



# Noise-driven bifurcations in a neural field system modelling networks of grid cells

José A. Carrillo<sup>1</sup> · Helge Holden<sup>2</sup> · Susanne Solem<sup>3</sup>

Received: 16 September 2021 / Revised: 1 February 2022 / Accepted: 4 May 2022  
© The Author(s) 2022

## Abstract

The activity generated by an ensemble of neurons is affected by various noise sources. It is a well-recognised challenge to understand the effects of noise on the stability of such networks. We demonstrate that the patterns of activity generated by networks of grid cells emerge from the instability of homogeneous activity for small levels of noise. This is carried out by analysing the robustness of network activity patterns with respect to noise in an upscaled noisy grid cell model in the form of a system of partial differential equations. Inhomogeneous network patterns are numerically understood as branches bifurcating from unstable homogeneous states for small noise levels. We show that there is a phase transition occurring as the level of noise decreases. Our numerical study also indicates the presence of hysteresis phenomena close to the precise critical noise value.

**Keywords** Grid cells · Noise-driven bifurcations · Neural field models · Partial differential equations

**Mathematics Subject Classification** Primary 92B20; Secondary 92C20 · 35Q92

---

✉ José A. Carrillo  
carrillo@maths.ox.ac.uk

Helge Holden  
helge.holden@ntnu.no  
<https://www.ntnu.edu/employees/holden>

Susanne Solem  
susanne.solem@nmbu.no

- <sup>1</sup> Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK
- <sup>2</sup> Department of Mathematical Sciences, NTNU Norwegian University of Science and Technology, NO-7491 Trondheim, Norway
- <sup>3</sup> Department of Mathematics, Norwegian University of Life Sciences, 1433 Ås, Norway

## 1 Introduction

By now it is well established that grid cells, and the characteristic hexagonal firing patterns they create in physical space, play an important role in the navigational system of mammalian brains (McNaughton et al. 2017). Since grid cells were discovered in Hafting et al. (2005), there has been extensive activity in order to understand their precise behaviour, see (Rowland et al. 2016; McNaughton et al. 2017) and the references therein. The main challenges ten years after the discovery of grid cells, such as how grid cells are organised and how they are connected to other cell types in the brain, were highlighted in Rowland et al. (2016). In particular, as the brain is inherently noisy (Rolls and Deco 2010), the lack of understanding of the effect of noise on grid cells was emphasised as a challenge. The recent results in Gardner et al. (2022) has provided insight into the organisation of grid cells by showing that the activity of the network (called a module in Gardner et al. 2022) is arranged on a torus. The question regarding the effects of noise on grid cells, however, remains open.

In accordance with previous experimental studies and general belief in the field, the results in Gardner et al. (2022) provided further evidence in favour of describing the grid cell network by continuous attractor network dynamics through a system of neural field models (Ermentrout and Terman 2010). The first attractor network models for grid cells were presented in McNaughton et al. (2006), Burak and Fiete (2009), Couey et al. (2013), which were based on the classical papers Wilson and Cowan (1972, 1973) and Amari (1977), see also Pinto et al. (1996). In Burak and Fiete (2009), Couey et al. (2013), the grid cells are assumed to have orientation preferences in four different directions. The hexagonal grid cell patterns are then generated by a system of  $4N^2$  neural field ordinary differential equations

$$\tau_i^\beta \frac{ds_i^\beta}{dt} + s_i^\beta = \Phi \left( \sum_{\beta'} \sum_j W_{ij}^{\beta'} s_j^{\beta'} + B_i^\beta(t) \right), \quad (1.1)$$

with  $\beta = 1, \dots, 4$ . Here  $s_i^\beta \geq 0$  represents the activity level of neuron  $i$  with orientation preference  $\beta$ , and  $\tau_i^\beta$  is its relaxation time. The right-hand side of (1.1) represents the firing rate of the neuron, see Bressloff (2012). The function  $\Phi$  is a given activation function, often of the form of a ReLU or sigmoid function. The firing rate of neuron  $i$  depends on an external input  $B_i^\beta(t)$  and the response of the network. It is assumed that the neurons are arranged on a square, which we will denote  $\Omega$  and call the neural sheet, according to the strength of their pairwise connection. The position of neuron  $i$  is denoted by  $\mathbf{x}_i$ . The strength of the connectivity between neuron  $i$  of type  $\beta$  and  $j$  of type  $\beta'$  is  $W_{ij}^{\beta'} = W(\mathbf{x}_i - \mathbf{x}_j - \mathbf{r}^{\beta'})$ , where  $W(\mathbf{x})$ ,  $\mathbf{x} \in \Omega$ , is assumed to be even in each coordinate, in  $\Omega$ . The connectivity is shifted in the direction of the orientation preference of neuron  $j$  of type  $\beta'$  with  $\mathbf{r}^{\beta'}$  which is given by shifts of equal length in the four cardinal directions: north, south, east and west. It has been commonly considered, and, as mentioned, recently shown in Gardner et al. (2022), that the network of neurons creates a torus connectivity. This is realised in the model by assuming that  $W$  is extended periodically outside  $\Omega$ .

By modelling the movement of a rat traversing physical space through the input  $B_i^\beta(t)$ , (1.1) can recreate the hexagonal patterns in physical space produced by the firing of a single grid cell as observed in experiments (Couey et al. 2013). In this model, the patterns in physical space are a consequence of the patterns generated on the neural sheet  $\Omega$ . However, (1.1) being a deterministic model, it does not offer much insight into the effects of noise.

Works on understanding noisy neural fields have in general been lacking (Bressloff 2012, Sect. 6) until recently (Touboul 2012; Kilpatrick and Ermentrout 2013; Kilpatrick 2014; MacLaurin and Bressloff 2020; Touboul et al. 2012; Bressloff 2019; Byrne et al. 2019). In Burak and Fiete (2012) fundamental limits on how information dissipates in networks of noisy neurons were derived. The author in Kilpatrick (2014) presents a study of two coupled noisy neural field models with a focus on the consequences of the coupling on the neural activity waves, while Kilpatrick and Ermentrout (2013) and MacLaurin and Bressloff (2020) investigate the effect of noise on stationary bumps in one-dimensional spatially extended networks.

Taking a different perspective than Burak and Fiete (2012); Kilpatrick and Ermentrout (2013); Kilpatrick (2014) and MacLaurin and Bressloff (2020), by adding noise to the common model of a grid cell network (1.1), the main goal of this work is to analyse the robustness of the hexagonal patterns in the activity level (Ermentrout and Cowan 1979) observed in (1.1) with respect to noise strength. We show that the stationary spatial patterns of the activity level emerge from the instability of homogeneous brain activity as the noise level diminishes. By upscaling the model (1.1) with noise to a system of Fokker–Planck-like partial differential equations, our analysis gives an estimate on the noise strength above which there is no coherent activity pattern. We also numerically explore the different branches of inhomogeneous stationary patterns bifurcating from the homogeneous state depending on the noise for several activation functions  $\Phi$  indicating the presence of hysteresis phenomena.

Instabilities of homogeneous steady states of noisy neural fields were also investigated by Byrne et al. (2019) utilising a partial differential equation (PDE) description. However, the PDEs were of a very different form than the ones presented in this manuscript. A Fokker–Planck-like system describing a network of noisy neurons can be found in Bressloff (2019), where neural variability in a coupled ring network was studied.

It is classical to analyse the behavioural change of neural fields without noise in the form of ordinary differential equations (ODEs) by standard bifurcation analysis (Murray 2002; Bressloff 2012; Veltz et al. 2015; Kilpatrick and Poll 2017; Schmidt and Avitabile 2020). Finding noise-driven bifurcations is more challenging, and one has to rely on other technical tools unless the coupling of the network has a particular structure where closed ODEs for the mean and the variance are available (Touboul et al. 2012; Touboul 2012). The system (1.1) with noise has, in addition, a nonlinear coupling, and the activity levels must remain nonnegative due to their physical interpretation, leading to technical additional constraints on the stochastic processes involved. In the following we deal with these challenges by analysing a system of Fokker–Planck-like partial differential equations with boundary conditions describing the space-time evolution of the law of the stochastic processes with respect to the noise level.

## 2 The PDE system: derivation, main goal, and numerical experiment

For the sake of the reader, we start by discussing the simplest classical case of no spatial connectivity, see Hopfield (1984); Bressloff (2012) and the references therein. Let us consider the classical neural field stochastic dynamical system for a network of  $M$  coupled neurons given by

$$\tau ds_k + s_k dt = \Phi\left(\frac{W_0}{M} \sum_{k'} s_{k'} + B(t)\right) dt + \sqrt{2\sigma} d\mathcal{W}_k. \quad (2.1)$$

Here, the neurons are considered indistinguishable and all-to-all coupled with equal strengths given by  $W_0 \in \mathbb{R}$  whose sign depends on the type of neurons considered: inhibitory or excitatory. We also consider that the relaxation time for all neurons is the same and equal to  $\tau$ .  $B(t)$  is the external input for this neural network and  $\sigma > 0$  is the strength of the noise  $\mathcal{W}_k$ . We have considered independent Brownian motion for each neuron in the network. Classical stochastic analysis implies that we can derive a Fokker–Planck equation for the evolution of the probability density of neurons with activity level  $s$  at time  $t$  in the large population limit  $M \rightarrow \infty$ , i.e., the law of the limiting stochastic process follows the PDE

$$\tau \frac{\partial f}{\partial t} = \frac{\partial}{\partial s} ([s - \Phi(W_0 \langle f \rangle + B(t))] f) + \sigma \frac{\partial^2 f}{\partial s^2}, \quad (2.2)$$

where  $f = f(t, s)$  denotes the probability to observe the activity  $s$  at time  $t$ , and  $\langle f \rangle$  denotes the mean value of the activity level  $s$

$$\langle f \rangle = \int_0^\infty s f(s) ds.$$

Notice that the noise can drive the activity level to be negative in (2.1), which is clearly not desirable from the modelling viewpoint. In order to avoid this, it is common practise to consider the Fokker–Planck equation (2.2) on  $s \in [0, \infty)$  with no-flux boundary conditions

$$\left(\Phi(W_0 \langle f \rangle + B(t)) f - \sigma \frac{\partial}{\partial s} f\right) \Big|_{s=0} = 0. \quad (2.3)$$

This ensures that particles cannot escape from non-negative values of the activity level variable  $s$  at the PDE level while keeping an evolution of a probability density, see Carrillo et al. (2011), Carrillo et al. (2013) for instance.

**Remark 2.1** (Microscopic Model) Reflective boundary conditions for stochastic processes have been incorporated at the stochastic differential equation (SDE) level in order to avoid particles to escape a fixed domain (Sznitman 1984; Lions and Sznitman 1984; Faugeras and Inglis 2015). One can produce a microscopic stochastic process

by adding an additional process counting when particles touch the boundary of the domain. The law of the rigorous mean-field limit,  $M \rightarrow \infty$ , of the following system

$$\tau ds_k + s_k dt = \Phi \left( \frac{W_0}{M} \sum_{k'} s_{k'} + B(t) \right) dt + \sqrt{2\sigma} d\mathcal{W}_k - dl_k, \tag{2.4}$$

$$l_k(t) = -|l_k|(t), \quad |l_k|(t) = \int_0^t 1_{\{s_k(\zeta)=0\}} d|l_k|(\zeta), \tag{2.5}$$

$k = 1, \dots, M$ , follows the evolution of (2.2)–(2.3) under suitable smoothness assumptions on  $\Phi$ , see Lions and Sznitman (1984), Sznitman (1991).

The next step in the modelling is to reinterpret  $M$  as the number of neurons in each of the cortical columns of a neural sheet of  $N$  columns. Given space points  $x_1, \dots, x_N$  in the region  $\Omega$  of the neural cortex, the interaction among  $NM$  neurons stacked in  $N$  columns at locations  $x_i$  with  $M$  neurons each, where  $s_{ik}^\beta$  represents the activity level with orientation  $\beta$  of the  $k^{th}$  neuron at location  $x_i$  is given for  $i = 1, \dots, N$  and  $k = 1, \dots, M$  by

$$\begin{aligned} \tau ds_{ik}^\beta + s_{ik}^\beta dt &= \Phi \left( \frac{1}{4NM} \sum_{\beta'=1}^4 \sum_{j=1}^N \sum_{m=1}^M W^{\beta'}(x_i - x_j) s_{jm}^{\beta'} + B^\beta(t) \right) dt \\ &+ \sqrt{2\sigma} d\mathcal{W}_{ik}^\beta - d\ell_{ik}^\beta, \end{aligned} \tag{2.6a}$$

$$\ell_{ik}^\beta(t) = -|\ell_{ik}^\beta|(t), \quad |\ell_{ik}^\beta|(t) = \int_0^t 1_{\{s_{ik}^\beta(r)=0\}} d|\ell_{ik}^\beta|(r) \quad \text{for } \beta = 1, 2, 3, 4. \tag{2.6b}$$

Here, we consider the same periodic setting, imposed through the periodicity of the interactions  $W^\beta$  for  $\beta = 1, \dots, 4$ , as in Burak and Fiete (2009), Couey et al. (2013). Moreover, the neurons are inhibitory (Couey et al. 2013) and the activity in the network is modulated by a time dependent external input as in (1.1). We are dealing with a population network of neurons structured by their orientation preference  $\beta = 1, \dots, 4$  corresponding to the four cardinal points (north, west, south, east). The network population includes the localised in space cross inhibition of neurons with different orientations modulated by the shape function  $W^\beta$ , where  $W^\beta(\mathbf{x}) = W(\mathbf{x} - \mathbf{r}^\beta)$ . Following the approach outlined above in the case of one population, we can formally write a Fokker–Planck type equation, in the limit  $N, M \rightarrow \infty$ , for the evolution of the probability density  $f^\beta(t, \mathbf{x}, s)$  of finding neurons of type  $\beta$  at position  $\mathbf{x}$  on the neural sheet  $\Omega$  with activity level  $s \geq 0$  at time  $t \geq 0$ . We refer to Cai et al. (2006) for a similar approach in conductance-voltage models. The system of equations reads

$$\tau \frac{\partial f^\beta}{\partial t} = -\frac{\partial}{\partial s} \left( \left[ \Phi^\beta(\mathbf{x}) - s \right] f^\beta \right) + \sigma \frac{\partial^2 f^\beta}{\partial s^2}, \tag{2.7}$$

where  $\Phi^\beta(\mathbf{x})$  is given by

$$\Phi \left( \frac{1}{4} \sum_{\beta'} \int_{\Omega} W^{\beta'}(\mathbf{x} - \mathbf{y}) \langle f^{\beta'} \rangle(t, \mathbf{y}) d\mathbf{y} + B^\beta(t) \right),$$

with

$$\langle f^\beta \rangle(t, \mathbf{x}) = \int_0^\infty s f^\beta(t, \mathbf{x}, s) ds, \quad \beta = 1, \dots, 4,$$

periodic boundary conditions in  $\mathbf{x}$ , and the no-flux boundary conditions at  $s = 0$  given by

$$\left( \Phi^\beta(\mathbf{x}) f^\beta - \sigma \frac{\partial}{\partial s} f^\beta \right) \Big|_{s=0} = 0, \quad \beta = 1, \dots, 4, \tag{2.8}$$

for each position  $\mathbf{x}$  in the square sheet  $\Omega$ . To realise the torus connectivity, we assume that  $W$  is periodic with respect to  $\Omega$  and even in each coordinate on  $\Omega$ . The function  $\Phi$  is typically a smooth approximation of the ReLU activation function  $(x)^+ = \max\{0, x\}$  or a sigmoid function. The initial probability density of the system (2.7) is denoted by  $f_0^\beta$ .

**Remark 2.2** The system of Fokker–Planck equations (2.7)–(2.8) can be rigorously derived from the microscopic stochastic processes (2.6) under suitable assumptions. The rigorous proof of this mean-field limit for the spatially extended system (2.7)–(2.8) has recently been obtained in Carrillo et al. (2021) by a generalisation of the coupling method of Sznitman (1991). This rigorous passage to the limit is a very interesting area of mathematical research on its own with a multitude of different models and limiting systems derived under different assumptions on the ingredients of the network. For instance, we refer to the works Moynot and Samuelides (2002), Faugeras et al. (2009), Faugeras and Inglis (2015), Touboul (2012), Touboul et al. (2012), and Cabana and Touboul (2018) in which the authors deal with spatially extended systems of neural networks modelled by their voltage with random connectivity interactions using large deviation principles (Arous and Guionnet 1995; Guionnet 1997).

To summarise, the main goal of this work is to focus on the biological information carried by the system of PDEs (2.7)–(2.8). More precisely, we study how noise affects the dynamics of (2.7)–(2.8) under the following assumptions:

- (A1) The grid cells are arranged on a torus, realised by setting the neural sheet  $\Omega = [-0.5, 0.5]^2$  and extending  $W$  periodically outside  $\Omega$ .
- (A2) The inhibitory (Couey et al. 2013) connectivity function  $W \leq 0$  is at least in  $L^2(\Omega)$ , and is an even function in each coordinate on  $\Omega$ . Furthermore, we define  $W_0 = \int_{\Omega} W(\mathbf{x}) d\mathbf{x}$ . In the numerical experiments  $W$  satisfies  $W(\mathbf{x}) = W(|\mathbf{x}|)$  in addition.
- (A3) The modulation function  $\Phi$  is in  $C^1$  (unless otherwise stated).

- (A4) There are four orientation preferences,  $\beta = 1$  (north), 2 (west), 3 (south), and 4 (east), where the shifts are of equal size  $z$  in each direction, i.e.,  $\mathbf{r}^\beta = z\mathbf{e}_\beta$ , where  $\mathbf{e}_\beta$  is the unit vector in direction  $\beta$ .

It is well-known that grid cell firing is strongly connected to mammals' navigation, but unknown exactly how the grid cell network communicates with other networks in the brain. We will therefore simply assume in the numerical experiments that the external input  $B^\beta(t)$  in (2.7)–(2.8) depends on the velocity at time  $t$ ,  $v(t)$ , of a moving animal in the following manner (Burak and Fiete 2009; Couey et al. 2013):

$$B^\beta(t) = B + \alpha v(t) \cos(\theta(t) - \theta^\beta), \quad (2.9)$$

where  $B > 0$  is a constant external excitatory input, assumed to be the same for different  $\beta$ ,  $\alpha$  the velocity modulation,  $\theta(t)$  the orientation of the animal at time  $t$  according to the reference frame, and  $\theta^\beta$  the orientation preference of the neurons of type  $\beta$  ( $\theta^1 = \frac{\pi}{2}$ ,  $\theta^2 = \pi$ ,  $\theta^3 = \frac{3}{2}\pi$ ,  $\theta^4 = 2\pi$ ). This particular form of the input, together with the right set of parameters in (1.1), has been shown to enable single cells of the ODE system (1.1) to create hexagonal firing patterns in physical space, see Burak and Fiete (2009), Couey et al. (2013).

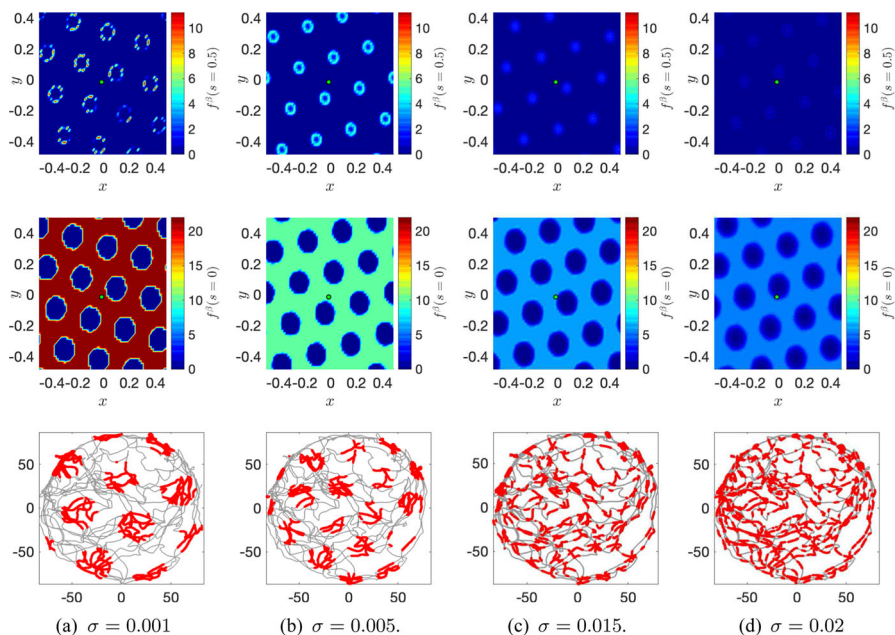
## 2.1 Numerical reproduction of the hexagonal patterns

We numerically demonstrate that the PDE system (2.7)–(2.8) with (2.9) is able to reproduce the characteristic single-cell hexagonal firing pattern as discovered by Hafting et al. (2005) for rats and see how this pattern depends on the noise strength  $\sigma > 0$ . For this, we use a numerical scheme that has been extensively utilised for Fokker–Planck like equations (Carrillo et al. 2015). For more details on the numerical approach and its validation, see Appendix 1. Before connecting the grid cell system (2.7)–(2.8) with the movement of a rat, we initialise the activity on the neuronal sheet  $\Omega$  by running the simulation with  $\alpha = 0$  until  $f^\beta$  has numerically stabilised into stationary patterns equal to the ones in the top and middle rows of Fig. 1, modulo translations.

Then we connect with the rats movement by setting  $\alpha = 0.3$  in (2.9) as in Couey et al. (2013). The velocity  $v(t)$  and orientation  $\theta(t)$  used in the numerical experiment are calculated using timestamped position data from the physical experiments in Hafting et al. (2005) where rats moved around in a circular enclosure with a radius of 80 cm. The shape of the enclosure can be seen in the plots in the bottom row of Fig. 1. The path of the rat after moving around for  $t = 5$  minutes, generated with the position data, is visualised in grey.

The red coloured areas in each plot in the bottom row in Fig. 1 make up the firing field of a grid cell in the network. The firing field consists of smaller red circular areas, which again are made up by even smaller red dots. Each dot marks a firing of the grid cell placed at position  $(0, 0)$  on the neuronal sheet  $\Omega$  as the velocity and orientation data is fed into the numerical method of (2.7)–(2.8) through (2.9). For simplicity, we have assumed in the numerical experiment that a neuron at a particular position  $\mathbf{x} \in \Omega$  fires as soon as the firing rate  $\Phi$  in (2.7)–(2.8) satisfies  $\Phi(\mathbf{x}) > 0$ .





**Fig. 1** Network patterns and single-cell responses after  $t = 5$  min for increasing noise strength  $\sigma$  (left to right  $\sigma = 0.001, 0.005, 0.015, 0.02$ ) with the modulation function used in Couey et al. (2013):  $\Phi(x) = (x)^+$  and (2.9) with  $I = 3$  and  $\alpha = 0.3$ . Top row: the probability density  $f^\beta(x, y, s = 0.5)$  for one  $\beta$ ,  $(x, y) \in \Omega$ , middle row:  $f^\beta(x, y, s = 0)$ , and bottom row: single-cell firing pattern in a circular enclosure traversed by a rat (units in cm) created by the cell at position  $(0, 0)$  on the neural sheet  $\Omega$  (green dot in top and middle plots)

Now, from left to right in the bottom row of Fig. 1, we see the pattern of the firing fields—the patterns created by the red dots over the path of the rat in the enclosure in physical space—for the grid cell at  $(0, 0)$  for increasing noise strength.

The top and middle rows in the figure display snapshots of the probability density  $f^\beta$  at  $s = 0.5$  and  $s = 0$ , respectively. As  $t$  increases, these patterns are translated in accordance with the movement of the rat.

As can be observed in the bottom row, the red fields generated with the PDE system (2.7)–(2.8) form hexagonal patterns similar to the ones observed in physical experiments. However, the distance between the activity bumps and the area they cover decreases as the noise strength increases. The second main observation is that by increasing the noise the firing becomes less and less localised. The numerical experiments support the existence of a critical value of the noise,  $\sigma_c > 0$ , at which a single cell could fire no matter where the rat is on its path.

The question we will address in the following is how stable the patterns observed in Fig. 1 are with respect to the noise strength  $\sigma$ .



### 3 Stability of the neural field system

In this section, we start by studying the spatially homogeneous solutions. We would like to understand the pattern formation in the neural field system (2.7)–(2.8) as a byproduct of the instability of these homogeneous solutions. We assume that solutions to (2.7) are sufficiently smooth and decay fast enough as  $s \rightarrow \infty$ . We further assume from now on that  $B^\beta = B$  is constant, i.e.,  $\alpha = 0$  in (2.9), in order to study the emergence of stationary network patterns of the system. Note that setting  $B^\beta$  to different constant values depending on  $\beta$  could yield non-stationary network patterns: a rat running with constant speed in one direction would give  $v(t) = \text{const} > 0$  and  $\theta(t) = \text{const}$  in (2.9), which results in different constant values of  $B^\beta$ . This would, with the right set of parameter values, consequently translate the network patterns in time. To avoid this technicality, we let  $B$  be identical for the four different direction preferences.

Homogeneous solutions  $f^\beta(t, \mathbf{x}, s) = f(t, s)$  for  $\beta = 1, \dots, 4$  to (2.7) satisfy

$$\tau \frac{\partial f}{\partial t} = \frac{\partial}{\partial s} ([s - \Phi(W_0\langle f \rangle + B)] f) + \sigma \frac{\partial^2 f}{\partial s^2}, \tag{3.1}$$

with no-flux boundary conditions (2.8) at  $s = 0$ , i.e.,

$$\left( \Phi^\beta(W_0\langle f \rangle + B) f^\beta - \sigma \frac{\partial f^\beta}{\partial s} \right) \Big|_{s=0} = 0, \quad \beta = 1, \dots, 4, \tag{3.2}$$

and

$$W_0 = \int_{\Omega} W(\mathbf{x}) d\mathbf{x}.$$

In order to find stationary spatially homogeneous states  $f_\infty$  we first assume that their mean  $\langle f_\infty \rangle$  is given. Denote by  $\Phi_0 = \Phi(W_0\langle f_\infty \rangle + B)$  the corresponding firing rate for simplicity. Thus, by integrating (3.1) and using the boundary condition (3.2), the stationary spatially homogeneous states  $f_\infty(s)$  are given by

$$f_\infty(s) = \frac{1}{Z} \exp\left(-\frac{(s - \Phi_0)^2}{2\sigma}\right), \tag{3.3}$$

with the mass normalisation factor  $Z$  such that  $\int_0^\infty f_\infty(s) ds = 1$ , i.e.,

$$Z = \sqrt{\frac{\pi\sigma}{2}} \left( 1 + \operatorname{erf}\left(\frac{\Phi_0}{\sqrt{2\sigma}}\right) \right), \tag{3.4}$$

where the error function has been defined as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-y^2) dy.$$

However, note that (3.3) is an implicit equation as  $\Phi_0$  depends on the mean  $\langle f_\infty \rangle$ . To show the existence of stationary solutions, we need to solve the consistency equation for the mean  $\langle f_\infty \rangle$  given by

$$\langle f_\infty \rangle = \Phi_0 + \sigma \frac{1}{Z} \exp\left(-\frac{\Phi_0^2}{2\sigma}\right). \quad (3.5)$$

We prove next that the stationary state exists and is unique by leveraging on (3.5) under suitable conditions on the firing rate function  $\Phi$ .

**Proposition 3.1** *Let  $\sigma > 0$ ,  $W_0 \leq 0$ , and  $0 \leq \Phi(x) \leq \Phi(B)$  for any  $x \leq B$ . Assume (A3), and that  $\Phi$  satisfies  $\Phi'(W_0 m + B) > \frac{1}{W_0}$  for all  $m \geq 0$ , then (3.1) has a unique stationary solution  $f_\infty$  defined by (3.3)–(3.5).*

**Proof** Define for  $m \geq 0$  and  $\sigma > 0$  the function

$$G(m, \sigma) = \Phi(W_0 m + B) + \sigma \frac{1}{Z} \exp\left(-\frac{\Phi^2(W_0 m + B)}{2\sigma}\right) - m.$$

First, notice that  $G(m, \sigma)$  satisfies  $G(0, \sigma) > 0$  and

$$G(m, \sigma) \leq \Phi(W_0 m + B) + \sqrt{\frac{\sigma}{2\pi}} - m \leq \Phi(B) + \sqrt{\frac{\sigma}{2\pi}} - m < 0$$

for  $m > \Phi(B) + \sqrt{\frac{\sigma}{2\pi}}$ . We now compute

$$\frac{\partial G}{\partial m}(m, \sigma) = -1 + \Phi'(W_0 m + B) W_0 g\left(\frac{\Phi(W_0 m + B)}{\sqrt{2\sigma}}\right)$$

with

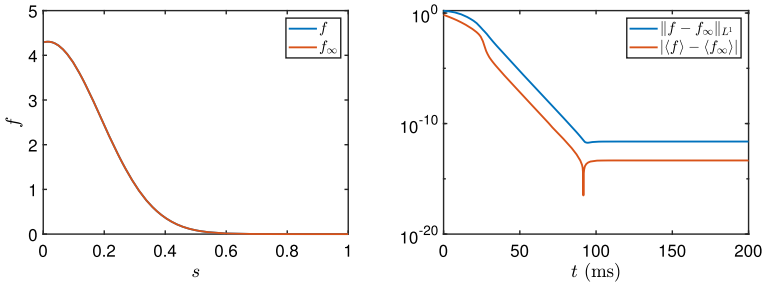
$$g(\eta) = \left(1 - \frac{2}{\sqrt{\pi}} \frac{\exp(-\eta^2)}{1 + \operatorname{erf}(\eta)} \left[\frac{1}{\sqrt{\pi}} \frac{\exp(-\eta^2)}{1 + \operatorname{erf}(\eta)} + \eta\right]\right).$$

It is not difficult to check that the supremum of  $g(\eta)$  over  $\eta \in [0, \infty)$  is given by

$$\alpha = \sup_{\eta \geq 0} \left(1 - \frac{2}{\sqrt{\pi}} \frac{\exp(-\eta^2)}{1 + \operatorname{erf}(\eta)} \left[\frac{1}{\sqrt{\pi}} \frac{\exp(-\eta^2)}{1 + \operatorname{erf}(\eta)} + \eta\right]\right) = 1.$$

Our assumptions on  $\Phi$  and  $W_0 \leq 0$  imply that  $\frac{\partial G}{\partial m}(m, \sigma) < 0$ , and then we obtain the desired unique zero of  $G$  defining our stationary state  $f_\infty$  through (3.3)–(3.5).  $\square$

**Remark 3.2** Notice that the previous proposition can also be applied for functions  $\Phi$  admitting negative values as  $\frac{\exp(-\eta^2)}{1 + \operatorname{erf}(\eta)}$  behaves like  $-\sqrt{\pi}\eta$  in the limit  $\eta \rightarrow -\infty$ , such



**Fig. 2** Stabilisation in time of the homogeneous problem. Left: numerical steady state  $f$  (blue) versus the fixed point steady state  $f_\infty$  (red) obtained from (3.3)–(3.5) as functions of  $s$ . Right: the  $L^1$ -error and difference in mean between  $f$  and  $f_\infty$  plotted as functions of time  $t$ . Parameters:  $\Phi_\varepsilon$  with  $\varepsilon = 0.01$ .  $W_0 = -20.6711$ .  $\sigma = 0.03$ . The interval  $[0, 3]$  is split into 512 grid points. Initial data: at random grid points  $s_j$ ,  $j = 1, \dots, 51$ ,  $f_0(s_j) = 512/153$  and zero elsewhere. Average of the slopes over 100 runs: 0.78 (difference in mean) and 0.57 ( $L^1$ -difference)

that one can show

$$0 \leq \left( 1 - \frac{2 \exp(-\eta^2)}{\sqrt{\pi} (1 + \operatorname{erf}(\eta))} \left[ \frac{1 \exp(-\eta^2)}{\sqrt{\pi} (1 + \operatorname{erf}(\eta))} + \eta \right] \right) \leq 1$$

for any  $\eta \in \mathbb{R}$ . For instance, the theorem is valid for the  $\varepsilon$ -approximation of  $\Phi(x) = (x)^+$  defined by

$$\Phi_\varepsilon(x) = 0.5x \left( 1 + \frac{x}{\sqrt{x^2 + \varepsilon}} \right) \tag{3.6}$$

for  $\varepsilon$  small enough such that  $\Phi'_\varepsilon(x) \geq \frac{1}{W_0}$ . The smooth approximation  $\tilde{\Phi}_\varepsilon(x) = 0.5(x + \sqrt{x^2 + \varepsilon})$  of  $\Phi(x) = (x)^+$  can also be used as  $\tilde{\Phi}_\varepsilon(x)$  trivially satisfies the hypotheses of Proposition 3.1 since  $\tilde{\Phi}_\varepsilon(x) > 0$ , and it is strictly increasing.

We have numerically analysed the stability of the stationary solutions obtained in the previous result among spatially homogeneous solutions of (3.1). In Fig. 2a, we illustrate that the computed stationary state and the numerical solution to the evolution problem after time  $t = 150\text{ms}$  are indistinguishable for the firing rate  $\Phi_\varepsilon$  with  $\varepsilon = 0.01$ . In Fig. 2b, we observe the convergence in time towards the stationary state by computing the difference in  $L^1$  and the difference in average between the stationary solution and the evolution problem. We conclude that the stationary state and the corresponding numerical solution to the evolution problem (3.1) are identical to machine precision after  $t = 150\text{ms}$ , and that the convergence in time is exponential with the rates computed by averaging over 100 runs.

Next, we focus on the linear stability of the spatially homogeneous solution as a solution of the nonlinear system (2.7)–(2.8). For comparison, we first find the condition for linear spatial stability in the case of zero noise ( $\sigma = 0$ ) following the classical approach as in Murray (2002). Let  $\hat{W}(\mathbf{k})$  be the two-dimensional Fourier transform

of  $W$  restricted to  $\Omega$ ,

$$\hat{W}(\mathbf{k}) = \int_{\Omega} W(\mathbf{x}) \exp(-i\mathbf{k} \cdot \mathbf{x}) d\mathbf{x},$$

with  $\mathbf{k} = (2\pi k_1 \ 2\pi k_2)^\top$ ,  $k_1, k_2 \in \mathbb{Z}$ . As an example, if  $W(\mathbf{x})$  is the characteristic of a ball with radius  $r$  fully contained in  $\Omega$ , then

$$\hat{W}(\mathbf{k}) = 2\pi \frac{J_1(r|\mathbf{k}|)}{|\mathbf{k}|},$$

where  $J_1(x)$  is a Bessel function. Now let  $\Phi'_0 = \Phi'(W_0(f_\infty) + B)$  with  $f_\infty$  given by (3.3), and define

$$F(\mathbf{k}) = \frac{1}{4} \Phi'_0 \hat{W}(\mathbf{k}) \sum_{\beta} \exp(-i\mathbf{k} \cdot \mathbf{r}^\beta). \tag{3.7}$$

**Remark 3.3** Given the assumptions on  $W$  in (A1) and (A2), the function  $F(\mathbf{k})$  is real-valued when the shifts  $\mathbf{r}^\beta$ ,  $\beta = 1, 2, 3, 4$ , satisfy the assumptions in (A4). With shifts of equal size  $z$ , one can check that

$$\sum_{\beta} \exp(-i\mathbf{k} \cdot \mathbf{r}^\beta) = 2 \cos(2\pi k_1 z) + 2 \cos(2\pi k_2 z)$$

with  $\mathbf{k} = (2\pi k_1 \ 2\pi k_2)^\top$  using Euler’s formula.

The following lemma presents a linear stability condition for the system (2.7) without noise.

**Remark 3.4** The proof of the mean field limit in Carrillo et al. (2021) relies on  $\sigma > 0$ . The mean field limit in the case  $\sigma = 0$  is easier to obtain and leads to a pure Vlasov equation. This is a very classical result in the smooth setting and is derived via estimates in transport distances, see Hauray and Jabin (2015), Cañizo et al. (2011), Jabin (2014), Hauray and Jabin (2007), Golse (2016) and the references therein.

**Lemma 3.5** Assume (A1)–(A4). Let  $F(\mathbf{k})$  be as in (3.7). Then the mean of the zero noise, i.e.,  $\sigma = 0$ , spatially homogeneous, stationary solution of (2.7)–(2.8) is linearly asymptotically stable if  $F(\mathbf{k}) < 1$ .

**Proof** By taking the mean of (2.7)–(2.8) with  $\sigma = 0$ , we find that the mean at position  $\mathbf{x} \in \Omega$ ,  $s^\beta(t, \mathbf{x}) = \langle f^\beta \rangle|_{\sigma=0}(t, \mathbf{x})$ , evolves according to (dropping the  $t$  dependence for ease of notation)

$$\tau \frac{d}{dt} s^\beta(\mathbf{x}) = \Phi \left( \frac{1}{4} \sum_{\beta'} \int_{\Omega} W^{\beta'}(\mathbf{x} - \mathbf{y}) s^{\beta'}(\mathbf{y}) d\mathbf{y} + B \right) - s^\beta(\mathbf{x}). \tag{3.8}$$

We linearise (3.8) around the mean stationary, spatially homogeneous solution  $s_\infty$ , defined through  $s_\infty = \Phi(W_0 s_\infty + B)$ , and get

$$\tau \frac{d}{dt} h^\beta(\mathbf{x}) = \frac{\Phi'_0}{4} \sum_{\beta'} \int_{\Omega} W^{\beta'}(\mathbf{x} - \mathbf{y}) h^{\beta'}(\mathbf{y}) d\mathbf{y} - h^\beta(\mathbf{x}),$$

where  $h^\beta = s^\beta - s_\infty$  and  $\Phi'_0 = \Phi'(W_0 s_\infty + B)$ . Applying the ansatz  $h^\beta(\mathbf{x}, t) \propto \exp(\lambda t + i\mathbf{k} \cdot \mathbf{x})$  to the equation, we find that each mode  $\mathbf{k}$  has the characteristic polynomial

$$p_{\mathbf{k}}(\lambda) = \left( \lambda + \frac{1}{\tau} \right)^3 \left( \frac{\Phi'_0 \hat{W}(\mathbf{k})}{4\tau} \sum_{\beta} \exp(-i\mathbf{k} \cdot \mathbf{r}^\beta) - \frac{1}{\tau} - \lambda \right).$$

We see that three of the eigenvalues are stable as long as  $\tau > 0$ . The fourth eigenvalue is

$$\lambda = \frac{\Phi'_0 \hat{W}(\mathbf{k})}{4\tau} \sum_{\beta} \exp(-i\mathbf{k} \cdot \mathbf{r}^\beta) - \frac{1}{\tau} = \frac{1}{\tau} (F(\mathbf{k}) - 1),$$

which determines whether the linear system is stable or not. The eigenvalue is negative if  $F(\mathbf{k}) < 1$ . □

We now turn to the case with noise,  $\sigma > 0$ . Let  $h^\beta = f^\beta - f_\infty$ , where  $f_\infty$  is defined through (3.3)–(3.5). Notice that for the perturbations to be admissible, we need to ensure that  $\int_0^\infty f^\beta ds = 1$ , and consequently

$$\int_0^\infty h^\beta ds = 0.$$

In principle one needs  $f^\beta \geq 0$  for it to be a probability density. However, we will prove below that linear stability holds without any assumption on the sign of  $f^\beta$ .

After linearising (2.7) around the spatially homogeneous state  $f_\infty$ , we get

$$\tau h_t^\beta = -\partial_s f_\infty \frac{\Phi'_0}{4} \sum_{\beta'} \int_{\Omega} W^{\beta'}(\mathbf{x} - \mathbf{y}) \langle h^{\beta'} \rangle d\mathbf{y} - \partial_s [(\Phi_0 - s)h^\beta] + \sigma \partial_{ss} h^\beta, \tag{3.9}$$

$$\left( f_\infty \frac{\Phi'_0}{4} \sum_{\beta'} \int_{\Omega} W^{\beta'}(\mathbf{x} - \mathbf{y}) \langle h^{\beta'} \rangle d\mathbf{y} + \Phi_0 h^\beta - \sigma \partial_s h^\beta \right) \Big|_{s=0} = 0,$$

with  $\beta = 1, \dots, 4$ . By a change of variables  $h^\beta = f_\infty v^\beta$ , where  $f_\infty$  is the stationary state satisfying

$$\begin{aligned} -\partial_s([\Phi_0 - s]f_\infty) + \sigma \partial_{ss} f_\infty &= 0, \\ (\Phi_0 f_\infty - \sigma \partial_s f_\infty)|_{s=0} &= 0, \end{aligned}$$

we get

$$\tau v_t^\beta = -\frac{(\Phi_0 - s)}{\sigma} \frac{\Phi'_0}{4} \sum_{\beta'} \int_\Omega W^{\beta'}(\mathbf{x} - \mathbf{y}) \langle f_\infty v^{\beta'} \rangle d\mathbf{y} + (\Phi_0 - s)v_s^\beta + \sigma v_{ss}^\beta, \tag{3.10}$$

$$\left( \frac{\Phi'_0}{4} \sum_{\beta'} \int_\Omega W^{\beta'}(\mathbf{x} - \mathbf{y}) \langle f_\infty v^{\beta'} \rangle d\mathbf{y} - \sigma v_s^\beta \right) \Big|_{s=0} = 0,$$

with  $\beta = 1, \dots, 4$ . We now restrict the set of perturbations in  $L^2(\Omega \times [0, \infty))$  to the ones of the form

$$h^\beta(\mathbf{x}, s, t) = f_\infty(s) \sum_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{x}) u_{\mathbf{k}}^\beta(s, t), \tag{3.11}$$

where  $u_{\mathbf{k}}^\beta(s, t)$  is sufficiently smooth. We can then reduce (3.10) to one Fourier mode. Dropping the subscript  $\mathbf{k}$  we set  $v^\beta(\mathbf{x}, s, t) = \exp(i\mathbf{k} \cdot \mathbf{x}) u^\beta(s, t)$ , where  $u^\beta$  may be complex-valued, and  $U = \sum_{\beta} \exp(-i\mathbf{k} \cdot \mathbf{r}^\beta) u^\beta$ , such that (3.10) turns into

$$\begin{aligned} \tau u_t^\beta &= -\frac{(\Phi_0 - s)}{\sigma} \frac{\Phi'_0}{4} \hat{W}(\mathbf{k}) \langle f_\infty U \rangle + (\Phi_0 - s)u_s^\beta + \sigma u_{ss}^\beta, \\ \left( \frac{\Phi'_0}{4} \hat{W}(\mathbf{k}) \langle f_\infty U \rangle - \sigma u_s^\beta \right) \Big|_{s=0} &= 0, \end{aligned} \tag{3.12}$$

for  $\beta = 1, \dots, 4$ . An equation for the time evolution of  $U$  can also be derived, namely

$$\tau U_t = -\frac{(\Phi_0 - s)}{\sigma} F(\mathbf{k}) \langle f_\infty U \rangle + (\Phi_0 - s)U_s + \sigma U_{ss}, \tag{3.13}$$

where  $F(\mathbf{k})$  is defined in (3.7). Denote

$$M_\infty = M_\infty(\sigma) = \int_0^\infty (s - \langle f_\infty \rangle)^2 f_\infty ds, \tag{3.14}$$

where the dependence on  $\sigma$  enters through  $f_\infty$  defined by (3.3)–(3.5). We remark that we cannot obtain closed ODE equations for the moments of the distribution in the activity variable  $s$ , and thus a similar analysis as in Touboul (2012); Touboul et al. (2012) is not possible here. Building on (3.12), we can prove the following result.

**Theorem 3.6** *Assume (A1)–(A4). Let  $\Phi$  satisfy the assumptions in Proposition 3.1 and let  $F(\mathbf{k})$  be as in (3.7) and satisfy the condition in Remark 3.3. Then the spatially homogeneous steady solution  $f_\infty$  to (2.7)–(2.8) is linearly asymptotically stable in  $L^2(\Omega \times [0, \infty))$  for admissible perturbations of the form (3.11) as long as*

$$F(\mathbf{k}) < \frac{\sigma}{M_\infty} \quad \text{for all } \mathbf{k}. \tag{3.15}$$

**Remark 3.7** Using the relations (3.3)–(3.5), one can check after some tedious computations that  $M_\infty(\sigma)$  satisfies  $M_\infty(0) = 0$  and  $M'_\infty(0) = 1$  assuming  $\Phi_0 > 0$ . Thus,  $\frac{\sigma}{M_\infty} \rightarrow 1$  as  $\sigma \rightarrow 0$ , yielding the condition in Lemma 3.5 in the zero noise limit.

**Proof (Proof of Theorem 3.6)** The proof is split into three parts. First, we obtain an upper bound for the time derivative of  $\int_0^\infty f_\infty |U|^2 ds$  (Part I). Recall that  $U = \sum_\beta \exp(-i\mathbf{k} \cdot \mathbf{r}^\beta) u^\beta$ . To obtain a time decaying estimate from the bound, we need to separate the linear part of  $U$  from the nonlinear (Part II). Finally, we establish the stability of  $U$ , and consequently  $u^\beta$ , which then yields the asymptotic stability of  $f_\infty$  (Part III).

**Part I:** Note that

$$\int_0^\infty f_\infty U ds = \sum_\beta \exp(-i\mathbf{k} \cdot \mathbf{r}^\beta) \int_0^\infty h^\beta ds = 0, \tag{3.16}$$

such that

$$\int_0^\infty \partial_s f_\infty U ds = \frac{1}{\sigma} \int_0^\infty (\Phi_0 - s) f_\infty U ds = -\frac{1}{\sigma} \langle f_\infty U \rangle. \tag{3.17}$$

We multiply (3.13) with  $\bar{U}$  (the complex conjugate of  $U$ ), and integrate over  $[0, \infty)$  with respect to  $f_\infty$ :

$$\begin{aligned} \frac{\tau}{2} \frac{d}{dt} \int_0^\infty f_\infty |U|^2 ds &= -F(\mathbf{k}) \int_0^\infty \frac{\Phi_0 - s}{\sigma} f_\infty \bar{U} ds \langle f_\infty U \rangle \\ &+ \int_0^\infty (\Phi_0 - s) f_\infty U_s \bar{U} ds + \sigma \int_0^\infty f_\infty U_{ss} \bar{U} ds. \end{aligned} \tag{3.18}$$

After applying (3.17) to the first term on the right-hand side and integrating the last term by parts, we find that  $U$  satisfies (where  $\Re(z)$  denotes the real part of the complex



number  $z$ )

$$\begin{aligned} \frac{\tau}{2} \frac{d}{dt} \int_0^\infty f_\infty |U|^2 ds &= \sigma F(\mathbf{k}) \left| \int_0^\infty \partial_s f_\infty U ds \right|^2 - \sigma \Re(f_\infty \bar{U} \partial_s U)|_{s=0} \\ &\quad - \sigma \int_0^\infty f_\infty |U_s|^2 ds \\ &= \sigma (F(\mathbf{k}) - 1) \left| \int_0^\infty \partial_s f_\infty U ds \right|^2 - \sigma \Re(f_\infty \bar{U} \partial_s U)|_{s=0} \\ &\quad + \sigma \left| \int_0^\infty \partial_s f_\infty U ds \right|^2 - \sigma \int_0^\infty f_\infty |U_s|^2 ds. \end{aligned}$$

Using the boundary condition on the second term above and by again utilizing the equivalence (3.17), we get

$$\begin{aligned} \frac{\tau}{2} \frac{d}{dt} \int_0^\infty f_\infty |U|^2 ds &= \sigma (F(\mathbf{k}) - 1) \left| \int_{R_+} \partial_s f_\infty U ds \right|^2 \\ &\quad + \sigma F(\mathbf{k}) \Re \left( (f_\infty \bar{U})|_{s=0} \int_0^\infty \partial_s f_\infty U ds \right) \\ &\quad + \sigma \left| \int_0^\infty \partial_s f_\infty U ds \right|^2 - \sigma \int_0^\infty f_\infty |U_s|^2 ds \\ &= \frac{1}{\sigma} (F(\mathbf{k}) - 1) |\langle f_\infty U \rangle|^2 - F(\mathbf{k}) \Re \left( (f_\infty \bar{U})|_{s=0} \langle f_\infty U \rangle \right) \\ &\quad + \sigma \left| \int_0^\infty \partial_s f_\infty U ds \right|^2 - \sigma \int_0^\infty f_\infty |U_s|^2 ds. \end{aligned}$$

Performing an integration by parts on the second to last integral and then yet again using (3.17),

$$\begin{aligned} \left| \int_0^\infty \partial_s f_\infty U ds \right|^2 &= \left| (f_\infty U)|_{s=0} + \int_0^\infty f_\infty U_s ds \right|^2 \\ &= (f_\infty^2 |U|^2)|_{s=0} + 2\Re \left( (f_\infty \bar{U})|_{s=0} \int_0^\infty f_\infty U_s ds \right) \\ &\quad + \left| \int_0^\infty f_\infty U_s ds \right|^2 \\ &= -(f_\infty^2 |U|^2)|_{s=0} + \frac{2}{\sigma} \Re \left( (f_\infty \bar{U})|_{s=0} \langle f_\infty U \rangle \right) \\ &\quad + \left| \int_0^\infty f_\infty U_s ds \right|^2, \end{aligned}$$

we arrive at

$$\begin{aligned} & \frac{\tau}{2} \frac{d}{dt} \int_0^\infty f_\infty |U|^2 ds \\ &= \frac{1}{\sigma} (F(\mathbf{k}) - 1) |\langle f_\infty U \rangle|^2 + (2 - F(\mathbf{k})) \Re((f_\infty \bar{U})|_{s=0} \langle f_\infty U \rangle) - \sigma (f_\infty^2 |U|^2)|_{s=0} \\ & \quad + \sigma \left| \int_0^\infty f_\infty U_s ds \right|^2 - \sigma \int_0^\infty f_\infty |U_s|^2 ds. \end{aligned}$$

From (3.12) and the definition of  $U$ , it can be shown that  $\langle f_\infty U \rangle$  follows

$$\frac{\tau}{2} \frac{d}{dt} |\langle f_\infty U \rangle|^2 = (F(\mathbf{k}) - 1) |\langle f_\infty U \rangle|^2 + \sigma \Re((f_\infty \bar{U})|_{s=0} \langle f_\infty U \rangle),$$

with  $F(\mathbf{k})$  as in (3.7). We add the two equalities,

$$\begin{aligned} & \frac{\alpha\tau}{2} \frac{d}{dt} \int_0^\infty f_\infty |U|^2 ds + \frac{\beta\tau}{2\sigma} \frac{d}{dt} |\langle f_\infty U \rangle|^2 \\ &= \frac{\alpha + \beta}{\sigma} (F(\mathbf{k}) - 1) |\langle f_\infty U \rangle|^2 \\ & \quad + (\alpha(2 - F(\mathbf{k})) + \beta) \Re((f_\infty \bar{U})|_{s=0} \langle f_\infty U \rangle) - \alpha\sigma (f_\infty^2 |U|^2)|_{s=0} \\ & \quad + \alpha\sigma \left| \int_0^\infty f_\infty U_s ds \right|^2 - \alpha\sigma \int_0^\infty f_\infty |U_s|^2 ds. \end{aligned}$$

By setting  $\alpha = 1$  and  $\beta = -F(\mathbf{k})$ , we get

$$\begin{aligned} & \frac{\tau}{2} \frac{d}{dt} \left( \int_0^\infty f_\infty |U|^2 ds - \frac{F(\mathbf{k})}{\sigma} |\langle f_\infty U \rangle|^2 \right) = - \left| (F(\mathbf{k}) - 1) \frac{1}{\sqrt{\sigma}} \langle f_\infty U \rangle + \sqrt{\sigma} (f_\infty U)|_{s=0} \right|^2 \\ & \quad + \sigma \left| \int_0^\infty f_\infty U_s ds \right|^2 - \sigma \int_0^\infty f_\infty |U_s|^2 ds. \end{aligned} \tag{3.19}$$

In the above derivation we have used that

$$2(1 - F(\mathbf{k})) \Re((f_\infty \bar{U})|_{s=0} \langle f_\infty U \rangle) = (1 - F(\mathbf{k})) ((f_\infty \bar{U})|_{s=0} \langle f_\infty U \rangle + (f_\infty U)|_{s=0} \langle f_\infty \bar{U} \rangle).$$

By applying the Cauchy–Schwarz inequality to  $\left| \int_0^\infty f_\infty U_s ds \right|^2$  with  $u_1 = f_\infty^{1/2}$ ,  $u_2 = f_\infty^{1/2} U_s$ , we see that the right-hand side of (3.19) is non-positive. However, it is not straightforward to determine whether the right-hand side is strictly negative or if there is a Grönwall type decay estimate to be obtained from this expression. In particular, the Cauchy–Schwarz inequality with the chosen functions  $u_1 = f_\infty^{1/2}$ ,  $u_2 = f_\infty^{1/2} U_s$

is an equality,

$$\left| \int_0^\infty f_\infty U_s ds \right|^2 = \left( \int_0^\infty f_\infty ds \right) \left( \int_0^\infty f_\infty |U_s|^2 ds \right) = \int_0^\infty f_\infty |U_s|^2 ds,$$

whenever  $u_2 = c(t)u_1$  for any function  $c(t)$ , which here means that  $U = c(t)(s - \langle f_\infty \rangle)$  when we make sure that the requirement  $\int_0^\infty U f_\infty ds = 0$  holds.

**Part II:** To separate the linear part from the nonlinear part, we split  $U$  into  $U = V + c(t)(s - \langle f_\infty \rangle)$ . We have that  $\int_0^\infty V f_\infty ds = 0$ . Next, we choose  $c(t)$  such that  $V$  and  $s - \langle f_\infty \rangle$  are orthogonal with respect to the measure  $f_\infty ds$ , i.e.,  $\int_0^\infty V(s - \langle f_\infty \rangle) f_\infty ds = 0$ . For this we choose

$$c(t) = \langle f_\infty U \rangle \left( \int_0^\infty s^2 f_\infty(s) ds - \langle f_\infty \rangle^2 \right)^{-1} = \frac{\langle f_\infty U \rangle}{M_\infty}, \tag{3.20}$$

where  $M_\infty$  is defined in (3.14). We can summarise this as

$$\begin{aligned} \int_0^\infty f_\infty |U|^2 ds &= \int_0^\infty f_\infty |V|^2 ds + M_\infty |c(t)|^2, \\ \langle f_\infty U \rangle &= M_\infty c(t), \\ (f_\infty U)|_{s=0} &= f_\infty(0)(V(0) - c(t)\langle f_\infty \rangle), \end{aligned}$$

such that (3.19) becomes

$$\begin{aligned} &\frac{\tau}{2} \frac{d}{dt} \left( \int_0^\infty f_\infty |V|^2 ds + \left( 1 - M_\infty \frac{F(\mathbf{k})}{\sigma} \right) M_\infty |c(t)|^2 \right) \\ &= - \left| (F(\mathbf{k}) - 1) \frac{1}{\sqrt{\sigma}} M_\infty c(t) + \sqrt{\sigma} f_\infty (V - c(t)\langle f_\infty \rangle) \Big|_{s=0} \right|^2 \\ &\quad + \sigma \left| \int_0^\infty f_\infty V_s ds \right|^2 - \sigma \int_0^\infty f_\infty |V_s|^2 ds. \end{aligned} \tag{3.21}$$

From the orthogonality  $\int_0^\infty V(s - \langle f_\infty \rangle) f_\infty ds = 0$ , we get

$$\begin{aligned} (f_\infty V)|_{s=0} &= - \int_0^\infty \partial_s (f_\infty V) ds = - \int_0^\infty \partial_s f_\infty V ds - \int_0^\infty f_\infty V_s ds \\ &= \int_0^\infty \frac{s - \Phi_0}{\sigma} f_\infty V ds - \int_0^\infty f_\infty V_s ds = - \int_0^\infty f_\infty V_s ds. \end{aligned}$$

Inserting this into the first square in (3.21) and then expanding the square, (3.21) turns into

$$\begin{aligned} & \frac{\tau}{2} \frac{d}{dt} \left( \int_0^\infty f_\infty |V|^2 ds + \left( 1 - M_\infty \frac{F(\mathbf{k})}{\sigma} \right) M_\infty |c(t)|^2 \right) \\ &= -R^2 |c(t)|^2 + 2\sqrt{\sigma} R \Re \left( c(t) \int_0^\infty V_s f_\infty ds \right) - \sigma \int_0^\infty f_\infty |V_s|^2 ds \\ &= \int_0^\infty \left( -R^2 |c(t)|^2 + 2\sqrt{\sigma} R \Re (c(t) V_s) - \sigma |V_s|^2 \right) f_\infty ds \\ &= - \int_0^\infty |Rc(t) - \sqrt{\sigma} V_s|^2 f_\infty ds, \end{aligned}$$

with  $R = (F(\mathbf{k}) - 1) \frac{M_\infty}{\sqrt{\sigma}} - \sqrt{\sigma} f_\infty(0) \langle f_\infty \rangle$ . We now apply the Poincaré inequality with respect to the measure  $f_\infty(s) ds$  (Muckenhoupt 1972; Roustant et al. 2017), given by

$$C \int_0^\infty \left| G - \int_0^\infty G f_\infty ds \right|^2 f_\infty ds \leq \int_0^\infty |G_s|^2 f_\infty ds.$$

Let  $G_s = Rc(t) - \sqrt{\sigma} V_s$ . Then  $G = Rc(t)(s - \langle f_\infty \rangle) - \sqrt{\sigma} V$  and  $\int_0^\infty G f_\infty ds = 0$ , such that

$$\begin{aligned} & \frac{\tau}{2} \frac{d}{dt} \left( \int_0^\infty f_\infty |V|^2 ds + \left( 1 - M_\infty \frac{F(\mathbf{k})}{\sigma} \right) M_\infty |c(t)|^2 \right) \\ & \leq -C \int_0^\infty |Rc(t)(s - \langle f_\infty \rangle) - \sqrt{\sigma} V|^2 f_\infty ds \\ & = -CR^2 M_\infty |c(t)|^2 - C\sigma \int_0^\infty |V|^2 f_\infty ds. \end{aligned} \tag{3.22}$$

In the above calculation, we have used the orthogonality of  $V$  and  $c(t)(s - \langle f_\infty \rangle)$  with respect to  $f_\infty ds$ .

**Part III:** Defining

$$Q(t) := \int_0^\infty f_\infty |V|^2 ds, \quad D(t) := M_\infty H |c(t)|^2, \quad H := 1 - M_\infty \frac{F(\mathbf{k})}{\sigma},$$

the inequality (3.22) reads

$$\begin{aligned} \frac{d}{dt} (Q(t) + D(t)) & \leq -\frac{2}{\tau} C \frac{R^2}{H} D(t) - \frac{2}{\tau} C \sigma Q(t) \\ & \leq -\frac{2}{\tau} C \min \left\{ \frac{R^2}{H}, \sigma \right\} (Q(t) + D(t)), \end{aligned}$$

due to (3.15). Note that also due to (3.15),  $D(t) \geq 0$ . It can be checked after some tedious computations using the explicit expression of  $f_\infty$  in (3.3)–(3.5) that  $R < 0$

when  $\Phi \geq 0$ . This leads to the exponential decay by an application of Grönwall’s inequality

$$Q(t) + D(t) \leq (Q(0) + D(0)) \exp\left(-\frac{2}{\tau} C \min\left\{\frac{R^2}{H}, \sigma\right\} t\right). \tag{3.23}$$

Thus, we can conclude that  $U$  is asymptotically stable. What remains to show is that the same holds for  $u^\beta$ . We multiply (3.12) with  $\bar{u}^\beta$  and integrate over  $[0, \infty)$ ,

$$\begin{aligned} \frac{\tau}{2} \frac{d}{dt} \int_0^\infty f_\infty |u^\beta|^2 ds &= -\frac{\Phi'_0}{4} \hat{W}(\mathbf{k}) \int_0^\infty \frac{\Phi_0 - s}{\sigma} f_\infty \bar{u}^\beta ds \langle f_\infty U \rangle \\ &\quad + \int_0^\infty (\Phi_0 - s) f_\infty u_s^\beta \bar{u}^\beta ds + \sigma \int_0^\infty f_\infty u_{ss}^\beta \bar{u}^\beta ds. \end{aligned}$$

As done for  $U$ , we integrate the last term by parts and use the boundary condition such that

$$\begin{aligned} \frac{\tau}{2} \frac{d}{dt} \int_0^\infty f_\infty |u^\beta|^2 ds &= \\ &- \frac{\Phi'_0}{4} \hat{W}(\mathbf{k}) \left( \int_0^\infty \partial_s f_\infty \bar{u}^\beta ds + \bar{u}^\beta f_\infty|_{s=0} \right) \langle f_\infty U \rangle - \sigma \int_0^\infty f_\infty |u_s^\beta|^2 ds \\ &= \frac{\Phi'_0}{4} \hat{W}(\mathbf{k}) \left( \int_0^\infty f_\infty \bar{u}_s^\beta ds \right) \langle f_\infty U \rangle - \sigma \int_0^\infty f_\infty |u_s^\beta|^2 ds \\ &\leq \frac{|\Phi'_0|}{4} |\hat{W}(\mathbf{k})| \left( \frac{1}{2\alpha} |\langle f_\infty U \rangle|^2 + \frac{\alpha}{2} \left| \int_0^\infty f_\infty \bar{u}_s^\beta ds \right|^2 \right) \\ &\quad - \sigma \int_0^\infty f_\infty |u_s^\beta|^2 ds, \end{aligned}$$

where  $\alpha$  is to be determined. We apply the Cauchy–Schwarz inequality to the middle term and rearrange,

$$\begin{aligned} \frac{d}{dt} \int_0^\infty f_\infty |u^\beta|^2 ds &\leq \underbrace{\frac{|\Phi'_0|}{4\tau\alpha} |\hat{W}(\mathbf{k})| |\langle f_\infty U \rangle|^2}_{=: C_1} \\ &\quad - \underbrace{\left( -\alpha \frac{|\Phi'_0|}{4\tau} |\hat{W}(\mathbf{k})| + \frac{2}{\tau} \sigma \right)}_{=: C_2} \int_0^\infty f_\infty |u_s^\beta|^2 ds. \end{aligned}$$

We now choose  $\alpha > 0$  such that  $C_2 > 0$ , and then apply Poincaré’s inequality to the integral,

$$\begin{aligned} \frac{d}{dt} \int_0^\infty f_\infty |u^\beta|^2 ds &\leq C_1 |\langle f_\infty U \rangle|^2 - C_2 \int_0^\infty f_\infty |u^\beta|^2 ds \\ &\leq \tilde{C}_1 \exp(-C_3 t) - \tilde{C}_2 \int_0^\infty f_\infty |u^\beta|^2 ds, \end{aligned}$$

where the exponential decay (3.23) is applied in the last step. This leads to

$$\begin{aligned} \int_0^\infty f_\infty |u^\beta|^2(t) ds &\leq \frac{\tilde{C}_1}{\tilde{C}_2 - C_3} \left( \exp(-C_3 t) - \exp(-\tilde{C}_2 t) \right) \\ &\quad + \int_0^\infty f_\infty |u^\beta|^2(0) ds \exp(-\tilde{C}_2 t). \end{aligned}$$

One can avoid  $\tilde{C}_2 = C_3$  by choosing  $\alpha$  appropriately.

The asymptotic stability of  $f_\infty$  in  $L^2(\Omega \times [0, \infty))$  for the set of perturbations given by (3.11) now follows by an application of Parseval’s identity with respect to  $\mathbf{x}$  to

$$\int_0^\infty \int_\Omega |h^\beta|^2 d\mathbf{x} ds \leq \frac{1}{Z} \int_0^\infty \int_\Omega |h^\beta|^2 d\mathbf{x} \frac{1}{f_\infty(s)} ds,$$

the identity (3.11), and the estimate above. Notice that  $f_\infty(s) \leq \frac{1}{Z}$  from (3.4). □

**Remark 3.8** In principle, the linear stability analysis is valid only for smooth  $\Phi$ . However, the stability condition (3.15) of the linearised problem does only depend on  $\Phi'$  such that the result holds for the linearised system with  $\Phi(x) = (x)^+$ . Notice that condition (3.15) is continuous with respect to the regularisation parameter  $\varepsilon$  in Remark 3.2 for which the linearisation is valid.

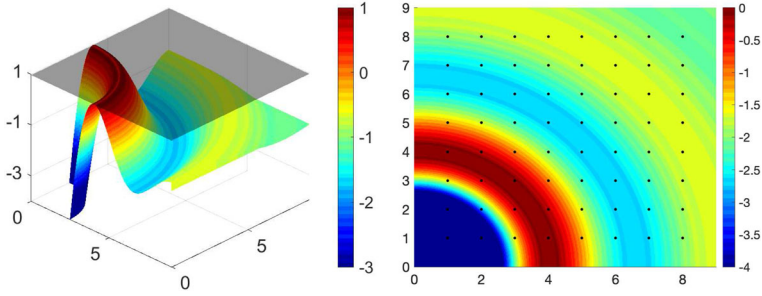
We also remark that the value of  $R$  in the proof above remains strictly negative as long as  $\varepsilon$  is small enough despite the fact that  $\Phi_\varepsilon$  may give negative values.

## 4 Bifurcation diagrams and phase transitions

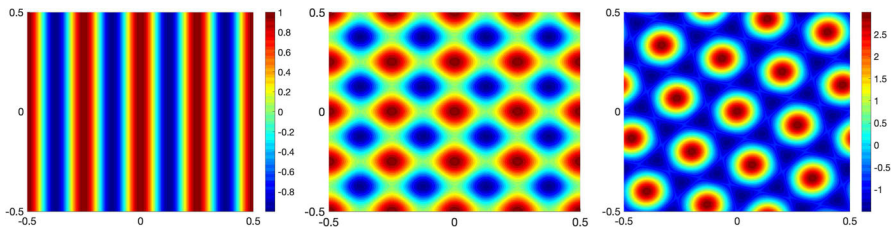
With our linear stability analysis at hand, we will now investigate how the stationary patterns of the nonlinear PDE system (2.7)–(2.8) change as we vary the noise parameter  $\sigma$ . We do this by numerically computing bifurcation branches from the spatially homogeneous solution  $f_\infty$  for various choices of the modulation function  $\Phi$ . The numerical procedure is described in more detail in Sect. 4.2.

### 4.1 Instability of the linearised system

First, to connect the linear stability analysis in Sect. 3 with the patterns we observe for the full system (2.7)–(2.8), we start by investigating the dominant Fourier modes



**Fig. 3** Plots of the linear stability condition on  $\frac{M_\infty}{\sigma} F(\mathbf{k})$  for  $\Phi(x) = \Phi_\varepsilon(x)$ ,  $\varepsilon = 0.01$ , with  $W(|\mathbf{x}|) = -0.005 \cdot 128^2 (1 + \tanh(10 - 50|\mathbf{x}|))$ . Left:  $\frac{M_\infty}{\sigma} F(\mathbf{k})$  at the minimal value of the noise  $\sigma$  for linear stability (3.15). Right: Associated contour plot highlighting the Fourier modes  $\mathbf{k} = 2\pi (k_1 \ k_2)^\top$  with black dots at  $(k_1, k_2)$



**Fig. 4** Linear combinations of  $\cos(\mathbf{k} \cdot \mathbf{x})$  are plotted against  $x$  horizontally and  $y$  vertically. The Fourier modes  $\mathbf{k}$  are chosen to be the points with the largest values of  $F(\mathbf{k})$  in Fig. 3b with  $W(|\mathbf{x}|) = -0.005 \cdot 128^2 (1 + \tanh(10 - 50|\mathbf{x}|))$ . From left to right:  $\mathbf{k} = 2\pi(4, 0)$ ,  $\mathbf{k} = 2\pi(4, 0), 2\pi(0, 4)$  and  $\mathbf{k} = 2\pi(4, 1), 2\pi(1, 4), 2\pi(3, -3)$

$\mathbf{k}$  of the perturbations  $h^\beta(\mathbf{x}, s, t) = f_\infty(s) \sum_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{x}) u_{\mathbf{k}}^\beta(s, t)$ , where  $u_{\mathbf{k}}^\beta$  satisfies (3.12). The stability condition (3.15) of Theorem 3.6 is visualised by plotting the function  $\frac{M_\infty}{\sigma} F(\mathbf{k})$  for the case  $\Phi(x) = \Phi_\varepsilon(x)$ ,  $\varepsilon = 0.01$  (cf. (3.6)), with  $W(|\mathbf{x}|) = -0.005 \cdot 128^2 (1 + \tanh(10 - 50|\mathbf{x}|))$ , against the modes  $\mathbf{k}$  in Fig. 3. The figure illustrates that the maximum points over the lattice  $\mathbf{k} = (2\pi k_1 \ 2\pi k_2)^\top$ ,  $k_1, k_2 \in \mathbb{Z}$  for this particular  $W$  are among  $\{(4, 0), (4, 1), (3, 3), (1, 4), (0, 4)\}$  and their reflections by symmetries with respect to the origin,  $k_1 = 0$ , and  $k_2 = 0$ . Note that the modulation function  $\Phi$  of the firing rate has no effect on the maximum points of  $F(\mathbf{k})$  since it enters through as an amplification factor in (3.7). As a consequence, we may expect that the patterns leading the instability of the homogeneous in space stationary state  $f_\infty$  are driven by a combination of these Fourier modes. Examples of possible patterns generated as a sum of cosines depending on the dominant modes, i.e., the maximum points of  $F(\mathbf{k})$  over the lattice  $\mathbf{k} = (2\pi k_1 \ 2\pi k_2)^\top$ ,  $k_1, k_2 \in \mathbb{Z}$ , are depicted in Fig. 4. Notice that the rightmost plot displays a hexagonal pattern similar to the ones generated by the nonlinear PDE system (2.7)–(2.8) in the top and middle row of Fig. 1. See a similar strategy to this for a related problem in Murray (2002, Ch. 12).



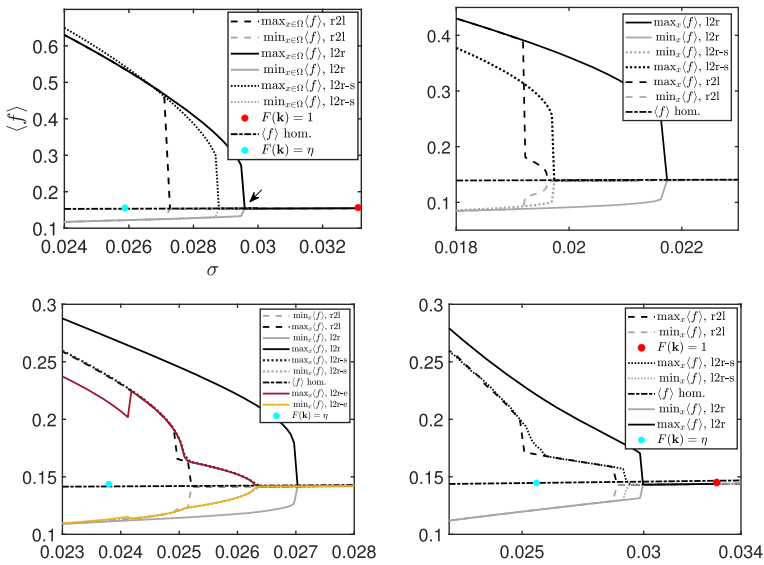
## 4.2 Bifurcations and phase transitions of the nonlinear PDE system

We now continue with our examination of the stability of the spatial patterns, generated by the full system (2.7)–(2.8), with respect to  $\sigma$ . In Fig. 5, we numerically compute bifurcation branches from the spatially homogeneous solution  $f_\infty$  for different modulation functions  $\Phi$  with  $W(|\mathbf{x}|) = -0.005 \cdot 128^2 (1 + \tanh(10 - 50|\mathbf{x}|))$  for the nonlinear problem (2.7)–(2.8). This is done by using a continuation based method on  $\sigma$  over an accurate numerical solver for the evolution in time of Fokker–Planck like equations developed in Carrillo et al. (2015); further details are given in Appendix 1. The continuation method starts either at the largest or the smallest noise value  $\sigma$  of the interval under consideration and it solves for the evolution in time of (2.7)–(2.8) up to stabilisation to a steady value. This allows for recursive computation of the stationary states for smaller or larger values of the noise by taking as initial data the already computed steady state. With this procedure we ensure, up to numerical accuracy, that we compute the stable stationary states, either by sweeping the noise values from left-to-right (l2r) or from right-to-left (r2l).

Each subplot in Fig. 5 shows the maximum and minimum over space  $\mathbf{x}$  of the average activity rate  $\langle f \rangle(\mathbf{x}) = \sum_\beta \langle f^\beta \rangle(\mathbf{x})$  of the computed steady states for each noise value  $\sigma$ . We show both the spatial maximum and minimum of  $\langle f \rangle(\mathbf{x})$  to illustrate the fact that the computed stationary states are not spatially homogeneous, in other words, that they lead to spatial patterns. We also plot the spatially homogeneous branch numerically solving the implicit expression (3.5) as reference. The red dots indicate the stability threshold in  $\sigma$  for the condition  $F(\mathbf{k}) < 1$ , as in Lemma 3.5, to hold.

In Fig. 5a, we observe the bifurcation branches for the sigmoid function  $\Phi(x) = 1/(1 + \exp(-15x))$ . All of them show a sharp discontinuity at different noise values. We restrict the discussion to the lines representing the spatial maximum. We first focus on the full line (l2r) and the dashed line (r2l) that connect two bifurcation branches at different noise values corresponding to a hexagonal-like pattern similar to Fig. 6a. This clearly indicates that there is a discontinuous phase transition near the noise value indicated by the arrow. The fact that the l2r and r2l curves do not coincide further indicates that there is a hysteresis phenomenon. This conclusion is supported by the fact that the blue dot, the minimum noise value for linear stability (3.15), is to the left of both branches. This allows the possibility of branches of dynamically unstable steady states bending backwards in noise at the phase transition point. Unstable branches are not computable with our numerical approach. Finally, we find a second bifurcation branch given by the dotted line (l2r-s) in Fig. 5c corresponding to a stripe-like pattern similar to Fig. 6b. This branch was found by imposing a particular symmetry on the initial data, i.e., enforcing a horizontal band with activity level one.

In Fig. 5b–d, we show analogous computations for the case of the modulation function given by  $\Phi(x) = (x)^+$  and its regularisations  $\Phi(x) = \Phi_\varepsilon(x)$  with  $\varepsilon = 0.1$  and  $\varepsilon = 0.01$ . Similarly to Fig. 5a, we observe a discontinuous phase transition for the full line (l2r) and the dashed line (r2l), and the linear stability blue dot is also to the left of the phase transition point as above. We remark that the blue and the red dots may lie outside the noise intervals in Fig. 5b–d, but they follow the same order.



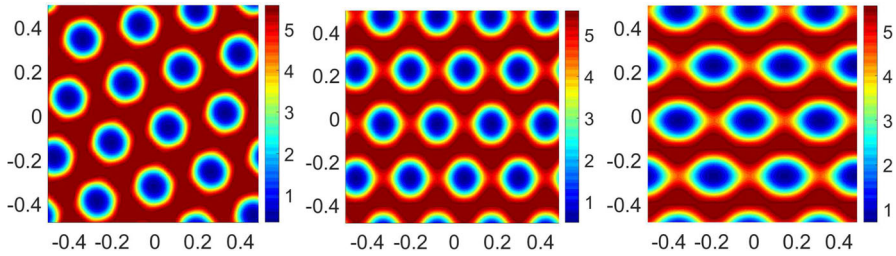
**Fig. 5** Bifurcation plots of  $\langle f \rangle$  with respect to  $\sigma$ , where  $W(|x|) = -0.005 \cdot 128^2 (1 + \tanh(10 - 50|x|))$  for different modulation functions  $\Phi$ . The red dots show the stability threshold  $F(k) = 1$  for no noise while the blue dots correspond to the stability threshold (3.15) with noise, where  $\eta = \frac{\sigma_c}{M_\infty(\sigma_c)}$ , and  $\sigma_c$  is the threshold value for linear stability. Top row:  $\Phi(x) = \frac{1}{1 + \exp(-15x)}$  (left)=(a),  $\Phi = \Phi_\varepsilon, \varepsilon = 0.1$  (right)=(b), bottom row:  $\Phi = \Phi_\varepsilon, \varepsilon = 0.01$  (left)=(c), and  $\Phi(x) = (x)^+$  (right)=(d)

Similar conclusions as above lead to hysteresis phenomena and the possible existence of unstable branches not obtainable with our present numerical approach.

The case of  $\varepsilon = 0.1$  in Fig. 5b resembles the behaviour observed for the sigmoid function in Fig. 5a. The hexagonal-like patterns are the preferred stable configurations both for generic initial data, full line (l2r), and starting with small perturbations of the homogeneous stationary state, dashed line (r2l). Again stripe-like patterns are obtained by choosing specific initial data. Similar branches and the numerical observation that the hexagonal-like pattern is the most stable configuration has already been reported for a neural field model without noise (Veltz et al. 2015).

This behaviour changes in Fig. 5c, d. The hexagonal-like patterns are still the preferred stable configurations for generic initial data, full line (l2r). However, starting with small perturbations of the homogeneous stationary state, dashed line (r2l), we connect to the stripe-like bifurcation branch, dotted line (l2r).

The bifurcation branches and their dynamics gets richer as the regularisation parameter gets smaller. We observe that for  $\varepsilon = 0.01$  in Fig. 5c, there is an additional branch, dark red line (l2r-e), leading to eye-like patterns as in Fig. 6c. This branch jumps to the stripe-like pattern for larger noise values. It is difficult to extract information on the range of noise values  $\sigma \in [0.025, 0.0265]$  since the branch, dark red line (l2r), does not show a sharp transition point while the dashed line (r2l) does. However, this becomes much clearer in the limiting case of the positive part in Fig. 5d. We observe two sharper discontinuous transition points in the stripe-like branch, lead-



**Fig. 6** Stationary patterns of  $f^\beta(s = 0)$  at  $\sigma = 0.022$  for  $\Phi(x) = \Phi_\varepsilon(x)$ ,  $\varepsilon = 0.01$  with  $W(|\mathbf{x}|) = -0.005 \cdot 128^2 (1 + \tanh(10 - 50|\mathbf{x}|))$ : hexagonal-like (a), stripe-like (b), eye-like (c). Left to right: black, dotted, and red line in Fig. 5c

ing to an intermediate pure-stripe branch,  $\sigma \in [0.025, 0.028]$ , before jumping to the homogeneous state, see the dashed line (r2l) and the dotted line (l2r).

### 5 Concluding remarks

The first conclusion of our analysis is that the mean-field limit of the grid cell model (1.1) with constant external input introduced by Burak and Fiete (2009), Couey et al. (2013), and Burak and Fiete (2012), presents phase transitions driven by the noise strength as demonstrated in Figs. 5 and 6. This behaviour resembles the phenomena appearing in the classical Kuramoto model (Kuramoto 1981; Sakaguchi et al. 1988; Acebrón et al. 2005; Carrillo et al. 2019, 2020) for synchronisation and other neural field models Touboul et al. (2012) in the computational neuroscience literature.

It is shown that the homogeneous in space stationary state is linearly unstable for small noise strength, similarly to basic ring and neural field models (MacLaurin and Bressloff 2020; Kilpatrick and Ermentrout 2013; Byrne et al. 2019). We numerically analysed the bifurcation diagram of stationary patterns showing the appearance of different branches identified by their symmetries, see Figs. 5 and 6. Our numerical experiments with random initial data demonstrate that the stationary hexagonal-like pattern in space of the activity level of neurons in Fig. 6a, leading to the solid black bifurcation branches in Fig. 5, has the largest basin of attraction. Moreover, the numerical simulations indicate that there is a sharp transition in the mean activity level together with a hysteresis phenomenon suggesting a discontinuous phase transition. Whether more stationary network patterns exist is another interesting topic.

The crucial implication of this phase transition on the rats' navigation path is that the larger the noise the less localised are the spatial firing fields of each neuron. This can be observed in the bottom row of Fig. 1 which shows that the firing field (coloured in red) gets denser as the noise increases. Moreover, there is a sharp value of the noise after which there is no localisation at all, leading the rats to not being able to orientate themselves in physical space. In other words, the point of transition from a homogeneous pattern (all neurons have the same mean activity level) to a non-homogeneous pattern (neurons at different locations in the network have different mean activity levels) gives an upper bound for the noise strength for which single grid

cells no longer can fire in a hexagonal pattern in physical space when connected with the rats movements through (2.9).

With a hexagonal network activity configuration as in Fig. 6a, a single neuron can create hexagonal neural field patterns in physical space as in the third row of Fig. 1. Exactly how the firing fields in physical space of a single neuron are affected by initial network activity patterns as the ones in Fig. 6b–c remains to be investigated.

From the methodological viewpoint, we remark that as the bifurcation branches are computed using a numerical approximation of the PDE (2.7)–(2.8), they can differ slightly from the actual branches of the PDE itself. To study bifurcations and phase transitions of the nonlinear PDE analytically will require sophisticated mathematical tools, and they will be investigated elsewhere.

From the computational neuroscience viewpoint, we expect that noise driven phase transitions will also naturally appear in related attractor dynamic models as the ones in Burak and Fiete (2012), Agamon and Burak (2020). Instability of homogeneous stationary network patterns should also play an important role therein. Additional investigations of more realistic models of coupled place and grid cells are needed. This will allow to connect with experiments and further contribute to the challenge of how noise affects network dynamics in Rowland et al. (2016, Future Issue 3).

**Acknowledgements** The authors are grateful to Edvard I. Moser for helpful discussions and advice. JAC was supported by the Advanced Grant Nonlocal-CPD (Nonlocal PDEs for Complex Particle Dynamics: Phase Transitions, Patterns and Synchronization) of the European Research Council Executive Agency (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 883363). HH was supported in part by the project Waves and Nonlinear Phenomena (WaNP), no 90104100, from the Research Council of Norway. Parts of this work was conducted while SS was employed at the Department of Mathematical Sciences, NTNU.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

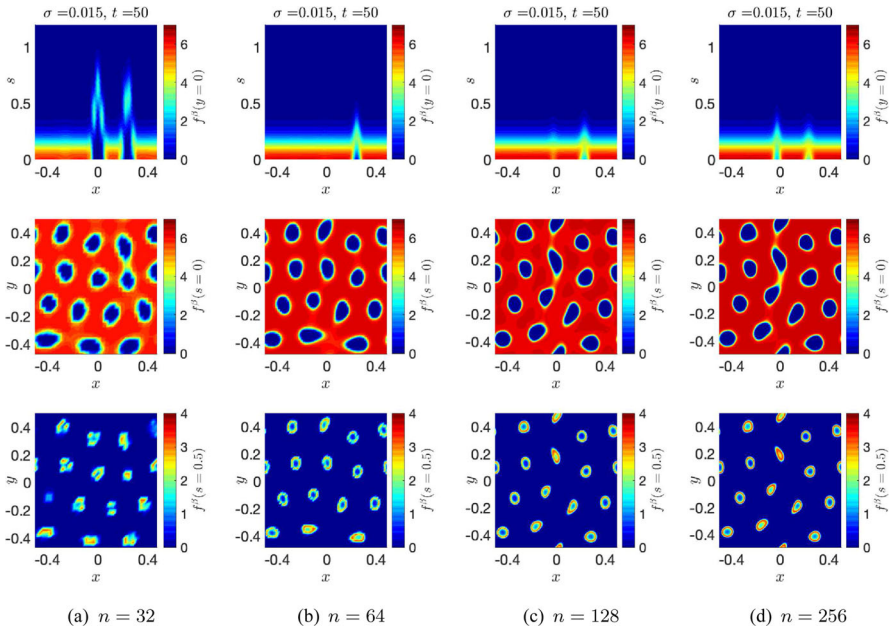
## Appendix: Numerical approach

This section contains details on the numerical approach. We re-write the system (2.7) as

$$\tau \frac{\partial f^\beta}{\partial t} = \frac{\partial}{\partial s} \left( f^\beta \frac{\partial}{\partial s} \left[ \frac{1}{2} (\Phi(W, f, B^\beta) - s)^2 + \sigma \log f^\beta \right] \right),$$

and discretise the system using the first-order (in time  $t$  and in  $s$ ) version of the finite volume numerical method in Carrillo et al. (2015) at every  $\mathbf{x} \in \Omega$  on the spatial mesh.

In all figures in this paper, the time scale is set to  $\tau = 10\text{ms}$ , and the computations are run on a  $64 \times 64 \times 64$  ( $\mathbf{x}, s$ )-grid for each neuron type  $\beta$ . The activity  $s$  is calculated



**Fig. 7** Grid refinement for  $\sigma = 0.015$ ,  $B = 3$ ,  $W(|\mathbf{x}|) = -0.005 \cdot 128^2(1 + 1 \tanh(50(0.2 - |\mathbf{x}|)))$ , and  $\Phi(x) = 1/(1 + \exp(-15x))$ , at  $t = 50\text{ms}$

on the interval  $[0, 1.3]$  and  $\mathbf{x} \in [-0.5, 0.5]^2$ . We impose no-flux boundary conditions in  $s$  and periodic in  $\mathbf{x}$ . We carefully checked that the support of the distribution in  $s$  is essentially inside the interval  $[0, 1.3]$  for all times such that imposing no-flux boundary condition on the right end is a good approximation for  $s \in [0, \infty)$ . The initial distribution is prescribed by randomly setting the activity at one percent of the grid points  $\{\mathbf{x}_n\}_n$ , denoted  $\{\mathbf{x}_{n_i}\}_i$ , in the four sheets to 1, i.e.,

$$f_0^\beta(\mathbf{x}_n, s) = \sum_i \delta_{\mathbf{x}_{n_i}, 1}(\mathbf{x}, s) + \sum_{n \notin \{n_i\}_i} \delta_{\mathbf{x}_n, 0}(s),$$

with  $\beta = 1, \dots, 4$ , where the positions  $\mathbf{x}_{n_i}$  are chosen randomly. The spatial preference is set by shifting the connectivity matrix determined by  $W$  one cell to the north, south, west, or east, respectively.

All figures in Sect. 3 are computed with  $B = 3$ . The value of each bifurcation branch in Fig. 5 is, for each  $\sigma$ , found by a continuation method in  $\sigma$  running the simulation for a minimum time of 2000ms. The calculation of the branch value is stopped when either the numerical time derivative satisfies the numerical equivalent of

$$\frac{d}{dt} \left\| \sum_\beta f^\beta \right\|_{L^1(\Omega \times [0, \infty))} \leq 10^{-8},$$

**Table 1**  $L^1$  and  $L^2$  (in  $\mathbf{x}$ ) errors and order of convergence (OOC) of the mean for  $n = 32, 64, 128$ 

| $n$ | $L^1$  | $L^2$  | OOC $L^1$ | OOC $L^2$ |
|-----|--------|--------|-----------|-----------|
| 32  | 0.1033 | 0.1993 | —         | —         |
| 64  | 0.0565 | 0.1085 | 0.87      | 0.88      |
| 128 | 0.0098 | 0.0175 | 2.53      | 2.64      |

or the maximum time of 6000ms is reached. The respective noise intervals are divided into at least 100 points.

### A.1 Grid refinement of the numerical method

To briefly check the robustness of the numerical method described in the manuscript, we performed a standard grid refinement analysis within a computationally feasible computational range from  $n = 64$  to  $n = 256$  cells in  $\mathbf{x}$  and  $s$ . For  $n = 256$ , the initial data  $f_{\Delta}^{\beta,0}(x)|_{s=1}$  is randomly set to  $1/256$  at 1% of the locations  $x$ , and  $f_{\Delta}^{\beta,0}(x)|_{s=0}$  is set to  $1/256$  on the complement. The rest is set to zero. On the coarser grids, we have used piecewise averages of this initial data. In Fig. 7, the solutions on a grid consisting of  $n^3$  cubes, with  $n = 32, 64, 128$ , and  $256$ , are plotted. The corresponding  $L^1$  and  $L^2$  errors can be found in the table below. The coarser solutions ( $n = 32, 64, 128$ ) are compared to the one computed on the grid consisting of  $256^3$  cubes (Table 1). It can be observed from the table and the figure that the numerical method used is stable when refining the grid.

## References

- Acebrón JA, Bonilla LL, Pérez Vicente CJ, Ritort F, Spigler R (2005) The Kuramoto model: a simple paradigm for synchronization phenomena. *Rev Mod Phys* 77:137–185
- Agamon 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
- Amari S-I (1977) Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol Cybern* 27(2):77–87
- Arous GB, Guionnet A (1995) Large deviations for Langevin spin glass dynamics. *Probab Theory Relat Fields* 102(4):455–509
- Bressloff PC (2012) Spatiotemporal dynamics of continuum neural fields. *J Phys A* 45(3):033001
- Bressloff PC (2019) Stochastic neural field model of stimulus-dependent variability in cortical neurons. *PLoS Comput Biol* 15(3):1–33
- Burak Y, Fiete I (2009) Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput Biol* 5(2):e1000291
- Burak Y, Fiete I (2012) Fundamental limits on persistent activity in networks of noisy neurons. *PNAS* 109:17645–17650
- Byrne Á, Avitabile D, Coombes S (2019) Next-generation neural field model: the evolution of synchrony within patterns and waves. *Phys Rev E* 99(1):012313
- Cañizo JA, Carrillo JA, Rosado J (2011) A well-posedness theory in measures for some kinetic models of collective motion. *Math Models Methods Appl Sci* 21(3):515–539
- Cabana T, Touboul JD (2018) Large deviations for randomly connected neural networks: I. Spatially extended systems. *Adv Appl Probab* 50(3):944–982
- Cai D, Tao L, Rangan AV, McLaughlin DW (2006) Kinetic theory for neuronal network dynamics. *Commun Math Sci* 4(1):97–127



- Carrillo JA, Chertock A, Huang Y (2015) A finite-volume method for nonlinear nonlocal equations with a gradient flow structure. *Commun Comput Phys* 17(1):233–258
- Carrillo JA, Choi Y-P, Pareschi L (2019) Structure preserving schemes for the continuum Kuramoto model: phase transitions. *J Comput Phys* 376:365–389
- Carrillo JA, Clini A, Solem S (2021) The mean field limit of stochastic differential equation systems modelling grid cells. [arXiv:2112.06213](https://arxiv.org/abs/2112.06213)
- Carrillo JA, Cordier S, Mancini S (2011) A decision-making Fokker–Planck model in computational neuroscience. *J Math Biol* 63(5):801–830
- Carrillo JA, Cordier S, Mancini S (2013) One dimensional Fokker–Planck reduced dynamics of decision making models in computational neuroscience. *Commun Math Sci* 11(2):523–540
- Carrillo JA, Gvalani RS, Pavliotis GA, Schlichting A (2020) Long-time behaviour and phase transitions for the McKean–Vlasov equation on the torus. *Arch Ration Mech Anal* 235(1):635–690
- Couey JJ, Witoelar A, Zhang S-J, Zheng K, Ye J, Dunn B, Czajkowski R, Moser M-B, Moser EI, Roudi Y, Witter MP (2013) Recurrent inhibitory circuitry as a mechanism for grid formation. *Nat Neurosci* 16:318–324
- Ermentrout GB, Cowan JD (1979) A mathematical theory of visual hallucination patterns. *Biol Cybern* 34(3):137–150
- Ermentrout GB, Terman DH (2010) *Mathematical foundations of neuroscience*, vol 35. *Interdisciplinary Applied Mathematics*. Springer, New York
- Faugeras O, Inglis J (2015) Stochastic neural field equations: a rigorous footing. *J Math Biol* 71(2):259–300
- Faugeras O, Touboul J, Cessac B (2009) A constructive mean-field analysis of multi population neural networks with random synaptic weights and stochastic inputs. *Front Comput Neurosci* 3
- Gardner RJ, Hermansen E, Pachitariu M, Burak Y, Baas NA, Dunn BA, Moser M-B, Moser EI (2022) Toroidal topology of population activity in grid cells. *Nature*
- Golse F (2016) On the dynamics of large particle systems in the mean field limit. In: *Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity*, volume 3 of *Lect. Notes Appl. Math. Mech.* Springer, Cham, pp 1–144
- Guionnet A (1997) Averaged and quenched propagation of chaos for spin glass dynamics. *Probab Theory Relat Fields* 109(2):183–215
- Hafting T, Fyhn M, Molden S, Moser M-B, Moser EI (2005) Microstructure of a spatial map in the entorhinal cortex. *Nature* 436:801–806
- Hauray M, Jabin P-E (2007)  $N$ -particles approximation of the Vlasov equations with singular potential. *Arch Ration Mech Anal* 183(3):489–524
- Hauray M, Jabin P-E (2015) Particle approximation of Vlasov equations with singular forces: propagation of chaos. *Ann Sci Éc Norm Supér (4)* 48(4):891–940
- Hopfield JJ (1984) Neurons with graded response have collective computational properties like those of two-state neurons. *Proc Natl Acad Sci* 81(10):3088–3092
- Jabin P-E (2014) A review of the mean field limits for Vlasov equations. *Kinet Relat Models* 7(4):661–711
- Kilpatrick ZP (2014) Coupling layers regularizes wave propagation in stochastic neural fields. *Phys Rev E* 89:022706
- Kilpatrick ZP, Ermentrout B (2013) Wandering bumps in stochastic neural fields. *SIAM J Appl Dyn Syst* 12(1):61–94
- Kilpatrick ZP, Poll DB (2017) Neural field model of memory-guided search. *Phys Rev E* 96:062411
- Kuramoto Y (1981) Rhythms and turbulence in populations of chemical oscillators. *Phys A* 106(1–2):128–143
- Lions P-L, Sznitman A-S (1984) Stochastic differential equations with reflecting boundary conditions. *Commun Pure Appl Math* 37(4):511–537
- MacLaurin JN, Bressloff PC (2020) Wandering bumps in a stochastic neural field: a variational approach. *Phys D* 406(132403):9
- McNaughton B, Moser E, Moser M-B (2017) Spatial representation in the hippocampal formation: a history. *Nat Neurosci* 20:1448–1464
- McNaughton BL, Battaglia FP, Jensen O, Moser EI, Moser M-B (2006) Path integration and the neural basis of the ‘cognitive map’. *Nat Rev Neurosci* 7(8):663–678
- Moynot O, Samuelides M (2002) Large deviations and mean-field theory for asymmetric random recurrent neural networks. *Probab Theory Relat Fields* 123(1):41–75
- Muckenhoupt B (1972) Hardy’s inequality with weights. *Studia Math* 44:31–38
- Murray J (2002) *Mathematical biology*. Springer-Verlag, New York



- Pinto D, Brumberg JC, Simons DJ, Ermentrout B (1996) A quantitative population model of whisker barrels: Re-examining the Wilson-Cowan equations. *J Comput Neurosci* 3:247–264
- Rolls ET, Deco G (2010) *The noisy brain: stochastic dynamics as a principle of brain function*. Oxford University Press, Oxford
- Roustant O, Barthe F, Iooss B (2017) Poincaré inequalities on intervals-application to sensitivity analysis. *Electron J Stat* 11(2):3081–3119
- Rowland DC, Roudi Y, Moser M-B, Moser EI (2016) Ten years of grid cells. *Annu Rev Neurosci* 39:19–40
- Sakaguchi H, Shinomoto S, Kuramoto Y (1988) Phase transitions and their bifurcation analysis in a large population of active rotators with mean-field coupling. *Progress Theor Phys* 79(3):600–607
- Schmidt H, Avitabile D (2020) Bumps and oscillons in networks of spiking neurons. *Chaos* 30(3):033133
- Sznitman A-S (1984) Nonlinear reflecting diffusion process, and the propagation of chaos and fluctuations associated. *J Funct Anal* 56(3):311–336
- Sznitman A-S (1991) Topics in propagation of chaos. In: *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math*. Springer, Berlin, pp 165–251
- Touboul J (2012) Mean-field equations for stochastic firing-rate neural fields with delays: derivation and noise-induced transitions. *Physica D* 241(15):1223–1244
- Touboul J, Hermann G, Faugeras O (2012) Noise-induced behaviors in neural mean field dynamics. *SIAM J Appl Dyn Syst* 11(1):49–81
- Veltz R, Chossat P, Faugeras O (2015) On the effects on cortical spontaneous activity of the symmetries of the network of pinwheels in visual area V1. *J Math Neurosci* 5:Art. 11
- Wilson H, Cowan J (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J* 12:1–24
- Wilson H, Cowan J (1973) A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biol Cybern* 13:55–80

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.