Reverse engineering a grid cell network

Author

Maria Mørreaunet

Supervisor

Yasser Roudi

Norges Teknisk-Naturvitenskapelige Universitet (NTNU)

Kavli Institute for Systems Neuroscience and Centre for Neural Computation

Dedication

I would like to dedicate this work to my dear cat Zorro.

Your are, and always will be, missed.

1 Acknowledgements

First of all, I would like to thank my supervisor, Yasser Roudi, for your guidance, for having faith in me, and for introducing me to the fantastic world of science!

Furthermore, I would like to thank Benjamin Dunn, for all your ideas, and the countless hours of helpful discussion and feedback. I am also grateful to Aree Witoelar for help and feedback!

Thanks to Tor and Hanne Stensola for providing the dataset, and for all your help.

The journey through neuroscience has been the most interesting two years of my life, spent at the most interesting place, both in terms of people and science. For this, I am grateful to all the excellent people at at CNC, and to all of my classmates - especially those of you who sat with me on this rollercoaster trying to fit the puzzle pieces.

Bjørn, thank you for both the helpful and the hilarious discussions. And thank you, Torgeir, for always (literally always, as in 24h) being there - especially through the last months. Simply put, I would not have gotten through this summer without all those laughs. That goes for Rine and Eline as well: Thank you!

And last, but certainly not the least: Thank you to all the people who were always there for me when I needed it the most, which to the greatest extent undoubtedly was Daniel. While I gave more than I had to science, you gave more than you had to me.
You are the wind at my back.

Contents

Abstract

The neural circuitry comprising what is assumed to be the brain's spatial navigation system has been investigated by scientists for decades. Research has revealed how the neural networks underlying this cognitive function might operate, and the grid cell network of the medial entorhinal cortex has been suggested to function as a path integrator. How neural networks, such as those of grid cells, operates, is likely to depend on the connectivity within the network. Much is known about the physiological connectivity between the cells in the medial entorhinal cortex, but the question of how the grid cell network is functionally wired is still debated.

Here, we have taken a statistical approach to this question. We have employed a statistical model - the kinetic Ising model - to reverse engineer the functional connectivity of a network of grid cells recorded *in vivo*. Different versions of the model were constructed to allow it to explain spatio-temporal variations in the firing of the cells. This was done by including components assumed to modulate the cells in the external fields of the kinetic Ising model. We have investigated how the differences in external fields of the model affect the inferred functional connectivity between the cells, and how the inferred functional connectivity relates to what has been found experimentally.

Our main findings are that including pairwise correlations in the model increases the model quality for all models, and that the inferred connectivity of the network remains remarkably stable when altering the external field. In addition, we find indications of stronger functional connectivity within grid cell modules than between, and indications that the within-module functional connectivity decreases with phase distance.

Abbreviations

EC - entorhinal cortex
MEC - medial entorhinal cortex
GLM - generalized linear model
HD - head direction
CAN - continuous attractor network
LNP - linear-nonlinear Poisson
LFP - local field potential

In the following, the word 'cell' refers to neurons unless specified otherwise.

2 Introduction

Imagine a ship in the middle of the Pacific Ocean, surrounded by nothing but sea. No land in sight to navigate by. How does it find its way?

Centuries ago, marine navigation was based solely on dead reckoning, the process of estimating the change in one's position based on velocity. Dead reckoning, or *path integration*, is very sensitive to error accumulation - especially at sea, where winds and ocean currents cause errors in estimating one's velocity. This is obviously a detrimental factor influencing the ability to find one's way. To overcome this problem, people have, over time, invented increasingly complex tools to aid navigation. Satellite-based navigation systems, which sprung from ideas conceived in the days of the Space Race, gradually took over marine navigation during the 70s and 80s. Before the advent of such systems, however, the main tools used for navigation at sea were nautical charts, tools to estimate position based on the observation of celestial objects, and a compass enabling them to set course. Starting from a known position, the first mate could calculate an estimate of the future position, based on the speed and course - i.e. path integrate. At fixed time intervals, depending on the complexity of the waters, the position was checked, and, if necessary, corrected. This process was repeated during the voyage until the ship was safe in port.

With the advent of high-tech GPS-based navigation systems, humans have efficiently solved the challenge of navigating worldwide, and today the oceans are populated by cargo vessels continuously carrying goods around the globe. Despite being a magnificent result of centuries of engineering, the intercontinental freighter in transit is, in effect, just a modern representation of a basic animal ability: finding and remembering the location of food sources, getting there by the shortest route possible,

and bringing it back home safely. Ironically, while able to put GPS satellites in orbit and utilize them for navigation, mankind have yet to fully understand its own *innate* navigation system, developed over the course of evolution. How does the human brain, an intricate, biological system of 10^{11} interconnected neurons [1], navigate?

Humans and other mammals are able to find the shortest route to a goal in a familiar environment without ever having traveled that specific path before. This is what Tolman presented in his 1948 paper [2]. In one of the experiments he described, rats were first trained to follow a devious maze tunnel to a food reward. The tunnel known to the rat was then closed, and many straight radial arms were opened. The majority of the rats chose the arm pointing in the direction of the reward, indicating that the rats had an internal sense of the direction in which the reward was located. Through his work, Tolman disproved the postulate that navigation is based only on learned routes, and suggested that, instead, the mammalian brain holds internal, *cognitive maps* of space.

Learning and storing information, such as cognitive maps, requires memory systems. In the medial temporal lobe resides a brain structure essential to declarative memory formation: the hippocampus [1]. The hippocampus is an archicortical[1] structure onto which highly processed neocortical[2] input converges, and whose backprojections spreads out to virtually all of neocortex. Internally, hippocampus has a striking unidirectional excitatory connectivity, from the dentate gyrus (DG) through subdivisions of *cornu ammonis* (CA), two three-layered structures in which the principal cells are condensed in one layer. The direction of flow of information is shown in Fig. 1. Mossy fibers from dentate granule cells innervate pyramidal cells in CA3. Schaffer collaterals arising from pyra-

[1] Archicortex is a phylogenetically old type of two- or three-layered cortex [4].
[2] Neocortex is defined by six layers [5].

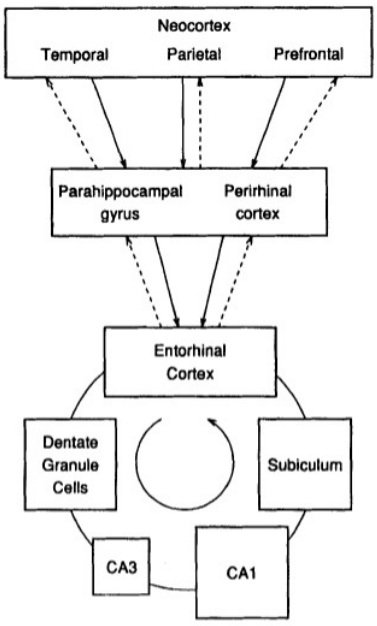

Figure 1: Hippocampus-entorinal cortex circuitry, with forward connections from (solid lines), and backprojections to (dashed lines), the neocortex via the parahippocampal and perirhinal cortex. Information enters and leaves the hippocampus via the entorhinal cortex. The directional connectivity within the hippocampal formation is shown; dentate granule cells project to pyramidal cells in CA3. These cells project to CA1, which project to subiculum. Reprinted from [3].

midal cells of CA3 synapse on to CA1 neurons, which in turn projects to subiculum [4]. The CA1 is a feed-forward network with sparse intrinsic excitatory connections, whereas the CA3 network is remarkably recurrently connected [4], a characteristic feature of memory networks [6].

The role of hippocampus in normal memory function was illuminated by Scoville and Milner in 1957 [7]. They reported on effects of hippocampectomy on memory in ten patients, with the most extreme case being H.M. H.M. suffered from severe epilepsy, and was treated with bilateral resection of parts of the hippocampus. The surgery was successful as an epilepsy treament, but after the surgery and until his death, he suffered from a severe case of anterograde amnesia - the inability to form new memories. This was a strong indication that hippocampus plays an important role in memory. In the following years, a great body of work was done in hippocampal research (reviewed in [8]). The observation

that hippocampal lesions resulted in poor performance in spatial tasks in animals [9–11], led O'Keefe and Dostrovsky [12] to record single-unit activity in hippocampus in a freely moving rat, and with it to a profoundly significant finding regarding the neural representation of space - namely the discovery of *place cells*, as described in more detail below.

2.1 Neural representation of space

Place cells, the first type of spatially modulated neurons to be discovered, constitute the major part of active principal cells in the CA region of the rodent hippocampus [14]. Their name derives from the fact that one place cell fires at one specific location in an environment of the individual. Place cells have been reported in rats, rabbits, mice, and bats [12, 14–18], and cells with similar properties have been reported in humans and nonhuman primates [19, 20]. This indicates that place-specific cells are part of the neural circuitry of hippocampus in all mammals.

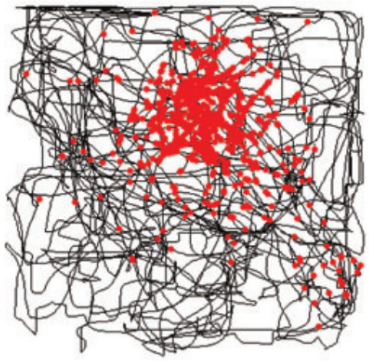

Figure 2: Spatial firing of a place cell recorded in a rat traversing a square box. The trajectory of rat is shown in black. Red circles indicates points where the cell emitted a spike. Adapted from [13].

In rodents, place cell firing shows a peculiar spatio-temporal spiking preference related to the local field potential (LFP)[3]. When the animal is not engaged in locomotion, the network is dominated by irregular LFP. In contrast, when the animal is changing its position in the environment, the network is dom-

[3]Extracellular electrical activity

inated by sinusoidal oscillations at 4-12 Hz, known as theta rhythm [21]. As an animal passes through the firing field of a place cell, the cell will fire at successively earlier phases of the theta rhythm. This phenomenon is known as phase precession. [22]

The total activity of a sufficiently large population of place cells will cover the entire surface of a given environment. If the environment is significantly altered, a phenomenon called remapping can be observed in the place cell population: the activity of the cells will change, both in terms of which cells are active, and the relative location of their firing fields [23]. The large number of place cells makes the number of possible combinations of active cells and the location of their firing fields close to infinite - such that each environment visited by the animal can be uniquely represented by the place cell network.

Although place cells of CA1 and CA3 have similar spatial firing characteristics, the representation of different environments is different in these two hippocampal subfields. The representation of two environments in CA3 is virtually orthogonal, in the sense that distinct subsets of cells are active in the two environments. In CA1, on the other hand, the active cell populations can overlap. In addition, upon exposure to a novel environment, the place cell firing stabilizes faster in CA1 than in CA3 [24]. This suggest a functional difference between the two place cell networks, providing yet another implication for the extensive recurrently connected CA3 subfield as a memory network.

Taken together, the observations described above points to the CA3 of hippocampus as a storage site for maps of different environments, but not necessarily the prime network for *computing* one's position on a given map. To investigate if this computation is performed outside of hippocampus, researchers turned their attention to a neocortical structure adjacent to, and extensively connected with, the hippocampus: the en-

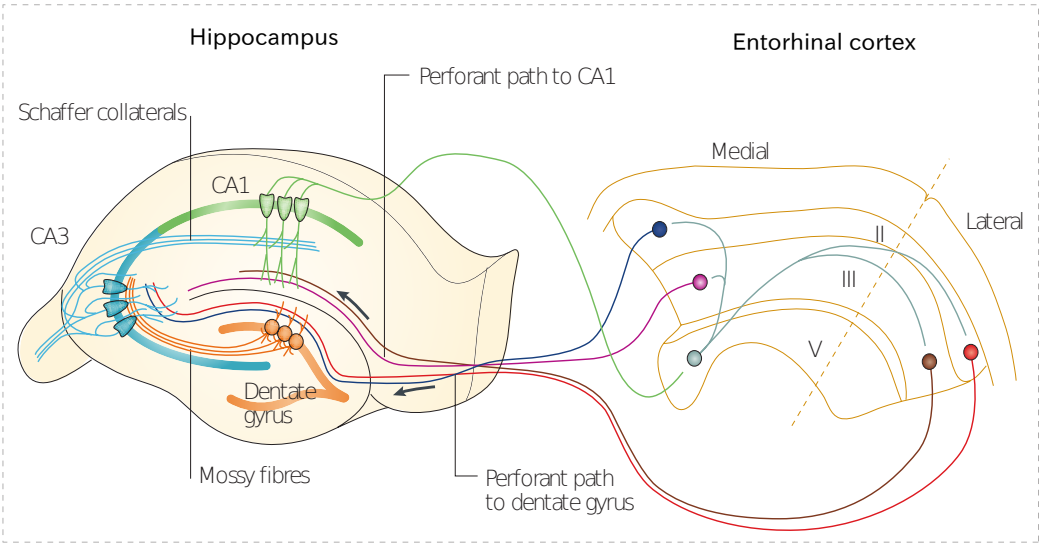

Figure 3: Representation of the layer-specific projections from and to the entorhinal cortex. Layer II projects to dentate gurys and CA3. Layer III projects to CA1, while layer V receives input from CA1. Adapted from [32].

torhinal cortex (EC). Cells in layer II of EC project to DG and CA3, whereas layer III cells project to CA1. EC receives hippocampal input mainly from CA1 to layer V (see Fig. 3). This reciprocal connection with the place cell network suggests a possible role of the entorhinal cortex in the mechanisms generating place cell activity. The first recordings in EC did not reveal cells with strong spatial modulation [25–27], indicating that position is calculated within the hippocampus. Further investigation of the hippocampal circuit, however, revealed persistent place cell firing in CA1 after lesioning dentate gyrus [28] or CA3 [29], leaving only EC as the likely input. This motivated researchers to revisit the EC, and in 2005, the existence of *grid cells* in the medial entorhinal cortex was established [30, 31].

2.2 Grid cells

The spatial firing of grid cells differs from that of place cells: as opposed to the single field of place cells, grid cells have multiple firing fields (shown in Fig. 4). The fields of one cell forms a hexagonal pattern that tiles the entire environment of the animal [31].

The grid pattern can be characterized by the distance between two neighboring firing fields, called the *spacing* of the cell, and the *orientation* of the pattern relative to an axis of the environment. The spacing and orientation is often similar for grid cells that are anatomically close, but no relationship between the *spatial phase* - the spatial offset relative to a reference point in the environment - and the anatomical distance between cells has been found [31, 33].

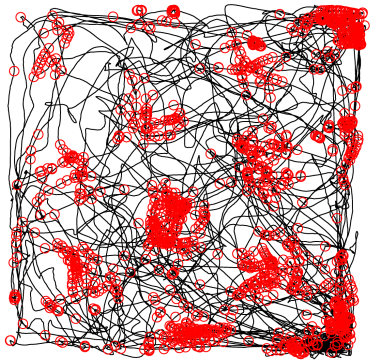

Figure 4: Spatial firing of a grid cell recorded from MEC of a rat traversing a square box. The trajectory of rat is shown in black. Red circles indicates points where the cell emitted a spike.

Although neighboring cells have similar spacing, the neighborhood size is limited: the spacing of the grid cells increases along the dorsoventral axis [30, 31, 33]. Whether this increase is discrete or continuous was recently investigated by Stensola et al. (2012) [34]. They recorded simultaneously from several grid cells, and found that grid cells cluster in modules with similar properties, such as spacing and orientation. The spacing increased in discrete steps between modules, of which up to four was found in one animal. Upon altering the environment, the change in the cell's spatial firing pattern was

coherent within modules, but differed across modules, suggesting that the modules can operate independently, and thus that the connectivity within a module might be stronger than that between modules [34].

Grid cells are not the only type of spatially modulated neurons in MEC. Border cells, which fire when the animal is close to borders of the environment [35], head direction cells, which fire preferentially when the animals is facing specific directions relative to the environment, and cells with conjunctive properties: grid x head direction modulation [33] and border x head direction modulation [36] are also part of the MEC cell population. The firing rates of head direction cells, conjunctive grid x head direction cells, and grid cells are speed modulated [33]. The *location* of a grid field, however, is unaffected by the velocity of the animal, and the spatial phase distance between the fields of two grid cells remains constant irrespective of the environment [37]. In addition, grid cells maintain their spatial firing pattern when visual cues are removed [31]. Taken together, this points to the MEC network as a metric system for keeping track of one's position by path integration.

In rodent MEC, the extracellular theta rhythm is associated with locomotion. Some grid cells show phase precession, but not all: It appears to be a layer specific property. Nearly all grid cells in layer II, and some in layer V, show phase precession. In layer III, however, phase precession is sparse, but phase locking can be observed in this layer [38]. In the next section, we will further discuss the different layers of MEC.

2.2.1 Internal connectivity of the entorhinal cortex

As shown in Fig.1, the EC receives hippocampal input, and projects back to hippocampus. The directional nature of the hippocampal-EC circuit is to some extent maintained within EC: the interlaminar connectivity

of EC is dominated by projections from deep layers to superficial layers [39]. Stimulation of subiculum cause excitatory responses in layer V of MEC [40]. Layer V projects to superficial layers, and about half of these projections terminate on putative interneurons, possibly providing feed-forward inhibition to principal neurons in superficial layers [41]. Projections from layer II to layer III, and from layer II and layer III to deeper layers have been reported, although sparse [39, 42].

Intralaminar anatomy has also been investigated. The principal cells of layer III and layer V are mainly pyramidal cells [4], and recurrent excitation has been found between these cell types within layers, with a relatively high probability of connections [43].

The anatomy of MEC layer II is significantly different from that of layer III and layer V. The main cell type in this layer is the stellate cell, constituting $\sim$67% of the principal cell population. Recently, the connectivity between stellate cells of MEC layer II was investigated, and found to be recurrent and inhibitory [44, 45]. Couey et al. (2013) [44] activated, by optogenetic retrograde labelling and patch clamping, two or more stellate cells and measured the response in other stellate cells. They found hyperpolarizing responses, indicating that the effective connections between layer II stellate cells, probably mediated by interneurons, is inhibitory. Excitatory connections from stellate cells to pyramidal cells were found, but no response in stellate cells were seen after stimulation of pyramidal cells. In layer II, about 50% of the principal cells are grid cells [33]. Although grid cells are also found in deeper layers, these grid cells are largely silent upon exploring a novel environment [46]. Taken together, these results point to the layer II stellate cells as a possible candidate for generation of the grid pattern.

2.2.2 Grid cell models

Models of the mechanism generating the grid pattern has been developed since shortly after the discovery of grid cells [47–49]. One class of grid cell models, the oscillatory interference models, arose as an extension of a place cell model that was originally motivated by the phase precession observed in place cells, and that the autocorrelation of the spike times of the cells had a higher frequency than the theta rhythm. The sum of the two oscillations, having slightly different frequencies, then interferet to generate phase precession [22]. In the first grid cell oscillatory-interference models, the grid cell output was modeled as a sum of different membrane oscillators modulated by velocity in directions spaced by 60°, in addition to a baseline oscillator. The cells in this model act independently of each other [50, 51]. The oscillatory-interference class of grid cell models was recently called into question by the discovery of grid cells in bats, as they lack the continuous theta oscillations observed in rats [52].

Another class of models, the attractor network models, rely on connections in the network to generate the grid cell activity. Attractor networks are dynamical models that have been used to model various neural computations. [53–58]. An attractor neural network is defined by an initial state, coupled dynamical equations, describing the collective behavior of nodes, and a connectivity matrix. The structure of this matrix enables the network to exhibit stable attractor states. An attractor state is (as shown in Fig. 5) a local energy minima in the state space of the network, or, in other words, a pattern of network activity which will remain stable over time. To maintain a stable pattern, the assembly of active neurons in that pattern must be recurrently connected. In one type of attractor network, the point attractors, each stored pattern is different from the others; they are distinct local energy minima in the state space of the system [6].

Two interesting properties that make attractor networks suitable for modelling memory networks are their ability to perform both pattern completion and pattern separation. Pattern completion results if the network at some point in time is in a state contained in the state space of the 'basin of attraction' of an attractor state (see Fig. 5). Over time, the network will then converge to the stored activity pattern - i.e. it is 'attracted' to that state. Depending on the number of nodes and excitatory connections, a biologically realistic point attractor can store a large number of attractor states in its state space [59], enabling it to perform pattern separation. Evidence exists that the CA3 place cell network of hippocampus operates as a point attractor network [60].

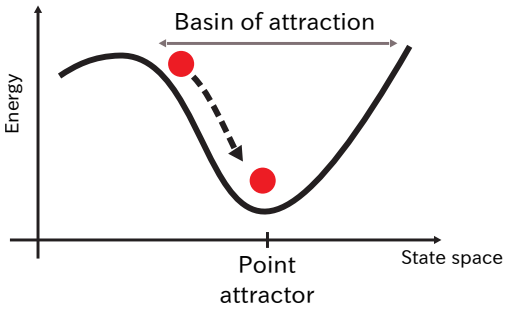

Figure 5: Simplified representation of attractor network dynamics. Attractor states are local energy minima in the state space of the system. If the system is in a position within the basin of attraction, it will converge to the stable attractor state. Adapted from [6].

One particular type of attractor network is the continuous attractor network. Rather than distinct attractor states, these networks have continuous attractor states [6], and are therefore suitable for modeling continuous features, such as eye movements [55] and head direction [58], and also location: both place cells [57, 61] and grid cells [48, 49, 54] have been modeled with continuous attractor models. In most of these grid cell models, cells that are close in spatial phase space have recurrent excitatory connections. The connection strength decreases with phase distance, so that cells with opposite phase inhibit each other [48, 49, 54]. This type of connectivity is known as the Mexican hat. Connecting the

cells according to spatial phase implicitly assumes that the cells have the same spacing and orientation. The activity of the network is driven by velocity input from head direction (HD) cells or HD x grid cells [44, 48, 54].

Recent findings, such as the modular arrangement of grid cells [34], which show that groups of cells have similar spacing and orientation, support the continuous attractor model. The recurrent inhibitory connectivity pattern revealed in [44], however, contradicts the Mexican hat connectivity. Whether a continuous attractor network with recurrent inhibition can generate grid patterns, was investigated in [44]. A continuous attractor model where, instead of excitation, each cell inhibits its phase space neighbors within a certain radius, was constructed. Providing the cells with a constant excitatory input generated grid firing, and the grid activity could be spatially driven by velocity-modulated head-direction input. The model also predicted that with loss of excitatory input, which grid cells are thought to receive from hippocampus [40], grid cells lose their spatial pattern and become head-directionally tuned - as was found experimentally [62].

2.3 Summary

The grid cell network of the medial entorhinal cortex is assumed to perform path integration. Just like the dead reckoning of the ship crossing the Pacific Ocean, estimating one's position through path integration accumulates errors without regular positional updates. The path integration and the update mechanisms are tightly connected on the ship, because one mechanism needs backup from the other: If one cannot keep track of where one is going between updates, continuous updating is needed, which requires a lot of attention. On the other hand, path integration alone is too prone to error accumulation. It is clear from the connectivity between what is assumed to be the path-integrator and its update system of the brain, namely the MEC and hippocampus, that they probably cooperate closely.

The connectivity of a network can tell a lot about what the network can and cannot do, implying that to understand the brain, studying connectivity is fundamental. Here, we have reverse engineered the functional connectivity of a network of grid cells from an *in vivo* recording from [34]. 'Functional connectivity' means the way in which a neuron affects another neuron in a statistical sense - which is not unlikely to be related to the real, underlying connectivity of the network. Ideally, then, our inference should yield results similar to experimental findings. To infer this connectivity, we use statistical models that will be described in more detail in the following section.

3 Statistical models

Neurons process and transmit information in the form of spikes. A neuron receives inputs from other neurons, or - if it is a sensory neuron - from sensory stimuli. Depending on the total input to a cell, and how the cell integrates it, the membrane voltage might reach threshold, resulting in the generation of an action potential, also known as a spike. The action potential is conducted down the axon to the synaptic terminal, where it triggers the release of neurotransmitters that diffuse across the synaptic cleft. The neurotransmitters bind to receptors in the postsynaptic neuron, causing a change in the membrane potential. This is generally assumed to be the fundamental mechanism by which the brain operates [1].

One approach to understanding the brain is to study how neurons, and networks of neurons, integrate information and translate it to the neural language of spikes. To answer these questions, one can record the response of one sensory neuron to a set of stimuli. From this, one can infer what the preferred stimulus of that cell is based on tuning curves: firing rate as a function of stimulus, or similarly, stimulus-triggered averages: average stimulus preceding a spike for one neuron. To do this, many trials are needed, as the response of a neuron is stochastic in nature.

In addition, the activity of each neuron is not independent of that of other neurons in the same network. To investigate coding at network level, one records the activity of many neurons simultaneously. Building the probability distribution of responses based on experimental results goes quickly from trivial to intractable as more neurons are considered,

because the activity of two neurons from the same network are likely to be correlated. Thus, the number of trials needed to construct the probability distribution increases exponentially with the number of neurons.

Ways of dealing with this 'curse of dimensionality' has become important to life sciences over the last few decades, due to a lot of technological breakthroughs, enabling the generation of vast amounts of high-dimensional data. Examples of such techniques are microarray techniques in cell biology, making it possible to analyze genetic [63] or protein-protein interactions [64], and multielectrodes in neuroscience, making it possible to analyze the structure and function of neuronal networks. A solution to the problem of analyzing high-dimensional data is to, instead of constructing the probability distribution experimentally, use statistical models. Statistical models make an assumption about what kind of probability distribution the observed data could have been drawn from. Based on this assumption, one can infer from the data what values of the parameters of that distribution are the most likely ones to generate the observed data. When applying this to neural data one can infer how the neurons influence each other in a statistical sense. This is what is meant by functional connectivity, which, as mentioned previously, can be related to the real connectivity in the network. In [65] the inferred couplings was well correlated with the true connectivity of a biologically realistic computational cortical network. Statistical models employed to reverse engineer neuronal networks, such as the one used in [65], will be described below.

3.1 Generalized linear models

GLMs are a flexible class of statistical models originally developed by Nelder and Wedderburn in 1972 [66], as an extension to linear regres-

sion. In the simple linear model, the response variable is assumed to be normally distributed, as it is a linear function of the input variable and a Gaussian noise term. When modeling response variables that are not normally distributed, such as spike counts in time bins, GLMs are an appropriate choice.

A generalized linear model consists of a conditional probability distribution of the response variable given the independent variables, a linear function of the independent variables, and a link function transforming the linear function to the expectation of the response variable [66].

GLMs have been employed to analyze neural data [67–69], where, often, the known sensory input to the neuron is added in the linear filter, and the probability distribution of spikes is assumed to be an inhomogenous Poisson process. This is known as the linear-nonlinear Poisson (LNP) cascade model, and it has a long history of application to neural data [68, 70]. Recently, this model was extended to include functional connectivity in the network [67]. In a LNP model with functional connectivity, the Poisson parameters $\lambda_i(t)$ is generally given by

$$\lambda_i(t) = f\Big[\sum_j \int J_{ij}(t-\tau)S_j(\tau)d\tau + h_i(t) + k_i x(t)\Big] \tag{1}$$

where J_{ij} is the time delay-dependent effect of neuron j on neuron i, $x(t)$ is the stimulus at time t, h_i is related to the cell's baseline firing rate at time t and $S_j(t-\tau)$ is the number of spikes at time $t-\tau$.

One is interested in inferring what values of the parameters of the model (here: J_{ij}, h_i and k_i) are the most likely to yield the observed data, assuming that the data was generated from this model. To answer this, one constructs a likelihood function: an expression for the probability of observing the data at hand depending on a set of parameter values. For Poisson distributed spike data with independent time bins, the like-

lihood function is

$$\mathcal{L}(\mathbf{S}, \boldsymbol{\lambda}) = \prod_{i=1}^{N} \prod_{t=1}^{T} \frac{(\lambda_i(t))^{S_i(t)} e^{-\lambda_i(t)}}{S_i(t)!} \tag{2}$$

The parameters can then be estimated by maximizing the likelihood function with respect to the parameters given the data. Usually, one maximizes the log of the likelihood function, since it is more convenient to work with. This is equivalent to maximizing the likelihood as $\log(x)$ is one-to-one for $x > 0$. For the GLM with Poisson distribution, the log-likelihood function is

$$\ln\mathcal{L}(\mathbf{S}) = \sum_{i=1}^{N} \sum_{t=1}^{T} (S_i(t) \ln(\lambda_i(t)) - \lambda_i(t) - \ln(S_i(t)!)) \tag{3}$$

The function is maximized by starting from a random set of parameters, and then moving a small step in the direction of the gradient. The process is repeated until the parameters converge; i.e. a maximum is reached. For convex likelihood functions, this is the global maximum. Given a close approximation of the actual system by the model, the inferred parameter values will reveal how the activity of the neurons depend on the input variables, and correlations between neurons depend on pairwise couplings.

3.2 Ising models

In large-scale neural recordings, the spike correlation between two neurons is often weak, suggesting that they act independently of one another [71]. However, the distribution of observed spike patterns at the population level does not agree with the distribution one would expect if the neurons act independently. Schneidman et al. (2006) [71] showed that

weak correlations between neurons can have a major effect on the population statistics. They did this by considering the spike pattern distribution given by the simplest possible model taking both pairwise correlations and firing rates into account, and that given by a model without interactions. Comparing the two distributions to the observed data, they found that the model that included correlations outperformed the one that assumed independence between neurons.

When asking what the simplest possible model that accounts for firing rate and pairwise correlations is, one arrives at the Ising model. This model is named after Ernst Ising, a physicist working in the field of statistical physics [72]: a discipline developed to bridge the microscopic and the macroscopic world. Macroscopic quantities, such as pressure and volume, can be calculated by applying statistical principles to the properties and configuration of the vast amount of microscopic units, e.g. molecules, constituting the system. This bridging relies on the principle that a system is most likely found in the state of maximum entropy, i.e. the state of maximum disorder [73].

Maximizing the entropy for a set of spike patterns with constraints that it should be consistent with observed means and pairwise correlations, one arrives at the Gibbs equilibrium distribution for the pairwise Ising model

$$P[\mathbf{S}] = \frac{1}{\mathrm{Z}}\exp\left[\frac{1}{2}\sum_{ij} J_{ij}S_iS_j + \sum_i h_iS_i\right] \tag{4}$$

where Z is a normalization factor, known as the partition function [74]. The J_{ij} is, equivalent to GLMs, the coupling of neuron j to neuron i, and h_i is related to the cell's baseline firing rate. In contrast to GLMs, S_i = 1 describes the presence of any number of spikes within the time bin, and J_{ij} is symmetric and time delay independent.

The model in (4) describes, for a given set of h_i and J_{ij}, the distribution

of spike patterns the system visits regardless of the order they are visited in. An intuitive problem with modeling neural data with this model is that networks of neurons are dynamical systems, where the spike pattern observed at a given time in most cases depends on the recent history of spike patterns. In addition, the assumption that the external input is constant in time is unlikely to hold for biological neural networks. Shared time-varying input to two neurons will in most cases cause a correlation between them and, if the varying input is not taken into account, the statistical model will interpret this correlation as the existence of a connection [65]. One can solve these problems by letting the state of each neuron at a given time depend on the state of the whole population in the previous time-step, as follows (Glauber, 1963) [76]

$$P(S_i(t+1) = 1|\{S_j(t)\}) = \frac{\exp\left[h_i(t) + \sum_j J_{ij}S_j(t)\right]}{2\cosh\left[h_i(t) + \sum_j J_{ij}S_j(t)\right]} = \frac{1}{2}\left[1+S_i(t+1)\tanh H_i(t)\right] \tag{5}$$

where

$$H_i(t) = h_i(t) + \sum_j J_{ij}S_j(t) \tag{6}$$

This is the kinetic Ising model, with the same parameters as (4). Note that, unlike the equilibrium model described at (4), J_{ij} can be asymmetric. The kinetic Ising model is closely related to GLMs; the kinetic Ising model is, in fact, a 'simple' GLM with a Bernoulli distributed response variable S_i, a simple linear filter without temporal integration ($\eta = H_i(t) = h_i + \sum_j J_{ij}S_j(t)$), and the link function relating η to the mean is $\frac{1}{2}(1+\tanh H_i(t))$.

3.2.1 Learning algorithms

In the *inverse* kinetic Ising model one asks what values of the parameters h_i and J_{ij} are the most likely to generate the observed data. Both exact and fast approximate algorithms for solving the inverse kinetic Ising model have been developed [77]. The exact solution is, like in the GLMs, found by gradient ascent on the log-likelihood function

$$L[\mathbf{S},\mathbf{J},\mathbf{h}] = \sum_{it} \left[S_i(t+1)H_i(t) - \log 2\cosh H_i(t)\right] \tag{7}$$

The learning rules are constructed from the gradient, and a learning rate η to ensure smooth convergence. For the simplest version, where the external fields h_i are constant, the learning rules are

$$\delta h_i = \eta\left[m_i - \langle \tanh H_i(t)\rangle_t\right] \tag{8}$$

$$\delta J_{ij} = \eta\left[\langle S_i(t+1)S_j(t)\rangle_t - \langle \tanh H_i(t)S_j(t)\rangle_t\right] \tag{9}$$

where $m_i = \langle S_i(t)\rangle_t$.

3.3 Applications to neural data

Both GLMs and Ising models have been applied to neural data. Recently, a lot of effort have been put in investigating the significance of pairwise correlations in networks of retinal ganglion cells.

As previously mentioned, Schneidman et al. (2006) [71] analyzed spike data from a network of salamander retinal ganglion cells using the equilibrium Ising model, and found that a model with pairwise correlations captures the statistics of the observed data significantly better than a model that assumes independent neurons. Similar results have been found in analysis of data from a monkey retina. Pillow et al. (2008) [67]

used a GLM with pairwise correlations in addition to the known stimulus, and found that this model fit the data better than a model without interactions. With a slightly different model, a 'stimulus-dependent maximum entropy model' (SDME), Granot-Atedgi et al. (2013) [78] also found that including connectivity in their model improved model fit on data from salamander retinal data.

Recently, using the non-stationary kinetic Ising model, the effect of a time-varying external input on the inferred connectivity for retinal data was investigated by Tyrcha et al. (2013) [65]. They analyzed a cortical model dataset with known connectivity, and a dataset from a salamander retina. The finding for the model data was that the kinetic Ising model not only fit the data better when including the couplings, but also that the inferred couplings was well correlated with the true connectivity of the network. The story for the retinal data, however, was different: Here, the finding was that the couplings were insignificant, and did not contribute significantly to the quality of the model.

Three of the studies described above concluded that correlations are important, while the fourth one [65] found that they are not. As the authors in [65] discussed, this is probably partly because they use a time-varying external field that will not only account for visual stimuli to the cells, but also all trial-to-trial reproducible correlated input to the recorded cells from unrecorded cells, theoretically leaving only noise correlations to be explained by couplings.

Here, we use the kinetic Ising model to analyze an *in vivo* 20-minute recording of 27 grid cells. We investigate how the significance of correlations in this dataset is affected by increasing the 'resolution' of the external field to account for the spatial variation in firing rate. We do this by defining the external fields of the kinetic Ising model in different ways. Through the external fields we alter the assumed input to the cells, either

spatially, or by other external input thought to modulate the firing of grid cells. We also study whether the inferred couplings reflect layer specific and modular structure that has been found experimentally.

4 Methods

4.1 Data

A dataset from a 20-minute (1253 seconds) recording of 65 neurons in the MEC area of a Long Evans male rat was analyzed [34], with the main focus on the 27 grid cells of the dataset. The activity of the cells and the local field potential (LFP) were recorded with a multisite approach with 10 tetrodes. 27 cells distributed on 7 tetrodes showed grid-like firing. Recording locations are shown in Appendix A.2. The grid cells were clustered according to spacing and orientation as described in [34].

Table 1: Distribution of grid cells on tetrode tracks

Cell no.	Tetrode track
1-3	2
4-8	3
9-13	4
14-15	6
16-20	7
21	10
22-27	12

The spikes were binned into 10 ms or 20 ms timebins. From this, a 'spike matrix' of -1's and 1's was constructed, where a '-1' indicated that the cell did not fire in time bin t, and a '1' indicated that the cell emitted one or more spikes in time bin t.

Because of missing data regarding true head direction, we used running direction as an estimate of head direction in the statistical test (Watsons U^2 test [79]) for head-direction modulation. We excluded all epochs with

running speed below 5 cm/s.

4.2 Model definition and inference

To infer the connectivity in the network, we used the kinetic Ising model, i.e. we assumed that observed spike trains comes from the probability distribution in (5). To infer external fields h_i and couplings J_{ij} we maximize the likelihood function given in (7) for the given dataset under the current model by gradient ascent. We construct different versions of the kinetic Ising model by varying the external field in the following ways:

1. Global:
To allow the external field of the kinetic Ising model to account for the spatial variations in the firing of the grid cells, we started by dividing the environment globally into K square boxes. We defined three models with increasing spatial resolution, with K = 2 x 2 in the first model, K = 3 x 3 in the second model, and K = 4 x 4 in the third. For each K, we defined external fields $h_{i,k}$, and a set of Cartesian coordinates, denoted A_k, defining the spatial extent of box k.
The total field in this model was given by

$$H_i(t) = h_{i,k(t)} + \sum_j J_{ij} S_j(t) \tag{10}$$

where $k(t)$ denotes that $x(t)$, $y(t) \in A_k$, and $x(t)$, $y(t)$ is the position of the rat at time t.

2. Circle fields:
We further increased the spatial resolution of the external fields by assigning each cell customized circular fields to closely match the cell's spatial differences in firing rate. The difference between the global and the customized fields is illustrated in Fig. 4.2. Each grid cell received an

individual number of circular fields, equal to, and in the same position as, its firing fields. The fields were defined as follows:
Local maxima on the ratemap were defined as field centers. The center of field k of cell i is denoted by $c_{i,k}$. The local maxima were defined as entries of the ratemap matrix with value greater than all its neighbouring entries within a radius of 6 cm. To determine the radius $r_{i,k}$ of field k for cell i, we calculated the average ratemap value on the circumference of circles, denoted by $\mathrm{E}_c(r)$, starting with r = 1 cm. Then, we increased r with 1 cm steps, and calculated $\mathrm{E}_c(r)$ in each step. This process was repeated until $\mathrm{E}_c(r) > \mathrm{E}_c(r-1)$. The radius of field k was set to half of the radius at this turning point ($r_{i,k} = \mathrm{E}_c(\frac{r}{2})$).

The total field of cell i at time t in this model is

$$H_i(t) = \sum_{k=2}^{n_i} \alpha_{i,k} \Phi\left[r_{i,k}^2 - \left((x(t) - x_{k,i})^2 + (y(t) - y_{k,i})^2\right)\right] + \alpha_{i,1} \prod_{k=2}^{n_i} \Phi\left[\left((x(t) - x_{k,i})^2 + (y(t) - y_{k,i})^2\right) - r_{i,k}^2)\right] + \sum_j J_{ij} S_j(t) \quad (11)$$

where Φ is the Heaviside function, $x(t), y(t)$ is the position of the rat at time t, n_i is the number of fields for cell i, and $c_{i,k}$ is the center of field k for neuron i.

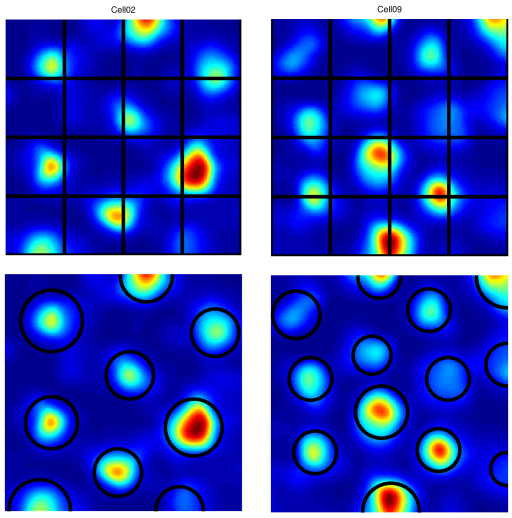

Figure 6: Illustration of difference between global and customized external fields

3. Gaussian fields:
Here, we defined the field of each cell as a sum of a constant field and two-dimensional Gaussian fields centered at the local maxima of the ratemap $(x_{k,i})$, and variance $r_{k,i}$. The total external field for cell i at time t is

$$H_i(t) = \sum_{k,i} \alpha_{k,i} \exp\left[-\left(\frac{(x(t) - x_{k,i})^2 + (y(t) - y_{k,i})^2}{r_{i,k}^2} \right) \right] + h_i + \sum_j J_{ij} S_j(t) \tag{12}$$

where $(x_{k,i}$,$y_{k,i})$ and $r_{i,k}$ is the field center and radius, respectively, of the fields we defined in the circle model for cell i.

4. Velocity field:
Motivated by the indications that grid cells perform path integration, we also defined a 'path integration' model, where we included a velocity field in addition to a constant external field. Because of missing data regarding true head direction of the animal, we used running direction as an estimate of head direction when constructing the input vector.

The total external field for cell i at time t in this model is

$$H_i(t) = h_i + \sum_k \alpha_{ik} v(t) \cos\big(\phi_{\text{HD}}(t) - \phi_{\text{HD},k}\big) + \sum_j J_{ij} S_j(t) \tag{13}$$

where $v(t)$ is the average speed in time bin t, $\phi_{\text{HD}}(t)$ is the average head direction in time bin t, and the set of ϕ_k's are k=2 angles spaced by $\frac{\pi}{2}$.

5. Theta (LFP) field

As mentioned in the introduction, it is known that some MEC cells show theta modulation. This was the case for most cells in the dataset analyzed here, as can be seen in Appendix A.1. This motivated us to construct a model with theta input added to the external field. The LFP data was fast-fourier transformed, and the maximum component between 4-12 Hz was set as the theta rhythm. From this, we constructed a theta input vector, where each element was the angular average of the theta phase in that time bin. In this model, we chose to use 10 ms time bins to get a better estimate of the actual phase of theta.

The total external field for cell i at time t in this model is

$$H_i(t) = h_i + \sum_k \alpha_{ik} \cos\big(\phi_{\text{LFP}}(t) - \phi_{\text{LFP},k}\big) + \sum_j J_{ij} S_j(t) \tag{14}$$

where $\phi_{\text{LFP}}(t)$ is the angular average of the theta phase in time bin t, and the set of $\phi_{\text{LFP}}k$'s are k=2 angular phases spaced by $\frac{\pi}{2}$.

5 Results

We have inferred the functional connectivities in an *in vivo* recording of 27 grid cells, and investigated how the inferred functional connectivity is affected by allowing the external field of a kinetic Ising model to vary in space and time. We defined 8 versions of the external field. In Fig. 7-12, histograms of the inferred couplings and external fields of the different model versions are presented.

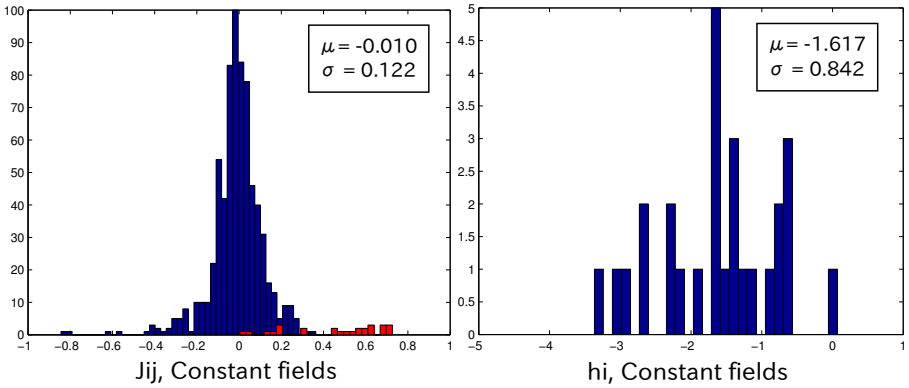

Figure 7: Histograms of inferred couplings J_{ij} (left, red: self-couplings) and external fields h_i (right) for the constant field model.

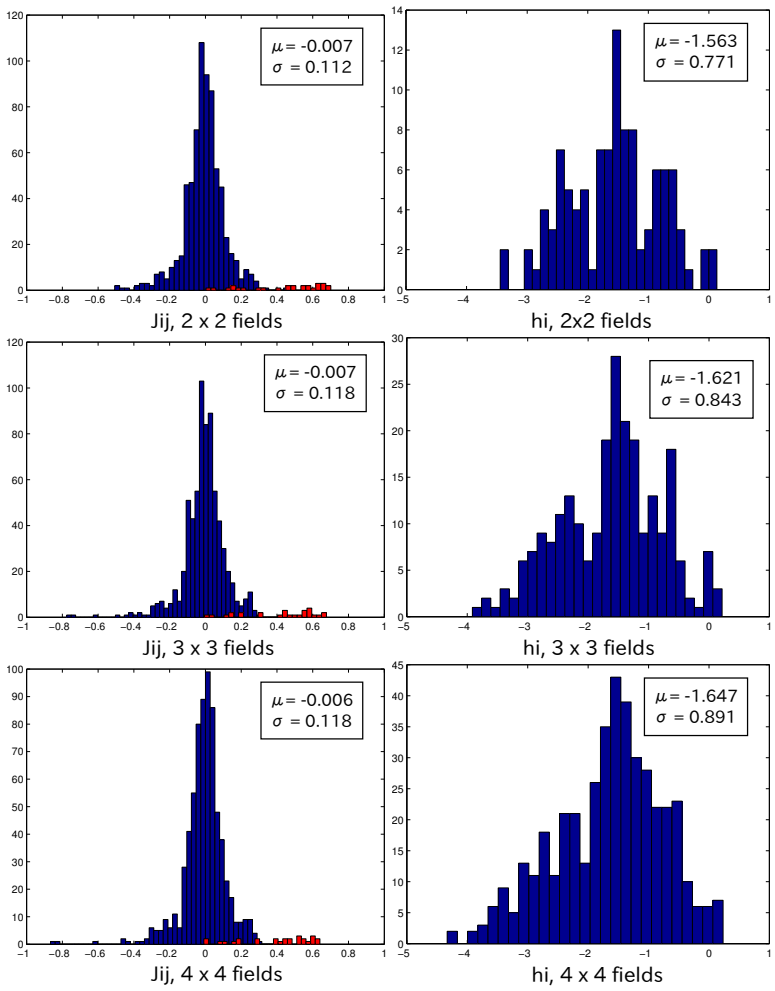

Figure 8: Histograms of inferred couplings J_{ij} (left, red: self-couplings) and inferred external fields h_i (right), for the 2 x 2 fields model (top row), 3 x 3 fields model (middle row), 4 x 4 fields model (bottom row).

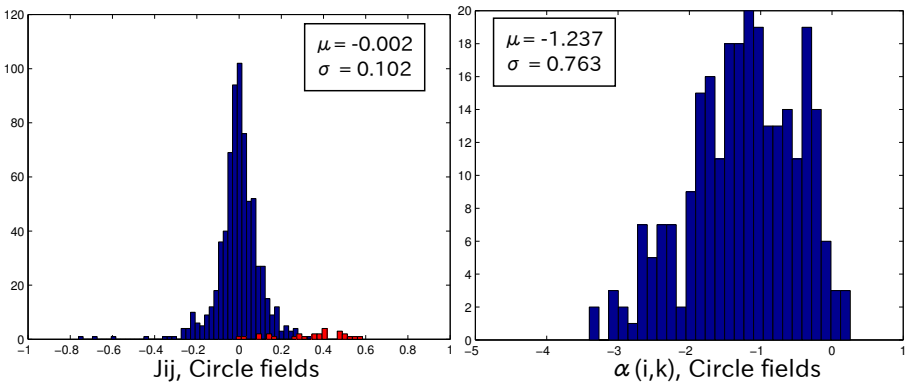

Figure 9: Histograms of inferred couplings J_{ij} (left, red: self-couplings) and external fields $\alpha_{i,k}$ (right) for the circle field model.

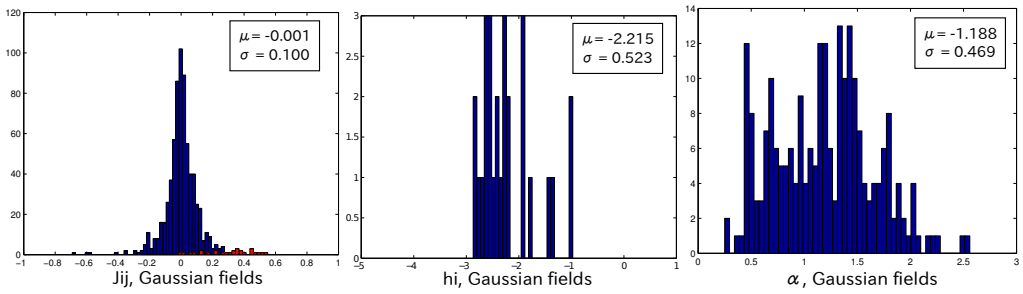

Figure 10: Histograms of inferred couplings J_{ij} (left column, red: self-couplings), external fields h_i (middle column), and α_{ik} (right column) for the Gaussian fields, where α_{ik} is the inferred strength of the k'th two-dimensional Gaussian field of cell i.

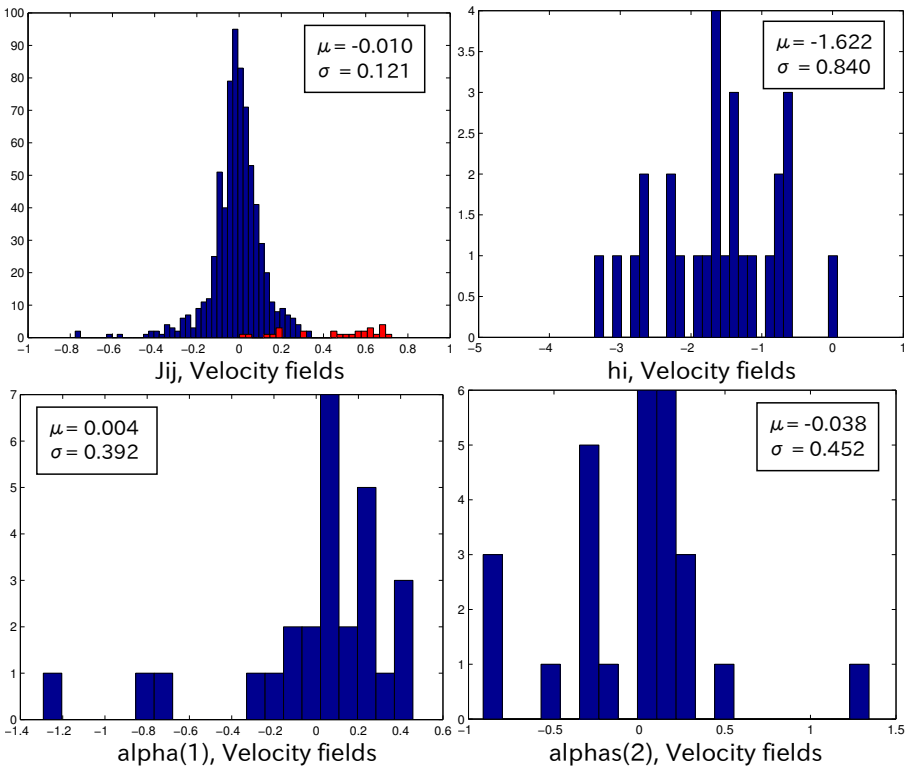

Figure 11: Histograms of inferred couplings J_{ij} (top left, red: self-couplings), external fields h_i (top right), $\alpha_{i,1}$ (bottom left) and $\alpha_{i,2}$ (bottom left) for the velocity field model. $\alpha_{i,1}$ and $\alpha_{i,2}$ is the inferred strength of head-direction modulation in two perpendicular directions.

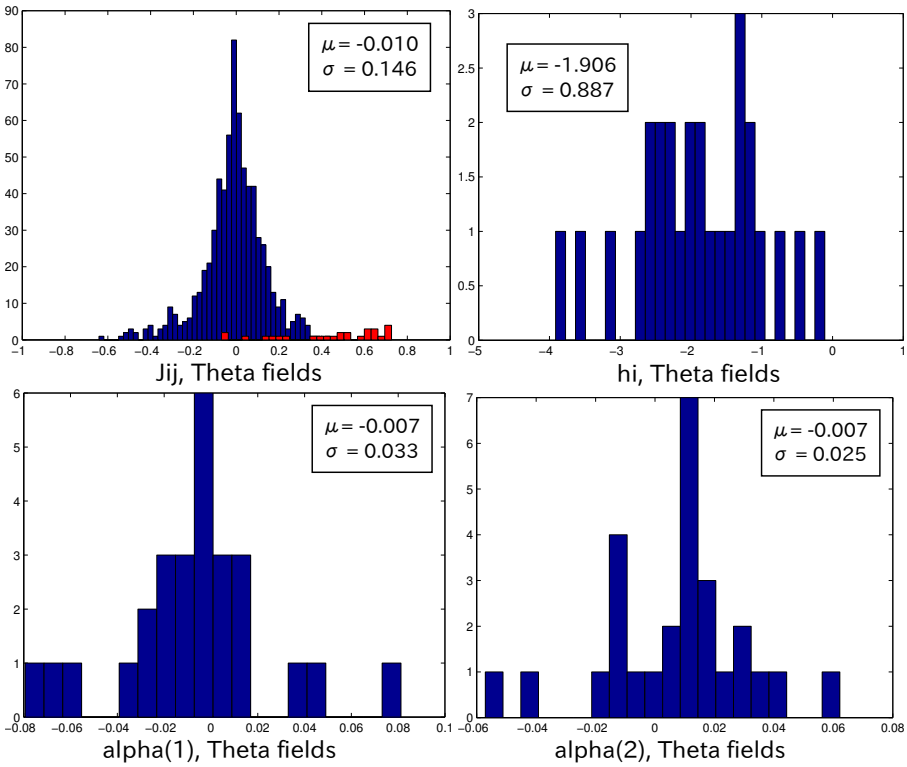

Figure 12: Histograms of inferred couplings J_{ij} (top left, red: self-couplings), external fields h_i (top right), α_{i1} (bottom left) and α_{i2} (bottom left) for the theta field model. $\alpha_{i,1}$ and$\alpha_{i,2}$ is the inferred strength of theta modulation in two perpendicular phases.

In all models, the external fields are negative and often strong, as one would expect for datasets with very low firing rates - which is the case for this dataset (see Fig. 13). The couplings are distributed around zero, and the self-couplings are positive. With the refractory period in mind, positive self-couplings might seem counterintuitive. However, the refractory period lasts for only a few ms, and we use 20 ms timebins. In addition, grid cells are active when the animal is in the cell's fields, and silent oth-

erwise, i.e. the state of a grid cell in a time bin is likely to be equal to the state in the previous time bin, which a stastistical model could interpret as a positive self-coupling.

5.1 Model comparison

5.1.1 External fields

We modified the external field of the kinetic Ising model to include possible sources of signal correlations, such as the spatial dependence of grid cell activity. We started by dividing the environment into subregions; 2 x 2, 3 x 3 and 4 x 4 boxes, and assigning each cell an external field per box. First, we checked if the inferred fields mimic the spatial difference in the firing of the cells. Below is a plot of average firing rate in a box vs. average inferred field in the box, where the boxes have been sorted by firing rate.

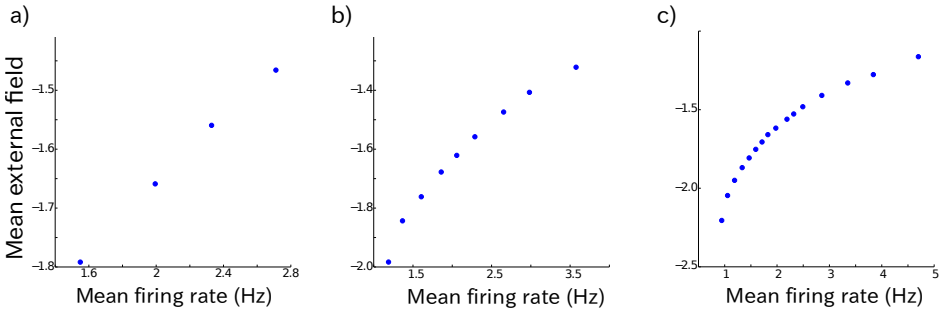

Figure 13: Average firing rate in box vs. average inferred field in the box. The boxes have been sorted by firing rate. a) 4 boxes, b) 9 boxes, c) 16 boxes.

Results are similar for the circle and Gaussian fields (not shown), i.e. the external fields capture at least part of the spatial variation in firing. We then asked whether this has any effect on the couplings we infer.

5.1.2 Couplings

In Fig. 14-15, the inferred couplings of all models with varying fields are plotted against the inferred couplings in the stationary model.

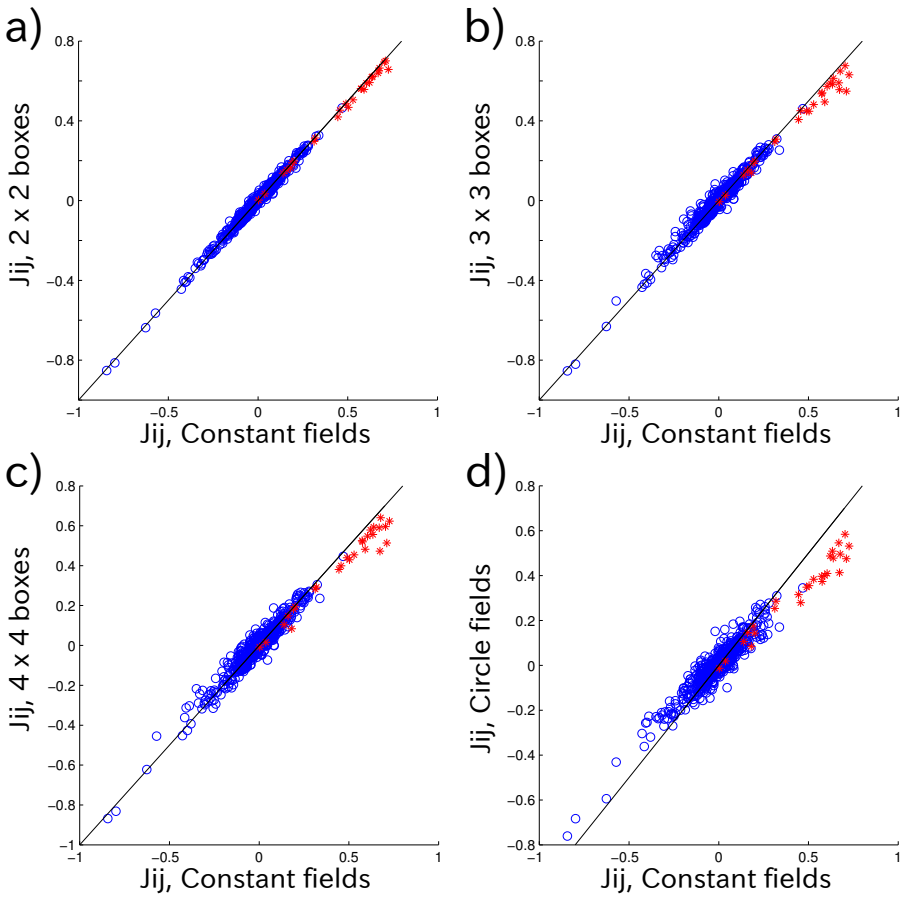

Figure 14: Inferred couplings from models with box fields (a-c) and circle fields (d) plotted against the couplings inferred with constant external field. Red: self-couplings.

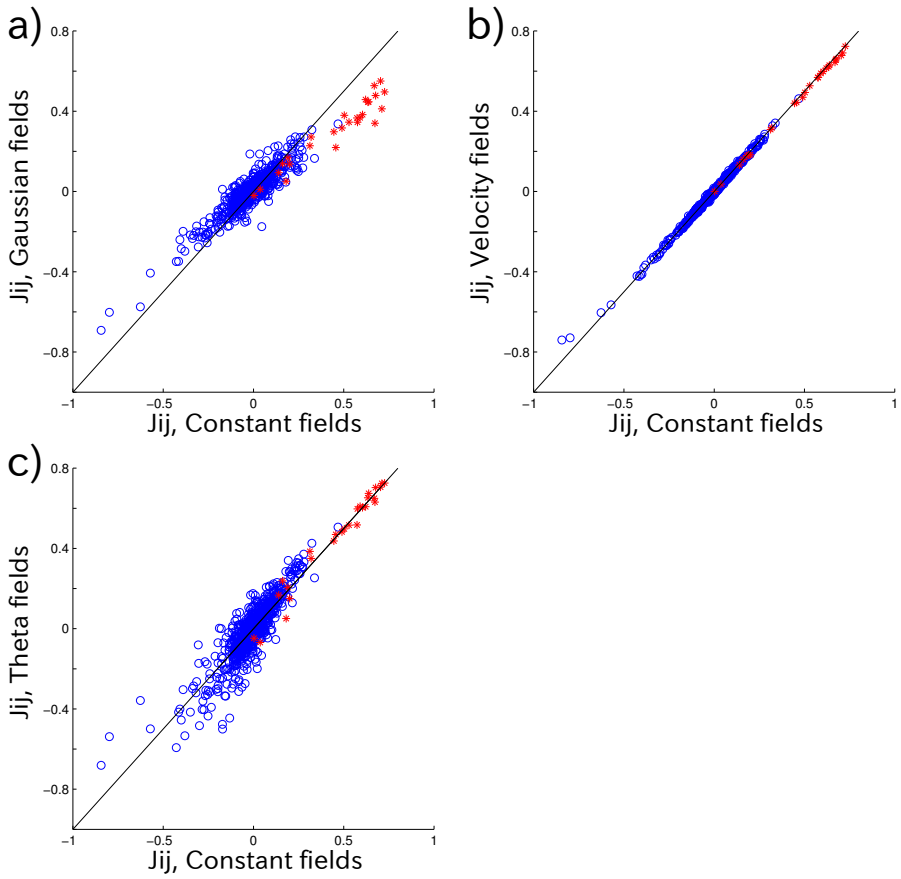

Figure 15: Inferred couplings from models with Gaussian fields (a), velocity fields (b), and theta fields (c), plotted against the couplings inferred with constant external field. Red: self-couplings.

The strength of the couplings decreases with increased 'spatial resolution' of the external fields. However, the couplings are clearly stable between models. We asked if this could be caused by the self-couplings which in most cases are strong and positive. As previously discussed, when the animal is in the firing field of a grid cell, the cell tends to fire, while when the animal is outside the cell's fields the cell is silent; i.e. the state of the cell in the previous timestep for most of the recording is equal to the next - which could be explained by a positive self-coupling. We

therefore tried setting all self-couplings to zero. First, we checked if the couplings change within a model without self-couplings (Fig. 16).

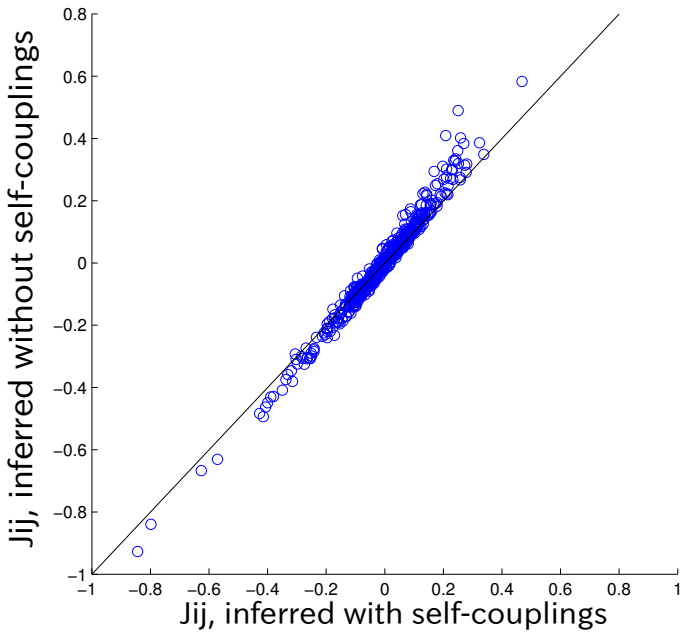

Figure 16: J_{ij} inferred with self couplings plotted against J_{ij} inferred without self couplings in the constant field model.

Removing self-couplings had little effects on the couplings between cells within a model. This indicates that the couplings are stable across models as well, which indeed is the case, as shown in Fig. 17 - 18:

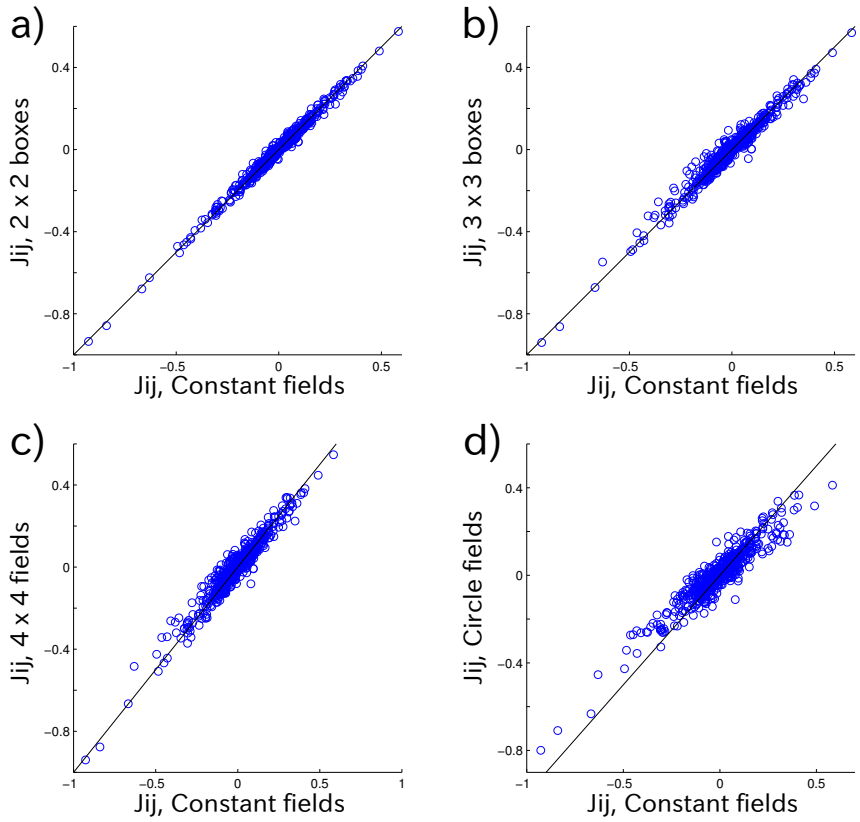

Figure 17: Inferred couplings from models with 2 x 2, 3 x 3, 4 x 4 and circle external fields plotted against the couplings inferred with constant external field. No self-couplings.

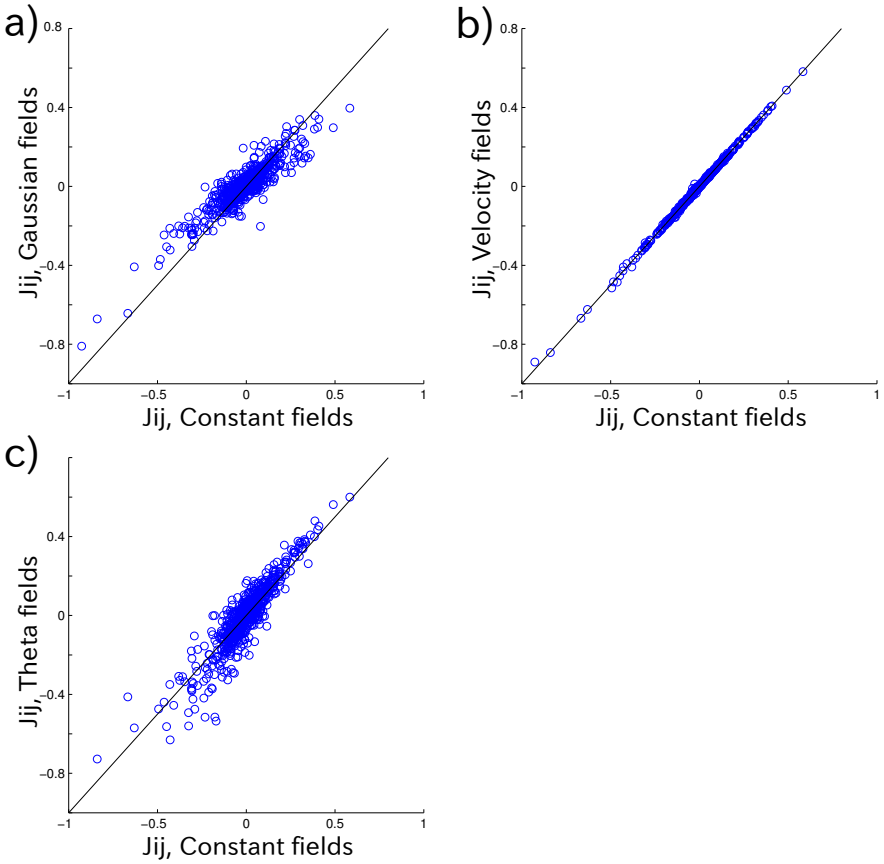

Figure 18: Inferred couplings from models with Gaussian, velocity and theta external fields plotted against the couplings inferred with constant external field. No self-couplings.

For the constant field model, we included all cells in the inference. Below is a scatter plot of the couplings inferred on only the grid cells vs. whole dataset.

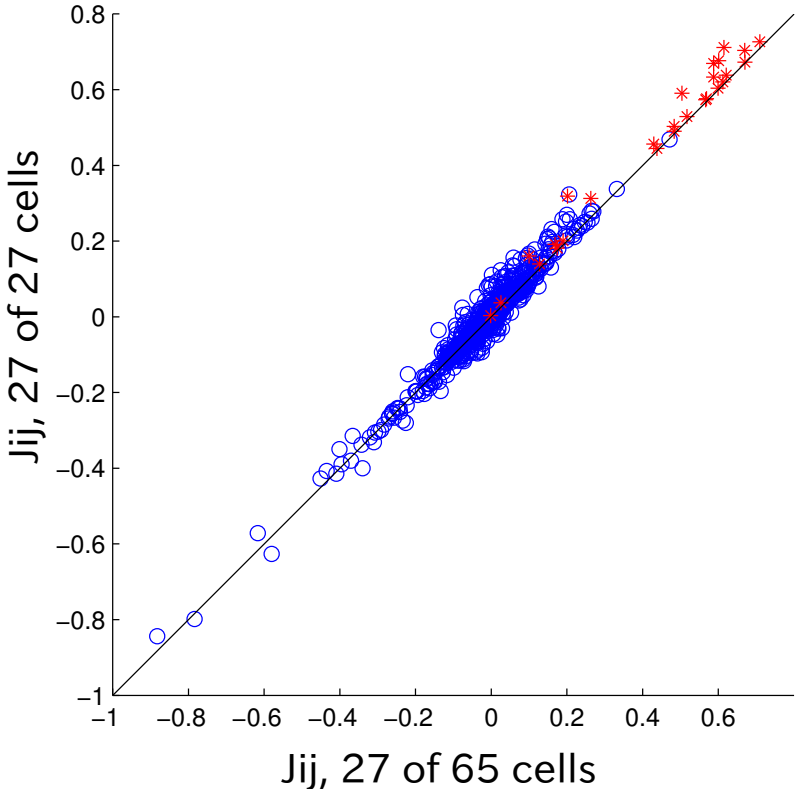

Figure 19: Couplings between 27 grid cells, inferred with constant external fields from dataset of 65 cells vs. inferred from a subset with grid cells only.

5.1.3 Model quality

The apparent significance of the couplings is further supported by the observation that the couplings inferred from the first half of the dataset is correlated with the ones inferred from the second half.

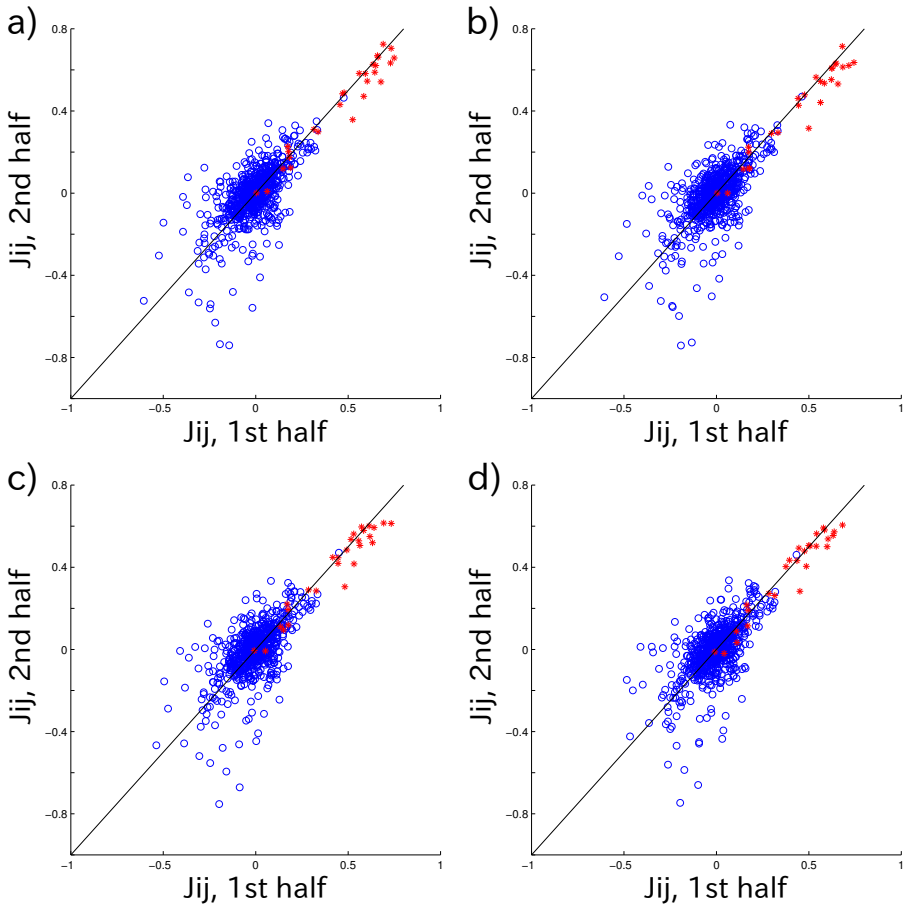

Figure 20: Couplings inferred on first half (10 minutes) of the dataset plotted against the couplings inferred on the second half of the dataset. a) Constant fields, b) 2 x 2 boxes, c) 3 x 3 boxes, d) 4 x 4 boxes.

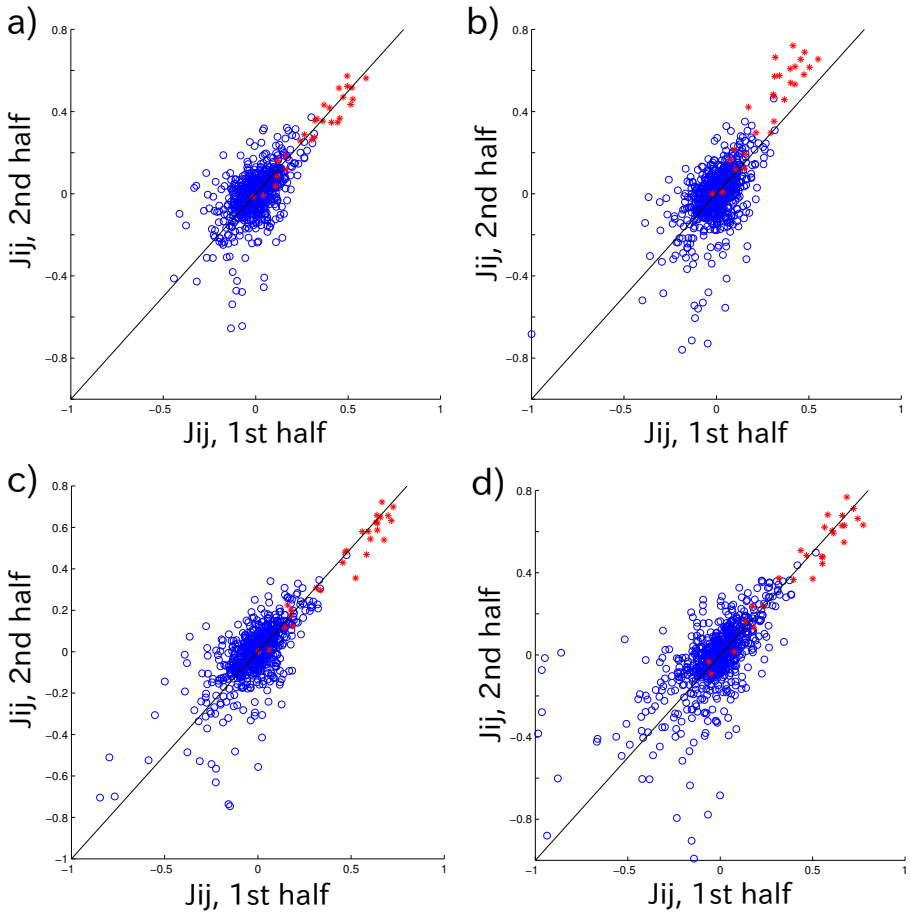

Figure 21: Couplings inferred on first half (10 minutes) of the dataset plotted against the couplings inferred on the second half of the dataset. a) circle fields, b) Gaussian fields, c) Velocity fields, and d) Theta fields.

We note that the behaviour of the animal is different in the two halves of the datasets: it is more active in the first half (see Fig. 22). This may to some extent explain the fact that the couplings inferred from the first and second half is not too similiar. A trend in the difference between the inferred couplings from the two halves is that some couplings are stronger in the second half. This is particularly evident in the model with

Gaussian fields (Fig. 21b.)

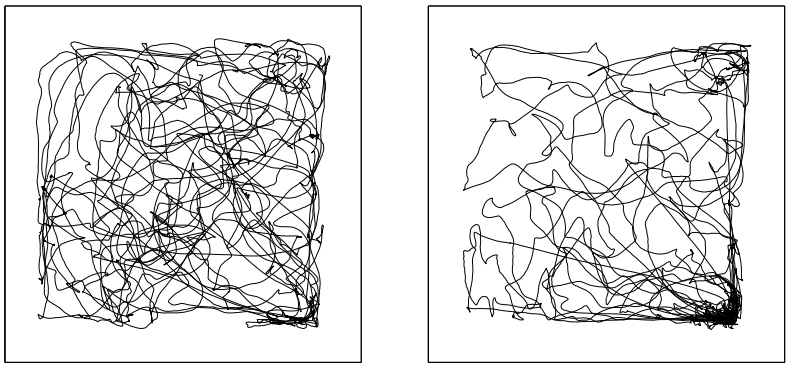

Figure 22: Trajectory of the animal in the first half (left) and second half (right).

When comparing model quality in terms of log-likelihood and Akaike-corrected[4] log-likelihood (Fig. 23-24), it is evident that including couplings in the models outperforms the models without couplings. This can be seen as a decrease in likelihood along the x-axis. Fig. 23 also shows (along the y-axis) that the quality of the models increase with increasing 'spatial resolution' of the external fields. It further shows that allowing spatial variation of the external fields improves the model fit more than velocity and theta fields.

[4]The Akaike correction is a correction penalizing models with many parameters to avoid over-fitting. The correction consists of subtracting the number of parameters [80].

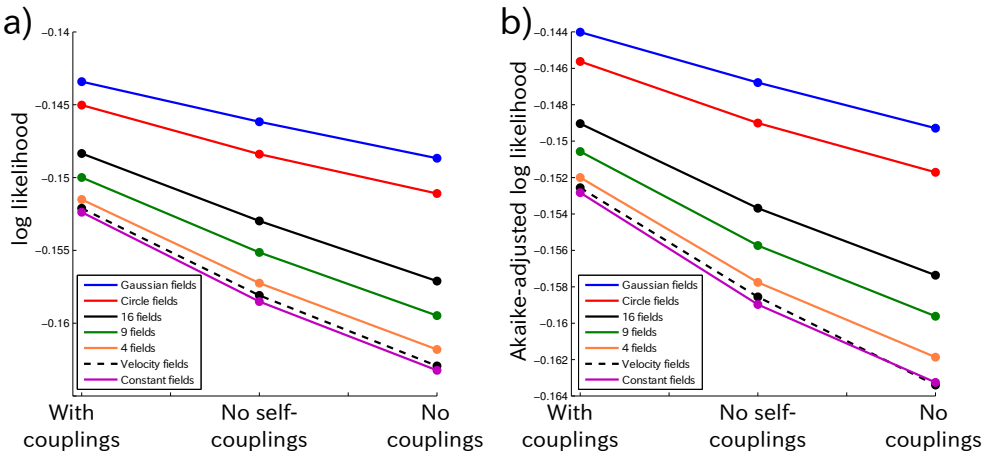

Figure 23: a) Log-likelihood per neuron per time bin for all models. b) Akaike-corrected log-likelihood per neuron per time bin for all models. Blue: Gaussian fields, red: Circle fields, solid black: 16 fields, green: 9 fields, orange: 4 fields, dotted black: velocity fields, magenta: constant fields.

In most of our analysis, we used 20 ms timebins. For the theta model, on the other hand, we chose 10 ms time bins to get a better approximation to the theta phase. Since we used smaller time bins for the model with theta fields, the model fit of this model is plotted separately below.

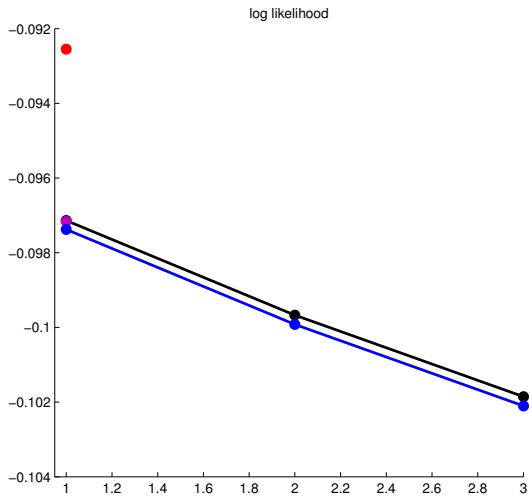

Figure 24: Log-likelihood per neuron per time bin for the model with theta fields (black), and Akaike-corrected log-likelihood for the model with theta fields (blue), both of which inferred with the spike data binned in 10 ms timebins. For comparison, the log-likelihood per neuron per time bin with 10 ms time bins is shown for the model with constant fields (magenta), and the model with circle fields (red).

5.2 Modules and connectivity

We asked how the inferred couplings relate to experimental results, and to what has been suggested in grid cell models. First, we compared the coupling strength within and between modules. One would expect, based on the experimental indications that modules can operate independently, that the coupling strength within modules is stronger than that between modules. To investigate this, we calculated the spacing and orientation of the cells. Based on this, we were able to separate three clusters of cells (see Fig. 25).

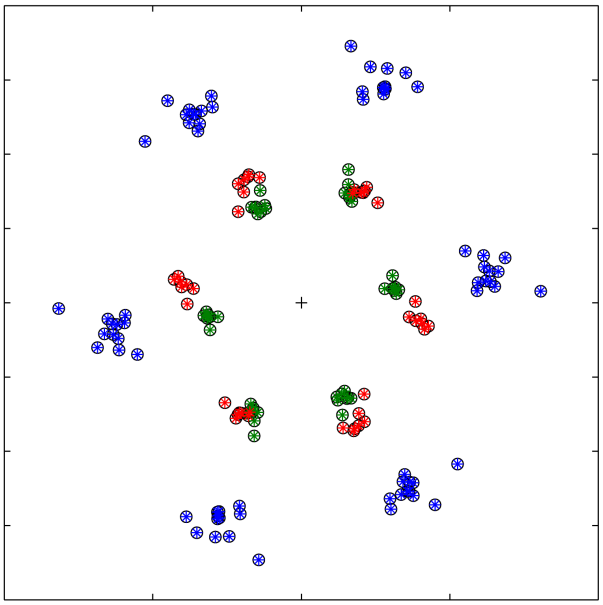

Figure 25: Center of mass of the six peaks nearest the center in the autocorrelograms for the 27 grid cells. Green: Module 1 (8 cells). Red: Module 2 (7 cells). Blue: Module 3 (12 cells).

We assumed that these clusters represent three modules. The comparison of coupling strength between and within modules is shown in Fig. 26:

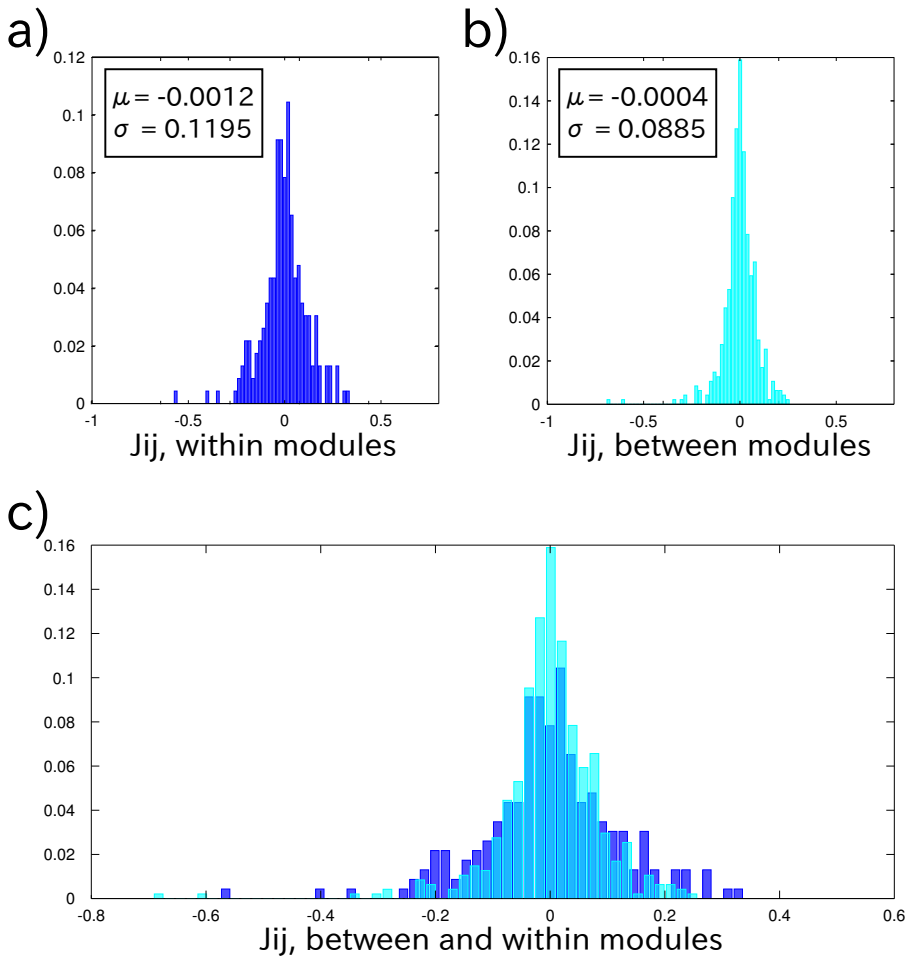

Figure 26: Histograms of couplings within (a) and between (b) modules (top). For clarity, the histograms are put on top of each other (bottom).

The histograms in Fig. 26 indicate that the inferred couplings are stronger within modules than between modules, supporting the hypothesis that each module represents a strongly connected network.

5.3 Spatial phase dependence of couplings

As discussed in the introduction, some attractor models of grid cells have suggested that the connection strength between grid cells should be correlated with spatial phase distance. This motivated us to check if there is a relationship between the inferred couplings and the spatial phase distance between the cells. This relationship is shown in Fig. 27:

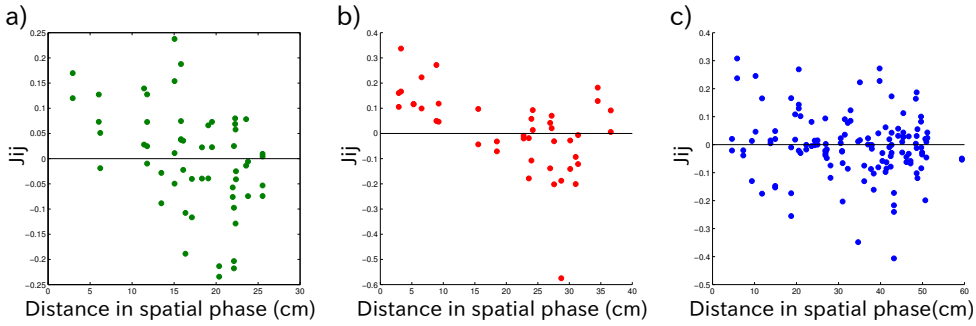

Figure 27: Couplings vs. phase distance in the three modules. a) Module 1 (8 cells). b) Module 2 (7 cells). c) Module 3 (12 cells). Colors correspond to colors in Fig. 25.

The relationship between coupling strength and spatial phase distance shown in Fig. 27 resembles a Mexican hat-like functional connectivity, with similar amounts of excitation and inhibition, contradicting experimental evidence of an all inhibitory network [44, 45]. The functional connections we have inferred, however, do not show any trend towards being negative. We will get back to this in the discussion.

5.4 Theta modulation

Since almost all the cells seem to have some theta phase preference, the theta oscillations could be a source of signal correlation. We looked for a relationship between preferred phase of theta and coupling strength. The cells could be clustered in two separate groups, with preference for opposite phases of theta. A histogram of couplings within and between these two 'theta clusters' are shown in Fig. 28.

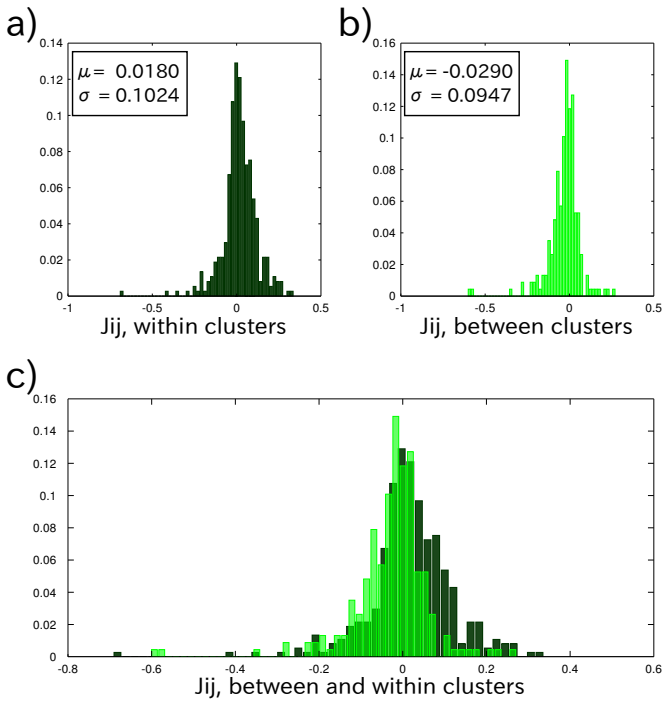

Figure 28: Histograms of couplings inferred with an external spatial Gaussian field within (a) and between (b) clusters of cell that have opposite theta phase preference (top). For clarity, the histograms are put on top of each other (bottom).

There is a trend in the inferred couplings of these clusters: the histogram of couplings between cells that prefer the same phase of theta is skewed towards positive values, whereas the opposite is true for couplings between cells that show preference to opposite phase of theta. This observation led us to construct the theta phase model. However, as shown in Fig. 29, this effect is not removed by adding an external theta field:

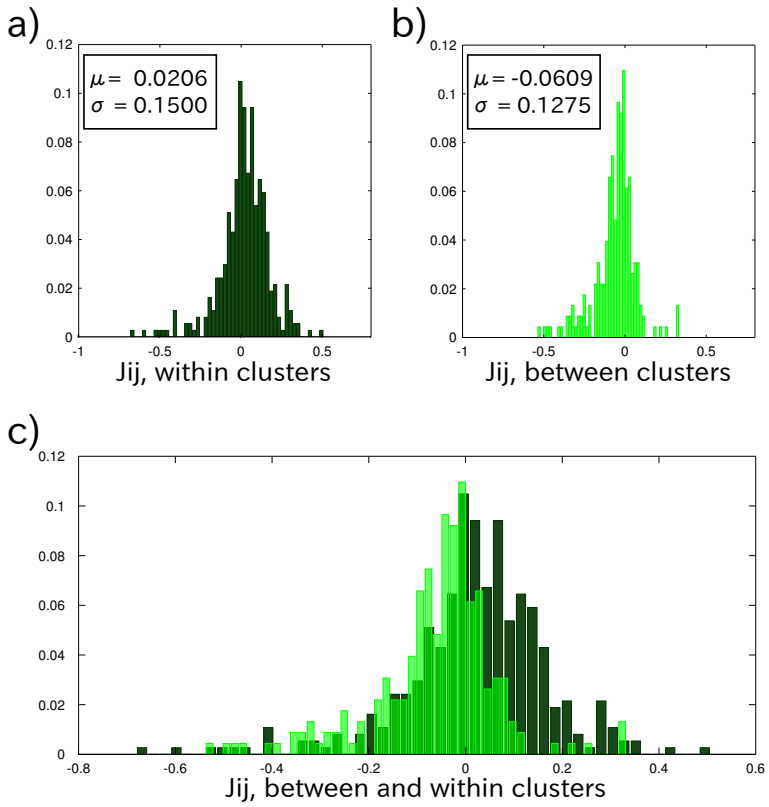

Figure 29: Histograms of couplings inferred with an external theta field within (a) and between (b) clusters of cell that have opposite theta phase preference (top). For clarity, the histograms are put on top of each other (bottom).

Adding a theta field does not remove the observed relationship between phase preference and sign of couplings. However, our attempt at modeling theta input was limited in that we have to use the average theta phase over 10 ms.

6 Discussion and Conclusion

In this project, the kinetic Ising model was used to reverse engineer the functional connectivity of a network of 27 grid cells recorded *in vivo*. To take the spatial firing variations into account, several versions of the model was developed by modifying the external field of the model. Components with different spatial resolution was added to the external field. In addition, we developed a model with a velocity field, and one with LFP theta fields, both of which are factors suggested to influence the activity of grid cells.

In all models, including the couplings improved the model quality. In other words, any model with couplings included did better at explaining the statistics of the dataset than the same model without couplings.

We found that the couplings remain stable across models. We asked if this could be caused by the positive self-couplings, which, as previously discussed, can explain some of the spatial variation in firing of the cells. Removing the self-couplings, however, did not affect the couplings between cells. Inferring on the entire dataset (65 cells) did not affect the couplings between grid cells either.

Further support for the significance of the couplings comes from the observation that the couplings inferred from the first half of the dataset are correlated with those inferred from the second half. We noted that the behaviour of the animal differs in the two halves of the dataset. The couplings tend to be weaker in the first half than in the second half (see Fig. 20-21; especially evident in Gaussian fields model (Fig. 21b)). Considering the firing of grid cells, this is not necessarily surprising. In the second half, the animal more frequently visits the lower left corner in the second half. If the animal stays more or less in the same location, one would expect the grid cells with fields in that location to fire, while the grid cells

that do not have a field in that location will likely be mostly silent, resulting in a stronger correlation between the active and between the silent cells.

The strongest effect on the couplings seems to be increasing the resolution of the spatial external fields. When we allow for spatially varying external fields, the strength of the couplings decreases with increased spatial resolution, and the model fit increases. At the maximum spatial resolution - the Gaussian field model - including the couplings still improves the quality of the model.

The stability of the couplings could be caused by an unobserved correlated input to the cells. As Tyrcha et al. (2013) [65] points out, including the known stimulus in the model, as in [67], will account for signal-correlations, while to account for all trial-to-trial-consistent input to the cells from the unobserved part of the network, one should infer with a 'full resolution' time-varying field, as was done in [65]. This requires several repeats of the same trial, which is difficult to make for an animal exploring a 2D environment.

The fact that the couplings are stable across models, and that they improve the quality of all models, indicates that they are significant. This significance can, as we just discussed, stem from that they explain unobserved correlated input to the cells. A candidate for such input is the LFP theta phase preference, which we found to be related to coupling strength. Our attempt at adding LFP theta to the external field did not remove this effect, however. Another possible reason that the inferred coupling are so stable is that they are related to the real functional connectivity of the network.

When comparing couplings within vs. between modules, we found indications of stronger couplings within than between modules. This is consistent with what one would expect from experimental indications that

modules can operate independently.

Our results also show indications of a relationship between coupling strength and spatial phase distance between firing fields. This supports the continuous attractor models, all of which assume some relationship between spatial phase distance and connectivity. In Fig. 27, where the coupling strength is plotted against phase distance, there seems to be a trend towards a 'Mexican hat'-like connectivity - in contrast to the recurrent inhibition found experimentally between the main grid cell candidate, namely the stellate cells of layer II. However, evidence exist that pyramidal cells of deeper layers also have grid pattern. Between these cells, direct excitatory connections have been indicated. It is not possible to confirm whether the grid cells in our dataset are stellate cells, but we used the available information to look for indications of what cell types could be present in the data. Considering the tetrode recording locations (see Appendix A.2), many of these cells are probably recorded from deeper layers. In addition, there are some layer-specific functional differences that can indicate what layer a grid cell may or may not belong to. As shown in Appendix A.1, some of the cells show head-direction preference, which is a property rarely observed in the LII population. Many of the cells show theta phase locking (see Appendix A.1) - a property observed in LIII cells. Taken together, this indicates that many of the cells in this dataset might not be stellate cells.

In this dataset, we have not detected an all-inhibitory network. This does not necessarily contradict any experimental findings, since the recurrent inhibition was found between stellate cells of layer II [44], which probably constitutes only a minority of the cells in this dataset. The dataset analyzed here likely contains a majority of grid cells in deeper layers. Our results could indicate that the functional connections between the grid cells of layer III and deeper layers decreases with spatial phase distance.

References

[1] Kandel ER, Schwartz JH, Jessell TM, Siegelbaum SA, and Hudspeth AJ. *Principles of Neural Science, Fifth Edition*. McGraw-Hill medical. McGraw-Hill Education, 2012.

[2] Tolman EC. Cognitive maps in rats and men. *Psychol Rev*, 55(4):189, 1948.

[3] Treves A and Rolls ET. Computational analysis of the role of the hippocampus in memory. *Hippocampus*, 4(3):374–391, 1994.

[4] Andersen P, Morris R, Amaral D, Bliss T, and O'Keefe J. *The hippocampus book*. Oxford University Press, 2006.

[5] Purves D, Augustine G, Fitzpatrick D, Hall W, LaMantia A, McNamara J, and White L. *Neuroscience*. Sinauer Associates, 2008.

[6] Rolls ET. Attractor networks. *WIREs Cogn Sci*, 1(1):119–134, 2010.

[7] Scoville WB and Milner B. Loss of recent memory after bilateral hippocampal lesions. *J Neurol Neurosurg Psychiatry*, 20(1):11–21, 1957.

[8] Douglas RJ. The hippocampus and behavior. *Psychol Bull*, 67(6): 416, 1967.

[9] Kaada BR, Rasmussen EW, and Kveim O. Effects of hippocampal lesions on maze learning and retention in rats. *Exp Neurol*, 3(4): 333–355, 1961.

[10] Kveim O, Setekleiv J, and Kaada B. Differential effects of hippocampal lesions on maze and passive avoidance learning in rats. *Exp Neurol*, 9(1):59–72, 1964.

[11] Thomas GJ and Otis LS. Effects of rhinencephalic lesions on maze learning in rats. *J Comp Physiol Psychol*, 51(2):161, 1958.

[12] O'Keefe J and Dostrovsky J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.*, 34(1):171–5, 1971.

[13] Moser EI, Kropff E, and Moser MB. Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.*, 31:69–89, 2008.

[14] O'Keefe J. A review of the hippocampal place cells. *Prog Neurobiol*, 13(4):419–439, 1979.

[15] O'Keefe J. Place units in the hippocampus of the freely moving rat. *Exp Neurol*, 51(1):78–109, 1976.

[16] Rotenberg A, Mayford M, Hawkins RD, Kandel ER, and Muller RU. Mice expressing activated CaMKII lack low frequency LTP and do not form stable place cells in the CA1 region of the hippocampus. *Cell*, 87(7):1351–1361, 1996.

[17] McHugh TJ, Blum KI, Tsien JZ, Tonegawa S, and Wilson MA. Impaired hippocampal representation of space in CA1-specific NMDAR1 knockout mice. *Cell*, 87(7):1339–1349, 1996.

[18] Ulanovsky N and Moss CF. Hippocampal cellular and network activity in freely moving echolocating bats. *Nat Neurosci*, 10(2):224–233, 2007.

[19] Hori E, Tabuchi E, Matsumura N, Tamura R, Eifuku S, Endo S, Nishijo H, and Ono T. Representation of place by monkey hippocampal neurons in real and virtual translocation. *Hippocampus*, 13(2):190–196, 2003.

[20] Ekstrom AD, Kahana MJ, Caplan JB, Fields TA, Isham EA, Newman EL, and Fried I. Cellular networks underlying human spatial navigation. *Nature*, 425(6954):184–188, 2003.

[21] Vanderwolf CH. Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalogr Clin Neurophysiol*, 26(4): 407–418, 1969.

[22] O'Keefe J and Recce ML. Phase relationship between hippocampal place units and the eeg theta rhythm. *Hippocampus*, 3(3):317–330, 1993.

[23] O'Keefe J and Conway D. Hippocampal place units in the freely moving rat: why they fire where they fire. *Exp Brain Res*, 31(4):573–90, 1978.

[24] Leutgeb S, Leutgeb JK, Treves A, Moser MB, and Moser EI. Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science*, 305 (5688):1295–1298, 2004.

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

[26] Quirk GJ, Muller RU, Kubie JL, and Ranck Jr JB. The positional firing properties of medial entorhinal neurons: description and comparison with hippocampal place cells. *J Neurosci*, 12(5):1945–1963, 1992.

[27] Frank LM, Brown EN, and Wilson M. Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron*, 27(1):169–178, 2000.

[28] McNaughton B, Barnes C, Meltzer J, and Sutherland R. Hippocampal granule cells are necessary for normal spatial learning but not for spatially-selective pyramidal cell discharge. *Exp Brain Res*, 76(3): 485–496, 1989.

[29] Brun VH, Otnæss MK, Molden S, Steffenach HA, Witter MP, Moser MB, and Moser EI. Place cells and place recognition maintained by direct entorhinal-hippocampal circuitry. *Science*, 296(5576):2243–2246, 2002.

[30] Fyhn M, Molden S, Witter MP, Moser EI, and Moser MB. Spatial representation in the entorhinal cortex. *Science*, 305(5688):1258–64, 2004.

[31] Hafting T, Fyhn M, Molden S, Moser M, and Moser E. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–6, 2005.

[32] Neves G, Cooke SF, and Bliss TV. Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. *Nat Rev Neurosci*, 9(1):65–75, 2008.

[33] Sargolini F, Fyhn M, Hafting T, McNaughton BL, Witter MP, Moser MB, and Moser EI. Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science*, 312(5774):758–62, 2006.

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

[35] Solstad T, Boccara CN, Kropff E, Moser MB, and Moser EI. Representation of geometric borders in the entorhinal cortex. *Science*, 322 (5909):1865–1868, 2008.

[36] Zhang SJ, Ye J, Miao C, Tsao A, Cerniauskas I, Ledergerber D, Moser MB, and Moser EI. Optogenetic dissection of entorhinal-hippocampal functional connectivity. *Science*, 340(6128), 2013.

[37] Fyhn M, Hafting T, Treves A, Moser MB, and Moser EI. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, 446 (7132):190–194, 2007.

[38] Hafting T, Fyhn M, Bonnevie T, Moser MB, and Moser EI. Hippocampus-independent phase precession in entorhinal grid cells. *Nature*, 453(7199):1248–1252, 2008.

[39] Köhler C. Intrinsic connections of the retrohippocampal region in the rat brain. ii. the medial entorhinal area. *J Comp Neurol*, 246(2): 149–169, 1986.

[40] Kloosterman F, Van Haeften T, Witter MP, and Lopes da Silva FH. Electrophysiological characterization of interlaminar entorhinal connections: an essential link for re-entrance in the hippocampal–entorhinal system. *Eur J Neurosci*, 18(11):3037–3052, 2003.

[41] van Haeften T, Baks-te Bulte L, Goede PH, Wouterlood FG, and Witter MP. Morphological and numerical analysis of synaptic interactions between neurons in deep and superficial layers of the entorhinal cortex of the rat. *Hippocampus*, 13(8):943–952, 2003.

[42] Canto CB, Wouterlood FG, and Witter MP. What does the anatomical organization of the entorhinal cortex tell us? *Neural Plast*, 2008, 2008.

[43] Dhillon A and Jones RS. Laminar differences in recurrent excitatory transmission in the rat entorhinal cortex *in vitro*. *Neuroscience*, 99 (3):413–422, 2000.

[44] Couey JJ, Witoelar A, Zhang SJ, Zheng K, Ye J, Dunn B, Czajkowski R, Moser MB, Moser EI, Roudi Y, and Witter MP. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nat Neurosci*, 16(3): 318–24, 2013.

[45] Pastoll H, Solanka L, van Rossum MC, and Nolan MF. Feedback inhibition enables theta-nested gamma oscillations and grid firing fields. *Neuron*, 77(1):141–154, 2013.

[46] Burgalossi A, Herfst L, von Heimendahl M, Förste H, Haskic K, Schmidt M, and Brecht M. Microcircuits of functionally identified neurons in the rat medial entorhinal cortex. *Neuron*, 70(4):773–86, 2011.

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

[48] Fuhs MC and Touretzky DS. A spin glass model of path integration in rat medial entorhinal cortex. *J Neurosci*, 26(16):4266–4276, 2006.

[49] McNaughton BL, Battaglia FP, Jensen O, Moser EI, and Moser MB. Path integration and the neural basis of the'cognitive map'. *Nat Rev Neurosci*, 7(8):663–678, 2006.

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

[51] Giocomo LM, Zilli EA, Fransén E, and Hasselmo ME. Temporal frequency of subthreshold oscillations scales with entorhinal grid cell field spacing. *Science*, 315(5819):1719–1722, 2007.

[52] Yartsev MM, Witter MP, and Ulanovsky N. Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature*, 479(7371):103–107, 2011.

[53] Stringer SM, Rolls ET, Trappenberg TP, and de Araújo IET. Self-organizing continuous attractor networks and motor function. *Neural Netw*, 16(2):161–182, 2003.

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

[55] Seung HS. How the brain keeps the eyes still. *Proc Natl Acad Sci U S A*, 93(23):13339–13344, 1996.

[56] Navratilova Z, Giocomo LM, Fellous JM, Hasselmo ME, and McNaughton BL. 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):772–789, 2012.

[57] Stringer S, Rolls ET, Trappenberg T, and De Araujo I. Self-organizing continuous attractor networks and path integration: two-dimensional models of place cells. *Network*, 13(4):429–446, 2002.

[58] Stringer S, Trappenberg T, Rolls ET, and Araujo I. Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells. *Network*, 13(2): 217–242, 2002.

[59] Roudi Y and Latham PE. A balanced memory network. *PLoS Comput Biol*, 3(9):1679–700, 2007.

[60] Wills TJ, Lever C, Cacucci F, Burgess N, and O'Keefe J. Attractor dynamics in the hippocampal representation of the local environment. *Science*, 308(5723):873–876, 2005.

[61] Romani S and Tsodyks M. Continuous attractors with morphed/correlated maps. *PLoS Comput Biol*, 6(8):e1000869, 2010.

[62] Bonnevie T, Dunn B, Fyhn M, Hafting T, Derdikman D, Kubie J, Roudi Y, Moser E, and Moser MB. Grid cells require excitatory drive from the hippocampus. *Nat Neurosci*, 16(3):309–17, 2013.

[63] Lezon TR, Banavar JR, Cieplak M, Maritan A, and Fedoroff NV. Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc Natl Acad Sci U S A*, 103(50):19033–19038, 2006.

[64] Weigt M, White RA, Szurmant H, Hoch JA, and Hwa T. Identification of direct residue contacts in protein–protein interaction by message passing. *Proc Natl Acad Sci U S A*, 106(1):67–72, 2009.

[65] Tyrcha J, Roudi Y, Marsili M, and Hertz J. The effect of nonstationarity on models inferred from neural data. *arXiv:1203.5673 [q-bio.QM]*, 2013.

[66] Nelder JA and Wedderburn RWM. Generalized linear models. *J R Stat Soc*, pages 370–384, 1972.

[67] Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky E, and Simoncelli EP. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

[68] Truccolo W, Eden UT, Fellows MR, Donoghue JP, and Brown EN. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol*, 93(2):1074–1089, 2005.

[69] Paninski L. Maximum likelihood estimation of cascade point-process neural encoding models. *Network*, 15(4):243–262, 2004.

[70] Chichilnisky E. A simple white noise analysis of neuronal light responses. *Network*, 12(2):199–213, 2001.

[71] Schneidman E, Berry II MJ, Segev R, and Bialek W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–12, 2006.

[72] Ising E. Beitrag zur theorie des ferromagnetismus. *Z Phys A-Hadron Nucl*, 31(1):253–258, 1925.

[73] Reif F and Reif F. *Fundamentals of statistical and thermal physics*, volume 1. McGraw-Hill New York, 1965.

[74] Hertz J, Roudi Y, and Tyrcha J. Ising models for inferring network structure from spike data. *arXiv:1106.1752 [q-bio.QM]*, 2011.

[75] Roudi Y, Nirenberg S, and Latham PE. Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't. *PLoS Comput Biol*, 5(5):e1000380, 2009.

[76] Glauber RJ. Time-dependent statistics of the ising model. *J Math Phys*, 4:294, 1963.

[77] Roudi Y and Hertz J. Mean field theory for nonequilibrium network reconstruction. *Phys Rev Lett*, 106(4):048702, 2011.

[78] Granot-Atedgi E, Tkačik G, Segev R, and Schneidman E. Stimulus-dependent maximum entropy models of neural population codes. *PLoS Comput Biol*, 9(3), 2013.

[79] Batschelet E. *Circular statistics in biology*. Academic Press London, 1981.

[80] Akaike H. A new look at the statistical model identification. *IEEE T Automat Contr*, 19(6):716–723, 1974.

A Appendix

A.1 Functional characteristics of the grid cells

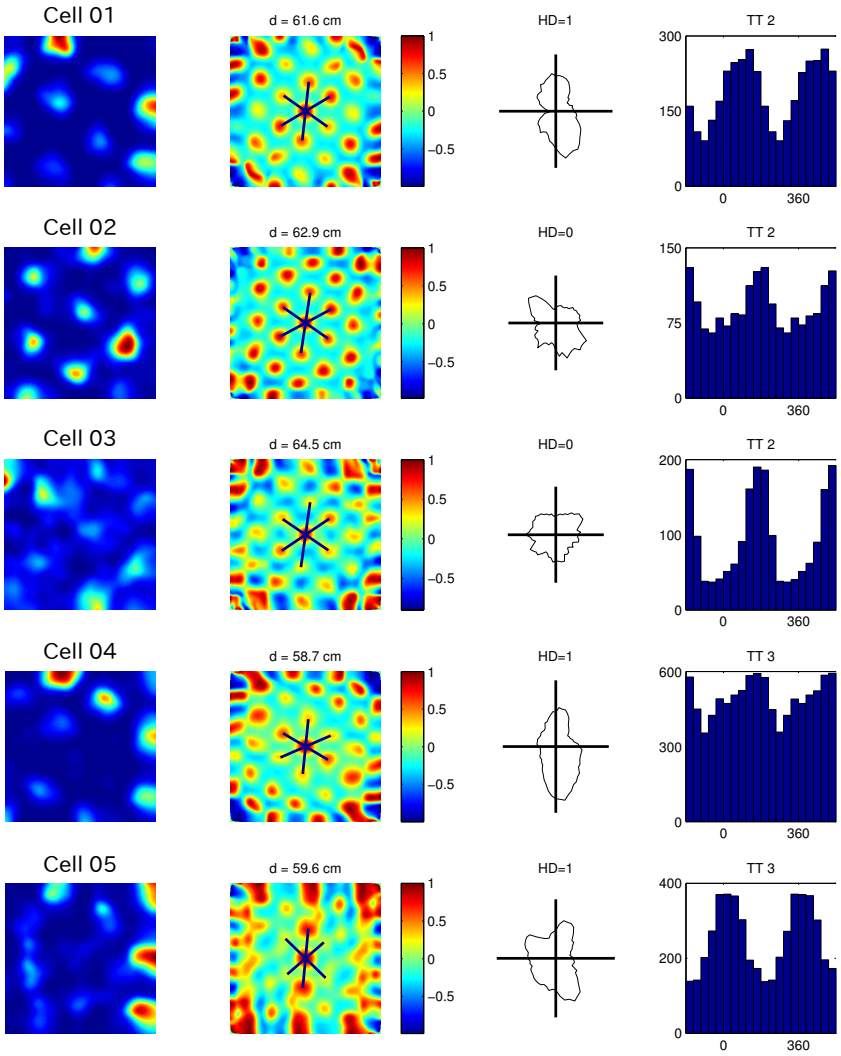

Figure 30: Ratemaps (column 1), autocorrelogram (column 2), head directional tuning (column 3) and distribution of spikes within the theta cycle (column 4) for cell 1-5.

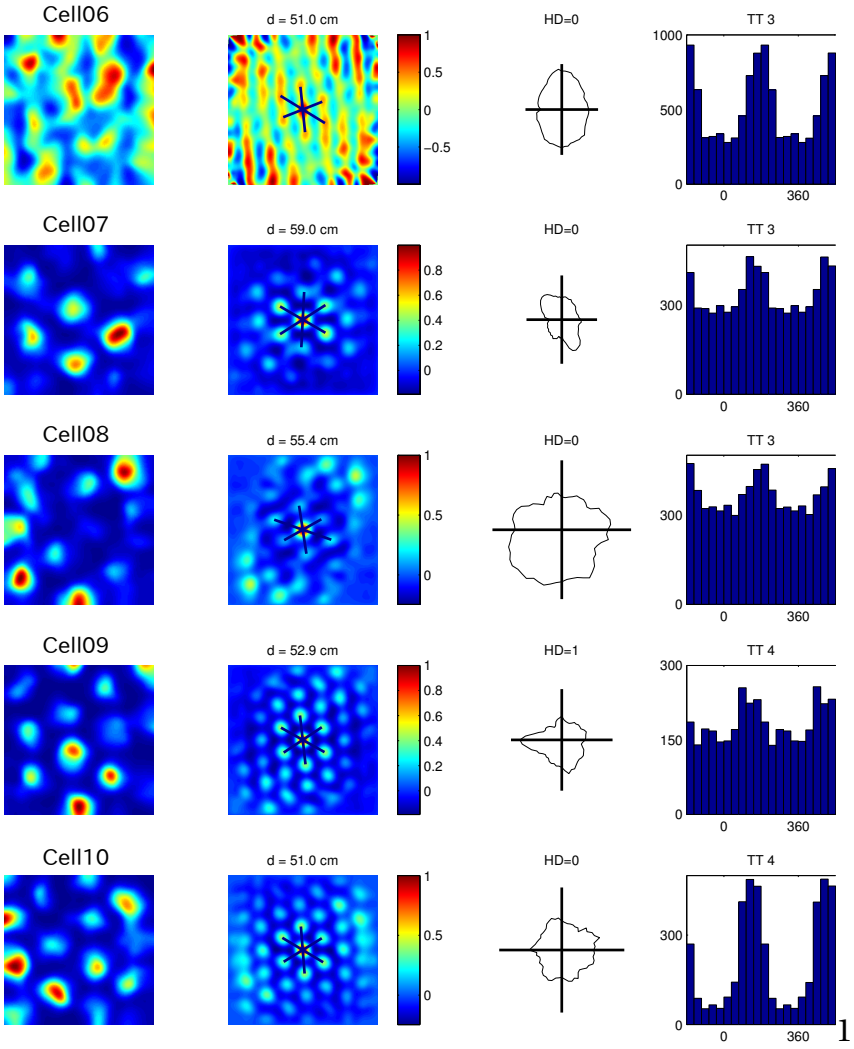

Figure 31: Ratemaps (column 1), autocorrelogram (column 2), head directional tuning (column 3) and distribution of spikes within the theta cycle (column 4) for cell 6-10.

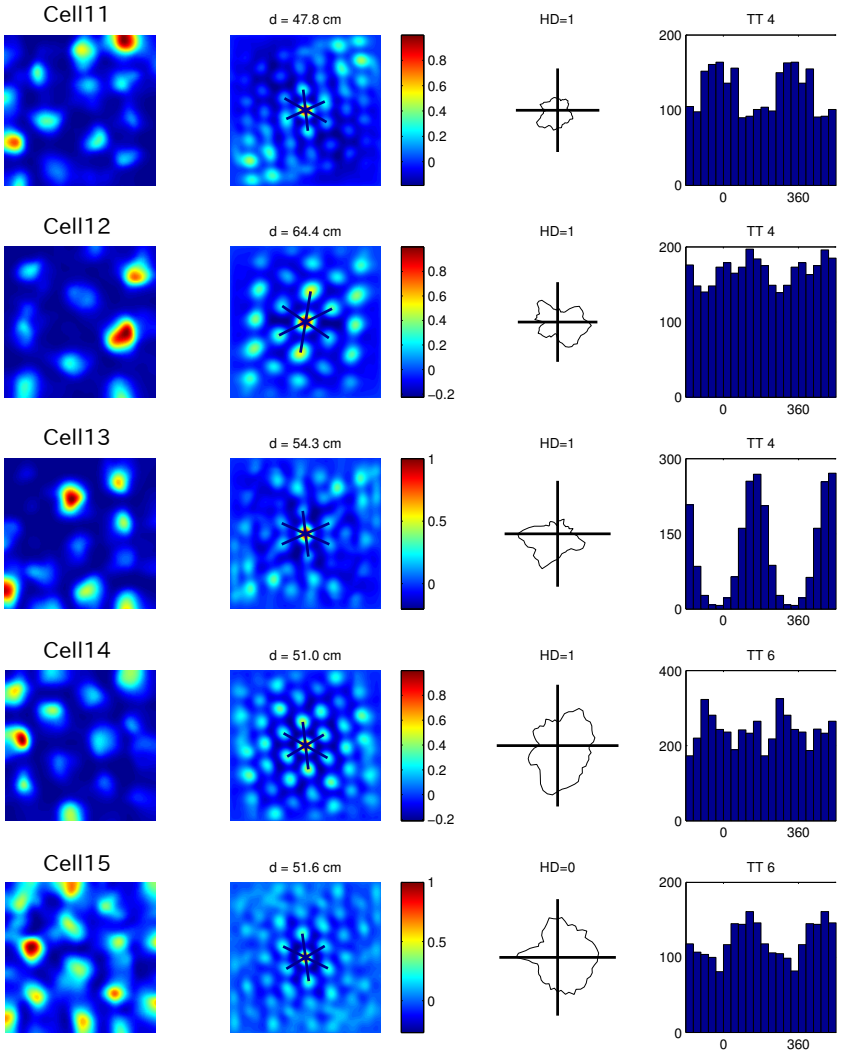

Figure 32: Ratemaps (column 1), autocorrelogram (column 2), head directional tuning (column 3) and distribution of spikes within the theta cycle (column 4) for cell 11-15.

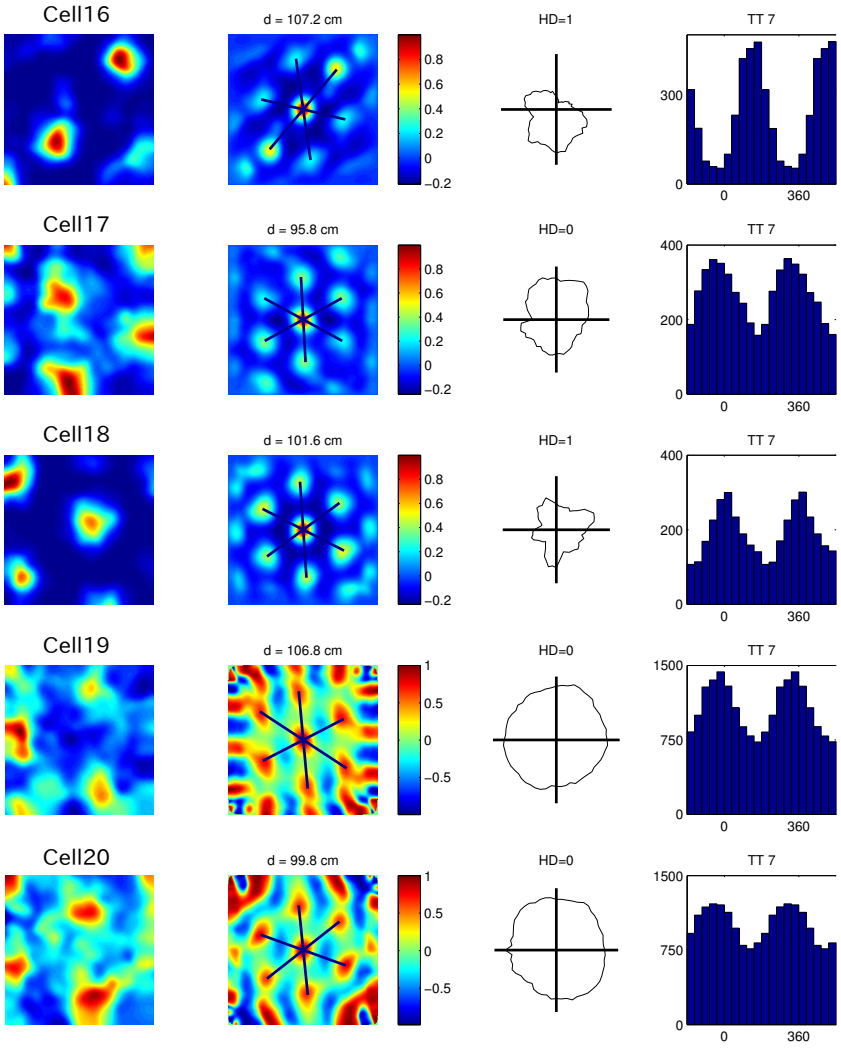

Figure 33: Ratemaps (column 1), autocorrelogram (column 2), head directional tuning (column 3) and distribution of spikes within the theta cycle (column 4) for cell 16-20.

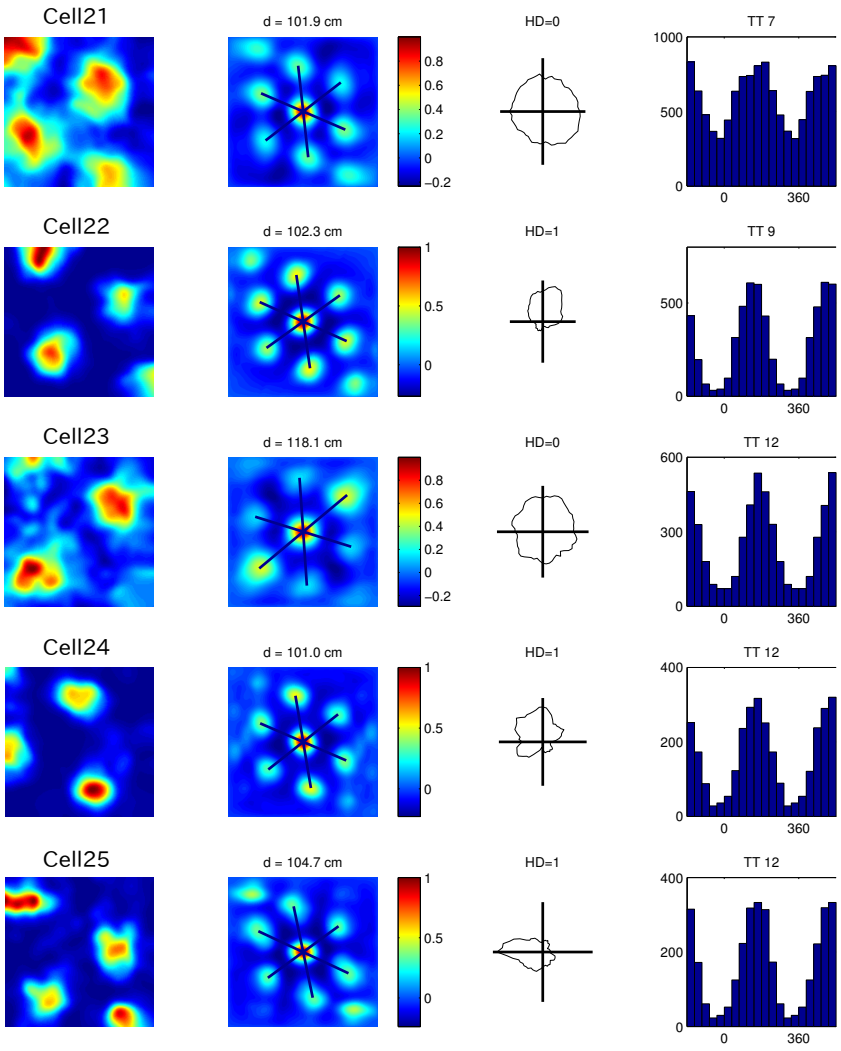

Figure 34: Ratemaps (column 1), autocorrelogram (column 2), head directional tuning (column 3) and distribution of spikes within the theta cycle (column 4) for cell 21-25.

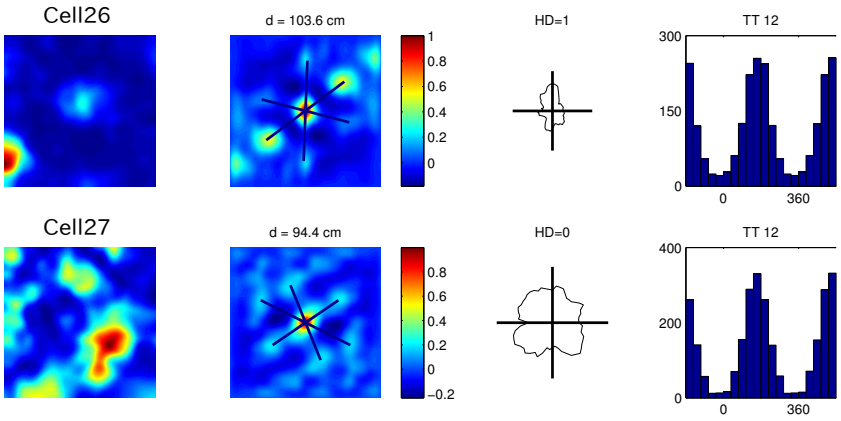

Figure 35: Ratemaps (column 1), autocorrelogram (column 2), head directional tuning (column 3) and distribution of spikes within the theta cycle (column 4) for cell 26-27.

A.2 Tetrode recording locations

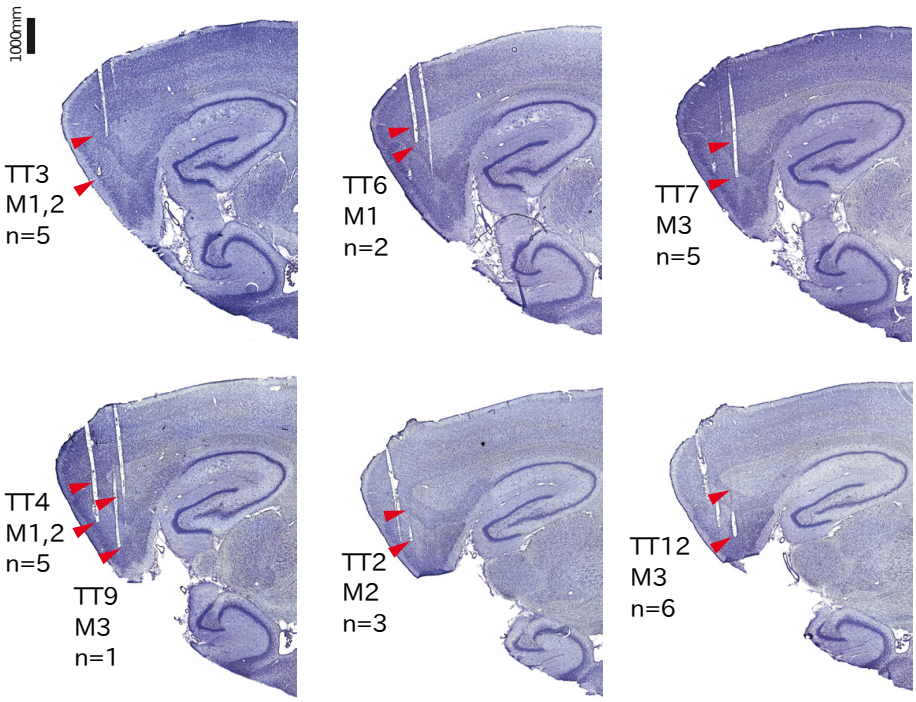

Figure 36: Nissl stained saggital sections showing the extent of grid cell recording locations. The exact location of the tetrode tracks (TT) at the time of the recording of the dataset analyzed in this project is not known, but the dorsoventral extent of the recording locations of grid cells is indicated by red arrowheads. The number of grid cells (n), and number of modules (M), associated with each tetrode track is shown. Adapted from [34].