

DTU Compute

Department of Applied Mathematics and Computer Science

**The dynamics of adaptive neuronal networks
A trip to topology and back**

Author

Simon Aertssen
s181603

Supervisor

Erik Martens
Poul Hjorth

February 1st 2020

Contents

Abstract	III
Acknowledgements	III
1 Nomenclature	IV
2 Introduction	5
3 The Theta Neuron Model	6
3.1 Canonical neuron models	6
3.2 Theta Neuron model description	6
3.3 Solutions for static currents	7
3.4 Numerical solutions	8
3.5 Frequency response	8
3.6 Phase response	9
4 Network Topologies	10
4.1 Representations and properties	10
4.2 Fixed-degree networks	10
4.3 Random / Erdős-Rényi networks	10
4.4 Scale-free networks	11
4.5 Networks of theta neurons	11
5 Mean Field Reductions	13
5.1 The Ott-Antonsen manifold	13
5.2 Simplifications for fixed-degree networks	14
5.3 Implications and challenges of the <i>MFR</i>	15
6 Investigation: Mean Field Reductions for undirected graphs	16
6.1 Directed graphs as permutations	16
6.2 Building the adjacency matrix	17
6.3 Initial conditions	17
6.4 Commutativity of complex vectors	18
6.5 Fixpoint iteration	19
6.6 A Newton-Raphson iteration for all fixpoints	19
6.7 Fixed-degree networks as a baseline	19
6.8 Results for arbitrary network topologies	20
7 Hebbian Learning and Synaptic Plasticity	23
7.1 Spike-timing dependant plasticity	23
7.2 Formulations of <i>STDP</i> as a model	24
7.3 Synaptic scaling	25
7.4 Intrinsic plasticity	26
8 Investigation: Emerging Network Topologies	27
8.1 <i>STDP</i> applied to networks of theta neurons	27
8.2 Results	30
9 Conclusion and Discussion	31

10 References	32
A Appendix	34
A.1 Transformation to the QIF model	34
A.2 Solutions to the QIF model	34
A.2.1 Solving for $I < 0$	34
A.2.2 Solving for $I = 0$	34
A.2.3 Solving for $I > 0$	34
A.3 Frequency response of the neuron models	35
A.4 Newton-Raphson root iteration	35

Abstract

Modelling is awesome.

Acknowledgements

1 Nomenclature

i, e (or \exp)	Imaginary unit. Euler's number.
$n, \deg(n)$	Network node. Degree of node n .
N	Network degree. The number of neurons in the network.
A_{ij}	Adjacency matrix. Models which neuron i is connected to neuron j and vice-versa.
$\langle k \rangle$	Average node degree in the network.
\mathbf{k}	Node degree. Vector of the in- and out-degree of a single node as $(k^{\text{in}}, k^{\text{out}})$.
$k^{\text{in}}, k^{\text{out}}$	Node degree vector of all in- and out degrees of the network.
M_k	Number of unique node degrees in the network.
$P(k), P(\mathbf{k})$	Univariate and bivariate network degree distribution.
k_{\min}, k_{\max}	Smallest and largest degree found in a network.
γ	Degree exponent of a scale-free network.
p	Probability threshold of forming a link in random networks.
c	Assortativity of the network.
$\theta(t)_i$	Phase variable function of the theta model (of neuron i).
$\mathcal{P}_n(\theta)$	Pulse shaped synaptic coupling function.
κ	Macroscopic coupling strength.
$\eta_i, I(t)_i$	Excitability threshold and input current (of neuron i).
$g(\eta \mathbf{k})$	Excitability threshold distribution with mean $\eta_0(\mathbf{k})$ and width $\sigma(\mathbf{k})$.
$Z(t)$	Kuramoto order parameter function.
$z(\mathbf{k}, t)$	Synchronisation dynamics function for nodes with degree \mathbf{k} .
$\bar{Z}(t)$	Mean field order parameter function for arbitrary networks.
$S^{\text{in}}(t)_i, S^{\text{out}}(t)_j$	Spike trains received and emitted by neuron i as a sum of delta functions in time.
K_{ij}	Synaptic connectivity matrix. Strength of the connections between neurons i and j .
Δt_{ij}	Time difference between spikes of neurons i and j .
$W(\Delta t_{ij})$	Learning window. Models the correlation between synaptic strength and spike times.
$\phi(\Delta t_{ij})$	IP learning function. Models correlation between excitability strength and spike times.
\mathbb{T}	Set of angles in $[-\pi, \pi[$.
\mathbb{K}	Set of unique degrees in a network, support of P .
\mathbb{R}	Set of real numbers.
\mathbb{C}	Set in the complex unit circle so that $\mathbb{C} = \{z : z \leq 1\}$.
$F(v), F^{-1}(v)$	Random permutation and inverse permutation of the elements of a vector v .

2 Introduction

In 2013, one of the largest scientific projects ever funded by the European Union was launched. With the Human Brain Project [1], scientists and researchers aimed to reconstruct the human brain through supercomputer-based models and to advance neuroscience, medicine, and computing. Across the globe different fields of science are drawing inspiration from the human brain, through different approaches.

One such approach is to model the behaviour of biological neurons and to quantify the information processes in the brain from stimuli from the senses or from electrical and chemical processes in the body. A given neuron receives hundreds of impulses in the form of neurotransmitters, almost exclusively on its dendrites and cell body. These stimuli add up to an excitatory or inhibitory influence on the membrane potential of the neuron, so that the potential spikes when excitation is higher than an internal threshold. At this point, the neuron releases its own neurotransmitter and joins the interneuronal communication [2]. The neuron dynamics are largely captured by this spiking behaviour, on which most efforts have been concentrated. In 1952, Hodgkin and Huxley described a mathematical model for the action potentials in neurons, using a set of nonlinear differential equations that approximates the electrical characteristics of the neuron elements. In 1963 the authors were awarded the Nobel Prize in Physiology or Medicine [3] for their work.

As the human brain contains more than 100 billion neurons [4] it is unfeasible to study complex models at this scale. The topology of neuronal networks displays traits of small-worldness, wiring optimisation, and heterogeneous degree distributions [5], for which it is difficult to pin down one type of network architecture. Through the mean-field reduction (*MFR*) proposed in [6] one can reduce a large network of indistinguishable neurons to a low-dimensional dynamical system, described by the attraction of a mean-field variable to a reduced manifold. In this work we will study the *MFR* of different types of networks of coupled Theta neurons using the generalisations found in [7].

Neurons communicate through *synapses* with electrical and chemical signals, in the form of action potentials and neurotransmitters respectively. We will speak of the presynaptic neuron as the neuron that sends a signal and of the postsynaptic neuron as the neuron that receives a signal. When the membrane voltage of a presynaptic neuron reaches an internal threshold, the neuron *spikes* (or *fires*) and an electrical signal travels down the neuron axons [2]. At the synapse, the electrical signal is converted into a chemical signal in the form of a neurotransmitter release of the presynaptic neuron, upon which the postsynaptic neuron receives the neurotransmitters and constructs its own electrical signal [8]. Most neurons in the central nervous system use either the excitatory neurotransmitter glutamate (AMPA or NMDA) or the inhibitory neurotransmitter GABA [9, 10].

The process that allows neurons to adjust the strength of their synapses is called *synaptic plasticity*. This makes neurons more susceptible to each others behaviour, and results in increased (or inhibited) synchronisation of brain waves. Using Hebb's postulate [11], on the behaviour of the neuronal network to form new or strengthen connections through the synaptic strength, one can quantify and model those changes to the network topology.

The work presented here is thus two-fold: we study the dynamics of pulse-coupled networks *on* networks, and the dynamics *of* such networks when they evolve over time.

3 The Theta Neuron Model

3.1 Canonical neuron models

A number of neuron model families have been identified, and often there exists a continuous change of variables from models of the same family into a *canonical* model that can represent the whole family [12]. As the transformation is not required to be invertible, we can study the universal neurocomputational properties of the family in a low dimensional model. It was Hodgkin [13] who classified neurons into two types based on their excitability, upon experimenting with the electrical stimulation of cells. Class 1 models begin to spike at an arbitrarily slow rate, and the spiking frequency increases when the applied current is increased. Class 2 models spike as soon as their internal threshold is exceeded and the spiking frequency stays relatively constant within a certain frequency band [12].

3.2 Theta Neuron model description

In [14], a Class 1 canonical phase model was proposed:

$$\dot{\theta} = (1 - \cos \theta) + (1 + \cos \theta) \cdot I \quad \theta \in \mathbb{T} \quad (1)$$

with I a bifurcation parameter on the supplied current. We can visualise the dynamics on the unit circle, like in Figure 1. The neuron produces a spike when θ surpasses π , upon which $\theta \leftarrow -\pi$.



Figure 1: SNIC bifurcation of the theta neuron model. A spike occurs when $\theta = \pi$. For $I < 0$, the neuron is in a rest state but *excitable*. For $I > 0$, $\dot{\theta} > 0$ so that θ moves continuously around the circle and we can observe *periodic* sustained spiking. The saddle-node bifurcation occurs at $I = 0$, so that θ will spike when it is larger than 0.

We can recognise the features of the class 1 model in Figure 2. This makes (1) the normal form of the *saddle-node-on-invariant-circle (SNIC)* bifurcation [15].

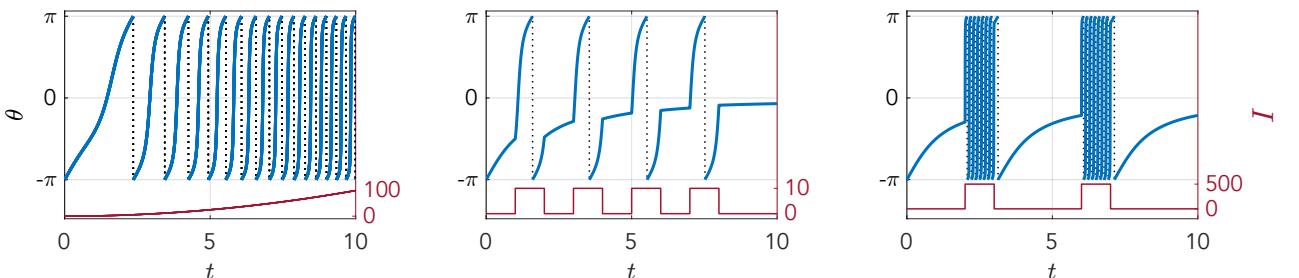


Figure 2: Properties of the theta neuron model, with solutions of (1) in blue, spikes marked in dotted lines, and the current I in red. Left: the spike frequency of θ increases as I is increased over time, which is the distinguishing feature of class 1 canonical models. Middle: spikes occur within a finite time period when $I > 0$ and within infinite time when $I = 0$. Right: when I is large, the neuron *bursts*.

Equilibria only exist for the *excitable* regime $I < 0$:

$$\begin{aligned} \dot{\theta} &= 1 - \cos \theta + I + I \cdot \cos \theta = (I + 1) + (I - 1) \cdot \cos \theta \\ \theta_{1,2}^* &= \pm \arccos \left(\frac{I+1}{1-I} \right) + 2\pi n \end{aligned}$$

We can find the stability of the equilibria through:

$$\frac{d}{d\theta}((1 - \cos \theta) + (1 + \cos \theta) \cdot I) = \sin \theta - \sin \theta \cdot I = (1 - I) \cdot \sin \theta$$

In the equilibria this yields:

$$\frac{d}{d\theta}(\theta_{1,2}^*) = \pm(1 - I) \cdot \sqrt{1 - \frac{I+1}{1-I}} = \pm(1 - I) \cdot \frac{2\sqrt{-I}}{1-I} = \pm 2\sqrt{-I}$$

This yields a stable equilibrium point for θ_1^* and an unstable for θ_2^* . This means that as θ gets perturbed above θ_2^* , a spike occurs and θ converges to θ_1^* . This is demonstrated in Figure 3.

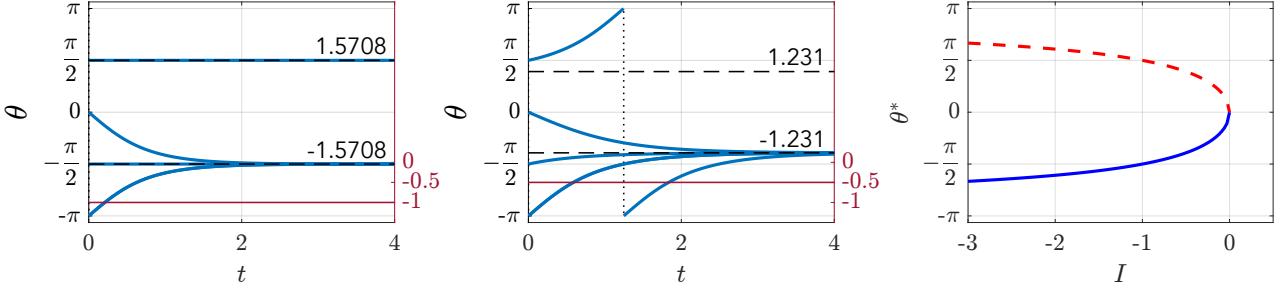


Figure 3: Equilibria θ^* for different values of I . Left: $I = -1$ yields $\theta_{1,2}^* = \pm \frac{\pi}{2}$, one of the simulations is started exactly on the unstable equilibrium. Middle: $I = -0.5$. Right: bifurcation diagram of the SNIC bifurcation, with the stable equilibria in blue, and the unstable in red.

3.3 Solutions for static currents

Gaining insight into (1) is hard, due to the difficulty of finding an analytical solution. However, it has been noted that there exists a simple transformation which yields (see A.1):

$$V \equiv \tan\left(\frac{\theta}{2}\right) \quad (2)$$

$$\dot{V} = V^2 + I \quad (3)$$

This model is called the *Quadratic Integrate and Fire model* (QIF). (3) models the membrane potential of a neuron, which spikes to $= \infty$ when the neuron spikes and is reset at $-\infty$. The transformation (2) is continuous between spikes, so insights from a solution for V can be transformed directly. The equilibria of the QIF model are simply $\pm\sqrt{I}$ so that we can express $\theta_{1,2}^* = 2 \cdot \arctan(\mp\sqrt{I})$, from [16].

The solution for the excitable regime $I < 0$ is :

$$V(t) = \frac{2\sqrt{-I}}{1 - e^{2t\sqrt{-I}}} - \sqrt{-I} \quad (4)$$

The solution at the bifurcation $I = 0$ is :

$$V(t) = \frac{-1}{t} \quad (5)$$

The solution for the periodic regime $I > 0$ is :

$$V(t) = -\sqrt{I} \cdot \cot(t\sqrt{I}) \quad (6)$$

These equations assume that at $t = 0$ a spike has occurred. The steps required to find (4)-(6) are described in A.2. Solutions for θ are found by taking the inverse of the transformation (2).

If the QIF model is so much simpler, then why bother using the Theta Model? Simulating the QIF model requires an artificial reset threshold, because we cannot expect a computer to represent infinity that easily. Finite thresholds make the analytical solutions more difficult and convoluted. By using the Theta Model we do not have to worry about these problems.

3.4 Numerical solutions

When I is not static, we need to revert to numerical solutions. For this work, a fixed-step 4-stage Runge-Kutta method (Dormand-Prince 45) was implemented to numerically solve all differential equations. A fixed-step algorithm makes it possible to finely tune the large memory demand of the systems presented in this work.

3.5 Frequency response

As we already saw in Figure 2, an increasing current increases the spiking frequency. We can compute this relationship by measuring how long it takes for V to reach a spike: we solve (6) for t at $V(t) = +\infty$ in A.3. This yields the oscillation period $T = \frac{\pi}{\sqrt{I}}$ which we can see in Figure 4. We know that when $\theta > \theta_2^*$ a spike occurs in the excitable regime, or in any case in the periodic regime. But the time that it takes to reach the spike can be arbitrarily long, depending on how far we are over θ_2^* . So, spikes will occur, but after a delay that is dependant on the stimulus. Explicitly, if we perturb $\theta(0) = \theta_2^* + \varepsilon$ we obtain from [16]:

$$T_{\text{spike}} = \frac{-\tanh^{-1}\left(1 + \frac{\varepsilon}{\sqrt{I}}\right)}{\sqrt{I}}$$

The delay to the spike blows up as $\varepsilon \rightarrow 0$ so that spikes may occur after a very large delay.

In most of our future work, I will not be a static current. We ask ourselves: how sensitively does T depend on I when I is perturbed? We can measure this as a *relative* perturbation using dI/I and dT/T [2] :

$$\left| \frac{dT}{dI} \frac{I}{T} \right| = \left| \frac{dT/T}{dI/I} \right| = \left| -\frac{\pi}{2} \left(\frac{1}{\sqrt{I}} \right)^3 \frac{I}{T} \right| = \left| \frac{\pi}{2} \left(\frac{T}{\pi} \right)^3 \frac{I}{T} \right| = \frac{1}{2} \left| \left(\frac{T}{\pi} \right)^2 \cdot \left(\frac{\pi}{T} \right)^2 \right| = \frac{1}{2}$$

Hence, a 1% change in I will result in a 0.5 % change in the period.

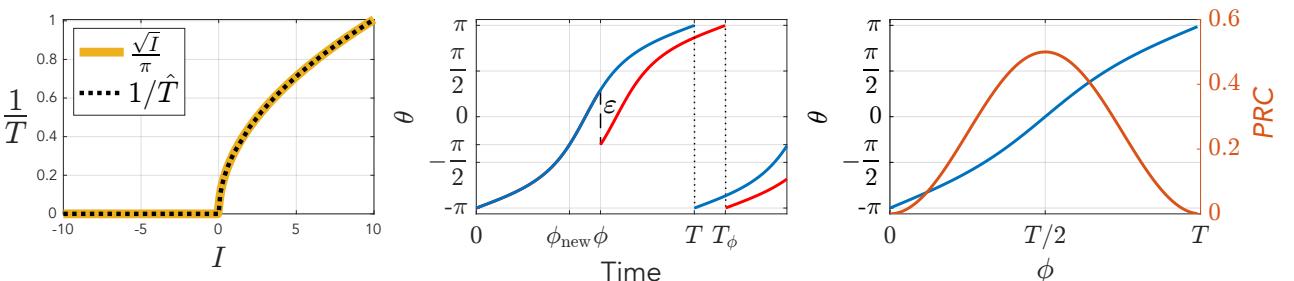


Figure 4: Response of the theta model to bifurcations on the frequency and the phase. Left: Frequency response of the theta model. For $I \leq 0$ the spike period is infinite, which is why we see the solutions to (1) approach $\theta = 0$ for $I = 0$. Middle: a bifurcation ε at time ϕ perturbs $\theta(t)$ (in blue) which results in a delayed spike (trajectory in red). Right: The PRC, (8) in red, with a solution for θ (in blue), to show when the model is the most susceptible to bifurcations.

3.6 Phase response

Perturbations on the period can also be understood from the perspective of the phase. Changes to the phase θ can delay or advance the event of a spike, and in general this depends on exactly when the stimulus occurs. The phase response curve (*PRC*) gives us exactly that relation [16, 17]. Let us define $\phi \in [0, T[$, which represents the time since the last event of a spike. When we add a small bifurcation $\varepsilon < 0$ to θ at time ϕ , a spike will occur at T_ϕ , and we have that $\theta(\phi_{\text{new}}) = \theta(\phi) + \varepsilon$. The time to the new spike is $T_\phi = T + (\phi - \phi_{\text{new}})$. The *PRC* can then be defined as:

$$PRC(\phi) = T_\phi - T \quad (7)$$

This process has been visualised in Figure 4, after [17]. For infinitesimally small perturbations to the phase, we can find the *PRC* as the *adjoint* of the solution, [16], as:

$$PRC(\phi) = \frac{1}{dV(\phi)/d\phi} = \frac{1}{2\sqrt{I}} \left(1 - \cos\left(\frac{2\pi}{T}\phi\right) \right) \quad (8)$$

We can use $\phi \in [0, T[$ and $\theta \in \mathbb{T}$ to see that (8) can be expressed as:

$$PRC(\theta) = \frac{1}{2\sqrt{I}} (1 + \cos \theta) \quad (9)$$

which is the magnitude with which I excites the model (1), [18]. Analysis of the *PRC* thus allows us to study how the bifurcation of θ with magnitude I occurs.

The *PRC* is always positive, which indicates that a positive bifurcation will advance the time of the spike, and vice versa. This has also been reported as a distinguishing feature of Class 1 models, [18].

4 Network Topologies

Networks consists of *nodes* n_j , $j \leq N$ connected by *links*. They arise in any context where objects are *related* to each other. In this section, we will look at the notation that is needed to represent a network, looking from different sides.

4.1 Representations and properties

We represent a finite network through the adjacency matrix: $A_{ij} = 1$ if there exists a relation from node j to node i and 0 otherwise. This means that A_{ij} can be *undirected* (symmetric) or *directed*. If we think of the relations between guests at a party, then the social network is directed, as people might not know each other mutually. However, the network of people having shaken hands is symmetric. Self-links are an edge-case that depends on the context, as one generally does not shake hands with himself.

The *degree* \mathbf{k} of a node is a two-vector of the number of links coming in to and going out of the node, $(k^{\text{in}}, k^{\text{out}})$. From A_{ij} we can compute the in- and out-degree vectors, which show how many links a node has coming in and out:

$$\mathbf{k}_i^{\text{in}} = \sum_{j=1}^N A_{ij} \quad \mathbf{k}_j^{\text{out}} = \sum_{i=1}^N A_{ij} \quad \deg(n_j) = \mathbf{k}_j = (\mathbf{k}_j^{\text{in}}, \mathbf{k}_j^{\text{out}}) \in \mathbb{K} \subset \mathbb{N} \quad (10)$$

The distribution of \mathbf{k}^{in} and \mathbf{k}^{out} is the most defining property of the network:

$$(\mathbf{k}^{\text{in}}, \mathbf{k}^{\text{out}}) \sim P(\deg(n) = \mathbf{k}) \quad (11)$$

The support of P is the set of unique degrees \mathbb{K} with cardinality M_k , which consists of integers. The average degree of the network is then:

$$\langle k \rangle = \frac{1}{N} \sum_{i,j=1}^N A_{ij} = \frac{1}{N} \sum_{i=1}^N \mathbf{k}_i^{\text{in}} = \frac{1}{N} \sum_{j=1}^N \mathbf{k}_j^{\text{out}} \quad (12)$$

For symmetric networks, $\mathbf{k}^{\text{in}} = \mathbf{k}^{\text{out}}$, so that P is really a univariate distribution. In this case, much of the coming analysis is heavily simplified, so we will start with univariate distributions.

4.2 Fixed-degree networks

A network consists of nodes, connected by links. The most simple network is one where all the nodes are connected, and so all nodes have a degree of N . In general, we can make networks where all nodes have the same degree, $\langle k \rangle$:

$$P(k) = \begin{cases} \langle k \rangle & \text{if } k = \langle k \rangle \\ 0 & \text{otherwise} \end{cases} \quad \mathbb{K} = \{\langle k \rangle\} \quad (13)$$

We will refer to these networks as fixed-degree networks. When $\langle k \rangle = N$, all nodes are self-coupled and connected to all other nodes in the network, so we speak of a *fully connected* network.

4.3 Random / Erdös-Rényi networks

In 1959 Erdös and Rényi published their work on random graphs [19], where links are established if a random uniformly distributed number is higher than a threshold p . The degrees follow a binomial distribution:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \quad \mathbb{K} = [0, N] \quad (14)$$

with a mean $\mu = p(N - 1)$ and standard deviation $\sigma = \mu(1 - p)$. For networks where $\langle k \rangle \ll N$, the network can be well approximated by a Poisson distribution:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \quad \mathbb{K} = [0, N] \quad (15)$$

with a mean $\mu = \langle k \rangle$ and standard deviation $\sigma = \sqrt{\langle k \rangle}$. Both (14) and (15) describe similar quantities, but the latter is used more often due to its analytical simplicity [20].

4.4 Scale-free networks

What we can often observe in nature is the preferential attachment to nodes with a high degree [5]: the rich or famous tend to get more rich or famous. This trait is also described as the 80/20 rule by Pareto. Networks with this property consist of a small number of highly connected nodes, and a large number of low degree nodes. We can represent this with a power law distribution:

$$P(k) = Ak^{-\gamma} \quad \mathbb{K} = [k_{\min}, k_{\max}] \quad (16)$$

with A is a constant so that $\sum_{k=1}^{\infty} P(k) = 1$. We can also see that $A \sum_{k=1}^{\infty} k^{-\gamma} = 1$ so that $A = \sum_{k=1}^{\infty} k^{-\gamma} = 1/\zeta(\gamma)$, the Riemann Zéta function [20].

Networks with a distribution like (16) are called *scale-free* networks, as they lack an internal scale to represent the magnitude of the network: we can observe (16) on different scales like the probability of two Hollywood actors appearing in a movie, or the connections between web pages on the internet [21]. One description that comes close is the *natural cutoff* k_{\max} , the expected degree of the largest degree in the network. As we only expect the largest hub to be the only hub in the domain $[k_{\max}, +\infty]$:

$$\int_{k_{\max}}^{\infty} P(k) dk = \frac{1}{N}$$

For (16) this results in:

$$k_{\max} = k_{\min} \cdot N^{\frac{1}{\gamma-1}} \quad (17)$$

which shows that there might be large differences in size between the nodes.

There are constraints on γ to yield a scale-free network. When $0 < \gamma < 2$ the largest hub grows faster than N , so once its degree exceeds $N - 1$ there are no more new nodes to connect to and the network will not be able to grow according to (16). A rigorous proof is given in [22]. For $\gamma = 2$, the system grows linearly, as we can see in (17). When $2 < \gamma \leq 3$ we find the most scale-free networks, as for $\gamma > 3$ hubs are not sufficiently large and numerous to have much influence on the network [20].

4.5 Networks of theta neurons

We can easily extend the model to networks of neurons:

$$\dot{\theta}_i = (1 - \cos \theta_i) + (1 + \cos \theta_i) \cdot [\eta_i + I_i(t)] \quad \theta_i \in \mathbb{T}^N \quad (18)$$

$$I_i(t) = \frac{\kappa}{\langle k \rangle} \sum_{j=1}^N A_{ij} \cdot \mathcal{P}_n(\theta_j) \quad (19)$$

where the excitability η_i allows neuron i to adjust in which regime it is situated, and $\eta_i \sim g(\eta|\eta_0, \sigma)$. κ is the *synaptic* or *coupling* strength, and $\mathcal{P}_n(\theta) = a_n(1 - \cos \theta)^n$ models synaptic coupling by a pulse-shaped signal, emitted when a neuron fires. As discussed in Chapter 2, there are conversions

from the action potential to a neurotransmitter and back, but this process will be captured by using only \mathcal{P} as the action potential and κ as the "efficiency" of the conversions. n models the sharpness of the pulse, and a_n is a normalisation constant so that $\int_{\mathbb{T}} \mathcal{P}_n d\theta = 2\pi$. We will take $n = 2$ from here, as in [7, 15, 23].

In (18) we see everything come together: changes to the phase θ_i are induced by $\dot{\theta}_i$ which in turn depends on the bifurcation of θ with magnitude I_i which depends on all neurons in the network.

Studying a set of differential equations like (19) is not feasible, as we are quickly approaching thousands of neurons. And in the end, the dynamics of a single neuron are not of interest. Instead, we wish to capture and study how the network behaves as a whole. One aspect, synchrony, can be captured by the Kuramoto order parameter:

$$Z(t) = \frac{1}{N} \sum_{j=1}^N e^{i\theta_j} \quad Z(t) \in \mathbb{C} \quad (20)$$

Z is a complex variable, consisting of a radius $r = |Z|$ and argument $\psi = \arg(Z)$, so that $Z(t) = r(t)e^{i\psi(t)}$. When all phases are uniformly distributed across the unit circle \mathbb{T} , then $|Z| = 0$, resulting in a network with no synchronisation. When all phases are exactly the same, $|Z| = 1$, the network is fully synchronised. (20) describes the *mean-field* of the network, a simpler model that describes the average behaviour of the whole network. Analysis is simply conducted either on $|Z(t)|$ versus time, or in the complex unit circle as $\text{Re}(Z(t))$ versus $\text{Im}(Z(t))$.

Different works on the dynamics of (20) have been published [15, 23], and we will build on that analysis in the following chapters.

5 Mean Field Reductions

5.1 The Ott-Antonsen manifold

The *mean-field reduction* (*MFR*) is the theory that predicts the dynamics of the order parameter (20). In [6, 24, 25] such a method was published for fully connected networks of indistinguishable oscillators. In [26] the authors extended their work to include networks with arbitrary degree distributions, applied to the Kuramoto model. Later this analysis was extended to networks of the Theta Neuron model [7].

We will now consider the limit $N \gg 1$ and formulate an exact *MFR* for different types of networks, following the method in [7]. To simplify notation, the authors incorporate the network size in to P so that $\sum_{\mathbf{k} \in \mathbb{K}} P(\mathbf{k}) = N$. To specify the probability of a link from a node of degree \mathbf{k}' to one of degree \mathbf{k} we can define an assortativity function:

$$a(\mathbf{k}_j \rightarrow \mathbf{k}_i) = 0 \leq \frac{k_j^{\text{out}'} k_i^{\text{in}}}{N \langle k \rangle} \leq 1 \quad (21)$$

where we have chosen a neutral assortativity [7]. (21) is constrained so that the number of links in the network, $N \langle k \rangle$, remains constant [26]:

$$\sum_{\mathbf{k}' \in \mathbb{K}} \sum_{\mathbf{k} \in \mathbb{K}} P(\mathbf{k}') a(\mathbf{k}' \rightarrow \mathbf{k}) P(\mathbf{k}) = N \langle k \rangle$$

We can now assume that the state of all neurons can be represented by a probability density function $f(\theta, \eta | \mathbf{k}, t)$. Hence, the marginal distribution

$$\int_{\mathbb{R}} \int_{\mathbb{T}} f(\vartheta, \eta' | \mathbf{k}, t) d\vartheta d\eta'$$

gives the fraction of nodes of degree \mathbf{k} with a phase in \mathbb{T} at time t . Also, we assume η_i do not change over time, so that $\int_{\mathbb{T}} f(\vartheta, \eta' | \mathbf{k}, t) d\vartheta = g(\eta | \mathbf{k})$ gives the excitability distribution.

To describe the global synchronisation of the network of theta neurons (18) we have introduced the order parameter (20). It is now hypothesized that $Z(t)$ can be approximated by a mean-field order parameter, defined by the continuum limit:

$$\bar{Z}(t) = \frac{1}{N} \sum_{\mathbf{k} \in \mathbb{K}} P(\mathbf{k}) \int_{\mathbb{R}} \int_{\mathbb{T}} f(\vartheta, \eta' | \mathbf{k}, t) e^{i\vartheta} d\vartheta d\eta' \quad (22)$$

Here, f is constrained by a continuity equation, as the number of oscillators is conserved:

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial \theta} (v_\theta f) = 0 \quad (23)$$

with v_θ a continuum version of (18):

$$v_\theta = (1 - \cos \theta) + (1 + \cos \theta)[\eta + I(\mathbf{k}, t)]$$

$$I(\mathbf{k}, t) = \frac{\kappa}{\langle k \rangle} \sum_{\mathbf{k}' \in \mathbb{K}} P(\mathbf{k}') a(\mathbf{k}' \rightarrow \mathbf{k}) \times \left[\int_{\mathbb{R}} \int_{\mathbb{T}} f(\vartheta, \eta' | \mathbf{k}', t) a_n (1 - \cos \vartheta)^n d\vartheta d\eta' \right]$$

In [6] it is shown that there exists a manifold of invariant probability densities for the continuity equation. The exact *MFR* is obtained by expanding f as a Fourier series, and expanding the pulse P_n using the binomial theorem. When assuming η_i is distributed according to a Lorenz distribution:

$$g(\eta | \mathbf{k}) = \frac{1}{\pi} \frac{\sigma(\mathbf{k})}{(\eta - \eta_0(\mathbf{k}))^2 + \sigma(\mathbf{k})^2} \quad (24)$$

the set of reduced equations then takes a particularly simple form, as (22) can be evaluated the poles of g using the Cauchy residue theorem for the integration of complex variables and find a closed form expression. We can now capture the dynamics by $z(\mathbf{k}, t)$, the mean-field variable for nodes of degree \mathbf{k} :

$$\begin{aligned} \frac{\partial z(\mathbf{k}, t)}{\partial t} &= -i \frac{(z(\mathbf{k}, t) - 1)^2}{2} + \frac{(z(\mathbf{k}, t) + 1)^2}{2} \cdot I(\mathbf{k}, t) \quad z(\mathbf{k}, t) \in \mathbb{C}^{M_{\mathbf{k}}} \\ I(\mathbf{k}, t) &= -\sigma(\mathbf{k}) + i\eta_0(\mathbf{k}) + iH_2(\mathbf{k}, t) \\ H_2(\mathbf{k}, t) &= \frac{\kappa}{\langle k \rangle} \sum_{\mathbf{k}' \in \mathbb{K}} P(\mathbf{k}') a(\mathbf{k}' \rightarrow \mathbf{k}) \cdot \left(1 + \frac{z(\mathbf{k}', t)^2 + (z(\mathbf{k}', t)^c)^2}{6} - \frac{4}{3} \operatorname{Re}(z(\mathbf{k}', t)) \right) \end{aligned} \quad (25)$$

with z^c the complex conjugate. H is a legacy term and has been computed in [23]. The mean-field order parameter can now be expressed in terms of $z(\mathbf{k}, t)$. Using the constraints on f and g we can now solve (22) as:

$$\bar{Z}(t) = \frac{1}{N} \sum_{\mathbf{k}} P(\mathbf{k}) z(\mathbf{k}, t) \quad \bar{Z}(t) \in \mathbb{C} \quad (26)$$

which clearly reflects the network architecture through P . We have now formulated the evolution on the invariant manifold by a reduced set of ordinary differential equations. The MFR is computationally efficient, and in [7] many methods for improving this efficiency further are treated.

5.2 Simplifications for fixed-degree networks

In the case of a fixed-degree network, every node has $\deg(\theta_i) = (\langle k \rangle, \langle k \rangle)$ so:

$$\frac{1}{\langle k \rangle} \sum_{\mathbf{k}' \in \mathbb{K}} P(\mathbf{k}') a(\mathbf{k}' \rightarrow \mathbf{k}) = \frac{1}{\langle k \rangle} N \left(\frac{\langle k \rangle \langle k \rangle}{N \langle k \rangle} \right) = 1$$

This is an identical formulation as in [15] and [23] and for any fixed-degree network, eqs. (25) and (26) reduce to a single complex differential equation:

$$\dot{Z}(t) = -i \frac{(Z - 1)^2}{2} + \frac{(Z + 1)^2}{2} \cdot \left(-\sigma + i\eta_0 + i\kappa \cdot \left(1 + \frac{Z^2 + (Z^c)^2}{6} - \frac{4}{3} \operatorname{Re}(Z) \right) \right) \quad (27)$$

This is an identical formulation as in [15] and [23] and for any fixed-degree network, and as it is a complex-valued function, the reduced system is two-dimensional, with three bifurcation parameters η_0, σ and κ . We will start our analysis with (27).

Three distinct macroscopic states can be identified. In the partially synchronous rest state (*PSR*) we can observe in Figure 5a, $Z(t)$ settles onto a stable node. Most neurons can be found in a resting state $\eta_0 + \sigma \lesssim 0$, and inhibit one another through $\kappa < 0$. Most neurons are therefore inactive, though some spiking neurons from the tail of g are present but have a negligible effect.

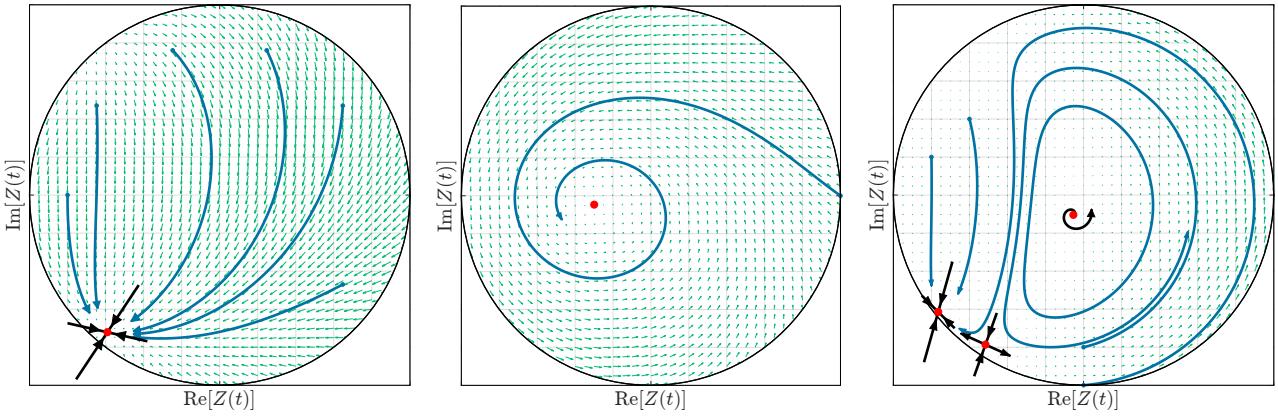
In Figure 5b we can observe a partially synchronous spiking (*PSS*) state, where we can see how $Z(t)$ settles onto a stable focus. This happens predominantly when $\eta_0 - \sigma \gtrsim 0$ and most neurons inherently spike, with the coupling being either excitatory or weakly inhibitory. Although most neurons are active, the network is partially synchronous and organized such that phase cancellation occurs by continuous spiking among the neurons.

Lastly, in the collective periodic wave state (*CPW*) we can observe a limit cycle of the mean field, in Figure 5c. Most neurons are active and inhibitory: $\eta_0 > 0$ and $k < 0$. The collective

oscillation emerges from the interplay between the neurons' inherent tendency to spike and the strong suppressive network interaction. *CPW* states are mediated through Hopf bifurcations and homoclinic bifurcations of $Z(t)$. We can also see the occurrence of a saddle-node bifurcation in the lower hand corner, for low σ . We will continue to study the *CPW* due to their interesting properties.

A more detailed discussion of the different regimes and bifurcations can be found in [15].

QUESTION: I have worked on the analysis presented in [15] to make my own figures on the bifurcations. However, I cannot seem to get them plotted. Should I continue on this?



(a) PSR state for $\eta_0 = -0.9, \sigma = 0.8$ and $\kappa = -2$. The mean field settles onto a stable node.
(b) PSS state for $\eta_0 = 0.5, \sigma = 0.7$ and $\kappa = 2$. The mean field settles onto a stable focus.
(c) CPW state for $\eta_0 = 10.75, \sigma = 0.5$ and $\kappa = -9$. The mean field settles onto a stable limit cycle.

Figure 5: Three macroscopic states observed in the *MFR* inside the imaginary unit circle $|Z(t)| = 1$. Green arrows mark the phase space vector field and blue trails mark solution curves. Red points indicate equilibrium points, with black arrows marking the direction of the eigenvectors in that point, scaled according to the magnitude of the corresponding eigenvalues.

5.3 Implications and challenges of the *MFR*

The advantages of using the *MFR* can be found in the number of equations we now have left to investigate. As there are M_k equations in (25), instead of N equations for N neurons, the reduction becomes more and more efficient for larger networks. As we have seen in (27) this yields a single equation for a fixed-degree network.

While the *MFR* gives us the opportunity to use any arbitrary univariate distribution $P(k)$ for undirected, symmetric networks or any bivariate distribution $P(\mathbf{k})$ for directed, asymmetric networks, none of the publications on the *MFR* have treated directed networks. The challenge is that now the support \mathbb{K} is a much larger set, as $\mathbb{K} = \mathbf{k}^{\text{out}} \times \mathbf{k}^{\text{in}}$. For example, the scale-free distribution (16) has $M_k = k_{\max} - k_{\min}$ number of degrees in its support. In [7], $M_k = 1250$. For 10.000 neurons, that is a reduction of 12,5%. When we wish to extend (16) to a bivariate distribution, M_k grows to $(k_{\max} - k_{\min})$. A bivariate distribution would need about 1.56×10^6 equations for 10.000 neurons. It is not feasible to solve this many equations at once.

6 Investigation: Mean Field Reductions for undirected graphs

We will now investigate the questions that were raised after deriving the *MFR*. How do we deal with the curse of dimensionality concerning the degree distribution? But there are also other questions to be answered. If the synchronisation dynamics of the network of Theta neurons (18) can be predicted by the Ott-Antonsen reductions eqs. (25) to (27), then it can also be measured by the order parameter (20). These systems describe the same quantity, but how can we show that?

6.1 Directed graphs as permutations

So how can we use the *MFR* efficiently when the network is a directed graph with an asymmetrical adjacency matrix? Let's investigate.

- Sampling k^{in} and k^{out} from a bivariate distribution requires us to find the marginal distribution of P for k^{in} , sampling k_i^{in} , and then sampling k_j^{out} from P while keeping k_i^{in} fixed. This is a cumbersome process. And what relation would there be between k^{in} and k^{out} ?
- However, if we assume that the marginal distributions for k^{in} and k^{out} are independent, there is a simplification to be found. We can even assume that the two marginal distributions are identical univariate distributions.
- Hence, we can sample k^{in} from a univariate distribution and find $k^{\text{out}} = F(k^{\text{in}})$ so that the total number of links remains constant.

This hypothesis can be tested: we assume that $P(\mathbf{k}) = P(k^{\text{in}}) \cdot P(k^{\text{out}})$ so that P consists of two identical and independent distributions, given by the distributions presented in Chapter 4. Then, we sample $k^{\text{in}} \sim P(k^{\text{in}})$ and perform a permutation to find all node degrees k_j . The surface given by P and the histogram of k_j have been plotted in Figure 6. As we can see, the variates follow the distribution well.

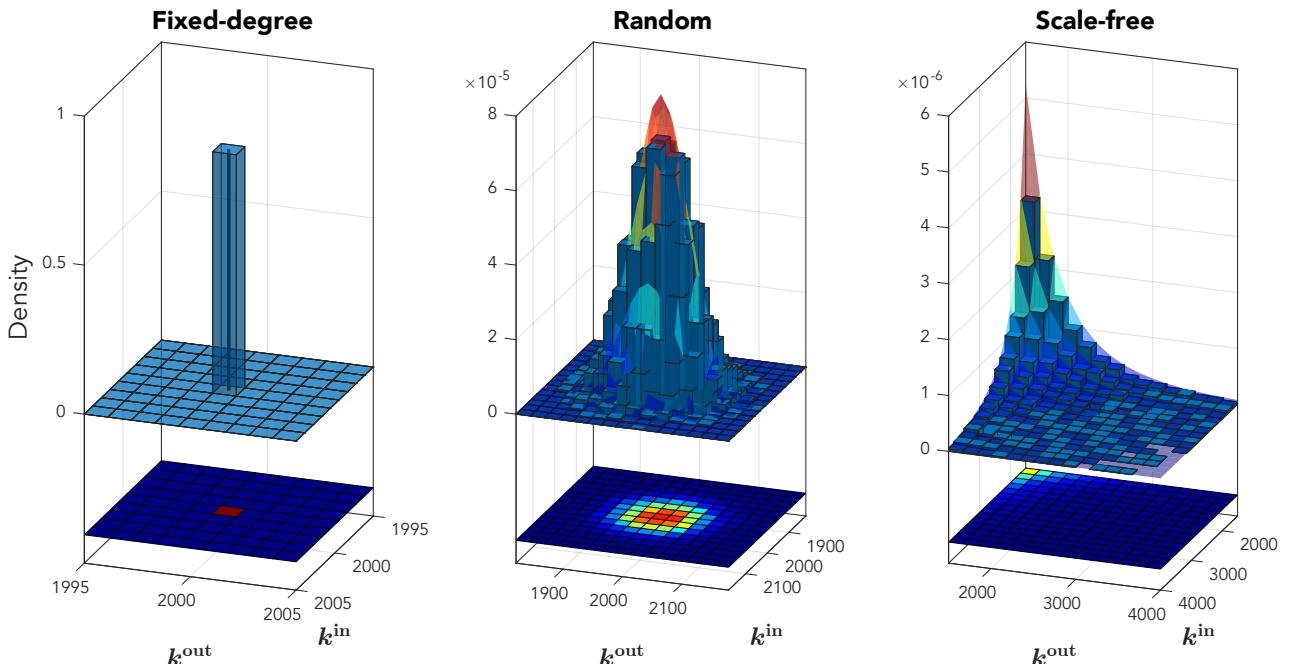


Figure 6: Bivariate distributions for different network topologies, using 10^4 number of samples. The surface given by $P(\mathbf{k})$ is well approximated by the histogram of variates sampled from a univariate distribution. $\langle k \rangle = 2 \times 10^3$ for all topologies, $p \approx 0.2$ for the random network and $\gamma = 4.3$ for the scale-free network.

Hence, we can use univariate distributions in our simulations of the Ott-Antonsen Mean Field (26).

6.2 Building the adjacency matrix

If we want to simulate the network of theta neurons (19) we need to construct the adjacency matrix. We can find an exact solution for A given the degree vectors in (11). A_{ij} represents a directed graph, but $A_{ij} \neq A_{ji}$ is not a necessary condition. For the elements of A_{ij} we need to find N^2 number of variables. We have the following constraints:

1. The column- and row-sums of A_{ij} must be equal to \mathbf{k}^{in} and \mathbf{k}^{out} , see (10). $2N$ constraints.
2. Self-coupling is mandatory: $A_{ii} = 1$. N constraints.
3. The total number of links is constant: $\sum_{i=1}^N \mathbf{k}_i^{\text{in}} \equiv \sum_{j=1}^N \mathbf{k}_j^{\text{out}} \equiv \sum_{i,j=1}^N A_{ij}$. 1 constraint.

This means that there are $N^2 - (3N + 1)$ variables to find. Once a solution has been found, A_{ij} can be switched with element A_{ic} if $A_{ij} \neq A_{ic}$ and A_{rj} with A_{rc} , which yields a new feasible solution. The number of switches one can make is high, and therefore we can simply try a stochastic approach to obtain A :

1. Choose a random row $i \in [1, N]$. $A_{i,i} = 1$, so we need $m = \mathbf{k}_i^{\text{in}} - 1$ elements that are 1.
2. Perform $F(\mathbf{k}_j^{\text{out}}, j \neq i)$ and therein find the indices ℓ of the m first largest elements.
3. Set $A_{il} = 1 \forall l \in F^{-1}(\ell)$.

Algorithms that find the largest value in a vector start from the first or the last element. The permutation allows us to find different maxima every time by shuffling the vector.

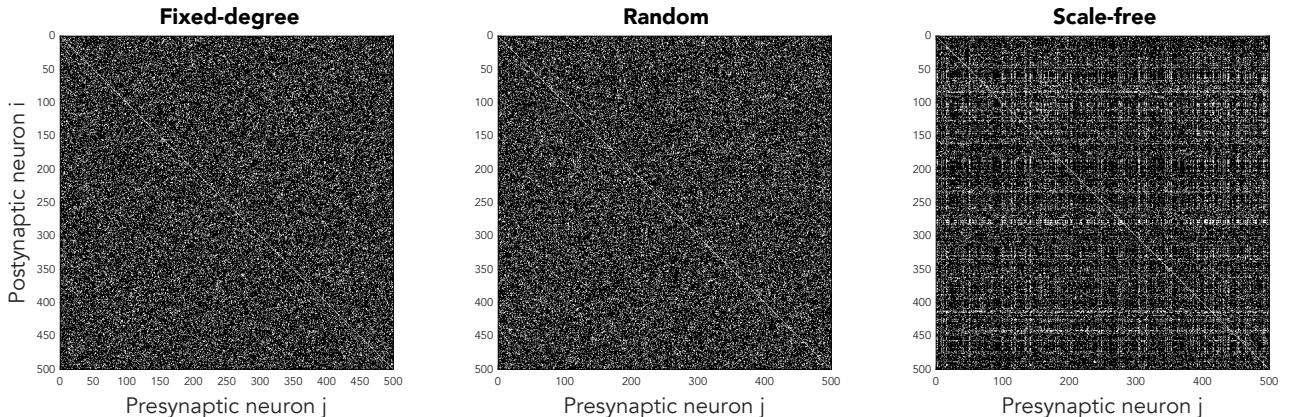


Figure 7: Adjacency matrices for different types of networks with $N = 500$ and $\langle k \rangle = 100$. We can see how the fixed-degree network is quite homogeneous, while the random network shows some more clustering. The scale-free network has a low number of nodes with a very high degree, which is why we see vertical and horizontal stripes in the adjacency matrix.

6.3 Initial conditions

As the systems in eqs. (20), (26) and (27) describe the same dynamics for fully connected networks, it is important to be able to transform initial conditions between systems. If we expect their behaviour to be the same, then we need to test that by starting from the same point in time. Hence we can test whether *macroscopically* we can find the same equilibria, but we can also test *microscopically* whether the systems arrive at those points at the same time.

As we can only really study behaviour of Z and \bar{Z} in the complex unit circle, the most important relation we need to find is the transformation from \mathbb{C} to \mathbb{T}^N and from \mathbb{C} to \mathbb{C}^{M_k} , which yield the phase angles $\theta(0)_i$ and the degree dynamics $z(\mathbf{k}, 0)$ from $Z(0)$ and $\bar{Z}(0)$ respectively.

Let us start with the simplest transformation. Given an initial phase angle $\theta(0)_i$ or initial degree dynamics $z(\mathbf{k}, 0)$ we wish to find their resulting description in the complex unit circle. Mapping operations onto the order parameter is straightforward using (20) and (26):

$$\theta(0)_i \longrightarrow Z(0) = \frac{1}{N} \sum_{j=1}^N e^{i\theta(0)_j} \quad (28)$$

$$z(\mathbf{k}, 0) \longrightarrow \bar{Z}(0) = \frac{1}{N} \sum_{\mathbf{k} \in \mathbb{K}} P(\mathbf{k}) z(\mathbf{k}, 0) \quad (29)$$

Starting from an initial synchronization $Z(0)$ and taking the inverse transformation, we can make use of the fact that the average of a set of identical values is the value itself. This is simple for $\theta(0)_i$: we can take all phase angles to be the same at $t = 0$. For $z(\mathbf{k}, 0)$ we have a weighed average which we need to undo, making sure that the whole sums up to N .

$$Z(0) \longrightarrow \theta(0)_i = -i \cdot \log(Z(0)) \quad (30)$$

$$Z(0) \longrightarrow z(\mathbf{k}, 0) = \frac{Z(0) \cdot n(\mathbf{k})}{P(\mathbf{k})} \quad (31)$$

Then, transforming between θ_i and $z(\mathbf{k})$, we need to filter θ_i per degree as there exist $P(\mathbf{k})$ number of nodes with $\deg(\theta_i) = \mathbf{k}$:

$$z(\mathbf{k}, 0) \longrightarrow \theta(0)_i = -i \cdot \log\left(\frac{z(\mathbf{k}) \cdot P(\mathbf{k})}{n(\mathbf{k})}\right) \quad \forall \theta \in \{\theta \mid \deg(\theta) = \mathbf{k}\} \quad (32)$$

$$\theta(0)_i \longrightarrow z(\mathbf{k}, 0) = \sum_{\mathbf{k}} e^{i\vartheta_{\mathbf{k}}} \quad \forall \vartheta_{\mathbf{k}} \in \{\vartheta_{\mathbf{k}} = \sum_{\mathbf{k}} \theta \mid \deg(\theta) = \mathbf{k}\} \quad (33)$$

The reason that these transformations only hold for the initial state is because it is currently unknown what distributions $\theta(t)_i$ and $z(\mathbf{k}, t)$ should have. That information is lost when taking the (weighed) average in (20) and (27).

6.4 Commutativity of complex vectors

It is important to notice that in (25) and (26) and many other equations in this work, we compute an inner vector product, which is non-commutative for complex numbers:

$$a \cdot b = \overline{b \cdot a} \quad a, b \in \mathbb{C}^r \quad (34)$$

This is the result of the *Conjugate* or *Hermitian* symmetry of the inner product. This is especially important in the MATLAB implementation, as one needs to remain consistent with left-hand or right-hand products.

6.5 Fixpoint iteration

In [7] a fixpoint iteration is suggested to find attractive fixpoints of the system (25). If we set $\frac{\partial z(\mathbf{k}, t)}{\partial t} = 0$ we can solve the following system:

$$\begin{aligned} i \frac{(z(\mathbf{k}, t) - 1)^2}{2} &= \frac{(z(\mathbf{k}, t) + 1)^2}{2} \cdot I(\mathbf{k}) \\ i \left(\frac{z(\mathbf{k}, t) - 1}{z(\mathbf{k}, t) + 1} \right)^2 &= I(\mathbf{k}) \\ \frac{z(\mathbf{k}, t) - 1}{z(\mathbf{k}, t) + 1} &\equiv b(\mathbf{k}, t) \\ z(\mathbf{k}, t) - 1 &= b(\mathbf{k}, t)z(\mathbf{k}, t) + b(\mathbf{k}, t) \\ z(\mathbf{k}, t) \cdot (1 - b(\mathbf{k}, t)) &= b(\mathbf{k}, t) + 1 \end{aligned}$$

We can then obtain the stable equilibria from:

$$ib(\mathbf{k}, t)^2 = I(\mathbf{k}) \quad z(\mathbf{k}, t)_\pm = \frac{1 \pm b(\mathbf{k}, t)}{1 \mp b(\mathbf{k}, t)} \quad (35)$$

where the signs are chosen so that $|z(\mathbf{k}, t)| \leq 1$. This works well, and in general this method converges fast.

6.6 A Newton-Raphson iteration for all fixpoints

The fixpoint iteration (35) only gives us the stable equilibria of the MFR. We can obtain all equilibria and their stability through the Jacobian from a Newton-Raphson iteration, which has been described in A.4. However, finding the Jacobian is a challenge, as (25) is non-holomorphic: $H_2(\mathbf{k}, t)$ does not satisfy the Cauchy-Riemann equations.

$$\begin{aligned} z(\mathbf{k}, t) &= x(\mathbf{k}, t) + i \cdot y(\mathbf{k}, t) \quad x, y \in \mathbb{R}^{M_k} \\ f(z(\mathbf{k}, t)) &= u(x(\mathbf{k}, t), y(\mathbf{k}, t)) + iv(x(\mathbf{k}, t), y(\mathbf{k}, t)) \\ &= \frac{1}{\langle k \rangle} \sum_{\mathbf{k}'} P(\mathbf{k}') a(\mathbf{k}' \rightarrow \mathbf{k}) \cdot \left(1 + \frac{z(\mathbf{k}', t)^2 + \bar{z}(\mathbf{k}', t)^2}{6} - \frac{4}{3} \operatorname{Re}(z(\mathbf{k}', t)) \right) \\ &= \frac{1}{\langle k \rangle} \sum_{\mathbf{k}'} P(\mathbf{k}') a(\mathbf{k}' \rightarrow \mathbf{k}) \cdot \left(1 + \frac{x(\mathbf{k}', t)^2}{3} - \frac{4}{3} x(\mathbf{k}', t) \right) \end{aligned}$$

This leaves us with only u defined as a real-valued function, so that the Cauchy-Riemann equations do not hold as v is zero. This is tricky, as expressing the Jacobian as a $M_k \times M_k$ matrix of complex numbers reduces the complexity.

QUESTION: explain how the Jacobian can be found as a $2M_k \times 2M_k$ matrix by using $z(\mathbf{k}, t) = [x(\mathbf{k}, t), y(\mathbf{k}, t)]$ and interweaving the x and y dimension in the matrix, see [27]. This has been implemented but only stable results for the PSR state. Right now I still can only find attractive fixpoints. Should I continue on this?

6.7 Fixed-degree networks as a baseline

Now we have all the necessary tools to simulate networks of theta neurons. First, we will use a fixed-degree network, as this is the most simple instance of the different topologies. The results are shown in Figure 8. There are small differences between simulation and theory, but these are most likely due to a finite network size and a finite integration step. As a matter of fact, the systems

(26) and (27) yield the exact same behaviour. This test benchmarks the lowest amount of error we can observe between simulation and theory, as for fixed-degree networks (25) consists of a single equation.

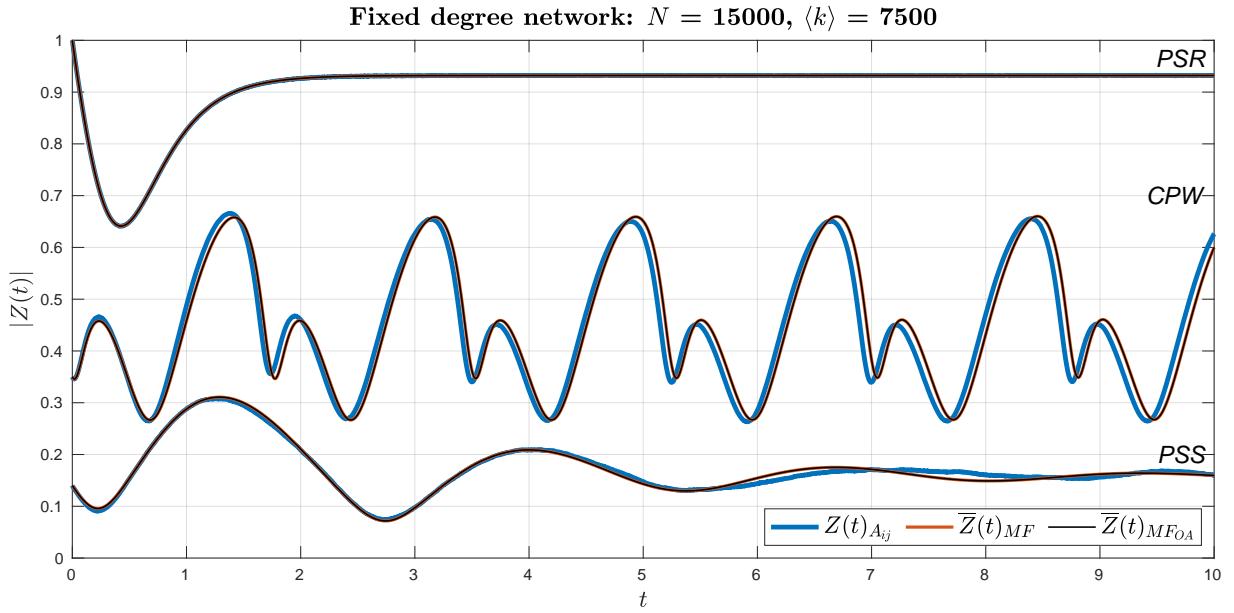


Figure 8: Comparison of the simulation of a fixed-degree network of Theta neurons and the Ott-Antonsen theory by the magnitude of the order parameter. We see that the same three macroscopic states are found by the three descriptions.

6.8 Results for arbitrary network topologies

The dynamics of random networks seem to be very similar to fixed-degree networks, when looking at the unit circle. We can see in Figure 11c that the limit cycle is a little larger. When looking at the dynamics over time the results in Figure 10 are also consistent, with a little more deviation between simulation and theory in the CPW state.

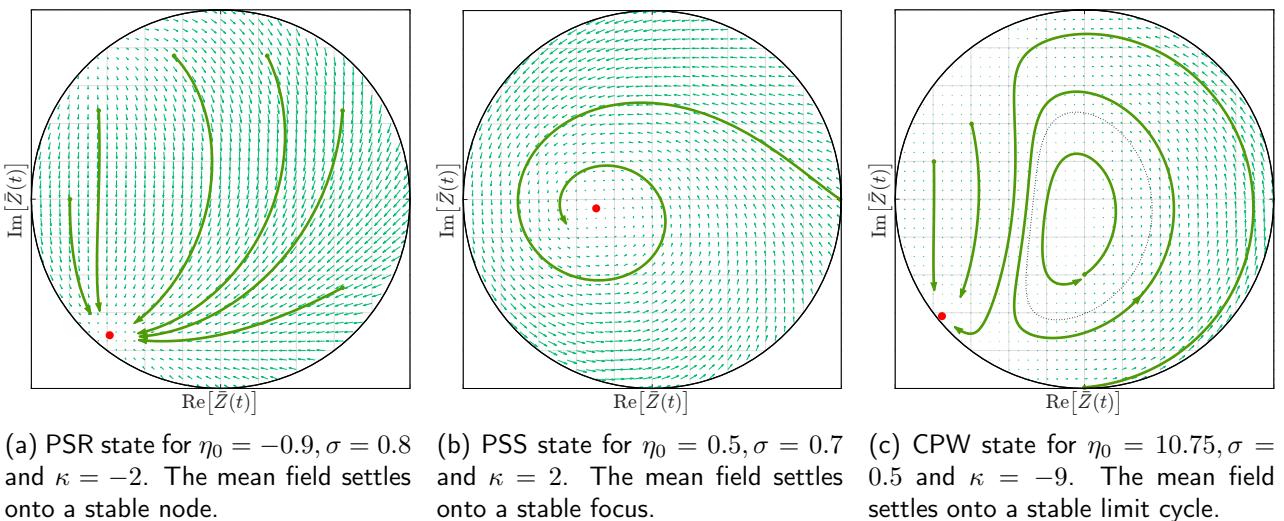


Figure 9: Three macroscopic states observed in the MFR inside the imaginary unit circle $|Z(t)| = 1$. Green arrows mark the phase space vector field and blue trails mark solution curves. Red points indicate equilibrium points, with black arrows marking the direction of the eigenvectors in that point, scaled according to the magnitude of the corresponding eigenvalues. The dotted line in the CPW state is the limit cycle of the fixed-degree networks, added for reference.

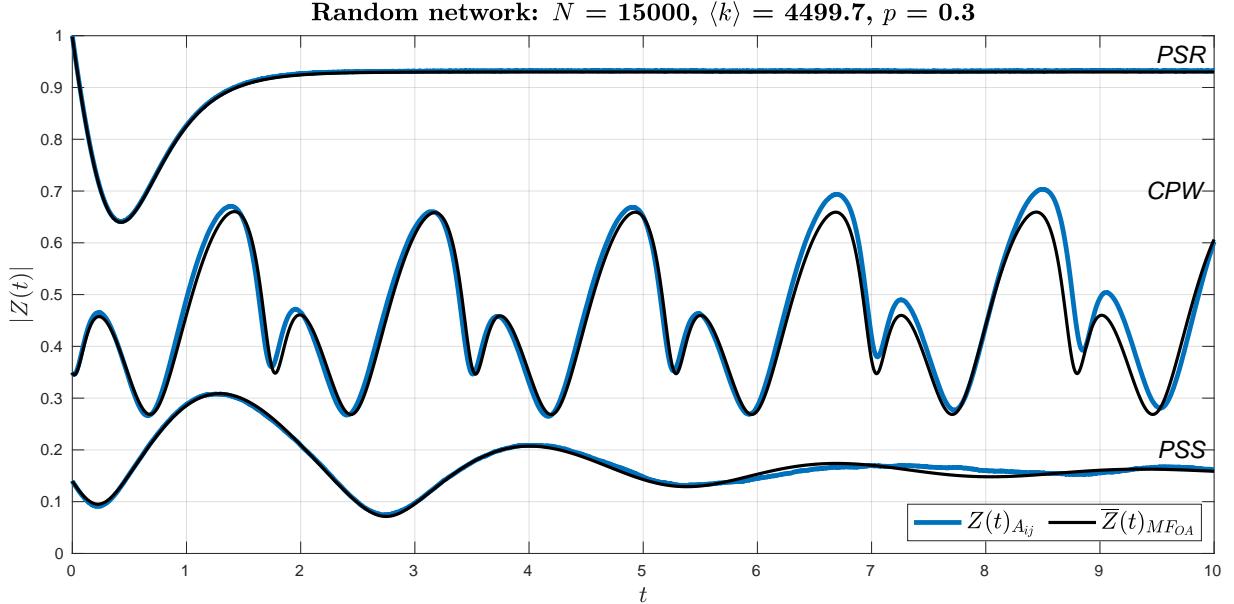
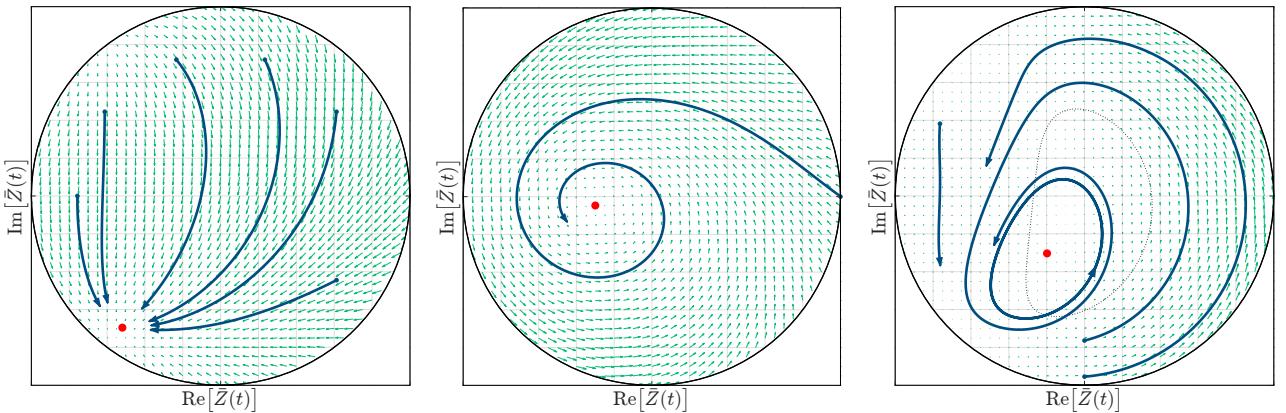


Figure 10: Comparison of the simulation of a random network of Theta neurons and the Ott-Antonsen theory by the magnitude of the order parameter.

For scale-free networks, we can see that again the three macroscopic states continue to exist, Figure 11. However, it seems like there is a fairly large discrepancy between simulation and theory, in Figure 12. The stable node in the *PSR* state is found at different locations, and the limit cycle in the *CPW* state seems to be very different, but with a similar period. Indeed, if we look at the limit cycle to which the dynamics are attracted to in Figure 13 we can indeed see two distinct cycles.

TODO: argue why this is the case, suspicions on the assortativity function not being assortative.



(a) PSR state for $\eta_0 = -0.9$, $\sigma = 0.8$ and $\kappa = -2$. The mean field settles onto a stable node.

(b) PSS state for $\eta_0 = 0.5$, $\sigma = 0.7$ and $\kappa = 2$. The mean field settles onto a stable focus.

(c) CPW state for $\eta_0 = 10.75$, $\sigma = 0.5$ and $\kappa = -9$. The mean field settles onto a stable limit cycle.

Figure 11: Three macroscopic states observed in the *MFR* inside the imaginary unit circle $|Z(t)| = 1$. Green arrows mark the phase space vector field and blue trails mark solution curves. Red points indicate equilibrium points, with black arrows marking the direction of the eigenvectors in that point, scaled according to the magnitude of the corresponding eigenvalues. The dotted line in the *CPW* state is the limit cycle of the fixed-degree networks, added for reference.

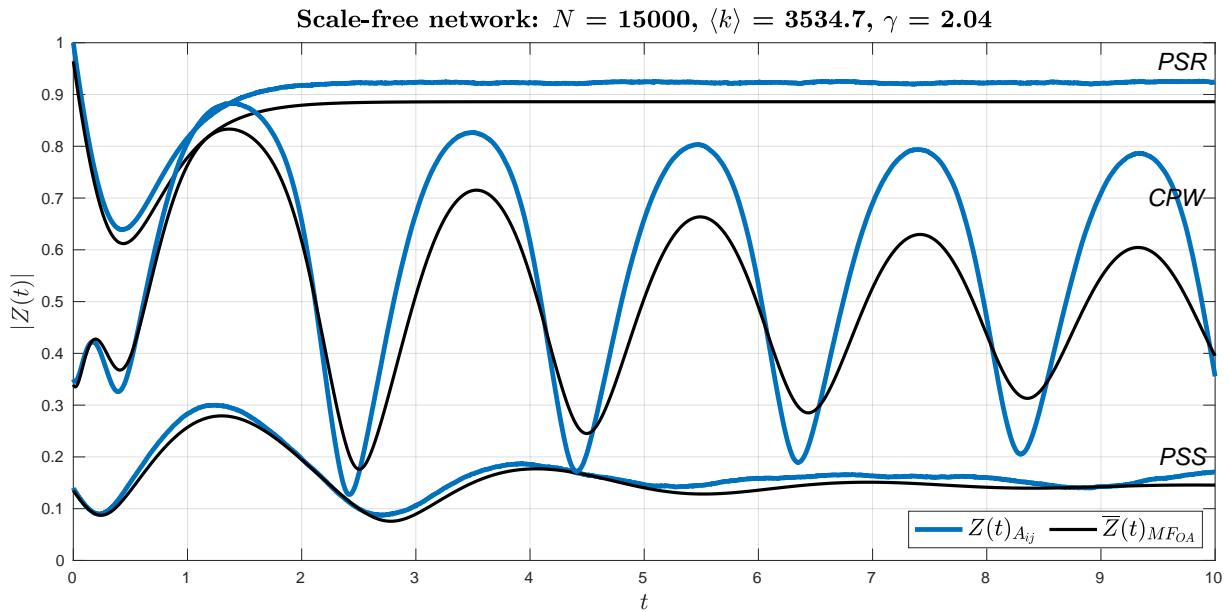


Figure 12: Comparison of the simulation of a scale-free network of Theta neurons and the Ott-Antonsen theory by the magnitude of the order parameter.

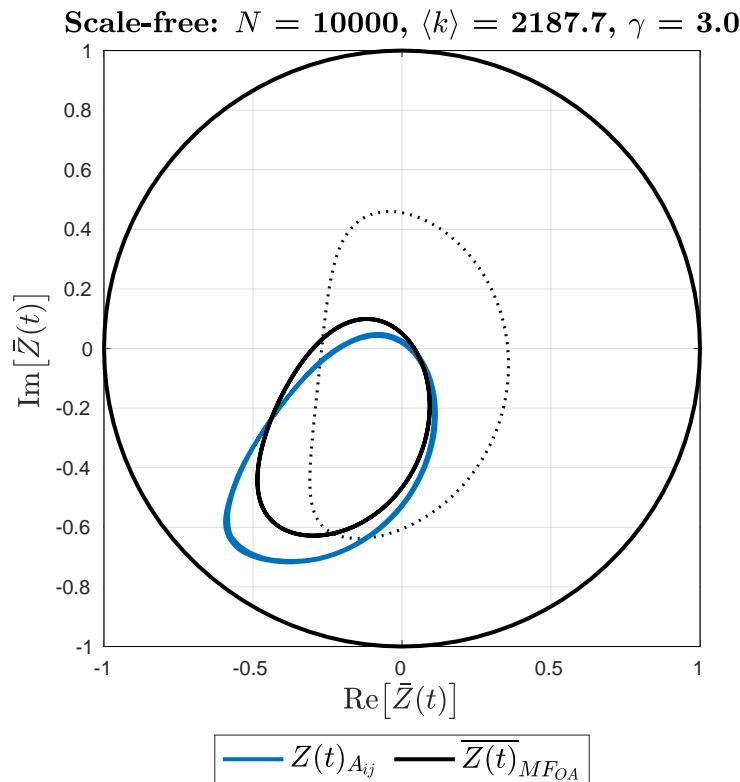


Figure 13: Comparison of the limit cycles found by theory and simulation.

7 Hebbian Learning and Synaptic Plasticity

When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased. [11]

This quote from Hebb has influenced the neuroscientific community since 1949. In its essence, Hebb postulated that neurons that *fire together, wire together*. It has since become known as *Hebbian learning*, and is simply modelled as a positive correlation between the action potentials of spiking neurons. It has been proven *in vivo* in many studies [28], just like its counterpart, *anti-Hebbian learning*, where a negative correlation can be found.

7.1 Spike-timing dependant plasticity

One specific temporal interpretation of these ideas is *spike-timing-dependent plasticity (STDP)*, where the relative timing of action potentials from the pre- and postsynaptic neuron determine causality [29, 30]. If the postsynaptic neuron fires right after the presynaptic neuron, then we can expect the synaptic strength from post- to presynaptic neuron to increase, and vice versa.

Let us say that neuron θ_i spikes at time t_i and that neuron θ_j spikes at t_j . Taking the time difference Δt_{ij} as $t_j - t_i$, we can say that when $\Delta t_{ij} > 0$ the spikes are correlated (there exists a temporally causal relation), and we can model an increase in synaptic strength of the connection from θ_i to θ_j , which we will gather in the *coupling matrix* K_{ij} . In the same fashion we can decrease K_{ji} when $\Delta t_{ij} < 0$ as there is no causal relation.

We will find an expression for ΔK_{ij} in function of Δt_{ij} so that at each time-step we can update $K_{ij} \leftarrow K_{ij} + \Delta K_{ij}$.

We can think of the coupling matrix as the continuous interpretation of κA_{ij} , where synaptic strength and network topology go hand in hand. This also means that we need to redefine some concepts:

$$\mathbf{k}_i^{\text{in}} = \sum_{j=1}^N |K_{ij}| \quad \mathbf{k}_j^{\text{out}} = \sum_{i=1}^N |K_{ij}| \quad (36)$$

$$\langle k \rangle = \frac{1}{N} \sum_{i,j=1}^N |K_{ij}| \quad \langle \hat{k} \rangle = \frac{1}{N} \sum_{i,j=1}^N K_{ij} \quad (37)$$

The absolute value ensures that we capture the magnitude of the coupling strength. The reason we want to distinguish between $\langle k \rangle$ and $\langle \hat{k} \rangle$ is that for some parts of the investigation it is beneficial to study how inhibitive and excitatory coupling strengths influence each other.

The functions $W(t)$ that relate Δt_{ij} to ΔK_{ij} are called *learning windows*, as they define a range in which K_{ij} is able to adapt or *learn*, and also when learning is optimal. When signals between neurons show a very large time difference (negative or positive) we do not expect them to be correlated. Because the learning windows are generally not symmetrical we can also expect the coupling matrix to be asymmetrical.

Another characteristic is the integral over the learning window. A window with a negative integral directs synaptic strengths mostly towards inhibitory behaviour, and vice versa with a positive integral. An integral of zero would mean that both inhibitory and excitatory synapses are stimulated equally. It has been proven that $\int W(\tau) d\tau$ is the magnitude of the correlation between signals [30].

This approach simplifies modeling the neuronal back-propagation, where another pulse is generated as an echo of the action potential which travels through the neuron dendrites (so, backwards). This behavior is believed to adjust the presynaptic weights, though it is a controversial subject [30].

In recent years, criticism on *STDP* has been growing, as experimental data has shown that *STDP* is usually accompanied by homeostatic plasticity of the neuron excitability and the synaptic strengths. Processes like *intrinsic plasticity* (IP), where one neuron's excitability changes over time as to self-regulate sensitivity to incoming action potentials, or *synaptic scaling*, where synapse characteristics are adjusted in unison to counteract positive feedback loops, have proven to stabilize the firing rate [28, 31]. When *STDP* and IP are combined, it seems like the two process balance each other out and stable network topologies can be found.

7.2 Formulations of *STDP* as a model

Following the notation in [29], we will denote the spike train coming from each neuron θ_i as $S_i^{\text{out}}(t) = \sum_n \delta(t - t_i^n)$, where t_i^n is the time that θ_i has fired. Similarly, we will denote the spike train coming into each neuron θ_i as $S_i^{\text{in}}(t) = \sum_f \delta(t - t_i^f)$ with t_i^f being the time that a neighbouring neuron has spiked. Now we can say that the synaptic strengths are adjusted as:

$$\Delta K_{ij} = \int_t^{t+\mathcal{T}} w^{\text{out}} S_i^{\text{out}}(\tau) + w^{\text{in}} S_j^{\text{in}}(\tau) d\tau + \iint_t^{t+\mathcal{T}} W(\tau' - \tau) S_i^{\text{out}}(\tau) S_j^{\text{in}}(\tau') d\tau d\tau' \quad (38)$$

$$= \sum_{t_i^n \in \mathcal{T}} w^{\text{out}} + \sum_{t_j^f \in \mathcal{T}} w^{\text{in}} + \sum_{t_j^f, t_i^n \in \mathcal{T}} W(t_j^f - t_i^n) \quad (39)$$

with \mathcal{T} the period over which learning occurs. $w^{\text{in}} > 0$ and $w^{\text{out}} < 0$ are small weights on the in- and outgoing action potentials and are necessary for K_{ij} to reach an equilibrium. This is proven from the average learning dynamics, [29]. In (38) we can recognise the correlation between signals as a convolution over the learning window.

The following learning window is proposed:

$$W(t)_K = \alpha \begin{cases} \left[A_p \left(1 - \frac{t}{\tilde{\tau}_p} \right) + A_n \left(1 - \frac{t}{\tilde{\tau}_n} \right) \right] \cdot \exp\left(\frac{t}{\tau_{\text{syn}}}\right) & \text{for } t \leq 0 \\ A_p \cdot \exp\left(-\frac{t}{\tilde{\tau}_p}\right) + A_n \cdot \exp\left(-\frac{t}{\tilde{\tau}_n}\right) & \text{for } t > 0 \end{cases} \quad (40)$$

Here t is the delay between presynaptic spike arrival and postsynaptic firing, α is a small learning parameter and all τ are time constants. Numerical values are usually $\alpha = 0.05$, $\tau_{\text{syn}} = 5 \text{ ms}$, $\tau_p = 1 \text{ ms}$, $\tau_n = 20 \text{ ms}$ and $A_p = 1$ and $A_n = -1$. $\tilde{\tau}_p \equiv \tau_{\text{syn}} \tau_p / (\tau_{\text{syn}} + \tau_p)$ and $\tilde{\tau}_n \equiv \tau_{\text{syn}} \tau_n / (\tau_{\text{syn}} + \tau_n)$. $\int W(t)_K d\tau = 2.56 \times 10^{-4}$.

Using different processes to generate spike trains, the learning equation (39) converges to a stable equilibrium [29]. The question will now be whether that is still the case when we will apply *STDP* to our network of Theta neurons, as changes to the coupling strength will also influence the spiking dynamics between neurons, which in turn will affect the learning again.

Another formulation of *STDP* as a mathematical model can be found in [32]. It is postulated without being concerned about the biological aspect too much, simplifying some of the ideas of [29]. The synaptic strengths are simply updated with:

$$\Delta K_{ij} = K^{\max} \cdot \sum_{t_j^f, t_i^n \in \mathcal{T}} W(t_j^f - t_i^n) \quad (41)$$

where K^{\max} is the maximum allowed synaptic strength, so that we can think of (41) as taking a percentage of the maximum coupling. The authors also constrain $0 \leq K_{ij} \leq K^{\max}$. In their further work on *STDP* and IP the authors booked remarkable progress, and their work is very interesting for our application [33].

The learning window is then again defined as a discontinuous function:

$$W(t)_S = \begin{cases} A_p \cdot \exp\left(\frac{-t}{\tau_p}\right) & \text{for } s > 0 \\ A_n \cdot \exp\left(\frac{t}{\tau_n}\right) & \text{for } s \leq 0 \end{cases} \quad (42)$$

where we will use $A_p = 0.005$, $A_n = -0.00525$ and $\tau_p = \tau_n = 20$ ms. $\int W(\tau)_S d\tau = -3.70 \times 10^{-4}$ so we expect the weights to be suppressed towards a negative value.

Recently, triphasic learning windows have been used to account for when it takes too long for the postsynaptic neuron to fire, and thus to decorrelate the relation between neurons. These learning windows are curves that were fitted to experimental data of the cortex and the hippocampus [28].

Extending the work of [32] we can find a brief investigation of network topology and clustering using triphasic windows, [34]. The method is as in (42), with the following learning window:

$$W(t)_C = A_p \cdot \exp\left(\frac{-(t-15)^2}{\tau_p}\right) - A_n \cdot \exp\left(\frac{-(t-20)^2}{\tau_n}\right) \quad (43)$$

where $A_p = 0.23$, $A_n = 0.15$, $\tau_p = 200$ and $\tau_n = 2000$. $\int W(s)_C ds = -60.0 \times 10^{-4}$.

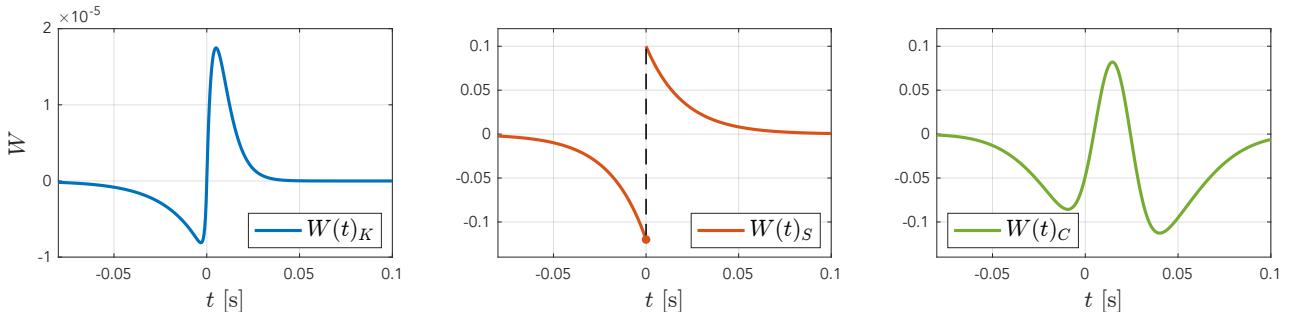


Figure 14: Three different biphasic learning windows. Left: in W_K the learning happens much slower, as the magnitude of the window is very small. Middle: we can see how in $W(t)_S$ a larger weight is put on the anti-Hebbian learning. Right: a triphasic window can punish the learning when signals arrive either too early or too late.

The learning windows generally have $W(t^*) = 0$ for $t^* \geq 0$. This means that no learning will take place when the delay between neuron spikes is exactly $t^* \geq 0$. The triphasic windows show two of those points.

It is important to use a time-step lower than 0.01 when implementing *STDP*. Otherwise we might miss important details in the curve of the learning window.

7.3 Synaptic scaling

There is no upper or lower bound on the synapse strength, and generally connection strengths are nonzero. Positive reinforcement loops might disturb the learning process, which would not be beneficial for the model. One technique we can apply to keep the strengths within a definitive

range is to scale homeostatically - a method where any increases in synaptic strength will balance out any decreases by scaling:

$$K_{ij}^s = K_{ij} \frac{\frac{1}{N} \sum_{i,j} K_{ij}}{\sum_i K_{ij}} \quad (44)$$

In this way, the out-degrees will remain constant. Using this approach, something has to remain constant, whether that is $\langle k \rangle$, or $\langle k \rangle^2$ or any other property of the adjacency matrix. However, this property is not what we are after: we want a method that is able to change the network topology entirely.

7.4 Intrinsic plasticity

Instead of scaling the weights to preserve a certain quantity in the network, we can allow the neurons to adjust their sensitivity to incoming signals. So when some synaptic strengths are increased, we can reduce the excitability, and vice-versa. This should counteract positive feedback. In [33] such a method is introduced in detail. We can simply update $\eta_i \leftarrow \eta_i + \eta_{\max} \cdot \phi_i$, where:

$$\phi_i(t) = \begin{cases} -\alpha \cdot \exp\left(\frac{T_{\min}-t}{T_{\min}}\right) & \text{for } t < T_{\min} \\ \alpha \cdot \exp\left(\frac{t-T_{\max}}{T_{\max}}\right) & \text{for } t > T_{\max} \\ 0 & \text{otherwise} \end{cases} \quad (45)$$

The argument t represents the time that has passed between successive spike of the same neuron, the *inter-spike interval* (ISI). This is always a positive number.

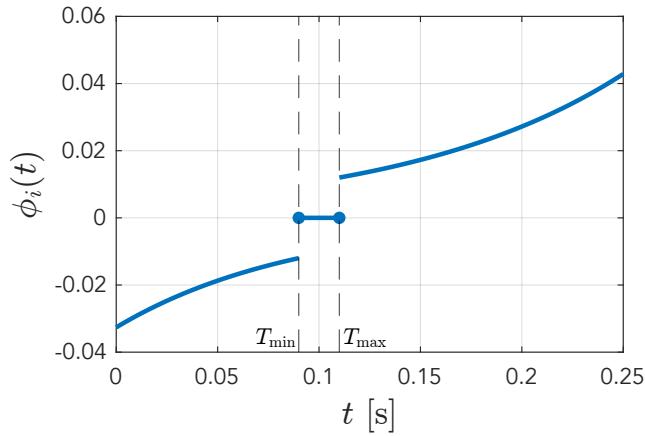


Figure 15: The learning behaviour for intrinsic plasticity as proposed in [33].

QUESTION: is it necessary to include more background theory on Hebbian Learning?

8 *Investigation: Emerging Network Topologies*

8.1 *STDP applied to networks of theta neurons*

TODO: explain in detail how the different methods eqs. (39) and (41) will be implemented, using STDP coupled with IP.

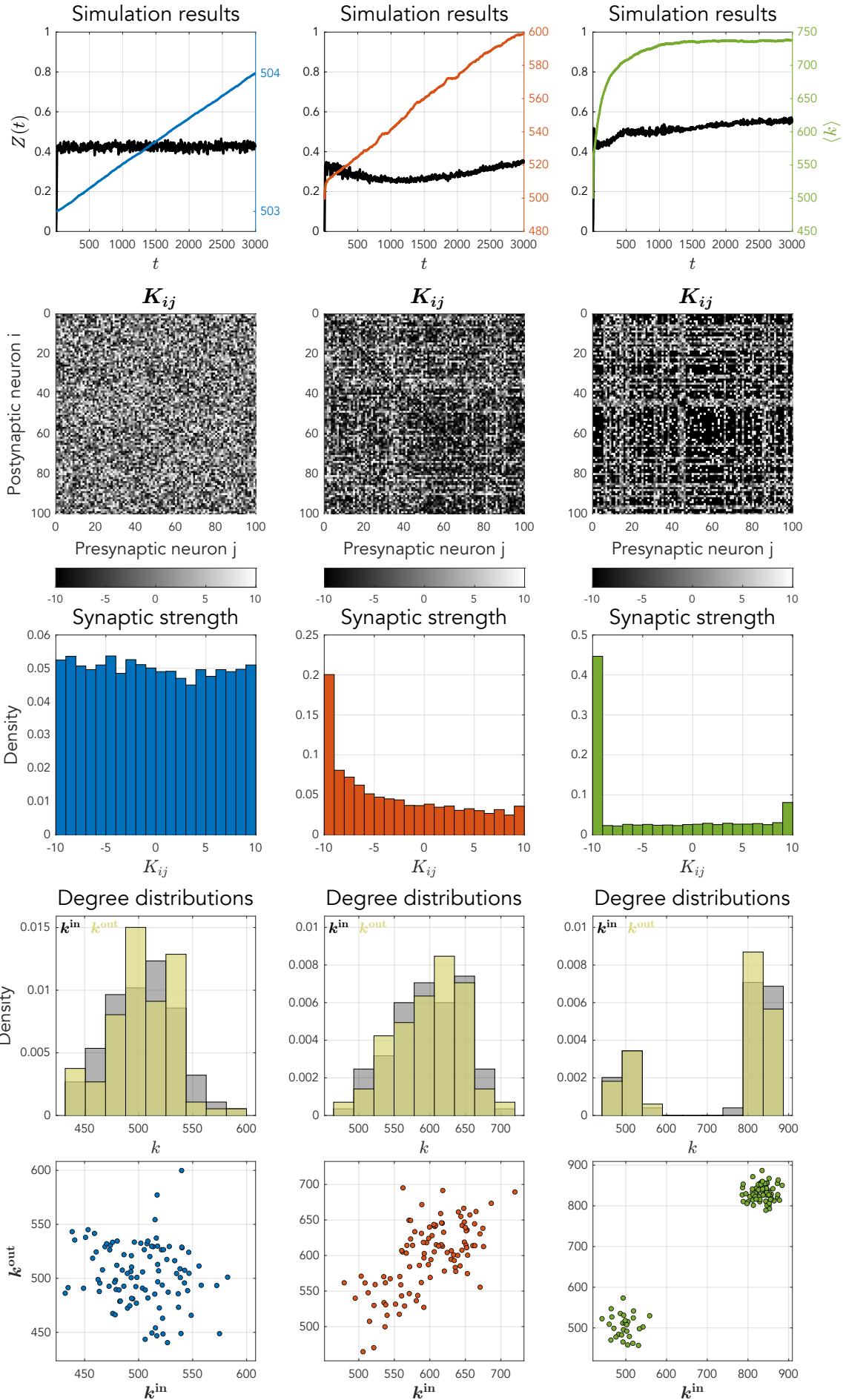


Figure 16: Results of the STDP learning.

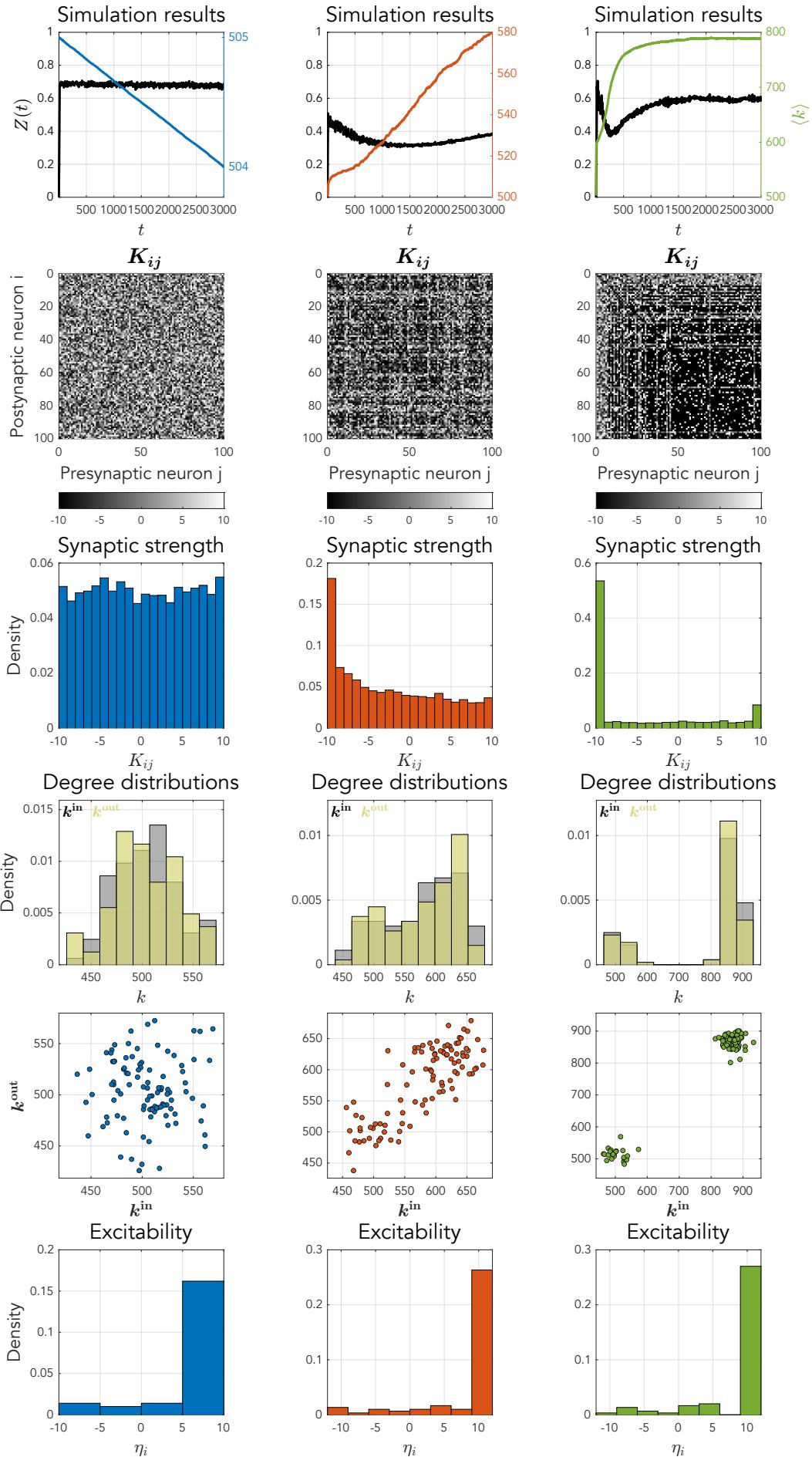


Figure 17: Results of the STDP learning with IP learning.

8.2 Results

TODO: *describe the emergent behaviour.*

9 Conclusion and Discussion

TODO: *what did we achieve with this work?*

10 References

- [1] Human Brain Project, 2017. <https://www.humanbrainproject.eu/en/>. Accessed: 28.05.2020.
- [2] C. Börgers, *An Introduction to Modeling Neuronal Dynamics*. Texts in Applied Mathematics. Springer International Publishing, 2018.
- [3] The Nobel Prize in Physiology or Medicine 1963, 2009. <https://www.nobelprize.org/prizes/medicine/1963/speedread/>. Accessed: 28.05.2020.
- [4] S. Herculano-Houzel, *The Human Brain in Numbers: A Linearly Scaled-up Primate Brain*. *Frontiers in human neuroscience* **3** (11, 2009) 31.
- [5] E. Bullmore and D. Bassett, *Brain Graphs: Graphical Models of the Human Brain Connectome*. *Annual review of clinical psychology* **7** (04, 2010) 113–40.
- [6] E. Ott and T. M. Antonsen, *Low Dimensional Behavior of Large Systems of Globally Coupled Oscillators*. [arXiv:0806.0004 \[nlin.CD\]](https://arxiv.org/abs/0806.0004).
- [7] S. Chandra, D. Hathcock, K. Crain, T. Antonsen, M. Girvan, and E. Ott, *Modeling the Network Dynamics of Pulse-Coupled Neurons*. *Chaos (Woodbury, N.Y.)* **27** (03, 2017) 10.
- [8] The University Of Queensland Brain Institute, *Action potentials and synapses*. <https://qbi.uq.edu.au/brain-basics/brain/brain-physiology/action-potentials-and-synapses>. Composed: 09/11/2017. Accessed: 02/10/2020.
- [9] B. Ermentrout and D. Terman, *The Mathematical Foundations of Neuroscience*. vol. Vol. 35. Springer, July, 2010.
- [10] X. Zhang and J. Feng, *Computational modeling of neuronal networks*. [arXiv:1203.0868 \[q-bio.NC\]](https://arxiv.org/abs/1203.0868).
- [11] D. O. Hebb, *The organization of behaviour: a neurophysical theory*. John Wiley and sons, 1949.
- [12] F. C. Hoppensteadt and E. M. Izhikevich, *Canonical Neural Models*., 2001.
- [13] A. Hodgkin, *The local electric changes associated with repetitive action in a non-medullated axon*. *The Journal of physiology* **107** no. 2, (March, 1948) 165—181. <https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/16991796/?tool=EBI>.
- [14] B. Ermentrout and N. Kopell, *Parabolic Bursting in an Excitable System Coupled with a Slow Oscillation*. *Siam Journal on Applied Mathematics - SIAMAM* **46** (04, 1986) 233–253.
- [15] T. Luke, E. Barreto, and P. So, *Complete Classification of the Macroscopic Behavior of a Heterogeneous Network of Theta Neurons*. *Neural Computation* **25** (12, 2013) 1–28.
- [16] B. Gutkin, *Theta-Neuron Model*. Springer New York, New York, NY, 2013. https://doi.org/10.1007/978-1-4614-7320-6_153-1.
- [17] F. A. Pérez, *Phase-responsiveness transmission in a network of quadratic integrate-and-fire neurons*. Universitat Politècnica de Catalunya, Facultat de Matemàtiques i Estadística (January, 2020) . Bachelor Thesis.
- [18] B. Ermentrout, *Type i Membranes, Phase Resetting Curves, and Synchrony* *Neural Comput.* **8** no. 5, (July, 1996) 979–1001. <https://doi.org/10.1162/neco.1996.8.5.979>.
- [19] P. Erdos and A. Rényi, *On Random Graphs I*. *Publicationes Mathematicae* **6** (1959) 290–297.
- [20] A. Barabási, *Network Science*. Cambridge University Press, 2016. <https://books.google.dk/books?id=iLtGDQAAQBAJ>.

- [21] A.-L. Barabási and E. Bonabeau, *Scale-Free Networks*. *Scientific American* **288** no. 60-69, (2003) 50–59. <http://www.nd.edu/~networks/PDF/Scale-Free%20Sci%20Amer%20May03.pdf>.
- [22] C. I. Del Genio, T. Gross, and K. E. Bassler, *All Scale-Free Networks Are Sparse*. *Phys. Rev. Lett.* **107** (Oct, 2011) 178701. <https://link.aps.org/doi/10.1103/PhysRevLett.107.178701>.
- [23] C. Bick, M. Goodfellow, C. Laing, and E. Martens, *Understanding the dynamics of biological and neural oscillator networks through exact mean-field reductions: a review*. *Journal of Mathematical Neuroscience* **10** no. 1, (Dec., 2020) .
- [24] E. Ott and T. M. Antonsen, *Long Time Evolution of Phase Oscillator Systems*. [arXiv:0902.2773 \[nlin.CD\]](https://arxiv.org/abs/0902.2773).
- [25] E. Ott, B. R. Hunt, and T. M. Antonsen, *Comment on "Long Time Evolution of Phase Oscillator Systems"* [*Chaos* 19, 023117 (2009), arXiv: 0902.2773]. [arXiv:1005.3319 \[nlin.CD\]](https://arxiv.org/abs/1005.3319).
- [26] J. Restrepo and E. Ott, *Mean field theory of assortative networks of phase oscillators* *EPL (Europhysics Letters)* **107** (07, 2014) .
- [27] D. J. Cross, *On the Relation between Real and Complex Jacobian Determinants*, 05, 2018.
- [28] J. Chrol-Cannon and Y. Jin, *Computational Modeling of Neural Plasticity for Self-Organization of Neural Networks*. *Bio Systems* **125** (04, 2014) .
- [29] R. Kempter, W. Gerstner, and L. van Hemmen, *Hebbian learning and spiking neurons*. *Phys. Rev. E* **59** (04, 1999) .
- [30] W. Gerstner and W. Kistler, *Mathematical Formulations of Hebbian Learning*. *Biological cybernetics* **87** (01, 2003) 404–15.
- [31] H. Lee and A. Kirkwood, *Mechanisms of Homeostatic Synaptic Plasticity in vivo*. *Frontiers in Cellular Neuroscience* **13** (Dec., 2019) .
- [32] S. Song, K. Miller, and L. Abbott, *Competitive Hebbian learning through spike timing-dependent plasticity* *Nature neuroscience* **3** (10, 2000) 919–26.
- [33] X. Li, W. Wang, F. Xue, and Y. Song, *Computational modeling of spiking neural network with learning rules from STDP and intrinsic plasticity*. *Physica A-statistical Mechanics and Its Applications* **491** (2018) 716–728.
- [34] J. Chrol-Cannon, A. Grüning, and Y. Jin, *The emergence of polychronous groups under varying input patterns, plasticity rules and network connectivities* in *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6, IEEE. 2012.

A Appendix

A.1 Transformation to the QIF model

We prove that the transformation (2) holds from the QIF model (3) to the Theta model (1).

$$V \equiv \tan\left(\frac{\theta}{2}\right) \quad \rightarrow \quad \frac{dV}{dt} = \frac{1}{2 \cos^2\left(\frac{\theta}{2}\right)} \frac{d\theta}{dt}$$

Insert into $\frac{dV}{dt} = V^2 + I$:

$$\frac{d\theta}{dt} = 2 \left(\cos^2\left(\frac{\theta}{2}\right) \cdot \tan^2\left(\frac{\theta}{2}\right) + \cos^2\left(\frac{\theta}{2}\right) \cdot I \right) = 2 \left(\sin^2\left(\frac{\theta}{2}\right) + \cos^2\left(\frac{\theta}{2}\right) \cdot I \right)$$

Using $\cos^2\left(\frac{\theta}{2}\right) = \frac{1+\cos(\theta)}{2}$ and $\sin^2\left(\frac{\theta}{2}\right) = \frac{1-\cos(\theta)}{2}$:

$$\dot{\theta} = 2 \left(\frac{1 - \cos \theta}{2} + \left(\frac{1 + \cos \theta}{2} \right) \cdot I \right) = (1 - \cos \theta) + (1 + \cos \theta) \cdot I$$

This proves that the transformation (2) is correct.

A.2 Solutions to the QIF model

Depending on the value of I , we can distinguish multiple solutions [17]. In all cases we can integrate through the separation of variables. Solutions are bound to start at $V(t_0)$, right after a spike has occurred at $t = t_0$.

A.2.1 Solving for $I < 0$

$$\begin{aligned} \int_{V(t_0)}^{V(t)} \frac{dv}{v^2 - \tilde{I}^2} &= \int_{V(t_0)}^{V(t)} \frac{dv}{(v + \tilde{I})(v - \tilde{I})} = \frac{1}{2\tilde{I}} \int_{V(t_0)}^{V(t)} \frac{dv}{v - \tilde{I}} - \frac{1}{2\tilde{I}} \int_{V(t_0)}^{V(t)} \frac{dv}{v + \tilde{I}} \\ &= \frac{1}{2\tilde{I}} \log\left(1 - \frac{2\tilde{I}}{v + \tilde{I}}\right) \Big|_{V(t_0)}^{V(t)} = \int_{t_0}^t d\tau = t - t_0 \\ V(t) &= \lim_{V(t_0) \rightarrow -\infty} \frac{2\sqrt{-I}}{1 - \left(1 - \frac{2\sqrt{-I}}{V(t_0) + \sqrt{-I}}\right) \cdot e^{2(t-t_0)\sqrt{-I}}} - \sqrt{-I} \\ &= \frac{2\sqrt{-I}}{1 - e^{2(t-t_0)\sqrt{-I}}} - \sqrt{-I} \end{aligned}$$

A.2.2 Solving for $I = 0$

$$\begin{aligned} \int_{V(t_0)}^{V(t)} \frac{dv}{v^2} &= \frac{1}{v} \Big|_{V(t_0)}^{V(t)} = -\frac{1}{V(t)} + \frac{1}{V(t_0)} = \int_{t_0}^t d\tau = t - t_0 \\ V(t) &= \lim_{V(t_0) \rightarrow -\infty} \frac{V(t_0)}{1 - V(t_0)(t - t_0)} \stackrel{\text{H}}{\underset{\infty}{\approx}} \frac{-1}{t - t_0} \end{aligned}$$

A.2.3 Solving for $I > 0$

$$\begin{aligned} \int_{V(t_0)}^{V(t)} \frac{dv}{v^2 + I} &= \int_{V(t_0)}^{V(t)} \frac{I}{\left(\frac{v}{\sqrt{I}}\right)^2 + 1} dv \stackrel{x=\frac{v}{\sqrt{I}}}{=} \int_{\frac{V(t_0)}{\sqrt{I}}}^{\frac{V(t)}{\sqrt{I}}} \frac{I}{x^2 + 1} dx = \frac{1}{\sqrt{I}} \arctan(x) \Big|_{\frac{V(t_0)}{\sqrt{I}}}^{\frac{V(t)}{\sqrt{I}}} \\ &= \frac{1}{\sqrt{I}} \left(\arctan\left(\frac{V(t)}{\sqrt{I}}\right) - \arctan\left(\frac{V(t_0)}{\sqrt{I}}\right) \right) = \int_{t_0}^t d\tau = t - t_0 \\ V(t) &= \lim_{V(t_0) \rightarrow -\infty} \sqrt{I} \cdot \tan\left((t - t_0)\sqrt{I} + \arctan\left(\frac{V(t_0)}{\sqrt{I}}\right)\right) = \sqrt{I} \cdot \tan\left((t - t_0)\sqrt{I} - \frac{\pi}{2}\right) \\ &= \sqrt{I} \cdot \cot\left((t - t_0)\sqrt{I}\right) \end{aligned}$$

A.3 Frequency response of the neuron models

The integral is solved like before, but now with the conditions of the spike:

$$\begin{aligned} T &= \lim_{a \rightarrow \infty} \int_{-a}^a \frac{I}{\left(\frac{v}{\sqrt{I}}\right)^2 + 1} dv \stackrel{x = \frac{v}{\sqrt{I}}}{=} \lim_{a \rightarrow \infty} \int_{\frac{-a}{\sqrt{I}}}^{\frac{a}{\sqrt{I}}} \frac{I}{x^2 + 1} dx = \lim_{a \rightarrow \infty} \frac{1}{\sqrt{I}} \arctan(x) \Big|_{\frac{-a}{\sqrt{I}}}^{\frac{a}{\sqrt{I}}} \\ &= \frac{1}{\sqrt{I}} \left(\frac{\pi}{2} - \left(-\frac{\pi}{2} \right) \right) = \frac{\pi}{\sqrt{I}} \end{aligned}$$

So the frequency of oscillation is proportional to \sqrt{I} .

A.4 Newton-Raphson root iteration

We define the equilibria $\mathbf{x}^* \in \mathbb{R}^n$ of a multivariate function $\mathbf{f}(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ with $\mathbf{f}(\mathbf{x}) = \mathbf{0}$. Expanding \mathbf{f} as a Taylor series, we obtain:

$$f_i(\mathbf{x} + \delta\mathbf{x}) = f_i(\mathbf{x}) + \sum_{j=1}^n \frac{\partial f_i(\mathbf{x})}{\partial x_j} \delta x_j + O(\delta\mathbf{x}^2) \approx f_i(\mathbf{x}) + \sum_{j=1}^n \frac{\partial f_i(\mathbf{x})}{\partial x_j} \delta x_j, \quad (i = 1, \dots, n)$$

We can also write this in vector notation, by setting $\mathbf{J}(\mathbf{x}) = \nabla \mathbf{f}(\mathbf{x}) = \frac{d}{d\mathbf{x}} \mathbf{f}(\mathbf{x}) \in \mathbb{R}^{n \times n}$

$$\mathbf{f}(\mathbf{x} + \delta\mathbf{x}) \approx \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_N(\mathbf{x}) \end{bmatrix} + \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial x_1} & \cdots & \frac{\partial f_N}{\partial x_N} \end{bmatrix} \begin{bmatrix} \delta x_1 \\ \vdots \\ \delta x_N \end{bmatrix} = \mathbf{f}(\mathbf{x}) + \mathbf{J}(\mathbf{x}) \delta\mathbf{x}$$

By assuming $\mathbf{f}(\mathbf{x} + \delta\mathbf{x}) = 0$ we can find that $\delta\mathbf{x} = -\mathbf{J}^{-1}(\mathbf{x})\mathbf{f}(\mathbf{x})$ so that $\mathbf{x} + \delta\mathbf{x} = \mathbf{x} - \mathbf{J}^{-1}(\mathbf{x})\mathbf{f}(\mathbf{x})$. This expression converges to \mathbf{x}^* . When the equations are nonlinear, the equations converge to the real root as $\mathbf{x}_k = \mathbf{x}_k - \mathbf{J}^{-1}(\mathbf{x}_k)\mathbf{f}(\mathbf{x}_k)$.