Synchrony and Asynchrony in Neural Networks

Fabian Kuhn
Faculty of Informatics
University of Lugano (USI), Switzerland
fabian.kuhn@usi.ch

Konstantinos Panagiotou
Max-Planck-Institute for Computer Science
Saarbrücken, Germany
kpanagio@mpi-inf.mpg.de

Joel Spencer
Department of Computer Science
Courant Institute, New York
spencer@courant.nyu.edu

Angelika Steger
Institute for Theoretical Computer Science
ETH Zurich, Switzerland
steger@inf.ethz.ch

Abstract

The dynamics of large networks is an important and fascinating problem. Key examples are the Internet, social networks, and the human brain. In this paper we consider a model introduced by DeVille and Peskin [6] for a stochastic pulse-coupled neural network.

The key feature and novelty in their approach is that they describe the interactions of a neuronal system as a *discrete*-state stochastic dynamical network. This idealization has two benefits: it captures essential features of neuronal behavior, and it allows the study of spontaneous *synchronization*, an important phenomenon in neuronal networks that is well-studied but unfortunately far from being well-understood. In synchronous behavior the firing of one neuron leads to the firing of other neurons, which in turn may set off a chain reaction that often involves a substantial proportion of the neurons.

In this paper we rigorously analyze their model. In particular, by applying methods and tools that are frequently used in theoretical computer science, we provide a very precise picture of the dynamics and the evolution of the given system. In particular, we obtain insights into the coexistence of synchronous and asynchronous behavior and the conditions that trigger a "spontaneous" transition from one state to another.

1 Introduction

Mathematical models in biology and other fields have several different uses. On the one hand, the goal is to make the model as *realistic* as possible, so that it can be used as an alternative to the actual system. Typically, such models are very complicated and include many relevant details, which makes a rigorous analysis of them difficult, if not impossible. On the other hand, a model can be designed to *isolate* just a particular mechanism. In this case, certain features of the real system are omitted, and the aim is to investigate how a subset of the actual ingredients produces a specific physiological phenomenon. A benefit of such idealized models is that it might be possible to analyze them rigorously, thus obtaining more insights into the actual behavior of the real system. This is the main topic of our work. In particular, we consider a model introduced by DeVille and Peskin [6] for a stochastic pulse-coupled neural network.

Understanding the human brain is a fascinating topic, as it is difficult to tackle with our current level of knowledge. Almost all aspects provide more questions than answers. This reaches from understanding functionality from a high level, down to developing models describing specific phenomena observed by neuroscientists. Concerning the latter, it is for example well-known, see e.g. [3], that dynamic and in particular *oscillatory* behavior of large collections of neurons plays a fundamental role in the functionality of our brain. However, the details and, more importantly, their functional consequences are still far from being well-understood. Over the last decades this topic was intensively studied, see e.g. [2, 5, 7, 11] for some recent work, and many references therein.

In this work we consider a discrete-state neural model introduced recently by DeVille and Peskin [6]. The key result and important insight of this work is that they show experimentally and by some estimates for the expected behavior of the model that the network can exhibit both *synchronous* and *asynchronous* behavior, a property which is omnipresent in the human brain [3]. Here, synchrony is achieved if there are massive interactions between the neurons. In particular, when a neuron accumulates a sufficiently high internal potential, it *fires*, which increases the potential of neighboring neurons. Some of those could now fire as well, which might start an immense chain reaction, a so-called "big burst". Alternatively, the network is in an asynchronous state when there are only few interactions between the neurons, and only small bursts occur. DeVille and Peskin also exhibit ranges for the parameters of the model for which the network switches spontaneously between synchrony and asynchrony.

The precise setting in [6] is the following. The system consists of n identical neurons. Each neuron has k levels, which we denote by $0, \ldots, k-1$. Moreover, there is an auxiliary level k, which is useful for describing neurons that are currently firing. The system has two overall modes, which are called the "burst" mode and the "interburst" mode. During the interburst mode, the neurons are completely independent, and in each time step a random neuron will be selected and promoted by one level. We shall sometimes say that this neuron was promoted due to the *outside impulse*. When a neuron in level $k-1$ is selected, it *fires*. At that moment, the interburst mode ends, and the burst mode begins. While in burst mode we repeat the following rule: for each neuron that fires, all neurons that have not fired yet are promoted by one level with probability p. If any neuron is promoted to level k while the system is the burst mode, it also fires, and its effect is computed recursively. During a burst, any particular neuron is allowed to fire at most once. That is, all neurons that reach level k can be thought of as being placed in one of two additional buckets. Each neuron that reaches level k is first placed in bucket W (the "*waiting*" neurons, or the *queue*) where it stays until the above rule is performed for it, we also say until it "fires". As soon as a neuron has fired, it is removed from W and placed into bucket F. The neurons in F are collectively put back into the system at level 0 once the burst is over (i.e., $|W| = 0$), and the system returns to the interburst mode. Note that while we described the burst mode as a recursive procedure, one should actually think of it as taking place in a *single* instant of time.

The model described above is clearly of the idealized variety, trying to focus on and to investigate a specific type of neuronal interactions and activities. Particularly, the *discrete* nature of the model captures nicely the interneuronal communication through vesicle release at synapses, i.e., "connections" between the neurons. As DeVille and Peskin argue, the purpose of the model is to allow a systematic study of the synchronization of pulse-coupled oscillators in a new setting that captures certain key features of neuronal interaction observed by neuroscientists. We refer the reader to [6] for a very elaborate discussion on this topic.

Our Results The typical approach used in biology to measure the quality of a model is to investigate the so-called mean-field limit and to perform (numerical) simulations. In particular, mean-field models for the dynamics of systems of neurons have been studied extensively in the past in various different settings, see e.g. [1, 4, 8] for some recent examples. However, no guarantees are given that the actual model behaves like the associated mean-field limit with reasonable confidence.

The first main contribution of our work is to show rigorously that the "typical behavior" of the discrete probabilistic system described above is very close to the trajectories described by an associated mean-field model. In particular, using techniques that were designed for the analysis of randomized algorithms and, more generally, random systems, we prove that the trajectory of our system will deviate significantly from some deterministic trajectory with only exponentially small probability. A major difficulty that we have to overcome here is that the system "jumps" between two inherently different states: the burst and the interburst mode. While the system is in either of this two modes, by applying a well-understood tool (the so-called "Differential Equation Method", which is used to study the behavior of certain random processes like e.g. Markov chains) we obtain after some technical work that with probability very close to one the actual behavior is not far from a solution of a set of differential equations. However, handling the "transitions" between the two modes is a challenging task, which requires the development of new techniques. As a side remark, we want to stress here that our mean-field description is different from the one given in [6], as we insist on a rigorous analysis.

The second main contribution of our work is to analyze rigorously the mean-field approximation. In particular, we answer the questions of DeVille and Peskin about the actual parameter settings and thresholds for which changes between synchronous and asynchronous behavior occur. In addition to that, we give a very precise picture for the dynamics of the system, depending on the values of the crucial parameters p and k of the model.

Before we state our results in detail we first try to convey some intuition. Assume for the moment (the unrealistic setting) that $k = 1$. Then each neuron that is promoted one level in the interburst mode starts to fire immediately. The following burst mode can then be viewed as follows. Think of a random graph $G_{n,p}$ (in the Erdős-Rényi sense, with n nodes/neurons and edge probability p). Then the neuron that fires is in some connected component of $G_{n,p}$, and the other neurons that will fire in this burst phase are exactly the other nodes in that component. That is, whether we will see a "big" burst (linear in n) or only "small" bursts (of size $o(n)$) depends on the relation of p and n. More precisely, by applying well-known results from random graph theory, see e.g. [9], we will see only small bursts if $p \ll 1/n$, while in the case $p \gg 1/n$ *huge* bursts also happen.

If $k > 1$ things get much more interesting. Clearly, we can still think of the neurons at level $k-1$ forming a random graph with edge probability p, and we know that all neurons in the component that contains a firing neuron will also fire. However, due to the presence of the smaller levels also other neurons may fire. Assume, for example, that a neuron from level $k-2$ is promoted during a burst phase to level $k-1$. It then has to be integrated into the random graph – and by that it may combine two connected components into a larger one.[1]

At this point the following should be plausible. Let $p = \beta k/n$. If $\beta > 1$ then regardless of the starting configuration of the system, we will reach a point where level $k-1$ contains at least $1/p$ neurons – and a big burst is thus likely to occur. More precisely, we have the following statement.

THEOREM 1.1. *Suppose that $p = \beta k/n$, where $\beta > 1$. Then, for sufficiently large k, regardless of the starting configuration, we will observe with high probability after finitely many time steps a big burst in which $\Theta(n)$ neurons fired.*

Now let us consider the case $\beta < 1$. If we start with a configuration in which all levels contain roughly n/k neurons then nothing exciting is going to happen: we will probably never experience any big bursts and the system stays in the state where all levels contain roughly the same number of neurons. If on the other hand we start in a configuration in which all neurons are in the same level, say at level 0, then we show that interestingly, also this type of state is *preserved*. That is, the neurons move "simultaneously" up towards level $k-1$, and only "tiny" bursts are observed. Once this level contains a sufficient number of neurons, a big burst starts – and it brings most neurons back to level 0; and a new cycle starts from the beginning. More precisely, we show the following.

[1] Actually, we use the analogy with the threshold phenomena in random graphs only to motivate the qualitative type of results that we should expect: for the actual proofs we rely on different techniques, see below.

THEOREM 1.2. *There exists a $c > 0$ such that the following is true. Suppose that $p = \beta k/n$, where*

$$\beta \geq \frac{2^{5/4}}{(k \ln k)^{1/4}} \cdot \left(1 + \frac{c}{\sqrt{\ln k}}\right).$$

Then for k sufficiently large, when starting with all neurons in level 0, the system converges to a stable state where bursts of size at least $(1-o(1)) \cdot n/2$ occur at least every kn time units.

In addition to this, we provide a very precise characterization of the stable state, and determine the asymptotic fraction of neurons that are involved in a big burst; see Section 4 for the technical details. Note that the above theorem only applies if $\beta < 1$ is not too small. The next results says that the lower bound from Theorem 1.2 is essentially sharp.

THEOREM 1.3. *There exists a $c > 0$ such that the following is true. Suppose that $p = \beta k/n$, where*

$$\beta \leq \frac{2^{5/4}}{(k \ln k)^{1/4}} \cdot \left(1 - \frac{c}{\sqrt{\ln k}}\right).$$

Then, for k sufficiently large, the following holds with high probability. If the system starts with all neurons in level 0, then there will be a finite number of big busts, and after that the system will remain in an asynchronous state.

Note that while we show that in the range "$\beta < 1$ and n, k not too small" both the uniform configuration and the "all-in first level" configuration are stable, we cannot confirm the observation from [6] that we should expect spontaneous transitions between these two extreme states. In fact we show, that once the system is in any of the two configurations, it stays there with very high probability. We assume that the observation in [6] is due to the fact that they ran their experiments for relatively small n. In fact, we believe that the transition between the two extreme states will actually happen due to *external* input during the interburst phase, which is an observation that opens up the possibility of interesting biological consequences. In particular, it seems that by combining external input with the property that neurons *tire* after periods of heavy firing allows to describe a system that can switch periodically between the two states.

Outline of the Paper In the remainder of this section we give a formal description of our system, and present the associated mean-field model. The mean-field model that we use is similar to the one derived in [6], but has a few differences that become important in our rigorous analysis. In Section 3 we present the crucial steps that are needed to show that the mean-field approximation captures the system dynamics with high probability. Finally, in Section 4.1 we analyze the mean-field behavior of the system and give the proofs of our main theorems.

1.1 Formal Description of the Model and the Mean-Field Equations

Let us introduce some notation before we actually describe the system dynamics. Let

$$\Omega = \left\{ \vec{x} = (x_0, \ldots, x_{k-1}, w) \mid x_i \geq 0 \text{ and } \sum_{i=0}^{k-1} x_i + w \leq 1 \right\}.$$

We will sometimes write $\vec{x} = (x_0, \ldots, x_{k-1})$, where we will implicitly assume that $w = 0$ *and* that $\sum_{i=0}^{k-1} x_i = 1$. Moreover, we will say that our system is state/configuration $\vec{x}$, if the number of neurons in level i equals $x_i n$, and the number of neurons that are waiting to fire (i.e., the neurons in the queue) is wn. Note that in the latter case the number of neurons that *have* already fired is precisely $(1 - \sum_{i=0}^{k-1} x_i)n$. Finally, we will denote by $B(\vec{x})$ the (random) size of the burst when the system starts in configuration $\vec{x}$.

The Interburst Phase Given the state $(x_0, \ldots, x_{k-1})$ of the system, define the family of random variables $X_0, \ldots, X_{k-1}$ by

$$\forall 0 \leq i \leq k-1 : X_i(t_0) = x_i n.$$

In the interburst phase the system dynamics are as follows. Let $J(t)$ be a random variable such that

$$\Pr[J(t) = i] = \frac{X_i(t_0 + t)}{n},$$

where

$$(1.1)\quad X_i(t_0 + t + 1) = \begin{cases} X_i(t_0 + t), & \text{if } J(t) \notin \{i-1, i, k-1\}, \\ X_i(t_0 + t) - 1, & \text{if } J(t) = i, \\ X_i(t_0 + t) + 1, & \text{if } J(t) = i-1. \end{cases}$$

Note that $X_i(t_0 + t + 1)$ is not defined if $J(t)$ is equal to $k-1$. Let

$$T = \min\{t : \ J(t) = k-1\}$$

be the first point in time where a neuron from level $k-1$ was selected. At time $T+1$ the system then is in the state $(X_0(t_0+T), \ldots, X_{k-1}(t_0+T)-1, 1)$, and switches to the burst phase described below.

The Burst Phase Given the state $(x_0, \ldots, x_{k-1}, w)$ of the system, define the family of random variables $X_0, \ldots, X_{k-1}, W$ by

$$\forall 0 \leq i \leq k-1 : X_i(0) = x_i n \ \text{ and } \ W(0) = wn.$$

Recall that during the burst phase a waiting neuron is selected and it fires, and thus promotes any neuron in the levels $0, \ldots, k-1$ with probability p. Hence, the number of promoted neurons from any level to the directly next level is binomially distributed. Set

$$\forall 0 \leq i \leq k-1 : Z_i(t) \sim \mathrm{Bin}\left(X_i(t), p\right).$$

Having this, we obtain the dynamics of the number of waiting neurons through the relation

$$W(t+1) = \begin{cases} 0, & \text{if } W(t) = 0 \\ W(t) - 1 + Z_{k-1}(t), & \text{otherwise.} \end{cases}$$

Informally, $W(t) + t$ is the *total* number of neurons that have been promoted from level $k-1$, where t is the number of neurons that have already fired yet. Moreover, the number of neurons in each level evolves as

$$(1.2)\quad \begin{aligned} &X_0(t+1) = X_0(t) - Z_0(t) \\ &X_i(t+1) = X_i(t) - Z_i(t) + Z_{i-1}(t), \forall 0 \leq i \leq k-1. \end{aligned}$$

The burst ends at time T, where

$$(1.3)\qquad T = \min\{t \ : \ W(t) = 0\}.$$

The Mean-Field Equations In this section we will present the mean-field description (i.e., how the system behaves “in average”) of our stochastic system. Due to the dynamics of the system this description contains two essential parts: the first part captures the typical evolution of the system in the asynchronous state, i.e., when no or just small bursts occur. One the other hand, the second part captures the behavior of the system while a big burst is taking place.

Suppose first that we are in the asynchronous state, and recall the dynamics given in (1.1). In order to obtain an intuition for the evolution of the system, suppose that we start at a point in time $\lfloor tn \rfloor$. How does the system look like at time $\lfloor (t+\varepsilon)n \rfloor$, for some arbitrarily small ε? As we are in the asynchronous state, we may assume that no or only very small bursts occur. In particular, if we assume that the effect of the bursts is negligible[2], then we obtain that the expected value of

[2]This assumption is unfortunately not completely true, although there are only very small bursts, as their *total* contribution might become big. Nevertheless, as we show below, our derived equations describe precisely the typical evolution of the system.

the number of neurons $X_i(t)$ in level i at time tn evolves roughly as

$$\mathbb{E}\left[X_i\left(t+\frac{1}{n}\right)\right] \approx \mathbb{E}\left[X_i(t)\right] - \frac{1}{n}\mathbb{E}\left[X_i(t)\right] + \frac{1}{n}\mathbb{E}\left[X_{i-1}(t)\right]$$

Noting that $X_i(t+\frac{1}{n}) \approx X_i(t) + \frac{1}{n}X_i'(t)$ we obtain that $\mathbb{E}\left[X_i'(t)\right] \approx -\mathbb{E}\left[X_i(t)\right] + \mathbb{E}\left[X_{i-1}(t)\right]$. Motivated by this we define the following system of differential equations. The *Merry-Go-Round Equation (MGR)* is given by the system

$$\begin{aligned} x_i'(t) &= x_{i-1}(t) - x_i(t) \text{ for } 1 \le i \le k-1, \\ x_0'(t) &= x_{k-1}(t) - x_0(t), \end{aligned} \tag{1.4}$$

and initial conditions $x_i(0) = x_i$ for $0 \le i \le k-1$. We will prove later (see Lemma 3.1) that this system approximates precisely the system dynamics in the asynchronous phase with high probability. Note that MGR has the "explicit" solution

$$x_i(t) = \sum_{j=0}^{k-1} x_j \cdot \Pr\left[\mathrm{Po}(t) \equiv i-j \pmod{k}\right] \tag{1.5}$$

for $i = 0, \ldots, k-1$. This is a fact that will become useful later.

Now suppose that the system is in the synchronous state, and recall the actual dynamics given by (1.2). If we denote by $X_i(t)$ the number of neurons at level i after tn neurons have fired, then we readily obtain for $1 \le i \le k-1$,

$$\mathbb{E}\left[X_i\left(t+\frac{1}{n}\right)\right] = \mathbb{E}\left[X_i(t)\right] - p\mathbb{E}\left[X_i(t)\right] + p\mathbb{E}\left[X_{i-1}(t)\right].$$

Again, by plugging into this the first two terms of the Taylor-series of X_i, we obtain with $p = \frac{\beta k}{n}$ that $\mathbb{E}\left[X_i'(t)\right] \approx -\beta k\mathbb{E}\left[X_i(t)\right] + \beta k\mathbb{E}\left[X_{i-1}(t)\right]$. As similar relation can be derived for X_0. This motivates the definition of the following system. The *Burst Equation (BST)* is the system of differential equations on $x_0, \ldots, x_{k-1}$ and the auxiliary variable x_k

$$\begin{aligned} x_0'(t) &= -x_0(t), \\ x_i'(t) &= x_{i-1}(t) - x_i(t) \text{ for } 1 \le i \le k-1, \\ x_k'(t) &= x_{k-1}(t), \end{aligned} \tag{1.6}$$

and the initial conditions $x_i(0) = x_i$ for $0 \le i \le k-1$ and $x_k(0) = w$. Note that we have omitted everywhere the factor βk in the definition of BST – this is done for solely technical reasons and a simple linear transformation of the solution gives us the solution to the actual system. In Section 3.2 we will show that BST approximates the actual system dynamics with high probability. Note that the BST system can be solved in closed form – we will come to this fact later when we need it. We will call the two systems MGR and BST the *mean-field* approximation of the actual system.

2 Preliminaries

One important tool in our analysis will be the following result by Seierstad [10], which strengthens Wormald's Differential Equation Method [12]. This powerful results state under which conditions certain stochastic processes can be approximated by suitably defined differential equations, and how good the obtained approximation indeed is. It will be a basic tool in showing that the mean-field approximation of our system is indeed a very good approximation.

Let us introduce some notation first. Let $(\Omega_n, \mathcal{F}_n, P_n)$ be a sequence of probability spaces. Let $m(n) = \mathcal{O}(n)$ be an integer function of n and suppose that a filtration $\mathcal{F}_{n,0} \subseteq \mathcal{F}_{n,1} \subseteq \cdots \subseteq \mathcal{F}_{n,m(n)} \subseteq \mathcal{F}_n$ exists for every $n \ge 1$. Let k be a fixed integer. For every $n \ge 1$ and $1 \le i \le k$ we consider random variables $(X_{n,m,i})_{0 \le m \le m(n)}$, such that $X_{n,m,i}$ is $\mathcal{F}(n,m)$-measurable. Finally, let $D \subseteq \mathbb{R}^k$, and define the *stopping time* T_D as the minimum value of m such that $n^{-1}X_{n,m,i} \notin D$, for some $1 \le i \le k$.

THEOREM 2.1. *Assume that there is a $C_0 > 0$ such that $X_{n,m,i} \le C_0 n$ for all n, $0 \le m \le m(n)$ and $1 \le i \le k$. Let $f_i : \mathbb{R}^k \to \mathbb{R}$ be functions and assume that the following conditions hold in some bounded connected open set D containing the closure of*

$$\big\{(z_1, \ldots, z_k) \mid \Pr\left[\forall 1 \le i \le k :\ X_{n,0,i} = z_i n\right] \ne 0 \text{ for some } n\big\}.$$

1. *For some functions $\beta = \beta(n) \ge 1$ and $\gamma = \gamma(n)$ with $\gamma = o(n^{-3/2})$ we have for all $1 \le i \le k$ and $1 \le m \le T_D$*
$$\Pr\left[|X_{n,m,i} - X_{n,m-1,i}| \le \beta \;\middle|\; \mathcal{F}_{m-1}\right] \ge 1 - \gamma.$$
2. *For some function $\lambda_1 = \lambda_1(n) = o(n^{-1/2})$ and all $1 \le i \le k$ and $1 \le m \le T_D$*
$$\Big|\mathbb{E}\left[X_{n,m,i} - X_{n,m-1,i} \;\middle|\; \mathcal{F}_{m-1}\right] - f_i(n^{-1}X_{n,m-1,1}, \ldots, n^{-1}X_{n,m-1,k})\Big| \le \lambda_1.$$
3. *The functions $(f_i)_{1 \le i \le k}$ are continuous and satisfy a Lipschitz condition.*

Then the following is true.

(a) For $(\hat{z}_1, \ldots, \hat{z}_k) \in D$ the system of differential equations
$$\frac{dz_i}{dt} = f_i(z_1, \ldots, z_k), \quad 1 \le i \le k$$
has a unique solution in D passing through $z_i(0) = \hat{z}_i$ for $1 \le i \le k$, which extends to points arbitrarily close to the boundary of D.

(b) *Let $\lambda > \lambda_1 + C_0\gamma n$ with $n^{-1/2} \ll \lambda \ll 1$. For some constants $C, d > 0$ and any function $\omega = \omega(n)$ tending to infinity, with probability $1 - \mathcal{O}(\gamma n + \exp(-d\frac{n\lambda^2}{\beta^2}) + n^2 \exp(-\frac{n^{1/4}}{\beta^3\omega}))$,*

$$X_{n,m,i} = nz_i(m/n) + \mathcal{O}(\lambda n)$$

uniformly for $0 \le m \le \sigma n \le m(n)$ and for each $1 \le i \le k$, where $\sigma = \sigma(n)$ is supremum of those m to which the solution can be extended before reaching within L_∞-distance $C\lambda$ of the boundary of D.

In Section 4, we will use the following upper bounds on the tail probabilities on Poisson variables $\mathrm{Po}(\lambda)$ with parameter λ.

LEMMA 2.1. *Let x, λ be two positive integers. For $x \ge \lambda$, it holds that*

$$\frac{x+1}{x+1-\lambda} \cdot \left(1 - \frac{\lambda^2}{(x+1)(x+1-\lambda)^2}\right) < \frac{\Pr\left(\mathrm{Po}(\lambda) \ge x\right)}{\Pr\left(\mathrm{Po}(\lambda) = x\right)} < \frac{x+1}{x+1-\lambda}.$$

If $x < \lambda$, the following bounds hold:

$$\frac{\lambda}{\lambda - x} \cdot \left(1 - \frac{x}{(\lambda-x)^2}\right) < \frac{\Pr\left(\mathrm{Po}(\lambda) \le x\right)}{\Pr\left(\mathrm{Po}(\lambda) = x\right)} < \frac{\lambda}{\lambda - x}.$$

We will also use the following Chernoff tail bounds. For $0 < \xi < \lambda$, we have

$$\begin{aligned} \Pr\left(\mathrm{Po}(\lambda) < \lambda - \xi\right) &< e^{-\xi^2/(2\lambda)}, \\ \Pr\left(\mathrm{Po}(\lambda) > \lambda + \xi\right) &< e^{-\xi^2/(3\lambda)}. \end{aligned} \tag{2.7}$$

3 The Mean Field Approximation

3.1 Merry Go Round – The Loading Phase

The main result of this subsection states that the MGR system approximates the behavior of the actual system well, provided that the $(k-1)$st level does not contain too many neurons. More precisely, we will show that whenever $x_{k-1} < (\beta k)^{-1}$, the system's behavior can be modeled by a subcritical branching process, as the *expected* number of neurons that are promoted from level $k-1$ to the waiting queue is in this case < 1.

LEMMA 3.1. *Let $(x_i(t))_{0 \le i \le k-1}$ be the solution of (1.4) with initial condition $x_i(0) = x_i$ for all $0 \le i \le k-1$. Let $0 < t_0 = \mathcal{O}(k)$ be such that $\max_{0 \le t < t_0} x_{k-1}(t) < (\beta k)^{-1} - \delta$, where $\delta > 0$, and $x_{k-1}(t_0) = (\beta k)^{-1} - \delta$. Then, with high probability there is a unique $t^* = t^*(t_0) > 0$ such that*

$$\forall 0 \le i \le k-1 : \; X_i(\lfloor t^* n \rfloor) = (1 + \Theta(n^{-1/3}))x_i(t_0)n,$$

and there was no burst larger than $\log^2 n$ until time $\lfloor t^* n \rfloor$.

Before we proceed with the proof of the above lemma, let us state an important implication for the limiting behavior of our actual system.

COROLLARY 3.1. *Suppose that the system starts in state $(x_0, \ldots, x_{k-1})$, where $x_{k-1}\beta k < 1$. Let $(x_i(t))_{0 \le i \le k-1}$ be the solution of (1.4) with initial condition $x_i(0) = x_i$ for all $0 \le i \le k-1$, and set $x_i^* = x_i(t^*)$, where $t^* = \min_t\{x_{k-1}(t)\beta k = 1\}$. Denote by N_i the number of neurons at the end of the asynchronous state. Then, if $t^* < \infty$, with high probability*

$$\forall 0 \le i \le k-1 \colon \lim_{n \to \infty} \frac{N_i}{n} = x_i^*.$$

In order to prove Lemma 3.1 we need a few observations about the probability that we observe a burst of a given size. The first statement characterizes the probability that such a "small" burst occurs, for any admissible $\vec{x}$ such that $x_{k-1} > 0$. Recall that we denote by W the set of waiting neurons, i.e., neurons that have been promoted from level $k-1$ and have not fired yet, and that we write $B(\vec{x})$ for the size of the burst if the system starts in state $(x_0 n, \ldots, x_{k-1}n, 1)$.

LEMMA 3.2. *Let $\vec{x}$ be such that $x_{k-1} > 0$. Then, for any $1 \le b \le n^{1/3}$, as $n \to \infty$*

$$\Pr\left[B(\vec{x}) = b \mid |W| = 1\right] = \left(1 + \mathcal{O}(n^{-1/6})\right) \cdot e^{-x_{k-1}\beta k b} \cdot \frac{(x_{k-1}\beta k b)^{b-1}}{b!}. \tag{3.8}$$

Proof. Let us begin with an auxiliary observation on the evolution of the number of neurons in level $k-1$. Suppose that the ith neuron in W fires, where $1 \le i \le b$. Then the number of neurons N_i that are promoted from level $k-1$ to W is distributed like $\mathrm{Bin}(Y_i, p)$, where Y_i is the number of neurons on level $k-1$ the moment in time the ith neuron fires. We write $Y_i = x_{k-1}n + Z_1 + \cdots + Z_{i-1} - L_{i-1}$, where Z_j denotes the number of neurons that were promoted from level $k-2$ to level $k-1$ due to the firing of the jth neuron in W, and L_{i-1} is the number of neurons promoted from level $k-1$ to W due to the firing of $i-1$ neurons. Note that the Z_j's are dominated by independent random variables distributed like $\mathrm{Bin}(n, p)$, as there are trivially at most n neurons in level $k-2$. Moreover, for all admissible i we have $0 \le L_{i-1} \le b$. Now, a straightforward application of the Chernoff bounds yields with plenty of room to spare that for all $1 \le i < b$, with probability at least $1 - e^{-\Omega(n^{1/2})}$, it holds $|Y_i - x_{k-1}n| \le n^{1/2}$. This implies that

$$\Pr\left[\forall 1 \le i \le b \colon Y_i = \left(1 + \mathcal{O}(n^{-1/2})\right)x_{k-1}n\right] \ge 1 - e^{-\Omega(n^{1/2})}. \tag{3.9}$$

Let us now turn to the proof of (3.8). We will write $B(\vec{x}) = \sum_{p\geq 0} \widetilde{W}_p$, where $\widetilde{W}_0 = |W| = 1$, and $\widetilde{W}_p$ is the number of waiting neurons after $\widetilde{W}_{p-1}$ neurons have been processed. In other words, we split up the burst in *phases*, where the pth phase begins when all neurons of the $(p-1)$st phase have been processed, and denote by $\widetilde{W}_p$ the number of neurons that are waiting at the beginning of phase p. Note that the number of phases P^* is at most b, as the burst ends as soon as $\widetilde{W}_i = 0$ for some i.

Let $\delta = \delta(n) = \mathcal{O}(n^{-1/2})$. Assume that for any $0 \leq w_{i+1} \leq b$ we could show

$$\text{(3.10)} \qquad \Pr\left[\widetilde{W}_{i+1} = w_{i+1} \mid \widetilde{W}_i\right] \leq$$

$$\left(1 + \mathcal{O}\Big(\frac{w_{i+1}^2}{n}\Big)\right) e^{-(x_{k-1}+\delta)\beta k \widetilde{W}_i} \frac{((x_{k-1}+\delta)\beta k \widetilde{W}_i)^{w_{i+1}}}{w_{i+1}!}.$$

We shall prove this fact later. Then the proof of the upper bound in the lemma completes as follows. Let

$$\mathcal{W}_b = \Big\{(w_1, \dots, w_\ell) \;\Big|\; \sum_{i=1}^{\ell} w_i = b-1 \;\wedge\; \ell \in \mathbb{N} \;\wedge\; \forall j \in [\ell] : w_i \in \mathbb{N}.\Big\}.$$

Denote for a sequence $\vec{w} \in \widetilde{\mathcal{W}_b}$ by $\ell(\vec{w})$ the index of the *last* (non-zero) entry in $\vec{w}$. Clearly,

$$\Pr[B(\vec{x}) = b \mid |W| = 1] = \sum_{\vec{w}\in\mathcal{W}_b} \Pr[\widetilde{W}_1 = w_1, \widetilde{W}_2 = w_2, \dots, \widetilde{W}_{\ell(\vec{w})} = w_{\ell(\vec{w})}, \widetilde{W}_{\ell(\vec{w})+1} = 0].$$

Abbreviate $\xi = (x_{k-1}+\delta)\beta k$, and $w_0 = 1$, and note that by using (3.10) we obtain

$$\begin{aligned}
\Pr[\widetilde{W}_1 = w_1, \dots, \widetilde{W}_{\ell(\vec{w})+1} = 0] &= \prod_{i=0}^{\ell(\vec{w})} \Pr[\widetilde{W}_{i+1} = w_{i+1} \mid \widetilde{W}_i = w_i] \\
&\leq \left(1 + \mathcal{O}\Big(\frac{b^2}{n}\Big)\right) \cdot \prod_{i=0}^{\ell(\vec{w})} e^{-\xi w_i} \frac{(\xi w_i)^{w_{i+1}}}{w_{i+1}!} \\
&= \left(1 + \mathcal{O}\Big(\frac{b^2}{n}\Big)\right) \cdot e^{-\xi b} \xi^{b-1} \cdot \prod_{i=0}^{\ell(\vec{w})-1} \frac{w_i^{w_{i+1}}}{w_{i+1}!}.
\end{aligned}$$

The proof of the upper bound completes with our assumption $\delta = \mathcal{O}(n^{-1/2})$ and the observation

$$\begin{aligned}
\sum_{\vec{w}\in\mathcal{W}_b} \prod_{i=0}^{\ell(\vec{w})-1} \frac{w_i^{w_{i+1}}}{w_{i+1}!} &= \frac{1}{(b-1)!} \sum_{\vec{w}\in\mathcal{W}_b} \binom{b-1}{w_1, \dots, w_{\ell(\vec{w})}} \prod_{i=0}^{\ell(\vec{w})-1} w_i^{w_{i+1}} \\
&= \frac{b^{b-2}}{(b-1)!},
\end{aligned}$$

as the sum in the middle counts precisely the number of labeled trees on b vertices. To see this, let T be any tree on b vertices, and let $L_d(T)$ be the set of vertices that have distance d from the vertex with label 1. So, $L_0(T) = 1$, $L_1(T)$ contains all neighbors of 1, and so on. The number of trees is then the number of ways to choose the sets $L_1, L_2, \dots$, i.e., $\binom{b-1}{|L_1|,\dots,|L_x|}$, for some $x \geq 1$, times the number of ways to connect the vertices with distance $d+1$ to the vertices with distance d, i.e. $|L_d|^{|L_{d+1}|}$, for $d < x$.

We finally prove the last missing statement (3.10), i.e., we estimate the probability of the event "$\widetilde{W}_{i+1} = w_{i+1}$", assuming the quantity $\widetilde{W}_i$. Recall that the $\widetilde{W}_i$ neurons fire sequentially and *independently* of each other. This means that there are non-negative integers $j_1, \dots, j_{\widetilde{W}_i}$ such that $\sum_{s=1}^{\widetilde{W}_i} j_s = w_{i+1}$, and the sth neuron that fired initiated j_s neurons from level $k-1$ to join the waiting queue. By exploiting (3.9), the probability for this events can be bounded from above by at most

$$\begin{aligned}
&\text{(3.11)} \\
&\Pr\left[\widetilde{W}_{i+1} = w_{i+1} \mid \widetilde{W}_i\right] \\
&\leq \sum_{\sum_{s=1}^{\widetilde{W}_i} j_s = w_{i+1}} \prod_{s=1}^{\widetilde{W}_i} \Pr\left[\operatorname{Bin}\left((x_{k-1}+\delta)n, p\right) = j_s\right] \\
&\qquad + \Pr[\exists 1 \leq i \leq b \colon Y_i = \big(1 + \mathcal{O}\big(n^{-7/9}\big)\big) x_{k-1} n] \\
&= \sum_{\sum_{s=1}^{\widetilde{W}_i} j_s = w_{i+1}} \prod_{s=1}^{\widetilde{W}_i} e^{-\Theta(\frac{j_s^2}{n})} \cdot e^{-(x_{k-1}+\delta)\beta k} \frac{((x_{k-1}+\delta)\beta k)^{j_s}}{j_s!} \\
&\qquad + e^{-\Omega(n^{2/9})} \\
&= \left(1 + \mathcal{O}\Big(\frac{w_{i+1}^2}{n}\Big)\right) \cdot e^{-(x_{k-1}+\delta)\beta k \widetilde{W}_i} \\
&\qquad \cdot ((x_{k-1}+\delta)\beta k)^{w_{i+1}} \cdot \sum_{\sum_{s=1}^{\widetilde{W}_i} j_s = w_{i+1}} \prod_{s=1}^{\widetilde{W}_i} \frac{1}{j_s!}.
\end{aligned}$$

Note that the last term equals the w_{i+1}th coefficient in the Taylor expansion of the function $e^{\widetilde{W}_i z}$ around $z = 0$. Putting everything together yields (3.10), as desired.

To obtain the lower bound in (3.8) observe that we only have to show a corresponding lower bound for (3.10). A closer look at the calculation above reveals that this is easily obtained: replace $x_{k-1}+\delta$ by $x_{k-1}-\delta$, and reverse all inequalities.

A final ingredient for the proof of Lemma 3.1 is the following statement, which says that big bursts are highly improbable whenever $x_{k-1} < (\beta k)^{-1}$. Note that this does not follow directly from Lemma 3.2, as the error term $\mathcal{O}(n^{-2/3})$ is too big for that purpose.

LEMMA 3.3. *Let $\vec{x}$ be such that $x_{k-1}\beta k = 1-\delta$, where $\delta = \delta(n) > 0$. Then, for any $1 \ll b \ll n$*

$$\Pr\left[B(\vec{x}) \geq b \mid |W| = 1\right] \leq e^{-\frac{1}{5}\delta^2 b}. \tag{3.12}$$

Proof. Suppose the the ith neuron in W fires, where $i < b$. Then the number of neurons N_i that are promoted from level $k-1$ to W is distributed like $\mathrm{Bin}\,(Y_i, p)$, where Y_i is the number of neurons on level $k-1$ the moment in time the ith neuron fires. Note that $Y_i \leq x_{k-1}n + Z_1 + \cdots + Z_{i-1}$, where Z_j denotes the number of neurons that were promoted from level $k-2$ to level $k-1$ due to the firing of the jth neuron in W.

First, note that the Z_j's are dominated by independent random variables distributed like $\mathrm{Bin}\,(n,p)$, as there are trivially at any point in time at most n neurons in level $k-2$. Hence, by a straightforward application of the Chernoff bounds we obtain with plenty of room to spare that for all $1 \leq i < b$ we have with probability at least $1 - e^{-\Theta(b)}$ that $Y_i \leq x_{k-1}n + 2\beta k b = (1+o(1))x_{k-1}n$.

Second, the above observation implies that with high probability the N_i's are dominated by independent $\mathrm{Bin}\,((1+o(1))x_{k-1}n, p)$ variables. Moreover, note that the event "$B(\vec{x}) \geq b$" implies that "$\sum_{i=1}^{b} N_i \geq b$". Again, the Chernoff bounds yield for large b that

$$\begin{aligned}&\Pr[B(\vec{x}) \geq b] \leq \\ &\qquad \Pr[\mathrm{Bin}\,((1+o(1))x_{k-1}n \cdot b, p) \geq b] \leq e^{-\frac{1}{4}\delta^2 b}.\end{aligned}$$

The proof is completed.

We are now ready to prove Lemma 3.1. Here we will in particular exploit a result by Seierstad [10], which strengthens Wormald's Differential Equation Method [12].

Proof. [Proof of Lemma 3.1] First, suppose that there was no neuron firing, i.e., $p = 0$. Then, the expectation of the number of neurons $X_i(\lfloor tn \rfloor)$ at level i after $\lfloor tn \rfloor$ time steps would be given, up to lower order terms, by the right-hand of (1.5), multiplied by n. By the virtue of the Chernoff bounds, the X_i's are sharply concentrated around this values in this very special case.

Now suppose that $p = \frac{\beta k}{n}$. Let us first consider the expected size of a burst at time for any given $\vec{x}$ such that $x_{k-1}\beta k < 1$, and initially $|W| = 0$. Note that with probability x_{k-1} we have $|W| = 1$, and otherwise the expected size of the burst is 0. By applying Lemma 3.2 and Lemma 3.3 we obtain with $\xi = \xi(\vec{w}) = x_{k-1}\beta k$

$$\mathbb{E}[B(\vec{x})] = x_{k-1} \cdot \sum_{b=1}^{\log^2 n} b \cdot e^{-\xi b}\frac{(\xi b)^{b-1}}{b!} + \mathcal{O}(n^{-1/6}).$$

Observe that $T'(x) = e^{T(x)} + xT'(x)e^{T(x)}$, which implies that $T'(x) = \frac{T(x)}{x(1-T(x))}$. We obtain

$$\begin{aligned}\mathbb{E}[B(\vec{x})] &= x_{k-1} \cdot e^{-\xi}T'(\xi e^{-\xi}) + \mathcal{O}(n^{-1/6}) \\ &= x_{k-1} \cdot e^{-\xi}\frac{T(\xi e^{-\xi})}{\xi e^{-\xi}(1-T(\xi e^{-\xi}))} + \mathcal{O}(n^{-1/6}) \\ &= \frac{x_{k-1}}{1-\xi} + \mathcal{O}(n^{-1/6}).\end{aligned} \tag{3.13}$$

Let $B(\tau)$ denote the size of the burst that occurred at time $0 \leq \tau \leq t$ ($B(\tau) = 0$ is admissible here), where with "*time*" we denote the number of neuron promotions that were initiated from the outside impulse, i.e. the time does not increment when neurons fire. In particular, any burst is considered as taking place in a single unit of time. Moreover, let us denote by $Y_i(\tau)n$ the number of neurons on level i at time τ. Note that if $\mathbf{Y}(\tau) = (Y_0(\tau), \ldots, Y_{k-1}(\tau))$ is such that $\beta k Y_{k-1}(\tau) < 1$, then for $1 \leq i \leq k-2$

$$\begin{aligned}&\mathbb{E}\left[Y_i(\tau+1) \mid \mathbf{Y}(\tau)\right] = \\ &\qquad Y_i(\tau) + \big(1 + pn\mathbb{E}\left[B(\tau)\right]\big)(Y_{i-1}(\tau) - Y_i(\tau)) \\ &\qquad +\mathcal{O}\big(p^2 n\mathbb{E}\left[B(\tau)\right]\big),\end{aligned}$$

which implies with (3.13) for large n

$$\begin{aligned}&\mathbb{E}\left[Y_i(\tau+1) - Y_i(\tau) \mid \mathbf{Y}(\tau)\right] = \\ &\qquad \left(1 + \frac{\beta k Y_{k-1}(\tau)}{1-\beta k Y_{k-1}(\tau)}\right)(Y_{i-1}(\tau) - Y_i(\tau)) + \mathcal{O}(n^{-1/6}).\end{aligned}$$

Similarly we obtain for $i = 0$

$$\begin{aligned}&\mathbb{E}\left[Y_0(\tau+1) - Y_0(\tau) \mid \mathbf{Y}(\tau)\right] = \\ &\qquad -\left(1 + \frac{\beta k Y_{k-1}(\tau)}{1-\beta k Y_{k-1}(\tau)}\right)Y_0(\tau) \\ &\qquad +\frac{Y_{k-1}(\tau)}{1-\beta k Y_{k-1}(\tau)} + \mathcal{O}(n^{-1/6}),\end{aligned}$$

and for $i = k-1$

$$\mathbb{E}\left[Y_{k-1}(\tau+1) - Y_{k-1}(\tau) \mid \mathbf{Y}(\tau)\right] = \left(1 + \frac{\beta k Y_{k-1}(\tau)}{1-\beta k Y_{k-1}(\tau)}\right) Y_{k-2}(\tau) - \frac{Y_{k-1}(\tau)}{1-\beta k Y_{k-1}(\tau)} + \mathcal{O}(n^{-1/6}).$$

Based on the above relations define the system of differential equations with unknown functions $(y_i(\tau))_{0\le i\le k-1}$

$$y_i' = f_i(y_0, \ldots, y_{k-1}), \tag{3.14}$$

where

$$f_i(y_0, \ldots, y_{k-1}) = \frac{y_{i-1} - y_i}{1 - \beta k y_{k-1}},$$

where we wrote for brevity $y_{-1} = y_{k-1}$. Moreover, note that Lemma 3.3 implies with room to spare

$$\Pr[B(\tau) \ge n^{1/20} \mid \beta k Y_{k-1}(\tau) < 1] \le e^{-n^{1/21}}.$$

Finally, let

$$D = \Big\{(y_0, \ldots, y_{k-1}) \;\big|\; \forall 0 \le i \le k-1 : y_i \ge 0, \sum_{i=0}^{k-1} y_i = 1,\ \beta k y_{k-1} < 1\Big\}.$$

Now, using all above facts we easily check that the conditions of Theorem 2.1 are fulfilled if we choose $X_{n,m,i} = Y_i(m)$, $\beta = n^{1/20}$, $\gamma = e^{-n^{1/21}}$, and $\lambda_1 = \mathcal{O}(n^{-1/6})$. Hence, we obtain for any $n^{-1/2} \ll \lambda \ll 1$ such that $\lambda \gg \lambda_1 + n\gamma$, say $\lambda = n^{-1/7}$, that with probability at least $1 - e^{-n^{\Theta(1)}}$ it holds

$$Y_i(\tau) = n y_i(\tau/n) + \mathcal{O}(n^{6/7}).$$

for all τ such that $\mathcal{O}(n^{-1/3}) \le y_{k-1}(\tau/n) \le 1 - \mathcal{O}(n^{-1/3})$, where the y_i are given by the solution to (3.14), with initial condition $y_i(0) = x_i$.

The above result describes precisely the evolution of the Y_i over time. But what is the connection to the solution of the simple system (1.4)? To establish this connection, let $x_0, \ldots, x_{k-1}$ be the solution of (1.4), and set

$$x_i(t) = y_i(h(t)),$$

where $h(t)$ is an unknown function. Then we easily see that $x_i'(t) = h'(t) y_i'(h(t))$, and therefore $h(t)$ satisfies the relations

$$h'(t) = 1 - \beta k x_{k-1}(t) = 1 - \beta k y_{k-1}(h(t)).$$

Note that h' is positive for all t such that $x_{k-1}(t)\beta k < 1$, and hence its inverse function $h^{-1}(t)$ is well-defined for all such t. This yields that

$$x_i(h^{-1}(\tau)) = y_i(\tau),$$

and the proof is completed if we choose τ by solving the equation $t = h^{-1}(\tau)$.

3.2 The Beginning of the Burst

The first result in this section gives us information about the evolution of the number of neurons in each level during a burst.

LEMMA 3.4. *Suppose that the system is in state $(x_0, \ldots, x_{k-1}, w)$. For any $t \le T$, where T is given in (1.3)*

$$\forall 0 \le i \le k-1 : \mathbb{E}[X_i(t)] = (1 + o(1)) x_i(pt) n, \tag{3.15}$$

where the X_i are given in (1.2), and the x_i are the solution to (1.6). Furthermore, for any $0 \le \varepsilon = \varepsilon(n) \le 1$

$$\Pr\left[|X_i(t) - \mathbb{E}[X_i(t)]| \ge \varepsilon \mathbb{E}[X_i(t)]\right] \le e^{-\varepsilon^2 n^{1/3}}.$$

Proof. The definition of the X_i's imply for any $t < T$ and $1 \le i \le k-1$

$$\mathbb{E}[X_i(t+1)] = \mathbb{E}[X_i(t)] - p\mathbb{E}[X_i(t)] + p\mathbb{E}[X_{i-1}(t)]$$

and

$$\mathbb{E}[X_0(t+1)] = \mathbb{E}[X_0(t)] - p\mathbb{E}[X_0(t)].$$

From this and the fact $T \le n$ we readily deduce that

$$\mathbb{E}[X_0(t+1)] = (1-p)^t x_0 n = (1 + \mathcal{O}(n^{-1})) x_0(pt) n,$$

as by solving (1.6) we obtain $x_0(z) = e^{-z} x_0$. Then by applying induction over k we obtain

$$\mathbb{E}[X_i(t+1)] = (1-p)\mathbb{E}[X_i(t)] - (1 + \mathcal{O}(n^{-1})) x_{i-1}(pt) n,$$

or equivalently,

$$\begin{aligned}
\mathbb{E}[X_i(t)] &= (1-p)^t \mathbb{E}[X_i(0)] + (1 + \mathcal{O}(n^{-1})) \cdot \sum_{j=0}^{t-1} (1-p)^{t-1-j} x_{i-1}(pj) \\
&= (1 + \mathcal{O}(n^{-1})) \cdot \left(e^{-pt} x_i + \int_0^t e^{p(z-t)} x_{i-1}(pz)\,dz\right) n \\
&= (1 + \mathcal{O}(n^{-1})) x_i(pt) n,
\end{aligned}$$

where the last step follows from $x_i(z) = e^{-z} x_i + \int_0^z e^{\xi - z} x_{i-1}(\xi)\,d\xi$. This completes the proof of (3.15). To show the concentration result, first note that the number of neurons that are promoted due to a firing neuron is dominated from above by a $\mathrm{Bin}(n,p)$ variable. This implies that with probability at least $1 - e^{-n^{1/3}}$ the number of promoted neurons is at most $n^{1/3}$, for any $t \le T \le n$. The proof completes with an application of the Azuma-Hoeffding inequality.

By exploiting the knowledge about the number of neurons in each level during a burst, we now are able to infer information about the actual size of the burst.

LEMMA 3.5. *Suppose that the system is in state* $(x_0, \dots, x_{k-1}, w)$*, where* $w > 0$*. Let* $s = 1 - (\sum_{i=0}^{k-1} x_i + w)$*, and set*

$$\tau = \min\{t > 0 :\ t - x_k(\beta k t) = s\}, \tag{3.16}$$

where x_k *is given in* (1.6)*, and* $x_k(0) = w$*. Then, the system will have with high probability a burst of size* T*, where* $|T - \tau n| = o(n^{4/5})$*.*

Proof. First, note that for any $t \le T$ we have

$$n = t + W(t) + \sum_{i=0}^{k-1} X_i(t) \overset{W(T)=0}{\Longrightarrow} n = T + \sum_{i=0}^{k-1} X_i(T).$$

By applying Lemma 3.4 for $t = T$ we obtain that $X_i(T) = (1 + o(n^{-1/5}))x_i(pT)n$ is true with high probability, for all $0 \le i \le k-1$. If we write $T = \alpha_T n$, then the above relation becomes

$$1 = \alpha_T + (1 + o(n^{-1/5})) \sum_{i=0}^{k-1} x_i(\beta k \alpha_T). \tag{3.17}$$

Note that for all z

$$\sum_{i=0}^{k} x_i(z) = \sum_{i=0}^{k-1} x_i + w = 1 - s$$
$$\Longrightarrow \sum_{i=0}^{k-1} x_i(z) = 1 - s - x_k(z),$$

where x_k is given in (1.6). The proof finishes by plugging this into (3.17).

Our final task in this section is to show under which conditions we will arrive in a state in which a big burst will occur. The next lemma answers this question.

LEMMA 3.6. *Let* $\vec{x}$ *be such that* $x_{k-2}\beta k > 1$ *and* $x_{k-1}\beta k \ge 1 - \varepsilon$*, for some arbitrarily small* $\varepsilon > 0$*. Then there are functions* $(\varepsilon_i(\varepsilon))_{1 \le i \le 4}$ *such that* $e_i \to 0$ *when* $\varepsilon \to 0$ *with the following properties. Suppose that the system starts in* $(x_0, \dots, x_{k-1})$*. Then, with high probability, the system will arrive after at most* $\varepsilon_1 n$ *time steps in a state* $(x'_0, \dots, x'_{k-1}, w)$*, such that*

- $x'_i = x_i \pm \varepsilon_2$*,*
- $w = \Theta(\varepsilon_3)$*, and*
- *there was no burst of size larger than* $\varepsilon_4 n$*.*

In words, the system will arrive in a configuration that will generate a big burst, if x_{k-1} is arbitrarily close to $(\beta k)^{-1}$, and simultaneously $x_{k-2} > (\beta k)^{-1}$. Note that in such a configuration the expected number of neurons that are promoted when a neuron fires from level $k-1$ is very close to one, while the expected number of neurons that are promoted from level $k-2$ to level $k-1$ is > 1. Hence the number of neurons in level $k-1$ increases quickly, and the big bursts gains unbounded momentum.

4 Thresholds for Synchronous and Asynchronous Behavior

In this section, we analyze the bursting behavior of the neural network model introduced in Section 1 by using the mean field approximation developed in Section 3. For simplicity, we ignore low probability events in the lemmas of Section 3 and make deterministic statements by assuming that the system behaves exactly as specified by the mean field approximation.

The mean field system has two phases. The *loading phase* that is described by the MGR differential equation (1.4) and the *burst phase* that behaves according to the BST differential equation (1.6). We assume that the system starts in the loading phase, i.e., we assume that initially $x_{k-1} < 1 - \delta$ for some $\delta > 0$ and $w = 0$. By Lemma 3.1, in this setting, the MGR equation describes the system well. Specifically, for every t, there is a 'real' time τ such that $\vec{x}(t)$ approximates the system at time τ arbitrarily closely. The solution of the MGR equation is given by Equation (1.5). We will sometimes argue about the contribution of only x_0 (the initial level 0) to the state at time t in the loading phase. Let $\vec{g}(t) = (g_1(t), \dots, g_{k-1}(t))$ denote this contribution.[3] For $i = 0, \dots, k-1$, we have

$$g_i(t) = x_0(0) \cdot \Pr\left(\mathrm{Po}(t) \equiv i \pmod{k}\right). \tag{4.18}$$

We set

$$\Omega_{\mathrm{BB}} = \{\vec{x} \in \Omega : kx_{k-1}\beta = 1, x_{k-2} > x_{k-1}\}$$

to define the set of states when the system changes from the loading phase to the burst phase.[4] Technically, by Lemma 3.6, the burst starts when $x_{k-2} > 1/(\beta k)$ and $x_{k-1} \ge 1 - \varepsilon$ for a sufficiently (arbitrarily) small $\varepsilon > 0$. For simplicity, we ignore this technicality and assume

[3] The letter 'g' stands for 'Gaussian wave' as the distribution of mass on the levels at time t is essentially Gaussian distributed with mean t.

[4] To cover all cases, Ω_{BB} should also contain those $\vec{x} \in \Omega$ with $kx_{k-1}\beta = 1$ and, for some $s \ge 2$, $x_{k-1} = x_{k-2} = \ldots = x_{k-s}$ and $x_{k-s-1} > x_{k-1}$. We do not need them and thus do not discuss them in this extended abstract.

that the burst phase starts as soon as the loading phase reaches a state in Ω_{BB}. All proofs of this section can be easily adapted by adding small error terms at a number of places.

By Lemma 3.4, during the burst phase, the system can accurately by described by the BST equation (1.6). The burst stops as soon as $x_k(t_B) = t_B/(\beta k)$ where t_B is the time since the start of the burst. Note that the size of a burst that finishes in time t_B is $t_B\beta k$ (cf. Lemma 3.5). Assume that there is a burst of size B, let $\vec{y} = (y_0, \ldots, y_{k-1}) \in \Omega_{\text{BB}}$ be the state at the beginning of the burst (i.e., at the end of the loading phase), and let $\vec{z} = (z_1, \ldots, z_{k-1})$ be the sate at the end of the burst before the next loading phase starts. For $i = 0, \ldots, k-1$, the vector $\vec{z}$ after the burst can be computed as

(4.19)
$$z_i = (B, 0, 0, \ldots, 0) + \sum_{j=0}^{k-1} y_j \cdot \Pr\left(\text{Po}(B \cdot \beta k) = j - i\right).$$

When analyzing the mean field behavior of our neural network system, we will need the following definition.

Definition 4.1. *We call a non-negative vector $\vec{x} \in \mathbb{R}^k$* safe *if the MGR with initial condition $\vec{x}(0) = \vec{x}$ has $x_{k-1}(t) \cdot \beta k \leq 1/2$ for* all $t \geq 0$.

Before studying the more interesting $\beta < 1$ case, let us first prove Theorem 1.1.

Proof. [Proof Idea] Note that if $\beta > 1$, the average x_i is larger than $1/(\beta k)$. This together with Equation (1.5) implies that regardless of the starting configuration, the loading phase will always reach a state in Ω_{BB} within a finite amount of time. Whenever the state of the system is in Ω_{BB}, there is a burst of size $\Theta(n)$.

4.1 Synchronous Behavior for $\beta < 1$

The main objective of this section is to prove Theorem 1.2, i.e., we will show that for β significantly smaller than 1, the system has a stable synchronous state. Throughout the section, we denote the initial state of the system by $\vec{x} = (x_0, \ldots, x_{k-1})$ and the system after time t in the loading phase by $\vec{x}(t)$, i.e., $\vec{x}(0) = \vec{x}$. Note that we measure time in the loading phase according to the MGR equation. The system remains in the loading phase until time t^* where t^* is the earliest time such that $\vec{x}(t^*) \in \Omega_{\text{BB}}$. At time t^*, there is a big burst. We denote the state of the system before the burst starts by $\vec{y} = (y_0, \ldots, y_{k-1})$, i.e., $\vec{y} = \vec{x}(t^*)$. Further, we denote the state after the burst at the beginning of the next loading phase by $\vec{z} = (z_0, \ldots, z_{k-1})$.

Before proving Theorem 1.2, we give a short overview of the main ideas. Assume that the system starts with all neurons in level 0. Then during the loading phase, all the neurons are gradually shifted to higher levels. While shifting the neurons, they are spread across a set of levels. At time t of the MGR equation, the distribution of the neurons are essentially Gaussian distributed with expectation and variance t. The system switches from the loading phase to the burst phase when it reaches a state in Ω_{BB}. At this time, the right tail of the Gaussian distribution has already (cyclically) moved from high levels back to level 0. Most of the remaining mass will be part of the first burst. During the burst, the mass from the tail that is in levels $0, \ldots$ at the beginning of the burst is moved to levels around βk. In every burst, the mass of the right tail of the big 'Gaussian wave' initiating the burst is moved to levels around βk and the tail neurons of the previous burst are promoted by roughly βk levels. When the system stabilizes, at the beginning of the loading phase, most of the mass is in level 0 and there is some additional mass around levels $i\beta k$ for $i = 1, 2, \ldots$. We formally define such a state (called a nice state) in Definition 4.2 and show that if the loading phase is started in a nice state, the state after the next burst will be nice again (according to the mean field approximation).

For simplicity, we will assume that $\beta \leq 1$ throughout the following considerations. Note that a much simpler argument is sufficient to show an analogous statement to Theorem 4.1 for $\beta > 1$.

For $i \in \{0, \ldots, k-1\}$, we define $\eta_i = i/k^4$ and we call $\vec{\eta} = (\eta_0, \ldots, \eta_{k-1})$ the noise vector. For a system state $\vec{x} = (x_0, \ldots, x_{k-1}) \in \mathbb{R}^k$, let $\vec{x}' = (x'_0, \ldots, x'_{k-1})$ be the vector that is obtained by setting $x'_0 = 0$ and $x'_i = \max\{0, x_i - \eta_i\}$ for $i > 0$. Hence, we essentially subtract the noise vector from the given state.

Definition 4.2. *We call a state $\vec{x} = (x_0, \ldots, x_{k-1}) \in \mathbb{R}^k$ ε-nice for a value $\varepsilon > 0$ if there is an integer $\ell \geq 1$ such that the mass in $\vec{x}'$ can be partitioned into ℓ vectors $\vec{x}'_1, \ldots, \vec{x}'_\ell$ where $\vec{x}'_i = (x'_{i,0}, \ldots, x'_{i,k-1})$ such that $\vec{x}' = \vec{x}'_1 + \cdots + \vec{x}'_\ell$ and the following conditions are satisfied.*

1. *There are positive integers $0 < r_1 < \cdots < r_\ell \leq k$ such that for $i \geq 1$,*

(4.20)
$$\begin{aligned} (1-\varepsilon)\beta k \;&\leq\; r_1 \;\leq\; \beta k \\ \textit{and} \quad (1-\varepsilon)\beta k \;&\leq\; r_{i+1} - r_i \;\leq\; \beta k \end{aligned}$$

and such that for all $i \in [\ell]$, $x'_{i,j} = 0$ for all $j < r_i - \Delta_i$ and for all $j > r_i + \Delta_i$, where $\Delta_i = 12 \cdot i \cdot \sqrt{k \ln k}$.

2. The total mass in every vector $\vec{x}'_i$ is at most

$$\forall i \in [\ell] : \sum_{j=0}^{k-1} x'_{i,j} = \sum_{j=r_i-\Delta_i}^{r_i+\Delta_i} x'_{i,j} \;\le\; \frac{\sqrt{2}+\frac{D}{\sqrt{\ln k}}}{\beta \cdot \sqrt{k \ln k}}. \tag{4.21}$$

where $D = D(\varepsilon) \ge 0$ is a suitably chosen constant.[5]

3. The vector $\vec{w} = (w_0, \ldots, w_{k-1})$ with $w_0 = 0$ and $w_i = x_i$ for $i > 0$ is safe. Further, for $i > r_1 + \Delta_1$, $x'_i \le 1/(3\beta k)$.

Note that the vector $\vec{x} = (1, 0, 0, \ldots, 0)$ is clearly ε-nice for every $\varepsilon > 0$. Based on Definition 4.2, we can state our the main technical theorem.

Theorem 4.1. *Let $\varepsilon > 0$ be a positive constant and assume that*

$$\beta \ge \frac{2^{1/4}}{\sqrt{\varepsilon(1-\varepsilon)} \cdot (k \ln k)^{1/4}} \cdot \left(1 + \frac{c}{\sqrt{\ln k}}\right) \tag{4.22}$$

for a sufficiently large constant c. Then, for k sufficiently large, when starting in an ε-nice state at time 0, the system stays in the loading phase until time $t^* = k - o(k)$. At time t^*, there is a burst of size at least $1 - \varepsilon$ such that after the burst, the system is back in an ε-nice state.

Note that Theorem 1.2 directly follows as a corollary of Theorem 4.1. Interestingly, Theorem 4.1 indicates that in order to obtain a stable synchronous system with bursts of size $b < 1/2$, the value of β has to be larger than for a system with bursts of size $1/2$. In order to prove Theorem 4.1, we need a bunch of technical lemmas.

Lemma 4.1. *For $\varepsilon > 0$, let $\vec{x} = (x_0, \ldots, x_{k-1})$ be an ε-nice state. If β is chosen as in Theorem 4.1 (Inequality (4.22)), for k sufficiently large, $\sum_{i=1}^{k-1} x_i < \varepsilon - 1/k^{1/4}$ and thus $x_0 \ge 1 - \varepsilon + 1/k^{1/4}$.*

Proof. By the construction of the vectors $\vec{\eta}$ and $\vec{x}'$, we have $\sum_{i=1}^{k-1} x_i \le \sum_{i=0}^{k-1} \eta_i + x'_i$ as well as $\sum_{i=0}^{k-1} \eta_i \le 1/(2k^2)$. It is therefore sufficient to show that $\sum_{i=0}^{k-1} \vec{x}'_i < \varepsilon - 1/k^{1/4} - 1/(2k^2)$ for k sufficiently large.

Let us consider the number ℓ of vectors $\vec{x}'_i$ into which $\vec{x}'$ is partitioned. Clearly, $r_\ell < k + 12 \cdot \ell \cdot \sqrt{k \ln k}$. Because by Definition 4.2, $r_1 \ge (1-\varepsilon)\beta k$ and $r_{i+1} - r_i \ge (1-\varepsilon)\beta k$ for $i \ge 1$, we therefore get

$$\ell \le \frac{k + 12 \cdot \ell \cdot \sqrt{k \ln k}}{(1-\varepsilon) \cdot \beta k}$$

and thus

$$\ell \le \frac{1}{(1-\varepsilon) \cdot \beta} \cdot \left(1 + \mathcal{O}\left(\sqrt{\frac{\log k}{k}}\right)\right).$$

Hence, the total mass in vector $\vec{x}'$ is at most

$$\begin{aligned} \ell \cdot \frac{\sqrt{2}+\frac{D}{\sqrt{\ln k}}}{\beta \cdot \sqrt{k \ln k}} &\le \frac{\left(\sqrt{2}+\frac{D}{\sqrt{\ln k}}\right) \cdot \left(1 + \mathcal{O}\left(\sqrt{\frac{\log k}{k}}\right)\right)}{(1-\varepsilon)\beta^2 \cdot \sqrt{k \ln k}} \\ &\le \frac{\varepsilon\left(\sqrt{2} + \mathcal{O}\left(\frac{1}{\sqrt{\log k}}\right)\right)}{\sqrt{2}} \left(1 + \frac{c}{\sqrt{\ln k}}\right)^{-2}. \end{aligned}$$

This is smaller than $\varepsilon - 1/k^{1/4} - 1/(2k^2)$ for sufficiently large k and a sufficiently large constant c.

The following three lemmas follow from the solution of the MGR differential equation given by Equation (1.5) and from the Poisson tail probability bounds given in Inequality (2.7). The first lemma states that the infinite sums of Poisson probabilities that determine the values of $x_i(t)$ in Equation (1.5) are either negligibly small or dominated by a single term.

Lemma 4.2. *Let $\vec{x}(t) = (x_0(t), \ldots, x_{k-1}(t))$ be the state after time $t \le k$ in the loading phase. We have*

$$x_i(t) = \sum_{j=0}^{k-1} x_j(0) \cdot \Pr(\mathrm{Po}(t) = \nu_{i,j}) + \mathcal{O}(e^{-k/16})$$

where $\nu_{i,j}$ is chosen such that $\nu_{i,j} \equiv i - j \pmod{k}$ and $\nu_{i,j} \in [t - k/2, t + k/2)$.

Lemma 4.3. *For sufficiently large k, after time $t \le k$ of the loading phase, all but a $1/k^5$-fraction of the mass starting in level $i \in \{0, \ldots, k-1\}$ ends up in levels j such that $j - i \equiv h \pmod{k}$ for $t - 5\sqrt{k \ln k} < h < t + 5\sqrt{k \ln k}$.*

For the following lemma, note that we only consider $\beta \le 1$ and that therefore $\beta k \le k$.

Lemma 4.4. *Assume that the burst starting with state $\vec{y}$ has size B. At the end of the burst (i.e., in vector $\vec{z}$), all but a $1/k^5$-fraction of the mass starting in level $i \in \{0, \ldots, k-1\}$ ends up in level 0 or in levels j for $B \cdot \beta k - 5\sqrt{k \ln k} < j < B \cdot \beta k + 5\sqrt{k \ln k}$.*

Lemma 4.5. *For β as given by Inequality (4.22), the system for the first time switches from the loading phase to the burst phase at a time t^* in the range $k - \sqrt{k \ln k} - \mathcal{O}(\sqrt{k}) \le t^* \le k - \sqrt{k \ln(k)/2} + \mathcal{O}(\sqrt{k})$. Hence, time t^* is the first time where the system reaches a state in Ω_{BB}. In particular, $x_{k-2}(t^*) > x_{k-1}(t^*) = 1/(\beta k)$.*

[5]The constant D can be chosen to be twice the hidden constant in the $\mathcal{O}(\cdot)$-notation in the statement of Lemma 4.5. We will not explicitly determine the value of D.

Proof. We have to show that such a time t^* exists such that $\vec{x}(t^*) \in \Omega_{\mathrm{BB}}$ and such that $\vec{x}(t) \notin \Omega_{\mathrm{BB}}$ for $t < t^*$. Hence, we have to find the smallest t^* such that $x_{k-1}(t^*) = 1/(\beta k)$ and we have to show that $x_{k-2}(t^*) > x_{k-1}(t^*)$.

Let us first argue why the smallest t^* for which $x_{k-1}(t^*) = 1/(\beta k)$ is in the given range. Because the initial state $\vec{x}$ is nice, the vector $\vec{x} - (x_0, 0, 0, \ldots, 0)$ is safe and thus $x_{k-1}(t) - g_{k-1}(t) < 1/(2\beta k)$ for all t. We therefore need $g_{k-1}(t) \geq 1/(2\beta k)$ to have $x_{k-1}(t) = 1/(\beta k)$. Clearly, $x_{k-1}(t) \geq 1/(\beta k)$ as soon as $g_{k-1}(t) \geq 1/(\beta k)$. Let us therefore consider the earliest times t_1 and t_2 for which $g_{k-1}(t_1) \geq 1/(2\beta k)$ and $g_{k-1}(t_2) \geq 1/(\beta k)$. By Lemma 4.2 and Stirling's formula, we have

$$\begin{aligned} g_{k-1}(t) &= x_0 \cdot \frac{t^{k-1}}{(k-1)!} \cdot e^{-t} + \Theta(e^{-k/16}) \\ &= \left(x_0 \pm \frac{1}{\Theta(k)}\right) \cdot \left(\frac{t}{k-1}\right)^{k-1} \cdot \frac{e^{k-t-1}}{\sqrt{2\pi k}}. \end{aligned}$$

For $t = k - 1 - q$ with $q = \rho\sqrt{k \ln k}$ and $\rho > 0$, this becomes

$$\begin{aligned} g_{k-1}(t) &= \left(x_0 \pm \frac{1}{\Theta(k)}\right) \cdot \left(1 - \frac{q}{k-1}\right)^{k-1} \cdot \frac{e^q}{\sqrt{2\pi k}} \\ &= \left(x_0 \pm \frac{1}{\Theta(k)}\right) \cdot e^{-q-\frac{q^2}{2(k-1)}+\mathcal{O}\left(\frac{q^3}{k^3}\right)} \cdot \frac{e^q}{\sqrt{2\pi k}} \\ &= \left(\frac{x_0}{\sqrt{2\pi}} \pm \frac{1}{\Theta(k)}\right) \cdot e^{-\frac{q^2}{2(k-1)}} \\ &= \left(\frac{x_0}{\sqrt{2\pi}} \pm \frac{1}{\Theta(k)}\right) \cdot e^{-\frac{1+\rho^2}{2}\cdot \ln k} \\ &= \left(\frac{x_0}{\sqrt{2\pi}} \pm \frac{1}{\Theta(k)}\right) \cdot k^{-\frac{1+\rho^2}{2}}. \end{aligned}$$

By Lemma 4.1, we have $x_0 > 1 - \varepsilon$ where $\varepsilon > 0$ is a constant. In order to obtain $g_{k-1}(k - 1 - \rho\sqrt{k \ln k}) = C/(\beta k)$ for a constant C, we therefore need to choose ρ such that $(1 + \rho^2)/2 = \ln(\beta k)/\ln k \pm \Theta(1/\log k)$. For β between $\Theta((k \ln k)^{-1/4})$ and 1, this yields $\sqrt{k \ln(k)/2} - \mathcal{O}(\sqrt{k}) \leq q \leq \sqrt{k \ln k} + \mathcal{O}(\sqrt{k})$ as claimed.

It remains to show that at time t^*, $x_{k-2}(t^*) > x_{k-1}(t^*)$. We first show a lower bound on $g_{k-2}(t^*) - g_{k-1}(t^*)$. We have

$$\begin{aligned} \frac{g_{k-2}(t^*)}{g_{k-1}(t^*)} &= \frac{x_0 \cdot \Pr(\mathrm{Po}(t^*) = k-2)}{x_0 \cdot \Pr(\mathrm{Po}(t^*) = k-1)} \\ &= \frac{k-1}{t^*} = 1 + \Theta\left(\sqrt{\frac{\log k}{k}}\right). \end{aligned}$$

From $g_{k-1}(t^*) = C/(\beta k)$ for a constant C, we therefore get that

(4.23)
$$g_{k-2}(t^*) - g_{k-1}(t^*) = \frac{1}{\beta k} \cdot \Theta\left(\sqrt{\frac{\log k}{k}}\right) = \Theta\left(\frac{\sqrt{\log k}}{\beta k^{3/2}}\right)$$

Let $\vec{w} = \vec{x}(t^*) - \vec{g}(t^*)$. In order to complete the proof, we show that $\vec{w}_{k-1} < \vec{w}_{k-2} + \mathcal{O}(k^{-3})$. By Lemma 4.3, from the mass starting in levels $5\sqrt{k \ln k}, \ldots, k-2-5\sqrt{k \ln k}$ at time 0, at most $1/k^5$ ends in levels $k-2$ or $k-1$ after time t^* of the loading phase. By Definition 4.2 (ε-nice), in an ε-nice state and thus at time 0, the mass in levels $i = 1, \ldots, r_1 - \Delta_1 - 1$ is at most i/k^4. Hence, the total mass in levels $1, \ldots, 5\sqrt{k \ln k}$ at time 0 is at most $\Theta(\sqrt{\log k}/k^{3.5})$. Except for the contribution from levels $k - 5\sqrt{k \ln k}, \ldots, k-1$, we therefore have $\vec{w}_{k-1} < \vec{w}_{k-2} + \mathcal{O}(1/k^3)$. Let $x_{i,j}(t^*) = x_j(0) \cdot \Pr(\mathrm{Po}(t^*) \equiv i - j \pmod{k})$ be the contribution of level j at time 0 to level i at time t^*. Analogously to the reasoning about $g_{k-2}(t^*)$ and $g_{k-1}(t^*)$, we have $x_{k-2,j}(t^*) > x_{k-1,j}(t^*)$ for $j = k - 5\sqrt{k \ln k}, \ldots, k-1$ (since the mass is even further away from the expected value of the respective Poisson distribution, the ratio between $x_{k-2,j}(t^*)$ and $x_{k-1,j}(t^*)$ is even larger than the ratio between $g_{k-2}(t^*)$ and $g_{k-1}(t^*)$ for the considered values of j).

The next lemma specifies the number of neurons in the tail of the big wave that are in the first few levels when the burst starts. Interestingly, the number of neurons hardly depends on the size of the 'wave' (i.e., the number of neurons starting the loading phase in level 0). In fact, if the 'wave' is smaller, the constant D in the following lemma becomes even slightly larger.

Lemma 4.6. *At the end of the loading phase, for a sufficiently large constant D and a sufficiently large value k, the mass in the first $5 \cdot \sqrt{k \ln k}$ levels of $\vec{y}$ is at most*

$$\sum_{i=0}^{\lfloor 5\cdot\sqrt{k \ln k}\rfloor} y_i \leq \frac{\sqrt{2} + \frac{D}{\sqrt{\ln k}}}{\beta \cdot \sqrt{k \ln k}}.$$

Proof. We consider the mass contributed from $\vec{g}(t^*)$ and $\vec{x}(t^*) - \vec{g}(t^*)$ separately. Note that from the definition of t^*, it holds that $g_0(t^*) \leq g_{k-1}(t^*) \geq 1/(\beta k)$. By Lemmas 4.2 and 2.1, we thus have

(4.24)
$$\begin{aligned} \sum_{i=0}^{\lfloor 5\cdot\sqrt{k \ln k}\rfloor} g_i(t^*) &\leq \sum_{i=0}^{\lfloor 5\cdot\sqrt{k \ln k}\rfloor} x_0 \cdot \Pr(\mathrm{Po}(t^*) = k+i) + e^{-k/16} \\ &\leq \frac{k+1}{k+1-t^*} \cdot \frac{1}{\beta k} + e^{-k/16}. \end{aligned}$$

In the partition of the vector $\vec{x}'$ into vectors $\vec{x}'_1, \ldots, x'_\ell$, only vectors $\vec{x}'_i$ for which $r_i + \Delta_i \geq k -$

$1 - 5\sqrt{k \ln k}$ contribute to the mass in levels $k - 1 - 5\sqrt{k \ln k}, \ldots, k-1$ at time 0. Because $r_{i+1} - r_i = \Theta(\beta k)$, $\vec{x}'$ is partitioned into $\mathcal{O}(k/(\beta k) = \mathcal{O}(1/\beta)$ vectors $\vec{x}_i$ and thus $\Delta_i = \mathcal{O}(\sqrt{k \log k}/\beta)$. The condition $r_i + \Delta_i \geq k - 1 - 5\sqrt{k \ln k}$ is therefore only satisfied for

$$\mathcal{O}\left(1 + \frac{\sqrt{k \log k}}{\beta} \cdot \frac{1}{\beta k}\right) = \mathcal{O}\left(1 + \frac{\sqrt{\log k}}{\beta^2 \sqrt{k}}\right)$$

vectors $\vec{x}'_i$. The total mass in levels $k - 1 - 5\sqrt{k \ln k}, \ldots, k-1$ at time 0 therefore is at most

$$\mathcal{O}\left(1 + \frac{\sqrt{\log k}}{\beta^2 \sqrt{k}}\right) \cdot \mathcal{O}\left(\frac{1}{\beta \sqrt{k \log k}}\right) = \mathcal{O}\left(\frac{\sqrt{\log k}}{\beta \sqrt{k}}\right). \tag{4.25}$$

Analogously to the proof of Lemma 4.5, the total mass in levels $0, \ldots, 12\sqrt{k \ln k}$ at time 0 is negligible (at most $\mathcal{O}(k^{-3})$). By Lemma 4.3, all mass that starts the loading phase in levels between $10\sqrt{k \ln k}$ and $k - 1 - 5\sqrt{k \ln k}$, contributes at most $1/k^5$ to the levels $0, \ldots, 5\sqrt{k \ln k}$ at time t^*. Combined with Inequalities (4.24) and (4.25), we thus obtain that the total mass in levels $0, \ldots, 5\sqrt{k \ln k}$ at the end of the loading phase is at most

$$\begin{aligned} \sum_{i=0}^{\lfloor 5\sqrt{k \ln k} \rfloor} x_i(t^*) &\leq \frac{1}{\beta(k - t^*)} + \mathcal{O}\left(\frac{\sqrt{\log k}}{\beta \sqrt{k}}\right) \\ &\leq \frac{\sqrt{2}}{\beta \cdot \sqrt{k \ln k}} + \mathcal{O}\left(\frac{1}{\beta \cdot \sqrt{k} \cdot \log k}\right). \end{aligned}$$

The second inequality follows from the upper bound on t^* given by Lemma 4.5. The lemma now follows by choosing the constant D sufficiently large.

The following lemma shows that except for the mass (specified by Lemma 4.6) that finishes the loading phase in the first few levels, almost all neurons the start the loading phase in level 0 participate in the burst at time t^*.

Lemma 4.7. *The burst at time t^* has size at least $1 - \varepsilon$.*

Proof. By Lemma 4.1, at the beginning $x_0 \geq 1 - \varepsilon + 1/k^{1/4}$. We show that at least $1 - \varepsilon$ of this mass is part of the big burst at time t^*. By Lemma 4.3, at most $1/k^5$ of the mass starting in level 0 is in levels $5\sqrt{k \ln k}, \ldots, t^* - 5\sqrt{k \ln k}$ at the end of the loading phase. Further, by Lemma 4.6, at most $\mathcal{O}(\beta^{-1}(k \log k)^{-1/2}) = \mathcal{O}((k \log k)^{-1/4})$ of the mass starting in x_0 is in levels $0, \ldots, 5\sqrt{k \ln k}$ at time t^*. Hence, when the burst starts, the total mass in levels $t^* - 5\sqrt{k \ln k}, \ldots, k - 1$ is at least $1 - \varepsilon + \Theta(1/k^{1/4})$. Recall that the state at the beginning of the burst is denoted by $\vec{y} = \vec{x}(t^*)$. Let $\vec{y}(t)$ be the state after time t of the burst. The burst clearly does not end as long as $y_{k-1}(t) \geq 1/(\beta k)$. For $t^* + t \geq k - 1$, we have

$$\begin{aligned} y_{k-1}(t) &= \sum_{i=0}^{k-1} y_i \cdot \Pr(\mathrm{Po}(t) = k - 1 - i) \\ &\geq \sum_{i=0}^{k-1} x_0 \cdot \Pr(\mathrm{Po}(t^*) = i) \cdot \Pr(\mathrm{Po}(t) = k - 1 - i) \\ &= x_0 \cdot \Pr(\mathrm{Po}(t^* + t) = k - 1) \\ &= x_0 \cdot \frac{(t^* + t)^{k-1}}{(k-1)!} \cdot e^{-t^* - t} \\ &= \left(x_0 - \frac{1}{\Theta(k)}\right) \cdot \left(1 + \frac{t^* + t - (k-1)}{k - 1}\right)^{k-1} \\ &\qquad \cdot e^{k-1-t^*-t} \\ &\geq \left(x_0 - \frac{1}{\Theta(k)}\right) \cdot e^{(t^* + t - k - 1)^2/2} \\ &= \left(x_0 - \frac{1}{\Theta(k)}\right) \cdot e^{(t - \Theta(\sqrt{k \log k}))^2/2}. \end{aligned}$$

Hence, we need to set $t = \Omega(\sqrt{k \log k})$ to obtain $y_{k-1}(t) = 1/(\beta k)$. Let $\rho(t) = x_0 \Pr(\mathrm{Po}(t^* + t) \leq k - 1) + e^{-k/16}$ be the mass that started the loading phase in level 0 and is still in levels $0, \ldots, k-1$ after time t of the burst. Using Lemma 2.1, $\beta = \Omega((k \log k)^{-1/4})$, and $t = \Omega(\sqrt{k \log k})$, we get that

$$\begin{aligned} \rho(t) &< \frac{t + t^*}{t + t^* - k - 1} \cdot \frac{1}{\beta k} + e^{-k/16} \\ &= \mathcal{O}\left(\frac{1}{\beta \sqrt{k \log k}}\right) = \mathcal{O}\left(\frac{1}{(k \log k)^{1/4}}\right). \end{aligned} \tag{4.26}$$

Because at the beginning of the burst, at least $1 - \varepsilon + \Theta(1/k^{1/4})$ of the mass x_0 starting the loading phase in level 0 is in levels $-5\sqrt{k \ln k}, \ldots, k - 1$, Inequality (4.26) implies that all except $\mathcal{O}((k \log k)^{-1/4})$ of this mass has to be part of the burst. For sufficiently large k, this is at least $1 - \varepsilon$.

Based on Lemmas 4.1–4.7, we can now prove Theorem 4.1.

Proof. [Proof of Theorem 4.1] Lemma 4.5 shows that if the system starts in an ε-nice state, for sufficiently large k it stays in the loading phase until a time $t^* = k - o(k)$ and then changes to the burst phase. Lemma 4.7 proves that the size of the burst at time t^* is at least $1 - \varepsilon$ if k is sufficiently large. It remains to show that when the burst ends, the system is again in an ε-nice state.

Recall the initial state $\vec{x} = \vec{x}(0)$ (by ε-niceness) can be decomposed as $\vec{x} \leq (x_0, 0, 0, \ldots, 0) + \vec{\eta} + \vec{x}'_1 + \cdots \vec{x}'_\ell$. We have to show that the state $\vec{z}$ immediately after the burst can be decomposed as $\vec{z} \leq (z_0, 0, 0, \ldots, 0) + \vec{\eta} +$

$\vec{z}_1'+\cdots\vec{z}_{\ell'}'$ such that the following conditions hold: There are integers $0 < r_1' < \cdots < r_{\ell'}'$ such that

1. $(1-\varepsilon)\beta k \leq r_1' \leq \beta k$ and $\forall i \in \{2,\ldots,\ell'\}$: $(1-\varepsilon)\beta k \leq r_i' - r_{i-1}' \leq \beta k$
2. $\forall i \in [\ell']\forall j \notin [r_i' - \Delta_i, r_i' + \Delta_i] z_{i,j}' = 0$ and for a sufficiently large constant D,

$$\forall i \in [\ell'] : \sum_{j=0}^{k-1} z_{i,j}' \leq \frac{\sqrt{2}+\frac{D}{\sqrt{\ln k}}}{\beta \cdot \sqrt{k \ln k}}. \tag{4.27}$$

3. The vector $\vec{z} - (z_0, 0, 0, \ldots, 0)$ is safe and for $i > r_1 + \Delta_1$, $z_i' \leq 1/(3\beta k)$.

Let $b \in [1-\varepsilon, 1]$ be the size of the burst at time t^*. By Lemma 4.3, during the loading phase, all but $1/k^5$ of the mass in $\vec{x}$ is cyclically shifted by $t^* = k - \Theta(\sqrt{k \log k})$ and spread out by at most $5\sqrt{k \ln k}$. In the burst phase, all but $1/k^5$ of the mass is non-cyclically shifted by $b\beta k$ and spread out by at most $5\sqrt{k \ln k}$. We can handle the different components of the start vector $\vec{x}$ independently.

Let us first consider how the vector $\vec{\eta}$ is shifted during the loading phase and the burst phase. The total mass of the noise vector $\vec{\eta}$ is $\sum_{i=0}^{k-1} \eta_i = \binom{k}{2}/k^4 < 1/(2k^2)$. By Lemma 4.3, all but a $1/k^5$-fraction of this mass (i.e., all but $1/(2k^7)$) is shifted to the right by t^* and spread to levels at most $5\sqrt{k \ln k}$ from there. Equivalently, everything is shifted to the left by $k - 1 - t^*$. Because all the mass is shifted an spread in the same way and because the coordinates of $\vec{eta}$ are monotonically increasing, we can shift $\vec{\eta}$ to the left by $k - t^* - 1 + 5\sqrt{k \ln k}$ (and sufficiently extend it to the right) to get an upper bound on the shifted vector η. Let us just consider the part of $\vec{eta}$ that is moved to levels $5\sqrt{k \ln k}+1, \ldots, k-1$ during the loading phase (the part that is moved to the first $5\sqrt{k \ln k}$ is part of the mass bounded by Lemma 4.6 and will be incorporated into the vector $\vec{z}_1'$). Let η_i' be the part of $\vec{\eta}$ that ends up in level i at the end of the loading phase. For $i > 5\sqrt{k \ln k}$, we get $\eta_i' \leq (i + k - 1 - t^* + 5\sqrt{k \ln k})/k^4 + 1/(2k^7)$. For $i \leq 5\sqrt{k \ln k}$, we define $\eta_i' = 0$ and let $\vec{\eta}' = (\eta_0', \ldots, \eta_{k-1}')$. During the burst phase, the vector $\vec{\eta}'$ is (non-cyclically) shifted to the right by $b\beta k \geq (1-\varepsilon)\beta k$ and spread to levels at most $5\sqrt{k \ln k}$ from there. Let η_i'' be the mass of $\vec{\eta}'$ ending in level i after the burst. Using the same argumentation as before, we have

$$\begin{aligned} \eta_i'' &\leq \min\left\{\frac{1}{k^7}, \frac{i + k - 1 - t^* - b\beta k + 10\sqrt{k \ln k}}{k^4} + \frac{1}{k^7}\right\} \\ &\leq \min\left\{\frac{1}{k^7}, \frac{i - (1-\varepsilon - o(1))\beta k}{k^4}\right\}. \end{aligned} \tag{4.28}$$

We now show how to construct the vectors $\vec{z}_i'$ into which the state $\vec{z}$ after the burst is partitioned. Consider a vector $\vec{x}_i'$ of the initial state. Because we start in an ε-nice state at time 0, $\vec{x}_{i,j}' = 0$ except for $j \in [r_i - \Delta_i, r_i + \Delta_i]$. Hence by Lemma 4.3, all but at $1/k^5$-fraction of the mass of $\vec{x}_i'$ is moved to levels in $\big[r_i - \Delta_i - (k-1-t^*) - 5\sqrt{k \ln k}, r_i + \Delta_i - (k-1-t^*) + 5\sqrt{k \ln k}\big]$ during the loading phase and to levels in

$$\begin{aligned} &\big[r_i - \Delta_i - (k-1-t^*) + b\beta k - 10\sqrt{k \ln k}, \\ &\quad r_i + \Delta_i - (k-1-t^*) + b\beta k + 10\sqrt{k \ln k}\big] \\ &\qquad \in \big[r_i - \Delta_i + b\beta k - 11\sqrt{k \ln k} - \mathcal{O}(\sqrt{k}), \\ &\qquad\quad r_i + \Delta_i + b\beta k + 10\sqrt{k \ln k}\big] \\ &\qquad \in \big[r_i + b\beta k - \Delta_{i+1}, r_i + b\beta k + \Delta_{i+1}\big] \end{aligned}$$

for sufficiently large k. For $i \geq 2$, we thus define $r_i' = r_{i-1} + b\beta k$ and let $\vec{z}_i'$ contain all the mass of $\vec{x}_{i-1}$ that is moved to levels between $r_i' - \Delta_i$ and $r_i' + \Delta_i$. The vector $\vec{z}_1'$ is constructed by using the mass that starts the burst in the first $5\sqrt{k \ln k}$ levels and is specified by Lemma 4.6. All but a $1/k^5$-fraction of this mass is moved to levels $b\beta k - \Delta_1, \ldots, b\beta k + \Delta_1$. We can therefore set $r_1' = b\beta k \in [(1-\varepsilon)\beta k, \beta k]$. We clearly have $r_{i+1}' - r_i \in [(1-\varepsilon)\beta k, \beta k]$. Further, the total mass of each vector $\vec{z}_i{}'$ is bounded as given by Inequality (4.27) because of Lemma 4.6 and because the total mass of each vector $\vec{x}_i{}'$ is bounded accordingly. Hence, all but a $1/k^5$-fraction of the mass of vector $\vec{x}'$ ends up in vector $\vec{z}'$.

So far, we have analyzed what happens to vectors $\vec{\eta}$ and $\vec{x}'$ during the loading phase and during the burst. Most of the mass starting in level 0 either goes to vector $\vec{z}_1'$ or participates in the burst and thus is in level 0 after the burst. The mass that does not end up in z_0 or $\vec{z}_1'$ can be bounded by

$$x_0 \left(\Pr(\mathrm{Po}(t^* + b\beta k) \leq k-1) + \Pr(\mathrm{Po}(t^*) > 5\sqrt{k \ln k})\right)$$

which is at most $2x_0 k^{-5}$. In order to satisfy all necessary conditions, we need to show that all the mass that is not in level 0 or in one of the vector $\vec{z}_i'$ (and thus in vector $\vec{z}'$) after the burst is smaller than $\vec{\eta}$. For sufficiently large k, this is true due to Inequality (4.28).

To complete the proof, it remains to show that the vector $\vec{w} = \vec{z} - (z_0, 0, 0, \ldots, 0)$ is safe and that $z_i' \leq 1/(3\beta k)$ for $k > r_1' + \Delta_1$. To simplify the proof of the latter, let us subtract $\max\{z_i', 1/k^5\}$ from every z_i' (and the corresponding $z_{i,j}'$). The mass that does not end up in z_0 or a vector $\vec{z}'$ is still upper bounded by $\vec{\eta}$. A closer look at Equation (1.5) reveals that the value of $x_i(t)$ at time t in the loading phase is a convex combination of the values $x_0, \ldots, x_{k-1}$. Hence, for all times $t \geq 0$, $\max_i x_i(t) \leq \max_i x_i(0)$. Let $\vec{w}(0) = \vec{w}$ and $\vec{w}(t)$ be the state after time t in the loading phase when starting in state $\vec{w}$. Because for $i > r_1 + \Delta_1$, $x'_i \leq 1/(3\beta k)$ for $t \leq k - 1 - r_1 - \Delta_1 - 5\sqrt{k \ln k}$, $w_{k-1} \leq 1/(3\beta k) + \eta_{k-1} + 1/k^5$. Because the total mass in levels $r_1 \pm \Delta_1$ levels of $\vec{w}$ is only $\mathcal{O}(1/(\beta\sqrt{k \log k})) = \mathcal{O}((k \log k)^{-1/4})$, the maximum contributed to any w_i by levels $r_1 \pm \Delta_1$ is at most

$$\mathcal{O}\left(\frac{\Pr\left(\text{Po}(k - \Theta(\beta k)) = k - \Theta(\beta k)\right)}{(k \log k)^{1/4}}\right) = \mathcal{O}\left(\frac{1}{k^{3/4} \log^{1/4} k}\right) = o\left(\frac{1}{\beta k}\right).$$

Therefore, at time $t = k - 1 - r_1 - \Delta_1 - 5\sqrt{k \ln k}$, all coordinates $w_i(t)$ are clearly smaller than $1/(2\beta k)$ (for sufficiently large k) and thus, the vector $\vec{w}$ is safe.

To show that $z_i \leq 1/(3\beta k)$ consider the vector resulting from $\vec{x}'$ after time t^* in the loading phase. Using similar arguments to above, no coordinate of this vector is larger than $1/(3\beta k) + 1/k^5$. Because all mass in $\vec{z}'$ in levels larger than $r_1 + \Delta_1$ comes from $\vec{x}'$ and because we subtracted $1/k^5$ from every z'_i, this implies that $z_i \leq 1/(3\beta k)$ for $i > r_1 + \Delta_1$.

4.2 Below the Threshold

The key to understand the value of the threshold for β is to understand the role of the part of the mass that starts the loading phase in level 0 and does not participate in the burst. Lemma 4.6 quantifies the number of neurons that are already in levels $0, \ldots$ when the system switches from the loading phase to the burst phase. Interestingly, up to lower order terms, the bound given by Lemma 4.6 is not only an upper bound but also a lower bound. If we start the system with all neurons in level 0, in every burst, roughly $\sqrt{2}n/(\beta\sqrt{k \ln k})$ neurons that are in level 0 at the beginning of the preceding loading phase are not part of the burst. All mass that is in level i at the beginning of the loading phase ends up in levels around $i + b\beta k$ after the next burst, where b is the size of this burst. If β is chosen as given by Theorem 1.3, the size of the burst reaches 1/2 after a finite number $(\Theta(\beta\sqrt{k \log k}))$ of iterations.

Let us call the mass in the first $5\sqrt{k \ln k}$ before burst i, the lump of burst i. Once the size of each burst is at most 1/2 the distance (in levels) between lumps of subsequent bursts is at most $\beta k/2$. If $\beta < 2^{1/4}/(k \ln k)^{1/4}$, the total mass of the lumps is $2/\beta \cdot \sqrt{2}/(\beta\sqrt{k \ln k}) > 1/2$. Hence, the size of the bursts will become smaller than 1/2. If the burst becomes smaller than 1/2, the total mass of the lumps becomes even larger and it can be shown that then, the big bursts become smaller quickly and finally disappear.

References

[1] F. Apfaltrer, C. Ly, and D. Tranchina. Population density methods for stochastic neurons with a 2-d state space: Application to neural networks with realistic synaptic kinetics. *Network: Computation in Neural Systems*, 17:373–418, 2006.

[2] P. C. Bressloff and S. Coombes. Desynchronization, mode locking, and bursting in strongly coupled integrate-and-fire oscillators. *Phys. Rev. Lett.*, 81(10):2168–2171, 1998.

[3] G. Buzsáki. *Rhythms of the Brain.* Oxford University Press, 2006.

[4] D. Cai, L. Tao, M. Shelley, and D.W. McLaughlin. An effective kinetic representation of fluctuationdriven neuronal networks with application to simple and complex cells in visual cortex. *Proc. Natl. Acad. Sci.*, 101(20):7757–7762, 2004.

[5] S. R. Campbell, D. L. L. Wang, and C. Jayaprakash. Synchrony and desynchrony in integrate-and-fire oscillators. *Neural Computation*, 11(7):1595–1619, 1999.

[6] R. E. L. DeVille and C. S. Peskin. Synchrony and asynchrony in a fully stochastic neural network. *Bulletin of Mathematical Biology*, 70(6):1608–1633.

[7] P. Goel and B. Ermentrout. Synchrony, stability, and firing patterns in pulsecoupled oscillators. *Physica D: Nonlinear Phenomena*, 163(3–4):191216, 2002.

[8] E. Haskell, D.Q. Nykamp, and D. Tranchina. Population density methods for large-scale modelling of neuronal networks with realistic synaptic kinetics: cutting the dimension down to size. *Network: Computation in Neural Systems*, 12(2):141–147, 2004.

[9] S. Janson, T. Łuczak, and A. Rucinski. *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[10] T. G. Seierstad. Stronger large deviation bounds for wormald's differential equation method. Submitted, `http://folk.uio.no/taralgs/artikler/wldb.pdf`, 2008.

[11] W. Senn and R. Urbanczik. Similar nonleaky integrate-and-fire neurons with instantaneous couplings always synchronize. *SIAM J. Appl. Math.*, 61(4):1143–1155 (electronic), 2000/01.

[12] N. Wormald. The differential equation method for random graph processes and greedy algorithms. In M. Karonski and H.J. Proemel, editors, *Lectures on Approximation and Randomized Algorithms*, pages 73–155. 1999.