Reviewing the Attractor Model of Grid Cells

Albesa, A., Awasthi, J., Boulaabi, M., Chawla, D., & Paques, M.

05/03/2021

Abstract

Grid cells are location modulated neurons, which are activated periodically when the position of a subject coincides any of the equally spaced fields of activity. The model used for reproduction and analysis, proposed an only-inhibition attractor model for grid cells activity (Burak & Fiete, 2009). In this paper, the authors reproduced network hexagonal patterns, single neuron activity associated to a real rat path and confirmed several of the predictions of their model in the context of Attractor Theory. The objective of this work has been to investigate this groundbreaking model that was proven successful to describe grid cells activity. The aforementioned model relies on a continuous attractor network. At the same time, two variations in boundary conditions (periodic & aperiodic) and in base model (firing rate and spiking) are studied. We have been able to reproduce most of the results included in (Burak & Fiete, 2009), with the exception of single neuron activity for long ($\approx$ 20 min) path integration, due to computational limitations. On the other hand, we have included several elements of additional analysis, as can be linear stability, network response to parameters variation and robustness of the model to noise and neuron dropout. The limitations of the model in both mathematical and biological terms are discussed and show how this model is likely to be incomplete. In this sense, a last section describing recent extensions and findings regarding this model is included.

Introduction

History of Grid Cells

Since the 1950s, neuroscience researchers have been fascinated by how neural cells selectively respond to certain features in the environment and the underlying maps that seems to exist between external features, population activity and single cells (Hubel & Wiesel, 1959). During this time, computational and neural models of sensory information were developed to increase our understanding motivation behind goal and reward directed spatial behaviour (Hubel & Wiesel, 1962). With this, we have been able to understand the transformation of sensory information from receptor level to the cortex .

To understand the transformation of sensory system we must focus on The region responsible for mediating sensory input, neural characterisations and behavioural representations is the association cortex. Within this area, one of the few parts we have relatively good understanding of is the medial entorhinal corext (MEC). The MEC is an area in the brain located in the medial temporal lobe that mediates as a wide functional network for memory and navigation. In particular, activity in the MEC has shown correlational activity with position and orientation within the environment (Knierim et al., 2014). John O'Keefe and Dostrovsky were the firsts to discover elements of a "localization system" in the brain, which allows orientation in environment (O'Keefe & Dostrovsky, 1971). The authors discovered cell activity in the hippocampus when animals were in specific locations within the environment (O'Keefe & Dostrovsky, 1971). This was a collection of localised cells, which covered specific parts of the environment, known as *place field* (O'Keefe, 1976). This raised the idea of *place cells* that aided spatial neural mapping representing the place field to aid navigation abilities. Building upon this work, Moser and Moser explored the presence of place cells outside of the hippocampus, by implanting electrodes in the dorsocaudal part of MEC (Hafting et al., 2005). This region was selected as MEC projects to dorsal hippocampus, where majority of the place cell activity occurs (Witter et al., 1989). It was found that cells did not fire at random locations in the environment but followed a periodic pattern instead. In 2005, Moser and others described that the firing fields of cells formed hexagonal-grid-like patterns throughout the surface space, giving the term "grid cells". Grid cells apply a modulo operation in

the physical space, such that a cell firing would mean that the animal is in one of the *potentially infinite* positions that maps to a particular class. Thus, grid cells equip the physical space with an equivalence relation that depends on the geometry of this triangular lattice. Grid cells have been found to be involved in measuring movement and distance in mice and other animals (Fyhn et al., 2008; Yartsev et al., 2012). These findings earned John O'Keefe, May-Britt Moser and Edvard Moser the Nobel prize in Physiology and Medicine in 2014. The prize went to John O'Keefe for the discovery of place cells in the hippocampus and to May-Britt Moser and Edvard Moser for the discovery of grid cells in the MEC.

Biology of Grid Cells

It is known that the MEC is a part of the dorsal pathway and allows for processing of spatial information (Mishkin et al., 1983). The ventral and dorsal pathways converge and distribute spatial information within the hippocampus, by forming spatial representations in episodic memory (Witter et al., 1993). Though unlike place cells the hexagonal grid cells map is activated across environments irrespective of landmarks, suggesting the role of path integration (McNaughton et al., 2006). Path integration is calculated with geometric vectors that are stored and continue to be updated with position within the environment.
Further evidence of path integration in grid cells is unveiled as a recent research shows that disruption of grid cell activity and impairment of path integration is achieved by removing NMDA glutamate receptors in new born mice (Gil et al., 2018). This research revealed that spatial selectivity was reduced and disrupted normal firing rates of grid cells. Furthermore, the results also revealed that global theta frequency modulation, speed selectivity and head direction selectivity were preserved. These results provide a better understanding of how grid cells support spatial memory and navigation. More research into the MEC show that spatially sensitive cells are modulated by theta frequency oscillations by local field potential (Pastoll et al., 2013).
Furthermore, in the second layer of MEC is made up of two distinct populations of principle cells, the pyramidal cells and stellate cells and it is specifically the latter that are of interest. They provide a key excitatory input from the MEC to the hippocampus (Ray et al., 2014). Intercellular recordings of stellate cells in research on mice running within a linear virtual field and exploring open space showed grid cell activity (Rowland et al., 2018). This implies a role for stellate cells in the maintenance and formation of grid cell and place cell activity within the hippocampus, therefore, showing the importance of stellate cells in path integration (McNaughton et al., 2006). However, this questions the role of grid cell activity in spatial firing at the hippocampus level and whether stellate cells are indeed grid cells. Nevertheless the role of stellate cells still remains unclear as there is evidence which does not suggest an association between grid cell activity and setllate cells (Kropff & Treves, 2008).

Models of Grid Cells

Several approaches have been taken to describe and explain the behavior of grid cells in mathematical terms. In the literature, two different models that account for grid cells activity can be found (Giocomo et al., 2011). The first type are called *oscillatory - interference* (OI) models, and the second *attractor* models, which will be discussed more in depth later.
In OI models, the integration of the velocity inputs through geometric vectors (both models assume that this input exists and it is both proportional to speed and parallel to direction of movement) is done through interference between oscillators, the frequency of which depends on this quantity. This results in a particular oscillatory phase which is able to temporally integrate velocity (*rate-to-phase* transformation) at the single-neuron level. On the other hand, attractor models propose that the activity of these cells does not depend solely on the input of sensory cues related to velocity, but also on the input received by other grid cells within the network. Through a Recurrent Neural Network, the activity of each neuron rises from a collective neural state shaped by both external inputs and the connectivity profile.
Since the initial presentation of the models, each approach has been particularly well-suited for explaining certain experimental observations. This has pushed researchers to revise the models, with the intention of improving them and extending their biological plausibility. Nevertheless, as of today, a consensual model that can account for all experimental observations has yet to be established.

Methods

Network Models

Firing Rate (FR) Network

Burak and Fiete (2009) propose a firing rate model with a ReLU transfer function:

$$\tau \frac{dS}{dt} = -S + \Big[W \cdot S + B \Big]_+ \tag{1}$$

with $S = (s_1, s_2, ..., s_N)$ the firing rates of each neuron and $[x]_+ = \max(0, x)$. Thus, the neural population can be modelled as a system of N differential equations, such that the i^{th} neuron self-decays towards the i^{th} output of the transfer function.

The connectivity matrix is defined as follows:

$$W_{ij} = a\mathrm{e}^{-\gamma|\boldsymbol{x}_i - \boldsymbol{x}_j - l\cdot \boldsymbol{e}_{\theta_j}|^2} - \mathrm{e}^{-\beta|\boldsymbol{x}_i - \boldsymbol{x}_j - l\cdot \boldsymbol{e}_{\theta_j}|^2} \tag{2}$$

with

$$\boldsymbol{x}_k = \Big(-n/2, -n/2 \Big) + \Big(k \ mod \ n, (k - k \ mod \ n)/n \Big) \ , \ n = \sqrt{N} \tag{3}$$

Equation (3) maps the label of each neuron $k \in [1, N] \subset \mathbb{N}$ to a position in a discrete grid $\boldsymbol{x}_k \in [-n/2, n/2) \times [-n/2, n/2) \subset \mathbb{Q}^2$. The distance defined by the norm in (2) corresponds to that within a torus, which plays a key role in providing the network with periodicity (by doing this, the opposite corners of the positional grid have a 0 distance). γ is chosen to be 1.05β, with $\beta = 1/\lambda^2$ (λ is called the *periodicity* of the lattice and will be shown to play an important role in its geometry). The vector $\boldsymbol{e}_{\theta_j}$ is called the *preference vector* of neuron j. Each neuron within the grid is assigned a *preferred direction* in repeated clusters of 4 units $\{(N, S, E, W) \rightarrow ((0, 1),\ (0, -1),\ (1, 0),\ (-1, 0))\}$

The network has a bias B that plays two main roles: coupling the system to an input velocity $\boldsymbol{v}(t) \in \mathbb{R}^2$ and defining an envelope that can be modified to introduce boundary conditions. The latter is achieved by driving the activity to zero near the edges of the grid defined by (3) and break toroidal topology:

$$B_i = A(\boldsymbol{x}_i) \cdot (1 + \alpha(\boldsymbol{v}(t) \cdot \boldsymbol{e}_{\theta_i})) \tag{4}$$

The second multiplicative term in (4) lies in the range $1 \pm \alpha|\boldsymbol{v}(t)|$, with the maximum corresponding to the velocity being parallel to the preference vector and the minimum to these being anti-parallel. $A(\boldsymbol{x})$ is defined as $\mathbf{1}$ for the periodic conditions of the network, and as follows for the aperiodic case,

$$A(\boldsymbol{x}) = \begin{cases} 1 & |\boldsymbol{x}| < R - \Delta r \\ \exp\Big[-a_0 \big(\frac{|x| - R + \Delta r}{\Delta r} \big)^2 \Big] & \text{otherwise} \end{cases}$$

which defines a region of diameter $R - \Delta r$ wih a value identical to the periodic case and a complementary region that exhibits Gaussian decay.

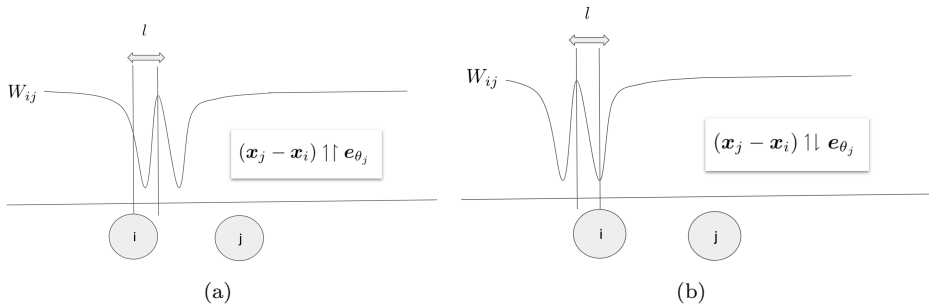


Figure 1: The connectivity matrix is defined to create a driving force in all four directions. In the absence of a velocity input, the symmetry of the system makes the net force acting on the network zero. When the system is coupled to a velocity different than zero, the symmetry is broken and the pattern evolves accordingly.

Implementation The network has been modeled in MATLAB®, with a backend layer written in C++ for some matrix and vector operations. We use *ode23* to dynamically evolve a vector S according to (1). This is done with a parameter *MaxStep* set to 20 ms, corresponding the maximum resolution in time for the velocity input. This resolution depends on the timestamps for the biological recordings described in (Hafting, 2014).

Spiking (SP) Network

The spiking network is an extension of the FR network that takes into account the presence of noise in the neural system. In this context, the network follows spiking dynamics (equation (5)), where stochasticity is achieved by sampling the firing times t_μ from an *Inhomogenous Poisson Process*.

$$\tau \frac{dS_i}{dt} = -S_i + \sum_\mu \delta(t - t_\mu) \tag{5}$$

The dynamics are described by an ODE composed of a self inhibition term and a gain term. If the neuron spikes at $t = t_\mu$, its activity is incremented by $1/\tau$. Otherwise, in the activity experiences exponential self-decay (Algorithm 1, lines 15-19). Spikes are sampled from an inhomogeneous Poisson distribution, which is characterised by having a time-dependent rate constant. In this model, the rate for each neuron at time t depends on neural activity $r_t^{(i)} = M\big[f(\sum_j W_{ij} s_j(t) + B_i(t))\big]$. This can be implemented discretely by splitting the time into intervals Δt with a rate $r_{\Delta t}^{(i)}$. The firing rate is thus updated for each interval but remains constant within them, thus generating a globally inhomogeneous but locally homogeneous PP.

In the spiking model, boundary conditions highly influence the network performance in generating a coherent trajectory. However to obtain satisfying results with the aperiodic bounds we had to reduce the variance of the firing train (not simulated, information retrieved from (Burak Fiete, 2009)). To meet this condition of inter-spikes regularity we resort to a decimation method. Each time interval Δt is divided into M sub-intervals of size $\Delta t/M$ (line 8). An homogeneous PP with rate $M \cdot r_{\Delta_t}^{(i)}$ is then generated on each sub-interval to produce a spike train (line 10-14). Finally, every spikes but the M^{th} one is decimated (line 15-16). This sub-Poisson process guarantees a control of the coefficient of variation of the inter-spike interval following $CV = 1/\sqrt{M}$.

Algorithm 1 Spiking Network

1: **procedure** SPIKING POISSON PROCESS($M, \Delta t$)
2: Initialize S_i in $\mathcal{N}(0,\, 0.1)$
3: Initialize $time\_intervals = [..., [t_i, t_i + \Delta t], ...]$
4: Initialize $spikes\_count^{(i)} 0$
5: **for** *interval* **in** *time_interval* **do**
6: $r_{\Delta t}^{(i)} = M\big[f(\sum_j W_{ij} \cdot s_j(t) + B_i(t))\ \big]$
7: map $interval \longrightarrow [..., [t_i + (k-1)\frac{\Delta t}{M}, t_i + k\frac{\Delta t}{M}], ...]$
8: sample M numbers from $U(0, 1)$
9: $r_{\delta t}^{(i)} = r_{\Delta t}^{(i)}/M$
10: **for** each sub-interval $\delta t^{(m)}$ **do**
11: **if** $random^{(m)} \leq r_{\delta t}^{(i)}$ **then**
12: $spikes\_count^{(i)}$ += 1
13: **end if**
14: **end for**
15: **if** $spikes\_count^{(i)} \geq M$ **then**
16: Remove M from $spikes\_count^{(i)}$
17: $S_i(t + \Delta t) = S_i(t) + 1$
18: **else**
19: $S_i(t + \Delta t) = S_i(t) \cdot \exp(-\Delta t/\tau)$
20: **end if**
21: **end for**
22: **end procedure**

Analysis of Results and Code for Reproduction

There is a GitHub repository available for reproducing results presented in this report (Albesa et al., 2021). The project structure can be found in the README.md file. Some of the methods and algorithms are outlined now:

Steady States and Linear Stability

A numerical approach can be followed in order to find the steady states of the FR system. This can be done by setting the temporal derivative to zero in (1). A first approximation can do this by assuming that in the steady state the input of the transfer function is non-negative for all neurons, so that the ReLU transformation is actually the identity. In this case:

$$\frac{dS}{dt} = 0 \Longleftrightarrow S = WS + B \Longleftrightarrow (I - W)S = B \Longleftrightarrow S = (I - W)^{-1}B \tag{6}$$

The computation of the S vector is straightforward. Another, more general, approach consists on finding the zeros of the temporal derivative (and capturing the nonlinearity induced by $[x]_+$) with a function such as *fsolve*, found in the Optimization Toolbox$^{\text{TM}}$.

We perform Linear Stability Analysis (LSA) on both the steady states found numerically by the procedure above described and experimentally by letting the network evolve until stabilising. We denote these as S_{ss} and S_{exp} respectively. We define small perturbations of the kind $S = S_0 + u(t)$, and assume $u(t) = e^{\omega t}v$. By doing this, one can arrive to a time-independent equation that describes the ω spectrum in terms computable terms:

$$(\tau\omega + 1 - \gamma)v = 0 \Longrightarrow \omega_i = \frac{\gamma_i - 1}{\tau} \tag{7}$$

where γ_i correspond to the N eigenvalues of the *modified weight matrix* [$\tilde{W_{ij}} = f'(X_i)W_{ij}$, $f' = (ReLU)' = \Theta$ (Heaviside function), $X = WS_0 + B$].
In this context, the state S_0 is stable when the amplitude of the complex oscillations defined by $e^{\omega_i t}$ decreases with time, i.e. when $\text{Re}(\omega_i) < 0$ for all ω_i. Should exist values $\omega_i > 0$, each would drive the network in the direction of the associated eigenvector.

In Vivo Data Processing

We have used the open database (Hafting, 2014) in all our results, that includes recorded positions and times for a rat moving freely inside a box, together with the firing times of different neurons in the MEC. In order to couple the position of the rat to the network we have discretely computed the velocity of the rat at each interval. Data processing also included removing NaN values and performing a firings$(t) \rightarrow$ firings(x, y) mapping. This is done in process_path.m script.

Lattice Analysis

The neural activity steady state appears as an hexagonal grid composed of bumps arranged in an hexagonal pattern. Bumps represent "peaks" of higher activity. This characteristic pattern can be described by the 3 main directions of the hexagonal grid $\{\pi_1, \pi_2 \text{ and } \pi_3\}$ and the average distance between the aforementioned bumps. Two functions (orientation.m and lattice_distance.m respectively) return these values. The first performs a 2-dimensional discrete Fourier transform on the image resulting from the pattern in order to obtain $\{\pi_i\}$ (it should be noted that in the Fourier space these are shifted by $\pi/2$). The second function smooths (averages) maxima belonging to the same bump) and then computes the average distance between the maxima in the pattern.

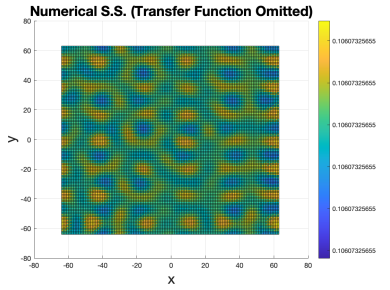
Path Reconstruction

This process is divided in two steps. In the first, we measure the correspondence between pattern and real path displacements (conversion factor). In the second, we measure pattern displacements of a network coupled to the velocity of a real path and reconstruct it by using the conversion factor found in the first step. In both cases we measure pattern displacements by selecting a bump from the pattern and keeping track of its center.

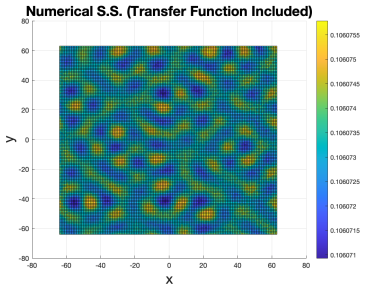
Results

Steady State and Linear Stability Analysis

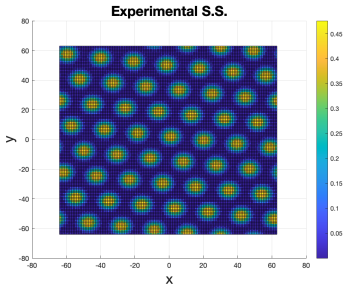
We have obtained a numerical steady state using two different approaches: matrix inversion (transfer function approximated to be the identity) and *zero search* approach through *fsolve* function. Both approaches have found solutions close to the trivial (figures 2(a), (b)) and have failed to find any of the stable patterns observed experimentally by letting the network evolve in time until stabilizing (figure 2(c)). Linear stability analysis confirms this by showing a contrast in the spectral decomposition of the exponential solutions of the equation: in figure 2(f) all eigenvalues are strictly less than zero, in opposition to figures 2(d), (e) where one can see how some of the values cross the 0 frontier.



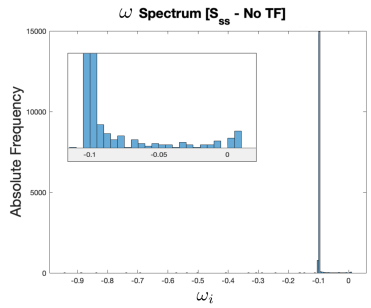
(a) numerical s.s. representation (matrix inversion)



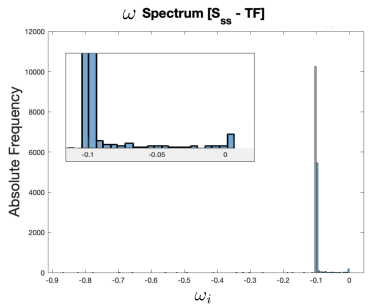
(b) numerical s.s. representation (*fsolve*)



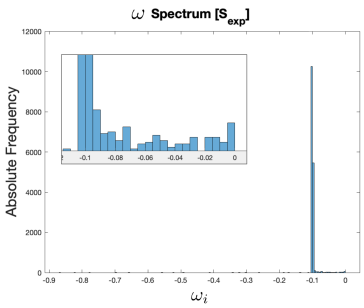
(c) experimental s.s. (pattern stable in time)



(d) eigenvalues distribution in LSA (s.s. in 2(a))



(e) eigenvalues distribution in LSA (s.s. in 2(b))



(f) eigenvalues distribution in LSA (s.s. in 2(c))

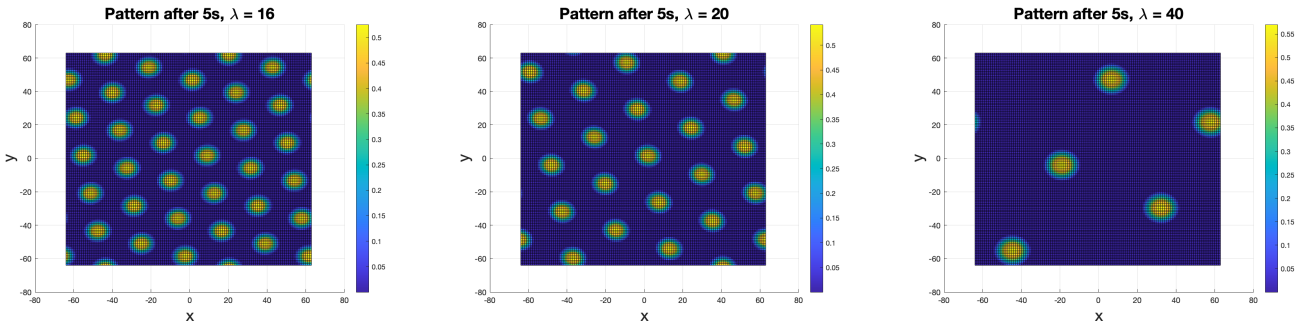
Figure 2: Numerically approximated steady states (a, b) and computationally obtained stable pattern (c) network activity. Spectral distribution of exponential solutions to perturbations (d, e, f).

Different λ values affect lattice dimensions

It is reasonable to think that the λ parameter has an impact on the geometry of the lattice, as it defines the range for inhibition of each neuron. Moreover, by inspecting (2), one can see a how inhibition depends on a *scaled distance* $|\boldsymbol{x}|/\lambda$:

$$W_{ij} = e^{-\frac{1.05}{\lambda^2}|\boldsymbol{x}|^2} - e^{-\frac{1}{\lambda^2}|\boldsymbol{x}|^2} \approx \frac{|\boldsymbol{x}|^2}{\lambda^2} - 1.05\frac{|\boldsymbol{x}|^2}{\lambda^2} \propto \left(\frac{|\boldsymbol{x}|}{\lambda}\right)^2 \tag{8}$$

where the approximation corresponds to a Taylor expansion of the exponential terms around 0 (for values $\gg 0$ both terms are approximately equal and $W_{ij} \approx 0$).
Figure 3 shows that increasing λ does correspond to a dilated distance and the subsequent increment in the neuronal distance between bumps. We computed the average distance $\bar{d}$ (in the neuronal grid space, units in neurons) between maxima within a lattice for different λ values (table 1) and observed the predicted linear relation.



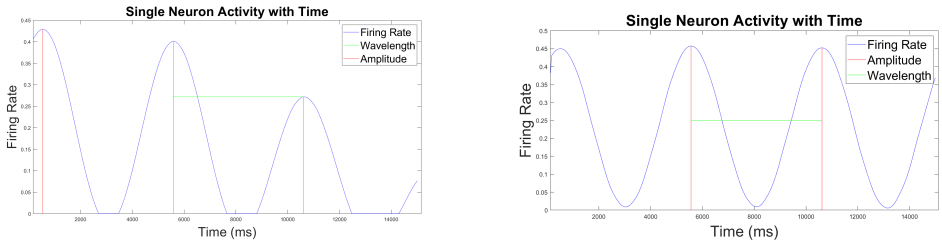
(a) (b) (c)

Figure 3: Influence of the periodicity parameter λ on the stable steady state. From left to right: $\lambda = 16$, $\lambda = 20$ and $\lambda = 40$.

Table 1: Lattice analysis

lattice response to λ					
λ	13	16	20	25	40
d (neurons)	18.85	23.46	28.61	35.43	61.04
regression parameters					
intercept ($\hat{\beta}_0$)	slope ($\hat{\beta}_1$)	R^2	$\hat{s}_{\hat{\beta}_0}$	$\hat{s}_{\hat{\beta}_1}$	
-2.1	1.560	0.997	1.3	0.053	

Single Neuron shows periodic activity for a constant velocity input



(a) SN activity, input velocity in x axis (b) SN activity, input velocity in pi_1 axis

Figure 4: SN activity for a constant velocity in a horizontal (a) and π_1(b) directions.

Figure 4 shows the activity profile with time of a particular neuron pinned in the network. In both subfigures there was a constant velocity input of magnitude 0.3 m/s, with the difference that in 4(a) the velocity was parallel to the horizontal axis and in 4(b) parallel to the axis associated to π_1. In the 4(b), one can see the effect of translating with time the ideal spatial sinusoids that appear in the preferred directions of the lattice, whereas in the 4(a) a damping effect is observed. Nevertheless periodicity is maintained, as one would expect from visual inspection of the generated pattern. We performed a *sin* fit to the curve plotted in 4(a) and obtained a figure that perfectly overlapped the original graph.

Spiking Neural Network is sensitive to multiple parameters

The stochasticity of the spiking network results in drift and rotation of the grid pattern even in the absence of velocity input. Both aperiodic and periodic networks show a significant drift of 20cm after a dozen of seconds (figure 5(a)). In figure 5(b), only the aperiodic network undergoes perturbation of the grid orientation. The regularity of the Poisson process described in the methods can be modified through M, which

influences the coefficient of variation of the spike train.

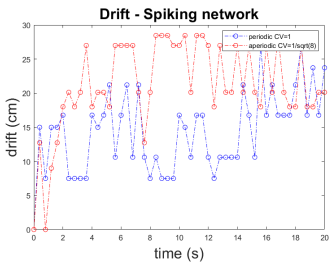
Figure 5(c) shows that reducing the CV of spiking allows to increase the network stability. Plotting the mean square drift versus the time elapsed reveals a linearity between the two values. We call D_{trans} the proportionality coefficient:

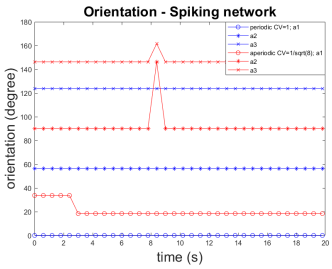
$$\langle \Delta x^2 \rangle = D_{trans} \Delta t \tag{9}$$

Moreover, the translation coefficient D_{trans} shows proportionality with CV^2 and inverse proportionality with the number of neurons (figure 5(d)).

$$D_{trans} \propto CV^2/N \tag{10}$$

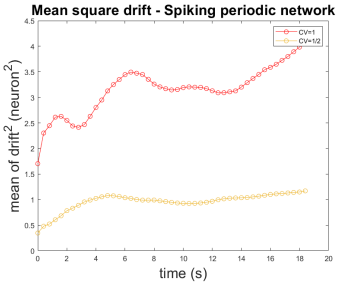
Finally, it has to be said that equivalent relations as (9) and (10) can be found regarding the network orientation evolution for the aperiodic spiking network. However, showing these relations would have asked to run simulations for a substantial time (3000s) as the orientation evolution is a far slower mechanism than the drift. Computational limitations restrained us to avoid such simulation times and therefore to prove any of the aforementioned relations.

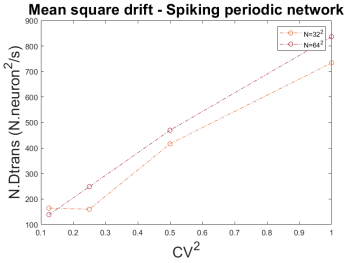




(a) Drift of the reconstructed path: Aperiodic Periodic

(b) Orientation evolution of the grid: Aperiodic Periodic



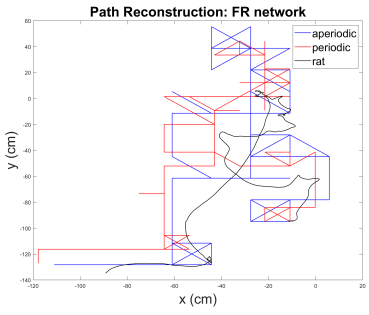


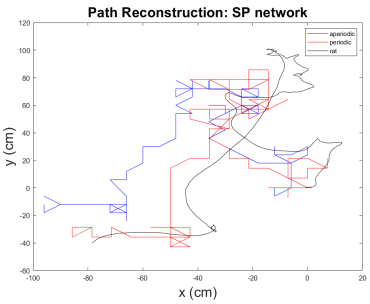
(c) Mean of the $drift^2$ of the grid: CV=1 CV=0.5

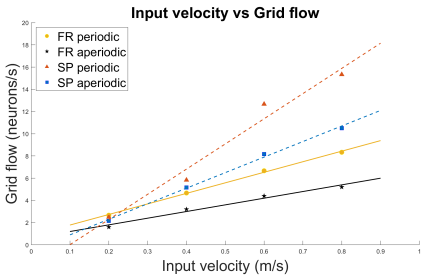
(d) Translation coefficient vs CV^2 : $N = 32^2 N = 64^2$

Figure 5: Stability analysis of the Spiking Network in absence of velocity inputs for N=64^2

In Vivo path can be reconstructed by decoding Neural Activity







(a) Path Reconstruction - FR Network

(b) Path Reconstruction - SP Network

(c) Response to Velocity Input

Figure 6: Reconstructed path from decoded activity in the FR (a) and SP (b) networks. Regression curves for network flow sensitivity to velocity input (c)

The path can, within reasonable precision, be reconstructed from network activity decoding. Adding aperiodic boundaries makes the network resistant to translations in activity, thus resulting in a bigger number of cm per neuron translated with respect to the periodic case (table 2). The distance from maximum to maximum in the lattice (d_{latt}) can be interpreted as the value λ_α discussed in the following section, and thus can be interpreted as the order of magnitude for which this particular network input to place cell is informative. It is worth mentioning that this is also the reason why path reconstruction has been possible: the range of motion of the rat is approximately the value of d_{latt}. Should it be the case d_{latt} was much bigger than the motion range the error could be bigger than the displacements, and should it be much smaller there would be constant discontinuities in the decoded position.

Table 2: Network sensitivity to input velocity

network	neurons/s vs m/s slope ($t_{0.05;2}$)	cm/neuron	d_{latt} (cm)
FR - Periodic	9.5±1.0	10.5±1.1	198±21
FR - Aperiodic	6.0±2.2	16.7±6.2	310±120
SP - Periodic	22.7±9.7	4,4±1.9	83.17±36
SP - Aperiodic	14.0±2.0	7.1±1.0	134±20

Model Limitations

Lattice and Network Dimensions vs Network Capacity

Grid cells within a particular network lack of a system of reference: by decoding population activity one can only infer where the rat (or any other agent) is within an infinite number of possibilities. In Fiete et al. (2008), joint networks of different geometries are studied as systems capable of storing the information about a rat position within a remarkable degree of precision. In this model, the network performs a modulo operation on the real position of the rat. To have a better understanding of the model, it is relevant to review a 1-dimensional example. In this context, given a network α with a λ_α distance between maxima of activity, and a position x, one can define the following map:

$$x \to x_\alpha = x \bmod \lambda_\alpha \tag{11}$$

where the mod (modulo) operation outputs the remainder of the integer division between x and λ_α. For N different networks, the position x can be mapped to $x = (x_1, x_2, ..., x_N)$ This representation of quantities has been studied in mathematics and is called Residue Number System (RNS). In this sense, a result that can be of high interest is the Chinese Remainder Theorem (CRT) which, given a set of co-prime numbers $\{\lambda_\alpha\}$, ensures a unique decomposition x for every value between 0 and $\left[\prod_{\alpha=1}^{N} \lambda_\alpha - 1\right]$
We have found that the minimum λ_α that the network can resist without alternative solutions to become more stable than the hexagonal pattern, although decreases, quickly saturates with N^2. In biological terms, this exposes a neural limitation on the precision of the representational space that comes from the number of neurons that participate in the network.

Robustness of the Attractor Model

This section has two main goals: (i) to to test the robustness of the network to the imperfections that one would want in a realistic biological model and (ii) to test if results are consistent with predictions of the attractor model.

Adding Gaussian Noise to the Weights Matrix

We have tested the model's robustness to noise in the network connections by randomly altering the weight matrix. This was implemented by using a perturbed weight matrix:

$$W'_{ij} = W_{ij} + \eta_{ij} \ , \ \eta_{ij} \sim \mathcal{N}(0, \sigma^2) \tag{12}$$

Here σ^2 corresponds to a percentage of the variation of the original weight matrix. We ran simulations from 0.1% to 5% of the variation and observed that up-til 2% applied noise the model could recover from the perturbations and return to the hexagonal steady state. However beyond 2% noise value the model exhibited 1000 fold firing activity which is not possible in case of biological neurons. These results indicate that the limit of tolerance to noise is around 2% of variation. In Seung (1996), the timescale for activity decay (τ_{act}) is related to the synaptic timescale (τ_{syn}) and percentage of weight miss-tuning (Δw): $\tau_{act} = \frac{\tau_{syn}}{\Delta w}$ The obtained limit value for Δw is likely to increase with neuron size, giving rise to a more robust network. If one considers the evidence regarding the role of NAMDA currents ($\tau_{syn} \approx 100ms$, Wang (2001)) we find that the long timescales of this network are consistent with its resistance to noise.

Artificial Neuron Dropout

We propose here to investigate to what extent the network can resist the withdrawal of neurons. In biological terms, this can be understood as the effect of a neuron dying in a network. Dropping neurons can also take into account the imperfect geometry that biological networks have by inducing asymmetry in the neuron arrangement. We found that even a very slight dropout rate engenders a progressive and irreversible deterioration of the steady state (Figure 8). Other steady states can be reached but the activity pattern is highly disturbed.

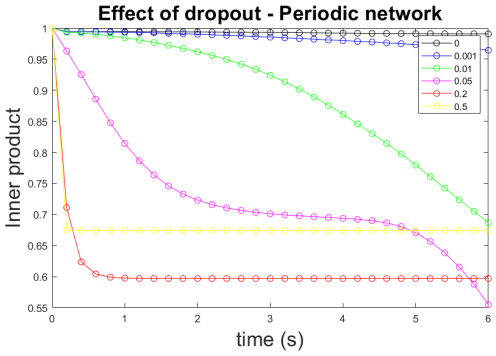


Figure 7: Similarity between S_{exp} (initial conditions) and network state with time for different dropout rates. The similarity metrics used is the normalized inner product. FR Network with $N = 64^2$

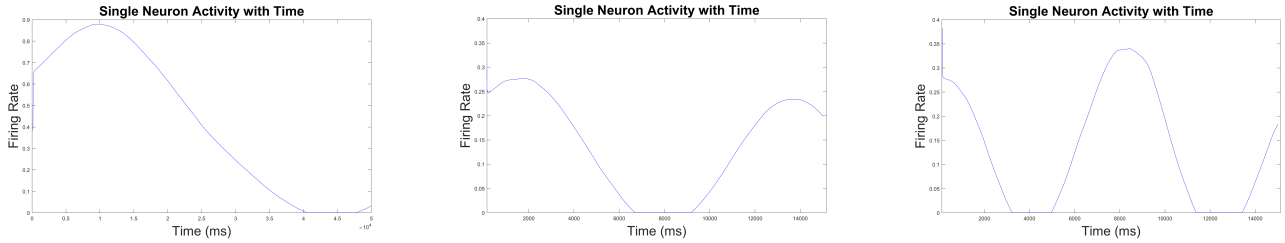
Perturbations in Initial State

In (Brody et al., 2003) the attractor states of a network are understood in terms of wells inside a *Lyapunov*(L) function. A way of characterising the shape of this L-function (also called *energy* (E) term) is by observing the effect of perturbations in different directions of the state space. We have tested the stability of the surroundings of the (experimental) steady state.

- Homotheties (dilations): This operation is performed by projecting the pattern matrix to a matrix $kN \times kN$(k is the ratio of the dilation, kN rounded to integer values). Later, only the center of dimensions $N \times N$ is conserved. The network activity recovers the original state for a ratio ≤ 1.5.
- Translations: Translations in the position plane remain stable. This confirms the hypothesis that translations lie inside a manifold of degenerate states that equally minimise E.
- Rotations: For the periodic network, given an initial stable state, only rotations of modulo $\pi/2$ remain stable. In this scenario, given two stable states S_0 and S_{90} such that one is a 90 degree rotation of the other, one can generate 2 attractor manifolds separated by an energy wall. These can be obtained by applying translations in S_0 and S_{90} (respectively). In the aperiodic network, on the contrary, there exist intermediate stable states (there are more manifolds with a smaller energy barrier in between).

Tuning of Directions Preference

The model predicts that tuning direction preference at the Single Neuron (SN) level should not have an effect on stable patterns geometry. We have tested this by reducing the preference vectors for East and West by a factor k. The effect observed is that the final pattern remains equal, while SN response to velocities in the x direction is dilated (reduces its frequency, figure 8).



(a) SN Activity, $k = 0.25$ (b) SN Activity, $k = 0.5$ (c) SN Activity, $k = 0.75$

Figure 8: SN activity for different direction preference tunings

Strengths and Weaknesses as a Biological Model

We have reviewed so far some of the limitations of the model regarding its biological plausibility. In addition to this, we have found of interest outlining some further considerations of the model in biological terms. A

key component of our model is that it relies on a perfect velocity vector being coupled to the network. While many cells show tuning to head direction alone, that is not the case of movement direction (Raudies et al., 2015) which our model relies on.
Another consideration is that this model is constructed using velocity as unique input, while this assumption is likely to be false for any neural network in the brain. An isolated model cannot account for differences in SN activity from one location to the other when the animal is lying in the same positional class (Dunn et al., 2017; Stensola et al., 2012). Efforts in extending the connectivity of the network to other regions is discussed in the following section and could help explain some of these evidence.
Attractor models have been traditionally criticised for not accounting for theta oscillations and precession, together with the absence of rebound current observed in stellate cells (Hasselmo, 2014). It should be noted, however, that the role of these elements in the model remains unclear, with some results suggesting that could even be processes independent of grid activity (Yartsev et al., 2012).

The Evolution of the Grid Cells Model

This last section has the intention of presenting the state of the art regarding grid cells modelling.

Agmon and Burak (2020) propose recurrent connectivity between the different grid networks and place cells. Differences in activity in a single neuron when the rat is found in two equivalent positions of the spatial grid (Dunn et al., 2017) can be thus explained by these receiving inputs from place cells. In addition, in this recurrent model the possibility of inference *jumps* in position that can occur due to noise in grid networks is damped in favor of a continuous flow in place network activity. The model presented in this article also includes some differences in the dynamics of grid networks, as can be the use of a sub-linear transfer function ($\phi(x) = [\sqrt{x}]_+$) and including an external current to account for depolarization and hyperpolarization in the cells.

Persistent Cohomology (De Silva et al., 2011) is a Topological Data Analysis method that assigns a lifetime (or *relevance*) to the number of holes of different dimensions found in a cloud of data points. In the recently published (*in press*) (Gardner et al., 2021) this technique was used on recorded grid cells activity. Results indicated, as expected and assumed in the attractor grid cell model, that the grid cells activity present toroidal topology.

Conclusions

We have successfully reproduced grid-cell-like activity with an attractor model in its different variations (periodic / aperiodic boundareis, FR / SP model). Furthermore, experiments testing its robustness and biological plausibility have shown that, while not invalidated, this model still needs to be extended to account for experimental observations. In this last sense, future research should benefit both from computational advances and integration of new experimental findings in order to reach a model capable of reproducing and explaining the biology of grid cells.

References

Agmon, H., & Burak, Y. (2020). A theory of joint attractor dynamics in the hippocampus and the entorhinal cortex accounts for artificial remapping and grid cell field-to-field variability. *Elife*, *9*, e56894.

Albesa, A., Awasthi, J., Boulaabi, M., Chawla, D., & Paques, M. (2021). *Grid Cells Repository.* `https://github.com/albertalbesa/grid-cells`. GitHub.

Brody, C. D., Romo, R., & Kepecs, A. (2003). Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. *Current opinion in neurobiology*, *13*(2), 204–211.

Burak, Y., & Fiete, I. R. (2009). Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput Biol*, *5*(2), e1000291.

De Silva, V., Morozov, D., & Vejdemo-Johansson, M. (2011). Persistent cohomology and circular coordinates. *Discrete & Computational Geometry*, *45*(4), 737–759.

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

Fiete, I. R., Burak, Y., & Brookings, T. (2008). What grid cells convey about rat location. *Journal of Neuroscience*, *28*(27), 6858–6871.

Fyhn, M., Hafting, T., Witter, M. P., Moser, E. I., & Moser, M.-B. (2008). Grid cells in mice. *Hippocampus*, *18*(12), 1230–1238.

Gardner, R. J., Hermansen, E., Pachitariu, M., Burak, Y., Baas, N. A., Dunn, B. A., ... Moser, E. I. (2021). Toroidal topology of population activity in grid cells. *bioRxiv*. doi: 10.1101/2021.02.25.432776

Gil, M., Ancau, M., Schlesiger, M., Neitz, A., Allen, K., De Marco, R., & Monyer, H. (2018). Impaired path integration in mice with disrupted grid cell firing. *Nature Neuroscience*, *21*(1), 81–91.

Giocomo, L. M., Moser, M.-B., & Moser, E. I. (2011). Computational models of grid cells. *Neuron*, *71*(4), 589–603.

Hafting, T. (2014). *Grid cell data hafting et al 2005 https://doi.org/10.11582/2014.00001.* Centre for the Biology of Memory.

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

Hasselmo, M. E. (2014). Neuronal rebound spiking, resonance frequency and theta cycle skipping may contribute to grid cell firing in medial entorhinal cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1635), 20120523.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, *148*(3), 574–591.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160*(1), 106–154.

Knierim, J. J., Neunuebel, J. P., & Deshmukh, S. S. (2014). Functional correlates of the lateral and medial entorhinal cortex: objects, path integration and local–global reference frames. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1635), 20130369.

Kropff, E., & Treves, A. (2008). The emergence of grid cells: Intelligent design or just adaptation? *Hippocampus*, *18*(12), 1256–1269.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M.-B. (2006). Path integration and the neural basis of the'cognitive map'. *Nature Reviews Neuroscience*, *7*(8), 663–678.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in neurosciences*, *6*, 414–417.

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

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

Pastoll, H., Solanka, L., van Rossum, M. C., & Nolan, M. F. (2013). Feedback inhibition enables theta-nested gamma oscillations and grid firing fields. *Neuron*, *77*(1), 141–154.

Raudies, F., Brandon, M. P., Chapman, G. W., & Hasselmo, M. E. (2015). Head direction is coded more strongly than movement direction in a population of entorhinal neurons. *Brain research*, *1621*, 355–367.

Ray, S., Naumann, R., Burgalossi, A., Tang, Q., Schmidt, H., & Brecht, M. (2014). Grid-layout and theta-modulation of layer 2 pyramidal neurons in medial entorhinal cortex. *Science*, *343*(6173), 891–896.

Rowland, D. C., Obenhaus, H. A., Skytøen, E. R., Zhang, Q., Kentros, C. G., Moser, E. I., & Moser, M.-B. (2018). Functional properties of stellate cells in medial entorhinal cortex layer ii. *Elife*, *7*, e36664.

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

Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M.-B., & Moser, E. I. (2012). The entorhinal grid map is discretized. *Nature*, *492*(7427), 72–78.

Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in neurosciences*, *24*(8), 455–463.

Witter, M. P., Groenewegen, H., Da Silva, F. L., & Lohman, A. (1989). Functional organization of the extrinsic and intrinsic circuitry of the parahippocampal region. *Progress in neurobiology*, *33*(3), 161–253.

Witter, M. P., et al. (1993). Organization of the entorhinal-hippocampal system: a review of current anatomical data. *HIPPOCAMPUS-NEW YORK-CHURCHILL LIVINGSTONE-*, *3*, 33–33.

Yartsev, M. M., Witter, M. P., & Ulanovsky, N. (2012). Yartsev et al. reply. *Nature*, *488*(7409), E2–E2.