B) are ineffective in inducing a corresponding change (Figure 3.3A, blue). This is because the periodic connectivity completely fixes the period of the pattern. Thus, the perturbation will not affect the relative phase relationships between cells, and the DRPS is predicted to remain narrow, unimodal, and peaked at zero (Figure 3.3B-C, blue).

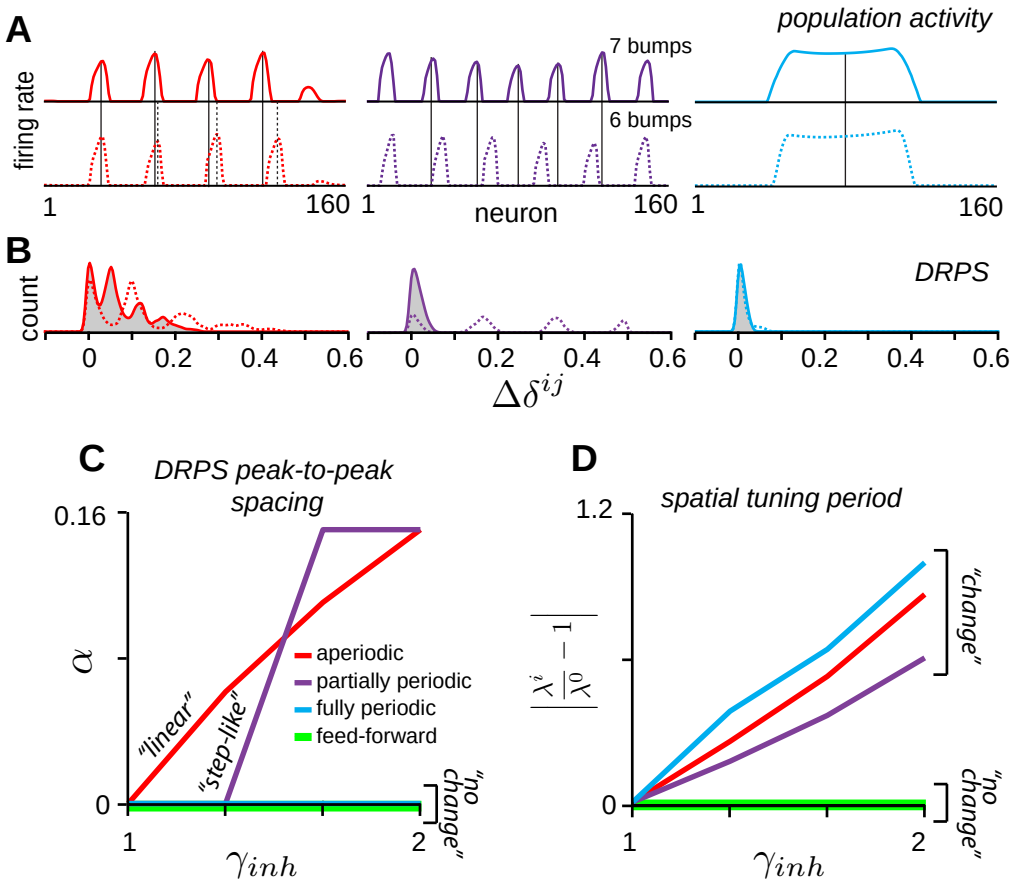

Figure 3.3: **Effects of perturbation on recurrent and feedforward neural networks and predictions for experiment.** (A-B) The effect of perturbing inhibitory weights in 1D neural network simulations of aperiodic (column 1), multibump periodic (column 2), and singlebump periodic (column 3) networks. (A) Population activity pattern pre- and post-perturbation. Vertical lines: bump centers in the unperturbed (solid) and perturbed (dotted) patterns. (B) DRPS of perturbed networks ($\gamma_{inh} = 1.33$ and 1.66, solid and dotted lines respectively) relative to the unperturbed ($\gamma_{inh} = 1$) network. (C) DRPS peak spacing (i.e., measured α) varies with perturbation strength (prediction – not simulated – for feedforward network in green). (D) Spatial tuning period (λ^i, where $i = \{0, 1, 2, 3\}$ as a function of perturbation strength (see Appendix B for how spatial tuning period is computed).

3.2.4 Discriminating feedforward from recurrent architectures

We've described how the three types of recurrent networks can be discriminated from one another based on the response of the DRPS to perturbations of increasing strength. If grid cells inherit their spatial tuning through feedforward summation [18, 78, 114] of spatial inputs as in the case of the feedforward models, Figure 3.1D-E, then perturbing the within only the grid cell layer will not affect the dynamics upstream, and thus not affect the grid cell responses. As a result, the DRPS should be narrow and centered at zero.

However, this response of the DRPS is identical to the multibump periodic network (Figure 3.3C, green line). How, then, can these two types of networks be discriminated?

In all recurrent model networks (Figure 3.1A-C), such perturbations induce a change in the efficacy with which the velocity inputs shift the population activity pattern over time, and thus change the periods of the spatial tuning curves (Figure 3.3D and Figure B.6). By contrast, in feedforward models the computations that subserve path integration occur upstream of the grid cells; thus the spatial tuning period should remain unchanged with perturbation (Figure 3.3D, green line). Thus, feedforward networks can be discriminated from recurrent networks, and particularly from singlebump periodic networks, based on whether or not the perturbation induces changes to the period of the spatial tuning curves.

3.2.5 Experimental feasibility of proposed method

We consider two key data limitations. First, it is not yet experimentally feasible to record from all cells in a grid cell network (GCN). Even a 100-cell sample would constitute a 1-10 % subsampling of the estimated size of a GCN. With present estimates that <20 % of cells in a local patch in MEC are grid cells [106], the yield would be a meager 20 grid cells. Is this sufficient to observe the predicted quantal structure in a phase shift distribution, if it were present? Fortunately, the proposed method is tolerant to severe sub-sampling of the population: a tiny random fraction of the population (10/1600 cells) can capture the essential structure of the full DRPS (or for each of the component DRPS's in 2D, as shown in Figure 3.4A).

Second, spatial tuning curves are estimated from neural responses during a random, finite exploration trajectory by the animal in which cells respond variably. Hence, spatial tuning parameters, including phase, are only known with a degree of uncertainty. In tests that depend only on the peak separation of the DRPS (e.g. Figure 3.3), this phase uncertainty is not a serious limitation.

However, more detailed questions about the number of bumps in the population activity pattern in an aperiodic network depend on estimating the number of peaks in the DRPS (or the component DRPS's in 2D), and here phase estimation uncertainty can be problematic: phase uncertainty will merge together peaks in the DRPS, Figure B.7. In the range of small perturbation strengths, the DRPS inter-peak spacing (in the aperiodic network) increases

with the stretch factor. Thus, the larger the perturbation, the more distinguishable the peaks are at a fixed phase error, Figure 3.4B and Figure B.7. Yet increasing the stretch factor is not without a tradeoff: The two-for-one relationship between number of peaks in the DRPS and the number of bumps in the population activity pattern per linear dimension holds when the total induced shift in phase is small for all bumps (as before, when $\alpha < \frac{1}{2M}$, with M now equal to the larger of the number of bumps along the two principal axes of the population activity pattern), Figure 3.4B. At larger stretch factors, the number of peaks in the DRPS is smaller than the number of bumps along the corresponding dimension of the pattern, and the discrepancy can be substantial.

Fortunately, the DRPS is computed from the relative phases between cells, which remain stable in a fixed network [125] (here "fixed" refers to the network while a given perturbation strength is stably maintained). This stability makes it possible to gain progressively better estimates of relative phase over time even if there is substantial drift in the spatial responses of cells, by computing the relative phase over short snapshots of the trajectory then averaging together the relative phase estimates from different snapshots across a progressively longer trajectory (similar to the methods used in [125] and [9]).

To distinguish $M = 5$ bumps per linear dimension based on peak-counting in the DRPS would require a stretch factor of no greater than $\alpha = 1/(2M) = 0.1$, and phase noise must be reduced to at least 0.02 (Figure B.7). Distinguishing 7 bumps would require $\alpha \leq 0.07$ and a phase noise of smaller

than about 0.01. Based on grid cell and trajectory data (accessed through http://www.ntnu.edu/kavli/research/grid-cell-data), this would require an approximately 10- and 50-minute recording, respectively (Figure 3.4C).

The proposed method therefore has high tolerance to subsampling and a more limited tolerance to phase uncertainty. It will require longer-than-usual but still realistic amounts of neural recordings to obtain adequately small error in relative phase estimation to test predictions that differentiate between models.

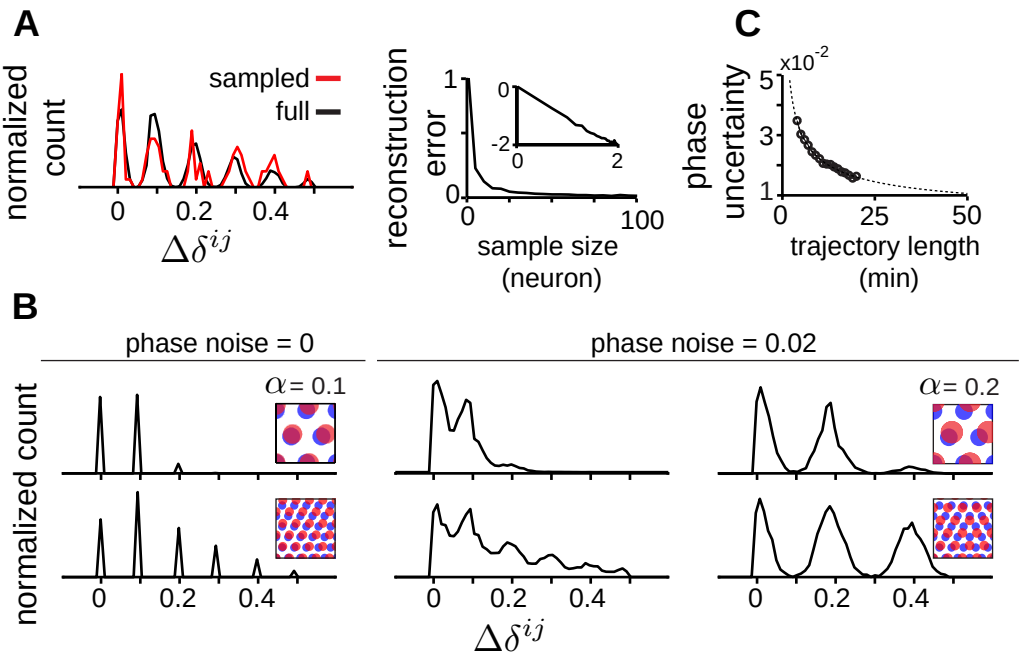

Figure 3.4: **Measurement limitations and the resolvability of predictions under such constraints.** (A) Left: DRPS, having sub-sampling (red) vs. fully-sampling (black) the population (here, DRPS is computed from single component of 2D relative phase – see Figure B.4). "Phase noise", i.e. i.i.d samples from $\mathcal{N}(0, \sigma^2_{phase})$, were added to each component of the relative phase vectors. Right: The L2-norm difference between the full and sampled DRPS as a function of number of sampled cells. Inset: log-log scale. (B) First and second columns: DRPS for different numbers of bumps in the population activity pattern and for different amounts of phase noise. Third column: Same as column 2, except for a larger stretch factor, $\alpha = 0.2$. (C) The uncertainty (standard deviation) in estimating relative phase, for different amounts of data (data from [48]), from bootstrap samples of the full dataset (see Appendix B for details), which follows $T^{-\frac{1}{2}}$ (dotted). *Parameters*: $\lambda_{pop,pre} = 40/3$ neurons (A), $= 20$ neurons (B, top row), $= 8$ neurons (B, bottom row); $\alpha = 0.1$; $\hat{e}_1 = [1, 0]$; $\hat{e}_2 = \hat{e}_1 + 60°$; network size: 40×40 neurons.

3.2.6 A decision tree for experimental design

We lay out a decision tree with an experimental workflow for discriminating between disparate networks, all of which exhibit 2D continuous attractor dynamics (Figure 3.5).

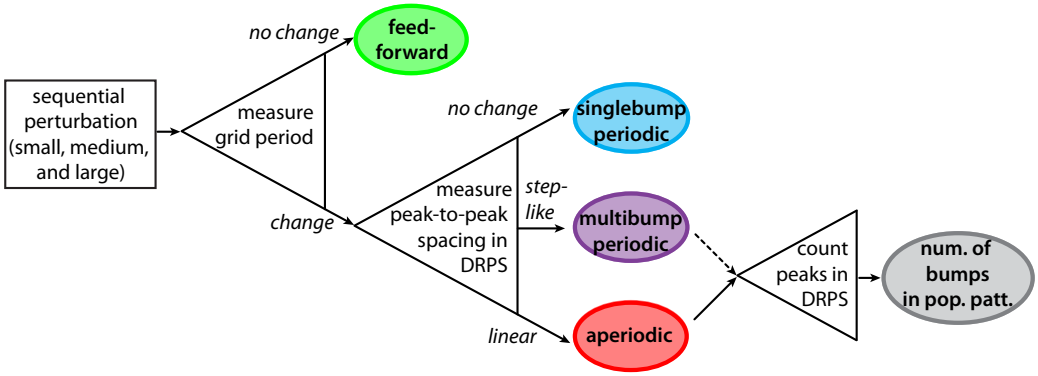

Figure 3.5: **Decision tree for experimentally discriminating circuit mechanisms.** For each of three circuit perturbations of increasing strength, both spatial tuning period and relative phase shifts are measured. Recurrent networks are discriminated from feedforward and feedforward-recurrent networks by the effects of the perturbation on spatial tuning period (first open triangle). Different recurrent networks can be discriminated based on how the peak spacing of the DRPS (α) varies with perturbation strength (second open triangle). The number of bumps in the multi-bump population activity patterns can be inferred by counting the peaks in the DRPS (third open triangle), though, for the multibump periodic, only a lower bound on the number of bumps can be established (dotted line).

The demands from experiment are to be able to stably induce a perturbation in one GCN, and to do so at 2-3 strengths. In all the cases, the term perturbation refers to a small change that leaves the network dynamics qualitatively unchanged while affecting its quantitative properties. The data to be collected are simultaneous recordings from several grid cells as the animal explores a familiar enclosure with no proximal spatial cues over about 20 minutes or more.

First, before applying perturbations, characterize the spatial tuning (periods) of the neurons, as well as cell-cell relationships (the relative phase). Next, apply a series of 2-3 perturbations of increasing strength. At each perturbation strength, characterize the spatial tuning of cells and cell-cell relationships. A change in the amplitude of the cells' response across the different perturbations signals that the perturbation is having an effect.

If further there is no change in the spatial tuning period, it follows that the perturbations produced no change in the population activity pattern and velocity responsiveness, thus the network must be feedforward, Figure 3.5 (green). Verify that cell-cell relationships remain unchanged across perturbations, as predicted for feedforward networks.

If there is a change in the spatial tuning period, characterize the cell-cell relationships in each perturbation condition. Plot the DRPS from each perturbed condition relative to the pre-perturbation condition, and obtain its peak spacing. If the peak spacing increases steadily and linearly with perturbation strength, that implies an aperiodic recurrent architecture, Figure 3.5

(red). If the DRPS peak spacing exhibits a step change, it is consistent with a multibump periodic recurrent network, Figure 3.5 (purple). A DRPS that remains narrowly peaked around zero, with no change in the peak spacing with perturbation strength, is consistent with a singlebump periodic network, Figure 3.5 (blue).

Finally, if the network is either aperiodic or multibump periodic, the underlying population activity pattern has multiple bumps. The number of peaks in the DRPS for each dimension of relative phase bounds from below the quantity $2M$, where M is the number of bumps in the population activity pattern along that dimension. When the stretch factor α times the number of bumps is smaller than $1/2$, and if the DRPS is quantal, the number of DRPS peaks equals the number of population activity bumps along the corresponding dimension.

3.2.7 Relating network parameters to experimental parameters

Changes in the strength of recurrent inhibition in our model can be mapped into changes in the gain of inhibitory synaptic conductances in the biological system, which can be induced, experimentally, by locally infusing allosteric modulators (e.g. benzodiazipines [94]; personal communication with C. Barry).

Changes in the time-constant of our model neurons can be mapped to changes in the excitatory post-synaptic potential (EPSP) time-constant in the biological system. Experimentally, the EPSP time-constant is sensitive to

temperature through the Arrhenius effect and can be lengthened by cooling [64, 80, 108].

While it is straightforward how to translate a perturbation of the inhibitory synaptic conductances of the biological system to changes in the inhibitory weights or our simplified model, it is much less so for the temperature perturbation. Therefore, to assess what to expect experimentally from a temperature perturbation and how to correctly include temperature effects in simpler neural models, we performed network simulations with cortical Hodgkin-Huxley neurons [90] while implementing documented temperature-dependent changes in all ionic and synaptic conductances (Methods, Appendix B, and Figure B.5). The effect of cooling on conductance amplitudes is to shrink the population period in an aperiodic network, but its effect on conductance time-constants is to expand the period. The net effect of cooling is an expansion because temperature changes have larger effects on conductance time-constants (larger Q10 factors) than amplitudes (smaller Q10 factors) [57, 108]. We therefore conclude that changes in temperature are reasonable to associate with changes in the time-constant of simple neuron models.

3.3 Discussion

3.3.1 Assumptions

The predictions made here assume that the network activity pattern in 2D is stable against rotations. Rotations of the population activity pattern would induce large changes in the DRPS, obscuring the predicted effects of

pattern period expansion in any recurrent network. The singlebump periodic network is not subject to rotations, but multibump periodic and aperiodic networks may be. In experimental data, the relative phase relationships between grid cells are indeed very stable across time and environments [125], suggesting that the population activity undergoes no rotation. It is unclear what features of the circuit stabilize the population activity pattern against rotation; it is possible that slight directional anisotropies in the outgoing connectivity of neurons pin its orientation.

The simplifying observation, that spatial responses may be used to estimate the DRPS, depends on other inputs not being able to overrule the new post-perturbation cell-cell relationships. For instance, external sensory inputs or hippocampal place cells that become associated with particular configurations of grid cells may keep resetting the grid networks to express old relative phase relationships. To avoid this possibility, it may be important to assess post-perturbation cell-cell relationships only in novel environments, for which there are no previously learned associations between external cues, place cell responses, and grid cell activity.

Finally, it is important to note that if in feedforward models one were to include feedback from the grid cell layer back to the spatially tuned inputs (as in [18]), the network would effectively become a type of recurrent circuit, and perturbing the grid cell layer may result in changes in grid period and cell-cell relationships.

3.3.2 Prior probabilities of different grid cell models being correct

From theoretical arguments, we believe the candidate grid cell mechanisms are not equally probable. In particular, the multibump periodic model is difficult to justify from the viewpoint of grid cell development. In [116] and Chapter 2, we see that activity-dependent rules acting on spatially informative feedforward inputs can lead to the formation of a network capable of path integration and with grid cell-like tuning. The network, post-development, has aperiodic structure. Under certain conditions, if network weights continue to undergo plasticity after the network has matured enough to expresses population activity patterning, the network can become a singlebump periodic network as neurons with the same spatial phase become wired together (Figure B.8). In fact, the addition of relatively weak coupling between neurons in nearest-neighbor activity bumps, or hotspots, in the activity pattern is sufficient to convert an aperiodic network into what is, functionally if not topologically, a singlebump periodic network (Figure B.8).

Thus, it is possible to imagine mechanisms for the development of the singlebump periodic and fully aperiodic networks. By contrast, a multibump periodic network involves local connectivity which does not depend on a neuron's spatial phase, but at the same time requires some mechanism for neurons at one end of the network to link with those at the opposite end in way that depends on spatial phase, Figure B.1. It is more difficult to imagine a plausible mechanism that can satisfy both constraints. By the same argument, in feedforward models, one would expect the 1D patterned inputs to grid cells to

involve singlebump periodic or aperiodic 1D networks.

3.3.3 Circuit inference through perturbation and sparse activity records: outlook and alternatives

It is interesting to compare the potential of our suggested approach with that of single synapse-level circuit reconstruction (a connectomics approach). A high-quality full-circuit connectome (i.e., the full network connectivity) can specify the topology of the network structure. In other words, it should be possible to reveal whether the circuit is intrinsically "local" (as in the aperiodic network of Figure 1A) [116], multibump periodic (with local center-surround-like connectivity and periodic boundary conditions as in Figure 1B), or single-bump periodic (with center-surround-like connectivity of a width that spans the entire network together with periodic boundary conditions). It may even be possible to infer the locality of structure in the aperiodic network from an unsigned connectome.

Network topology is, however, one ingredient in circuit mechanism: Determining whether the signed connections lead to population activity patterning still requires a large amount of inference (for instance, converting the connections into weights and inserting the matrix into an appropriate dynamical model). Even with further inference steps, whether the network actually performs certain functions like velocity-to-position integration or only inherits them is not answerable based on connectomics data. For instance, a network with lateral interactions between cells may generate position-dependent

responses *de novo* through integration (Figure 1A-C), or may act only to further pattern inputs that are already spatially tuned (Figure 1D-E) [18, 68, 78]. Despite these functional differences, both types of networks have similar connectivity and topologies.

On the other hand, single neuron-resolution records of activity within a GCN can be fruitfully used to understand the dimensionality and relationships of neural responses, but without perturbation, inferring actual connectivity and thus mechanisms from activity is problematic [59, 93]. Hence, activity records do not distinguish between different recurrent models. In short, while connectomics and large-scale recording can provide troves of useful information, they are not sufficient for discriminating between models; as we have shown here, they may also not be immediately necessary.

As we have seen, with a perturbation approach it is possible to localize where integration occurs: if the perturbed area is performing integration, the spatial tuning period is predicted to change. Generally speaking, perturbation modulates the effect of connectivity on dynamics, and the proposed readout is neural activity. This closed-loop approach allows for detailed tests of mechanistic neural models, whose very goal is to relate architecture and dynamics, in a way not easily rivaled by non-perturbative probes of connectivity or activity.

3.4 Methods

Figure 3.1 is schematic. Figure 3.2 is generated from ideal (imposed) periodic patterns but without dynamical neural network simulations. There-

fore, α is known exactly. In Figures 3.2, 3.4A,B, B.3, B.4, and B.7 relative phase is computed for convenience (to save the computational cost of generating spatial tuning curves, then deriving relative phases) from the population phases (thus, by setting $\delta^{ij} = \phi^i_{pop} - \phi^j_{pop}$). Figures 3.3, B.2, B.6, which distinguish between different recurrent architectures, are based on dynamical neural network simulations using the mature grid cell network described in Appendix B. Briefly, the model is a network of excitatory and inhibitory neurons (except in B.8 – see figure caption for details), with linear-nonlinear Poisson (LNP) spiking dynamics [16, 116]. For Figure B.5, we use Hodgkin-Huxley dynamics. Structured lateral interactions between neurons lead to pattern formation in the neural population. relative phases are explicitly computed from spatial tuning curves of cells, which are obtained from spike responses to 2-minute long simulated quasi-random trajectories. Velocity inputs drive shifts of the population activity pattern, resulting in spatially periodic tuning. Only cells from the simulation with good spatial tuning are included in the analysis of relative phase shifts: for fully and multibump periodic networks, this means all cells in the network, while for aperiodic networks this means cells in the central 3/4 of the network. Since the inhibitory and excitatory populations share similar population activity patterning and spatial tuning in these simulations, we made the arbitrary choice to display the inhibitory population. Perturbations are applied directly to network parameters τ_s and γ_{inh} (see Appendix B), as a result changing the measured spatial tuning curves, and thus leading to shifts in relative phase when compared across perturbations. The

stretch factor, α, is computed from the data as described in the subsection above.

Chapter 4

Derivation of the constraints governing the population activity pattern phenotype

4.1 Introduction

The last chapter introduced the notion that the period of the population pattern could be controlled by certain parameters of the network dynamics. In this chapter, we seek to understand this dependency, in particular, what selects or constrains the pattern phenotype. We start by analyzing the solution via linear stability analysis, followed by an alternative approach that takes into consideration the neural nonlinearity in order to derive self-consistency constraints on the pattern properties, and finally compare these analytical results with results from simulation.

4.2 Linear stability analysis

Our starting point is the equation for deterministic rate-based neurons in the continuum limit, given by:

$$\frac{ds(x,t)}{dt} + \frac{s(x,t)}{\tau_s} = f\left(\int_{-\infty}^{\infty} \gamma W(x-x')s(x',t)dx' + B(x)\right), \qquad (4.1)$$

where s describes the firing rate of the cell at location x in brain space at time t, τ_s is the neuron time constant, $B(x)$ is input external to the system, and f is a rectification nonlinearity: $f(u) = u$ if $u > 0$, else $f(u) = 0$. We will assume for now that the external is spatially uniform such that $B(x) = \overline{B}$. The coupling function, $W(x - x')$, describes the synaptic weight between cells located at positions x and x' in topological space and is translation invariant and symmetric. We will be varying two parameters: the prefactor γ, which controls the strength of the weights, and the time constant τ_s (similar to the last chapter).

The steady-state homogeneous solution, s_0, to this equation is given by

$$s_0 = \frac{\tau_s \overline{B}}{1 - \tau_s \gamma \overline{W}}, \tag{4.2}$$

where $\overline{W} = \int_{-\infty}^{\infty} W(x - x')dx'$, and we have assumed that $f(\gamma \overline{W} s_0 + \overline{B}) = \gamma \overline{W} s_0 + \overline{B}$ (i.e., $\gamma \overline{W} s_0 + \overline{B} > 0$). Consider a spatially dependent perturbation, $\epsilon(x, t)$, to the steady state solution. Substitution of $s(x, t) = s_0 + \epsilon(x, t)$ yields

$$\frac{d}{dt}\left(s_0 + \epsilon(x,t)\right) + \frac{s_0 + \epsilon(x,t)}{\tau_s} = f\left(\int_{-\infty}^{\infty} \gamma W(x - x')(s_0 + \epsilon(x', t))dx' + \overline{B}\right) \tag{4.3}$$

$$= f\left(\gamma \overline{W} s_0 + \overline{B} + \int_{-\infty}^{\infty} \gamma W(x - x')\epsilon(x', t)\right) \tag{4.4}$$

$$\approx f\left(\gamma \overline{W} s_0 + \overline{B}\right) + f'(\gamma \overline{W} s_0 + \overline{B}) \int_{-\infty}^{\infty} \gamma W(x - x')\epsilon(x', t) \tag{4.5}$$

where in the last line we have Taylor-expanded f about $\gamma \overline{W} s_0 + \overline{B}$. Canceling

terms yields an equation for the dynamics of the perturbation:

$$\frac{d\epsilon(x,t)}{dt} + \frac{\epsilon(x,t)}{\tau_s} = f'(\gamma\overline{W}s_0 + \overline{B}) \int_{-\infty}^{\infty} \gamma W(x-x')\epsilon(x',t)dx'. \tag{4.6}$$

Substituting the form $\epsilon(x,t) \sim e^{\alpha t + ikx}$, which is a single-mode perturbation, yields

$$\alpha + \frac{1}{\tau_s} = f'(\gamma\overline{W}s_0 + \overline{B}) \int_{-\infty}^{\infty} \gamma W(x-x')e^{-ik(x-x')}dx' \tag{4.7}$$

$$= f'(\gamma\overline{W}s_0 + \overline{B})\gamma\widetilde{W}(k) \tag{4.8}$$

$$= \gamma\widetilde{W}(k), \tag{4.9}$$

where we have used the definition of the Fourier transform of $W(x-x')$,

$$\widetilde{W}(k) = \int_{-\infty}^{\infty} W(x-x')e^{-ik(x-x')}dx', \tag{4.10}$$

and in the last line we have used the fact that $\gamma\overline{W}s_0 + \overline{B} > 0$, so that $f'(\gamma\overline{W}s_0 + \overline{B}) = 1$.

The condition for the *unstable* growth of the perturbation away from the homogenous solution is that the growth rate, $\mathrm{Re}(\alpha)$, be larger than zero. Using the fact that $\mathrm{Re}(\widetilde{W}(k)) = \widetilde{W}(k)$ when $W(x-x')$ is real and symmetric yields the following constraint for instability:

$$\widetilde{W}(k) > \frac{1}{\tau_s\gamma}. \tag{4.11}$$

The system is unstable to perturbations if there exists a mode of the synaptic weight function for which this condition is satisfied. The largest such mode is

defined as the *critical* mode, and determines the frequency of the pattern at the onset of instability.

For the special case in which the synaptic weight function is a difference-of-Gaussians (Figure 4.1A), namely

$$W(x) = \alpha_E e^{-x^2/2\sigma_E^2} - \alpha_I e^{-x^2/2\sigma_I^2}, \tag{4.12}$$

which has the following Fourier transform (Figure 4.1B)

$$\widetilde{W}(k) = \sqrt{2\pi}\alpha_E\sigma_E e^{-\sigma_E^2 k^2/2} - \sqrt{2\pi}\alpha_I\sigma_I e^{-\sigma_I^2 k^2/2}, \tag{4.13}$$

the critical mode predicted by linear stability analysis is given by the mode for which $d\widetilde{W}(k)/dk = 0$ (Figure 4.1B):

$$k^{*2} = \frac{2}{\sigma_E^2 - \sigma_I^2} \log\left(\frac{\alpha_E\sigma_E^3}{\alpha_I\sigma_I^3}\right). \tag{4.14}$$

Note that the particular parameters of interest, namely τ_s and γ, control only the threshold at which this mode goes critical, but not the critical mode itself. Thus, this does not explain the numerical observation that the wavelength of the solution *does* depend on these parameters (see Figure B.2). Next, we consider the effects of the nonlinearity.

4.3 Analysis of nonlinear solution

We know from numerical simulations that the steady state solution of the dynamics tends towards a rectified sinusoid, Figure 4.1C. Assume that the pattern is symmetric about and maximal at the origin. Let the period of

the pattern be a, and the point where the solution first crosses zero be x_0, Figure 4.1C.

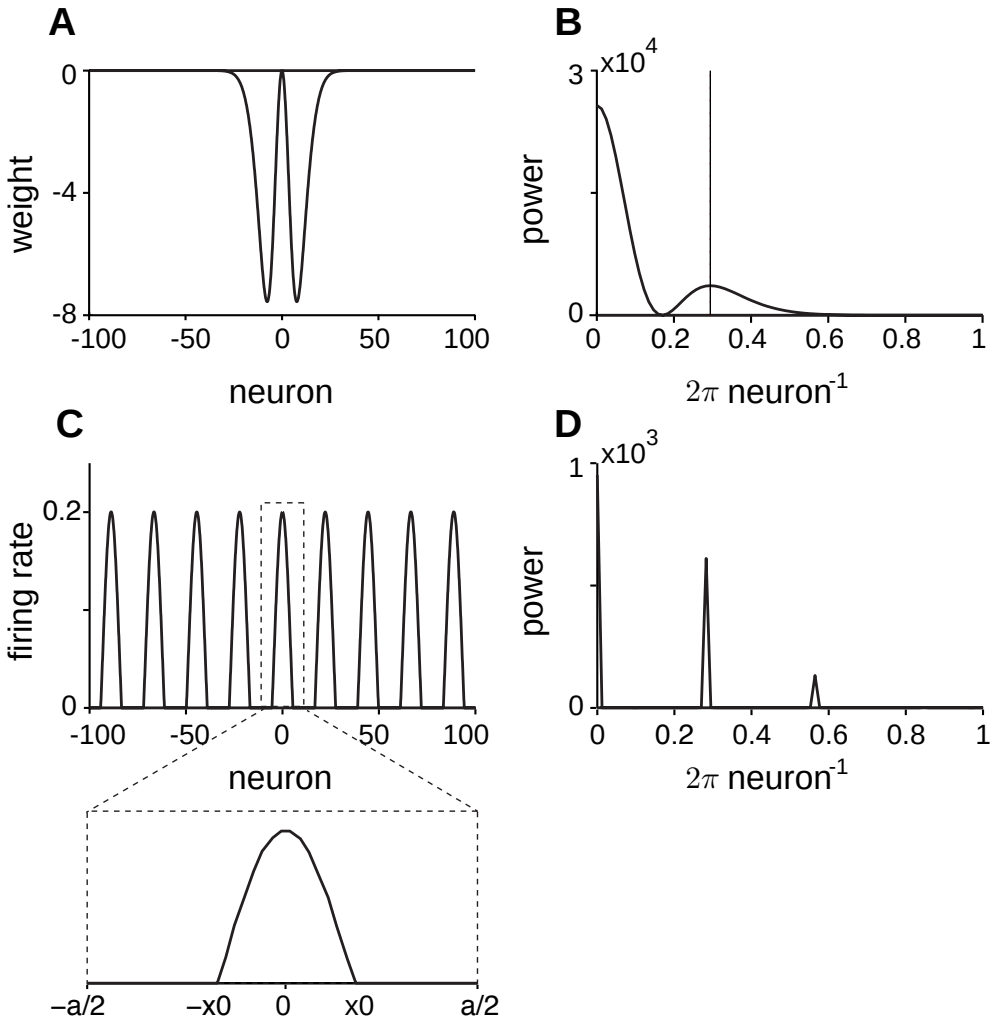

Figure 4.1: (A) Difference-of-Gaussians weight profile, Equation 4.12. $\alpha_E = 16$, $\alpha_I = 16$, $\sigma_E = 4$, $\sigma_I = 8$. (B) Power spectral density of profile in (A), with black line indicating critical mode. (C) Converged of the network dynamics using synaptic weights in (A) shows that the network activity is roughly a rectified sinusoid, with single-bump portion shown at bottom. (D) Power spectral density of the pattern in (C).

Accordingly, we will assume, in the presence of the nonlinearity, a steady-state solution of the form

$$\overline{s}(x) = \left[\sum_{l=-\infty}^{\infty} s_l \cos(klx) \right]_+ . \tag{4.15}$$

where $k = \frac{2\pi}{a}$. For the particular difference-of-Gaussian profile shown in Figure 4.1A, the power spectrum of the converged steady-state solution suggests the inclusion of 2 modes, Figure 4.1D. In steady state, the solution obeys the following equation:

$$\frac{\overline{s}(x)}{\tau_s} = f\left(\int_{-\infty}^{\infty} \gamma W(x-x')\overline{s}(x')dx' + \overline{B} \right). \tag{4.16}$$

The integral can be broken into intervals of length a:

$$\int_{-\infty}^{\infty} W(x-x')\overline{s}(x')dx' = \sum_{n=-\infty}^{\infty} \int_{a(n-\frac{1}{2})}^{a(n+\frac{1}{2})} W(x-x')\overline{s}(x')dx' \tag{4.17}$$

$$= \sum_{n=-\infty}^{\infty} \int_{-\frac{a}{2}}^{\frac{a}{2}} W(x-x'+na)\overline{s}(x'-na)dx' \tag{4.18}$$

$$= \sum_{n=-\infty}^{\infty} \int_{-\frac{a}{2}}^{\frac{a}{2}} W(x-x'+na)\overline{s}(x')dx'. \tag{4.19}$$

where, in the second to last line, we have used the change of variables, $x' \leftarrow x' - na$, and in the last line, we have used the fact that $s(x,t)$ is spatially periodic, with period a, so that $\overline{s}(x) = \overline{s}(x-na)$ for all n. Rewriting $W(x)$ in its Fourier basis as

$$W(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{W}(k)e^{ikx}dk, \tag{4.20}$$

and using the Poisson summation formula,

$$\sum_{n=-\infty}^{\infty} e^{ikna} = \frac{2\pi}{a} \sum_{l=-\infty}^{\infty} \delta(k - 2\pi l/a), \tag{4.21}$$

yields

$$\sum_{n=-\infty}^{\infty} \int_{-\frac{a}{2}}^{\frac{a}{2}} \gamma W(x - x' + na)\overline{s}(x')dx' = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \int_{-\frac{a}{2}}^{\frac{a}{2}} \int_{-\infty}^{\infty} \gamma \tilde{W}(k)\overline{s}(x')e^{ik(x-x'+na)}dkdx' \tag{4.22}$$

$$= \frac{1}{a} \sum_{l=-\infty}^{\infty} \int_{-\frac{a}{2}}^{\frac{a}{2}} \int_{-\infty}^{\infty} \gamma \tilde{W}(k)\overline{s}(x')e^{ik(x-x')}\delta(k - 2\pi l/a)dkdx' \tag{4.23}$$

$$= \frac{1}{a} \sum_{l=-\infty}^{\infty} \int_{-\frac{a}{2}}^{\frac{a}{2}} \gamma \tilde{W}(2\pi l/a)\overline{s}(x')e^{\frac{2\pi il}{a}(x-x')}dx' \tag{4.24}$$

$$= \sum_{l=-\infty}^{\infty} \gamma \tilde{W}(2\pi l/a)e^{\frac{2\pi ilx}{a}} \frac{1}{a} \int_{-\frac{a}{2}}^{\frac{a}{2}} \overline{s}(x')e^{\frac{-2\pi ilx'}{a}}dx' \tag{4.25}$$

$$= \sum_{l=-\infty}^{\infty} \gamma \tilde{W}(2\pi l/a)\hat{s}_l e^{\frac{2\pi ilx}{a}}, \tag{4.26}$$

where $\hat{s}_l$ is the Fourier series coefficient of $\overline{s}(x)$,

$$\hat{s}_l = \frac{1}{a} \int_{-\frac{a}{2}}^{\frac{a}{2}} \overline{s}(x)e^{\frac{-2\pi ilx}{a}}dx. \tag{4.27}$$

Thus, the steady-state equation has the form (replacing $\nu = \frac{2\pi}{a}$)

$$\frac{\overline{s}(x)}{\tau_s} = f\left(\sum_{l=-\infty}^{\infty} \gamma \tilde{W}(\nu l)\hat{s}_l e^{il\nu x} + \overline{B}\right). \tag{4.28}$$

Over the interval $[-x_0, x_0]$ where $\overline{s}(x)$ is positive, the Fourier coefficients become

$$\hat{s}_l = \frac{1}{a}\int_{-\frac{a}{2}}^{\frac{a}{2}} \left[\sum_{l'=-\infty}^{\infty} s_{l'}\cos(l'\nu x)\right]_+ e^{il\nu x}dx \tag{4.29}$$

$$= \frac{1}{a}\int_{-x_0}^{x_0} \left(\sum_{l'=-\infty}^{\infty} s_{l'}\cos(l'\nu x)\right) e^{il\nu x}dx \tag{4.30}$$

$$= \sum_{l'=-\infty}^{\infty} \frac{s_{l'}}{a}\int_{-x_0}^{x_0} \cos(l'\nu x)\cos(l\nu x)dx \tag{4.31}$$

$$= \sum_{l'=-\infty}^{\infty} M_{ll'}s_{l'}, \tag{4.32}$$

where

$$M_{ll'} = \frac{1}{a}\int_{-x_0}^{x_0} \cos(l'\nu x)\cos(l\nu x)dx \tag{4.33}$$

$$= \begin{cases} \frac{\sin(\nu(l-l')x_0)}{\nu a(l-l')} + \frac{\sin(\nu(l+l')x_0)}{\nu a(l+l')} & l \neq l' \\ \frac{x_0\nu}{2\pi} + \frac{\sin(2l\nu x_0)}{2l\nu a} & l = l', -l'. \end{cases} \tag{4.34}$$

Thus, over the interval $[-x_0, x_0]$, and using the ansatz for the steady-state solution in Equation 4.15, the steady-state equation 4.28 becomes

$$\frac{1}{\tau_s}\sum_{m=-\infty}^{\infty} s_m e^{im\nu x} = \sum_{l=-\infty}^{\infty} \gamma\tilde{W}(\nu l)\hat{s}_l e^{il\nu x} + \overline{B} \tag{4.35}$$

$$= \sum_{l=-\infty}^{\infty} \gamma\tilde{W}(\nu l)\left(\sum_{l'=-\infty}^{\infty} M_{ll'}s_{l'}\right) e^{il\nu x} + \overline{B}. \tag{4.36}$$

Matching components on the left and right sides, we generate a system of

equations for the steady-state solution coefficients, s_l,:

$$\frac{s_l}{\tau_s} = \sum_{l'=-\infty}^{\infty} \gamma \tilde{W}(\nu l) M_{ll'} s_{l'} + \overline{B}\delta_{l0} \tag{4.37}$$

$$\Rightarrow \sum_{l'=-\infty}^{\infty} \left(\frac{1}{\tau_s}\delta_{ll'} - \gamma \tilde{W}(\nu l) M_{ll'} \right) s_{l'} = \overline{B}\delta_{l0} \tag{4.38}$$

$$\Rightarrow \sum_{l'=-\infty}^{\infty} N_{ll'} s_{l'} = \overline{B}\delta_{l0} \tag{4.39}$$

$$\Rightarrow s_l = \overline{B} O_{l0}, \tag{4.40}$$

where O is the inverse of N such that $ON = 1$. We can derive an additional equation that constrains the relationship between a and x_0, by enforcing that the solution goes to zero at x_0:

$$\sum_{l=-\infty}^{\infty} s_l e^{il\nu x_0} = \quad 0 \tag{4.41}$$

$$\Rightarrow \sum_{l=-\infty}^{\infty} O_{l0} e^{il\nu x_0} = \quad 0. \tag{4.42}$$

For the difference-of-Gaussians weights used above, Equation 4.12, the solid lines in Figure 4.2A-B shows the manifold of solutions corresponding to the constraint in Equation 4.41, for different values of the parameter τ_s (holding γ fixed). Note that there are multiple branches in the solution space for each value of τ_s, and that for larger values of τ_s, the range of periods the solution can take increases.

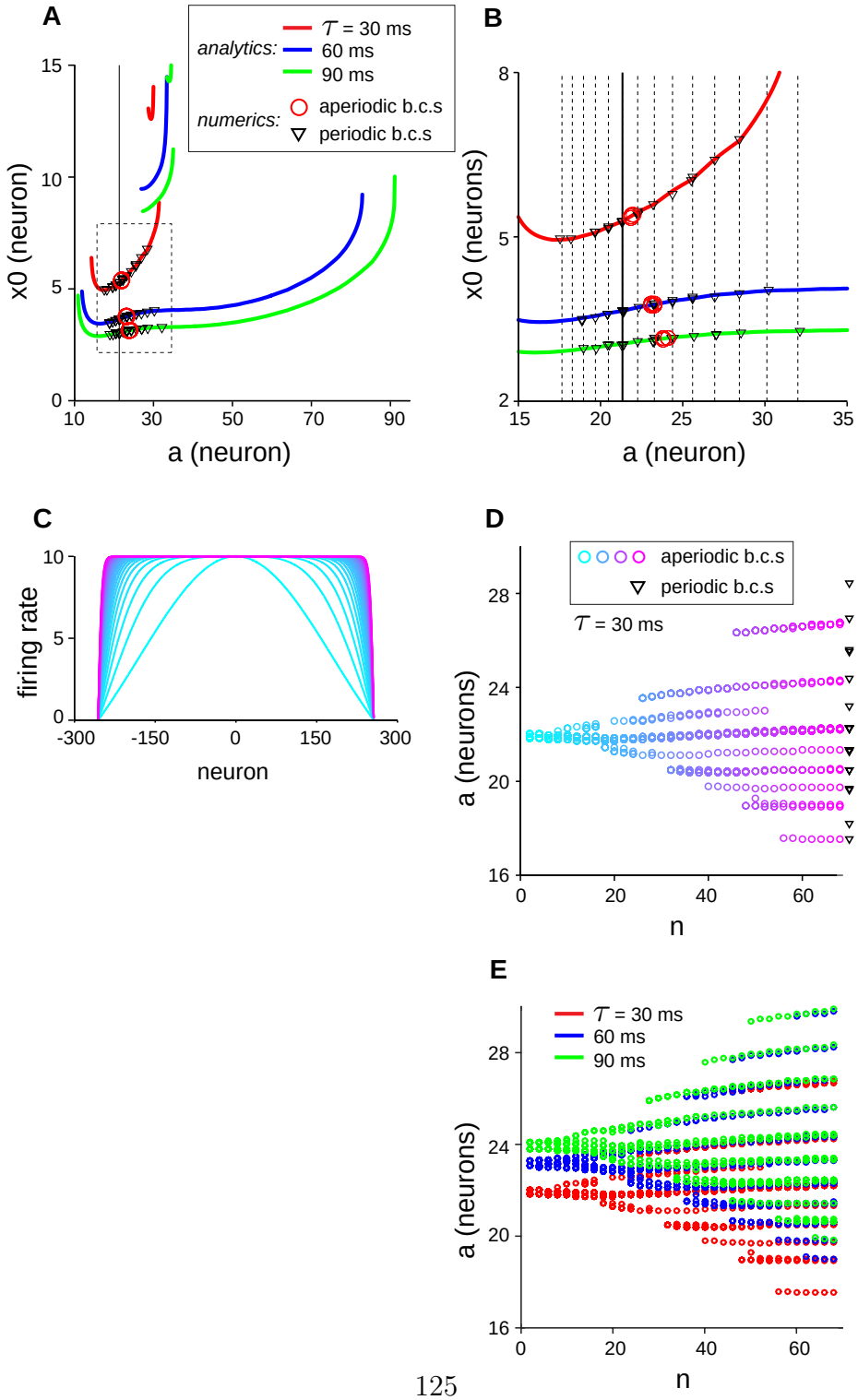
A
analytics:
$\mathcal{T}$ = 30 ms
60 ms
90 ms
numerics:
aperiodic b.c.s
periodic b.c.s
x0 (neuron)
a (neuron)
B
x0 (neurons)
a (neuron)
C
firing rate
neuron
D
aperiodic b.c.s
periodic b.c.s
$\mathcal{T}$ = 30 ms
a (neurons)
n
E
$\mathcal{T}$ = 30 ms
60 ms
90 ms
a (neurons)
n

Figure 4.2: (A) Comparison of analytical and numeric solutions for different boundary conditions. Solid lines: manifolds of analytic solutions, for different values of τ_s: 30 ms (red), 60 ms (blue), 60 ms (green). The converged numerical solutions for the periodic and aperiodic networks are shown as the blue asterisks (20 trials) and red circles (20 trials), respectively. Initial conditions span a large range of x_0's and a's (not shown). The black vertical line indicates the wavelength predicted from linear stability analysis. (B) Close-up of (A) (dotted box in (A)), showing local structure of manifold. The vertical black line is same as in (A). The dotted lines are predicted wavelengths based on the constraint that the number of bumps in the periodic network is an integer. (C) Envelopes used to generate data for aperiodic network in (D). (D) Bifurcation diagram showing converged pattern wavelengths for the aperiodic network as a function of the external input envelope shaping parameter, n (see Equation 4.44), for τ_s=30 ms, color-coded as in (C). The data for each n-value corresponds to the converged pattern wavelengths for 20 trials (trials initialized as in (A)). Black triangles are from a simulation with periodic boundaries and are the same data as on red curve in (A), here placed next to aperiodic simulations for comparison. (E) Bifurcation diagram for aperiodic network for different values of τ_s. (Red circles are the same data as shown in (D))

4.4 Numerical results

The discretized rate-equations have the form

$$\frac{ds_i}{dt} + \frac{s_i}{\tau_s} = f\left(\sum_j^N \gamma W_{ij} s_j + B_i\right). \tag{4.43}$$

We will consider both periodic and aperiodic network structures. For the *aperiodic* network, neurons are arranged on a 1D line, with the ith neuron located at position x_i, which ranges from $-N/2$ to $N/2$. The total size of the network is N. The synaptic weight matrix is given by $W_{ij} = W(x_i - x_j)$, where $W(x)$ is the difference-of-Gaussians function used above, Equation 4.12. For the *periodic* network, the neurons are arranged on a ring, so that the neurons at $-N/2$ and $N/2$ are identified.

Likewise, the external input, B_i, has a form that also depends on the boundary conditions of the network:

$$B_i = \begin{cases} \overline{B} & \text{periodic b.c.s} \\ \overline{B}(e^{-(x_i/(N/2))^n} - e^{-1})/(1 - e^{-1}) & \text{aperiodic b.c.s,} \end{cases} \tag{4.44}$$

where the parameter n controls the slope of the spatially modulated external inputs.

We numerically integrate Equation 4.43 for fixed network parameters and boundary conditions, and plot, for different initial conditions, the converged numerical solutions in Figure 4.2.

4.4.1 Periodic boundary conditions

Under periodic boundary conditions, because of the wide range of initial conditions, the numerical solutions are spaced out along the manifold at

intervals that correspond to constraining the number of bumps in the pattern to be an integer (Figure 4.2A-B, blue x's). In other words, the following relationship must hold:

$$a_m m = L, \tag{4.45}$$

where a_m and m are the wavelength and number of bumps in a pattern, and L is the size of the network. Using this constraint, one can predict the wavelengths of the other allowable solutions (dotted vertical lines in Figure 4.2B, which indicate a_m for $m = \{1, 2, 3, ...\}$. Moreover, the solutions are aligned across different values of τ_s (Figure 4.2B), and are therefore independent of τ_s.

4.4.2 Aperiodic boundary conditions

For aperiodic boundary conditions, in the special case of the external input, $B(x)$, in the form of a Gaussian, the numerical solutions converge to a *single* point on the manifold (Figure 4.2A-B, red circles). Consistent with our expectations from the last chapter, the wavelength of the solution increases as a function of τ_s (Figure 4.2A-B).

We can also vary the sharpness of the external input envelope. For increasingly rectangular envelopes (Figure 4.2C), the single unique solution in the case of the Gaussian input bifurcates into a multiplicity of solutions (Figure 4.2D). In the limit as $n \to \infty$, the aperiodic network behaves as a periodic network: the solutions are spaced in a way consistent with Equation 4.45 (the reason for the gradual expansion of the solution set is because the system size is effectively changing, due to the bulk interaction with the boundaries).

In Figure 4.2E, we see that for the hard, rectangular envelopes, the converged solution set across different values of τ_s are more or less equivalent. As the envelope becomes smoother, this solution overlap begins to diverge, so that for smooth, Gaussian envelopes, the solution is unique and depends on τ_s.

4.5 Discussion

In this chapter we have derived analytically the solution manifold that constrains the relationship between the pattern period and the width of the individual bumps in the pattern, assuming a 1D spatially-infinite continuum of neurons with translation invariant center-surround coupling. We have shown how this solution manifold depends on certain network parameters like the neuron time constant. In particular, for larger values of the time constant, the solution range along the period dimension increases. Numerically, for a finite system size, under both periodic and aperiodic boundary conditions, we have shown that the solutions converge to points along the manifold predicted by the analytics. In particular, for the network with periodic boundary conditions, the solutions are predictably spaced and correspond to different patterns with whole numbers of bumps. Moreover, the family of solution periods are independent of the neuron time constant. For aperiodic boundary conditions, the convergence of the solutions depends entirely on the shape of the external input, ranging from unique solutions in the case of smooth, tapering, Gaussian-like profiles to a multiplicity of solutions in the case of hard,

rectangular profiles. Intriguingly, in this latter case, the dispersion of solutions approaches the predictable spacing of a network with periodic boundary conditions, implying that in this limit, by this metric, the two networks are dynamically identical and thus similarly uninfluenced to changing the neuron time constant. Contrast this with the dynamics under a smooth tapering profile, where the unique solution, as a function of period, is very sensitive to the neuron time constant.

4.5.1 Implications for discriminating networks experimentally

In Chapter 3, we introduced the notion that the available experimental data on grid cells was consistent with a diversity of circuit designs. For multibump patterning networks, we distinguished between periodic and aperiodic boundary conditions, and showed how the two could be experimentally discriminated (from each other, and from a third category – singlebump periodic nets) based on how the population patterns responded to linearly increasing certain network parameters (like the neuron time constant). In that work, in simulations of the aperiodic network, we used smoothly-tapered Gaussian-like excitation profiles. Based on the insights from this chapter, namely that aperiodic networks with smooth boundaries are not representative of aperiodic networks in general (because aperiodic networks with hard boundaries behave like periodic networks), we now revise the three experimentally discriminable (based on the methods of Chapter 3) circuit architectures to read: 1) multibump aperiodic networks with smooth boundaries, 2) multibump pe-

riodic networks or aperiodic networks with hard boundaries, 3) singlebump periodic networks.

4.5.2 Future work

In this chapter, we set out to understand analytically the selection rule for the period of the pattern in the case of the aperiodic network, motivated from observations in Chapter 3. However, our approach reduces the problem to the interval over which the pattern is periodic. Thus, while the method yields solutions consistent with the numerics, we are unable to incorporate analytically the specific boundary profiles that may offer the extra constraints that narrow the analytic solutions to those obtained numerically. To do this may require a reformulation of the problem, which we leave for future work.

Chapter 5

Developmental plausibility and dynamical stability: determining the most plausible grid cell circuit architectures

A running theme of this thesis has been to try understand the circuit that underlies the grid cell response. In Chapter 3, we proposed an experimental strategy to distinguish between circuit architectures that are currently consistent with the available data: multibump periodic networks, multibump aperiodic networks with "hard" and "soft" boundaries, and single-bump periodic networks. In what follows, we consider the biological plausibility of each circuit, evaluated on the basis of the work presented in this thesis and its precursors.

One criterion for the biological plausibility of a circuit model is the plausibility of the required circuit architecture. Grid cells are not hardwired from birth [71, 119]; our model of grid cell development, laid out in Chapter 2, suggests that the topology of the space explored by the animal over the course of development constrains the architectural topology of the developing network. This implies that certain circuit models, constructed without reference to developmental constraints, may be more or less plausible.

A second criterion for biological plausibility is that the circuit dynamics should exhibit the same range of states as found experimentally, and be similarly stable. The grid cell code is believed to represent animal position by integrating self-motion cues. This representation is encoded in the grid cell population activity pattern phase, which depends on the pattern orientation and period. Detailed statistical analyses of experiments in which the spatial periods of grid cells change upon exposure to novel environments and distorted familiar environments has revealed that the spatial phase relationships between cell pairs do not change [3, 125]. This invariance implies that the underlying population pattern has not changed in orientation or period despite the changes in spatial tuning.

These findings are consistent with our expectations: to maintain an accurate representation of position, the orientation and period of the pattern remain fixed over time in a trial and on repeated visits to the same environment (i.e., across trials). In assessing circuit plausibility, it is important that the model circuit possess the same types of stability seen in experiment.

With these constraints in mind, we now discuss the plausibility of the three circuit designs in question.

5.1 Mutlibump periodic networks

- *Developmental plausibility:* In Chapter 2, Figure B.1, we argue that multibump periodic networks are developmentally implausible. This is based on the the necessity of precision long-range coupling of neurons

separated by multiple wavelengths of the population pattern (i.e., neurons whose location prefs during development are well separated). Thus, a developmental scheme for such networks is not known and the plausibility that such networks can form through exploration and plasticity is strongly doubtful.

- *Dynamical stability:* In both Chapters 3 and 4, it was shown that the population patterns of multibump periodic networks, across trials, have multiple discrete solutions in period. Simulations (unpublished) of 2D networks suggests that this is also the case for orientation of the pattern. This form of dynamical instability is inconsistent with the data [125]. Over the course of singe-trials, however, both the period and orientation are stable [16]. Thus, while the single-trial stability of these models is consistent with the data, their biological plausibility depends on whether the system is able to select the correct period and orientation across trials. If such stabilization occurs, it would likely involve feedback from the hippocampus or lateral entorhinal cortex.

We have argued that multibump periodic networks are developmentally implausible, and across trials are dynamically unstable.

5.2 Multibump aperiodic networks

We consider multibump aperiodic networks with "hard" and "soft" boundaries separately.

5.2.1 "Hard" boundaries

- *Developmental plausibility:* The developmentally most-probable circuit architecture, according to Chapter 2, is one with a planar topology, i.e., with aperiodic boundary conditions. This is because the network inherits its topology from the topology of the space in which the animal explores during development, and 2D space in an open or boxed-in arena is not periodic. Thus, the resulting network is also expected to be aperiodic. Multibump aperiodic networks with hard boundaries (without smoothly tapered input envelopes, which would require an additional learning mechanism to form) are developmentally plausible.

- *Dynamical stability:* Across trials, a given multibump aperiodic network architecture with hard boundaries can support multiple discrete solutions (in terms of period; Chapter 3), similar to multibump periodic networks. The question of across-trial stability of orientation is unclear, as this likely depends on the network boundary structure. For sufficiently square boundaries, the orientation is likely to take multiple discrete solutions, as in the multibump periodic networ. Over single trials, the period is stable, and is likely to be so for orientation. However, the multiplicity of solutions for both period and orientation, and potential variation across trials, puts into question the biological plausibility of such models.

We have argued that multibump aperiodic networks with hard boundaries are developmentally plausible, but suffer from non-uniqueness of the ac-

tivity period and orientation across trials.

5.2.2 "Soft" boundaries

- *Developmental plausibility:* In Chapter 2, it was assumed that during development the activity away from the network center tapered off smoothly. The biological plausibility of such an assumption is unclear. While cells called "border" cells are found in the vicinity of grid cells and are activated when the animal visits a particular environmental border, it is unclear if they are synaptically connected to grid cells. If so, as we argue in the discussion section of Chapter 2, such cells might provide the necessary inhibition to promote tapering of the activities near the edges, although there is an important distinction between the network edge and the edge of an explored environment. Thus, modeling the development of aperiodic networks with soft boundaries require more assumptions than do aperiodic networks with hard boundaries; however, the model is still aperiodic and roughly plausible.

- *Dynamical stability:* Dynamically, the period of the population pattern is stable within and across trials (Chapter 3). However, as was shown in [16], the orientation of the pattern tends to rotate slowly over the course of a trial. This is because the solution space in terms of orientation is continuous: all pattern orientations are allowed. This non-uniqueness of orientation within single trials is also problematic across trials. Unless there is some mechanism in place to uniquely select and stabilize the

orientation, such networks are not plausible.

We have argued that multibump aperiodic networks with soft boundaries are developmentally plausible (yet, slightly less so than aperiodic networks with hard boundaries), yet dynamically unstable given their tendency to rotate over the course of single trials.

5.3 Single-bump periodic network

- *Developmental plausibility:* In the appendix of Chapter 2, Figure A.4, we showed that further activity-dependent learning applied to a fully "mature" network can convert a multibump aperiodic network into effectively a single-bump periodic network. This occurs via the emergent coupling of co-active cells across single and multiple wavelengths of the population pattern. Therefore, single-bump periodic networks of this sort are developmentally plausible. We refer to this as the *developmental* single-bump periodic network, and distinguish it from the *hardwired* single-bump periodic networks considered in Chapter 3, *a la* [46]. In the latter case, it is as yet unclear developmentally how such a network should arise.

- *Dynamical stability:* As we showed in Chapter 4, both hardwired and developmental single-bump periodic networks are stable in period, both within and across trials. For the hardwired single-bump periodic network, in which there is only one bump in the network pattern, orien-

tation is not defined. For the case of the developmental single-bump periodic network, there are many bumps in the population pattern, even though each "domain" is periodically connected. Thus, orientation is well defined. While it is not clear to what extent the orientation varies within and across trials, it is likely to depend on the shape of the network boundary. In sum, both networks are stable in period, but the stability of the orientation is much less clear (or not even defined) and requires further investigation.

We have argued that single-bump periodic networks are developmentally plausible, dynamically stable in period, but in terms of orientation not well understood.

5.4 Conclusions

In summary, the multibump aperiodic and the single-bump periodic networks are the most developmentally plausible models of grid cell activity. The multibump aperiodic networks with soft boundaries and the single-bump periodic networks exhibit dynamics with a unique stable pattern period, for maximal consistency with the data. Periodic networks and aperiodic networks with hard boundaries are more stable in pattern orientation within a trial than are aperiodic networks with soft boundaries. However, in terms of trial-to-trial variability of orientation, all networks are likely unstable. Assuming some mechanism to stabilize orientation, the most biologically plausible networks,

developmentally and in terms of the uniqueness of the period, are the single-bump periodic and multibump aperiodic networks with soft boundaries.

Appendices

Appendix A

A model of grid cell development through spatial exploration and spike time-dependent plasticity

A.1 Scores and measures

Network velocity response: The GCN velocity response quantifies the translation velocity of the population pattern, v_{pop} (neuron/s), as a function of input animal velocity, v_{rat} (m/s) (see Figure A.2D, inset). Animal velocity is held constant for 10 seconds as the location of one of the peaks in the population pattern is tracked; animal velocity is incremented in steps (increments of 0.05 m/s). Within a step, the mean velocity of the population pattern is calculated over the 10-second window, after filtering the population pattern trajectory with a 2-second moving average filter. We fit a line to the plot of GCN velocity to animal velocity, for all data points with v_{pop} >0.1. *Velocity sensitivity* is the slope of this curve; the *velocity pinning threshold* is the x-intercept of this line.

Estimation of translation invariance: Translation invariance is measured as the inverse of the normalized local standard deviation of an off-diagonal row of the weight matrix (in this case, the middle half of the 5th off-diagonal of

the W_{II} matrix). Normalization is by the mean value.

Estimation of spatial tuning: The spatial tuning curve of a neuron is measured by first building a histogram of locations when the cell spikes (bin size = 1 cm). This histogram is then normalized (divided bin-by-bin) by the time spent in that spatial bin. The normalized histogram is smoothed by convolving with a boxcar filter (width = 5 bins).

Estimation of spatial tuning period and blob size: To measure *period*: For each cell, the power spectrum of the cell's spatial tuning curve for a single trial is computed. The period is taken to be the wavelength at which the power spectrum has the largest peak. To measure *blob size*: a Gaussian is fit to the central peak (i.e., the points between the troughs immediately to the left and right of the central peak) in the auto-correlogram of the cell's spatial tuning curve. The blob size is taken as the standard deviation of this Gaussian.

1D spatial gridness score: There is no clear 1D version of the commonly used 2D gridness score (see e.g. [95]). We developed a 1D gridness measure for spatial tuning. (Generalizing our 1D gridness score to 2D yields results that are similar to those derived from the common 2D gridness score, as in [95] (data not shown)). For each cell, we compute the power spectrum of its normalized spatial tuning curve (normalization involves mean subtraction and division by the standard deviation). The power spectrum is rescaled by $2/L^2$, where L is the number of bins in the tuning curve (1 cm bin size yields $L = 100$). The spatial gridness is taken to be the power of the largest frequency component in the power spectrum. This normalization yields a gridness score of 1 if the

the spatial tuning curve is a perfect sinusoid.

2D spatial gridness score: We use the same gridness score for 2D spatial tuning as used in [95]. Given the autocorrelogram of a cell's spatial tuning curve, we define an annular region with inner and outer radii, R_i and R_o (R_i and R_o are chosen by hand such that the annulus only contains the 6 peaks closest to the origin, excluding the peak at the center). Then, the Pearson correlation coefficient is computed between this map and the same map rotated by an angle $\phi = \frac{i}{180}$ degrees, where i is an integer that increments from 1 to 180; this defines a vector of correlation coefficients, $\rho(\phi)$. The gridness score is defined as

$$gridness = \min\{\rho(60^\circ), \rho(120^\circ)\} - \max\{\rho(30^\circ), \rho(90^\circ), \rho(150^\circ)\}. \quad \text{(A.1)}$$

Inter-trial stability score: For each cell, the Pearson correlation coefficient is computed between the spatial tuning curves associated with two different trials. Trials begin with the same initial populaton phase, and involve two separate trajectories selected from a larger random trajectory (see below on the generation of a the random trajectory) to have similar starting locations (with a difference no greater than 1 cm).

If n_i is the spike rate density in the i^{th} bin of the cell's spatial tuning curve for a given trial, then the cell's inter-trial stability, ρ_{ITS}, measured

between trials t_1 and t_2 is given by

$$\rho_{ITS} = \frac{\sum_{i=1}^{L}(n_i^{(t_1)} - \overline{n}^{(t_1)})(n_i^{(t_2)} - \overline{n}^{(t_2)})}{\sqrt{\sum_{i=1}^{L}(n_i^{(t_1)} - \overline{n}^{(t_1)})^2}\sqrt{\sum_{i=1}^{L}(n_i^{(t_2)} - \overline{n}^{(t_2)})^2}} \tag{A.2}$$

where L is the number of bins in the spatial tuning curve and $\overline{n}$ is the mean spike rate density in a given trial.

Spatial coherence: For each cell, the Pearson correlation coefficient is measured between the normalized spatial tuning curve and the same spatial tuning curve, but whose ith element is replaced by the mean of the 8 nearest elements (excluding itself). Formally, if n_i is the spike rate density in the ith bin of the cell's spatial tuning curve, then coherence, ρ_{coh}, is given by

$$\rho_{coh} = \frac{\sum_{i=1}^{L}(n_i - \overline{r})(m_i - \overline{m})}{\sqrt{\sum_{i=1}^{L}(n_i - \overline{n})^2}\sqrt{\sum_{i=1}^{L}(m_i - \overline{m})^2}} \tag{A.3}$$

where L is the number of bins in the spatial tuning curve and

$$m_i = \frac{1}{M}\sum_{\substack{j=-\frac{M}{2} \\ j\neq 0}}^{\frac{M}{2}} n_{i+j}, \tag{A.4}$$

and M is the number of bins included in the average.

Population activity gridness and period: Similar to the spatial gridness, the population activity gridness is taken to be the power of the largest frequency component of the power spectrum measured from a normalized snapshot of the population activity (normalized = mean subtracted, followed by division by standard deviation) The power spectrum is rescaled by the factor $2/\mathrm{L}^2$, where L is the number of bins in the population activity vector from which the

power spectrum was computed. The population activity vector is shortened to include only the middle one-half of the population, so that for the E^L population, L is 100. From the power spectrum, the population activity period is taken to be the wavelength at which the power spectrum has the largest peak.

Spatial phase and relative *phase*: The phase of a cell is the normalized offset in the central peak from zero in the the cross-correlation of the spatial response of the cell of interest and a reference spatial response (the reference is the first-trial spatial tuning of a particular cell in the same GCN; it is held fixed across trials). If d is the offset in the central peak of the cross-correlation, the spatial phase of the α^{th} cell in a given trial is defined as

$$\phi^{\alpha} = \frac{d}{\lambda} \bmod 1, \tag{A.5}$$

where λ is the cell's spatial tuning curve period (see above) measured from that trial, and $\phi^{\alpha} \in [0, 1)$. The *relative* spatial phase between cells α and β in a given trial is given by

$$\delta^{\alpha\beta} = (\phi^{\alpha} - \phi^{\beta}) \bmod 1. \tag{A.6}$$

For any phase ϕ, phase magnitude is defined as $|\phi| = \min\{\phi, 1 - \phi\}$, where $|\phi| \in [0, 0.5)$. If ϕ_{t_1} and ϕ_{t_2} are phases measured in two separate trials, then the change in phase is defined as

$$\Delta_t(\phi) = (\phi_{t_1} - \phi_{t_2}) \bmod 1. \tag{A.7}$$

To compute the development of relative spatial phase as in Figure 2.5E, at each developmental time point, and for each trial at that time point, the relative phase $\delta^{\alpha\beta_*}$ is computed for each cell α with respect to a fixed reference cell β_*, and then averaged over all trials.

Pairwise temporal correlation coefficient: The pairwise temporal correlation coefficient is the Pearson's correlation coefficient measured between spike trains for a pair of cells, after first convolving each spike train with with a Gaussian kernel ($\sigma = 15$ ms for the 1D GCN and $\sigma = 270$ ms for the 2D GCN).

Direction tuning: First, spikes are binned according to the animal's heading direction (in 1D, there are two directions and thus two bins, left and right; the binning is over $[0°, 360°)$, with bin size $\Delta = 6°$) and then normalized (divided bin-by-bin) by the time spent in that bin. This gives a directional spike rate density. The mean direction vector is

$$\rho_{dir} = \frac{\sum_{\theta=0}^{360} r_\theta e^{\imath\theta}}{\sum_{\theta=0}^{360} r_\theta}. \tag{A.8}$$

where r_θ is the spiking rate in the bin defined by $[\theta, \theta + \Delta]$. The mean vector length, i.e., the direction tuning strength, is given by $|\rho_{dir}|$.

Speed tuning: To compute the *speed tuning* of a given cell, spikes are binned by animal speed and running direction. Separate bins are assigned for the same speed if along different directions (the preferred and anti-preferred directions); bin = 0.1 m/s. Within each bin, we then compute the mean number of spikes. We then compute the slopes and intercepts of the resulting mean spike rate per bin versus speed bin plots, for each direction. For the speed tuning of

Figure A.6, in the manner of [118], firing rates are binned according to the absolute value of the animal's velocity, i.e., the preferred and anti-preferred directions are combined. The rest is the same as above.

A.2 Generating spike trains with CV < 1.

To generate stochastic spikes with a lower CV given by CV $= 1/\sqrt{M}$ (with M an integer greater than 1), we use the following procedure, unmodified, from [16]: First subdivide each time interval dt into M sub-intervals of length dt/M. In each of these finer time-steps $[t, t + dt)$, generate k spikes according to an inhomogeneous Poisson process with instantaneous rate $\lambda = Mr(t)$:

$$P(k|\lambda) = \frac{e^{-\lambda dt}(\lambda dt)^k}{k!}. \tag{A.9}$$

Finally, go through the spikes in order of emission, counting them and retaining only every Mth spike. This procedure generates a spike train with rate r and CV $= 1/\sqrt{M}$.

A.3 Generating a quasi-random 1D exploration trajectory.

The quasi-random trajectory is generated as follows. Starting from an initial location, a trajectory segment is drawn by picking a velocity and a time interval from uniform distributions on the intervals [-1,1] m/s and [0,0.02] seconds, respectively, and integrating them to obtain position coordinates over

the interval. If any part of this segment would extend outside the enclosure (i.e., outside the [0,1]-meter interval), a different velocity and time interval are picked, until one is found that respects the enclosure confines. At the end of that segment, another is picked, and so on. A 4-hour duration trajectory is assembled in this way. The resulting trajectory is smoothed with a moving average filter of width 1s (2000 bins). Because smoothing leads to possible expansion of the trajectory outside the bounds of the enclosure, the smoothed trajectory is globally rescaled so that the minimum and maximum are contained within [0,1] m. This process determines animal location $\vec{x}(t)$, and by its derivative, the velocity $\vec{v}(t)$ (see Figure A.1). The exploration speeds in this 1D trajectory are consistent with recorded animal exploration speeds in 2D.

Because of the steps taken to ensure that the trajectory remains within the enclosure, it is technically quasi-random rather than random. For simplicity, we refer to it as random in the main text and elsewhere.

The quasi-random 10-second trajectories used for GCN testing in Figures 2.4, and 2.5, and 2.6, were extracted from the longer developmental trajectory such that, within 10 seconds, the animal begins within the same spatial interval, $[x_0, x_0 + \delta x]$, where $x_0 = 0.1$ and $\delta x = 0.01$) and touches both boundaries. Also, the trajectories are selected to be non-overlapping. The 60-second trajectories used in Figure 2.5 are just the extensions of the set of 10-second trajectories taken from the developmental trajectory. The 2-minute long trajectory used in Figure 2.6 is taken to be the first 2 minutes of the developmental

training trajectory.

A.4 1D hardwired model with translation-invariant weights.

The weights going from population P' to P, and from cells i and j, are described as follows (see Figure A.2):

$$W^{PP'}(i,j) = \alpha A^{PP'}(i,j)\Theta(\mu(i-\gamma j))\Theta(|i-\gamma j|-\delta)\times \left[\exp\left[\frac{-(i-\gamma j-\Delta)^2}{2\sigma^2}\right] + \beta \exp\left[\frac{-(i-\gamma j+\Delta)^2}{2\sigma^2}\right]\right], \tag{A.10}$$

where $\gamma = \frac{N_P}{N_{P'}}$ (N_P is the size of population P), Θ is the Heaviside function ($\Theta(x) = 0$ for $x < 0$ and is 1 otherwise), and A is an envelope function that tapers the weights, where $A^{PP'}(i,j) = A_i^P A_j^{P'}$,

$$A_i^P = \begin{cases} 1 & r_i^P < \eta N_P \\ \exp\left[-a_0\left(\frac{r_i^P - \eta N_P}{(1-\eta)N_P}\right)^2\right] & \text{otherwise} \end{cases} \tag{A.11}$$

and $r_i^P = |i - \frac{N_P}{2}|$, $\eta = 0.28$, and $a_0 = 60$. The resulting weights are translation invariant (minus the tapering at the boundaries).

$E^L \to I$: $\alpha = 52$; $\beta = 0$; $\Delta =$ -1; $\sigma = 2$; $\mu = 0$; $\delta =$ -1;

$E^R \to I$: $\alpha = 52$; $\beta = 0$; $\Delta = 1$; $\sigma = 2$; $\mu = 0$; $\delta =$ -1;

$I \to E^L$: $\alpha =$ -9; $\beta = 0$; $\Delta = 4$; $\sigma = 5$; $\mu = 1$; $\delta = 3$;

$I \to E^R$: $\alpha =$ -9; $\beta = 0$; $\Delta =$ -4; $\sigma = 5$; $\mu =$ -1; $\delta = 3$;

$I \to I$: $\alpha =$ -60; $\beta = 1$; $\Delta = 2$; $\sigma = 3$; $\mu = 0$; $\delta = 3$;

In Figure A.3, the weight parameters are as follows:

$E^L \to I$: $\alpha = 4.6$; $\beta = 0$; $\Delta =$ -1; $\sigma = 4$; $\mu = 0$; $\delta =$ -1;

$E^R \rightarrow I$: α = 4.6; $\beta = 0$; $\Delta = 1$; $\sigma = 4$; $\mu = 0$; δ = -1;

$I \rightarrow E^L$: α = -7.5; $\beta = 0$; $\Delta = 4$; $\sigma = 5$; $\mu = 1$; $\delta = 3$;

$I \rightarrow E^R$: α = -7.5; $\beta = 0$; Δ = -4; $\sigma = 5$; μ = -1; $\delta = 3$;

$I \rightarrow I$: α = -2.2; $\beta = 1$; $\Delta = 2$; $\sigma = 3$; $\mu = 0$; $\delta = 3$;

$E^L \rightarrow E^L$: $\alpha = 13$; $\beta = 0$; Δ = -2; $\sigma = 2$; $\mu = 0$; δ = -1;

$E^R \rightarrow E^R$: $\alpha = 13$; $\beta = 0$; $\Delta = 2$; $\sigma = 2$; $\mu = 0$; δ = -1;

$E^L \rightarrow E^R$: $\alpha = 13$; $\beta = 0$; $\Delta = 0$; $\sigma = 2$; $\mu = 0$; δ = -1;

$E^R \rightarrow E^L$: $\alpha = 13$; $\beta = 0$; $\Delta = 0$; $\sigma = 2$; $\mu = 0$; δ = -1;

A.5 Rescaling the learning rate of inhibition or the inhibitory strength.

The scale of learning rate of inhibition, γ, (Figure 2.7) modifies the learning rates of the inhibitory learning rates: $\gamma_{II} \rightarrow \gamma\gamma_{II}$ and $\gamma_{EI} \rightarrow \gamma\gamma_{EI}$. Similarly, the strength of inhibition (Figure A.7) modifies the inhibitory synaptic weights: $W^{II} \rightarrow \gamma W^{II}$ and $W_{EI} \rightarrow \gamma W^{EI}$.

A.6 Development of periodic GCN.

If plasticity is enabled during the activation phase post pattern-formation, then neurons with similar phases should become coupled. If neurons across the GCN with a similar spatial tuning phase are wired up, and if this happens for neurons of all spatial phases, then the GCN topology is equivalent to a single-bump network with periodic boundary conditions (a ring network in 1D

or a single-bump twisted-torus network in 2D).

However, as discussed in the text, enabling plasticity during the activation phase is problematic because of the positive feedback loop between recurrent-driven activity and potentiation of the recurrent connections themselves. The tendency is that at whichever specific population pattern phase the recurrent activity-based plasticity is turned on, weights become strengthened and that specific population pattern becomes a discrete fixed-point of the network dynamics; the tendency is toward a degradation of the translation-invariance across phases of the population pattern.

To minimize and overcome this effect, we assume a sharply reduced learning rate for the recurrent synapses during this phase and incorporate neural spike frequency adaptation dynamics, which force the network pattern to flow rather than remain pinned at specific phases. The total synaptic current $I_i^P(t)$ is modified as follows:

$$I_i^P(t) = A_i^P \alpha^{P,vel}(v,t)(g_i^{P,rec}(t) + g_i^{P,loc}(t) + g_i^{P,adapt}(t) + g^0) + A_i^P g^{0\prime}, \quad \text{(A.12)}$$

where $g_i^{P,adapt}(t) = \alpha_{adapt} a_i^P(t)$ and

$$\frac{da_i^P(t)}{dt} + \frac{a_i^P(t)}{\tau_{adapt}} = \sum_b \delta(t - t_{i,b}^P). \quad \text{(A.13)}$$

For the results shown in Figure A.4, $\alpha_{adapt} = 4$ and $\tau_{adapt} = 120$ ms, and $\eta = 7.5 \times 10^{-5}$ s^{-1}. All other parameters and equations remain unchanged from those specified in the Methods.

A.7 Revealing low-dimensional local connectivity in the weight matrix: Sorting by simulated annealing.

We define a cost function on neural indices and weights, that imposes a penalty when cells that share are strong connection are assigned distant indices. Minimizing the cost function with respect to neural indices provides a sorting order for cells. We use this ordering to visualize structure in the randomized connectivity matrix of Figure A.2. The cost function is given by:

$$E = \sum_{i,j} W_{ij}^2 (i-j)^2. \tag{A.14}$$

To minimize the cost function, we employ simulated annealing [67]. First, index the N cells and construct a synaptic matrix from this ordering. In each step of the algorithm, the indices of two cells chosen at random are swapped, and the difference in the cost before and after the swap is evaluated. If $\Delta E \leq 0$, i.e., the cost has decreased, the swap is accepted; if $\Delta E > 0$, the swap is accepted with probability $P(\Delta E)$ given by the Boltzmann distribution:

$$P(\Delta E) = e^{-\Delta E/T}, \tag{A.15}$$

where T is a control parameter. T is initially large ($T_0 = 1 \times 10^5$) and is lowered in stages such that $T_n = 0.9^n T_0$ in the nth stage. Transition to the next stage is made after 1×10^4 flips have been accepted in the present stage. The process is terminated when the number of steps within any given stage reaches a maximum value, 1×10^5, before 1×10^4 flips have been accepted.

A.8 Supplemental figures

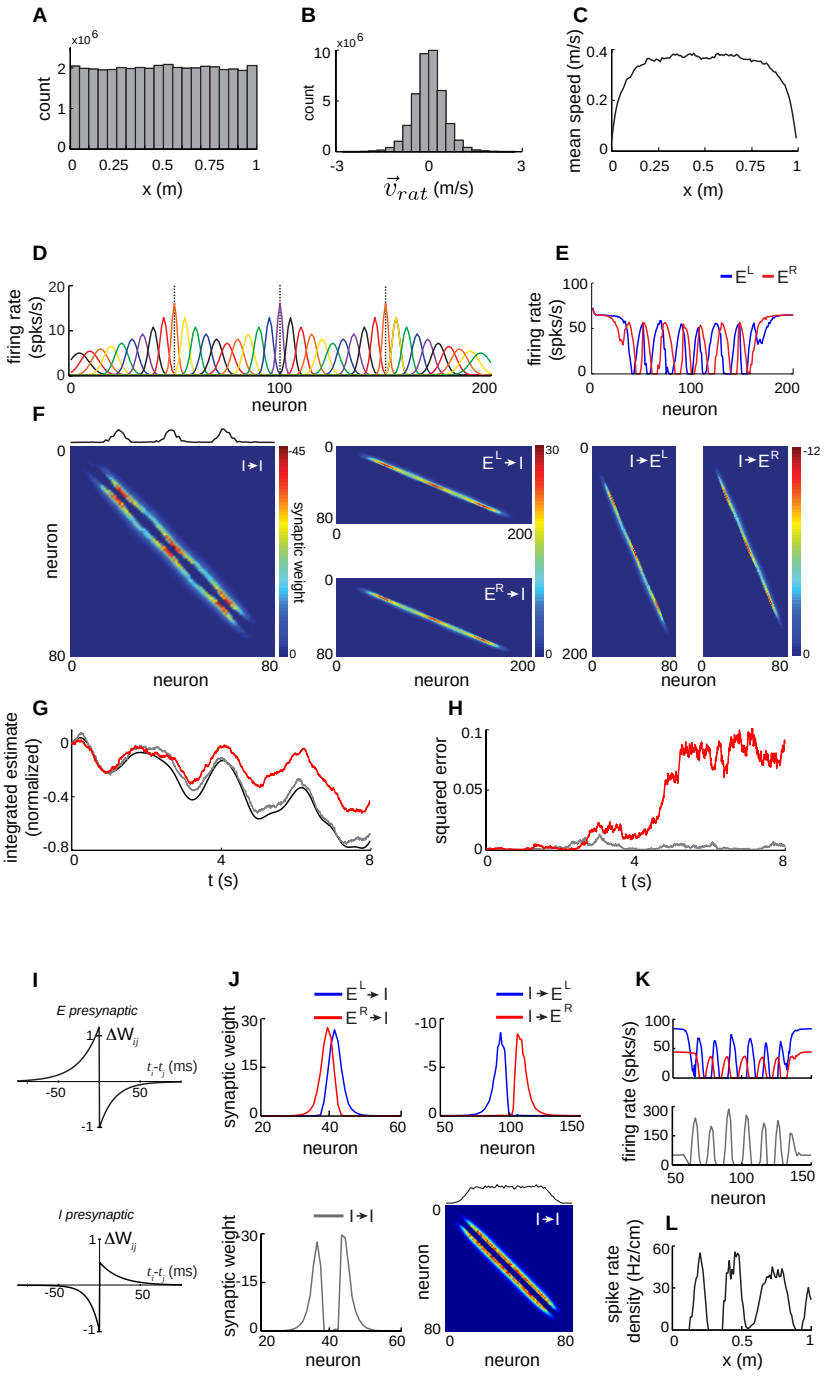
A
x10^6
count
x (m)
B
x10^6
count
$\vec{v}_{rat}$ (m/s)
C
mean speed (m/s)
x (m)
D
firing rate (spks/s)
neuron
E
E^L E^R
firing rate (spks/s)
neuron
F
I→I
E^L→I
E^R→I
I→E^L
I→E^R
neuron
synaptic weight
G
integrated estimate (normalized)
t (s)
H
squared error
t (s)
I
E presynaptic
ΔW_ij
t_i-t_j (ms)
I presynaptic
J
E^L→I
E^R→I
I→E^L
I→E^R
synaptic weight
neuron
I→I
K
firing rate (spks/s)
neuron
L
spike rate density (Hz/cm)
x (m)

Figure A.1: **Supplemental figures associated with Figure 2.1.** (A-C) *Statistics of the (quasi)random 1D trajectories generated for development.* (A) Histogram of visited locations in 1-meter enclosure for the entire training trajectory (4 hours). (B) Histogram of instantaneous animal velocities for training trajectory (mean speed, $\overline{v_{rat}} = 0.36$). (C) Mean speed as a function of position within enclosure, computed from training trajectory. (D-H) *Development with non-uniform location-specific inputs.* (D) Snapshots of the location-specific input during training. Both the width and amplitudes of the location-specific inputs are varied, across locations. The three vertical black lines mark landmark locations. The standard deviations of location-specific inputs are proportional to the distance from the nearest landmark. (E) Snapshot of the excitatory L and R population activities from the mature GCN, after training. (F) Synaptic weights of the mature GCN. The curve above the I-to-I matrix shows the off-diagonal, which reveals the non-uniformity in the weights. (G) Path integration performance: Red curve: neural representation of location over an 8-second animal trajectory (black curve), from the GCN trained on non-uniform location-specific inputs, as in (D). Gray curve: neural representation of location by mature GCN from main text (trained with uniform visual inputs). The red and gray curves represent integrated location in GCN coordinates; to find scale-factors to convert each curve into location coordinates, we simply performed a least-squares linear regression (with zero y-intercept) on the first second of each curve onto the actual location (black curve). (H) Squared error, measured point-by-point, between the normalized GCN trajectories in (G) and the true animal trajectory. (I-L) *GCN development with time-reversed plasticity windows: Hebbian STDP in I-to-E and anti-Hebbian STDP in E-to-I synapses.* (I) Time-reversed plasticity kernels (relative to Figure 2.1D) used for development. All other parameters identical to 1D model parameters from main text. (J) Top: E-to-I (left) and I-to-E (right) synaptic weight profiles in the mature GCN (at T = 4 hours). Note reversal of weight asymmetries as compared to Figure 2.2A-B, row 4 in main text. Bottom: I-to-I synaptic weight profile (left) and full connectivity matrix (right). (K) Snapshot of the E (top) and I (bottom) population activity patterns in the mature GCN in the activation phase. (L) Sample spatial tuning curve from a single neuron in the E population, measured over a 10-second trajectory (see Appendix A).

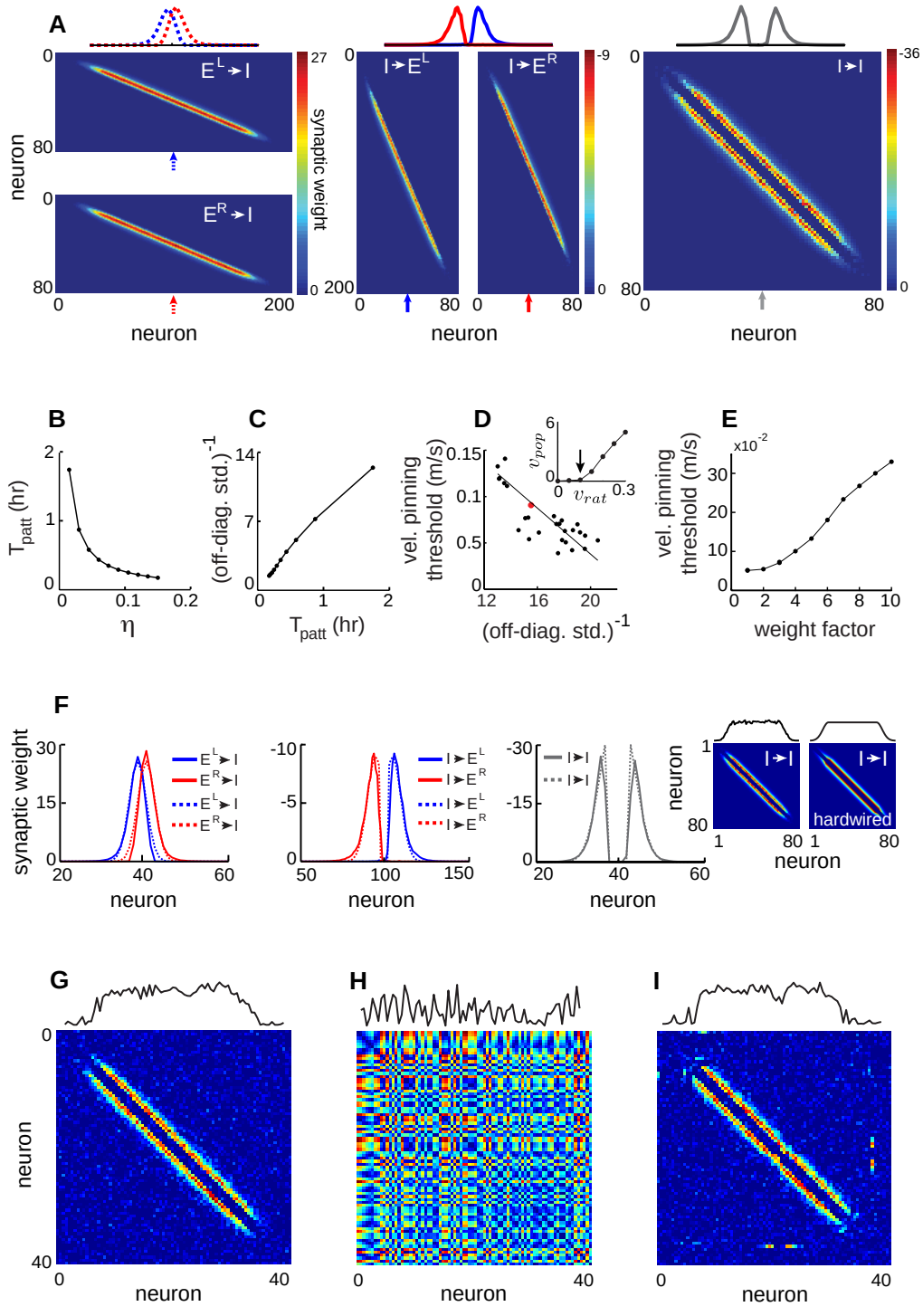
A
B
C
D
E
F
G
H
I
neuron
synaptic weight
vel. pinning threshold (m/s)
weight factor
hardwired

Figure A.2: **Supplemental figures associated with Figure 2.2.** (A) Synaptic weight matrices corresponding to the 1D GCN and figures in the main text, learned under conditions in which the visual cues are uniform (contrast with Figure A.1). Top: Slices from the weight matrices, taken at the locations of the color-coded arrows. (B-F) *Tradeoffs in translation invariance and learning time.* (B) The population patterning time, T_{patt} (see Figure 2.5), increases as the learning rate parameter, η, decreases. (C) Translation invariance (assessed by the inverse standard deviation of the off-diagonal weight band of the GCNat T_{patt}; see Appendix A) increases as T_{patt} increases. (D) The pinning threshold (defined as the maximum animal speed that fails to elicit a translation of the population activity pattern, and given by the threshold in the GCN's velocity response as a function of input animal speed, arrow in inset; see Appendix A), drops with increasing translation invariance. Black line: linear fit to results (Pearson's correlation: r = -0.75 ($p<0.01$)). (E) In a hardwired network with weights based on the mature GCN but with hand-imposed translation invariance (at least down to the the discreteness of single neurons; see (F) and Appendix A), the pinning threshold grows as the synaptic weights are scaled up in size. *Weight factor* is a multiplicative scaling; a 1 corresponds to weights of same strength as in the mature GCNof Figure 2.2. (F) Hardwired GCN with translation-invariant weights used in (E). Left three panels, dotted lines: Cross-sections of the E-to-I (first panel), I-to-E (second panel), and I-to-I (third panel) synaptic weight profiles in the hardwired network. Solid lines: the mature 1D GCN weights profiles from the main text. Rightmost two panels: weight matrices of the mature (left) and hardwired (right) GCNs. The hardwired network is translation-invariant (down to the resolution of single-neuron discreteness) by construction, but otherwise is matched to the mature GCN. (G-I) *Discovery of simple architectures in weight matrix data from a non-topographic GCN.* (G) I-to-I synaptic weight matrix organized according to inherited preferred location. To each element of the matrix, independent white noise samples (drawn from a normal distribution $\mathcal{N}(0, a^2)$, where a is 0.5 times the standard deviation of the distribution of weights in the I-to-I synaptic matrix) have been added. A cross section of the matrix is plotted above. (H) I-to-I synaptic weight matrix as it would appear if the inherited location preferences were unknown, as expected from a connectomics dataset generated from a network that is not topographically organized. (I) Matrix in B reordered via simulated annealing, sorted by synaptic weight size (see Appendix A). *Interpretation of results in* (B-E): The learning time T_{patt} can

be shortened with larger learning rate (B), but at a fixed weight threshold the result is less translation invariance (C) and more pinning during velocity integration (D). If the network weights continue growing after the weight threshold for pattern formation is crossed, to increase translation invariance, this can also hurt integration even if the final weights are translation-invariant (E), simply because stronger weights than necessary tend to enhance pinning. These observations illustrate the tradeoff between a fast approach to network patterning and the final quality of integration in the network, and suggest that it is optimal for the learning rate to be low enough so that at threshold for pattern formation, enough traversals have been performed for translation invariance.

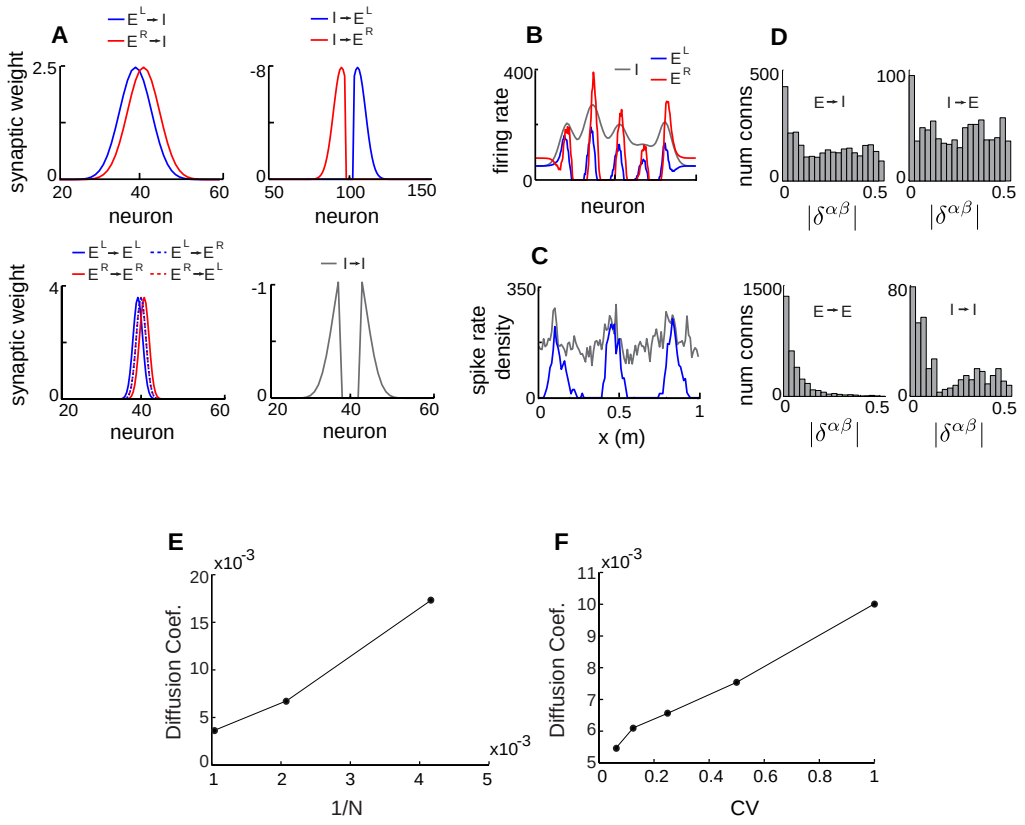

Figure A.3: **Supplemental figures associated with Figure 2.3.** (A-D) *E-to-E connectivity enables grid-like spatial tuning in the E population with little tuning in the I population (compare with Figure 2.3).* (A) E-to-I (top left), I-to-E (top right), E-to-E (bottom left), and I-to-I (bottom right) synaptic weight profiles in a modified version of the mature GCN of the main manuscript (see Appendix A). (B) Snapshot of the E (red and blue) and I (gray) population activity patterns. (C) Sample spatial tuning curves from non-edge neurons of the E (blue) and I (gray) populations, measured over a 10-second trajectory. (D) The number of connections between cell pairs as a function of the magnitude of the relative phase between the cells, $|\delta^{\alpha\beta}|$, for the E-to-I (top left), I-to-E (top right), E-to-E (bottom left), and I-to-I (bottom right) weights. A synapse whose strength is $>= 5\%$ of the largest strength for that synapse type counts as a connection. The distribution of spatial phases of the E cells that project to any I cell is broad and I cells display weak or no spatial tun-

ing, consistent with the findings in [14]. (E-F) *Parametric quantification of integration performance in the mature GCN.* (E) Diffusivity or the diffusion coefficient (i.e. the slope of the mean squared displacement of the population pattern trajectory, measured as the center of one of the activity bumps over time, as a function of temporal displacement) of the population pattern at zero input velocity, as a function of GCN size (N = 480 is the size of the GCN in the main text). Diffusivity characterizes how much (squared) error integration error is introduced by ongoing noise in the GCN per unit time. (F) Diffusion coefficient as a function of the coefficient of variance (CV) of the spiking process (see Methods and Appendix A).

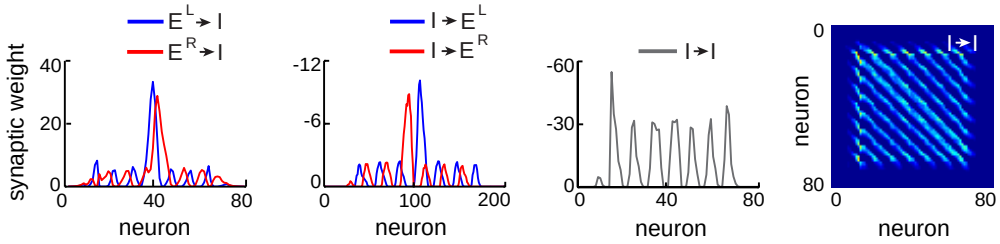

Figure A.4: **Supplemental figures associated with Figure 2.4.** *Development of a GCN with periodic connectivity.* Beginning with the mature 1D GCN, we allow synaptic modification to occur through the same STDP rules by releasing the restriction on plasticity during the activation phase. During this period, the learning rate is sharply diminished, and we have added adaptation to the neural dynamics (see Appendix A for details). These modifications are all applied at T = 4 hours in the original simulations. With these modifications, at T = 5 hours, cells with similar spatial phases all wire together, and the GCN becomes topologically periodic: E-to-I (first panel), I-to-E (second panel), and I-to-I (third panel) synaptic weight profiles, and the full I-to-I synaptic weight matrix (fourth panel).
Notes: This kind of learning is unstable because of strong positive feedback that results from the expression of recurrent weights during their activity-dependent plasticity. The instability creates biases in the previously symmetric or previously balanced weight profiles, driving pattern flow in the absence of velocity inputs. The added neural adaptation also works against faithful integration. Thus, it remains to be seen whether GCNs with periodic structure are plausible.

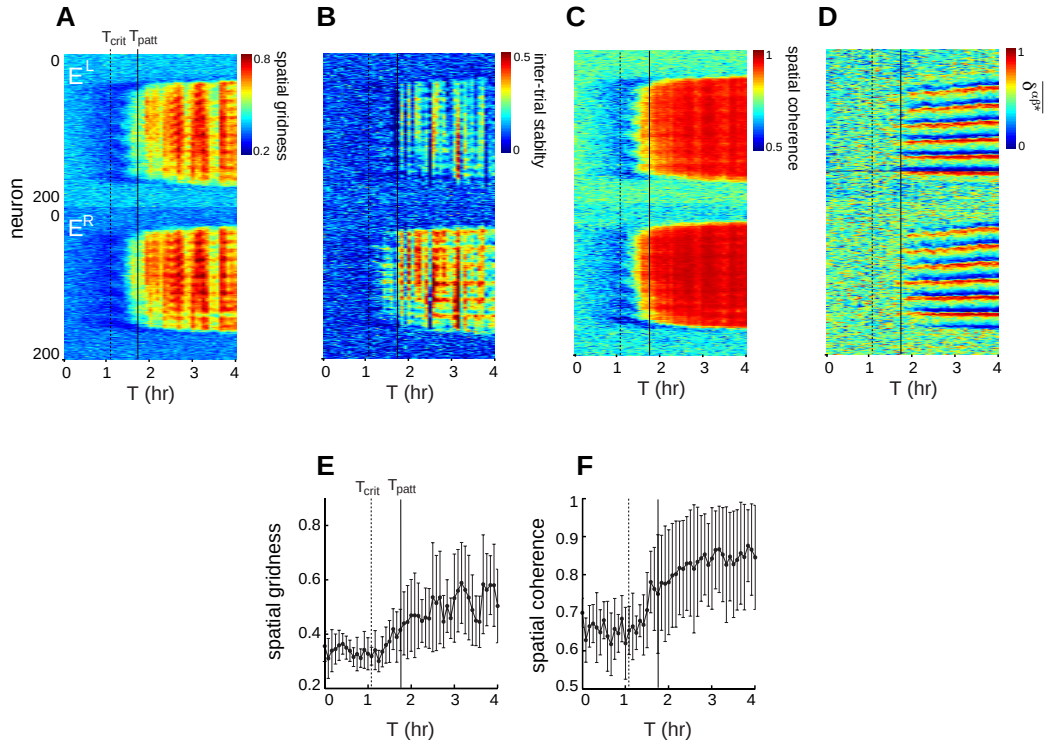

Figure A.5: **Supplemental figures associated with Figure 2.5.** (A-D) *Developmental scores for the entire E population.* (A) Spatial gridness, (B) inter-trial stability, (C) spatial coherence, and (D) relative phase, $\delta^{\alpha\beta_*}$ (β^* indicates fixed reference cell, which is marked by the dark horizontal line in the figure, see Appendix A), as a function of cell number in the E population (y-axis) and development time (x-axis). Each element in a matrix is the score for a particular cell at a particular development time point, averaged over 10 trials (trials are 10-second, random trajectories; the same set of random trajectories are used to test the GCN at each developmental time point). (E-F) *Developmental scores obtained from a small number of trials and neurons have poorer resolution in assessing the onset of pattern formation.* The development of (E) spatial gridness and (F) spatial coherence, averaged across randomly chosen cells in the E population (n = 15; drawn from the pool of cells with gridness score > 0.5) and over n = 4 trials. Vertical lines at same times as in Figure 2.5A. Compare with Figure 2.5.

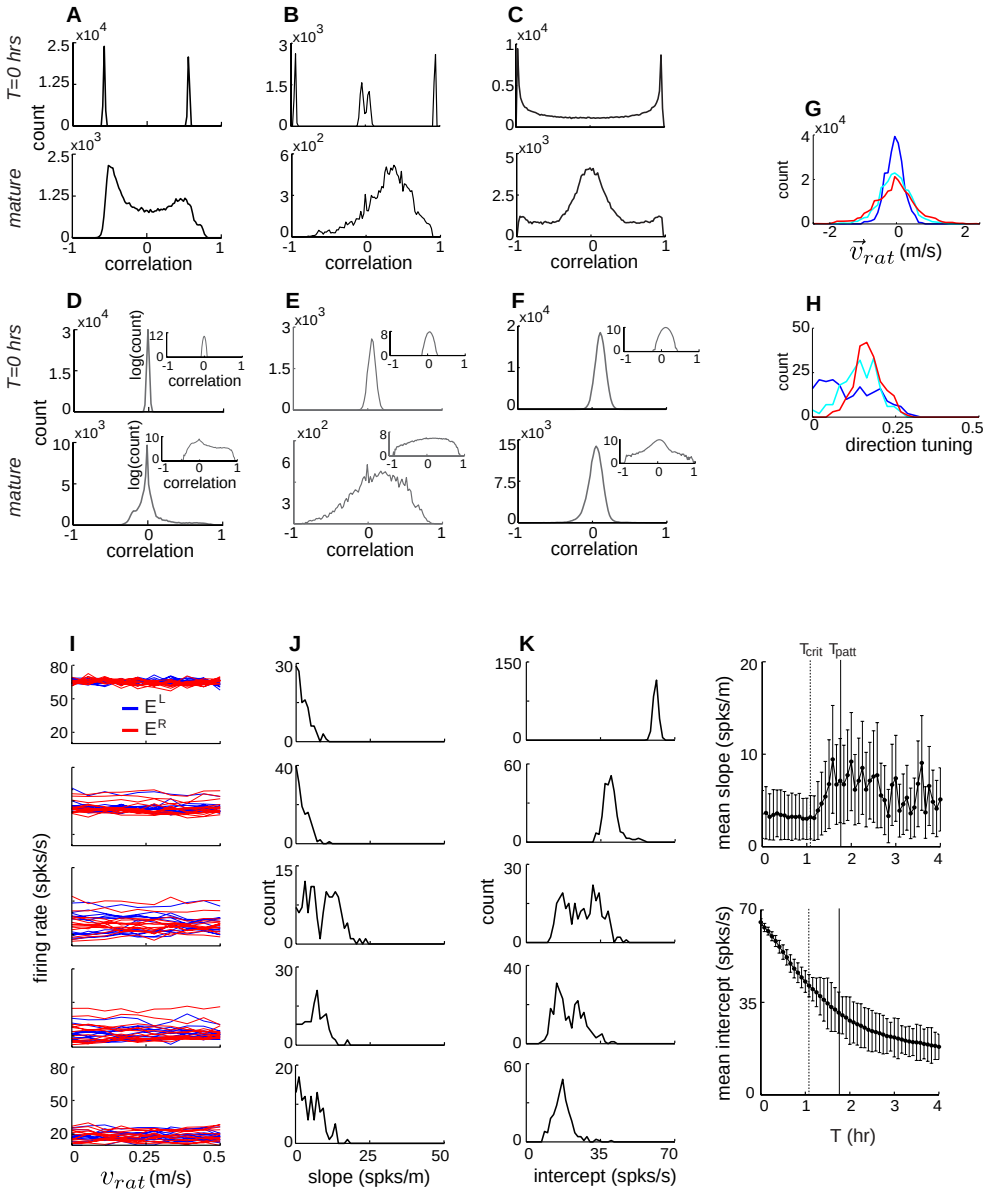

Figure A.6: **Supplemental figures associated with Figure 2.6.** (A-F) *The development of pairwise correlations between excitatory neurons in 1D and 2D GCNs.* The distribution of pairwise correlations in: left column: the 1D GCN; middle column: the 2D GCN with 4 different directionally tuned

populations; right column: the 2D GCN with omnidirectional tuning preferences. (A-C) *In vivo* correlations, in the presence of velocity inputs. (A) The immature 1D GCN (top) exhibits strong correlations and anticorrelations in the form of two sharp peaks, because of the common, dominant velocity inputs to the two populations. The mature 1D GCN (bottom; T = 4 hours) has gained a clear uniform component. (B) The immature 2D GCN with 4 directional populations exhibits three peaks (top): strong velocity-driven correlations within-population, strong anticorrelations between opposing direction populations, and zero correlations for orthogonal direction populations. The mature GCN (bottom) has gained a uniform component. (C) The immature 2D GCN with arbitrary direction preference in individual neurons already exhibits a uniform component (top), because of the evenly distributed velocity-driven direction preferences. In the mature GCN(bottom; after completion of the training trajectory), the velocity-driven components have shrunk and the uniform component has grown, but the gain in the uniform component in this GCN is quantitative rather than qualitative. (D-F) The same correlations as in (A-C), but in a simulated *in vitro* condition, in which the velocity inputs are removed. Here, the velocity-induced correlations have vanished, and all immature GCNs simply display a peak at zero (top plots). Similarly, all the mature GCNs display a uniform component related to pattern formation, that is uncomplicated by velocity correlations (bottom plots). The uniform component does not extend as far toward ± 1 as in the *in vivo* conditions, because in the absence of velocity drive, the pattern can only translate based on random fluctuations, and does not uniformly translate over all phases. [Correlation distributions are computed from a 2-minute long trajectory in which the animal executes movements (*in vivo*) (in 2D, the trajectory is experimentally derived, see Methods; in 1D, the trajectory is random, see Appendix A), or over trajectory segments where speed is lower than a threshold (boxed insets), or when the velocity input is zero (*in vitro*). Correlations in (A,D) are measured between E cells whose 1D spatial gridness scores in the mature GCN is greater than 0.5 (see Appendix A); correlations in (B,E) are measured between E cells whose 2D spatial gridness score (see Appendix A) in the mature GCN exceeds 0; correlations in (C,F) are measured between cells drawn randomly from the E population.] (G-H) *Mean direction tuning score as a measure is not independent of the statistics of the trajectory used to assess it.* (G) Distribution of velocities for three different 2-minute 1D trajectories. (H) The distribution of direction tuning scores measured in the mature GCN for the three trajectories

in (G) changes substantially with trajectory. Direction tuning scores should therefore only be compared across experiments with caution. (I-K) *Modulation of firing rate by animal speed, as computed in [118].* (I) Sample speed tuning curves, binned according the animal's speed (movement in the preferred and anti-preferred directions are conflated), for different cells in the E population, color-coded according to the cell's direction preference. Contrast with Figure 2.6C (see Appendix A). Cells and trajectory used to generate spike trains same as in Figure 2.6C. (J) Snapshots of the distribution of the slopes (left column) and intercepts (right column) of the regression lines used to fit speed tuning curves for the cells in the E population. Computed from same cells and trajectory as in Figure 2.6D. (K) Development of the mean of the speed tuning slope (top) and intercepts (bottom). *Interpretation of results in* (K): In [118], it is reported that the mean slopes and intercepts decrease with development. This is true of our data if considered from point T_{patt} onwards; after T_{patt}, spatial tuning is still evolving to become grid-like, and the mean slope and intercept do decrease with development (Figure 2.5B-D).

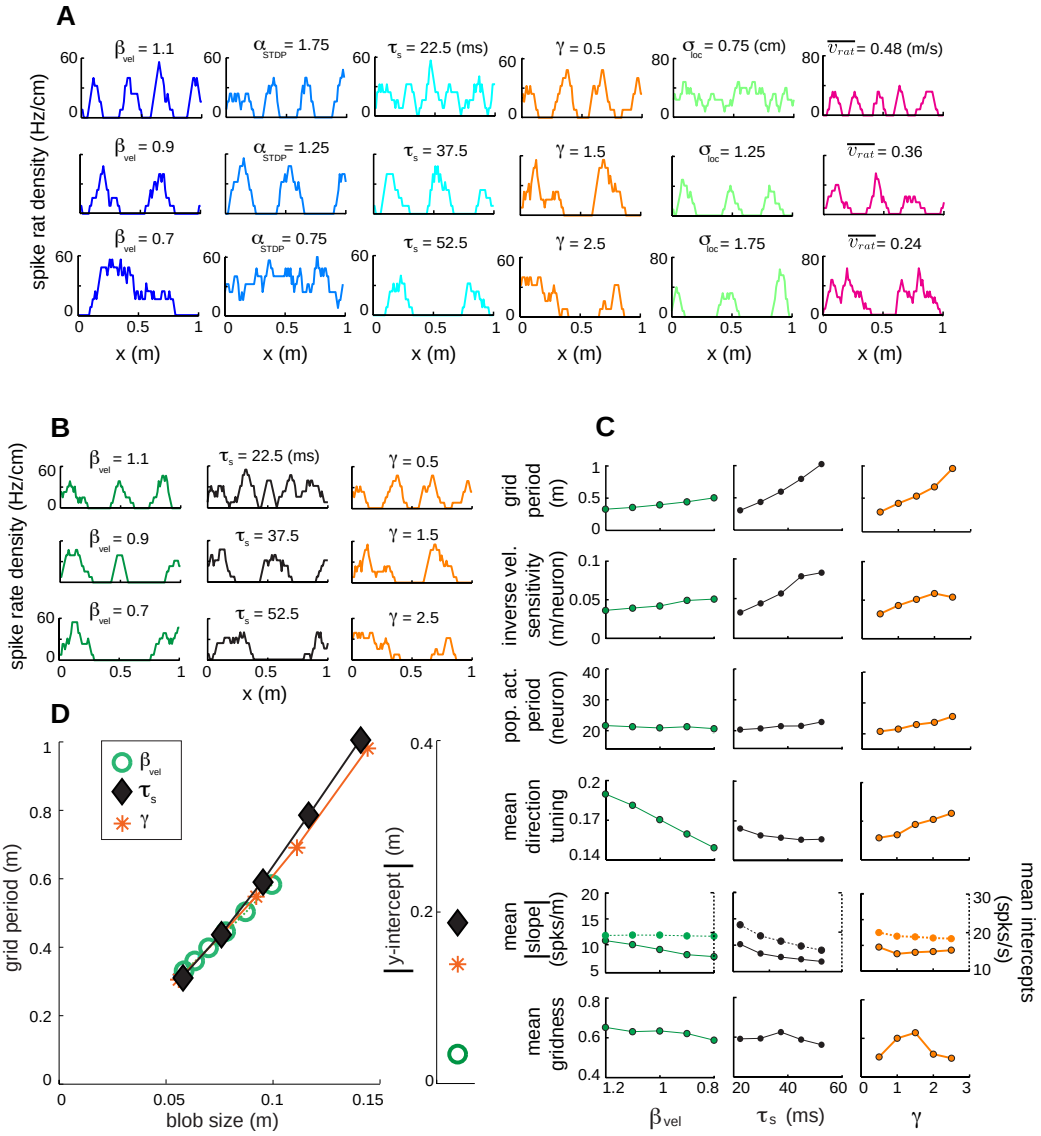

Figure A.7: **Supplemental figures associated with Figure 2.7.** (A) Sample spatial tuning curves for GCNs developed with different parameters, color-coded according to Figure 2.7. (B-D) *Rapid rescaling of grid cell tuning: possible mechanisms.* (B) Columns: The effects of varying individual parameters in the mature GCN. The gain of the velocity input (β_{vel}; column 1), the biophysical time-constant of neurons (τ_s; column 2), and the strength of inhibition (γ;

column 3) are varied individually (see Methods). (C) Systematic variation in properties of the GCN in response to changing parameters in the mature GCN. Rows: Mean grid period; inverse GCN velocity sensitivity (see Appendix A); population pattern period; average direction tuning score; average of the magnitudes of slopes (solid lines) and intercepts (dotted lines) of the regression lines that fit the speed tuning data; mean gridness score. (D) Relationship between blob size and grid period, as different parameters are individually varied.

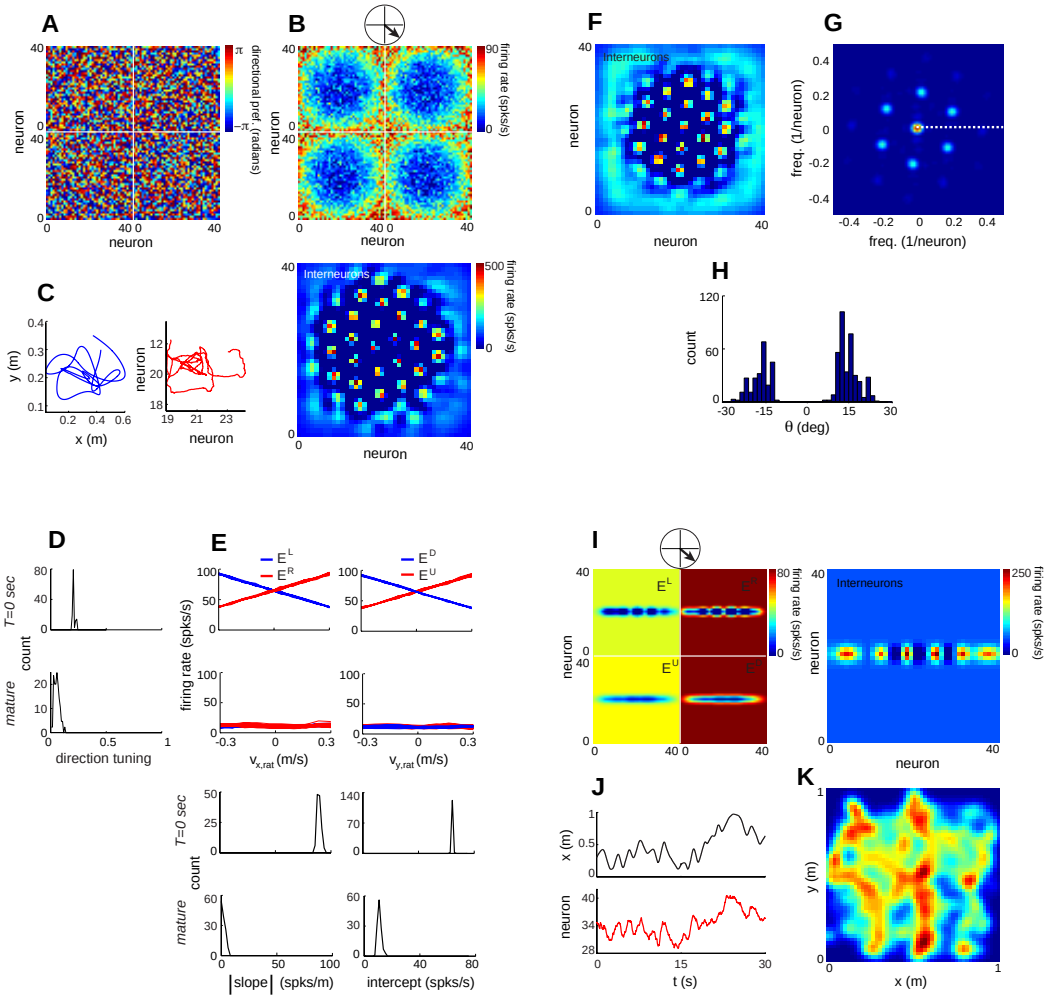

Figure A.8: **Supplemental figures associated with Figure 2.8.** (A-C) *2D GCN with randomly assigned, omnidirectional direction preferences.* (A) Direction preferences of the four excitatory populations (a distinction that is not meaningful now that a population does not correspond to a specific direction preference; we preserve a partitioning of cells into four groups simply to keep figures maximally similar for comparison). Each cell in each population is assigned a direction preference randomly drawn from the uniform distribution over the interval $[-\pi, \pi)$. (B) Snapshot of the population activities of the excitatory (top) and inhibitory (bottom) GCNs after training. Training trajectory and parameters identical to 2D GCN simulation details in Appendix

A. (C) Left: 30-second animal trajectory. Right: GCN estimate of location. Same parameters and training trajectory as described in Methods, except that $\beta_{vel} = 3.5$, $\gamma_{II} = 8.3$, $\gamma_{EI} = 0.25$, and $\gamma_{IE} = 20.8$. (D-E) *Development of direction and speed tuning statistics in the 2D grid cell GCN.* Statistics in the immature (at T = 0 s; top) and mature GCN (at the end of learning – see Methods; bottom) of Figure 2.8 from the main manuscript. Scores obtained from a 2-minute long trajectory in which the animal executes movements according to a recorded trajectory from a randomly foraging rat (see Methods). Data shown for cells with gridness score> 0 in the mature GCN (see Appendix A). (D) Direction score distribution. (E) Speed tuning. Top: Speed tuning curves sampled from the population of cells in (D) (speed is binned along the axis of the cell's preferred direction); the distribution of slopes (bottom left) and intercepts (bottom right) include all of the cells in (D). (F-H) *Systematic orienting of the 2D population activity pattern in the mature GCN.* (F) Snapshot of the inhibitory population activity pattern. (G) Power spectrum of the population activity pattern in (F). (H) Distribution of angles (measured with respect to the horizontal) of the two non-central peaks closest to the positive x-axis in the power spectrum, over several trials (n = 10) and over time within each trial (n=10 per trial). In each trial, the GCN activity is seeded with random initial conditions; as the population activity quickly becomes patterned, the GCN is driven by a random 10-second trajectory segment from a recorded animal trajectory (as in Figure 2.8). Angles are sampled every 0.5 seconds for the last 5 seconds of each trial). The distribution shows that *de novo* patterns form with a narrow range of orientations, and the patterns, once formed, rotate at most within a narrow range. The square-like GCN edges induce a particular orientation in the population pattern, directed away from the horizontal ($\approx \pm 15^{\circ}$). Intriguingly, grid cell responses in square environments, reported in [103]), tend to align to the square environment with approximately the same orientation ($\approx 7 - 18$ degrees). We wonder whether inhibitory border cells activated in the physical environment [104] somehow impose constraints on the allowable orientations in the population pattern by associating environmental boundaries with network edges, and thus impose a systematic orienting of the grid fields. (I-K) *2D GCN in animal confined to a narrow (1D) corridor during development.* (I) Snapshot of the population activities of the 2D GCN after development in a confined 1D environment (for example, a long, narrow track in which the external 2D world is not visible, where the animal can only move along a single line on the track, with textures and landmarks

to distinguish different parts of the track). The training trajectory consists of a sweep from left to right and back at speed $v_{rat} = 1$ m/s. (J) During random motion in an open 2D environment (30-second trajectory), the GCN is able to accurately integrate and thus track displacements along only one dimension of the animal's 2D trajectory. This dimension corresponds to displacements of the population activity pattern along its one extended dimension. Black line (top): x-coordinate of animal displacement. Red line (bottom): center of one of the activity bumps in bottom panel of (I). (K) Smoothed rate map of one cell during random navigation of the full 2D space (5 minutes). Note the banded structure of the spatial response. Same parameters and training trajectory as described in Methods, except that $\beta_{vel} = 3.5$, $\eta = 0.12$, $\gamma_{II} = 4$, $\gamma_{EI} = 0.25$, and $\gamma_{IE} = 21.6$.
Interpretation of results in (I-K): This result shows that if a GCN or a subpopulation within a GCN receives only 1D velocity inputs (even if the animal actually explores 2D space), that subpopulation will form stripe cells (resembling those seen in [70]). Raising an animal on non-Euclidean surfaces, for instance on a sphere (e.g. as done in [69]), should distort the relationship between velocity, time, and distance travelled and should lead to corresponding conformal distortions in GCN wiring and population patterning. These conformal distortions, as well as others that might arise from other unusual conditions during development (e.g. other non-Euclidean surfaces and radial-arm mazes), can result in complex changes in spatial tuning on flat 2D surfaces, that can be directly queried in our developmental model.

Appendix B

Cortical microcircuit determination through perturbation and sparse sampling in grid cells

B.1 Neural network simulations.

Below, we describe the two different neural dynamics models used in the paper: the linear-nonlinear-Poisson (LNP) model and the Hodgkin-Huxley conductance model.

Roman subscripts (e.g. i, j) refer to individual cells within population P. The population index P can take the values {I, E^R, E^L}. Integration in all simulations is by the Euler method with time-step dt.

B.1.1 Linear-Nonlinear-Poisson dynamics (all figures except Figure B.5).

The LNP model we use is identical to that used in [116]. Given a time-dependent summed input G_i^P to the (P, i)th cell, the instantaneous firing rate of the cell is

$$r_i^P = f(G_i^P), \tag{B.1}$$

with the neural transfer function f given by

$$f(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0. \end{cases} \tag{B.2}$$

Based on this time-varying firing rate, neurons fire spikes according to an inhomogeneous (sub-Poisson) point process with a coefficient of variance of CV = 0.5 (see [16] and [116] for details on generating a sub-Poisson point process).

The time-dependent activation s_i^P of synapses from the (P,i)th cell is given by

$$\frac{ds_i^P}{dt} + \frac{s_i^P}{\tau_{syn}} = \sum_b \delta(t - t_{i,b}^P), \tag{B.3}$$

where $t_{i,b}^P$ specifies the time of the bth spike of the cell and the sum is over all spikes of the cell.

The total input $G_i^P(t)$ into the (P,i)th cell is given by

$$G_i^P = A_i^P \alpha^{P,vel}(g_i^{P,rec} + g^0) + A_i^P g^{0'}, \tag{B.4}$$

where g^0 (g^0=50 for the E and I populations) and $g^{0'}$ ($g^{0'}$=15 for the E population; $g^{0'}$=0 for the I population) are small, positive, constant bias terms common to all cells, $g_i^{P,rec}$ are the recurrent inputs, $\alpha^{P,vel}$ are the velocity inputs, and A_i^P is an envelope that either suppresses activity near the network boundaries for the *aperiodic* network, or is flat and equal to unity for the *periodic* networks (see below). The recurrent input is

$$g_i^{P,rec} = \sum_{P'} \sum_{j=1}^{N^{P'}} W_{ij}^{PP'} s_j^{P'}, \tag{B.5}$$

where $W_{ij}^{PP'}$ are the recurrent weights and δ is the Kronecker delta function. The form of the envelope function, A_i^P, depends on the boundary conditions

of the network. For *aperiodic* networks, the envelope shape is a 1D version as that given by [16]:

$$A_i^{P,aper} = \begin{cases} 1 & r_i^P < \kappa N_P \\ \exp\left[-a_0\left(\frac{r_i^P - \kappa N_P}{(1-\kappa)N_P}\right)^2\right] & \text{otherwise} \end{cases} \tag{B.6}$$

where N_P is the size of the network, $r_i^P = |i - \frac{N_P}{2}|$, $\kappa = 0.3$ determines the range over which tapering occurs, and $a_0 = 30$ controls the steepness of the tapering. For *periodic* networks, the envelope is flat:

$$A_i^{P,per} = 1 \tag{B.7}$$

All cells in the Pth population (with preferred direction given by the unit vector $\hat{e}^P$) receive a common velocity input:

$$\alpha^{P,vel} = 1 + \beta^{vel}\vec{v} \cdot \hat{e}^P, \tag{B.8}$$

where $\vec{v}$ is instantaneous velocity of the animal and β^{vel} sets the gain of the velocity input; $\hat{e}^P$ = (0,0), (0,1), (0,-1) for the I, E^R, E^L populations, respectively. The velocity input, unless otherwise noted, is based on a 2-minute quasi-random trajectory derived with an algorithm identical to that described in [116]. Over the course of these trajectories, the stochastic dynamics leads to drift in the path-integrated estimate of animal location if uncorrected. To minimize this drift, the pattern phase is reset whenever the animal is in the vicinity of one of the 5 'landmarks' evenly spaced throughout the environment. During each encounter with a landmark, the pattern phase is corrected via strong feedforward inputs that impose a snapshot of the pattern at its

"correct" phase; "correct" pattern snapshots are captured from the population pattern during the animal's initial encounter with each of the landmarks.

Temperature/neuromodulation of LNP dynamics. Temperature-dependent modulations are modelled as a simple rescaling of the synaptic activation time constant, τ_{syn}. Modulations of network inhibition are modelled as a gain change in the efficacy of the synaptic weights projecting from inhibitory neurons, i.e., $W^{PI} \leftarrow \gamma_{inh} W^{PI}$, where γ_{inh} is the strength of inhibition.

B.1.2 Hodgkin-Huxley dynamics (only used in Figure B.5).

The model we use is identical to the reduced Hodgkin-Huxley "regular spiking (RS)" model of cortical neurons, as described in [90], supplemented with synaptic dynamics. The dynamics of the membrane potential of the (P, i)th neuron is given as

$$C_m \frac{dV_i^P}{dt} = -I_i^P, \tag{B.9}$$

where I_i^P is the summed input current and C_m is the capacitance of the membrane. The summed input current is given as

$$I_i^P = +\alpha^{P,vel} A_i^P \left(I_i^{P,rec} - I_{app} \right), \tag{B.10}$$

where the first term represents currents related to the ionic membrane conductances and the second and third terms represents synaptic and external conductances, respectively, gated by velocity inputs, $\alpha^{P,vel}$, and an envelope

function, A_i^P. The ionic current has the following form:

$$I_i^{P,ion} = \overline{g}_L(V_i^P - \overline{V}_L) + \overline{g}_K n^4(V_i^P - \overline{V}_K) + \overline{g}_M q(V_i^P - \overline{V}_K) + \overline{g}_{Na} m^3 h(V_i^P - \overline{V}_{Na}), \tag{B.11}$$

where the $\overline{g}$'s are the maximum conductance values and the $\overline{V}$'s the reversal potentials of the leak conductance (L), fast (K) and slow (M) potassium conductances, and the sodium conductance (Na). The dynamics and parameter settings of the gating variables n, m, q, h are described in [90] (note that we have replaced the "p" gating variable of [90] with "q"). The synaptic current based on recurrent connections within network is

$$I_i^{P,rec} = \overline{\eta} \sum_{P'} \sum_{j}^{N^{P'}} W_{ij}^{PP'} s_j^{P'} (V_i^P - \overline{V}^P), \tag{B.12}$$

where $s_j^{P'}$ is the synaptic activation of the (P', j) neuron (which has the same dynamics as described above in equation (3) – here, we define the time of a spike elicited by the jth neuron, t_j^{spk}, as when the voltage moves from below 0 mV to above it in a single-time step, within the interval $(t, t + \Delta t)$), $\overline{\eta}$ is a synaptic scaling factor shared by all synaptic weights, and $\overline{V}$ is the synapse-specific reversal potential ($\overline{V}^E$=0 mV and $\overline{V}^I$=-80 mV).

Temperature/neuromodulation of HH dynamics. To simulate temperature-dependent modifications, we used separate Q_{10} factors to modulate the time constant ($Q_{10}^{\tau} = 3$) and amplitudes ($Q_{10}^{a} = 1.3$) of the ionic/synaptic conductances [57, 64]. At temperature T (°C), the conductance amplitudes $\overline{g}(T)$ and

time constants $\tau(T)$ have the following form:

$$\overline{g}(T) \leftarrow \overline{g}(T_0)(Q_{10}^a)^{\frac{T-T_0}{10}} \tag{B.13}$$

$$\tau(T) \leftarrow \tau(T_0)/(Q_{10}^\tau)^{\frac{T-T_0}{10}}. \tag{B.14}$$

The conductance amplitude modulation was applied specifically to $\overline{g}_L, \overline{g}_K, \overline{g}_M$, $\overline{g}_{Na}, W_{ij}^{PP'}$. The conductance time constant modulation was applied to the gating variable time constants $\tau_n, \tau_q, \tau_m, \tau_h$ (for gating variable x, the time constant τ_x is defined as $\tau_x = 1/(\alpha_x + \beta_x)$, where α_x and β_x are the rate constants governing the gating variable's dynamics – see [90]) and the synaptic time constant τ_{syn}. For temperature perturbations of the ionic conductances only, $\overline{g}_L, \overline{g}_K, \overline{g}_M, \overline{g}_{Na}, \tau_n, \tau_q, \tau_m, \tau_h$ change with temperature, while $W_{ij}^{PP'}$ and τ_{syn} =16 ms are held constant. For temperature perturbations of the synaptic conductances only, $W_{ij}^{PP'}$ and τ_{syn} change with temperature, while the ionic conductance properties are held fixed.

The effects of specific neuromodulators targeting the inhibitory synapses was modelled in exactly the same as for the LNP model.

B.1.3 Synaptic weights for network of excitatory and inhibitory neurons (all figures except Figure B.8).

The detailed synaptic weights used in the simulations are based on the developmentally-inspired hardwired weights with aperiodic boundary conditions described in the Appendix A and the SI of [116], and therefore can be viewed as the plausible culmination of a developmental process. Compared to the LNP-based model used in [16], we chose to implement the model in [116]

because it is more realistic, incorporating separate populations of excitatory and inhibitory cells; however, both models give qualitatively similar results. (Note that while the description of the weights below is different than that specified in [116] in order to enable flexibility in setting the network boundary conditions, the weights used for the aperiodic network are identical to those specified in [116].) The weights from population P' to P, between cells i and j, are described as follows:

$$W_{ij}^{PP'} = \frac{\eta}{\rho} A_{ij}^{PP'} \Theta(c_0 - \delta\rho) \left[\Theta(-\mu x) + \Theta(\mu(x - \mu N_P/2))\right] \times \left[\exp\left(\frac{-c_-^2}{2(\sigma\rho)^2}\right) + \epsilon \exp\left(\frac{-c_+^2}{2(\sigma\rho)^2}\right)\right], \tag{B.15}$$

where $x = i - \gamma j$, $\gamma = \frac{N_P}{N_{P'}}$ (N_P is the size of population P), Θ is the Heaviside function ($\Theta(x) = 0$ for $x < 0$ and is 1 otherwise), $c_0 = \psi(x)$ and $c_\pm = \psi(x \pm \Delta\rho)$ where $\psi(x) = \min\left(N_p - |x \mod N_p|, |x \mod N_p|\right)$, and $A_{ij}^{PP'} = A_i^P A_j^{P'}$, where $A_i^P = A_i^{P,aper}$ for *aperiodic* networks and $A_i^P = A_i^{P,per}$ for *periodic* networks (see above for definition of $A_i^{P,per}$ and $A_i^{P,aper}$). ρ is a scale factor that controls the width of the synaptic weights, and therefore the number of bumps expressed in the pattern, whereas η is a synaptic scaling factor that modulates only the amplitudes.

(A note on terminology: the *partially periodic* network has an overall topology that resembles the *periodic* network of [16]. In our usage in the present work, periodic refers to a fully periodic network, in which the periodicity of connections matches that of activity pattern, whereas in the partially periodic network, the bulk of connectivity does not reflect the periodicity of the population activity pattern.)

B.1.4 Simulation parameters

Aperiodic network with LNP dynamics.

$P = E_L, E_R, I$; $N_{E_L} = N_{E_R} = 400$ neurons; $N_I = 160$ neurons; CV = 0.5; $dt =$ 0.5 ms; $\tau_{syn} = 30$ ms*; $\beta^{vel} = 2$; $A_i^P = A_i^{P,aper}$; $\rho = 2.2$. $\gamma_{inh} = 1$*;

$E^L \rightarrow I$ (i.e., W^{IE_L}): $\eta = 21$; $\epsilon = 0$; Δ = -1; $\sigma = 2$; $\mu = 0$; $\delta = 0$;

$E^R \rightarrow I$: $\eta = 21$; $\epsilon = 0$; $\Delta = 1$; $\sigma = 2$; $\mu = 0$; $\delta = 0$;

$I \rightarrow E^L$: $\eta = 8$; $\epsilon = 0$; $\Delta = 4$; $\sigma = 5$; μ = -1; $\delta = 3$;

$I \rightarrow E^R$: $\eta = 8$; $\epsilon = 0$; Δ = -4; $\sigma = 5$; $\mu = 1$; $\delta = 3$;

$I \rightarrow I$: $\eta = 24$; $\epsilon = 1$; $\Delta = 2$; $\sigma = 3$; $\mu = 0$; $\delta = 3$;

(* indicates that parameters can change through perturbation)

Partially periodic network with LNP dynamics.

Same parameters as aperiodic network, except that $A_i^P = A_i^{P,per}$, and $\rho = 2.2$.

Fully periodic network with LNP dynamics.

Same parameters as aperiodic network, except that $A_i^P = A_i^{P,per}$, and $\rho = 22$.

Aperiodic network with HH dynamics.

All ionic conductance parameters are identical to those described in [90] for the RS model; as noted there, the parameters are set to values corresponding to a temperature of $T_0 = 36°C$. $N_{E_L} = N_{E_R} = 400$ neurons; $N_I = 160$ neurons; $dt = 0.025$ ms; $\tau_{syn} = 15$ ms*; $\beta^{vel} = 0.8$; $C_m = 1$ μF/cm^2; $\overline{g}_L = 0.1$ ms/cm^2*; $\overline{g}_K = 5$ ms/cm^2*; $\overline{g}_M = 0.07$ ms/cm^2*; $\overline{g}_{Na} = 50$ ms/cm^2*; $\overline{V}_L$ = -70 mV; $\overline{V}_K$ = -90 mV; $\overline{V}_{Na} = 50$ mV; $I_{app} = = 3$ μA/cm^2; $\overline{\eta} = 0.0015$; $A_i^P = A_i^{P,aper}$;

$\rho = 2.2$. $\gamma_{inh} = 1^*$; Synaptic weights are identical to those described for the aperiodic network with LNP dynamics up to a constant, $\overline{\eta}$. (* indicates that parameters can change through perturbation)

B.2 Scores and Measures

Bootstrap resampling and phase uncertainty. Given an original spike map of M total spikes (with locations) from one cell, we created a new spike map of N ($N < M$) total spikes, by picking spikes (with their corresponding location coordinates) from the original map one at a time, at random, and with replacement. The same was done for a second, simultaneously recorded cell. From these sampled spike trains for a pair of cells, we estimated relative phase (by computing the location of the peak closest to the origin in the cross-correlation of the spatial maps of the two cells, as in [125]). The procedure was performed 100 times, generating 100 bootstrapped relative phase estimates per cell pair. Phase uncertainty was measured as the mean of the magnitudes of the bootstrapped relative phase estimates.

Spatial tuning curves. For a given cell and trajectory, we build a histogram of spike counts at each location (bin size = 1 cm), then normalize the count in each bin by the amount of time spent in it. The normalized histogram is smoothed by convolution with a boxcar filter (width = 5 bins) to yield a spatial tuning curve.

Spatial tuning period. The spatial tuning period is measured as the inverse of the spatial frequency with the highest peak in the power spectrum of the spatial tuning curve (excluding the peak at 0 frequency).

Population period. Given the last 500 snapshots (frames) of the population pattern from a given trial, the population period is measured as followed: For each frame, measure the inverse of the frequency with the highest peak in the power spectrum (as with the spatial tuning period) of the population pattern. The population period is the average of these estimates.

Velocity response. Velocity response is measured as the translation speed (neurons/sec) of the network pattern to fixed input velocity, computed by tracking the displacement of the pattern for 10 seconds, smoothing the resulting trajectory with an 4-second moving average filter, and then measuring the average speed of the middle-half of the trajectory.

2D relative phase. The displacement vector $\vec{d}$ is converted into a 2D phase $\vec{\delta}$ according to $\vec{\delta} = f(d_1^{proj}/\lambda_1 \bmod 1, d_2^{proj}/\lambda_2 \bmod 1)$, where $\vec{d}^{proj} = (d_1^{proj}, d_2^{proj})$ is the oblique projection of $\vec{d}$ onto the principal vectors $\lambda_1\hat{e}_1$ and $\lambda_2\hat{e}_2$, and

$$f(\vec{x}) = \begin{cases} (x_1 - 1, x_2 - 1) & \text{if } x_1 \geq 0.5 \text{ and } x_2 \geq 0.5 \\ (x_1 - 1, x_2) & \text{if } x_1 \geq 0.5 \text{ and } x_2 < 0.5 \\ (x_1, x_2 - 1) & \text{if } x_1 < 0.5 \text{ and } x_2 \geq 0.5 \\ (x_1, x_2) & \text{if } x_1 < 0.5 \text{ and } x_2 < 0.5. \end{cases} \tag{B.16}$$

B.3 Supplemental figures

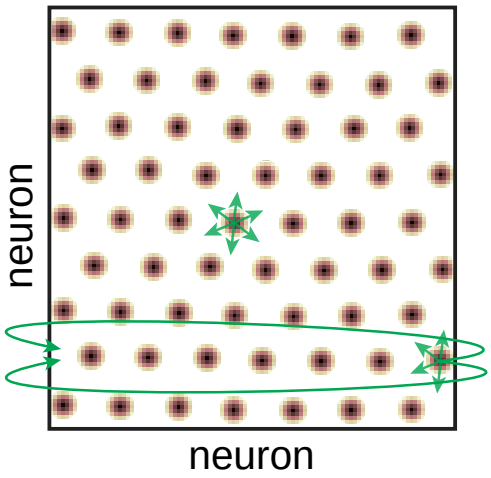

Figure B.1: **The *a priori* theoretical implausibility of partially periodic networks.** Population activity in the cortical sheet (yellow-black blobs), with schematic of connectivity (green). Note that in the bulk of the sheet, connectivity is local and not determined by the periodic activity in the sheet. However, the imposition of periodic boundary conditions requires that some neurons connect with others on the far edge of the sheet. Even if neurons are not topographically organized, the connectivity requires that a planar cortical sheet is somehow intrinsically connected as a torus. Activity-dependent weight changes that are based on the expression of periodic activity patterns could produce a torus-like connectivity, but then if the sheet is not topographically ordered it is likely that neurons in various bumps will connect to each other, producing a fully periodic rather than partially periodic network (see also Figure B.8).

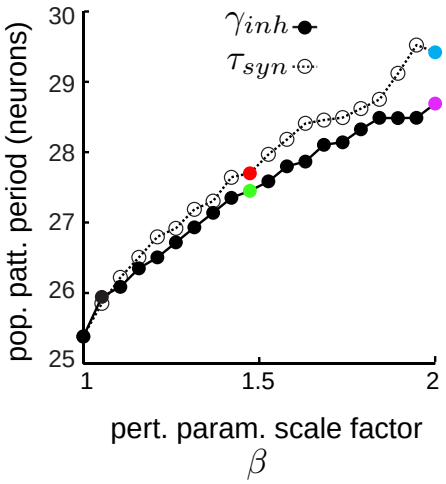

Figure B.2: **Dynamical simulations of the aperiodic network with LNP dynamics: gradual change in population period and peak-to-peak spacing of the DRPS with perturbation strength.** Change in population pattern period as the inhibition strength (filled circles) or the time-constant (open circles) are scaled up by the factor β (inhibition strength $\gamma_{inh} = \beta\gamma^*_{inh}$ with $\gamma^*_{inh} = 1$; time-constant $\tau_{syn} = \beta\tau^*_{syn}$, with $\tau^*_{syn} = 30$ ms) in the 1D aperiodic grid cell neural network (see Methods for simulation details and Appendix B for definition of population period).

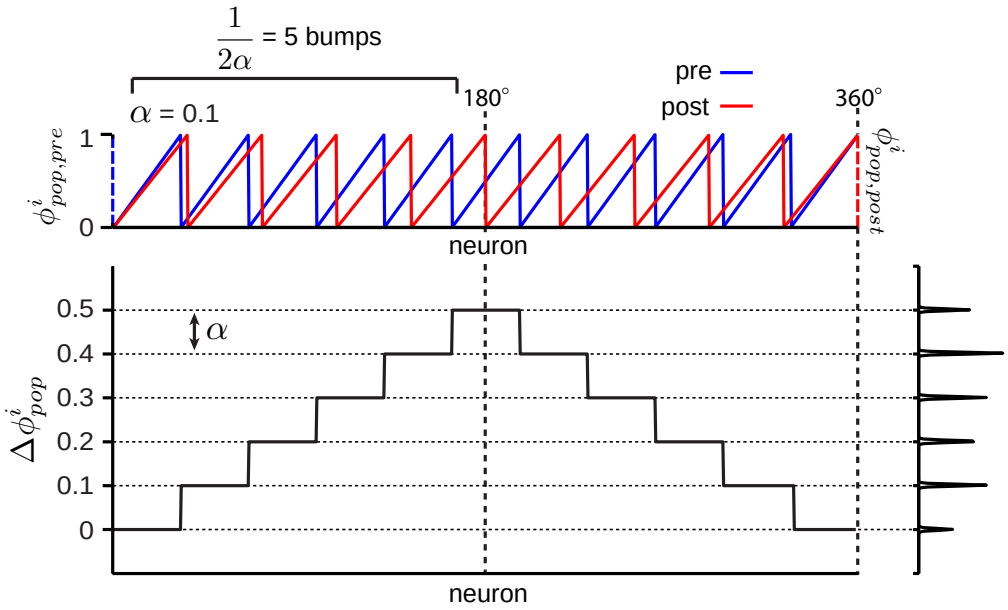

Figure B.3: **When the 2:1 relationship between number of peaks in the DRPS and the number of bumps in the population pattern breaks down.** Top: Schematic of the phase in a population pattern, pre- (blue) and post- (red) perturbation, for a large 1D network with many bumps. If the post-perturbation pattern is aligned to the first bump of the original pattern, the Mth bump is shifted by an amount $\lambda_{pop,pre}\alpha M$ away from the corresponding bump in the original pattern. When this shift equals $\lambda_{pop,pre}/2$, i.e. the perturbed bump is maximally out of phase with the original pattern, there can be no additional (farther out) quantal peaks in the DRPS. Thus, the number of DRPS peaks equals the number of bumps in the pattern only when $\lambda_{pop,pre}\alpha M < \lambda_{pop,pre}/2$, or equivalently, when $M\alpha < 1/2$. Bottom: Black curve: Difference in the pre- and scaled (by α) post-perturbation phases of cells. At right, the DRPS is aligned vertically with the y-axis of the phase shift plot, so that the origin of the DRPS peaks is more readily apparent. It is clear that once the two patterns reach counter-phase, the locations of DRPS peaks simply repeat. Thus, $M^* = 1/(2\alpha)$ bumps can accurately be discriminated (per linear dimension of the pattern) for a given stretch factor α; when $M > M^*$, the inference process suffers from systematic underestimation.

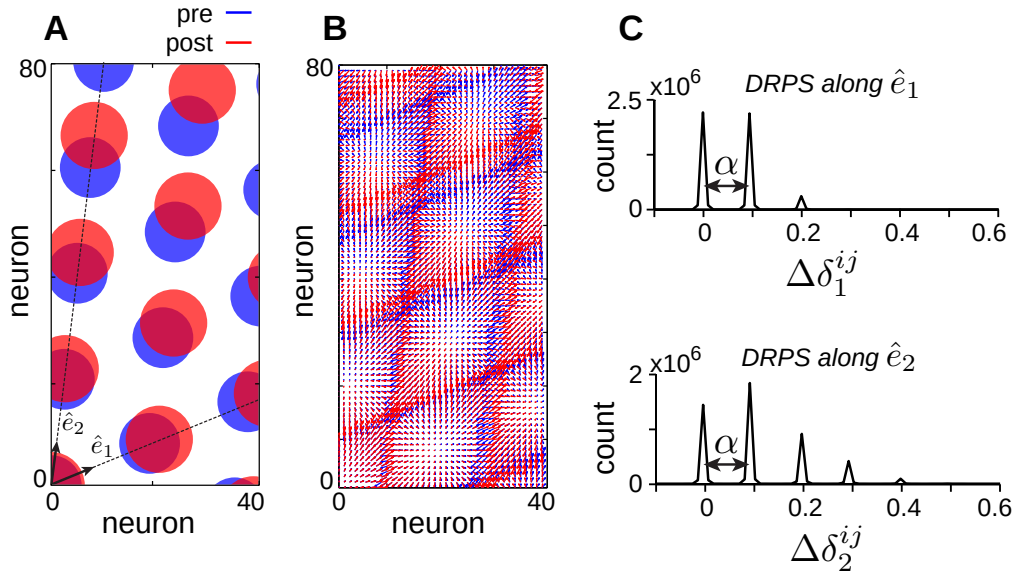

Figure B.4: **The DRPS in two dimensions.** (A) Schematic of 2D population activity pre- (blue) and post- (red) perturbation (not a dynamical neural network simulation). For illustration, the pattern is depicted topographically and pattern expansion is from bottom left. Subsequent predictions are independent of both choices. Dotted lines: The two principal axes of the pattern. (B) Population phase of each cell, depicted as an arrow (2D phase is a vector). (C) Two relative phase shift histograms computed separately for the two components of the vector phase, along the two principal axes of the lattice (gray: raw data; black: smoothed with 2-bin Gaussian). The DRPS for each phase component resembles the 1D DRPS. Data for each histogram: n = (3200 choose 2); bins = 200. *Parameters:* $\vec{\lambda}_{pop,pre} = \lambda_{pop,pre}(\hat{e}_1 + \hat{e}_2)$, $\lambda_{pop,pre}$ = 20 neurons, $\vec{\lambda}_{pop,post} = (1+\alpha)\vec{\lambda}_{pop,pre}$, $\alpha = 0.1$, $\hat{e}_1 = (\cos\theta, \sin\theta)$, $\theta = 23°$, $\hat{e}_2 = \hat{e}_1 + 60°$, network size: 80×40 neurons.

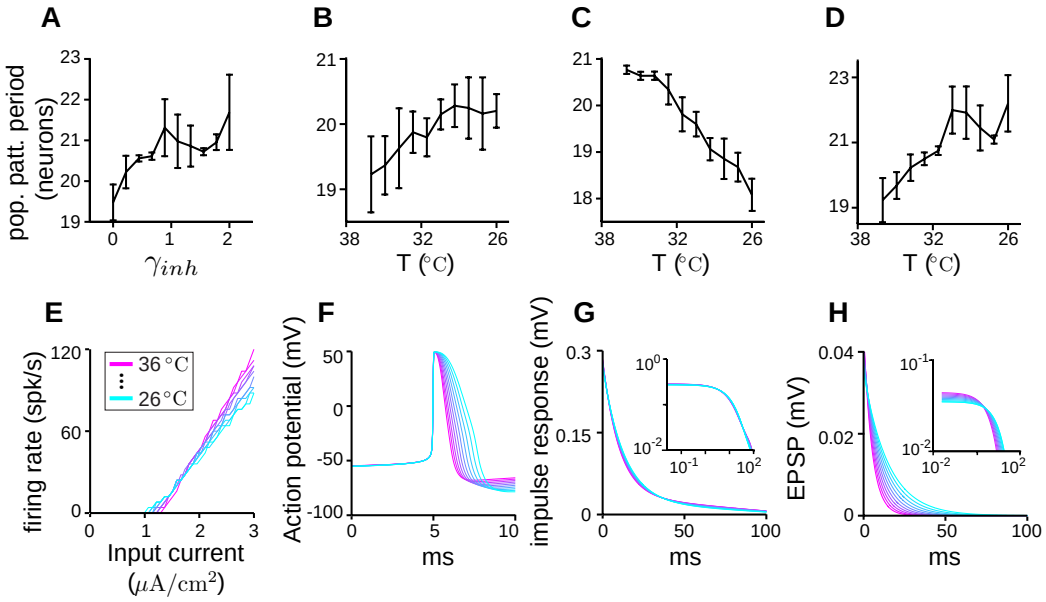

Figure B.5: **Dynamical simulations of the aperiodic network with HH dynamics: gradual change in population period.** (A-D) Population period as a function of perturbation strength (see Appendix B for description of HH dynamics and perturbation details). (A) Population period increases gradually with strength of inhibition. Each point is the population period averaged over 10 trials (with error bars equal to the standard deviation). Each trial is 1 second long, in which for the first third of the trial the population pattern is flowed with velocity input equal to 0.3 m/s, and for the remaining time allowed to relax with no velocity input (see Appendix B for definition of population period). (B) Population period increases as the temperature is stepped down from 36°C to 26°C. (C) Temperature perturbation of only the ionic conductances leads to decreasing population period with decreasing temperature. (D) Temperature perturbation of only the synaptic conductances leads to increasing population period with decreasing temperature. The net result of (C-D) is that the population period increases with decreasing temperature (B). (E-G) Temperature dependencies of single-cell properties, color-coded as a function of temperature. (E) Firing rate as a function of input current, (F) action potential shape, and (G) impulse response (i.e., subthreshold response of membrane potential to current pulse), with log-log plot in inset. (H) EPSP shape as a function of temperature. Though there is a slight decrease in am-

plitude of the EPSP with temperature (as shown by the log-log plot of the same data in the inset), it is small compared to the effect on the EPSP time constant.

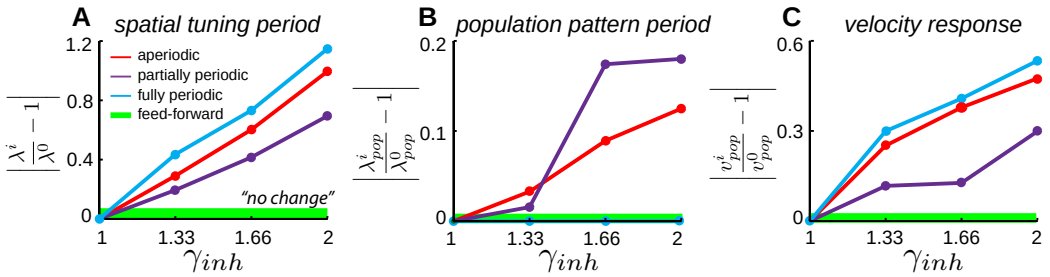

Figure B.6: **Changes in spatial tuning period in dynamical neural network simulations are due to changes in both the population period and the velocity response of the network.** (A) Spatial tuning periods (λ_x^i for different perturbation strengths indexed by i) of cells in the different recurrent networks, as the strength of inhibition is varied (data as in 3D). (B) The underlying period (λ^i_{pop}) of the population patterns in the same networks for the corresponding strengths of inhibition. (C) The velocity response (v^i_{pop}) of the networks, or efficacy with which a unit input velocity shifts the phase of the population pattern, as a function of inhibition strength. See Methods for simulation details, and Appendix B for definition of scores. It is clear that the spatial tuning period (A) is more strongly influenced by the velocity response (C) than by the population period (B).

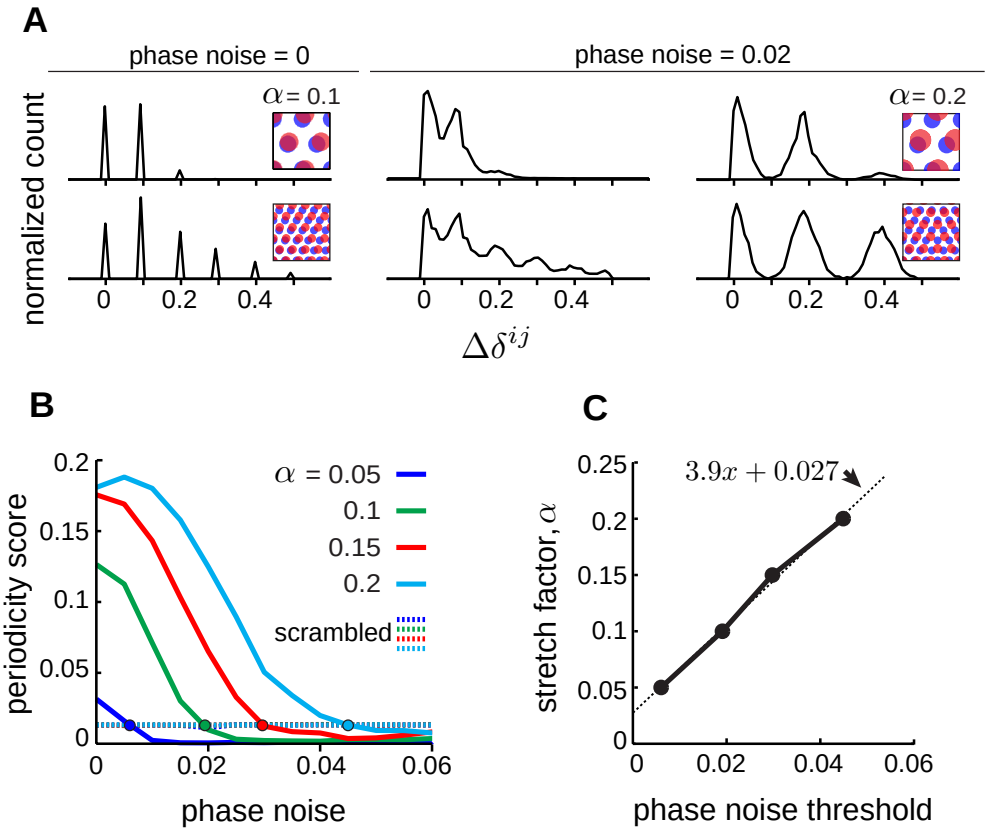

Figure B.7: **Effects of uncertainty in phase estimation.** (A) Copied from Figure 4B. First and second columns: DRPS (200 bins) for different numbers of population pattern bumps along the first principal axis of the pattern and for different amounts of phase noise (noise is sampled i.i.d. from a gaussian distribution, $\mathcal{N}(0, \sigma^2_{phase})$, and added to each component of the relative phase vector, $\vec{\delta}^{ij}$; "phase noise" is the same as σ_{phase}). Third column: Same as the second column, except for a larger stretch factor, $\alpha = 0.2$. Note that the peak-to-peak separation has increased so that the individual peaks are discernible. However, for the 5 bump network in the second row, inferring the number of bumps in the underlying population pattern would lead to an underestimate, since $M \times \alpha = 5 \times 0.2 > 1/2$. (B) Solid lines: Periodicity score (a measure of how well separated and equidistant are the peaks in the DRPS, and ranges between 0 and 1; see Appendix B) as a function of phase noise for

2-bump network in (A), for different values of the stretch factor, α (solid lines). Periodicity is measured for the DRPS along the first principal axis. Dashed lines: Same as solid lines, except computed by randomly shuffling the phase vectors post-perturbation. (C) Stretch factor, α, as a function of threshold phase noise (defined as the phase noise where the DRPS is indistinguishable from the DRPS when the phase vectors in the post-perturbation condition are reassigned randomly, i.e., the value of the phase noise when the colored curves in (B) cross the respective colored dashed lines).

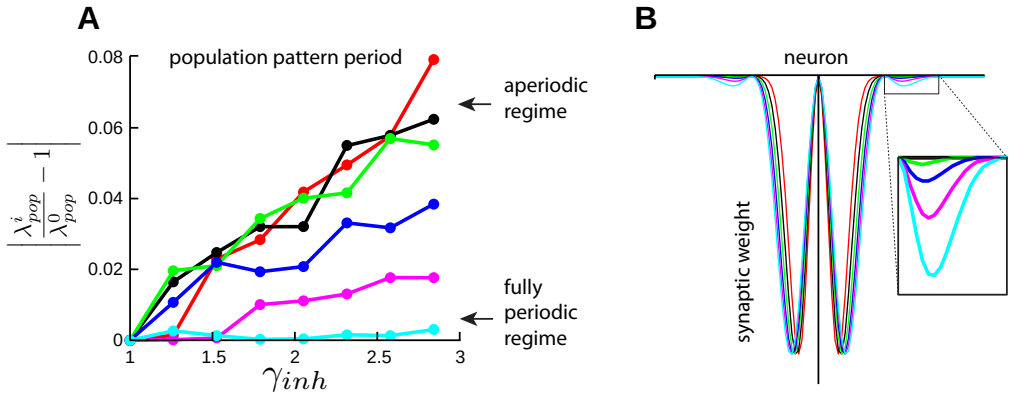

Figure B.8: **Weakly coupling neurons in different activity bumps in an aperiodic network results in behavior identical to single-bump fully periodic networks.** (A) The population pattern in the aperiodic network network exhibits a regime of continuous stretching (black curve) with increasing inhibition strength (γ_{inh}). The ordinate axis is the stretch-factor *alpha*, which quantifies the deviation of the period post-perturbation from that pre-perturbation, normalized by the pre-perturbation period (See Appendix B for definition of population period). However, adding even very weak synaptic connections between neurons in adjacent activity bumps in the aperiodic network (B) transforms the network into one that will not stretch at all (cyan curve), like the single-bump fully periodic network. The two constructions (the single bump network and the aperiodic network with the addition of between-bump connections) would be mathematically the same if there were strong coupling between all neurons of the same activity phase in the network; this numerical result shows that the addition of very small weights that reflect the periodicity of the population activity (but are nevertheless largely local within the network in the sense that only neighboring bump neurons are connected, not neurons in remote bumps) already transforms the aperiodic network into an effectively fully periodic, single-bump network. *Simulation details*: The network connectivity is as a hybrid of the aperiodic network in [16] with the fully periodic network of [38] (note that, while the model of [38] does not have explicit periodic boundary conditions, the multimodality of the synaptic weights couples adjacent activity bumps so that the network acts as a single-bump, fully periodic network). The dynamics are LNP-based (see Appendix B) and driven with inputs simulating animal motion at constant speed (v =

0.3 m/s) for 10 seconds. There are only two populations (call them R and L), distinct in their directional preferences ($\hat{e}^P$ = (0,1), (0,-1) for the R and L populations, respectively) and output synaptic asymmetries (see below). The shifted output weight profiles are sinusoids with gaussian envelopes, the latter which constrain the non-locality of the projections. For a narrow gaussian envelope, the weights resemble the purely local, center-surround profiles of [16], whereas for wide gaussian envelopes, the weights resemble the non-local, multimodal projections of [38]. The weights going from population P' to P and from cells i and j, are given by $W_{ij}^{PP'} = \frac{\eta}{C} \exp\left(\frac{-x^2}{2\sigma^2}\right)(\cos(ax) - 1)$, where $x = i - j + \Delta$ ($\Delta = \pm 1$ for $P' = R(+)$ and $P' = L(+)$), η is a scaling factor that modulates the amplitude of the weights, $C = \sqrt{2\pi\sigma^2}\left(\exp\left(\frac{-\sigma^2 a^2}{2}\right) - 1\right)$ is a normalization factor, σ determines the width of the gaussian envelope, and a determines the period of the underlying sinusoid.

Parameters. $N_R = N_L = 200$ neurons; CV = 0.5; $dt = 0.5$ ms; $\tau_{syn} = 30$ ms; $g^0 = 50$; $g^{0'} = 0$; $\beta^{vel} = 1$; $A_i^P = A_i^{P,aper}$; $a = 2\pi/20$; $\eta = 200$; $\sigma = 4 \rightarrow 12$;

Bibliography

[1] D. Baldauf and R. Desimone. Neural mechanisms of object-based attention. *Science*, 344:424–427, 2014.

[2] C. A. Barnes, B. L. McNaughton, S. J. Mizumori, B. W. Leonard, and L. H. Lin. Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing. *Prog Brain Res*, 83:287–300, 1990.

[3] C. Barry, L. L. Ginzberg, J. O'Keefe, and N. Burgess. Grid cell firing patterns signal environmental novelty by expansion. *Proc Natl Acad Sci U S A*, 109(43):17687–92, 2012.

[4] Caswell Barry, Robin Hayman, Neil Burgess, and Kathryn J Jeffery. Experience-dependent rescaling of entorhinal grids. *Nat Neurosci*, 10(6):682–4, Jun 2007.

[5] P. Beed, A. Gundlfinger, S. Schneiderbauer, J. Song, C. Bohm, A. Burgalossi, M. Brecht, I. Vida, and D. Schmitz. Inhibitory gradient along the dorsoventral axis in the medial entorhinal cortex. *Neuron*, 79:1197–1207, 2013.

[6] G Bi and M Poo. Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu. Rev. Neurosci.*, 24:139–166, 2001.

[7] T. L. Bjerknes, E. I. Moser, and M. B. Moser. Representation of geometric borders in the developing rat. *Neuron*, 82:1–8, 2014.

[8] 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):1239–1255, 2008.

[9] T. Bonnevie, B. Dunn, M. Fyhn, T. Hafting, D. Derdikman, J. L. Kubie, Y. Roudi, E. Moser, and M. Moser. Grid cells require excitatory drive from the hippocampus. *Nat Neurosci*, 16:309–317, Jan 2013.

[10] M. Brandon, A. Bogaard, C. Libby, M. Connerney, K. Gupta, and M. Hasselmo. Reduction of theta rhythm dissociates grid cell spatial periodicity from directional tuning. *Science*, 332(6029):595–9, Apr 2011.

[11] Mark P Brandon, Andrew R Bogaard, Chris M Andrews, and Michael E Hasselmo. Head direction cells in the postsubiculum do not show replay of prior waking sequences during sleep. *Hippocampus*, 22(3):604–618, Mar 2012.

[12] M. Brecht, S. Ray, A. Burgalossi, Q. Tang, H. Schmidt, and R. Naumann. An isomorphic mapping hypothesis of the grid representation. *Philos Trans R Soc Lond B Biol Sci*, 369, 2014.

[13] 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, Jan 2008.

[14] C. Buetfering, K. Allen, and H. Monyer. Parvalbumin interneurons provide grid cell-driven recurrent inhibition in the medial entorhinal cortex. *Nat Neurosci*, 17(5):710–8, May 2014.

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

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

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

[18] D. Bush and N. Burgess. A hybrid oscillatory interference/continuous attractor network model of grid cell firing. *J Neurosci*, 34(14):5065–5079, 2014.

[19] G. Buzsaki and E. I. Moser. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nat Neurosci*, 16:130–138, 2013.

[20] Cathrin B Canto, Noriko Koganezawa, Prateep Beed, Edvard I Moser, and Menno P Witter. All layers of medial entorhinal cortex receive

presubicular and parasubicular inputs. *J Neurosci*, 32(49):17620–31, Dec 2012.

[21] Cathrin B Canto, Floris G Wouterlood, and Menno P Witter. What does the anatomical organization of the entorhinal cortex tell us? *Neural Plast*, 2008:381243, 2008.

[22] P. Cisek and J. F. Kalaska. Neural mechanisms for interacting with a world full of action choices. *Annu Rev Neurosci*, 33:269–298, 2010.

[23] Laura Lee Colgin, Edvard I Moser, and May-Britt Moser. Understanding memory through hippocampal remapping. *Trends Neurosci*, 31(9):469–77, Sep 2008.

[24] T. A. Coogan and A. Burkhalter. Hierarchical organization of areas in rat visual cortex. *J Neurosci*, 13(9):3749–3772, 1993.

[25] J. J. Couey, A. Witoelar, S. Zhang, K. Zheng, J. Ye, B. Dunn, R. Czajkowski, M. Moser, E. Moser, Y. Roudi, and M. Witter. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nat Neurosci*, 16(3):318–24, Mar 2013.

[26] Dori Derdikman and Edvard I Moser. A manifold of spatial maps in the brain. *Trends Cogn Sci*, 14(12):561–9, Dec 2010.

[27] S. S. Deshmukh and J. J. Knierim. Influence of local objects on hippocampal representations: Landmark vectors and memory. *Hippocampus*, 23(4):253–267, Feb 2013.

[28] Sachin S Deshmukh and James J Knierim. Representation of non-spatial and spatial information in the lateral entorhinal cortex. *Front Behav Neurosci*, 5:1–33, 2011.

[29] A Dhillon and R S Jones. Laminar differences in recurrent excitatory transmission in the rat entorhinal cortex in vitro. *Neuroscience*, 99(3):413–22, 2000.

[30] J. J. DiCarlo, D. Zoccolan, and N. C. Rust. How does the brain solve visual object recognition. *Neuron*, 73:415–434, 2012.

[31] Christian F Doeller, Caswell Barry, and Neil Burgess. Evidence for grid cells in a human memory network. *Nature*, 463(7281):657–61, Feb 2010.

[32] C. Domnisoru, A. A. Kinkhabwala, and D. A. Tank. Membrane potential dynamics of grid cells. *Nature*, 495(7440):199–204, 2013.

[33] D. E. Feldman. The spike-timing dependence of plasticity. *Neuron*, 75:556–571, 2012.

[34] J. Ferbinteanu, R. M. Holsinger, and R. J. McDonald. Lesions of the medial or lateral perforant path have different effects on hippocampal contributions to place learning and on fear conditioning to context. *Behav Brain Res*, 101(1):65–84, May 1999.

[35] Ila R Fiete, Yoram Burak, and Ted Brookings. What grid cells convey about rat location. *J Neurosci*, 28(27):6858–6871, Jul 2008.

[36] Elodie Fino and Laurent Venance. Spike-timing dependent plasticity in the striatum. *Front Synaptic Neurosci*, 2:6, 2010.

[37] L. M. Frank, E. N. Brown, and M. Wilson. Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron*, 27:169–178, 2000.

[38] 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, Apr 2006.

[39] Marianne Fyhn, Torkel Hafting, Alessandro Treves, May-Britt Moser, and Edvard I Moser. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, 446(7132):190–4, Mar 2007.

[40] Marianne Fyhn, Torkel Hafting, Menno P Witter, Edvard I Moser, and May-Britt Moser. Grid cells in mice. *Hippocampus*, 18(12):1230–8, 2008.

[41] Marianne Fyhn, Sturla Molden, Menno P Witter, Edvard I Moser, and May-Britt Moser. Spatial representation in the entorhinal cortex. *Science*, 305(5688):1258–64, Aug 2004.

[42] Derek L F Garden, Paul D Dodson, Cian O'Donnell, Melanie D White, and Matthew F Nolan. Tuning of synaptic integration in the medial entorhinal cortex to the organization of grid cell firing fields. *Neuron*, 60(5):875–89, Dec 2008.

[43] L. M. Giocomo, T. Bonnevie, T. Stensola, T. Van Cauter, E. I. Moser, and M. B. Moser. Topographical organization of head direction signals in medial entorhinal cortex. 2013.

[44] L. M. Giocomo, T. Stensola, T. Bonnevie, T. Van Cauter, M. B. Moser, and E. I. Moser. Topography of head direction cells in medial entorhinal cortex. *Cell*, 24:252–262, 2014.

[45] Lisa M Giocomo, Eric A Zilli, Erik Fransén, and Michael E Hasselmo. Temporal frequency of subthreshold oscillations scales with entorhinal grid cell field spacing. *Science*, 315(5819):1719–1722, Mar 2007.

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

[47] Julie S Haas, Thomas Nowotny, and H D I Abarbanel. Spike-timing-dependent plasticity of inhibitory synapses in the entorhinal cortex. *J Neurophysiol*, 96(6):3305–13, Dec 2006.

[48] Torkel Hafting, Marianne Fyhn, Sturla Molden, May-Britt Moser, and Edvard I Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, Aug 2005.

[49] R H R Hahnloser. Emergence of neural integration in the head-direction system by visual supervision. *Neuroscience*, 120(3):877–891, 2003.

[50] M. E. Hasselmo, C. Bodelon, and B. P. Wyble. A proposed function of hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Comp.*, 14:793–817, 2002.

[51] Michael E Hasselmo and Mark P Brandon. A model combining oscillations and attractor dynamics for generation of grid cell firing. *Front Neural Circuits*, 6:30, 2012.

[52] 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):1252–71, 2007.

[53] D. J. Heeger. Normalization of cell responses in cat striate cortex. *Vis Neuroscience*, 9:181–197, 1992.

[54] J. Heys and D. Dombeck. Cellular resolution optical imaging of medial entorhinal cortex. In *SFN Poster*, 2014.

[55] G. E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Comp.*, 14(1771-1800), 2002.

[56] G. E. Hinton, P. Dayan, B. J. Frey, and R. M. Neal. The "wake-sleep" algorithm for unsupervised neural networks. *Science*, 268:1158–1161, 1995.

[57] A. L. Hodgkin, A. F. Huxley, and B. Katz. Measurement of current-voltage relatioships in the membrane of the giant axon of loligo. *J Physiol*, 116:424–448, 1952.

[58] Y. Holmgren, C. Zilberter. Coincident spiking activity induces long-term changes in inhibition of neocortical pyramidal cells. *J Neurosci*, 21(20):8270–8277, 2001.

[59] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J. P. Thiran, R. Meuli, and P. Hagmann. Predicting human resting-state functional connectivity from structural connectivity. *PNAS*, 106(6):2035–2040, 2009.

[60] Shiyong Huang, Richard L Huganir, and Alfredo Kirkwood. Adrenergic gating of hebbian spike-timing-dependent plasticity in cortical interneurons. *J Neurosci*, 33(32):13171–8, Aug 2013.

[61] 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, and Michael J Kahana. Direct recordings of grid-like neuronal activity in human spatial navigation. *Nat Neurosci*, 16(9):1188–90, Sep 2013.

[62] J. P. Johansen, C. K. Cain, L. E. Ostroff, and J. E. LeDoux. Molecular mechanisms of fear learning and memory. *Cell*, 147:509–524, 2011.

[63] E. R. Kandel, Y. Dudai, and M. R. Mayford. The molecular and systems biology of memory. *Cell*, 157:163–186, 2014.

[64] B. Katz and R. Miledi. The effect of temperature on the synaptic delay at the neuromuscular junction. *J Physiol*, 181:656–670, 1965.

[65] Kristin M Kerr, Kara L Agster, Sharon C Furtak, and Rebecca D Burwell. Functional neuroanatomy of the parahippocampal region: the lateral and medial entorhinal areas. *Hippocampus*, 17(9):697–708, 2007.

[66] Nathaniel J Killian, Michael J Jutras, and Elizabeth A Buffalo. A map of visual space in the primate entorhinal cortex. *Nature*, 491(7426):761–4, Nov 2012.

[67] S. Kirkpatrick, Jr. C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[68] Emilio Kropff and Alessandro Treves. The emergence of grid cells: Intelligent design or just adaptation? *Hippocampus*, 18(12):1256–69, 2008.

[69] I. U. Kruge, T. Wernle, E. I. Moser, and M. B. Moser. Grid cells of animals raised in spherical environments. In *SFN Poster*, 2013.

[70] Julija Krupic, Neil Burgess, and John O'Keefe. Neural representations of location composed of spatially periodic bands. *Science*, 337(6096):853–7, Aug 2012.

[71] 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, Jun 2010.

[72] S. Leutgeb, J. Leutgeb, A. Treves, M. Moser, and E. Moser. Distinct ensemble codes in hippocampal areas ca3 and ca1. *Science*, 305(5688):1295–8, Aug 2004.

[73] S. Leutgeb and J. K. Leutgeb. Pattern separation, pattern completion, and new neuronal codes within a continuous ca3 map. *Learn Mem*, 14(11):745–57, Nov 2007.

[74] Colin Lever, Stephen Burton, Ali Jeewajee, John O'Keefe, and Neil Burgess. Boundary vector cells in the subiculum of the hippocampal formation. *J Neurosci*, 29(31):9771–9777, Aug 2009.

[75] Jiang-teng Lu, Cheng-yu Li, Jian-Ping Zhao, Mu-ming Poo, and Xiaohui Zhang. Spike-timing-dependent plasticity of neocortical excitatory synapses on inhibitory interneurons depends on target cell type. *J Neurosci*, 27(36):9711–20, Sep 2007.

[76] B. L. McNaughton and C. A. Barnes. Physiological identification and analysis of dentate granule cell responses to stimulation of the medial and lateral perforant pathways in the rat. *J Comp Neurosci*, 175:439–453, 1977.

[77] 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, Aug 2006.

[78] Himanshu Mhatre, Anatoli Gorchetchnikov, and Stephen Grossberg. Grid cell hexagonal patterns formed by fast self-organized learning within entorhinal cortex. *Hippocampus*, 22:320–334, Dec 2012.

[79] M.-L. Mittelstaedt and H. Mittelstaedt. Homing by path integration in a mammal. *Naturwissenschaften*, 1980.

[80] E. I. Moser and P. Anderson. Conserved spatial learning in cooled rats in spite of slowing of dentate field potentials. *J Neurosci*, 14(7):4458–4466, 1994.

[81] G. M. Muir and D. K. Bilkey. Theta- and movement velocity-related firing of hippocampal neurons is disrupted by lesions centered on the perirhinal cortex. *Hippocampus*, 13(1):93–108, 2003.

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

[83] Zaneta Navratilova, Lisa M Giocomo, Jean-Marc Fellous, Michael E Hasselmo, and Bruce L McNaughton. Phase precession and variable spatial scaling in a periodic attractor map model of medial entorhinal grid cells with realistic after-spike dynamics. *Hippocampus*, 22(4):772–89, Apr 2012.

[84] J. OKeefe. Place units in the hippocampus of the freely moving rat. *Exp Neurol*, 51:78–109, 1976.

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

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

[87] C. Parron and E. Save. Comparison of the effects of entorhinal and retrosplenial cortical lesions on habituation, reaction to spatial and non-spatial changes during object exploration in the rat. *Neurobiol Learn Mem*, 82(1):1–11, Jul 2004.

[88] C. Parron and E. Save. Evidence for entorhinal and parietal cortices involvement in path integration in the rat. *Exp Brain Res*, 159(3):349–359, 2004.

[89] H. Pastoll, L. Solanka, M. C. W. van Rossum, and M. F. Nolan. Feedback inhibition enables theta-nested gamma oscillations and grid firing fields. *Cell*, 77:141–154, 2013.

[90] Martin Pospischil, Maria Toledo-Rodriguez, Cyril Monier, Zuzanna Piwkowska, Thierry Bal, Yves Frégnac, Henry Markram, and Alain Destexhe. Minimal hodgkin-huxley type models for different classes of cortical and thalamic neurons. *Biol Cybern*, 99(4-5):427–41, Nov 2008.

[91] G. J. Quirk, R. U. Muller, J. L. Kubie, and J. B. Ranck. The positional firing properties of medial entorhinal neurons: description and

comparison with hippocampal place cells. *J Neurosci*, 12:1945–1963, 1992.

[92] J. Rinzel and G. B. Ermentrout. *Methods of Neuronal Modeling*, chapter Analysis of neural excitability and oscillations. MIT Press, 1989.

[93] Y. Roudi, J. Tyrcha, and J. Hertz. Ising model for neural data: Model quality and approximate methods for extracting functional connectivity. *Phys Rev E*, 79, 2009.

[94] Uwe Rudolph and Hanns Möhler. Analysis of gabaa receptor function and dissection of the pharmacology of benzodiazepines and general anesthetics through mouse genetics. *Annu Rev Pharmacol Toxicol*, 44:475–98, 2004.

[95] 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):758–62, May 2006.

[96] Francesco Savelli, D Yoganarasimha, and James J Knierim. Influence of boundary removal on the spatial representations of the medial entorhinal cortex. *Hippocampus*, 18(12):1270–1282, 2008.

[97] C. Schmidt-Hieber and M. Hausser. Cellular mechanisms of spatial navigation in the medial entorhinal cortex. *Nature Neuroscience*, 16(3):325–331, 2013.

[98] M. N. Shadlen and R. Kiani. Decision making as a window on cognition. *Neuron*, 80:791–806, 2013.

[99] Trygve Solstad, Charlotte N. Boccara, Emilio Kropff, May-Britt Moser, and Edvard I. Moser. Representation of geometric borders in the entorhinal cortex. *Science*, 322:1865–1868, 2008.

[100] S. Sreenivasan and I.R. Fiete. Error correcting analog codes in the brain: beyond classical population coding for exponentially precise computation. *Nature Neuroscience*, 14:1330–1337, 2011.

[101] Hill-Aina Steffenach, Menno Witter, May-Britt Moser, and Edvard I Moser. Spatial memory in the rat requires the dorsolateral band of the entorhinal cortex. *Neuron*, 45(2):301–13, Jan 2005.

[102] H. Stensola, T. Stensola, T. Solstad, K. Frøland, M. Moser, and E. Moser. The entorhinal grid map is discretized. *Nature*, 492(7427):72–8, Dec 2012.

[103] T. Stensola, H. Stensola, M. B. Moser, and E. I. Moser. The grip map anchors systematically to environment geometry. In *SFN Poster*, 2013.

[104] S. Stewart, A. Jeewajee, T. J. Wills, N. Burgess, and C. Lever. Boundary coding in the rat subiculum. *Philos Trans R Soc Lond B Biol Sci*, 369, 2014.

[105] S. M. Stringer, T. P. Trappenberg, E. T. Rolls, and I. E. T. Araujo. Self-organizing continuous attractor networks and path integration: one-

dimensional models of head direction cells. *Network: Comput in Neural Sys*, 13:217–242, 2002.

[106] Q. Tang, A. Burgalossi, C. L. Ebbesen, S. Ray, R. Naumann, H. Schmidt, D. Spicher, and M. Brecht. Pyramidal and stellate cell specifity of grid and border representations in layer 2 of medial entorhinal cortex. *Neuron*, 84:1191–1197, Dec 2014.

[107] Jeffrey S Taube. The head direction signal: origins and sensory-motor integration. *Annu Rev Neurosci*, 30:181–207, 2007.

[108] S. M. Thompson, L. M. Masukawa, and D. A. Prince. Temperature dependence of intrinsic membrane properties and synaptic potentials in hippocampal ca1 neurons in vitro. *J Neurosci*, 5(3):817–824, 1985.

[109] E. C. Tolman, B. F. Ritchie, and D. Kalish. Studies in spatial learning. 1. orientation and the short-cut. *J Exp Psychol*, 36:13–24, 1946.

[110] Stephane Valerio and Jeffrey S Taube. Path integration: how the head direction signal maintains and corrects spatial orientation. *Nat Neurosci*, 15(10):1445–53, Oct 2012.

[111] T. Van Cauter, J. Camon, A. Alvernhe, C. Elduayen, F. Sargolini, and E. Save. Distinct roles of medial and lateral entorhinal cortex in spatial cognition. *Cereb Cortex*, 23:451–459, 2012.

[112] T. van Groen and J. M. Wyss. The postsubicular cortex in the rat: characterization of the fourth region of the subicular cortex and its connections. *Brain Research*, 529:165 – 177, 1990.

[113] H. Wan, J. P. Aggleton, and M. W. Brown. Different contributions of the hippocampus and perirhinal cortex to recognition memory. *J Neurosci*, 19:1142–1148, 1999.

[114] Adam C Welday, I Gary Shlifer, Matthew L Bloom, Kechen Zhang, and Hugh T Blair. Cosine directional tuning of theta cell burst frequencies: evidence for spatial coding by oscillatory interference. *J Neurosci*, 31(45):16157–76, Nov 2011.

[115] 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.

[116] J. E. Widloski and I. R. Fiete. A model of grid cell development through spatial exploration and spike time-dependent plasticity. *Neuron*, 83:481–495, 2014.

[117] T. Wills, C. Lever, F. Cacucci, N. Burgess, and J. O'Keefe. Attractor dynamics in the hippocampal representation of the local environment. *Science*, 308(5723):873–6, May 2005.

[118] Thomas J Wills, Caswell Barry, and Francesca Cacucci. The abrupt development of adult-like grid cell firing in the medial entorhinal cortex.

Front Neural Circuits, 6:1–13, 2012.

[119] Tom J Wills, Francesca Cacucci, Neil Burgess, and John O'Keefe. Development of the hippocampal cognitive map in preweanling rats. *Science*, 328(5985):1573–1576, Jun 2010.

[120] D. J. Willshaw and C. Von Der Malsburg. How patterned neural connections can be set up by self-organization. *Proc Natl Royal Soc London*, 194, 1976.

[121] M. P. Witter and D. G. Amaral. *The Rat Nervous System*, chapter Hippocampal formation, pages 635–704. Elsevier, 2004.

[122] Menno P Witter and Edvard I Moser. Spatial representation and the architecture of the entorhinal cortex. *Trends Neurosci*, 29(12):671–8, Dec 2006.

[123] R. O. Wong. Retinal waves and visual system development. *Annu Rev Neurosci*, 22:29–47, 1999.

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

[125] 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. *Nat Neurosci*, 16(8):1077–84, Aug 2013.

[126] B. J. Young, T. Otto, G. D. Fox, and H. Eichenbaum. Memory representation within the parahippocampal region. *J Neurosci*, 17:5183–5195, 1997.

[127] K. Zhang. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci*, 15:2112–2126, 1996.

[128] X. O. Zhu, M. W. Brown, and J. P. Aggleton. Neuronal signalling of information important to visual recognition memory in rat rhinal and neighbouring cortices. *Eur J Neurosci*, 7:753–765, 1995.

Vita

John Eric Widloski was born in Brookfield, Wisconsin on 13 December 1984, the son of Dr. Mark and Susan Widloski. Shortly after, he moved to Maryville, Tennessee, where he attended primary and secondary school. Growing up, his interests were in playing music, taking photographs, collecting records, and astronomy. Afterwards, he went to the Georgia Institute of Technology, whereupon he received a Bachelor of Science degree in Physics in 2003. He then started graduate studies in Physics at the University of Texas at Austin. In 2010, he joined Ila Fiete's lab to study the neural mechanisms of spatial navigation, with a special emphasis on modeling grid cells.

Permanent address: 1506 South Court St.
Maryville, Tennessee 37803

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.