

---

---

# Interneurons in A Two Population Grid Cell Network

---

AUTHOR

Ziwei Huang

SUPERVISORS

Benjamin Dunn

Trygve Solstad

NORGES TEKNISK-NATURVITENSKAPELIGE  
UNIVERSITET(NTNU)

KAVAL INSTITUTE FOR SYSTEMS NEUROSCIENCE  
AND CENTRE FOR NEURAL COMPUTATION



## Abstract

The hexagonal firing pattern of entorhinal grid cells could arise from a competitive mechanism mediated by interneurons. In this thesis, we asked if a two-population continuous attractor model, consistent with the proposed inhibitory connectivity pattern, could maintain grid cell firing even if interneurons lack spatial periodicity, as was recently observed in a sub-population of entorhinal interneurons.

First, using non-negative matrix factorization, we constructed two-population models with varying numbers of interneurons while maintaining the same effective connectivity between grid cells. Surprisingly, network drift decreased exponentially with the number of assumed interneurons. Networks having less than 5% interneurons were able to accurately path integrate. The resulting connectivity was patterned with each interneuron receiving projections from either many grid cells with similar spatial selectivity or cells that together formed an inverted grid pattern. In both cases, grid cells with inhomogeneous peak firing rates had lower grid scores than the corresponding interneurons.

Second, we considered a network where the connections from grid cells to interneurons were fixed to sparse random values, while back projections were found using NMF. We found that including variation in the fields of the grid cells resulted in aperiodic neurons in this randomly connected model and the spatial selectivity of interneurons decreased dramatically as the variance in grid field firing rates was increased. Although this network produced aperiodic interneurons similar to recordings, a considerably larger proportion of interneurons was required to reach the same level of stability which did not decrease exponentially as in the fully factorized case.

Furthermore, we analyzed 75 grid cells from a single freely behaving animal and found that the degree to which the firing fields vary was possibly consistent with the models in which gridness score decreased as the grid field variation increased. Take together, our result suggested that the firing patterns of interneurons were not limited to the grid or anti-grid pattern but fell somewhere on the spectrum between highly periodic to aperiodic.



## Acknowledgements

Thousand thanks to my supervisors, Benjamin Dunn and Trygve Solstad, for guiding me into the amazing world of theoretical neuroscience. I never could have come this far without your help and guidance.

Also, I would like to thank Daniel Wennberg, Christin Berndtsson, Yixin Tong for all the helpful discussions. Thanks Kavli Institute for Systems Neuroscience / Centre for Neural Computation for offering such a fantastic study and work environment, and funding my trip to COSYNE in United States.

And last, I would like to thank my family and my girlfriend Shanyi for all your love and unconditional supports.



# Contents

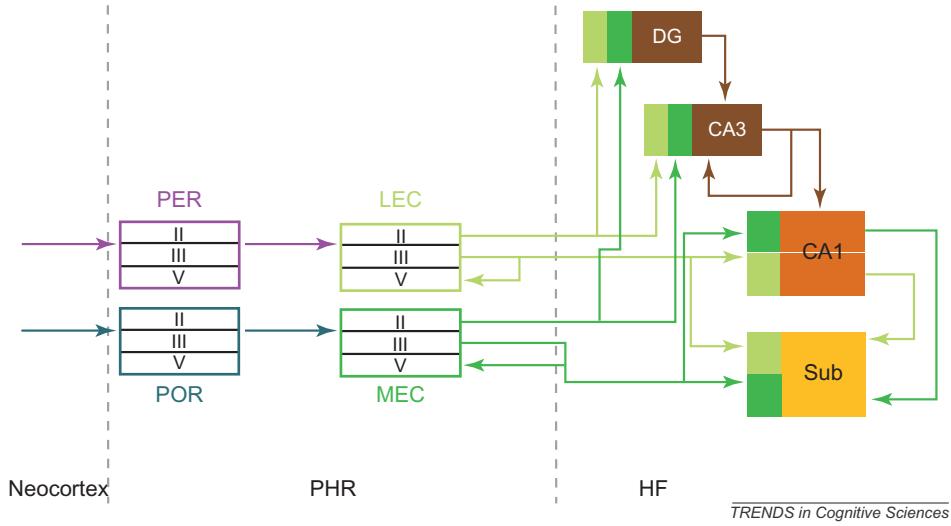
Abstract . . . . .	i
Acknowledgements . . . . .	iii
<b>1 Introduction</b>	<b>1</b>
1.1 Grid cells in entorhinal cortex . . . . .	2
1.2 Attractor network models . . . . .	4
1.3 Interneurons in grid cell network . . . . .	6
<b>2 Two population grid cell model</b>	<b>11</b>
2.1 Nonnegative matrix factorization . . . . .	13
2.2 Fully factorized model and randomly connected model . . . . .	14
<b>3 Methods</b>	<b>17</b>
3.1 Model implementation . . . . .	17
3.2 Data analysis and statistics . . . . .	18
3.2.1 Simulated Rate maps . . . . .	18
3.2.2 Gridness Score . . . . .	18
3.2.3 Spatial Sparsity Score . . . . .	19
3.2.4 Grid field variation . . . . .	19
3.3 Experimental data . . . . .	19
<b>4 Results</b>	<b>23</b>
4.1 Synaptic Connectivity . . . . .	23
4.2 Interneurons in I-surround configuration fully factorized model . . . . .	27

4.3	Interneurons in E-surround configuration fully factorized model . . . . .	31
4.4	Interneurons in randomly connected model . . . . .	35
4.5	Gridness score and grid field variation . . . . .	39
4.6	Grid field variation in experimental data . . . . .	42
4.7	Drift . . . . .	45
4.8	Path Integration . . . . .	46
<b>5</b>	<b>Discussion and Conclusion</b>	<b>51</b>
5.1	How were grid cells and interneurons wired together? . . . . .	52
5.2	Future Directions . . . . .	54
<b>References</b>		<b>57</b>
<b>A</b>	<b>Supplementary Figures</b>	<b>63</b>
A.1	Grid cell data . . . . .	63
A.2	Tetrode recording locations . . . . .	71
<b>B</b>	<b>Publications</b>	<b>73</b>

# 1 | Introduction

It had long been speculated that the mammal navigation was guided by a “cognitive map” [1]. O’Keefe and Dostrovsky discovered the first neural evidence of this map in 1970s: hippocampal place cells [2]. The most distinctive feature of place cells was that each cell fired in one unique location of the environment, and the same cell participated in encoding many different environments. How did place cells emerge? Were place cells responsible for all the spatial computation? These questions motivated neuroscientists to explore the mystery of how space was represented in the brain. O’Keefe speculated that, in order to form a complete cognitive map, each place cell needed to receive at least two different types of spatial input: first, external cues of the environment, and second, a metric system which integrated the animal’s self-motion information [3]. In behavioral level, this metric system was manifested as path integration[4], which was a intrinsic capacity of animals to keep track of their moving distance and angular motion, even in the situation without visual or other sensory inputs.

Thirty years later, Edvard Moser and May-Britt Moser discovered grid cells in the dorsal medial part of entorhinal cortex (dMEC) [5], which is the anatomical upstream of the hippocampus and project directly to CA3 (Figure 1.1). Each grid cell had multiple firing fields, forming a hexagonal pattern covering the whole arena the animal explored. The multiple firing locations for each grid cell were independent of both the velocity of the animal and the external/environmental cues, e.g. when the animal explored in the dark, suggesting that the generation of these cells activity is an intrinsic network phenomenon and could be maintained by self-motion information alone [5]. Subsequent research revealed that grid cells in the layer II of MEC processed directional signal from grid × head direction conjunctive cells in deeper layers of MEC [6], boundary



**Figure 1.1: Anatomical connections between hippocampus and entorhinal cortex.** **PER:** perirhinal cortex, **POR:** postrhinal cortex, **LEC:** lateral entorhinal cortex, **MEC:** entorhinal cortex, **DG:** dentate gyrus, **Sub:** subiculum. Adapted from [9].

information from border cells [7] as well as velocity input from speed cells [8], suggesting that grid cells were probably play an important role in path integration.

## 1.1 Grid cells in entorhinal cortex

There were three basic properties of grid cells (Figure 1.2): grid spacing (the shortest distance between two grid fields), grid orientation (the smallest angle of grid axes compared to an reference axis of the environment), and grid phase (the Cartesian locations of firing vertices relative to the reference location of the environment).

Grid cells recorded from the same tetrode tended to share the same spacing and orientation, and the spacing increased along the dorso-ventral axis of MEC [5]. Experimental data showed that this increase was stepwise instead of continuous [10], and could be clustered into 4 or 5 modules. Moreover, cells recorded from the same tetrode tended to have different phases, i.e. shifted away from the x-y reference coordinate independently, and without any topological properties.

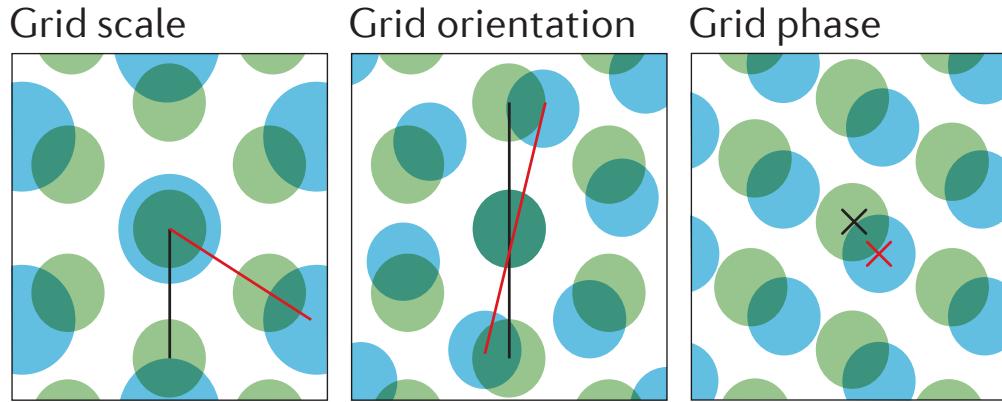


Figure 1.2: **Basic properties of grid cells.** Firing patterns of a pair of grid cells, indicating the difference between grid spacing, grid orientation and grid phase. Adapted from [11].

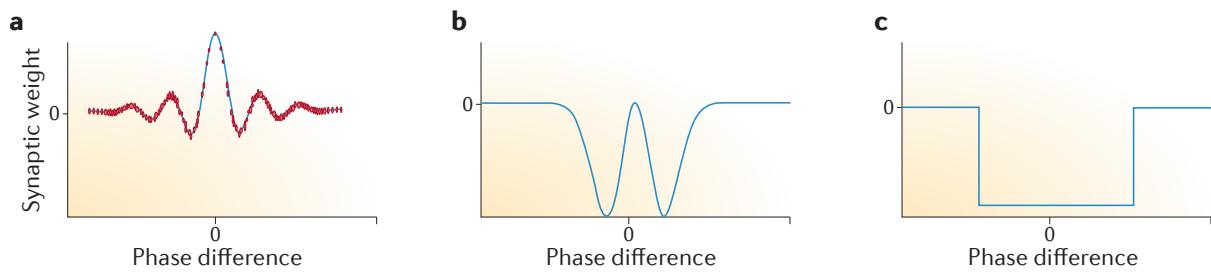


Figure 1.3: **Attractor models of grid cells with different connectivity patterns.** Left: Mexican-hat connectivity; Center: Mexican-hat like inhibitory connectivity; Right: Lincoln-hat inhibitory connectivity. Adapted from [11].

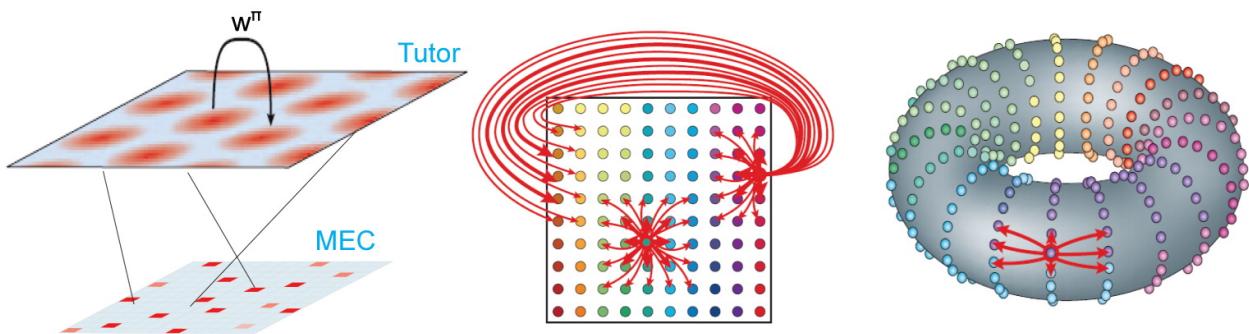


Figure 1.4: **Toroidal synaptic matrix of grid cells.** Left: The periodic tutor network trained a MEC network without topological arrangement. Center and right: a synaptic matrix with edges wrapped around to the opposite sides and formed as a torus. Adapted from [12].

## 1.2 Attractor network models

Since the discovery of grid cells, different theoretical models have been proposed to explain the generation of this phenomenon [13, 14, 15]. One major class of models called “continuous attractor network (CAN) models” emphasized that the hexagonal pattern of grid cells was a network phenomenon generated within the entorhinal cortex through local interaction.

One of the earliest model proposed by Fuhs and Touresky [16] suggested a topologically organized network where cells were arranged according to their phases similarity, that is, they assumed nearby grid cells in the brain shared similar or slightly shifted grid phases. Each cell received excitatory input from those nearby cells with similar phases, while the connection weights decayed progressively as the phases difference increased. A global inhibition applied to all cells in the network prevented the excitation from spreading, resulting in the formation of multiple “bumps” of activity arranged hexagonally on the neural sheet, so it was called “Mexican-hat connectivity” (See Figure 1.3a). These bumps were then moved by integrating head-directional and speed-dependent input from a “hidden layers” as the animal navigated through the environment, and this network activity would be reflected in each grid cell as the hexagonal pattern in the firing rate map. This model had two problems: first, the topologically organization of cells was at odds with the experimental observation that nearby grid cells rarely shared similar phases [5], and the second ones was known as the boundary problem, that is, the planar representation of space was limited by the number of neurons on the neural sheet, in other words, the finite number of cells in the network was far from enough to encode the almost infinite amount of locations in a real life situation.

To address these problems, a revised model was proposed by McNaughton et al. [12]. In this model, nearby grid cells in MEC did not necessarily share similar grid phases but those did with similar phases learned to connected to each other through the training of the assumed topographical arranged tutor network during early development. Due to the periodicity of this tutor network, the edges of the synaptic matrix would wrap around to the opposite side, forming a torus and ensuring that the activity of bumps would be periodic (see Figure 1.4). Whether

such a toroidal synaptic matrix or the topographical arranged tutor network exist is, of course, still a matter of debate given the lack of experimental support. A single bump network with such a configuration could produce a hexagonal grid on a twisted torus [17], whereas a multiple bumps network can also work in an untwisted torus. It should also be noted that a network with multiple bumps with aperiodic boundary conditions could perform precise path integration if carefully structured [18].

These Mexican-hat models were in agreement with experimental observations that conjunctive cells with both grid cell and head-direction cell responses [6] were found in the deeper layers of MEC, as were cells with speed modulation [8]. However, the main challenge of these models was the fact that stellate cells in layer II of MEC, where we could find the most confined and largest amount of grid cells [6, 19], lacked excitatory connections between each other almost completely [19, 20].

A pure inhibitory attractor network had been proposed earlier by Burak and Fiete [18], in which cells in the network formed a Mexican-hat like inhibitory connectivity, i.e. cells with similar phases inhibited each other less than those with larger phase difference, whereas inhibition from those far away would progressively go back to zero (See Figure 1.3b). Using pair-recording and optogenetics, Couey et al. [20] demonstrated that almost all functional connections in MEC layer II were inhibitory and the magnitude of inhibition exerted on two simultaneous recording stellate cells seemed to be constant. This finding lead to the inverted Lincoln-hat or all-or-none connectivity model (Figure 1.3c), in which cells exerted the same amount of inhibition to others within the fix radius but not to the cells beyond that.

The benefit of switching to pure inhibitory models, besides from being more biologically compatible, was that they are more computationally stable in perspective of excitatory-inhibition balance [21]. Since the inhibitory feedback was provided by the GABAergic interneurons within the network, there was no need to include a global inhibition, as the one applied in Mexican-hat model, to prevent runaway excitation. But this implementation lead to another problem: pure inhibitory models require external tonic excitation out of entorhinal cortex. Hippocampus would be one of the possible candidates since there is a hippocampal-entorhinal loop where

hippocampus could back project to the deeper layers of MEC (Figure 1.1). Bonnevie et al. [22] supported this assumption by blocking the hippocampal excitation to MEC and observed that grid cells lose its hexagonal pattern gradually and became dominated by the head-directional signal. This finding supplemented the speculation that the feedback projections from hippocampus, which probably contained spatial information such as path-integrator coordinates (mainly from place cells) and specific features of the environment, served as the error-correction mechanism for grid cells [23, 24]. It was known that recurrent connectivity contained intrinsic noise which will lead to unwanted drift in the activity of continuous attractor network [25]. This drift would accumulate over time as the animal explored the environment, resulting in the destruction of the grid pattern. By integrating the spatial information from hippocampus, entorhinal grid cells could reduce the accumulation of error and perform path integration over a long time accurately.

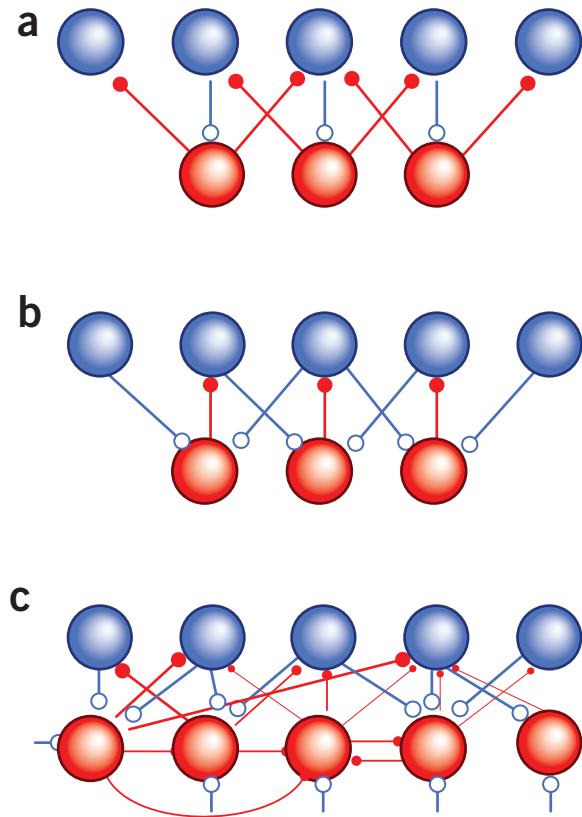
### 1.3 Interneurons in grid cell network

Previous models [20, 18] were called “effective inhibition models” because interneurons in those models were not implemented explicitly, even though they conceptualized that grid cells communicated with other grid cells indirectly by recruiting interneurons within the same network. The effective models ignored the specific configuration of how grid cells and interneurons are wired together, so they lacked the power of predicting what kind of neural activity these interneurons might have.

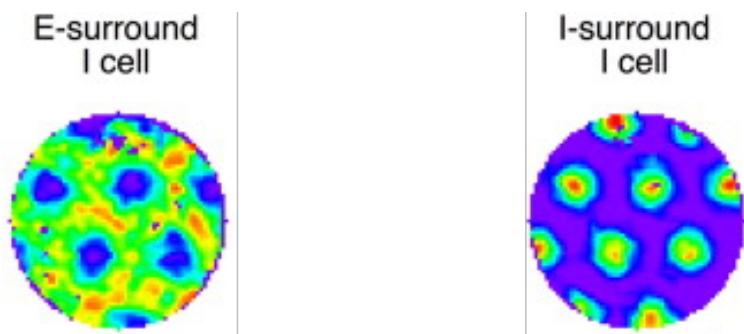
The physiology of GABAergic interneurons in MEC was not that well studied by far, but there was a consensus that this population was diverse in terms of morphology, anatomy and function [26]. The estimated amount of interneurons were around 13% to 16% of total neural population in layer II of MEC [27], and parvalbumin-expressing GABAergic cells ( $PV^+$  neuron) accounted for half of them [28, 29]. These  $PV^+$  neurons formed direct inhibitory synapses onto principal cells and other interneurons in layer II and III [27, 28, 30], suggesting that they were the one of the best candidate for supporting the recurrent inhibitory grid cells network. Calretinin immunore-

active neurons (CR) were another major type of interneurons in MEC, which believed to have no overlap with the PV<sup>+</sup> population [31]. Considerable amount of CR neurons in layer II expressed GABA, though the exact proportion was still in debate [31]. One interesting observation of CR neurons was that they rarely made contact with principal neuron, suggesting that their primary role was to mediate other interneurons [31], as their counterpart in Hippocampus [32]. Layer II of MEC was also rich in other cell types of interneurons expressed different neuropeptides, such as somatostatin (SOM), vasoactive intestinal polypeptide (VIP), enkephalin (ENK), Substance-P and corticotropin releasing factor (CRF), though their functions and physiology were less well studied.

So far, only one explicit two-population model had been published [33], in which two network configurations were investigated. In the E-surround configuration, inhibitory signals from interneurons were projected to adjacent excitatory grid cells, while grid cells signal were onto the surrounding ring of inhibitory neurons (equivalent to the Figure 1.5b). In this scenario, interneurons would have a anti-grid firing pattern (Figure 1.6 left). The second configuration, which termed I-surround configuration, was the opposite of the first one, that is, each grid cells made an excitatory connection to one inhibitory neuron, while this interneuron inhibit the surrounding grid cells (Figure 1.5a). In this case, interneurons would inherit the hexagonal pattern from grid cells (Figure 1.6 right).

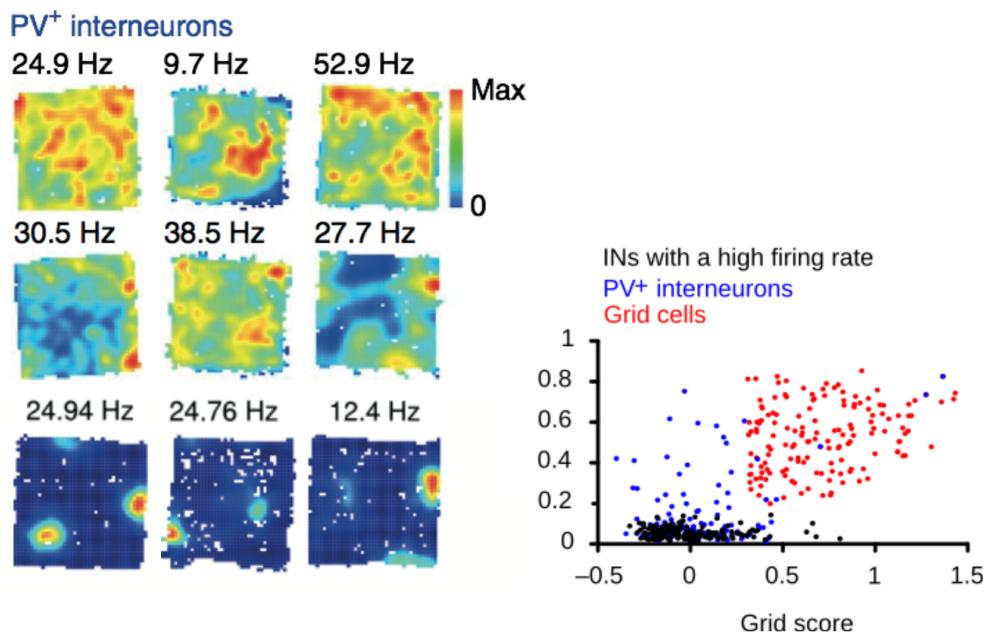


**Figure 1.5: Different connectivity between grid cells (blue) and interneurons (red).** Adapted from [34]



**Figure 1.6: Predicted interneuron firing fields for the E-surround and I-surround configuration.** Adapted from [33]

However, these predictions were challenged by a recent experimental observation. In their study, Buetfering et al. [35] explored the spatial firing properties of PV<sup>+</sup> cells in MEC, and found that their firing patterns were spatially stable and almost always, but not completely, aperiodic, instead of grid or anti-grid patterns as suggested (Figure 1.7). Using latency peak in cross-correlation analysis as an indicator of synaptic connection, Buetfering et al. [35] found that the rate maps of pairs of grid cells that made contacts to the same PV<sup>+</sup> interneurons were substantially different from each other in grid phases. They concluded that the aperiodic firing of interneurons resulted from integrating diverse inputs from grid cells with various phases. This observation was at odds with the assumption of the inhibitory attractor network, that is, grid cells shared similar grid phases coupled with each other via interneurons.



**Figure 1.7: Firing patterns of PV<sup>+</sup> interneurons in MEC.** Most of the PV<sup>+</sup> interneurons recorded were firing aperiodic, however, a small amount of them were showed confined grid patterns as grid cells. Left: Rate maps of PV<sub>+</sub> interneurons; Right: Spatial sparsity and grid score for grid cells, PV<sub>+</sub> interneurons and putative interneurons with a high firing rate Adapted from [35] and corresponding supplementary material.

As noted by Roudi and Moser [34] in response to this study, the presence of short latency peak in cross-correlation might not necessarily mean the existence of direct synaptic connections, i.e.

short-latency peaks might also show up in cross-correlation if pairs of cells shared a common excitatory input while one of them were activated at later time due to intrinsic properties, etc. [36].

Furthermore, the aperiodic firing interneurons were not necessarily contradicted with the continuous attractor network since synaptic configurations in the real biological system were rarely as ideal as the one wired in simulation of Pastoll et al. [33]. As shown in Figure 1.5c, it was possible that the effective inhibition was a happy coincidence of a randomly connected network where interneurons connected with multiple grid cells as well as each other, and multiple grid cells with different firing rates exerted their firing patterns onto multiple interneurons. And since interneurons inherited firing patterns from grid cells with various grid phases (and possibly with various peak firing rates in each grid field), their firing would become aperiodic.

So that, the main question of this thesis could be formulated like this: Can a two-population grid cell model, consistent with the effective inhibitory connectivity pattern, maintain grid cell firing even if interneurons lack spatial periodicity? In other words, what kind of synaptic configuration allowed grid cells performed path integration while interneurons fired in an aperiodic manner?

Starting from this question, we built a two-population continuous attractor model of grid cells using a technique called nonnegative matrix factorization (see next chapter). We found that, if grid cells and interneurons were connected in a random manner, including variation in the firing fields of the grid cells would result in aperiodic firing interneurons. By contrast, if the connectivity between grid cells and interneurons was well-structured, interneurons would fire periodically and the amount of interneurons could reduce to a surprisingly small number. Furthermore, we analyzed grid cell data from a single freely behaving animal and found that the degree to which the firing fields vary was consistent with the models in which gridness score decreased as the grid field variation increased. Take together, our result suggested that the firing patterns of interneurons might be not limited to the grid or anti-grid pattern but fell somewhere on the spectrum between highly periodic to aperiodic.

## 2 | Two population grid cell model

We derived the two population model from the effective model presented in [20]:

$$\tau \frac{ds_i}{dt} + s_i = g \left( \sum_j J_{ij} u_j + I_t \right)_+ \quad (2.1)$$

$$\tau \frac{du_i}{dt} + u_i = g \left( \sum_j K_{ij} s_j \right)_+ \quad (2.2)$$

where  $(.)_+$  is the threshold-linear function,  $g$  the gain,  $\tau$  the neuronal time constant, and  $J_{ij}$  and  $K_{ij}$  the strength of connection from inhibitory interneurons to grid cells and grid cells to interneurons, respectively. In other words,  $J$  is a pure negative matrix because interneurons here can only inhibit grid cells, while  $K$  is a pure positive matrix because grid cells can only excite interneurons.  $I_t$  represents external input, its content depends on the context. For the drift calculation in the result section,  $I_t$  is a constant external input, while for the path integration tasks we include a term that depends on the time-varying speed  $v_t$  and direction  $\theta_t$ ,

$$I_t = \text{constant} + \alpha v_t \cos(\theta_t - \theta_i) \quad (2.3)$$

where  $\alpha$  is the velocity modulation and  $\theta_i$  is the preferred direction of cell  $i$ . To determine how spatial periodicity of the interneurons changes with respect to the variance in spatial fields of the grid cells, we included a population of neurons with place cell-like coding and excitatory projections to the grid cells,

$$I_t = \text{constant} + \sum_j H_{ij} P_j(t) \quad (2.4)$$

The grid field peak rates are varied by drawing the strengths of the connections,  $H_{ij}$ , from a distribution with a variance related to the desired variability in peak firing rates.

To determine possible connectivity patterns corresponding to the inhibitory portion of an effective connectivity,  $W_{ij}$ , between principal cells, we note that, at equilibrium,  $du_i \rightarrow 0$  and we have

$$\sum_j J_{ij} u_j \approx g \sum_j J_{ij} \sum_k K_{jk} s_k \approx \sum_j W_{ij} s_j. \quad (2.5)$$

As in [18, 20, 33], we assumed a purely inhibitory connectivity with

$$W_{ij} = W_0 H(R_{max} - d_{ij}) H(d_{ij} - R_{min}), \quad (2.6)$$

where  $H$  is the Heaviside function,  $R_{max}$  the outer ring of the radial extent of the connectivity,  $R_{min}$  the inner ring,  $W_0$  the strength of the inhibitory interactions between connected neurons and  $d_{ij}$  the distance between cell  $i$  and  $j$ ,  $d_{ij}^2 = (x_i - x_j - l \cos \theta_i)^2 - (y_i - y_j - l \sin \theta_i)^2$  with  $x_i = 1 \dots N_x$  and  $y_i = 1 \dots N_y$  representing the position of neuron  $i$  in a two dimensional  $N_x \times N_y$  neural sheet with periodic boundary conditions and spatial offset  $l$ . That is, each grid cell in the model only connected with cells within the radius between  $R_{min}$  and  $R_{max}$ , whereas those that are too close or far apart are not coupled, forming a inhibitory ring connectivity.

The main challenge here is to find solutions to the problem  $W_{ij} = \sum_k J_{ik} K_{kj}$ . Since our model required  $K$  to be pure positive,  $J$  and  $W$  to be pure negative, using normal matrix factorization techniques would result in mixed values in one of the factorized matrix even when we forced the other one to be pure positive or negative. To solve this problem, we used tricks from non-negative matrix factorization [37, 38] (See next section), which first forced a nonnegative constraint on all matrices (assuming all  $J$ ,  $K$  and  $W$  are positive), then solve for the equations and

turn one of the factorized matrix into negative.

## 2.1 Nonnegative matrix factorization

Nonnegative matrix factorization (NMF) was first introduced by Finnish mathematicians Paatero and Tapper [39] as positive matrix factorization, and later developed and popularized by Lee and Seung in two papers published in 1999 and 2001 [37, 38].

An NMF problem can be defined as following: given a non-negative matrix  $W$ , find non-negative matrix factors  $J$  and  $K$  such that:

$$W \approx JK \quad (2.7)$$

where  $J$  and  $K$  can be attained by the multiplicative update rules

$$K_{a\mu} \leftarrow K_{a\mu} \frac{(J^T W)_{a\mu}}{(J^T J K)_{a\mu}} \quad \text{and} \quad J_{ia} \leftarrow J_{ia} \frac{(W K^T)_{ia}}{(J K K^T)_{ia}} \quad (2.8)$$

Iteration of these update rules converges to the local minima of the cost function

$$\|J - K\|^2 = \sum_{ij} (J_{ij} - K_{ij})^2 \quad (2.9)$$

The main difference between NMF and other matrix factorization methods, such as principal components analysis (PCA) and vector quantization (VQ) is that NMF uses nonnegativity constraints and this feature leads to a parts-based representation [37] due to the additive nature of the algorithm.

Lee and Seung [37] applied the NMF algorithm to an image dataset, which can be viewed as an  $n \times m$  matrix  $W$  and each column of matrix  $V$  is one of  $m$  images with  $n$  pixels. The factor matrix  $J$  is called basis matrix while  $K$  is called encoding matrix or coefficient matrix. The dot product

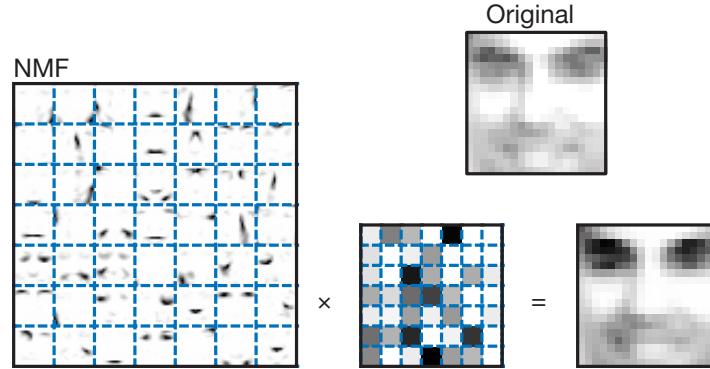


Figure 2.1: **NMF learns a parts-based representation of faces.** Adapted from [37].

of the basis matrix and a column of the coefficient matrix yields a approximate face image of the dataset (Figure 2.1). This parts-based representation is thought to be consistent with the psychological and physiological evidence in the brain [40].

NMF have been widely used in the field of machine learning, data mining, signal processing, to name a few [41]. Variations of NMF, such as Sparse NMF, Orthogonal NMF, Discriminant NMF [42], Semi-NMF [43] were also developed for diverse contexts.

## 2.2 Fully factorized model and randomly connected model

In this thesis, we constructed the two population model of grid cells based on two different uses of NMF.

First, both synaptic matrices  $J$  and  $K$  were initialized with random values, then learned the corresponding connectivity by applying the NMF algorithm to the effective inhibition matrix  $W$ . The algorithm was run till both  $J$  and  $K$  converged to the optimal condition, so that we called this model **fully factorized model**. This was a regular use of NMF as a unsupervised learning algorithm.

For the second case, synaptic matrices  $J$  and  $K$  were also initialized with random values but matrix  $K$ , which represents the projection from grid cells to interneurons, were constraint to

a given sparsity, such as 0.8 in our simulation, representing that only 20% of grid cells were connected to the interneurons in the network. Then the NMF algorithm were used to update the  $J$  given the effective inhibition matrix  $W$  without updating  $K$  at the same time, learning the backprojection from Interneurons to grid cells by the algorithm. This one-sided NMF can be regarded as a supervised learning algorithm and we called it **randomly connected model**.

The rank  $r$  of factorization was generally chosen to be smaller than  $n$  and  $m$  of  $W$  such that  $(n + m)r \leq nm$  due to dimension reduction. But in our case,  $r$  represented the number of interneurons in the network, it could be either smaller or bigger than the amount of grid cells.



# 3 | Methods

## 3.1 Model implementation

To investigate the firing properties of interneurons in the two population model, we constructed a network consisting of  $N_{\text{grid cells}} = 16 \times 16$  grid cells and  $N_{\text{interneurons}} = 26$  interneurons for both I-surround and E-surround configuration fully factorized model, whereas randomly connected model contained the same proportion of interneurons as grid cells ( $N_{\text{grid cells}} = N_{\text{interneurons}} = 16 \times 16$ ). All cells in these implementations were uniformly distributed on a twisted torus [17].

In the simulations where we considered network with grid cells having varying peak firing rates, we used  $\tau = 10\text{ms}$ ,  $I_{\text{constant}} = 1$ ,  $W_0 = -0.2$ ,  $R_{\min}$  and  $R_{\max}$  are normalized to be the  $0.6 \times \frac{16}{21}$  and  $0.6 \times \frac{36}{21}$  percent of minimal size of the neural sheets, respectively. The networks received no velocity input but were driven by place cells input. This place cells network with  $48 \times 48$  place cells were predefined, each place cell in the network encoded one unique position of the environment and projected to one grid cell in the grid cell network. The synaptic connection from place cells to grid cells were wired in a way that nine place cells arranged in hexagonal manner across the environment activated one grid cell, ensuring that the “bump” of activity on grid neural sheet was in the correct position. Then, the grid field peaks rates could be varied by drawing the strengths of the connections from place cells to grid cells,  $H_{ij}$  in equation 2.4, from a distribution with a variance related to the desired variability in peak firing rates.

To evaluate how the drift depends on the size of the inhibitory population, we used three different size fully factorized network ( $N_{\text{grid cells}} = 16 \times 13$ ,  $32 \times 27$ ,  $64 \times 55$ , respectively) and one

randomly connected network ( $N_{\text{grid cells}} = 32 \times 27$ ). Parameters were the same in all simulation:  $\tau = 10ms$ ,  $I_{\text{constant}} = 0.1$ ,  $W_0 = -0.1$ ,  $R_{\min}$  and  $R_{\max}$  are normalized to be the  $0.6 \times \frac{18}{21}$  and  $0.6 \times \frac{50}{21}$  percent of minimal size of the neural sheets, respectively.

For path integration, we switched to a multi-bumps model on a torus with  $N_{\text{grid cells}} = 64 \times 56$ ,  $N_{\text{interneurons}} = 200$ ,  $\tau = 10ms$ ,  $l = 4$ ,  $\alpha = 30$ ,  $I_{\text{constant}} = 0.1$ ,  $W_0 = -0.1$ ,  $R_{\min} = 15$  and  $R_{\max} = 20$  for fully factorized models and  $N_{\text{grid cells}} = 64 \times 56$ ,  $N_{\text{interneurons}} = 64 \times 56 \times 20$ ,  $\tau = 10ms$ ,  $l = 4$ ,  $\alpha = 20$ ,  $I_{\text{constant}} = 1$ ,  $W_0 = -0.3$ ,  $R_{\min} = 15$  and  $R_{\max} = 20$  for randomly connected model. No place cell input were used for error correction.

## 3.2 Data analysis and statistics

### 3.2.1 Simulated Rate maps

To construct rate maps, the position data from simulation were divided into 30 by 30 non-overlapping bins. Maps for number of spikes and time were smoothed with a Gaussian kernel function with the size of  $5 \times 5$  bins.

The firing rate map was then constructed by dividing the number of spikes in each bin by the time spent per bin.

### 3.2.2 Gridness Score

The degree of spatial periodicity (“gridness score”) was computed from the autocorrelogram of the rate maps [10]. A circular region of the autocorrelogram containing six nearest peaks from the center peak was defined. The Pearson correlation was calculated between the circular region with its rotation first for angles of  $60^\circ$  and  $120^\circ$ , then for angles of  $30^\circ$ ,  $90^\circ$ , and  $150^\circ$ . The gridness score then could be determined by highest minimum difference between two groups rotations.

### 3.2.3 Spatial Sparsity Score

Spatial sparsity score [35, 44] was calculated as following:

$$\text{sparsity} = 1 - \frac{(\sum_{i=1}^N p_i \lambda_i)^2}{\sum_{i=1}^N p_i \lambda_i^2}$$

where  $p_i$  is the occupancy probability of bin  $i$ , and  $\lambda_i$  is the mean firing rate for bin  $i$ .

### 3.2.4 Grid field variation

Grid fields variation of one grid cell was determined by the coefficient of variance (CV) between each grid field. Fields with their peaks on the border of the rate map were excluded. After identifying all the fields in the rate map, peaks of each field in the smoothed rate map were then used to compute the CV. See Figure 3.2 for the identified fields.

## 3.3 Experimental data

A dataset containing 176 putative grid cells identified as having spatially stable, periodic firing fields (from [10]) was analyzed for the grid field variation. The activity of the cells were recorded from 8 tetrodes.

The spikes were binned into 20 ms time bins, and all epochs with running speed less than 5 cm/s were excluded. The position data of the 150 cm  $\times$  150 cm environment was divided into non-overlapping 100 by 100 spatial bins. The path was smoothed with a Gaussian kernel function with the size of 5  $\times$  5 bins. The firing rate map was then constructed by dividing the number of spikes in each bin by the time spent per bin.

To ensure all grid cells in the analysis have clear grid fields, cells with gridness score less than 0.3 and spatial sparsity less than 0.4 were excluded from further analysis.

The remaining 75 grid cells could be assigned to four distinct modules using K-mean clustering [10]. A summary of the descriptive statistics of all data used in the analysis was presented in Table 3.1 and Figure 3.1, splitting into each module.

Table 3.1: Mean spacing and orientation for the 4 modules

<b>Modules</b>	<b>Mean Spacing ± std</b>	<b>Mean Orientation ± std</b>
1 (25 cells)	$34.0 \pm 2.1\text{cm}$	$-9.8 \pm 3.5^\circ$
2 (23 cells)	$40.9 \pm 2.5\text{cm}$	$5.3 \pm 2.5^\circ$
3 (20 cells)	$65.7 \pm 2.5\text{cm}$	$-7.1 \pm 3.0^\circ$
4 (7 cells)	$77.5 \pm 4.0\text{cm}$	$-11.5 \pm 2.7^\circ$

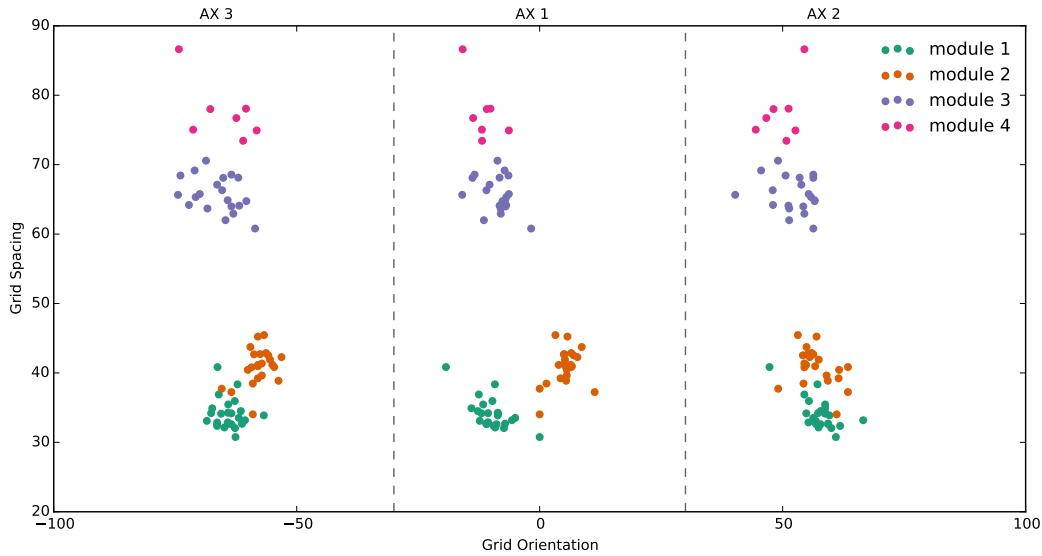


Figure 3.1: Grid spacing against grid orientation for all three grid axes.

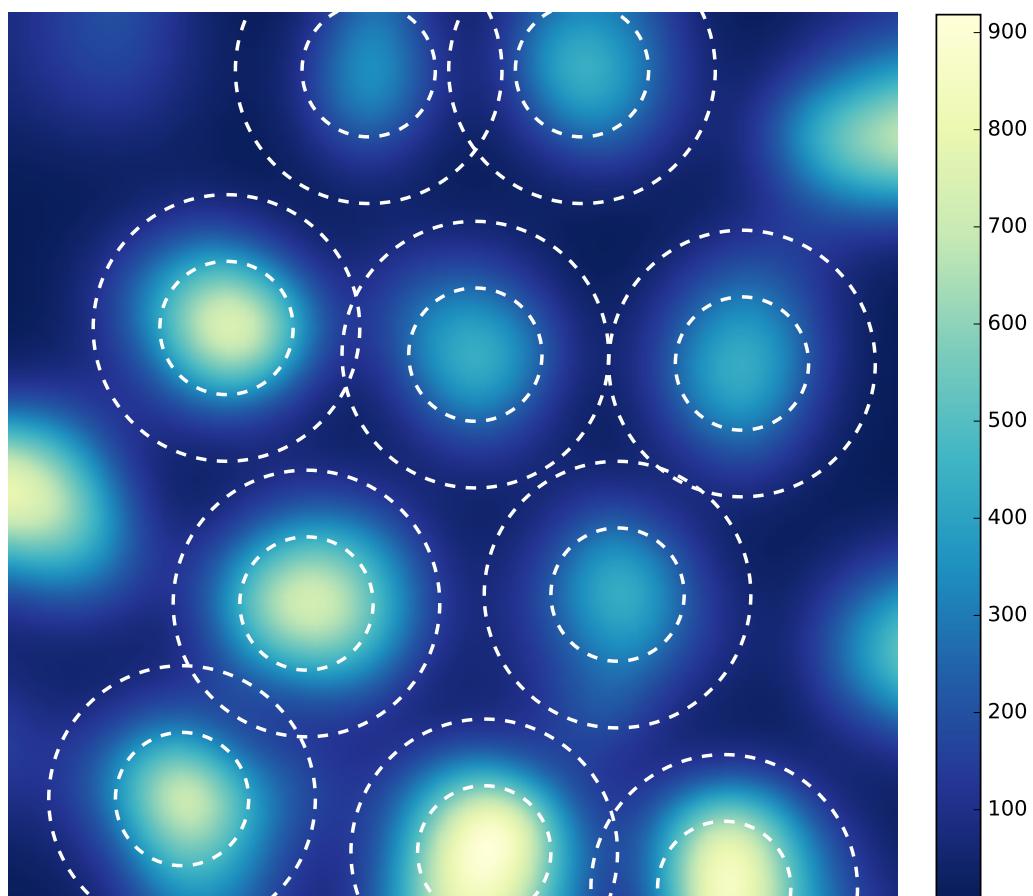


Figure 3.2: **Illustration of identified grid fields.**



# 4 | Results

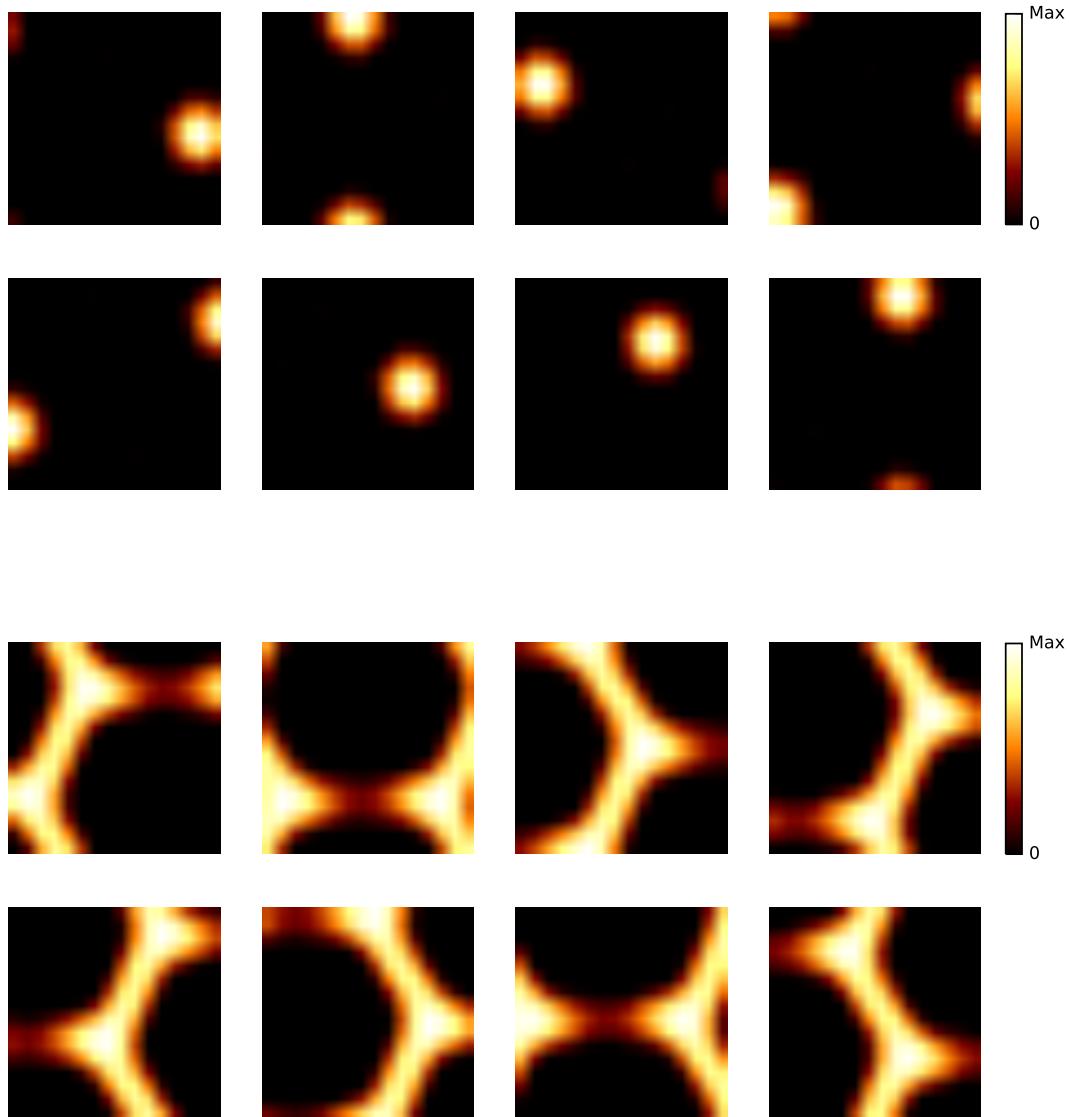
## 4.1 Synaptic Connectivity

The nonnegative matrix factorization of the effective inhibition matrix  $W$  resulted in three different cases of synaptic connectivity between grid cells and interneurons.

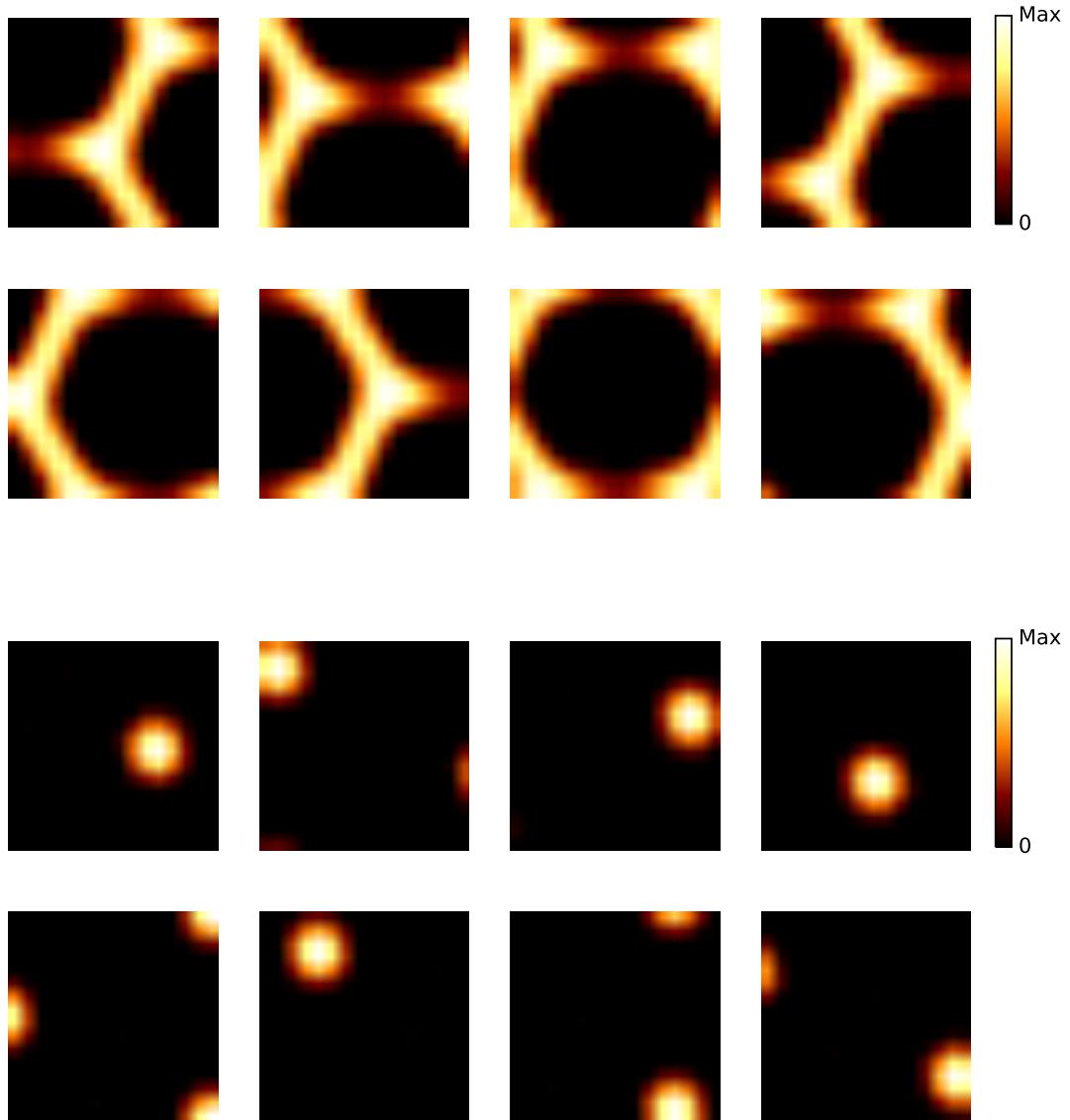
In the fully factorized model, each interneuron received input from a “bump” of grid cells and back-projected to a “ring” of grid cells (Figure 4.1), resembling the I-surround configuration in the study of Pastoll et al.[33]. Swapping and transposing the  $J$  and  $K$  in the equation 2.1 and 2.2, we got the second synaptic connectivity of the fully factorized model, which was the opposite of the first one: a “ring” of grid cells activated one interneuron and this interneuron projected inhibition to a “bump” of grid cells (Figure 4.2), resembling the I-surround configuration. [33].

For the randomly connected model, the projection from grid cells to interneurons was initialized randomly with a given sparsity and not updated (Figure 4.3 top), while the sparse back-projection from grid cells to interneurons was learned by the NMF algorithm (Figure 4.3 bottom).

These different synaptic configuration lead to different predictions of the firing patterns of interneurons. Similar to the study of Pastoll et al.[33], I-surround and E-surround configuration of fully factorized model caused interneurons to have grid and anti-grid pattern, respectively, while interneurons in randomly connected model would be fired with a less patterned activity because they received input from grid cells of randomly phases.



**Figure 4.1: I-surround synaptic connectivity of fully factorized model.** Visualization of the activated neurons on the grid neural sheet which projected to one interneuron (top) and the inhibited grid cells received input from one interneuron (bottom). Eight examples for each case were randomly selected from a network with 256 grid cells and 26 interneurons. Each interneurons received inputs from a “ring” of grid cells and back-projected to a “bump” of grid cells.



**Figure 4.2: E-surround synaptic connectivity of fully factorized model.** Visualization of the activated neurons on the grid neural sheet which projected to one interneuron (top) and the inhibited grid cells received input from one interneuron (bottom). Eight examples for each case were randomly selected from a network with 256 grid cells and 26 interneurons. Each interneurons received inputs from a “bump” of grid cells and back-projected to a “ring” of grid cells.

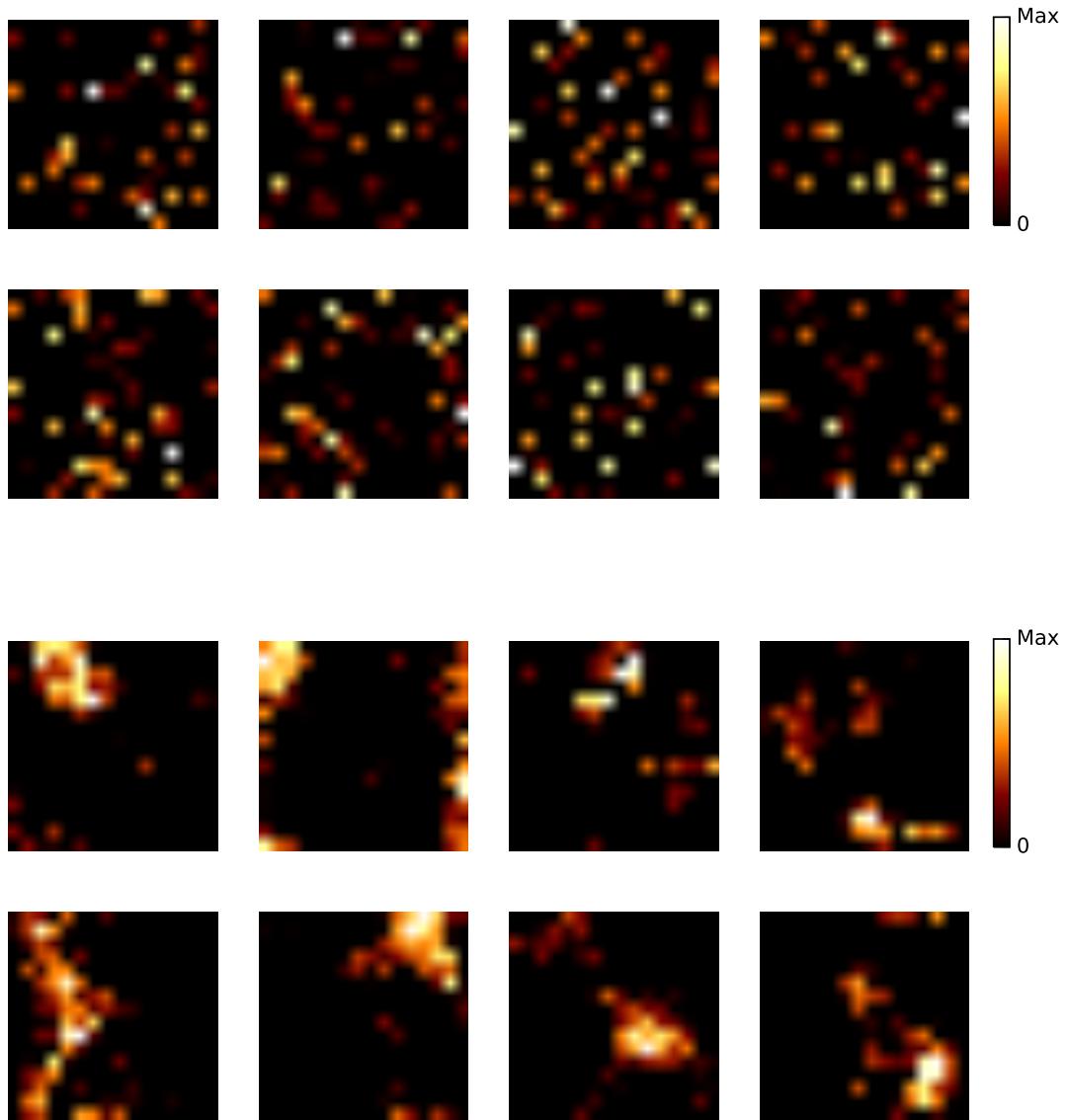


Figure 4.3: **Synaptic connectivity of randomly connected model.** Visualization of the activated neurons on the grid neural sheet which projected to one interneuron (top) and the inhibited grid cells received input from one interneuron (bottom). Eight examples for each case were randomly selected from a network with 256 grid cells and 256 interneurons. Grid cells and interneurons in the network connected with each other in a random manner.

## 4.2 Interneurons in I-surround configuration fully factorized model

As predicted, interneurons in I-surround configuration network had regular hexagonal firing pattern and also had a higher firing rate and a lower sparsity score than grid cells in the same network (Grid field variation, CV:  $0.04 \pm 0.02$ ; gridness score, interneurons:  $1.10 \pm 0.10$ , grid cells:  $1.09 \pm 0.12$ ,  $P = 0.57$ , *t*-test; firing rate, interneurons:  $2.74 \pm 0.56$  Hz, grid cells:  $0.26 \pm 0.04$  Hz,  $P < 0.001$ , *t*-test; spatial sparsity , interneurons:  $0.49 \pm 0.05$ , grid cells:  $0.58 \pm 0.05$ ,  $P < 0.001$ , *t*-test; Figure 4.5.)

The grid field variation increased as the variance of the place cells input increased, and gridness score of grid cells dropped accordingly while interneurons remained stable under the situation of high grid field variance (Grid field variation, CV:  $0.57 \pm 0.24$ ; gridness score, interneurons:  $1.09 \pm 0.15$ , grid cells:  $0.92 \pm 0.26$ ,  $P < 0.001$ , *t*-test; firing rate, interneurons:  $3.53 \pm 0.80$  Hz, grid cells:  $0.34 \pm 0.13$  Hz,  $P < 0.001$ , *t*-test; spatial sparsity , interneurons:  $0.55 \pm 0.05$ , grid cells:  $0.75 \pm 0.05$ ,  $P < 0.001$ , *t*-test;Figure 4.6.)

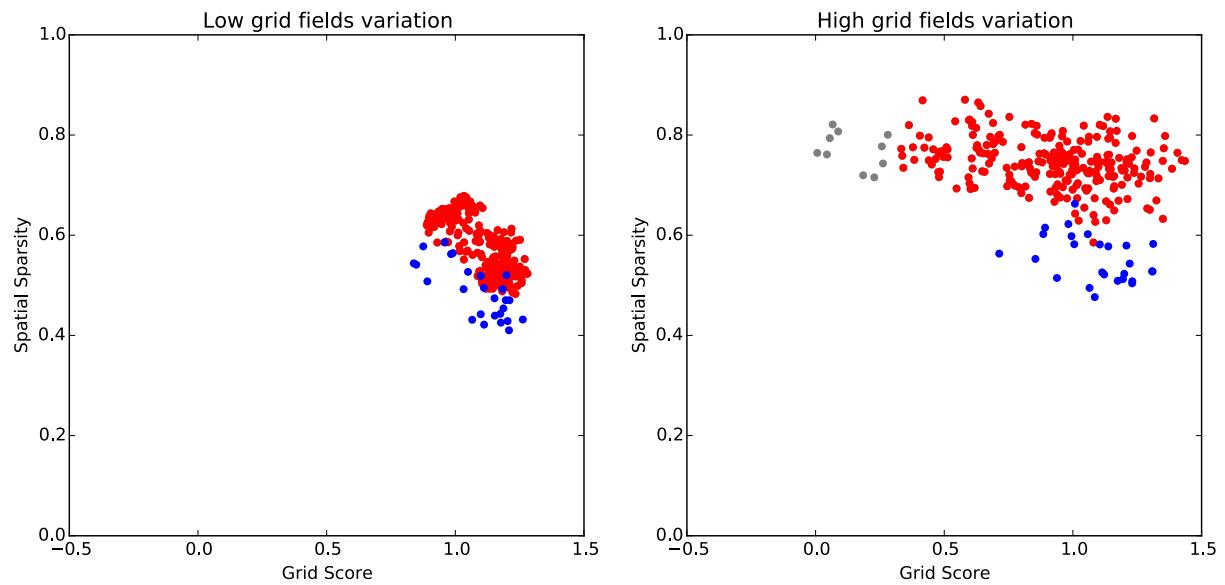


Figure 4.4: Spatial sparsity and grid score for simulated interneurons and grid cells in I-surround configuration fully factorized model. **Blue:** interneurons, **Red:** grid cells, **Gray:** grid cells with low gridness score ( $< 0.3$ )

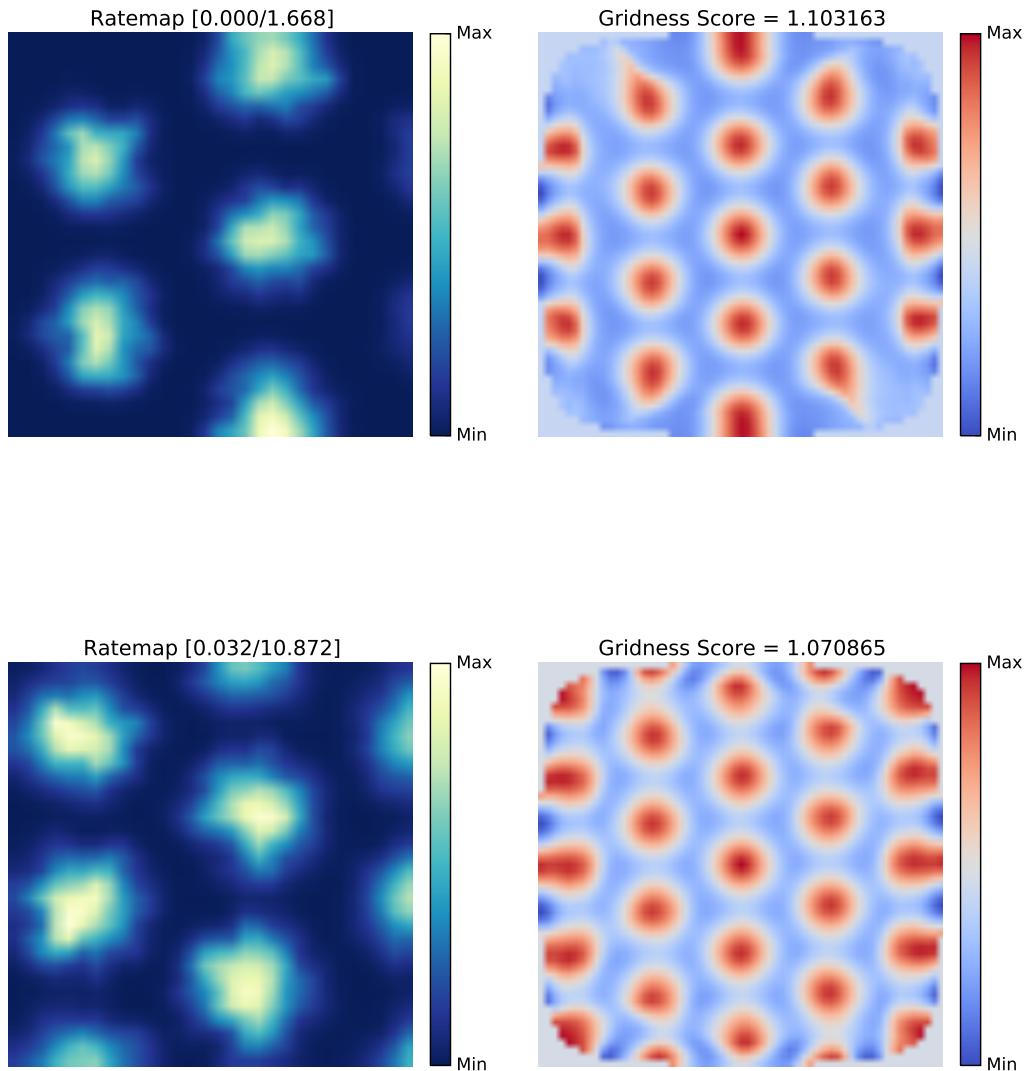


Figure 4.5: **Rate maps (left)** and **autocorrelograms (right)** of **grid cells (top)** and **interneurons (bottom)** in I-surround configuration fully factorized model with low grid field variation.

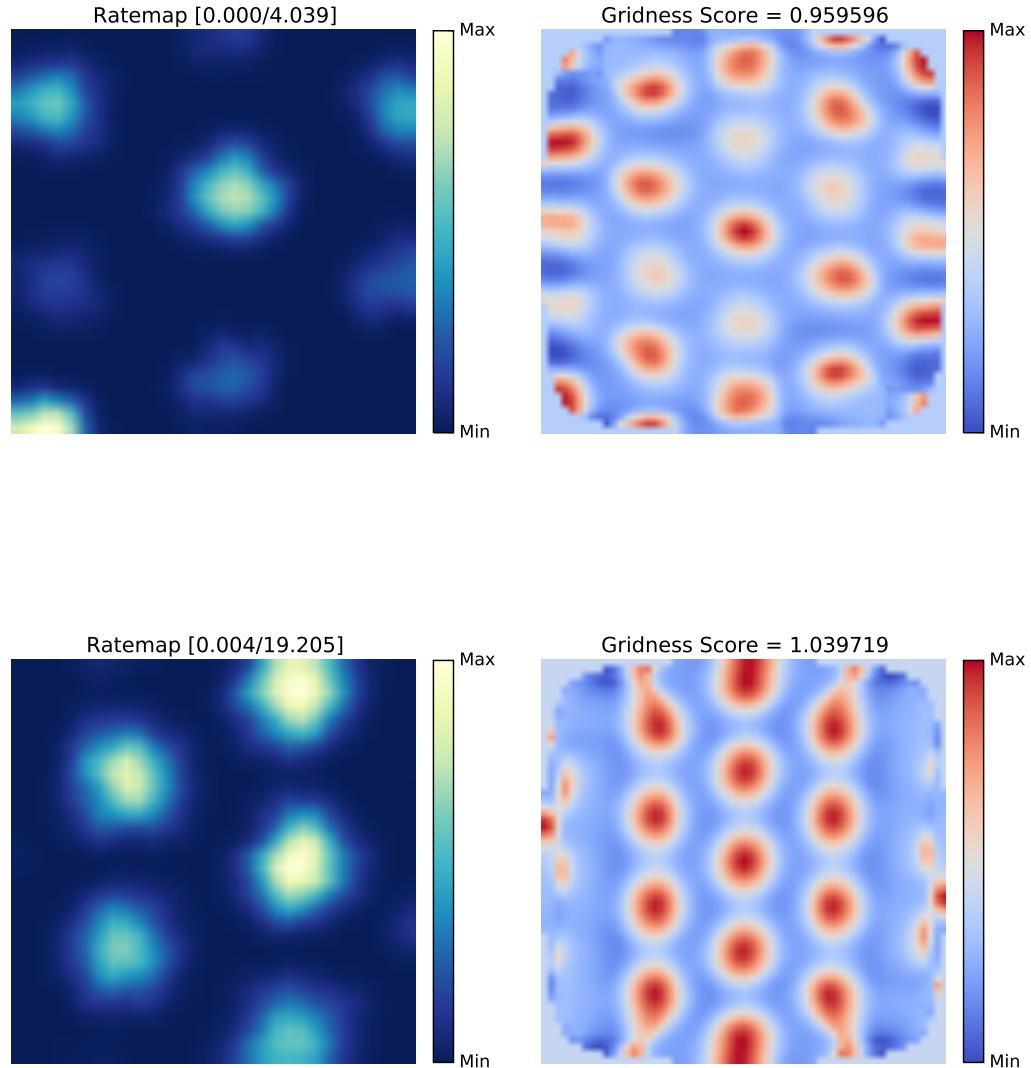


Figure 4.6: **Rate maps (left)** and **autocorrelograms (right)** of **grid cells (top)** and **interneurons (bottom)** in I-surround configuration fully factorized model with high grid field variation.

### 4.3 Interneurons in E-surround configuration fully factorized model

E-surround configuration network causes interneurons to have anti-grid pattern (Figure 4.8). Other properties of both grid cells and interneurons were similar to the I-surround configuration except the mean firing rate of interneurons was significantly greater than grid cells (Grid field variation, CV:  $0.05 \pm 0.02$ ; gridness score, interneurons:  $0.68 \pm 0.16$ , grid cells:  $1.14 \pm 0.11$ ,  $P < 0.001$ , *t*-test; firing rate, interneurons:  $21.75 \pm 3.47$  Hz, grid cells:  $0.26 \pm 0.05$  Hz,  $P < 0.001$ , *t*-test; spatial sparsity , interneurons:  $0.17 \pm 0.03$ , grid cells:  $0.57 \pm 0.05$ ,  $P < 0.001$ , *t*-test; Figure 4.8.)

In the scenario of high grid field variation, the anti-grid pattern of interneurons were still stable while grid cells lost their gridness gradually (Grid field variation, CV:  $0.55 \pm 0.24$ ; gridness score, interneurons:  $0.62 \pm 0.17$ , grid cells:  $0.88 \pm 0.22$ ,  $P < 0.001$ , *t*-test; firing rate, interneurons:  $27.06 \pm 5.13$  Hz, grid cells:  $0.34 \pm 0.14$  Hz,  $P < 0.001$ , *t*-test; spatial sparsity , interneurons:  $0.21 \pm 0.04$ , grid cells:  $0.75 \pm 0.06$ ,  $P < 0.001$ , *t*-test; Figure 4.9.)

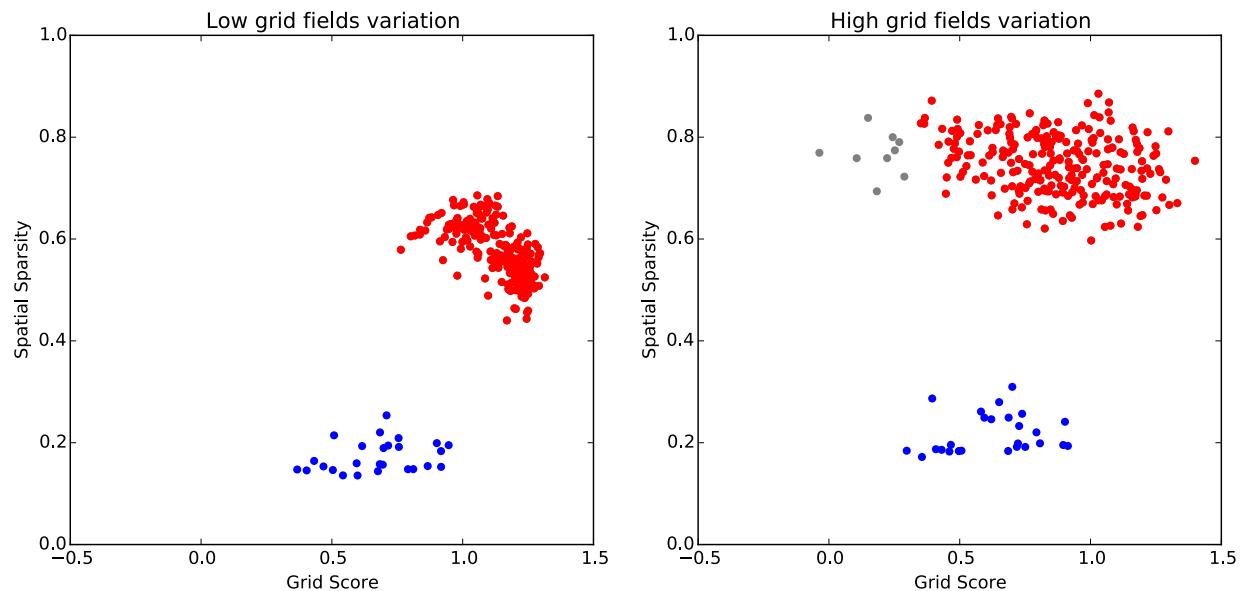


Figure 4.7: Spatial sparsity and grid score for simulated interneurons and grid cells in E-surround configuration fully factorized model. **Blue:** interneurons, **Red:** grid cells, **Gray:** grid cells with low gridness score ( $< 0.3$ )

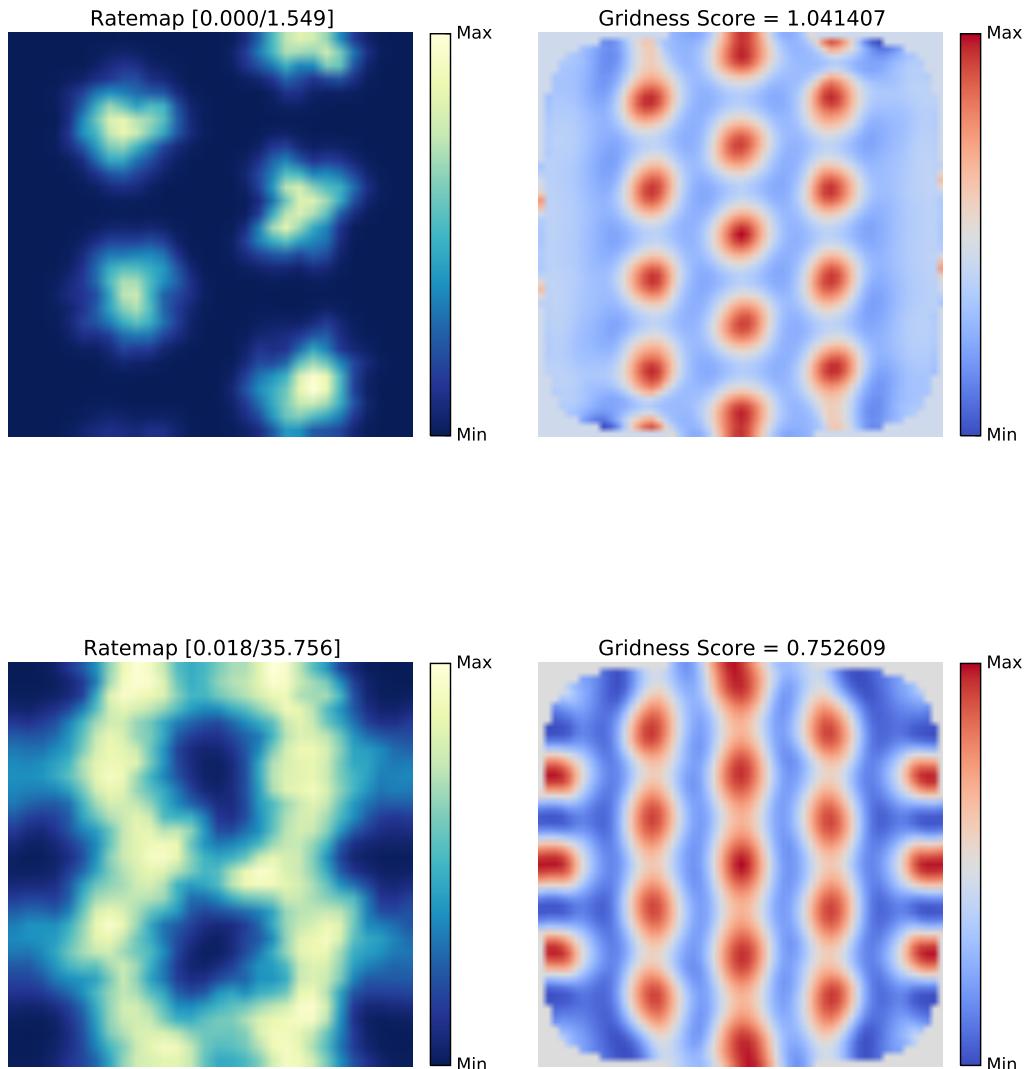


Figure 4.8: Rate maps (left) and autocorrelograms (right) of grid cells (top) and interneurons (bottom) in E-surround configuration fully factorized model with low grid field variation.

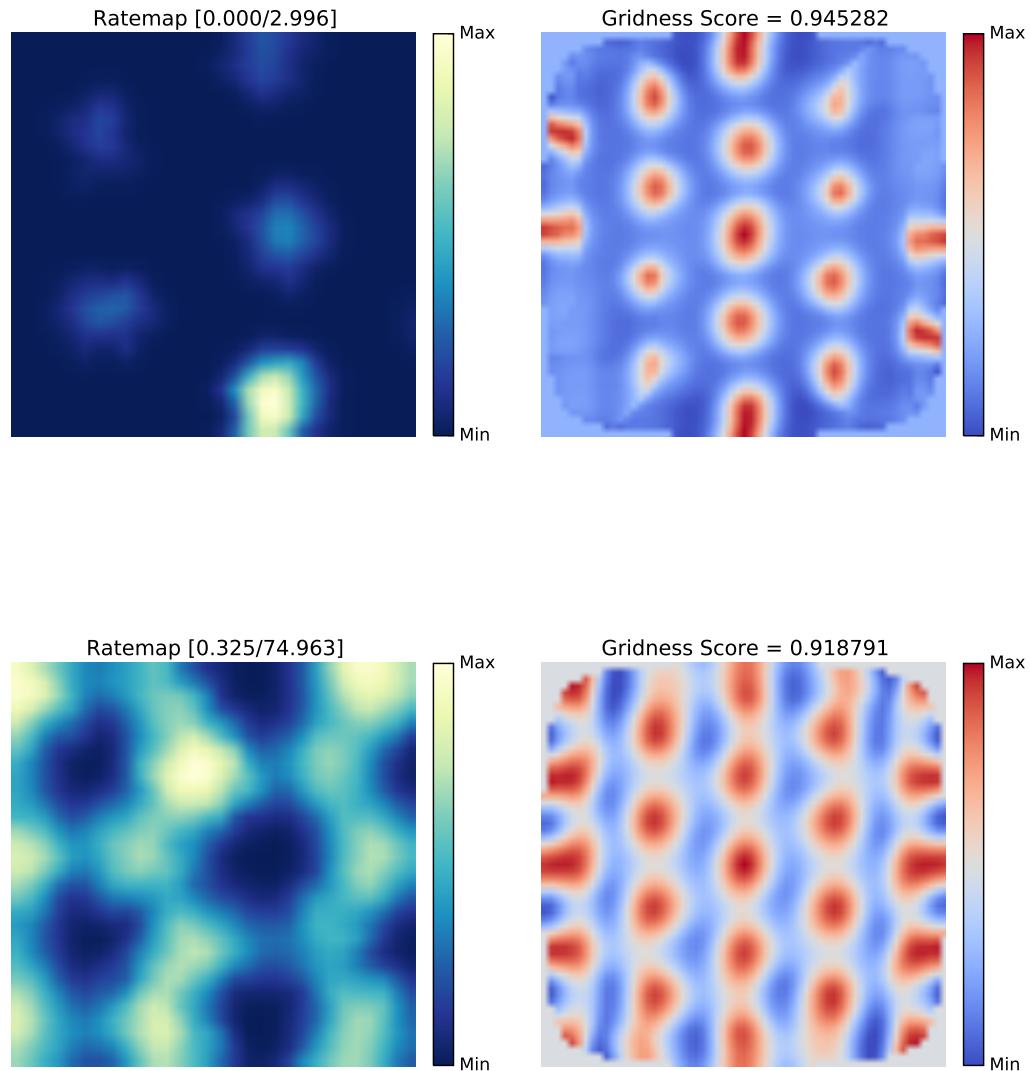


Figure 4.9: **Rate maps (left)** and **autocorrelograms (right)** of **grid cells (top)** and **interneurons (bottom)** in E-surround configuration fully factorized model with high grid field variation.

## 4.4 Interneurons in randomly connected model

Interneurons in randomly connected model fired in a more diverse manner than in fully factorized model since the patterns of interneurons were inherited from random grid cells with various phases (Grid field variation, CV:  $0.01 \pm 0.01$ ; gridness score, interneurons:  $0.69 \pm 0.26$ , grid cells:  $1.13 \pm 0.08$ ,  $P < 0.001$ , *t*-test; firing rate, interneurons:  $0.39 \pm 0.07$  Hz, grid cells:  $0.09 \pm 0.02$  Hz,  $P < 0.001$ , *t*-test; spatial sparsity, interneurons:  $0.05 \pm 0.03$ , grid cells:  $0.64 \pm 0.05$ ,  $P < 0.001$ , *t*-test). In a small possibility, if one interneuron integrate input from grid cells shared similar phases, then that interneuron could have a regular hexagonal (Figure 4.11) or reverse hexagonal pattern. But in most cases, grid cells with various phases projected to one interneurons, resulting in periodic but not hexagonal patterns (Figures not shown).

In the case of high grid field variation, the firing patterns of interneurons became aperiodic, in agreement with the experimental data reported by Buetfering et al. [35] (Grid field variation, CV:  $0.38 \pm 0.28$ ; gridness score, interneurons:  $0.18 \pm 0.30$ , grid cells:  $0.77 \pm 0.24$ ,  $P < 0.001$ , *t*-test; firing rate, interneurons:  $0.60 \pm 0.11$  Hz, grid cells:  $0.09 \pm 0.02$  Hz,  $P < 0.001$ , *t*-test; spatial sparsity, interneurons:  $0.17 \pm 0.04$ , grid cells:  $0.84 \pm 0.06$ ,  $P < 0.001$ , *t*-test; Figure 4.12).

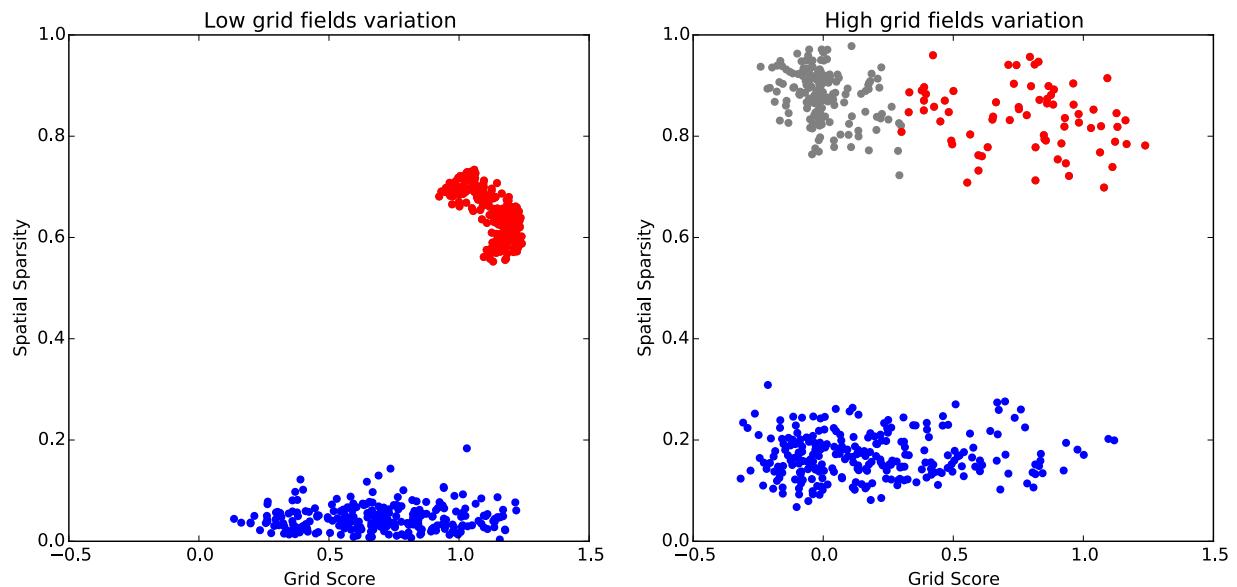


Figure 4.10: Spatial sparsity and grid score for simulated interneurons and grid cells in randomly connected model. **Blue:** interneurons, **Red:** grid cells, **Gray:** grid cells with low gridness score ( $< 0.3$ )

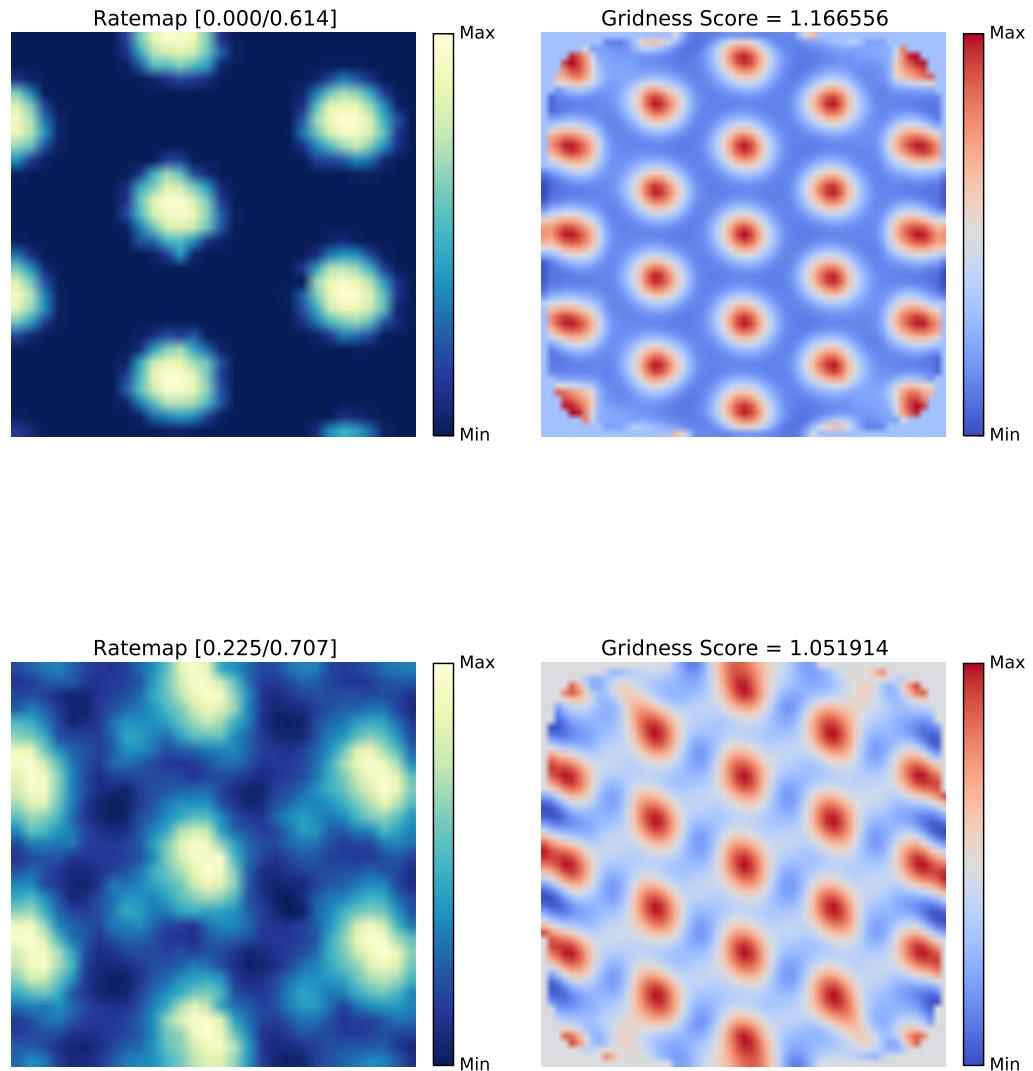


Figure 4.11: **Rate maps (left) and autocorrelograms (right) of grid cells (top) and interneurons (bottom) in randomly connected model with low grid field variation.**

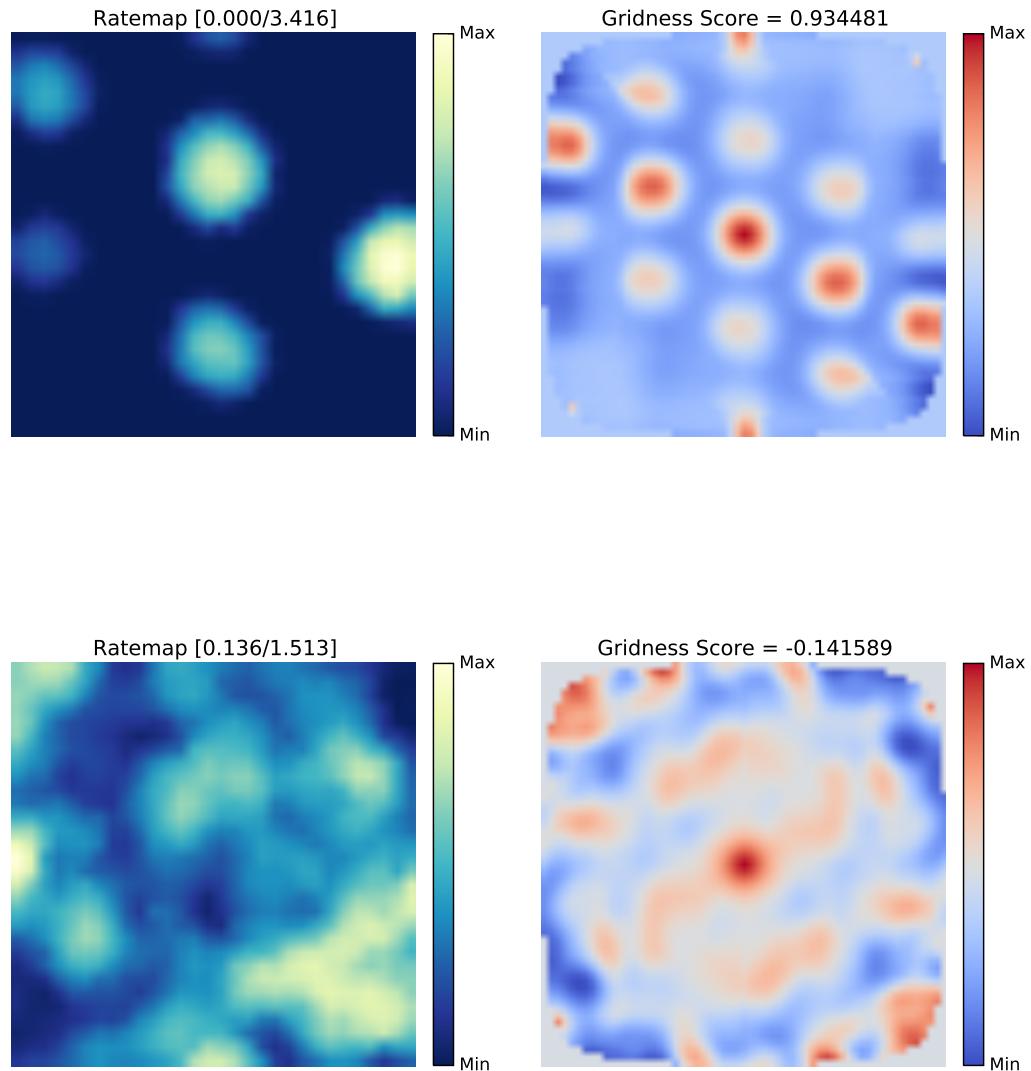


Figure 4.12: **Rate maps (left)** and **autocorrelograms (right)** of grid cells (top) and interneurons (bottom) in randomly connected model with High grid field variation.

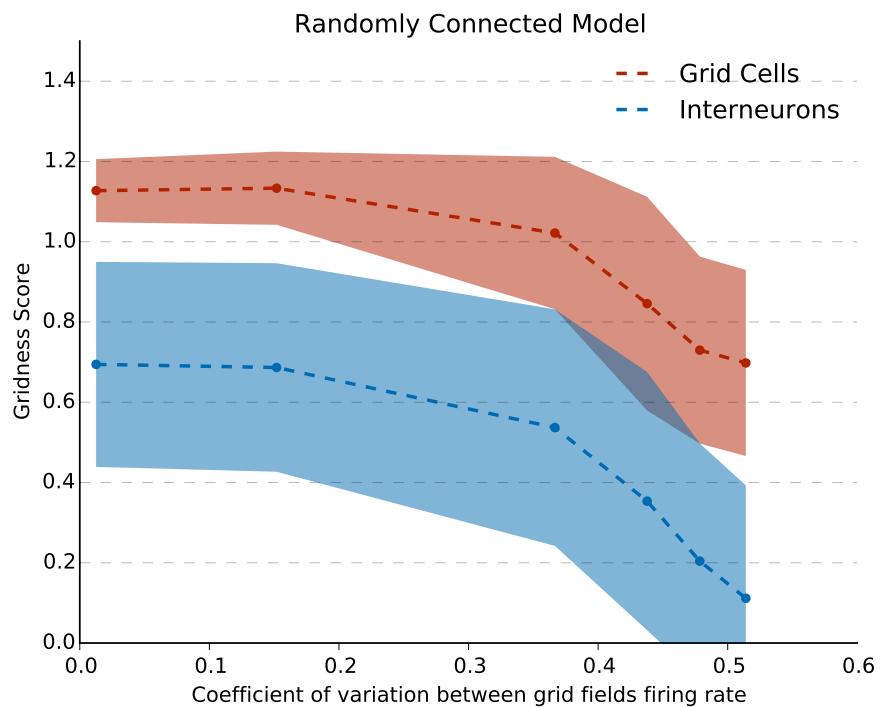
## 4.5 Gridness score and grid field variation

The gridness scores of both grid cells and interneurons were plotted as a function of grid field variation for all three synaptic connectivities. Grid field variation was manipulated by increasing the variance of the place cells input to the grid cells, as stated in Method section.

On the one hand, interneurons in randomly connected model had the same trend as grid cells and lost the grid pattern in high field variation situation (Figure 4.13).

One the other hand, for both I-surround and E-surround fully factorized model, gridness score of grid cells decreased as the variance increased while interneurons remained stable (Figure 4.14).

It should be noted that the average gridness scores in both E-surround fully factorized model and randomly connected model were lower than the ones in I-surround fully factorized model. This was probably due to the calculation of gridness score were only optimal for hexagonal pattern but not for anti-grid pattern, so that the gridness score cannot reflect the true degree of periodicity of cells for the anti-grid interneurons in E-surround fully factorized case.



**Figure 4.13: Randomly connected model.** Gridness score of both grid cells and interneurons decreased as grid field variation increased. In the situation of low grid field variation, interneurons in this configuration possessed mix patterns, such as grid, anti-grid and other periodic but not hexagonal firing pattern, resulting in the average gridness score around 0.7.

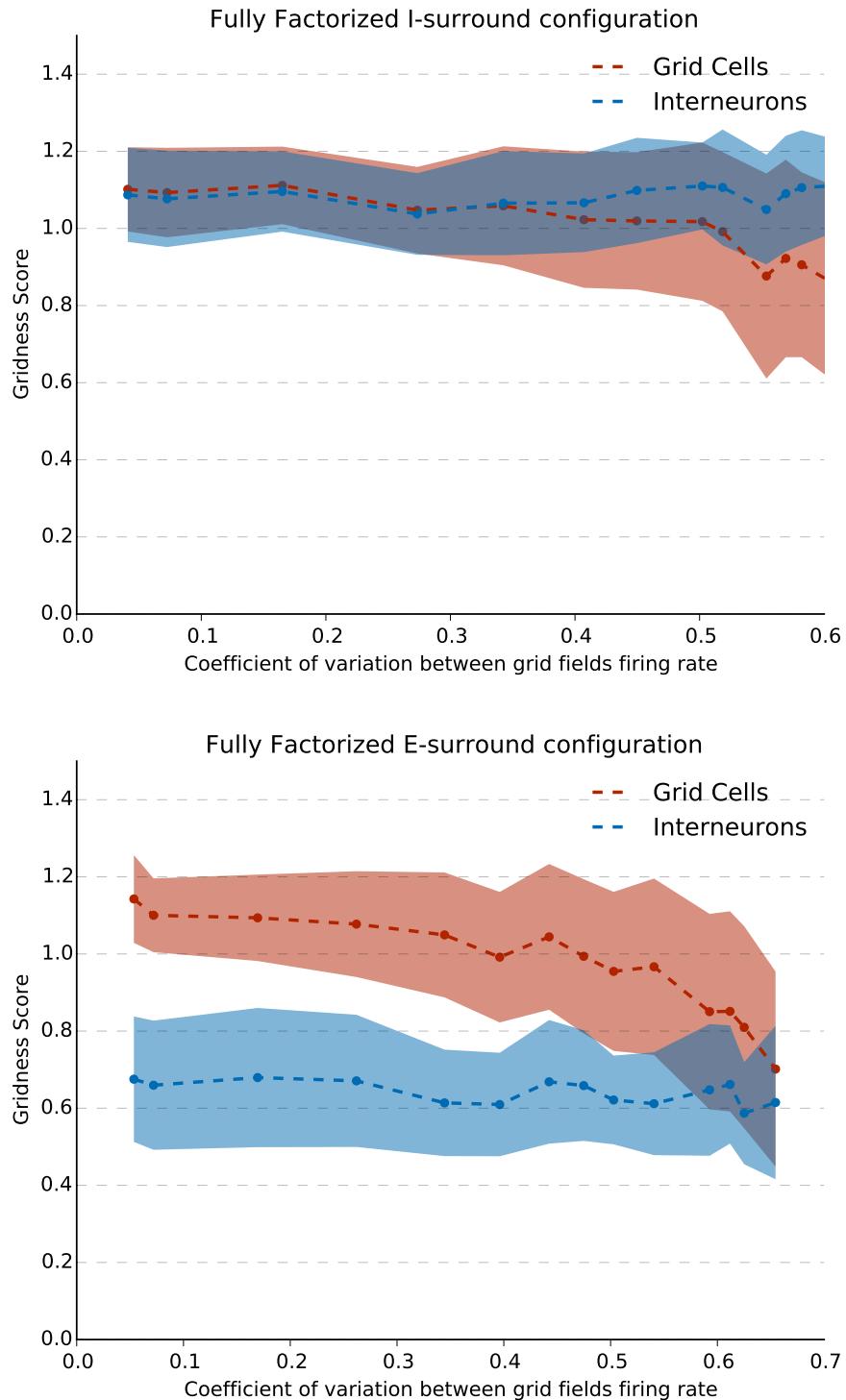
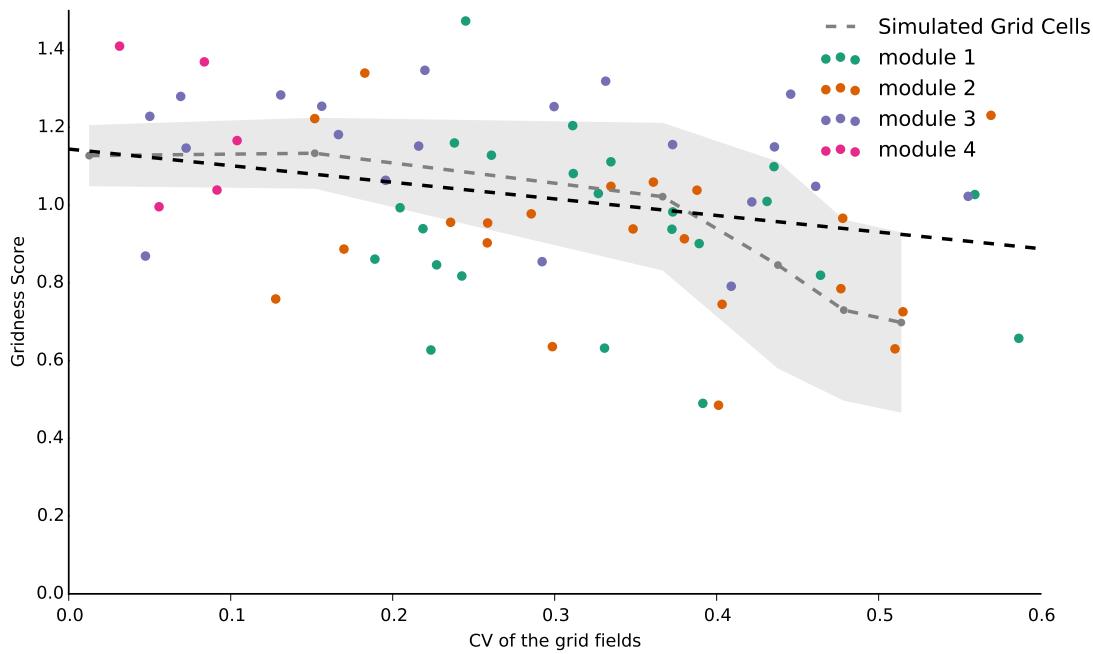


Figure 4.14: **I-surround (top)/ E-surround (bottom) configuration fully factorized model.** Interneurons stayed stable in high grid field variation situation.

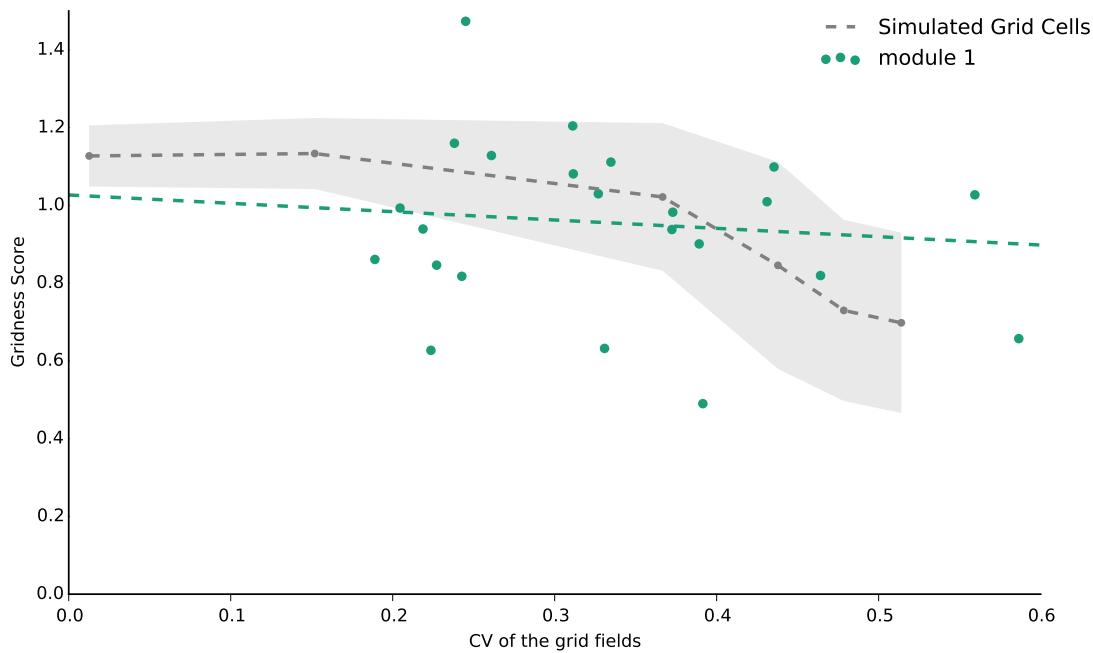
## 4.6 Grid field variation in experimental data

75 grid cells from Stensola et al [10] were analyzed to see if there is any connection between gridness score and grid field variations (Grid field variation, CV:  $0.33 \pm 0.17$ ; gridness score:  $1.00 \pm 0.22$ ). As shown in Figure 4.15, gridness score of grid cells were negative correlated with the coefficient of variance of grid field peak firing rate. The slope ( $\hat{\beta} = -0.43$ ) of the linear regression line are significantly different from 0 (t-test,  $P < 0.01$ ).

Since the data clustered into 4 different modules (module 4 is excluded due to small sample size), we also investigated if the same trend preserved in each module (module 4 was excluded due to small sample size). As indicated in Figure 4.16 and 4.17, all three modules showed a slightly negative correlation between gridness score and grid field variation, however, all slopes ( $\hat{\beta}_{\text{module 1}} = -0.21$ ,  $\hat{\beta}_{\text{module 2}} = -0.49$ ,  $\hat{\beta}_{\text{module 3}} = -0.24$ ) of the linear regression lines are not significantly different from 0 (t-test,  $P > 0.05$ ).



**Figure 4.15: Gridness score versus grid field variation of all cells.**



**Figure 4.16: Gridness score versus grid field variation of module 1.**

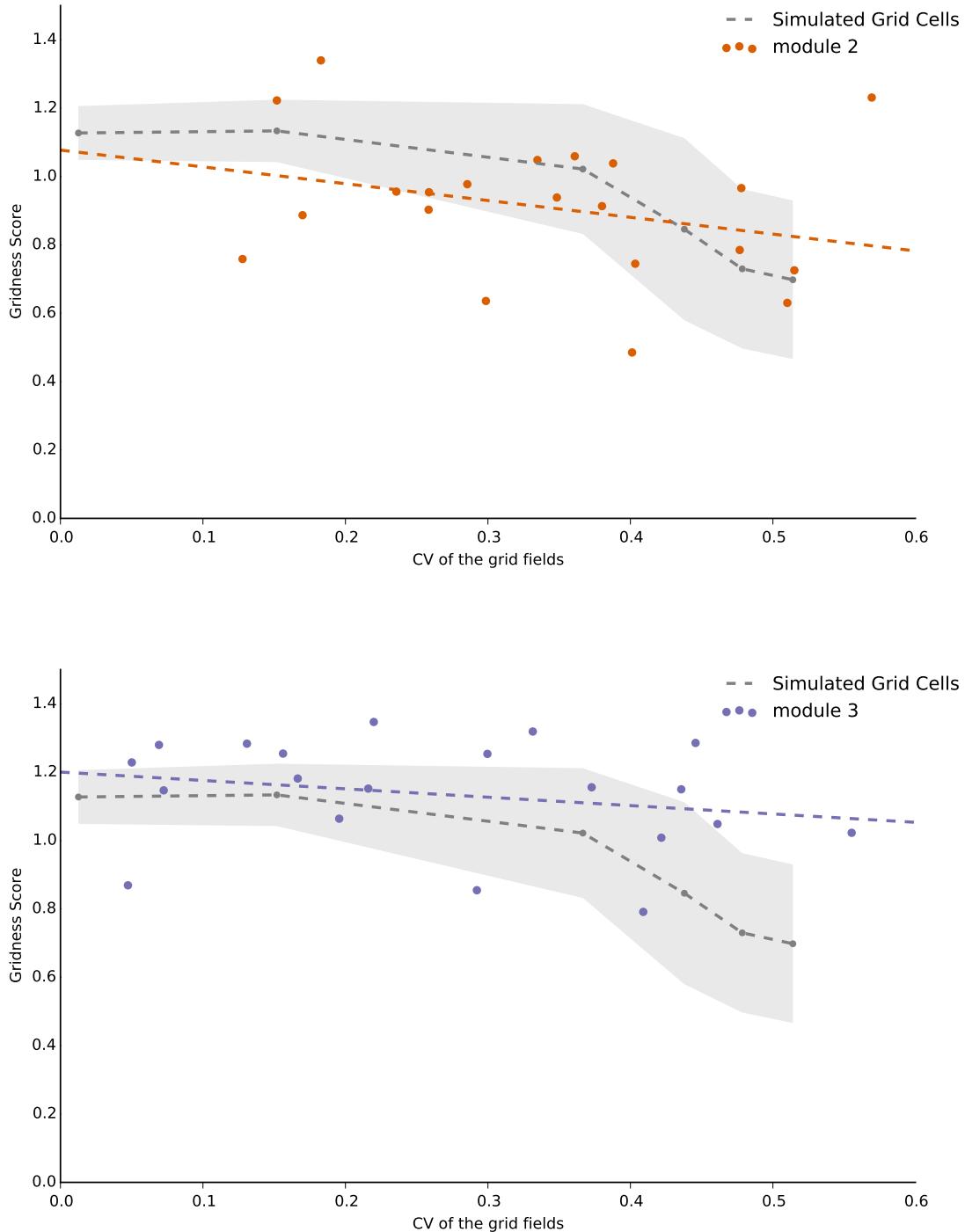


Figure 4.17: **Gridness score versus grid field variation of module 2 and 3.**

## 4.7 Drift

Here we used drift or drift velocity as a measure to evaluate the stability of different implementations of our two population model of grid cells. Drift in the network was measured by the average distance (over 200 repetition) that the activity on the neural sheet shifted during 100 time steps, starting from a random initial position.

As indicated in Figure 4.18, drift in the fully factorized model dropped exponentially as the ratio of the number of interneurons to grid cells and the size of the network was increased. For the case of randomly connected model, where interneurons received random input from grid cells, the network required a considerably higher proportion of interneurons to ensure a stable grid pattern.

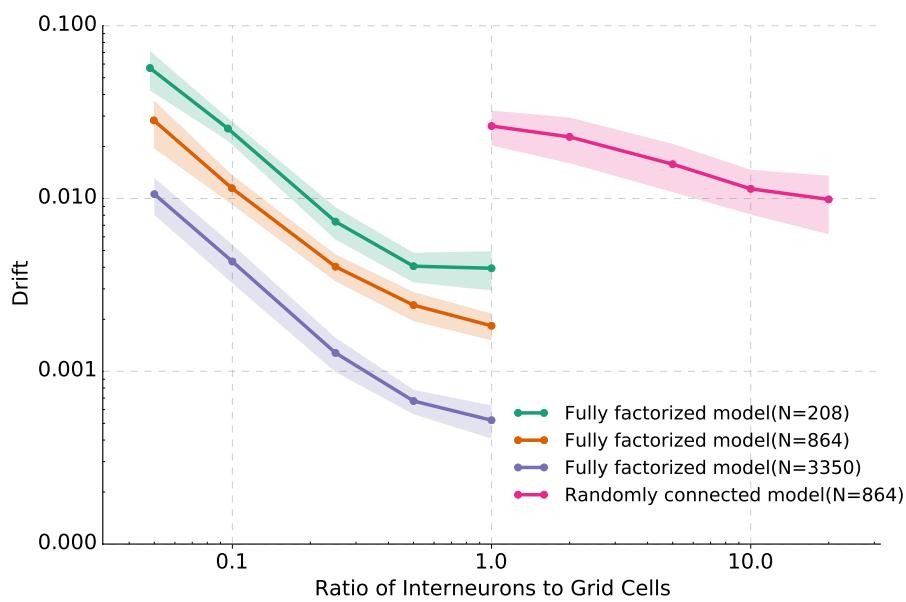


Figure 4.18: **Drift**. X-axis represented the gridness score, Y-axis the drift velocity (the unit of the drift velocity is neuron/time step, then normalized by the length of the neural sheet).

## 4.8 Path Integration

In this section, we explored the capacity of the two population model to perform path integration.

Both I-surround and E-surround fully factorized models with  $64 \times 56$  grid cells and 200 interneurons could path integrate accurately over 150000 time steps (Figure 4.22 and 4.23). However, the randomly connected model with  $64 \times 56$  grid cells and  $64 \times 56 \times 20$  interneurons was not able to form hexagonal patterns due to the accumulation of drift over time. Further investigation is needed to determine how many interneurons in the randomly connected model are required to lower the drift to the level so that path integration could be performed accurately.

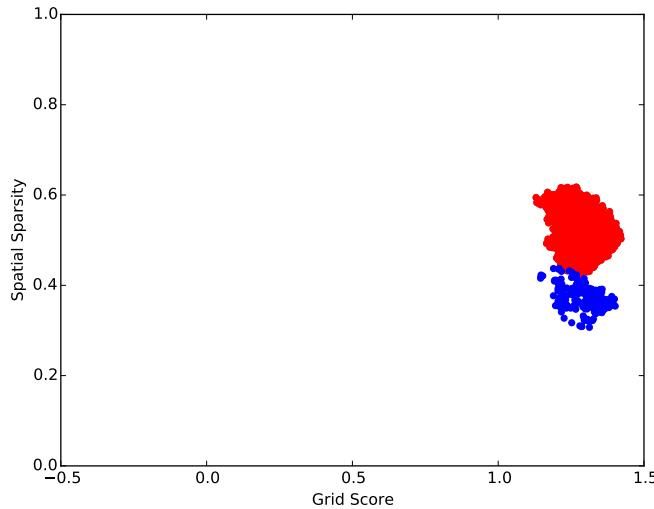


Figure 4.19: Spatial sparsity and grid score for path-integrated interneurons and grid cells in I-surround configuration fully factorized model. **Blue:** interneurons, **Red:** grid cells, **Gray:** grid cells with low gridness score ( $< 0.3$ )

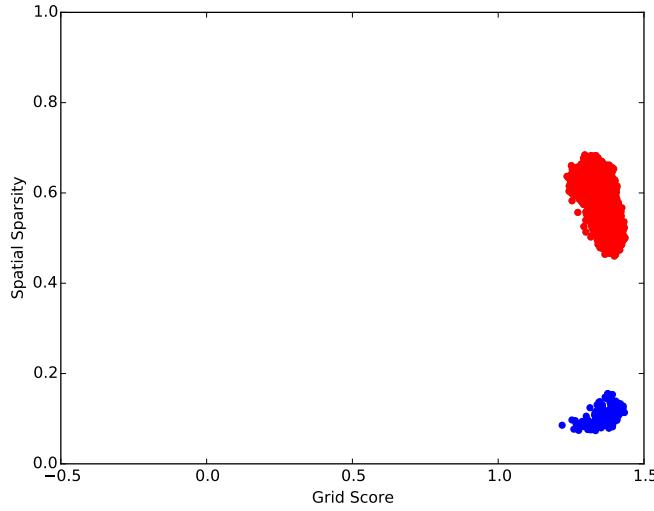


Figure 4.20: Spatial sparsity and grid score for path-integrated interneurons and grid cells in E-surround configuration fully factorized model. **Blue:** interneurons, **Red:** grid cells, **Gray:** grid cells with low gridness score ( $< 0.3$ )

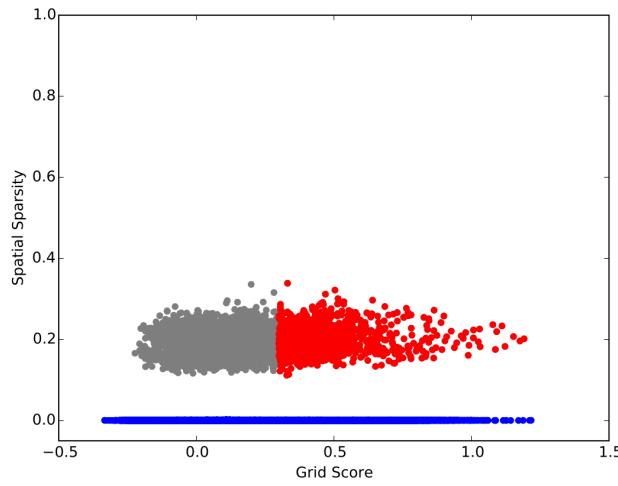


Figure 4.21: Spatial sparsity and grid score for path-integrated interneurons and grid cells in randomly connected model. The process of path integration was unsuccessful in this simulation. **Blue:** interneurons, **Red:** grid cells, **Gray:** grid cells with low gridness score ( $< 0.3$ )

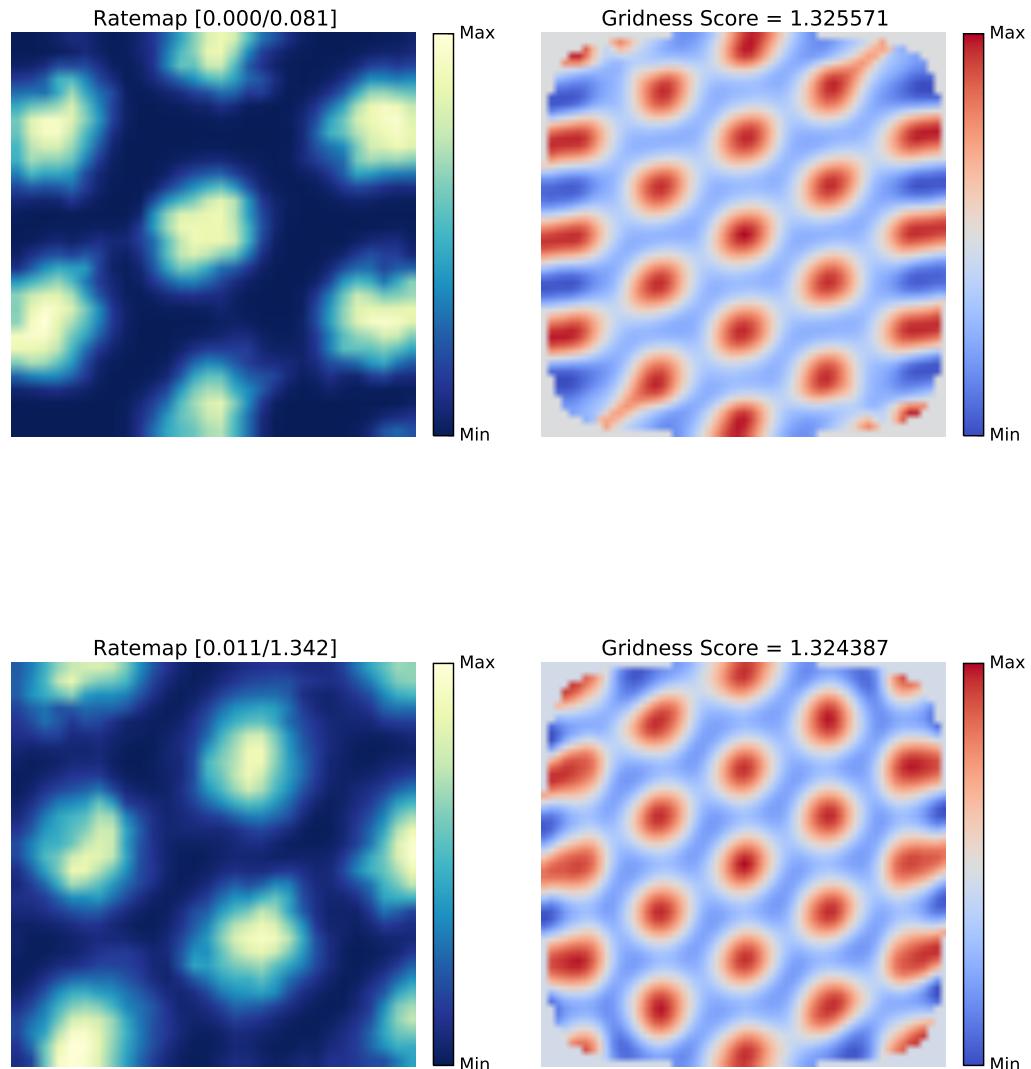


Figure 4.22: **Accurate path integration in I-surround configuration fully factorized model.**  
Top: grid cell; Bottom: interneuron.

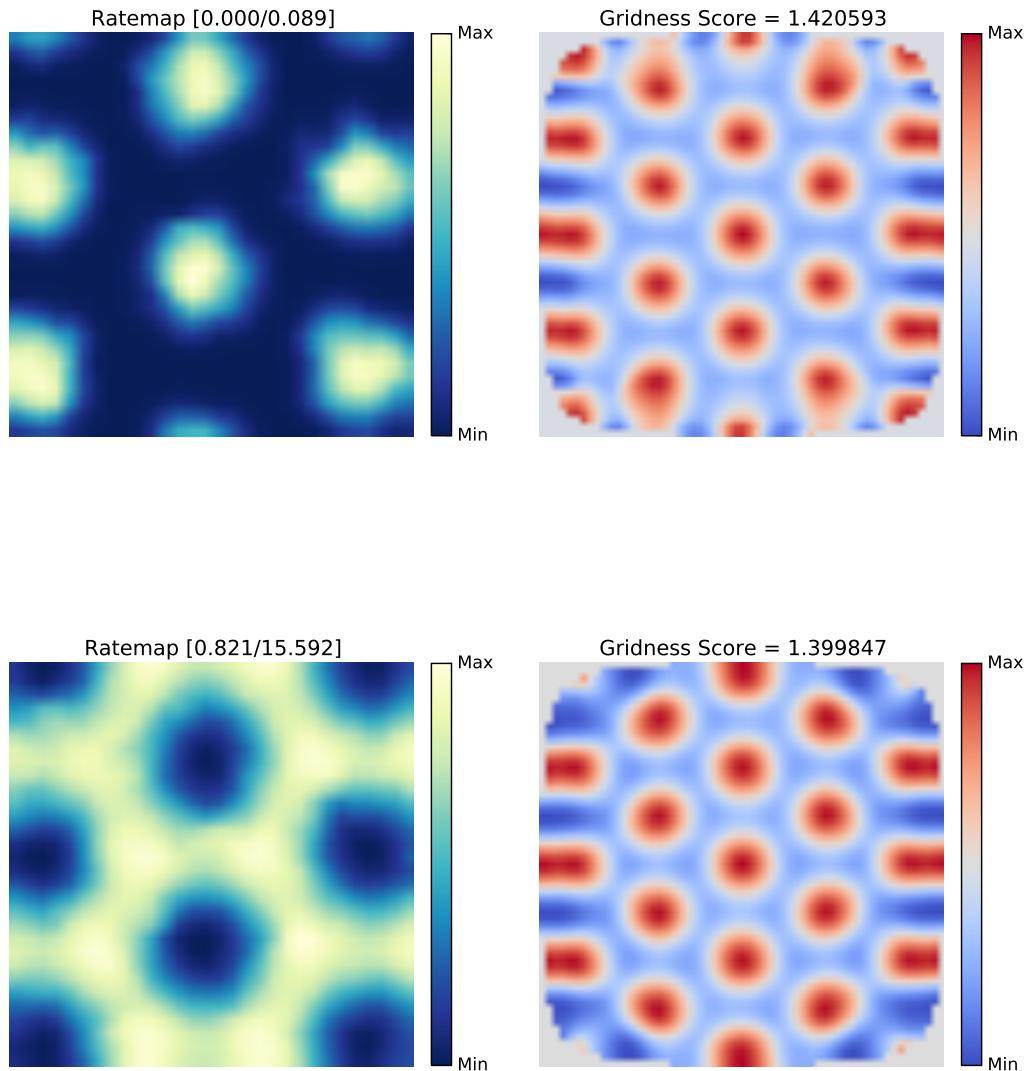


Figure 4.23: **Accurate path integration in E-surround configuration fully factorized model.**  
Top: grid cell; Bottom: interneuron.

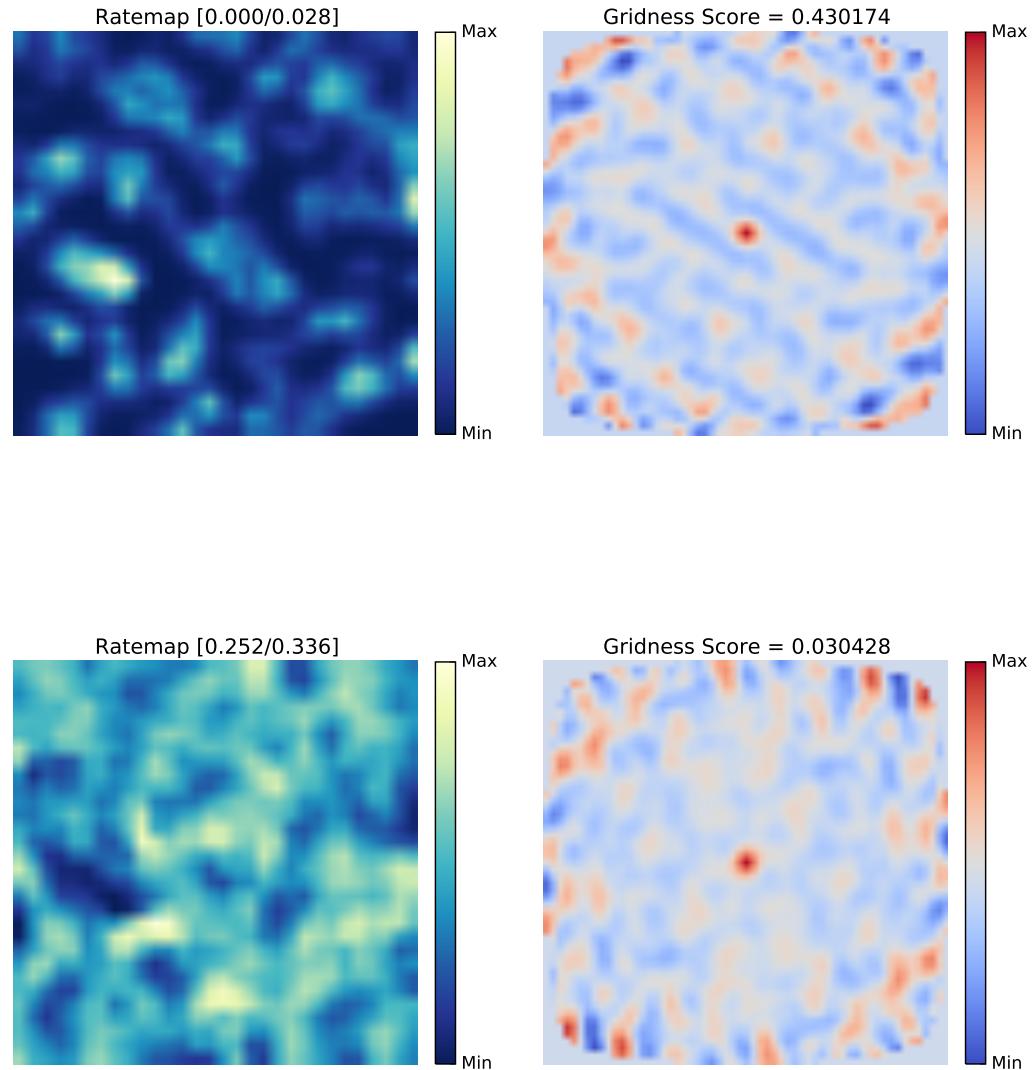


Figure 4.24: **Failed Path integration in randomly connected model.** Top: grid cell; Bottom: interneuron.

## 5 | Discussion and Conclusion

In this thesis, we explored the firing properties of interneurons in a two population grid cell model using nonnegative matrix factorization. We investigated different implementations of the two population model which lead to two opposite speculations of how grid cells and interneurons could be wired together in the recurrent inhibitory network:

- In the fully factorized model, the connectivity between grid cells and interneurons was well-structured: either I-surround configuration, that is, each interneuron received input from a “ring” of grid cells then backprojected to a “bump” of grid cells with similar grid phases, or E-surround configuration, i.e. each interneuron was recruited by many grid cells with similar spatial selectivity then inhibited other grid cells arranged in a ring on the neural sheet. In either case, the firing patterns of both grid cells and interneurons maintained a high degree of periodicity, and the proportion of interneurons could be significantly less than grid cells in the same network, for example, a network with interneurons account for less than 5% of total cells population could path integrate accurately;
- In the randomly connected model, the grid cells and interneurons were connected with each other in a random manner but resulted in the effective inhibitory connectivity. The two population network produced aperiodic firing interneurons similar to recordings when the peak firing rates of grid cells varied, however, a considerably larger proportion of interneurons was required to reach the level of network stability for path integration without spatial input as error correction.

Furthermore, we analyzed experiment data of 75 grid cells from a single freely behaving animal and found that the relationship between gridness score and grid field variation was consistent with the above models, that is, gridness score decreased as the grid field variation increased, when all data points were used. The same result did not hold true when we split the data into four modules, probably due to the small sample size for each module. Also, it should be noted that the true values of peak firing rate of grid fields were likely affected by the degree of smoothing when constructing the rate maps. So that, future experiments with better coverage should permit more accurate measurements. In our analysis, the expected variation in fields did vary considerably from cell to cell, so that the experimental data could be fit into both fully factorized model and randomly connected model. In both cases, the firing patterns of interneurons were inherited from the directly connected grid cells. Given that interneurons were recruited by grid cells with similar phases, even if the grid fields peak firing rates of those grid cells varied, the resulting patterns were still hexagonal. By contrast, majority of interneurons in randomly connected model integrated inputs from grid cells with various phases, the firing pattern of those interneurons would depend on the degree of both phase variability and the grid field peak firing rates variability, resulting in low spatial sparsity and aperiodic firing patterns similar to experimental observation.

## 5.1 How were grid cells and interneurons wired together?

The randomly connected model seem biologically impossible in the first glance. Even though interneurons in this scenario showed aperiodic firing patterns similar to experimental observation [35], the amount of interneurons required to drive the network was at odd with the fact that the estimated number of interneurons was roughly around 13% to 16% of total neural population in layer II of MEC [27, 45, 46], while, the amount of stellate cells was account for 55% to 67% [27, 45, 47]. However, there is one question needs to be answered, that is, how many of these stellate cells are grid cells? Currently, published data were contradicted with each other dramatically, ranging from just 3% [47] to more than 50% [6, 46]. In this sense, the randomly connected model was not completely hopeless if the exact amount of stellate cells as grid cells

were small.

Moreover, the potential of the randomly connected model were not exhausted in this thesis. For one thing, we still need to explore how many interneurons were enough to path integrate accurately in this scenario. As shown in section 4.4, some interneurons still possessed high spatial periodicity. How many interneurons had grid-like firing in this model? Were those interneurons playing a more important role than others, such as having more synaptic weight to grid cells? If we eliminate those high periodic interneurons, can the network still perform path integration, or otherwise, can this network path integrate only with those high periodic interneurons?

The last question raised a possibility that the randomly connected model was effectively the fully factorized model with most of the interneurons not providing any role. It was likely that the continuous attractor network in our randomly connected network was supported primarily by the few interneurons that, by chance, receive projections such that they do have periodic firing fields. This could be perceived as a weakness of this model, however, at the same time, it further demonstrated that the number of interneurons necessary to maintain this pattern was so few that even with just random connectivity the phenomenology can be maintained. Although the interneurons in the fully factorized model were not like the majority of those recorded in experiment [35], the simulated result suggested that a surprisingly small amount of interneurons would be sufficient to drive the network. Interestingly, more than 3 out of 140 interneurons sampled in the experiment showed grid-like firing and have low spatial sparsity similar to grid cells (see Supplementary Figure 3 in [35]). Further experimentation should help to determine if those periodic interneurons in [35] were just noise due to experimental error, or there are really small amount of periodic firing interneurons as predicted by the fully factorized model.

Taken the results of simulation and data analysis together, we proposed the following synaptic configuration of two population grid cell network: grid cells and interneurons connected with each other in a random manner and produced the effective inhibition required for continuous attractor network as a final result. A small proportion of interneurons were activated by grid cells with similar phases, resulting in the hexagonal or anti-hexagonal firing pattern. Those interneurons projected back to grid cells in the same network with a synaptic configura-

tion similar to the ones of fully factorized models, and they probably had a stronger connection strength than the rest of the population. Other interneurons, which were the majority of the population, integrated input from grid cells with various phases, and due to those grid cells tended to have different grid field variation, their firing patterns would become aperiodic and have low spatial sparsity. These interneurons could feedforward to other places, or project back to grid cells in the same network, however, since their connection strength were outweighed by the well-structured hexagonal interneurons, they would play a smaller role in the network mechanism.

For too long interneurons have been taken as simple mechanisms to keep the neural activity in balance [21]. Stronger and stronger evidence is emerging supporting interneurons as integral components of the neural code. However, in most grid cell models have considered only the “functional connectivity”, excitatory neurons have received all the attention and the role of interneurons was usually ignored. Here we demonstrate how this role could develop even if the projections to the interneuron population were drawn at random, resulting in what would appear as a sparse interneuronal code with just a few, key interneurons shaping the activity of the population of grid cells to emerge collectively into a neural code. With this in mind, the broader context of this work could be the thought that a potentially small population of key interneurons might be critical to this network, suggesting that we could disrupt the system disproportionately through manipulation of these key cells.

In order to crack the neural computation of navigation, one needs to understand how specific cell types contribute to this network mechanism. This work is our attempt to separate the roles of excitatory and inhibitory neurons in the MEC circuit.

## 5.2 Future Directions

The two population model in this thesis focused on the interactions between grid cells and interneurons. To test if this simplistic model would still hold true in a more biologically complicated setting, the lateral connection between interneurons should be taken into consideration,

so were other external inputs onto interneurons, such as velocity input or other neuromodulation. Furthermore, another important thing that was left out in this thesis was the interactions across layers within MEC and with other cells types such as border cells [7]. Also, it would be informative if the two population model were implemented as a integrate-and-fire spiking model, which could include a more detailed time-coordinated spike patterns and network oscillation.

Moreover, it would be interesting to develop a two population variation of the developmental model of grid cells [48]. Since the amount of interneurons required in the fully factorized model was so small, the synaptic connection needed to be learned between grid cells and interneurons would also be considerably smaller than the one in the effective model, e.g. in a network with  $100 \times 100$  grid cells, an effective model would have to learn  $100 \times 100 \times 9999$  synaptic connections while a two population with 500 interneurons only required  $100 \times 100 \times 500 \times 2$  synaptic connections. It would be interesting to see if the resulting synaptic connectivity between grid cells and interneurons were similar to the ones of nonnegative matrix factorization, and to see where would the limit of number of interneurons in the network be which could produce grid cells with confined grid pattern.



# References

- [1] Edward C. Tolman. Cognitive maps in rats and men. *Psychological review*, 55(4):189, 1948.
- [2] John O'Keefe and Jonathan Dostrovsky. The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(1):171–175, 1971.
- [3] John O'Keefe. Place units in the hippocampus of the freely moving rat. *Experimental Neurology*, 51(1):78–109, 1976.
- [4] M.-L. Mittelstaedt and H. Mittelstaedt. Homing by path integration in a mammal. *Naturwissenschaften*, 67(11):566–567, 1980.
- [5] 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, 2005.
- [6] Francesca Sargolini, Marianne Fyhn, Torkel Hafting, Bruce L. McNaughton, Menno P. Witter, May-Britt Moser, and Edvard I. Moser. Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science*, 312(5774):758–762, 2006.
- [7] Trygve Solstad, Charlotte N. Boccara, Emilio Kropff, May-Britt Moser, and Edvard I. Moser. Representation of geometric borders in the entorhinal cortex. *Science*, 322(5909):1865–1868, 2008.
- [8] Emilio Kropff, James E. Carmichael, May-Britt Moser, and Edvard I. Moser. Speed cells in the medial entorhinal cortex. *Nature*, advance online publication, 2015.

- [9] Menno P. Witter. Connectivity of the hippocampus. In Vassilis Cutsuridis, Bruce Graham, Stuart Cobb, and Imre Vida, editors, *Hippocampal Microcircuits*, number 5 in Springer Series in Computational Neuroscience, pages 5–26. Springer New York, 2010.
- [10] Hanne Stensola, Tor Stensola, Trygve Solstad, Kristian Frøland, May-Britt Moser, and Edvard I. Moser. The entorhinal grid map is discretized. *Nature*, 492(7427):72–78, 2012.
- [11] Edvard I. Moser, Yasser Roudi, Menno P. Witter, Clifford Kentros, Tobias Bonhoeffer, and May-Britt Moser. Grid cells and cortical representation. *Nature Reviews Neuroscience*, 15(7):466–481, 2014.
- [12] 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’. *Nature Reviews Neuroscience*, 7(8):663–678, 2006.
- [13] Edvard I. Moser, May-Britt Moser, and Yasser Roudi. Network mechanisms of grid cells. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1635):20120511, 2014.
- [14] Lisa M. Giocomo, May-Britt Moser, and Edvard I. Moser. Computational models of grid cells. *Neuron*, 71(4):589–603, 2011.
- [15] Eric A. Zilli. Models of grid cell spatial firing published 2005-2011. *Frontiers in Neural Circuits*, 6:16, 2012.
- [16] Mark C. Fuhs and David S. Touretzky. A spin glass model of path integration in rat medial entorhinal cortex. *The Journal of Neuroscience*, 26(16):4266–4276, 2006.
- [17] Alexis Guanella, Daniel Kiper, and Paul Verschure. A model of grid cells based on a twisted torus topology. *International Journal of Neural Systems*, 17(04):231–240, 2007.
- [18] Yoram Burak and Ila R. Fiete. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput Biol*, 5(2):e1000291, 2009.
- [19] A Dhillon and R. S. G Jones. Laminar differences in recurrent excitatory transmission in the rat entorhinal cortex in vitro. *Neuroscience*, 99(3):413–422, 2000.

- [20] Jonathan J. Couey, Aree Witoelar, Sheng-Jia Zhang, Kang Zheng, Jing Ye, Benjamin Dunn, Rafal Czajkowski, May-Britt Moser, Edvard I. Moser, Yasser Roudi, and Menno P. Witter. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nature Neuroscience*, 16(3):318–324, 2013.
- [21] Michael Okun and Ilan Lampl. Balance of excitation and inhibition. *Scholarpedia*, 4(8):7467, 2009.
- [22] Tora Bonnevie, Benjamin Dunn, Marianne Fyhn, Torkel Hafting, Dori Derdikman, John L. Kubie, Yasser Roudi, Edvard I. Moser, and May-Britt Moser. Grid cells require excitatory drive from the hippocampus. *Nature Neuroscience*, 16(3):309–317, 2013.
- [23] Edvard I. Moser, Emilio Kropff, and May-Britt Moser. Place cells, grid cells, and the brain’s spatial representation system. In *Annual Review of Neuroscience*, volume 31 of *Annual Review of Neuroscience*, pages 69–89, 2008.
- [24] Dávid Samu, Péter Erős, Balázs Ujfalussy, and Tamás Kiss. Robust path integration in the entorhinal grid cell system with hippocampal feed-back. *Biological Cybernetics*, 101(1):19–34, 2009.
- [25] Misha Tsodyks and Terrence Sejnowski. Associative memory and hippocampal place cells. *International journal of neural systems*, 6:81–86, 1995.
- [26] Floris G. Wouterlood. Spotlight on the neurones (i): cell types, local connectivity, microcircuits, and distribution of markers. *The Parahippocampal Region*, pages 61–88, 2002.
- [27] Csaba Varga, Soo Yeun Lee, and Ivan Soltesz. Target-selective gabaergic control of entorhinal cortex output. *Nature Neuroscience*, 13(7):822–824, 2010.
- [28] Floris G. Wouterlood, Wolfgang Härtig, Gert Brückner, and Menno P. Witter. Parvalbumin-immunoreactive neurons in the entorhinal cortex of the rat: localization, morphology, connectivity and ultrastructure. *Journal of Neurocytology*, 24(2):135–153, 1995.
- [29] Maria Miettinen, Esa Koivisto, Paavo Riekkinen Sr., and Riitta Miettinen. Coexistence of parvalbumin and gaba in nonpyramidal neurons of the rat entorhinal cortex. *Brain Research*, 706(1):113–122, 1996.

- [30] 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. *Journal of Neuroscience*, 32(49):17620–17631, 2012.
- [31] Floris G. Wouterlood, Jacqueline C.M. van Denderen, Theo van Haeften, and Menno P. Witter. Calretinin in the entorhinal cortex of the rat: Distribution, morphology, ultrastructure of neurons, and co-localization with -aminobutyric acid and parvalbumin. *The Journal of Comparative Neurology*, 425(2):177–192, 2000.
- [32] Attila I. Gulyás, Norbert Hájos, and Tamás F. Freund. Interneurons containing calretinin are specialized to control other interneurons in the rat hippocampus. *The Journal of Neuroscience*, 16(10):3397–3411, 1996.
- [33] Hugh Pastoll, Lukas Solanka, Mark C. W. van Rossum, and Matthew F. Nolan. Feedback inhibition enables theta-nested gamma oscillations and grid firing fields. *Neuron*, 77(1):141–154, 2013.
- [34] Yasser Roudi and Edvard I. Moser. Grid cells in an inhibitory network. *Nature Neuroscience*, 17(5):639–641, 2014.
- [35] Christina Buetefring, Kevin Allen, and Hannah Monyer. Parvalbumin interneurons provide grid cell-driven recurrent inhibition in the medial entorhinal cortex. *Nature Neuroscience*, 17(5):710–718, 2014.
- [36] Yasser Roudi, Benjamin Dunn, and John Hertz. Multi-neuronal activity and functional connectivity in cell assemblies. *Current Opinion in Neurobiology*, 32:38–44, 2015.
- [37] Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.
- [38] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In T. K. Leen, T. G. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems 13*, pages 556–562. MIT Press, 2001.

- [39] Pentti Paatero and Unto Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994.
- [40] Nikos K. Logothetis and David L. Sheinberg. Visual object recognition. *Annual review of neuroscience*, 19(1):577–621, 1996.
- [41] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-ichi Amari. *Nonnegative Matrix and Tensor Factorizations*. Wiley, 2009.
- [42] Yu-Xiong Wang and Yu-Jin Zhang. Nonnegative matrix factorization: A comprehensive review. *IEEE Transactions on Knowledge and Data Engineering*, 25(6):1336–1353, 2013.
- [43] Chris Ding, Tao Li, and Michael I. Jordan. Convex and semi-nonnegative matrix factorizations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(1):45–55, 2010.
- [44] William E. Skaggs and Bruce L. McNaughton. Theta phase precession in hippocampal. *Hippocampus*, 6:149–172, 1996.
- [45] C. W. Gatome, L. Slomianka, H. P. Lipp, and I. Amrein. Number estimates of neuronal phenotypes in layer ii of the medial entorhinal cortex of rat and mouse. *Neuroscience*, 170(1):156–165, 2010.
- [46] Charlotte N. Boccara, Francesca Sargolini, Veslemøy Hult Thoresen, Trygve Solstad, Menno P. Witter, Edvard I. Moser, and May-Britt Moser. Grid cells in pre- and parasubiculum. *Nature Neuroscience*, 13(8):987–994, 2010.
- [47] Qiusong Tang, Andrea Burgalossi, Christian Laut Ebbesen, Saikat Ray, Robert Naumann, Helene Schmidt, Dominik Spicher, and Michael Brecht. Pyramidal and stellate cell specificity of grid and border representations in layer 2 of medial entorhinal cortex. *Neuron*, 84(6):1191–1197, 2014.
- [48] Emilio Kropff and Alessandro Treves. The emergence of grid cells: Intelligent design or just adaptation? *Hippocampus*, 18(12):1256–1269, 2008.



# A | Supplementary Figures

## A.1 Grid cell data

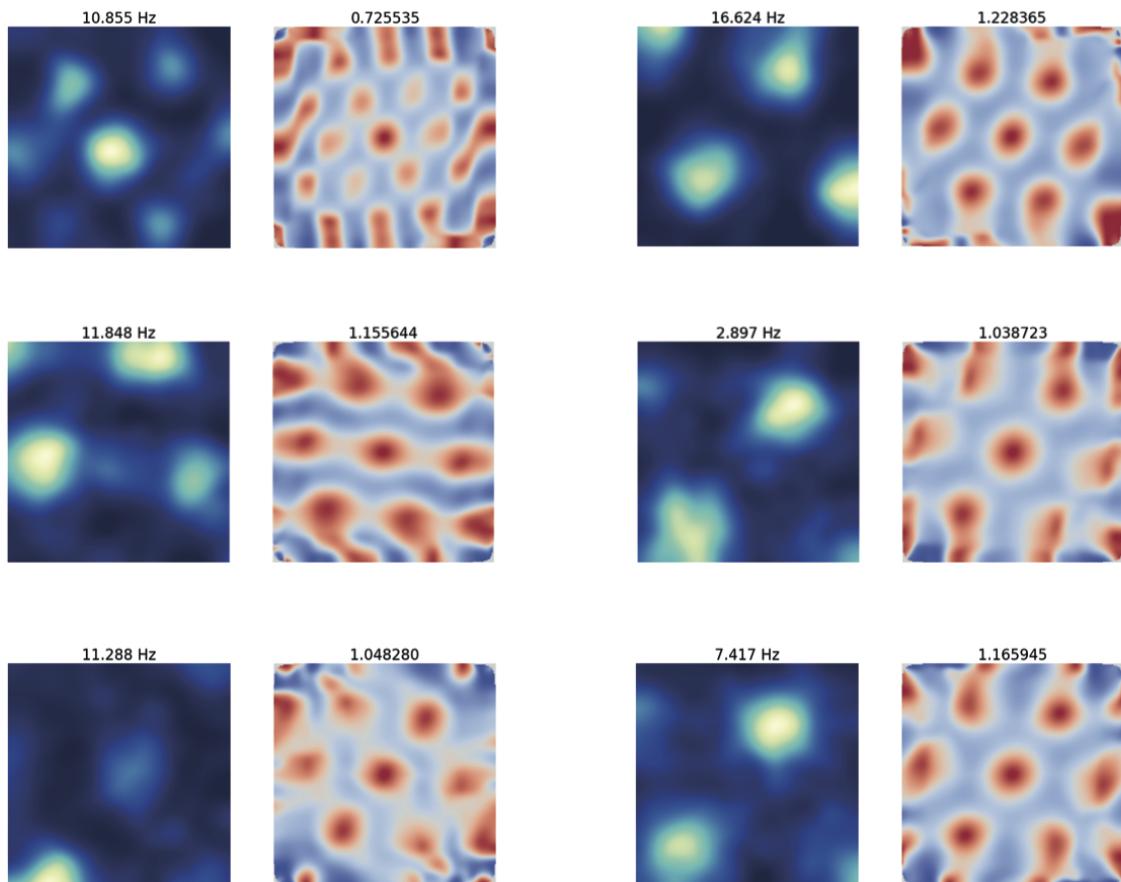


Figure A.1: Rate map and autocorrelogram of all grid cells included in data analysis. Part 1.

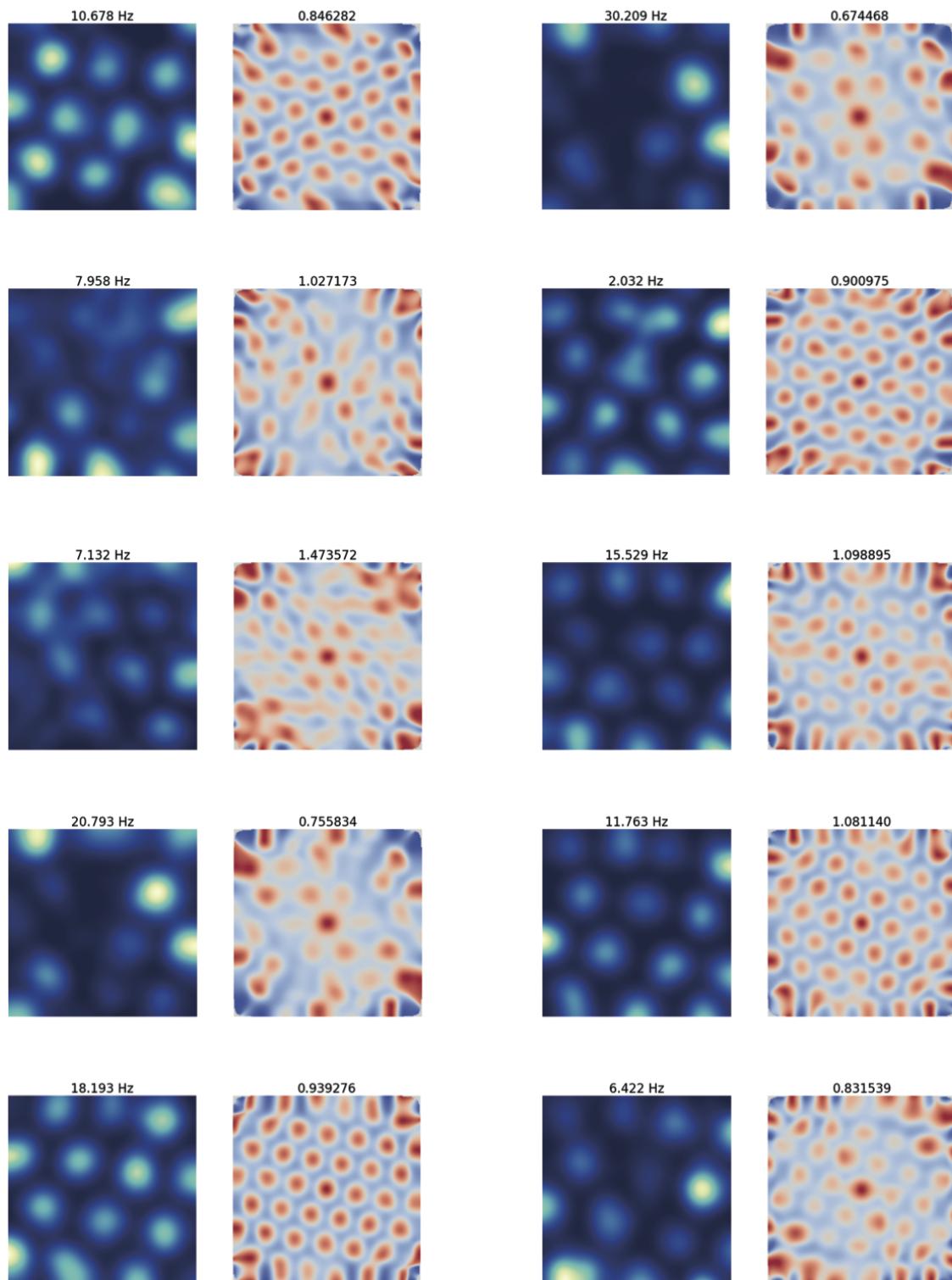


Figure A.2: Rate map and autocorrelogram of all grid cells included in data analysis. Part 2.

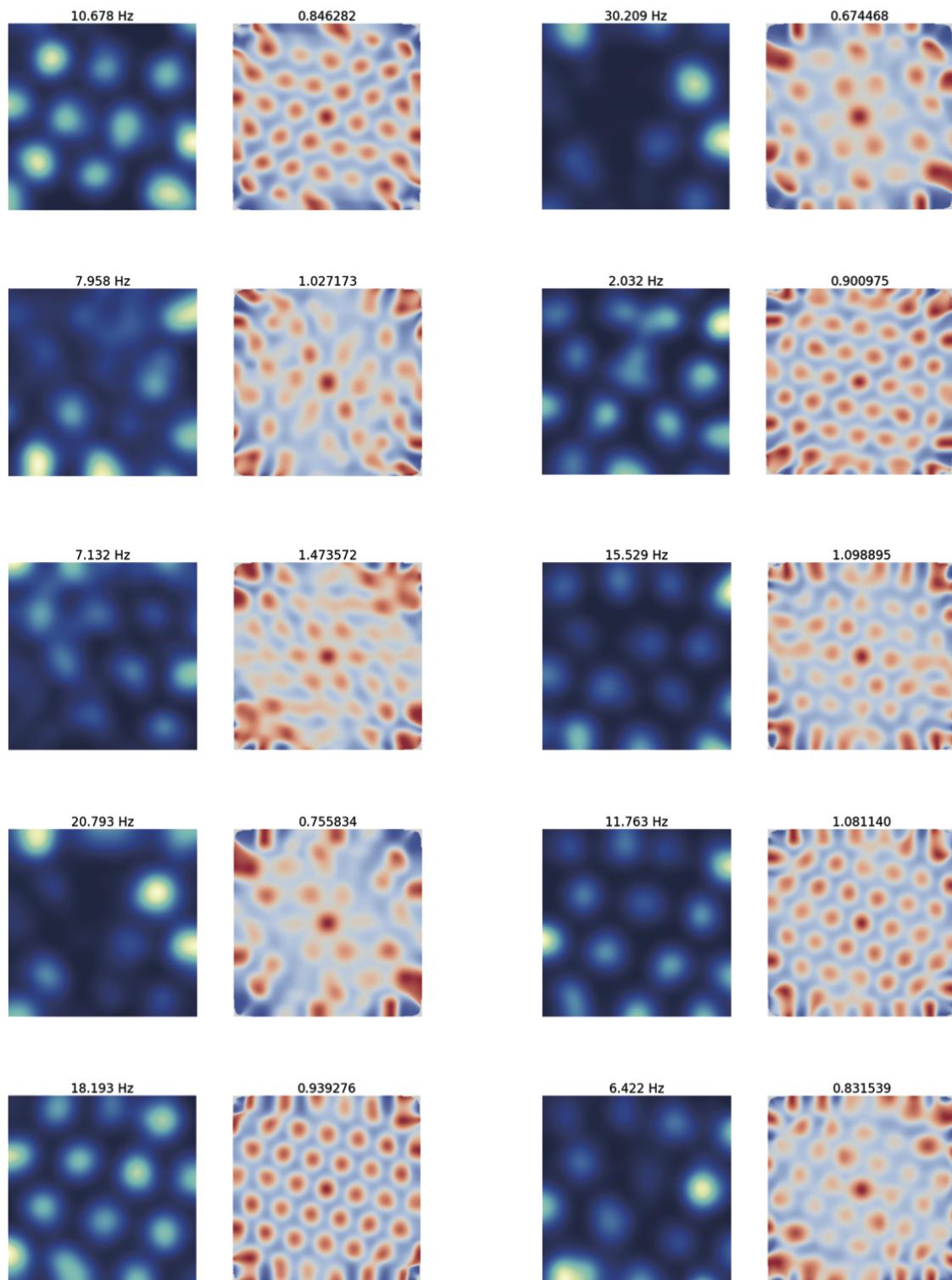


Figure A.3: Rate map and autocorrelogram of all grid cells included in data analysis. Part 3.

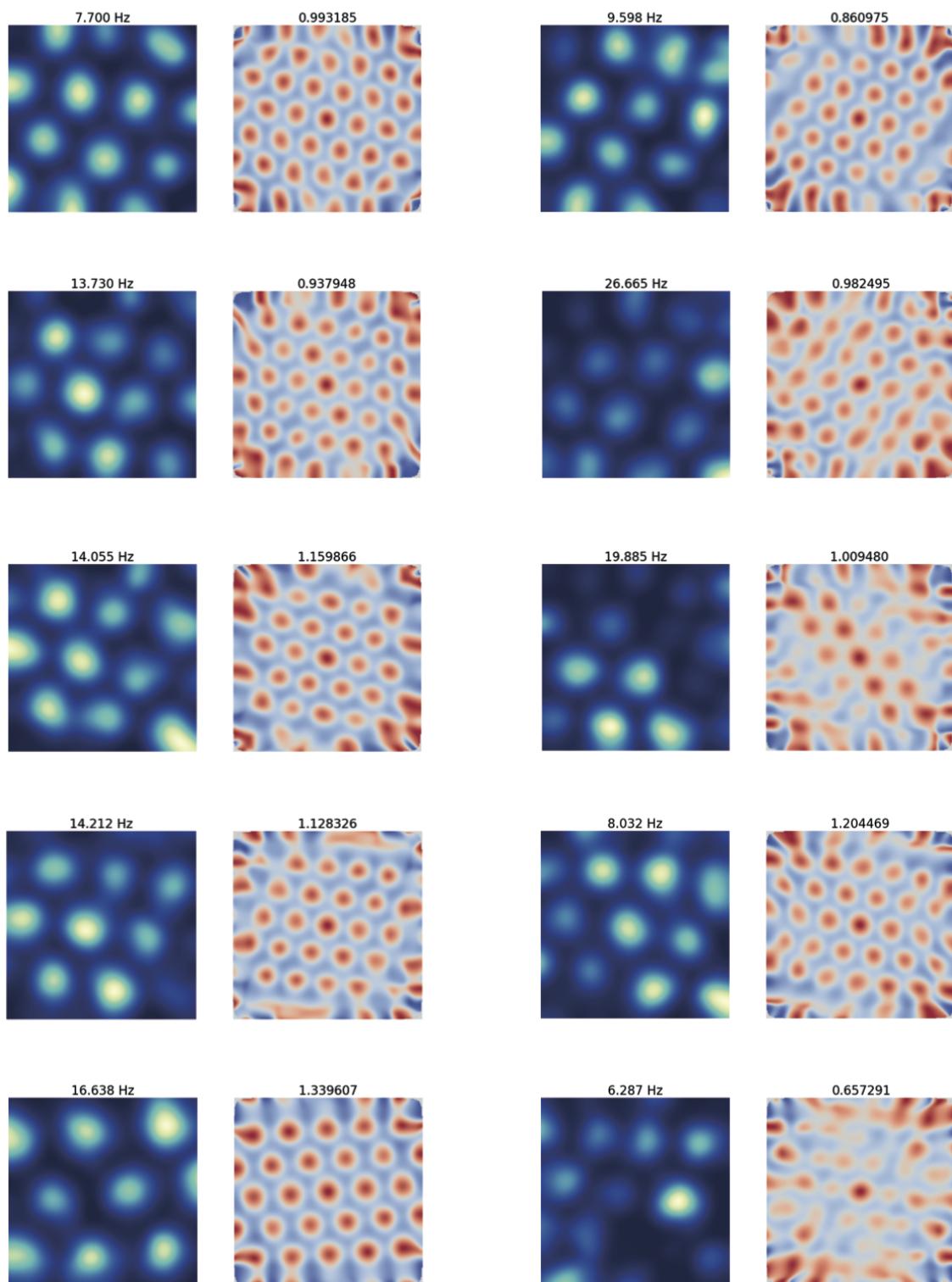


Figure A.4: Rate map and autocorrelogram of all grid cells included in data analysis. Part 4.

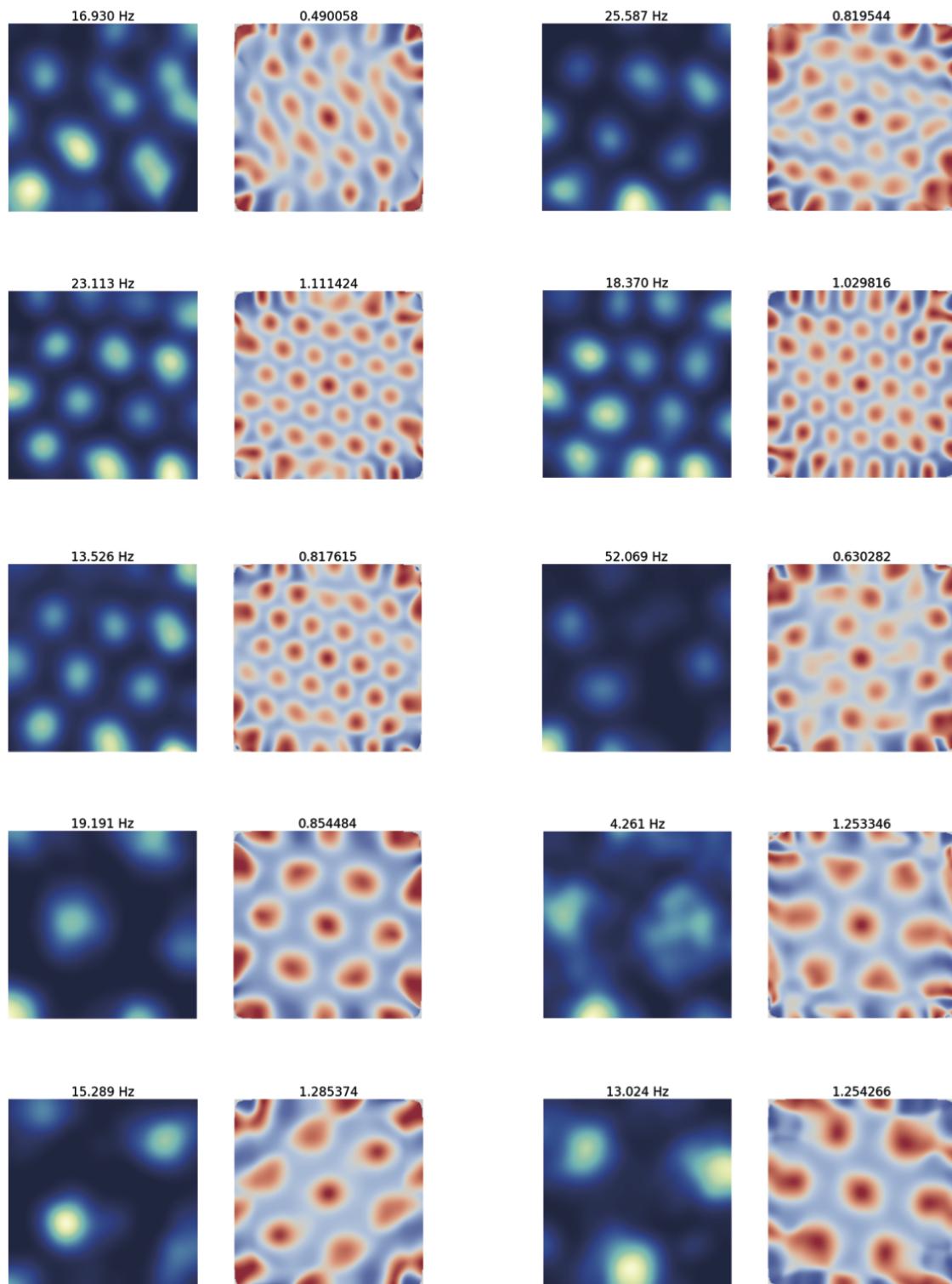


Figure A.5: Rate map and autocorrelogram of all grid cells included in data analysis. Part 5.

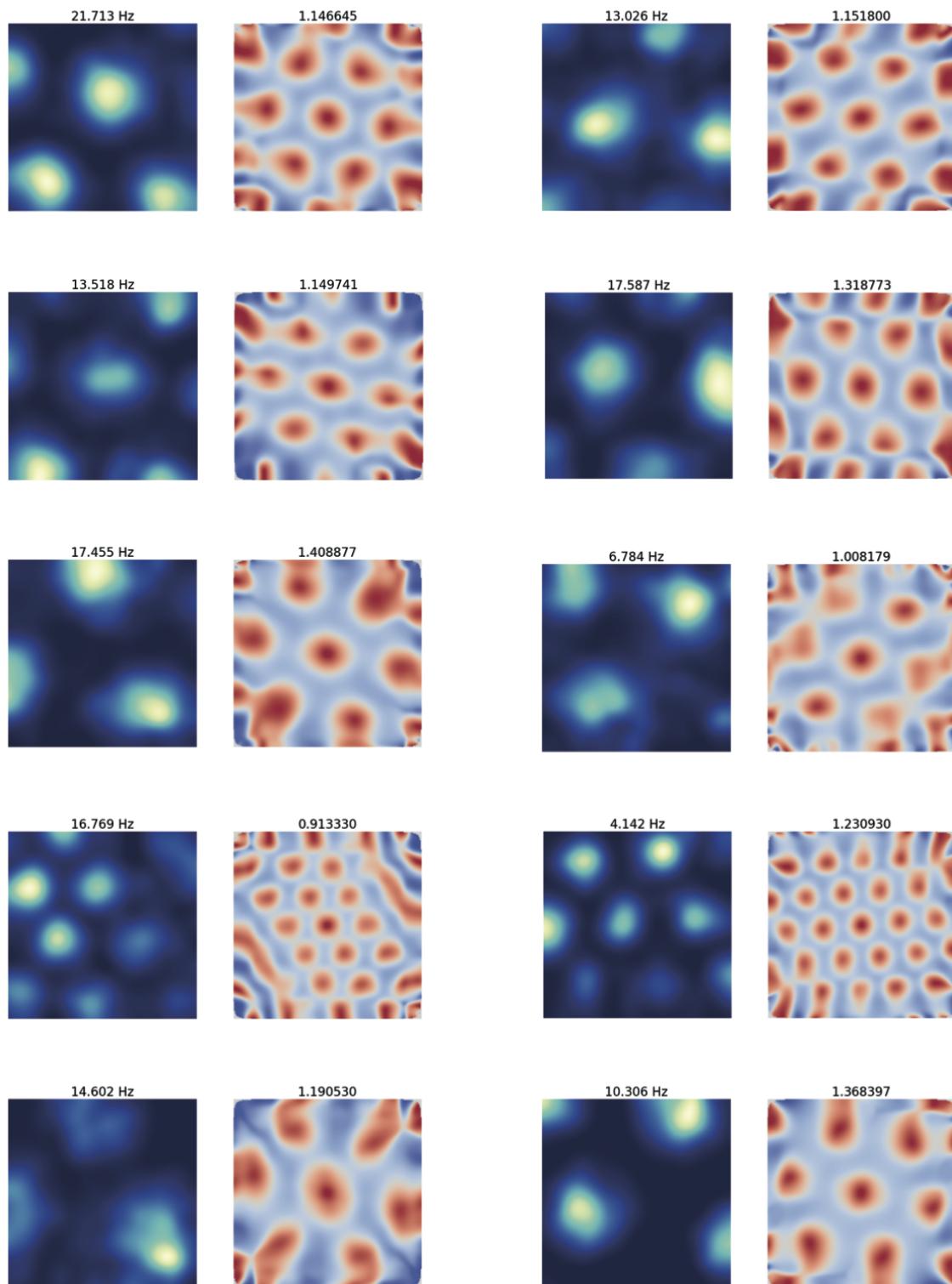


Figure A.6: Rate map and autocorrelogram of all grid cells included in data analysis. Part 6.

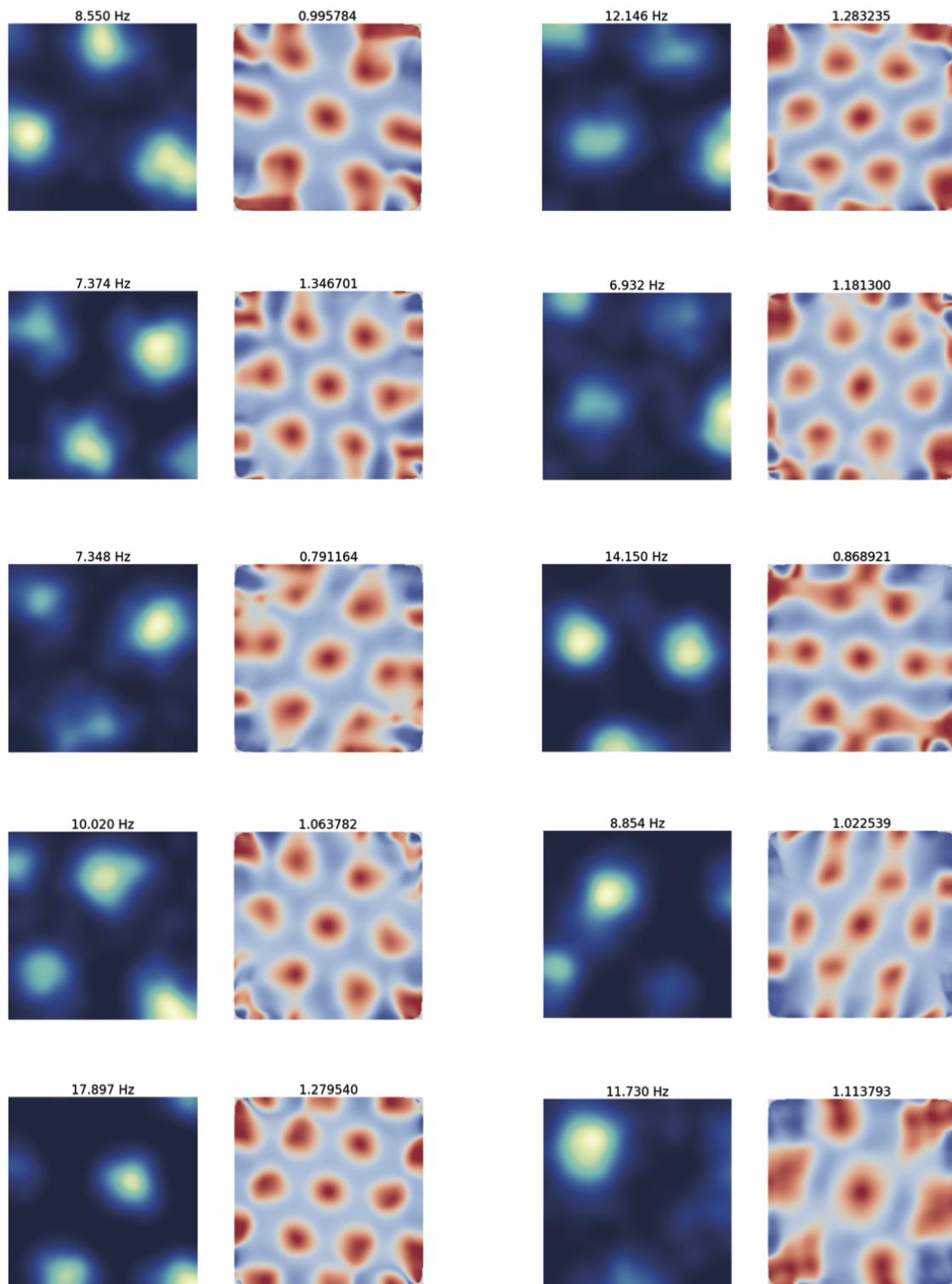


Figure A.7: Rate map and autocorrelogram of all grid cells included in data analysis. Part 7.

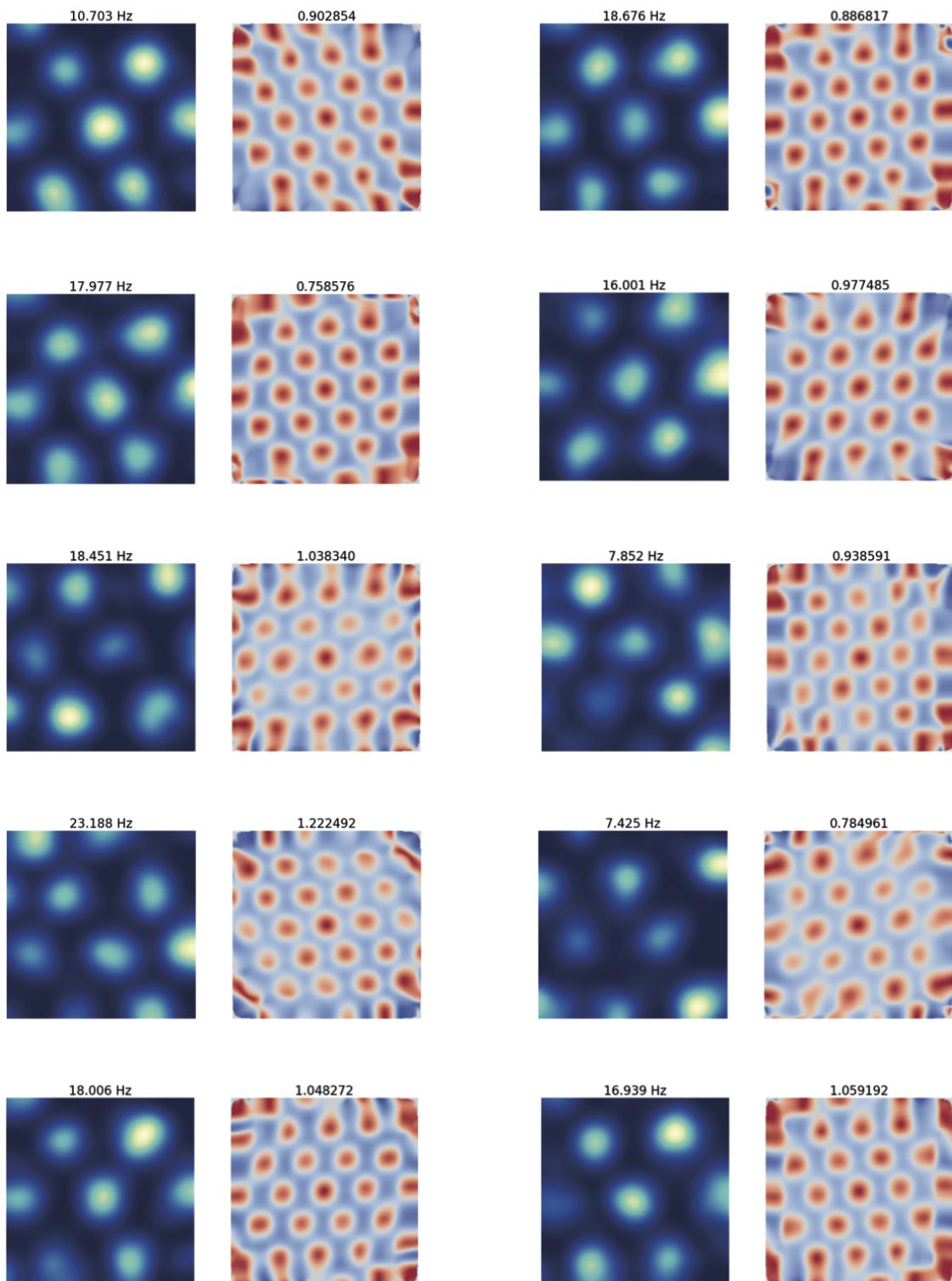


Figure A.8: Rate map and autocorrelogram of all grid cells included in data analysis. Part 8

## A.2 Tetrode recording locations

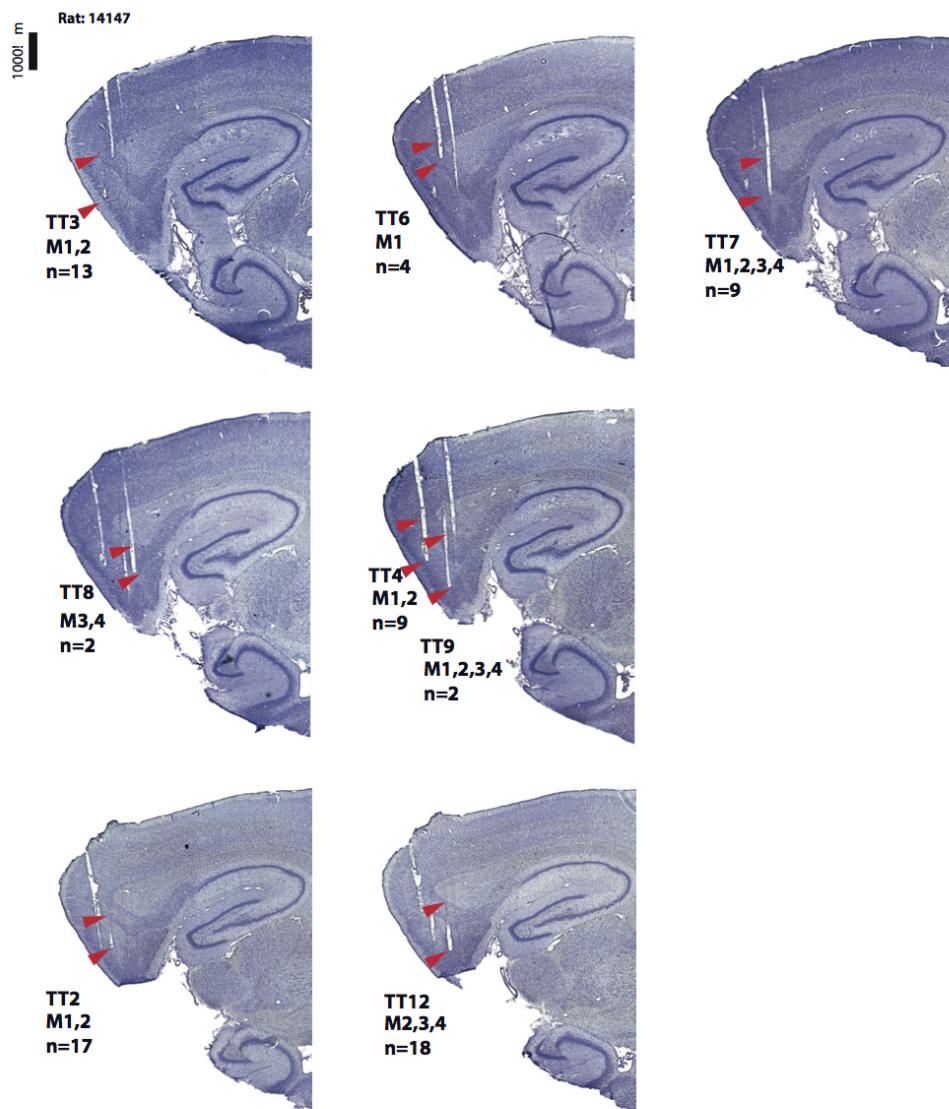


Figure A.9: The red arrow in this Nissl stained sagittal sections indicated the recording location of grid cells. The number of grid cells (**n**) and the number of modules (**M**) associated with each tetrode track (**TT**) is shown



## B | Publications

This work has been partially published as a peer-reviewed abstract and poster at Computational and Systems Neuroscience (Cosyne) 2015.

Huang, Ziwei; Solstad, Trygve; Dunn, Benjamin Adric. (2015) Interneurons in two population grid cell network. Cosyne Abstracts 2015, Salt Lake City USA.

# Interneurons in a two population grid cell network

Ziwei Huang \*

Trygve Solstad †

Benjamin Dunn \*

November 26, 2014

## Summary

The hexagonal firing pattern of entorhinal grid cells [1] could arise from a competitive mechanism mediated by interneurons [2, 3, 4]. Here we asked if a two-population continuous attractor model [5, 6], consistent with the proposed inhibitory connectivity pattern, could maintain grid cell firing even if interneurons (a) comprise less than 20% of the neural population and (b) lack spatial periodicity, as was recently observed in a sub-population of entorhinal interneurons [7, 8].

First, using non-negative matrix factorization (NMF) [9], we constructed two-population models with varying numbers of interneurons while maintaining the same effective connectivity between grid cells. Surprisingly, network drift decreased exponentially with the number of assumed interneurons and networks having less than 10% interneurons were able to accurately path integrate. The resulting connectivity was patterned with each interneuron receiving projections from either many grid cells with similar spatial selectivity or cells that together formed an inverted grid pattern. In both cases, grid cells with inhomogeneous peak firing rates had lower grid scores than the corresponding interneurons. Interestingly, thought to be outliers, a small number of interneurons with both high grid scores and spatial sparsity have also been observed experimentally [7].

Second, we considered a network where the connections from grid cells to interneurons were fixed to sparse random values, while back projections were found using NMF. In this case, the spatial selectivity of interneurons decreased dramatically as the variance in grid field firing rates

was increased. Although this network produced aperiodic interneurons similar to recordings, a considerably larger proportion of interneurons was required to reach the same level of stability which did not decrease exponentially as in the fully factorized case.

Further experiments should be able to determine if reality falls somewhere on the spectrum between these two simple cases.

## Additional details

We let each grid cell,  $s_i$ , and interneuron,  $u_i$ , follow the dynamics of a simple firing rate model

$$\tau \frac{ds_i}{dt} + s_i = g\left(\sum_j J_{ij}u_j + I_t\right)_+ \quad (1)$$

$$\tau \frac{du_i}{dt} + u_i = g\left(\sum_j K_{ij}s_j\right)_+ \quad (2)$$

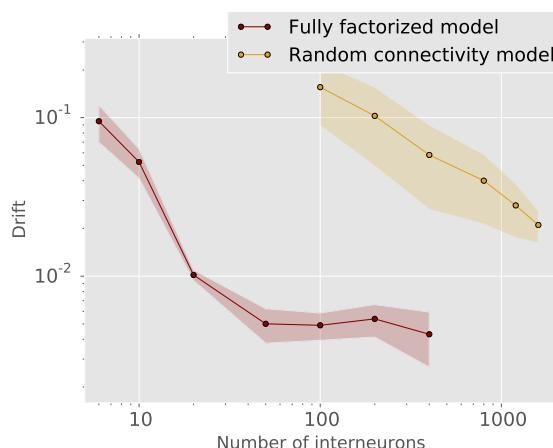
where  $(.)_+$  is the threshold-linear function,  $g$  the gain,  $\tau$  the neuronal time constant, and  $J_{ij}$  and  $K_{ij}$  the strength of connection from interneurons to grid cells and grid cells to interneurons, respectively. For the drift calculations  $I_t$  is a constant external input, while for the path integration tasks we include a term that depends on the time-varying speed,  $v_t$ , and direction  $\theta_t$ ,  $I_t = \text{constant} + \alpha v_t \cos(\theta_t - \theta_i)$ , where  $\alpha$  is the velocity modulation and  $\theta_i$  the preferred direction. To determine how spatial periodicity of the interneurons changes with respect to the variance in spatial fields of the grid cells, we included a population of neurons with place cell-like coding and excitatory projections to the grid cells  $I_t = \text{constant} + \sum_j H_{ij}P_j(t)$ . The grid field peak rates were varied by drawing the strengths of the connections,  $H_{ij}$ , from a distribution with increasing variance.

\*Kavli Institute for Systems Neuroscience and Centre for Neural Computation, NTNU

†Faculty of teacher and interpreter education, Sør-Trøndelag University College (HiST)

To determine possible connectivity patterns corresponding to the inhibitory portion of an effective connectivity,  $W_{ij}$ , between principal cells, we note that, at equilibrium,  $\frac{du_i}{dt} \rightarrow 0$  and we have  $\sum_j J_{ij}u_j \approx g \sum_j J_{ij} \sum_k K_{jk} s_k \approx \sum_j W_{ij} s_j$ . We can therefore find solutions to the problem  $W_{ij} = \sum_k J_{ik} K_{kj}$  for different numbers of interneurons using techniques from non-negative matrix factorization [9]. As in [2, 3, 4], we assumed a purely inhibitory connectivity with  $W_{ij} = W_0 H(R_{max} - d_{ij}) H(d_{ij} - R_{min})$ , where  $H(\cdot)$  is the Heaviside function,  $R_{max}$  the outer ring of the radial extent of the connectivity,  $R_{min}$  the inner ring,  $W_0$  the strength of the inhibitory interactions between connected neurons and  $d_{ij}$  the distance between cell  $i$  and  $j$ ,  $d_{ij}^2 = (x_i - x_j - l \cos \theta_i)^2 + (y_i - y_j - l \sin \theta_i)^2$  with  $x_i = 1 \dots N_x$  and  $y_i = 1 \dots N_y$  representing the position of neuron  $i$  in a two dimensional  $N_x \times N_y$  neural sheet with periodic boundary conditions and spatial offset  $l$ . For the networks with random projections to interneurons, each grid cell projected to each interneuron with a probability of 0.2. The strength of each connection was taken as the absolute value of a random number drawn from a normal distribution of variance 0.1. Using different values for the variance and sparsity of the random connections did not qualitatively change the results in figure 1.

For path integration we used  $N_x = 64$ ,  $N_y = 56$ ,  $l = 4$ ,  $I_{ext} = 0.1$ ,  $W_0 = -0.1$ ,  $R_{max} = 20$ ,  $R_{min} = 15$  and no spatial input. To evaluate how the drift depends on the size of the inhibitory population, we used a smaller model on a twisted torus [10] with  $N_x = 16$ ,  $N_y = 13$ ,  $l = 0$  and  $v_t = 0$ .



**Figure 1: Drift.** In the fully factorized case, drift decreased exponentially with the number of interneurons. For the case where interneurons received random input from grid cells, the network required a considerably higher proportion of interneurons to ensure a stable grid pattern. The relative proportion of interneurons necessary to acquire the same level of drift in the fully factorized model continued to decrease for larger networks of grid cells (not shown).

## References

- [1] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, E. I. Moser, *Nature* **436**, 801 (2005).
- [2] Y. Burak, I. R. Fiete, *PLoS computational biology* **5**, e1000291 (2009).
- [3] J. J. Couey, et al., *Nature Neuroscience* **16**, 318 (2013).
- [4] H. Pastoll, L. Solanka, M. C. van Rossum, M. F. Nolan, *Neuron* **77**, 141 (2013).
- [5] B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, M.-B. Moser, *Nature Reviews Neuroscience* **7**, 663 (2006).
- [6] M. C. Fuhs, D. S. Touretzky, *The Journal of Neuroscience* **26**, 4266 (2006).
- [7] C. Buetfering, K. Allen, H. Monyer, *Nature Neuroscience* **17**, 710 (2014).
- [8] Y. Roudi, E. I. Moser, *Nature neuroscience* **17**, 639 (2014).
- [9] D. D. Lee, H. S. Seung, *Nature* **401**, 788 (1999).
- [10] A. Guanella, D. Kiper, P. Verschure, *International journal of neural systems* **17**, 231 (2007).

