

Dynamics of Stochastic Neuronal Networks and the Connections to Random Graph Theory

R. E. Lee DeVille^{1*}, C. S. Peskin² and J. H. Spencer²

¹ Department of Mathematics, University of Illinois, Urbana, IL 60801

² Courant Institute of Mathematical Sciences, New York University, New York, NY 10012

Abstract. We analyze a stochastic neuronal network model which corresponds to an all-to-all network of discretized integrate-and-fire neurons where the synapses are failure-prone. This network exhibits different phases of behavior corresponding to synchrony and asynchrony, and we show that this is due to the limiting mean-field system possessing multiple attractors. We also show that this mean-field limit exhibits a first-order phase transition as a function of the connection strength — as the synapses are made more reliable, there is a sudden onset of synchronous behavior. A detailed understanding of the dynamics involves both a characterization of the size of the giant component in a certain random graph process, and control of the pathwise dynamics of the system by obtaining exponential bounds for the probabilities of events far from the mean.

Key words: neural network; neuronal network; synchrony; mean-field analysis; integrate-and-fire; random graphs; limit theorem

AMS subject classification: 05C80, 37H20, 60B20, 60F05, 60J20, 82C27, 92C20

1. Introduction

More than three centuries ago, Huygens first observed in his experiments on pendulum clocks that coupled nonlinear oscillators tend to synchronize [26, 5]. Since then, the study of synchronized nonlinear oscillators has had a long and storied history, and along the way has inspired many ideas in modern dynamical systems theory. Moreover, the study of oscillator synchronization has done much to explain the dynamics of many real systems, including cardiac rhythms [37, 23,

*Corresponding author. E-mail: rdeville@illinois.edu

22], circadian rhythms [14, 48], biochemical kinetics and excitable media [20, 27, 49, 13, 29], earthquakes [35, 15], a wide variety of neural systems [30, 1, 21, 47, 50, 7, 46, 51, 12, 24, 36], and even synchronization of fireflies [9, 34, 33]. See [43, 38, 52] for reviews.

It is of particular interest in mathematical neuroscience to study the dynamics of pulse-coupled oscillators, that is, oscillators that are coupled only when one of them “fires”. More specifically, each oscillator has no effect on its neighbors until it reaches some prescribed region of phase space known as the “firing regime”. The typical interaction between oscillators is that one oscillator’s firing leads to an advance or retardation of the phase of the oscillators on the receiving end of the coupling. There has been a large body of work on deterministic pulse-coupled networks [29, 30, 1, 21, 47, 50, 7, 46, 51, 12, 28, 37, 34, 39], much of which has been dedicated to the question of when such networks synchronize. The paradigm which has emerged from this work is that the presence of any one of several phenomena tends to cause such networks to synchronize, these include *leakiness* (the tendency of each oscillator to relax to an attractor) and *refractoriness* (any mechanism which imposes a minimum recovery time that must occur between the firing times of any oscillator). If we then add noise to such a network, then this would be expected to weaken the network’s ability to synchronize. If a network with synchronization mechanisms is perturbed by noise, then the two forces might balance, leading to the selection of certain coherent structures in the dynamics (much like the formation of coherent structures in certain nonlinear wave equations [45, 44]).

In [16], two of the authors considered a specific example of a network which contains refractoriness and noise; the particular model was chosen to study the effect of *synaptic failure* on the dynamics of a neuronal network, that is to say, the coupling between neurons was probabilistic and the arrival of an action potential at a given synapse would only have a certain probability of having any effect on the postsynaptic neuron. Using the language of neuronal models, the network we study is an all-to-all coupled neuronal network, whose neurons are fully stochastic discrete-state non-leaky integrate-and-fire oscillators, and whose synapses are faulty. See [16] for further discussion on the biological motivation for this model.

It was observed that one could choose parameters to make the network synchronize or desynchronize. For example, if the synapses were chosen to be very reliable, the network would synchronize and undergo very close to periodic behavior — in particular, the dynamics would be a series of large events where many neurons fired together, and the sizes of these events, and the intervals between them, had a small variance. If the synapses were chosen to be very unreliable, then the network would desynchronize and there would be small interneuronal correlation. Both of these behaviors were expected: strong coupling should lead to synchrony, and weak coupling in a noisy system should lead to decoherence. However, the most interesting observation was that when the reliability of the synapses was chosen in some middle regime, the network could support both synchronized and desynchronized behaviors, and would dynamically switch between the two.

Numerical evidence presented in [16] suggested the conjecture that if one takes the number of neurons in the network to infinity (and scales other parameters appropriately), then there is a deterministic hybrid system to which the stochastic network dynamics limit. (This deterministic limit is reminiscent of the population models commonly studied in neuroscience [28, 8, 41, 42, 25, 11, 10, 4].) Moreover, this deterministic system has (for some parameter values) two attractors, one

of which corresponds to synchrony and one to asynchrony; the real neuronal network model would then be a stochastic perturbation of this system whose variance scales inversely to the number of neurons, and thus a large network would spend most of its time near the attractors the limiting system and switch between them rarely, but at random times.

The fundamental component of the description of the synchronous state is a precise characterization of the “big bursts” in the dynamics, by which we mean bursts which scale like N , where N is the number of neurons in the network. What we show below is that there is a precise connection between big bursts in the network model considered here and the emergence of “giant components” in the Erdős-Rényi random graph model [18, 19, 3, 6]. What was done in [18] is to define the probability distribution on graphs which corresponds to independently choosing each of the $N(N - 1)/2$ undirected edges to be present with probability p . Once this distribution is defined, one can attempt to compute the probability distribution of the largest connected component of such a graph, or the size of the connected component of a vertex chosen at random. These are difficult combinatorial questions in general; however, the great breakthrough of [18] was to show that in the limit $N \rightarrow \infty, pN = \beta \in \mathbb{R}$, the solution to these questions has a relatively simple description. There is a critical value $\beta_{cr} = 1$ such that for $\beta < \beta_{cr}$ all of the components are very small (the largest has only logarithmic order), and for $\beta > \beta_{cr}$ the largest component scales like N . In fact, one can say more: there is a deterministic function $\rho(\beta)$ such that the proportion of vertices in this “giant component” (as Erdős and Rényi called it) is almost surely $\rho(\beta)$ in the limit. For $\beta = \beta_{cr}$ — the “phase transition” (in the language of mathematical physics) — the situation is very delicate. It turns out that we will not need to deal with the critical case below, which simplifies matters tremendously.

To compute the size of the big bursts in this model, we will need only a slight generalization of these ideas — noting that p in the random graph model corresponds to the probability of synaptic transmission in the neuronal network model. This approach will allow us to characterize the size of the big bursts as a function of the state of the system immediately prior to such a burst. The big bursts are coupled to the state of the system, so we will also need to understand the evolution of the state of the system when it is not undergoing big bursts. We can describe the stochastic dynamics as a Markov process, and we will be able to use a slightly modified version of Kurtz’ Theorem [40, 31, 32] to describe the interburst evolution.

Combining these two main ideas will show that in the limit $N \rightarrow \infty, pN \rightarrow \beta$, there exists a deterministic dynamical system $Q_{\beta,K}$ defined on \mathbb{R}^K such that the dynamics of our stochastic model converge pathwise to the dynamics of $Q_{\beta,K}$, and for some choices of β, K , the dynamical system $Q_{\beta,K}$ has at least two attractors, one of which corresponds to the synchronous dynamics, and one to the asynchronous dynamics. Moreover, the system $Q_{\beta,K}$ exhibits *critical phenomena*, or *bifurcations*: namely, for some K , there exist critical parameters $\beta_*(K)$ such that the structure of the attractors changes qualitatively as β passes through $\beta_*(K)$. We also see below that although this model, like the Erdős-Rényi random graph, has a critical transition from small bursts to large bursts, the phase diagram has some strikingly different features. In particular, the function $\rho(\beta)$ mentioned above which gives the size of giant components in the Erdős-Rényi is a continuous function of β , by which we mean $\lim_{\beta \rightarrow 1+} \rho(\beta) = 0$, and in fact is smooth for $\beta > 1$. In the model we consider here, we will show that the corresponding function which gives the “size of bursts”

in this model is sometimes discontinuous at the critical parameter value, leading to a “first-order phase transition”.

1.1. Plan of the paper

In Section 2., we give a precise description of the process we consider here, describe some phenomenological aspects of its dynamics, and make the “critical phenomena” ideas precise. In Section 3., we propose and motivate a deterministic dynamical system defined on \mathbb{R}^K which will serve as the “mean-field ($N \rightarrow \infty$) limit” of our stochastic dynamical system. In Section 4. we prove Theorem 1, which shows precisely that our network process limits pathwise onto the mean-field system as $N \rightarrow \infty$. In Section 5., we analyze the mean-field system and prove Theorem 5, which shows that for K sufficiently large, there exist values of β such that the mean-field has two attractors, and show how the two observed stochastic states in our network model correspond to these attractors.

2. Neural network model

2.1. Description of model

The network model we study has four parameters: $N, K \in \mathbb{Z}^+, p \in (0, 1), \rho \in \mathbb{R}^+$. The network consists of N neurons, and each neuron can be at any of the levels $\{0, 1, \dots, K - 1\}$. The rate ρ is the rate at which the neurons spontaneously move up one level as a result of exogenous input from neurons outside the network. More precisely, in any small time Δt , each neuron has a probability $\rho\Delta t + O(\Delta t^2)$ of being spontaneously promoted one level. At a sequence of times whose successive differences are chosen to be exponentially distributed with mean $(\rho N)^{-1}$, an event occurs. To implement an event in a stochastic simulation of the process, we choose one neuron at random and promote it one level. If this neuron was at any level other than $K - 1$, then the event stops. If the neuron was at level $K - 1$ prior to being promoted, then we say that this neuron “fires”, and this is the start of a “burst”. During the burst, whenever any neuron fires, we promote each neuron in the network which has not yet fired independently with probability p , and if any other neuron is raised past level $K - 1$ during a burst, it also fires. We continue this computation until the effect of each neuron which fired during the burst has been computed. Of course, the number of neurons which fire in a single burst can be any number from 1 to N — no neuron fires more than once in a burst. Finally, at the end of the burst, we set all of the neurons which fired to level 0, and the burst is idealized to take zero time.

A more complete discussion of this model and its relationship to physiological neuronal networks is given in [16]. Very briefly, however, the motivation behind this network is that each neuron is at some voltage level which is only allowed to be at the discrete states $k\Delta V$, $k = 0, \dots, K - 1$, where 0 is the reset voltage and $K\Delta V$ is the firing threshold. Whenever a neuron reaches threshold, it fires and activates the outgoing synapses. We assume that the network is all-to-all, so that each firing can potentially raise the voltage of every other neuron in the network; however, the synapses are failure-prone, and thus only with probability p will it propagate to the postsynaptic neuron the

information that an action potential has occurred in the presynaptic neuron. The refractoriness in this model is given by assuming that a neuron can only fire once during a burst — in particular, all of the neurons which fire together in one burst are subsequently clumped together at the end of the burst, even if they were in different states at the beginning of the burst.

To give a precise definition of our model, we define the state space of our system as the set $(\{0, 1, \dots, K-1\})^N$; picking a point in this space specifies the level of each neuron. To compute the bursting events below, we will find it convenient to append two states at levels K and $K+1$ — we think of K as the level where a neuron is “firing”, and $K+1$ is the level where a neuron which has already fired remains until the end of the burst. Now consider an initial (perhaps random) vector $X_0 \in (\{0, 1, \dots, K-1\})^N$, and define a cadlag process on the evolution of X_t as follows. Let $0 = t_0 < t_1 < \dots$ be a sequence of times such that $t_{i+1} - t_i$ is exponentially distributed with mean $(\rho N)^{-1}$, and define X_t to be constant on $[t_i, t_{i+1})$. Pick an index $n \in \{1, \dots, N\}$ uniformly and compute the following:

- if $X_{t_i, n} < K-1$, then $X_{t_{i+1}, n} = X_{t_i, n} + 1$, and $X_{t_{i+1}, j} = X_{t_i, j}$ for all $j \neq n$, i.e. promote only neuron n by one level,
- if $X_{t_i, n} = K-1$, then define a temporary vector $Y^{(0)}$ with $Y_j^{(0)} = X_{t_i, j}$ for all $j \neq n$ and $Y_n^{(0)} = K$. For each $r = 1, 2, \dots$, if there is no m with $Y_m^{(r-1)} = K$, then we say $Y^{(r)} = Y^{(r-1)}$. Otherwise, set

$$A^{(r)} = \{n \in \{1, \dots, N\} : Y_n^{(r)} = K\}.$$

Choose m uniformly out of $A^{(r)}$, and define $Z \in (\{0, 1\})^N$ as the random vector with $Z_m = 1$, $Z_j = 0$ for all j such that $Y_j^{(r-1)} \in \{K, K+1\}$, and all other entries of Z as independent realizations of the Bernoulli random variable which is 1 with probability p and 0 with probability $1-p$. Then define $Y^{(r)} = Y^{(r-1)} + Z$. Finally, define

$$k^* = \min_{k > 0} (Y_n^{(k)} \neq K \text{ for all } n)$$

and

$$X_{t_{k+1}, n} = \begin{cases} Y_n^{(k^*)}, & Y_n^{(k^*)} \neq K+1, \\ 0, & Y_n^{(k^*)} = K+1. \end{cases}$$

We have appended two additional levels to our state space: we can think of K as the “queue”, i.e. neurons awaiting processing, and $K+1$ as neurons which have been processed. Whenever we promote a neuron, if it isn't at level $K-1$, then no burst occurs. If it is, however, we move it to the queue. We then iterate the following algorithm until the queue is empty: pick a certain neuron at random from the queue, promote each neuron in the network which has not yet reached the queue by one level independently with probability p , leave the neurons which are already in the queue, or have been processed, where they are, and then move the neuron we picked from the queue into the processed bin.

As defined above, $Y^{(r)}$ is defined for all $r \in \mathbb{N}$, but in practice we only need consider Y until such time as the queue is empty. However, $k^* < N$, so for any burst, we need only compute a finite

set of $Y^{(r)}$. Also notice that at each stage in the burst, how we choose the index m from $A^{(r)}$ is not important either statistically or algorithmically: since all neurons are equivalent in this model, it is easier and equivalent to simply keep track of the number of neurons at each level. This would of course not be true in a network whose structure was inhomogeneous. Finally, a quick note about ρ : changing ρ doesn't qualitatively change anything about the stochastic process, and amounts only to a (deterministic) time change. We will set $\rho = 1$ below.

2.2. Phenomenology

We give a brief overview of the phenomenology of the network here; a much more detailed description is given in [16]. See Figure 1: what we observe is that the network's behavior is radically different depending upon p . When p is chosen large enough (top two panels) the network synchronizes. This can be seen in one of two ways: in (a), the burst sizes are close to periodic, or in (b), the auto-correlation of the burst sizes shows a clear signature of periodicity. However, when p is chosen small enough (panels (c,d)) the system is asynchronous and there are very poor correlations. Moreover, for p small, the maximum size of a burst is orders of magnitude smaller than for p large (in these simulations, 25 versus 800). However, one can also see that there are intermediate levels of p where the network supports both types of behavior (see Figure 2) and switches randomly between the two. The main result of this paper is to explain this switching.

2.3. Overview of mean-field

We define a mean-field approximation to this random system in Section 3., prove its validity in Section 4., and analyze many of its properties in Section 5.. The description of the mean-field is somewhat technical so we delay its explicit definition until below, but for now we give a schematic overview so that we can make the notion of "critical behavior" in this model precise.

We will define a deterministic hybrid dynamical system on a state space $S \subseteq \mathbb{R}^K$. Define some distinguished subset $T \subseteq S$, a map $G: T \rightarrow S \setminus T$, and a smooth vector field $f: S \setminus T \rightarrow \mathbb{R}^K$. Our mean-field then has the form: take any point in $S \setminus T$, flow under the ODE $x' = f(x)$ until the system reaches T , apply the map G , mapping back into $S \setminus T$, and repeat. If the flow never reaches T , then we just follow the flow for all time.

The correspondence between this and the system above is that the ODE flow corresponds to the times between big bursts, and the map G corresponds to the effect of a big burst on the system. Of course, it is *a priori* possible that some initial conditions never reach T under the flow, in which case we do not have a big burst.

This (loosely) defines a dynamical system, and the first main question is whether we get finitely many, or infinitely many, bursts. Define $T_0 = T$ and U_0 as the subset of $S \setminus T$ which flows into T under the ODE $x' = f(x)$. For each $k > 0$, define $T^k = G^{-1}(U_{k-1})$ and U^k as the subset of $S \setminus T$ which flows into T^k . From this definition, U^k is the set of initial conditions which have at least k big bursts. Finally, define $U^\infty = \bigcap_{k \geq 0} U^k$, the set of initial conditions which have infinitely many big bursts.

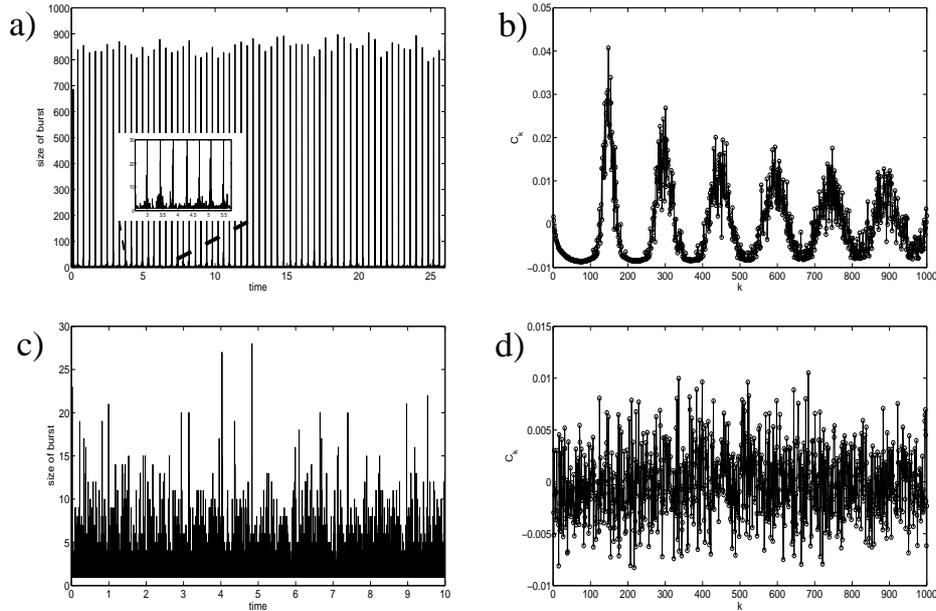


Figure 1: Network dynamics for synchronous and asynchronous states. In all of these simulations, we have chosen $N = 1000$ and $K = 10$. In panels (a,b) we chose $p = 1 \times 10^{-2}$, which leads to synchrony. Panel (a) is a plot of the size of the bursts versus time: all bursts here are either quite large (near the size of the whole network) or significantly smaller. The inset in panel (a) is a zoomed view showing that between the large bursts there is activity, but all these bursts are much smaller. In panel (b) we have plotted the auto-correlation of the burst numbers, which shows a clear periodic structure. Specifically, if we denote the size of the j th burst in the simulation as b_j , then we are plotting $c_k = \langle b_j b_{j+k} \rangle / \langle b_j^2 \rangle$ for various values of k . To obtain these statistics, we simulated the full network until 10^5 bursts were obtained, and then averaged over a sliding window of length 9.9×10^4 . Panels (c,d) show the same for $p = 5 \times 10^{-3}$ which leads to asynchronous dynamics. The size of the bursts are much smaller than in the synchronous regime, and the auto-correlation shows no temporal structure. (This data was also presented in [16].)

The particular mean-field system that we consider depends on several parameters. The limit we will consider is $N \rightarrow \infty, p = \beta/N, \rho = 1$, and K fixed, so our mean-field will depend on β, K . The crucial questions will involve the set $U_{\beta,K}^\infty$ and the dynamical system restricted to this set. We will be most interested in considering K fixed and letting β change. We will show that for β sufficiently small, $U_{\beta,K}^\infty = \emptyset$, and for β sufficiently large, $U_{\beta,K}^\infty$ is the entire state space S , i.e. when β is small enough, no initial conditions give rise to infinitely many big bursts, and when β is sufficiently large, all initial conditions do. This leads to the natural definition of two critical β values

$$\beta_{c,1}(K) = \inf\{\beta : U_{\beta,K}^\infty \neq \emptyset\}, \quad \beta_{c,2}(K) = \sup\{\beta : U_{\beta,K}^\infty \neq S\}. \quad (2.1)$$

By definition, if $\beta > \beta_{c,1}(K)$, then the system undergoes infinitely many big bursts for some

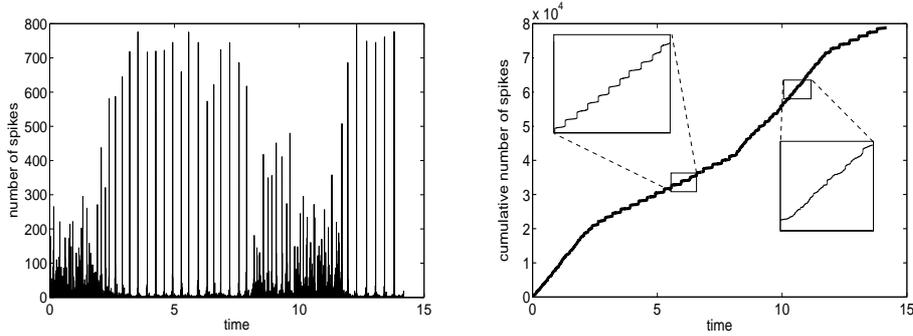


Figure 2: Here we have chosen $N = 1000$ and $K = 10$ as in Figure 1, but now $p = 9.35 \times 10^{-3}$. In the left panel we graph the size of the bursts as in the earlier figures. In the right panel, we are plotting the cumulative number of spikes up to time t as a function of t , where each spike is the firing of one neuron. Observe that the system switches between synchronous and asynchronous behavior. Also notice that the mean firing rate (slope of the cumulative plot) is significantly lower during the synchronous state even though the bursts in the synchronous state are large. We have plotted two blowups to show that in the asynchronous state, the system fires at a fairly constant rate, but in the synchronous state almost all of the firing takes place during the large bursts. (This data was also presented in [16].)

initial conditions. Also by definition, if $\beta < \beta_{c,2}(K)$, then the system does not undergo infinitely many big bursts for some initial conditions. Therefore, if $\beta_{c,1}(K) < \beta_{c,2}(K)$, then for all $\beta \in (\beta_{c,1}(K), \beta_{c,2}(K))$, the mean-field system can express at least two qualitatively different types of dynamics.

In Section 5., we will demonstrate that the flow map $x' = f(x)$ has an attracting fixed point for $\beta < \beta_{c,2}(K)$. Thus, if $\beta_{c,1}(K) < \beta_{c,2}(K)$, then there exists an interval of β 's such that the mean-field has at least two attractors: one solution with infinitely-many big bursts corresponding to the synchronous dynamics, and one with no big bursts corresponding to the asynchronous dynamics. We will further show that for any fixed $\theta > 1/2$, if K is sufficiently large, then $\beta_{c,1}(K) < \theta K$, so that the dynamical system is bistable for $\beta \in [\theta K, K]$.

The import of all of this is that it explains the switching seen in the simulations. In Section 4., we will show that the probability of a realization of the stochastic system being far away from the mean-field, on any finite time interval, is (almost) exponentially small in N . Thus the bistability in the mean-field allows for switching in the stochastic system, but only on long timescales — and this is exactly what has been observed numerically.

3. Mean-field ($N \rightarrow \infty$) limit — description

As seen in Section 2.2., this system essentially has two phases of behavior: a region in time where many “small bursts” occur (and sometimes this region lasts forever) and occasionally, a “big burst” in the system (where by “big” we loosely mean “of $O(N)$ ”).

One of the main results of this paper is that we can make this characterization precise, and show that the system is well approximated by a “mean-field limit” which consists of an ODE coupled to a map — the ODE corresponds to the dynamical behavior when there is no big burst, and the map corresponds to the effect of a big burst.

In this section we define and formally motivate the mean-field limit and give a short exposition on its properties (the proof of its validity is delayed until Section 4.). The mean field is the limit in which we fix $K \in \mathbb{N}, \rho = 1$, and let $N \rightarrow \infty, pN \rightarrow \beta$ and thus should only depend on β, K .

3.1. Generic mean-field

We consider a dynamical system with state space

$$S_K^+ = \left\{ x \in \mathbb{R}^K : 0 \leq x_k \leq 1, \sum_{k=0}^{K-1} x_k = 1 \right\}. \quad (3.1)$$

NB: We will always index our vectors starting at index 0 instead of 1, i.e. when we say $x \in \mathbb{R}^K$, we write its components as $(x_0, x_1, \dots, x_{K-1})$.

Define the “big burst domain” as

$$D(G_{\beta,K}) := \{x \in S_K^+ : x_{K-1} > \beta^{-1} \vee (\exists j : x_j > \beta^{-1} \wedge (x_k = \beta^{-1} \text{ for } j < k \leq K-1))\}. \quad (3.2)$$

To describe the mean-field system, choose two functions

$$G_{\beta,K} : D(G_{\beta,K}) \rightarrow (S_K^+ \setminus D(G_{\beta,K})), \quad f : (S_K^+ \setminus D(G_{\beta,K})) \rightarrow \mathbb{R}^K.$$

We will specify the exact form of these functions in Section 3.2. below; once these functions are chosen, we can define the mean-field solution with initial condition $\xi_0 \in S_K^+ \setminus D(G_{\beta,K})$. Define ϕ as the flow generated by f , namely $\phi(\xi_0, \tau) := x(\tau)$, where $x(t)$ satisfies the ODE

$$\frac{d}{dt}x(t) = f(x(t)), \quad x(0) = \xi_0.$$

Define

$$s^* : S_K^+ \setminus D(G_{\beta,K}) \rightarrow \mathbb{R} \cup \{\infty\}, \\ x \mapsto \inf_{\tau > 0} \{\tau : \phi(x, \tau) \in D(G_{\beta,K})\}.$$

If $s^*(x) < \infty$, then we define

$$F_{\beta,K}(x) = \phi(x, s^*(x)),$$

so

$$D(F_{\beta,K}) = \{x \in S_K^+ \setminus D(G_{\beta,K}) : s^*(x) < \infty\}.$$

It is possible that $D(F_{\beta,K}) \cup D(G_{\beta,K}) \subsetneq S_K^+$ and we will see that this is in fact true for many choices of β, K . If the dynamical system ever enters $S_K^+ \setminus (D(F_{\beta,K}) \cup D(G_{\beta,K}))$, then it will never burst again.

We can now define our dynamical system. Pick any initial condition $\xi_0 \in S_K^+ \setminus D(G_{\beta,K})$ and define

$$\xi(t) = \phi(\xi_0, t), \quad 0 \leq t < s^*(\xi_0), \quad (3.3)$$

and if $s^*(\xi_0) < \infty$, i.e. $\xi_0 \in D(F_{\beta,K})$, define

$$\xi(s^*(\xi_0)) = G_{\beta,K}(F_{\beta,K}(\xi_0)). \quad (3.4)$$

In short, we flow under the ODE until we hit the set $D(G_{\beta,K})$, and if and when we do so, we apply the map $G_{\beta,K}$. Since $\xi(s^*(\xi_0)) \in S_K^+ \setminus D(G_{\beta,K})$, we can then extend the definition of this mean-field for all time in the natural way: given $\xi_0 \in S_K^+ \setminus D(G_{\beta,K})$, define

$$0 = \tau_0 < \tau_1 < \dots < \tau_n < \dots$$

by $\tau_n - \tau_{n-1} = s^*(\xi(\tau_{n-1}))$, define

$$\xi(t) = \phi(\xi(\tau_{n-1}), t - \tau_{n-1}), \text{ for } t \in [\tau_{n-1}, \tau_n),$$

and

$$\xi(\tau_n) = G_{\beta,K}(\xi(\tau_n-)).$$

Of course, this definition allows one of the τ_j (and thus all subsequent ones) to be infinite, which will happen if the dynamical system ever leaves $D(F_{\beta,K}) \cup D(G_{\beta,K})$. The τ_j are the times of the “big bursts”: away from these τ_j the flow of $\xi(t)$ is smooth, and at the τ_j it is right-continuous. For bookkeeping purposes, we will find it convenient to keep track of the burst sizes, and we define $B_j := B(\tau_j) := t^*(\xi(\tau_j-))$.

3.2. Definition of special functions

We need to specify the functions $f, G_{\beta,K}$. To define f , we first define the matrix L_P to be the $K \times K$ matrix

$$(L_P)_{ij} = \begin{cases} -1, & i = j, \\ 1, & i = j + 1 \pmod K, \\ 0, & \text{else.} \end{cases}$$

To be consistent with the vector notation, we will index the rows and columns by $0, \dots, K - 1$. We can now define f by

$$\frac{d}{dt}x = f(x) := \mu(\gamma)L_Px, \quad (3.5)$$

where we denote $\gamma = \beta x_{K-1}$ and

$$\mu(\gamma) = \frac{1}{1 - \gamma}. \quad (3.6)$$

From (3.2), we see that $x \notin D(G_{\beta,K})$ implies $\gamma < 1$, so that (3.5) is finite on the domain of definition. We will find it convenient to change time by

$$\frac{d\tau}{dt} = \mu(\gamma)$$

in (3.5) giving the linear ODE

$$\frac{d}{d\tau}x = L_P x. \quad (3.7)$$

Equation 3.7 is the form in which we study the system in Section 5. below. Notice that the trajectories of (3.5) and (3.7) coincide (they are, however, traced out at different rates), so questions about whether a trajectory enters a given set are equivalent for the two systems. Writing out (3.7) in coordinates gives

$$x'_k = x_{k-1} - x_k,$$

where we define k periodically on the set $\{0, \dots, K-1\}$. The system is linear and subject to the constraint that $\sum x_k = 1$. Clearly the point (corresponding to an ‘‘equipartition of states’’)

$$x_{\text{ep}} := (K^{-1}, K^{-1}, \dots, K^{-1})$$

is a fixed point for (3.5). We will show below that x_{ep} is a globally-attracting fixed point for (3.5); in particular, this means that without the burst map all trajectories would converge to x_{ep} .

To define $G_{\beta,K}$, we first define

$$\chi_{\beta,K}(x, t) := -t + \sum_{i=1}^K x_{K-i} \mathbb{P}(Po(t\beta) \geq i) = -t + \sum_{i=1}^K x_{K-i} \left(1 - \sum_{j=0}^{i-1} \frac{t^j \beta^j}{j!} e^{-t\beta} \right), \quad (3.8)$$

where we denote $Po(\lambda)$ to be a Poisson random variable with mean λ . Then, for $x \in D(G_{\beta,K})$, we define

$$t^*(x) = \min_{t>0} \{t : \chi_{\beta,K}(x, t) = 0\}. \quad (3.9)$$

Define L_B as the $(K+1) \times (K+1)$ matrix (indexed $i, j = 0, \dots, K$) with

$$(L_B)_{i,j} = \begin{cases} -1, & i = j < K, \\ 1, & i = j + 1, \\ 0, & \text{else} \end{cases} \quad (3.10)$$

and define $G_{\beta,K}(x)$ to be the first K components of the vector

$$\exp(t^*(x)\beta L_B)x + t^*(x)e^{(0)}, \quad (3.11)$$

where we denote $e^{(j)}$ as the j th standard basis vector. Notice that (3.8) is the K th component of $\exp((\beta t)L_B)x$ minus t .

We now show that $0 < t^*(x) < 1$ for all $x \in D(G_{\beta,K})$. By definition, $t^*(x) > 0$. Since $\chi_{\beta,K}(x, 0) = 0$ for any x , the set $D(G_{\beta,K})$ is defined as those x so that $\chi_{\beta,K}(x, t)$ is positive in a small t neighborhood of 0. In fact, it is straightforward to check that $D(G_{\beta,K})$ is exactly those x such that the first nonzero coefficient in the Taylor series of $\chi_{\beta,K}(x, t)$ is positive. Finally, $\chi_{\beta,K}(x, t) < 1 - t$, and thus $t^*(x) < 1$.

3.3. Motivation of special functions

The definitions given above are complicated and not very intuitive, so we motivate their forms here. We first have

$$x'_k = \mu(\gamma)(x_{k-1} - x_k)$$

with periodic boundary conditions. In this model, neurons are only promoted upwards, so rate of change of the number of neurons in state k will grow as neurons come in from state $k - 1$ and deplete as they leave, and thus the positive term will be proportional to x_{k-1} and the negative term proportional to x_k . The reason for the prefactor is $\mu(\gamma)$ is that, as we will show below, the mean size of a burst is $\mu(\gamma)$. If a neuron is in state k , and we have a burst of size B , this will give B chances for a neuron to be promoted one step, so that the rate of promotion should be proportional to the mean size of the burst.

To motivate the definition of $G_{\beta,K}$, we use a minor generalization of the arguments of [18] for the size of the large component of a random graph. One way to think of this is to imagine the extended $(K + 2)$ -dimensional system given by the states

$$S_0, S_1, \dots, S_{K-1}, Q, P,$$

where the Q is the number of neurons in a “queue” and P is the number of “processed” neurons. (The levels Q and P correspond to the extra levels $K, K+1$ in the formal definition of the stochastic process.) We start with one neuron in the queue — this corresponds to a neuron firing. Then, every time we “process” a neuron — compute the neurons which it promotes because of its firing — we move it to P and move all the neurons which fired to state Q . The process terminates when $Q = 0$, i.e. the queue is empty. So, to compute the size of a burst, we need to compute the value of P when Q hits 0.

Assume $\gamma = \beta x_{K-1} > 1$. Recall that each firing neuron promotes, on average, β other neurons. The mean number of neurons promoted from state S_{K-1} to Q will then be $x_{K-1}\beta = \gamma$. Thus if $\gamma > 1$, the expected value of the queue is growing, at least at first. (If $\gamma < 1$, then we expect the queue to die out quickly with high probability, and to have a burst of $O(N)$ is exponentially unlikely.)

Now, let us assume that we have processed tN neurons through to state P , and ask how large we expect Q to be at this point in time. Clearly, Q will be made up of all neurons which started in state S_{K-i} and have been promoted i or more times, minus the ones we have processed. There are $x_{K-i}N$ neurons in S_{K-i} at the beginning, each neuron is promoted with probability β/N , and we have had tN promotion possibilities. The probability of a neuron being promoted i or more steps is exactly

$$b(\geq i; tN, \beta/N)$$

where $b(k; n, p)$ is the binomial distribution, and in the limit $N \rightarrow \infty$, $b(\cdot, tN, \beta/N) \rightarrow Po(t\beta)$. Thus the number of neurons which started in state S_{K-i} and were promoted i or more times should be roughly $x_{K-i}N\mathbb{P}(Po(t\beta) \geq i)$. Therefore we expect

$$Q + P = N \sum_{i=1}^K x_{K-i} \mathbb{P}(Po(t\beta) \geq i)$$

and thus Q should roughly be $N\chi_{\beta,K}(x, t)$ (since $P = tN$ by definition). The burst ends when this first equals zero. This would suggest $t^*(x)$ be defined as in (3.9) (we verify this below).

Now let $k > 0$. The number of neurons we expect to see in state S_k after a burst is the number which started in states S_{k-i} and were promoted exactly i times:

$$\sum_{i=0}^k x_{k-i} \mathbb{P}(Po(t^*\beta) = i).$$

Finally, the number in state S_0 after the burst will be the number which started in S_0 and did not get promoted plus the number of neurons which fired in the burst:

$$x_0 \mathbb{P}(Po(t^*\beta) = 0) + t^*.$$

4. Mean-field limit — proof of validity

Here we will show that the mean-field approximates the stochastic process quite well in the $N \rightarrow \infty$ limit. We will show three things, namely

- the times at which the “big bursts” occur for the stochastic process are close to those times when they occur in the mean-field,
- the sizes of these big bursts are close to the sizes predicted by the mean field,
- between big bursts, the stochastic process stays close to the ODE part of the mean field.

We will first state the theorem, then state the crucial three lemmas involved in the proof, and then finally give a proof of the theorem based on these lemmas. We delay the proof of the lemmas until Appendix A.

Theorem 1. *Consider any $x \in S_+^K \cap \mathbb{Q}^K$. For N sufficiently large, $Nx \in \mathbb{Z}^K$ and we can define the neuronal network process $X_t^{(N)}$ as defined above in Section 2.1. with initial condition $X_0^{(N)} = Nx$.*

Choose and fix $\epsilon, h, T > 0$. Let $\xi(t)$ be the solution to the mean-field with initial condition $\xi(0) = x$. Then there exists a sequence of times τ_1, τ_2, \dots at which the system has a big burst. Let $b_{min}(T) = \min\{B(\tau_k) : \tau_k < T\}$, i.e. b_{min} is the smallest big burst which occurs before time T , and let $m(T) = \arg \max_k \tau_k < T$, i.e. $m(T)$ is the number of big bursts in $[0, T]$.

Pick any $\alpha < b_{min}(T)$. Consider the stochastic process $X_t^{(N)}$ and denote by $T_k^{(N)}$ the times at which the $X_t^{(N)}$ has a burst of size larger than αN . Then there exists $C_{0,1}(\epsilon)$ such that for N sufficiently large,

$$\mathbb{P}\left(\sup_{j=1}^{m(T)} \left| T_j^{(N)} - \tau_j \right| > \epsilon\right) \leq C_0(\epsilon) N \exp(-C_1(\epsilon) N^{\psi(K)}), \quad (4.1)$$

and

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\sup_{j=1}^{m(T)} \sup_{t \in [\tau_j, (\tau_{j+1} \wedge T_{j+1}^{(N)}) - \epsilon]} \left| X_t^{(N)} - \xi(t - (T_j^{(N)} - \tau_j)) \right| > \epsilon \right) \leq C_0(\epsilon) N \exp(-C_1(\epsilon) N^{\psi(K)}), \quad (4.2)$$

where $\psi(K) = (K + 3)^{-1}$.

We will break up the proof of this theorem into several steps, given by Lemmas 2, 3, and 4 below. A few comments: the first estimate (4.1) simply states that the times at which the stochastic system have big bursts are close to the times predicted by the mean field. The second statement is a bit more complicated, but essentially simply says that if we make the appropriate time-shift so that the interburst times start at the same time for both processes, then in the interburst interval, the stochastic process is pathwise convergent to the mean field. Notice that in both cases, our error estimates are not quite exponential (they would be if $\psi = 1$) as the typically are in these sorts of convergence statements. These exponents are not one because the Markov chain itself is badly behaved when $N \rightarrow \infty$ simply due to the fact that as we increase N , the number of separate events which can occur grows (since a burst can be as large as N).

We denote throughout $\gamma = x_{K-1}\beta$. Clearly, if $\gamma < 1$, then $x \notin D(G_\beta)$, and if $\gamma > 1$, then $x \in D(G_\beta)$. We will characterize the burst size in the cases $\gamma < 1, \gamma > 1$ (we do not need to consider the critical case $\gamma = 1$ here). For any $x \in S_+^K \cap \mathbb{Q}^K$, for N sufficiently large $xN \in \mathbb{Z}^K$. Define $B^{(N)}(x)$ as the random variable which is the burst size given that we start the system in state

$$(x_0N, x_1N, \dots, x_{K-1}N - 1, 1, 0).$$

Lemma 2. *Let x be such that $\gamma < 1$. Then for any $b \geq 1$ and $\gamma < \gamma' < 1$, we have*

$$\mathbb{P}(B^{(N)}(x) = b) \leq \frac{b^{b-2}(\gamma')^{b-1}e^{-\gamma'b}}{(b-1)!} + O(N^{-N}). \quad (4.3)$$

From this it follows that:

$$\lim_{b \rightarrow \infty} \lim_{N \rightarrow \infty} \mathbb{P}(B^{(N)}(x) > b) = 0. \quad (4.4)$$

Equivalently: let $\alpha > \gamma e^{1-\gamma}$ (α can be chosen in $(0, 1)$ if $\gamma \in (0, 1)$). Then for b sufficiently large,

$$\mathbb{P}(B^{(N)}(x) > b) < \alpha^b. \quad (4.5)$$

Finally,

$$\lim_{N \rightarrow \infty} \mathbb{E}(B^{(N)}(x)) = \frac{1}{1-\gamma},$$

and in particular is finite for $\gamma \in [0, 1)$.

Lemma 3. *Consider x with $\gamma > 1$. Then there is a $p(\gamma) \in (0, 1)$ such that*

$$\lim_{b \rightarrow \infty} \lim_{N \rightarrow \infty} \mathbb{P}(B^{(N)}(x) > b) = p(\gamma). \quad (4.6)$$

From this it follows that there exists $\eta' > 0$ such that

$$\lim_{N \rightarrow \infty} \mathbb{P}(B^{(N)}(x) > \eta' N) > 0. \quad (4.7)$$

In fact, we can be more precise: $p(\gamma) = \tilde{\gamma}/\gamma$, where $\tilde{\gamma}$ is the unique solution in $(0, 1)$ of

$$\gamma e^{1-\gamma} = \tilde{\gamma} e^{1-\tilde{\gamma}}. \quad (4.8)$$

Define $t^*(x)$ as in (3.9). Then, for all $\eta > 0$, there exists $C_{0,1}(\eta) > 0$ such that for N sufficiently large,

$$\mathbb{P}(|B^{(N)}(x) - t^*(x)N| > \eta N | B^{(N)}(x) > \eta' N) \leq C_0(\eta) \exp(-C_1(\eta)N). \quad (4.9)$$

In fact, a slightly modified restatement of the last also holds true: say $\gamma < 1$ and there is $\eta' > 0$ so that $\chi_{\beta,K}(x, \eta') > 0$. Define $\tilde{t}_{\eta'}(x)$ to be the smallest zero of $\chi_{\beta,K}(x, t)$ with $t > \eta'$, then (4.9) holds with $t^*(x)$ replaced with $\tilde{t}_{\eta'}$.

As mentioned above, we will delay the proofs of these assertions until Appendix A, but we now discuss their content briefly. The content of Lemma 2 is simply that if $\gamma < 1$, then the probability of a burst of size B^* decays exponentially fast in B^* when N is large, so the probability of a big burst is exponentially small in N . Moreover, the expected size of a burst is bounded for any $\gamma < 1$, and, although this expectation blows up as $\gamma \rightarrow 1$ (as we would expect), it does not do so too badly. The content of Lemma 3 is that when $\gamma > 1$, we do get big bursts with some positive probability (although not with probability one!), but that whenever we do have a big burst, its size is close to its “expected size” $t^*(N)$ with probability exponentially close to one.

In short, these lemmas say that the probability of getting a big burst when the mean-field doesn't, or not getting one when the mean-field does, is exponentially small in N .

Lemma 4. Choose $x, X_t^{(N)}, T = \tau_1, T_1^{(N)}$ as in Theorem 1, where we also require that $x \notin D(G_\beta)$. By definition, for $t < \tau_1$, the solution $\xi(t)$ to the mean-field has no big bursts, and in particular, $\xi_{K-1}\beta < 1$ during this time. Define the events

$$A_1^\epsilon := \left\{ \omega : \left| T_1^{(N)}(\omega) - \tau_1 \right| > \epsilon \right\}, \quad (4.10)$$

$$A_2^\epsilon := \left\{ \omega : \sup_{t \in [0, \tau_1 \wedge T_1^{(N)} - \epsilon]} \left| \xi(t) - N^{-1} X_t^{(N)} \right| > \epsilon \right\}, \quad (4.11)$$

$$A_3^\epsilon := \left\{ \omega : \left| B(T_1^{(N)}) - t^*(\xi(\tau_1))N \right| > \epsilon N \right\}. \quad (4.12)$$

$$(4.13)$$

Then for all $\epsilon > 0$, there are $C_{0,1}(\epsilon) > 0$ so that for N sufficiently large,

$$\mathbb{P}(A_1^\epsilon \cup A_2^\epsilon \cup A_3^\epsilon) \leq C_0(\epsilon)N \exp(-C_1(\epsilon)N^{\psi(K)}), \quad (4.14)$$

where we define $\psi(K) = (K + 3)^{-1}$.

The lemma says that the probability of each of the three events occurring becomes exponentially unlikely for N large. The first event being unlikely says that the time of the first big burst for the stochastic system is likely close to that of the deterministic system. The second event being unlikely says that the stochastic process likely stays near to the mean field for all of this time, and finally the third event being unlikely says that the size of the first big burst, once it arrives, is likely close to that of the mean field. Notice that we excise some of the time domain in the definition of A_2^ϵ ; the reason for this is the non-uniformity of the process. As we approach the big burst, $\mu(\gamma) \rightarrow \infty$ (in graph-theoretic language, the random graph becomes “critical”) and this should lead to some interesting dynamics near criticality. We do not need to consider that here, however; from Lemma 3 we have good control on the size and time of the big burst when it does occur.

Proof of Theorem 1.

Finally, we see that given our three lemmas the proof of the theorem is clear. Consider a solution to the mean field with $\xi(0) \notin D(G_\beta)$ and pick $T > 0$. If the solution has no big bursts, then Lemma 4 gives the theorem directly, so assume that we have at least one big burst. However, in time T , the solution will have at most finitely many big bursts; denote this number $m(T)$. From Lemmas 2, 3, it is clear that (4.1) is satisfied for $j = 1$, and then it is clear from Lemma 4 that (4.2) holds for $j = 1$.

Now restart the process (with a reshifted time) at $\tau_2 \wedge T_2^{(N)}$. Of course, the initial conditions might differ by ϵ , but we can modify $\xi(\tau_1)$ to be equal to $X_{T_1^{(N)}}$, and using standard Gronwall’s estimates for the ODE the error introduced can be controlled. Then we bootstrap and apply Lemmas 2—4 again. Since we have at most $m(T)$ steps in this process, the estimates hold. \square

5. Analysis of mean-field

In this section we will study the mean-field system and the critical parameters $\beta_{c,1}(K), \beta_{c,2}(K)$ as defined in (2.1). We will show analytically that $\beta_{c,2}(K) = K$ and $\beta_{c,1}(K) \geq 1$ for all K , and that $\beta_{c,1}(K) < \beta_{c,2}(K)$ for K sufficiently large, thus giving a region of bistability in the mean-field system. We will also present the results of a numerical study which suggests that this bistability exists for $K \geq 6$.

5.1. Existence and stability of fixed point, value of $\beta_{c,2}(K)$

First we show that $x' = L_P x$ has a unique attracting fixed point in $S_K^+ \setminus D(G_{\beta,K})$. We see that $I + L_P$ is the $K \times K$ matrix which shifts vectors in \mathbb{R}^K one index to the right. Define $\zeta = e^{i2\pi/K}$, and

$$v_j = (1, \zeta^j, \zeta^{2j}, \dots, \zeta^{(K-1)j}),$$

and we have $(I + L_P)v_j = \zeta^{-j}$. The eigenvalues of L_P are thus $\lambda_j := -1 + e^{i2\pi k/K}$, $k = 0, \dots, K - 1$. All of these eigenvalues have negative real parts except $\lambda_0 = 0$. Thus the only fixed point of (3.5) is the vector whose entries are all the same, and subject to the normalization $\sum_k x_k = 1$, this gives x_{ep} . To see that x_{ep} is attracting, consider any perturbation $x_{\text{ep}} + \epsilon v$. If the

coefficients of $x_{\text{ep}} + \epsilon v$ sum to one, then the coefficients of v sum to zero. Since the system is linear, the evolution of the error is given by $v' = L_P v$ as well.

Let us write $v = \sum_{k=0}^{K-1} \alpha_k v_k$, where v_k are the eigenvectors defined above. Then

$$e^{tL_P} v = e^{tL_P} \left(\sum_{k=0}^{K-1} \alpha_k v_k \right) = \sum_{k=0}^{K-1} \alpha_k e^{t\lambda_k} v_k \rightarrow \alpha_0 v_0.$$

On the other hand, notice that if $v' = e^{tL_P} v$, then

$$\frac{d}{dt} \sum_{k=0}^{K-1} v_k = \sum_{k=0}^{K-1} v_{k-1} - \sum_{k=0}^{K-1} v_k = 0,$$

so we know the coefficients of $e^{tL_P} v$ must also sum to zero. Since the coefficients of v_0 sum to one, this means $\alpha_0 = 0$, and thus all perturbations decay to zero exponentially fast. (In fact, we can even get the decay rate, which is the largest real part of the nonzero eigenvalues, which is $1 - \cos(K^{-1})$.)

Therefore, if $\beta > K$, then any initial condition for (3.5) will enter $D(G_{\beta,K})$, whereas if $\beta < K$, some initial conditions will not, and thus $\beta_{c,2}(K) = K$.

5.2. Determination of $\beta_{c,1}(K)$, K large

The purpose of this section is to estimate $\beta_{c,1}(K)$ for large K . We will show that for K sufficiently large, $\beta_{c,1}(K) < K$, thus giving a bistability band; namely, for β in this region, some initial conditions will decay to the attracting fixed point, and others will undergo infinitely many big bursts. The main idea of the argument is to define a set $N_{\beta,K}$ and show that, for K sufficiently large and $\beta > K/2$, $N_{\beta,K} \subseteq U_{\beta,K}^\infty$ and also $N_{\beta,K} \neq \emptyset$.

We now define the set $N_{\beta,K}$. First, we define

$$O_{\beta,K} := \{x \in S_+^K \setminus D(G_{\beta,K}) : (\phi(x, t))_{K-1} < (10\beta)^{-1}, \forall t > 0\}.$$

One can think of these as initial conditions who never get close to a big burst, since under the flow of ϕ the $K-1$ 'st component never gets near β^{-1} . For notation, we now refer to the ‘‘left’’ and ‘‘right’’ modes of this system; specifically we define $L = \{1, 2, \dots, \lfloor \frac{K}{2} \rfloor\}$ and $R = \{\lfloor \frac{K}{2} \rfloor + 1, \dots, K-1\}$ (note that neither of these sets contain the 0 index!). We also define the projections P_L, P_R in the obvious way ($(P_L x)_i = \delta_{i \in L} x_i$, etc.). We then say $x \in S_+^K \setminus D(G_{\beta,K})$ is in $N_{\beta,K}$ if the following three conditions hold:

$$\sum_{i \in L} x_i < K^{-10}, \quad \sum_{i \in R} x_i < \frac{100}{\sqrt{\log(K)}}, \quad \text{and } P_R x \in O_{\beta,K}.$$

(From this it follows that $x_0 > 1 - K^{-10} - 100/\sqrt{\log(K)}$, as $\sum x_i = 1$. We will write $x_0 = 1 - \delta$ and use later that $\delta \rightarrow 0$ as $K \rightarrow \infty$.) In words, what we require is that there is very little mass in the left modes, little (but potentially much more) mass in the right modes, and if we simply flow under the mass contained in the right modes, we'd never precipitate a big burst. Clearly $e^{(0)} \in N_{\beta,K}$ and thus $N_{\beta,K}$ is not empty. All that remains to show is

Theorem 5. *Choose any $\theta > 1/2$. For K sufficiently large,*

$$N_{\theta K, K} \subseteq U_{\theta K, K}^{\infty}.$$

Thus $\beta_{c,1}(K) \leq \theta K$ when K is large enough, and this implies $\beta_{c,1}(K) < \beta_{c,2}(K) = K$. Basically, the proof consists of showing that for any $x \in N_{\beta, K}$, we have $s^*(x) < \infty$ (so that $F_{\beta, K}(x)$ is defined) and then we show that $G_{\beta, K} \circ F_{\beta, K}(x) \in N_{\beta, K}$. Thus every point in $N_{\beta, K}$ will eventually undergo a burst, and after the big burst map finds itself back in $N_{\beta, K}$ — thus each of these points undergo infinitely many big bursts. We delay the full proof until Appendix B below.

5.3. Calculation of $\beta_{c,1}(K)$, K small

Here we present some discussion for specific small values of K . One observation which we make in all cases is that when we choose $\beta > \beta_{c,2}(K)$, the system seems to be attracted to a periodic orbit where all the bursts are the same size, which size we denote $t^*(\beta, K)$. (Note that while $\beta > \beta_{c,2}(K)$ implies that there by infinitely many big bursts, we have not shown that there the burst sizes are independent of initial condition, or that they become a constant sequence, etc.)

We will first show analytically that $\beta_{c,2}(2) = 2$, that $t^*(\beta, 2)$ is well-defined for $\beta > 2$, and that $t^*(\beta, 2)$ is continuous at $\beta = 2$. This contrasts significantly with the case where K is large as shown in Section 5.2.. We will then present some numerical evidence detailing the transition from K “small” to K “large”; in particular, we will show that once K has reached 6, it seems to be the same as the K large case, and the transition from 2 to 6 shows some surprises as well.

To see that $\beta_{c,2}(2) = 2$, note that $K = 2$ is really just a one degree-of-freedom system, since $\sum x_k = 1$. So, consider any $\beta < 2$ and an initial condition $x \notin D(G_{\beta, K})$. One of the following three cases must be true: $x_1 > \beta^{-1}$, $x_0 < x_1 \leq \beta^{-1}$, or $x_0 \geq x_1$. In the first case, $x \in D(G_{\beta, K})$, and $G_{\beta, K}(x)$ is then in the third case. In either of the two last cases, it’s easy to see that the system will decay to $(1/2, 1/2)$, and thus have no further big bursts.

Now we show $t^*(\beta, 2)$ is continuous at $\beta = 2$, and in fact we will actually compute the Taylor expansion to the right of $\beta = 2$. Using (3.8) and noting that the big burst occurs when $x = ((\beta - 1)/\beta, 1/\beta)$, we have

$$\chi_{\beta, 2}(x, t) = -t + 1 - e^{-t\beta} - (\beta - 1)e^{-t\beta}$$

Writing $\beta = 2 + \epsilon$,

$$\begin{aligned} \partial_t \chi_{\beta, 2}(x, t)|_{t=0} &= 0, \\ \partial_{tt} \chi_{\beta, 2}(x, t)|_{t=0} &= 2\beta(\beta - 1) - \beta^2 = 2\epsilon + O(\epsilon^2), \\ \partial_{ttt} \chi_{\beta, 2}(x, t)|_{t=0} &= 3\beta^2(1 - \beta) + \beta^3 = -4 - 12\epsilon + O(\epsilon^2). \end{aligned}$$

Since the third derivative is of the opposite sign, and of larger order, than the second, we can find the root by the third-order Taylor series truncation, and thus we have

$$t^*(2 + \epsilon, 2) = \frac{\epsilon}{2} + o(\epsilon),$$

and thus $t^*(\beta, 2)$ is continuous at $\beta = 2$, and moreover

$$\frac{\partial}{\partial \beta} t^*(\beta, 2)|_{\beta=2+} = \frac{1}{2}.$$

For $K > 2$, we present a series of numerical experiments designed to compute $t^*(\beta, K)$, a synopsis of which is presented in Figure 3. We reiterate that $t^*(\beta, K)$ is not *a priori* well-defined for $K > 2$. All we have shown is that for $\beta > \beta_{c,1}(K)$, some initial conditions give rise to dynamics which include infinitely many big bursts. However, we always observe that whenever $\beta > \beta_{c,2}(K)$, the system seems to be attracted to a unique periodic orbit with all bursts of the same size, so we define $t^*(\beta, K)$ to be this observed size.

To compute $t^*(\beta, K)$, we first need to set three discretization parameters, `tol`, `intLen`, and `dt`. We then compute t^* as follows: We consider the initial condition $e^{(0)}$, and flow under (3.5) until we hit $D(G_{\beta,K})$. To determine whether this has occurred numerically, we discretize (3.5) using a first-order Euler method with timestep `dt`, and if we ever get within `tol` of $D(G_{\beta,K})$, we say we've had a burst, and if we ever integrate the ODE for time longer than `intLen`, we say we have decayed to the attracting fixed point. If we have determined that we have entered $D(G_{\beta,K})$, we need to determine t^* , but to do so we simply use a standard bisection method to find a smallest root of $\chi_{\beta,K}$, and we verify that this bisection method has converged to within `tol`. In all of the numerics presented here, we chose `dt` = 10^{-4} , `tol` = 10^{-6} , `intLen` = 10^2 .

As seen in Figure 3, as we increase K we have a transition from the $K = 2$ case to the K large case. We observe that for all $K \leq 5$, $\beta_{c,1}(K) = K$, i.e. that there is no bistable regime. However, something interesting seems to happen at $K = 5$: even though $\beta_{c,1}(5) = 5$, numerical evidence suggests that $\lim_{\beta \rightarrow 5+} t^*(\beta, 5) > 0$. (Contrast Figures 3(a) and (b).) Once we increase K to 6 (Figure 3(c)), $\beta_{c,1}(K)$ moves to the left of K , giving a bistable regime, but this bistable regime is not very large in β . Finally, as we increase K further, $\beta_{c,1}(K)/K$ decreases. The numerically computed values up to four decimal places are

K	$\beta_{c,1}(K)$	$\lim_{\beta \rightarrow \beta_{c,1}(K)+} t^*(\beta, K)$
4	4.000	0.0000
5	5.000	0.3901
6	5.973	0.5529
10	9.414	0.7402

6. Conclusions

It is well-known that many neuronal network models (and real neuronal networks!) have the ability to exist in several states (many networks exhibit both synchronous and asynchronous behaviors), and when perturbed by noise, will switch among these several states. We have considered a very idealized model of a neuronal network in this paper, but we have been able to rigorously explain why the network switches between competing states: it is a small stochastic perturbation of a deterministic system. We hope that the ideas contained in this paper can be extended to more

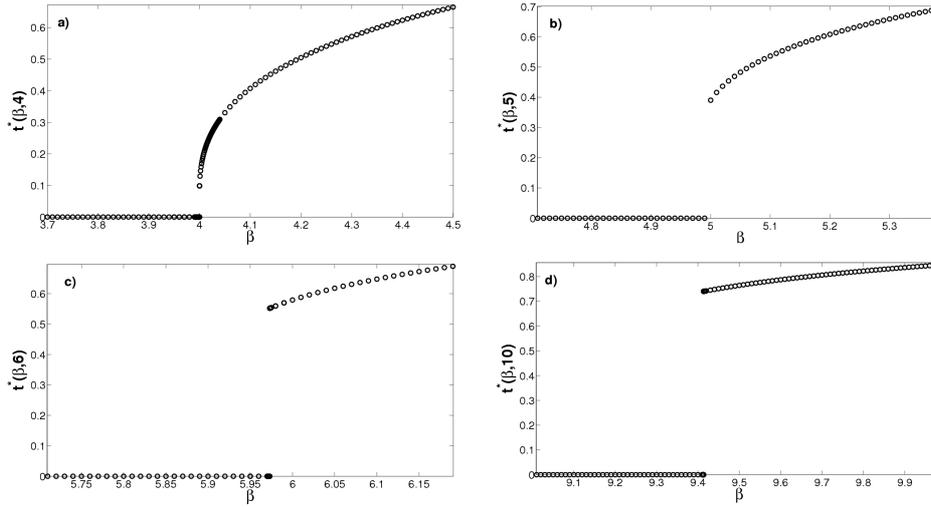


Figure 3: We plot the numerically-computed value of $t^*(\beta, K)$ for $K = 4, 5, 6, 10$ and various values of β . As mentioned in the text, it is observed that the mean-field only limits onto two attractors: the equipartition or the big burst solution. In each case, we are plotting the size of the big burst to which our system is attracted. One sees a wide variety of different behaviors depending on K : for K small enough, we have $\beta_{c,1}(K) = K$, and t^* is a continuous function of β , but for K large enough, the function t^* becomes discontinuous, and $\beta_{c,1}(K) < K$.

complicated network models and can explain the wealth of behavior observed there (e.g. it was shown in [17] that a finite-size effect can lead to non-trivial synchronization behavior in a more realistic network model.)

There are also several open questions left unresolved about this model. Numerical evidence suggests the conjecture that for all $\beta > \beta_{c,2}(K)$, the mean-field system always limits onto one special infinite-burst periodic orbit, thus leading to the definition of a canonical “burst size” $t^*(\beta, K)$ associated to β and K . Moreover, it is observed that t^* is always an increasing function of β for fixed K , and for K large enough $t^*(\beta, K)$ has a discontinuity at $\beta_{c,2}(K)$, or a first-order phase transition. This is in stark contrast with the Erdős-Rényi model (corresponding to $K = 1$), for which $t^*(\beta)$ is a continuous function of β , and actually right-differentiable there: $t^*(1+\epsilon, 1) = 2\epsilon + O(\epsilon^2)$. We saw numerically in Section 5.3. that for $K = 2, 3, 4$, the behavior of t^* is much like that for the Erdős-Rényi model, but for $K > 4$ it is quite different. Moreover, the evidence seems to suggest that as we increase K , $\beta_{c,1}(K)/K$ decreases, and moreover

$$\lim_{\beta \rightarrow \beta_{c,1}(K)+} t^*(\beta, K)$$

seems to increase as well. The numerical evidence suggests that

$$\lim_{K \rightarrow \infty} \frac{\beta_{c,1}(K)}{K} = 0, \quad \lim_{K \rightarrow \infty} \lim_{\beta \rightarrow \beta_{c,1}(K)+} t^*(\beta, K) = 1.$$

A Proof of Lemmas 2– 4.

To prove Lemmas 2, 3, we will need a few formulas concerning the size of the burst. We will compute the size of the burst by “stages”. We start with $Q_0 = 1$, with one neuron in the queue. To move from Q_n to Q_{n+1} , we process all of the Q_n neurons in the queue, and see how many are promoted. If this is ever zero, then the burst stops, and the total burst size is $\sum Q_i$. (This is formally different from the definition above, but since all the neurons are interchangeable, we can see that this is an equivalent formulation.)

First notice that whenever we process the effect of a neuron in the queue, we promote each neuron in state S_{K-1} with probability $p = \beta/N$. The total number of neurons which are promoted are then given by the binomial distribution; in particular, we promote k neurons with probability

$$b(k; x_{K-1}N, \beta/N) = \binom{x_{K-1}N}{k} \left(\frac{\beta}{N}\right)^k \left(1 - \frac{\beta}{N}\right)^{x_{K-1}N-k}.$$

Most of the calculations below involve getting upper and lower estimates for various convolutions of the Bernoulli distribution. We will find it convenient to use the notation

$$f(N) \lesssim g(N)$$

if we have

$$f(N) \leq g(N) \text{ for all } N, \quad f(N) - g(N) = O(N^{-1}) \text{ as } N \rightarrow \infty.$$

Thus this symbol corresponds to an explicit upper bound in one direction and asymptotic equivalence in the other direction.

Lemma 6. *For $C \geq 0$, we have*

$$b(k; x_{K-1}N - C, \beta/N) \lesssim \frac{\gamma^k e^{-\gamma}}{k!},$$

where $\gamma = x_{K-1}\beta$. Moreover, for any $\gamma' > \gamma$,

$$\mathbb{P}(Q_{n+1} = k | Q_n) \lesssim \frac{(\gamma Q_n)^k e^{-Q_n \gamma}}{k!} + O(N^{-N}).$$

Proof. To prove the first claim, we simply note that

$$\alpha N \log(1 - \beta/N) \lesssim -\alpha\beta,$$

from a Taylor series expansion, and of course

$$\prod_{i=0}^k (1 - i/N) \lesssim 1,$$

and the first claim follows. To prove the second, let us first make the assumption that the number of neurons in state S_{K-1} stays lower than $x_{K-1}N$ throughout the burst. Then we have:

$$\begin{aligned}
\mathbb{P}(Q_{n+1} = k | Q_n) &= \sum_{k_1 + \dots + k_{Q_n} = k} B(k_1; x_{K-1}N, \beta/N) B(k_2; x_{K-1}N - k_1, \beta/N) \times \\
&\quad \dots \times B(k_{Q_n}; x_{K-1}N - k_1 - k_2 - \dots - k_{Q_n-1}, \beta/N) \\
&\lesssim \sum_{k_1 + \dots + k_{Q_n} = k} \frac{\gamma^{k_1} e^{-\gamma}}{k_1!} \frac{\gamma^{k_2} e^{-\gamma}}{k_2!} \dots \frac{\gamma^{k_{Q_n}} e^{-\gamma}}{k_{Q_n}!} \\
&= \gamma^k e^{-Q_n \gamma} \sum_{k_1 + \dots + k_{Q_n} = k} \frac{1}{k_1! k_2! \dots k_{Q_n}!} \\
&= \gamma^k e^{-Q_n \gamma} \frac{Q_n^k}{k!} \\
&= \frac{(\gamma Q_n)^k e^{-Q_n \gamma}}{k!}.
\end{aligned}$$

Now consider $\gamma < \gamma' < 1$. Define ξ so that $\xi\beta = \gamma'$. If we know that the number of neurons in state S_{K-1} stays below ξN throughout the burst, then the previous estimate holds with γ replaced by γ' . So, we need to compute the probability that the number of neurons in state S_{K-1} ever exceeds ξN . Defining $\tilde{\xi} = \xi - x_{K-1}$, it is easy to see that this probability is bounded above by

$$b(\geq \tilde{\xi}N; (1 - x_{K-1})N, \beta/N) \leq \frac{(\beta - \gamma)^{\xi N} e^{-(\beta - \gamma)}}{(\xi N)!},$$

and using the very coarse version of Stirling's approximation that there is $C > 0$

$$A! \leq CA^{A+1/2} e^{-A},$$

shows that this probability is bounded above by a constant times N^{-N} . □

Proof of Lemma 2.

Throughout this proof, for ease of notation, we will neglect to write the error term of size $O(N^{-N})$ in every equation.

Given a random sequence Q_0, Q_1, \dots , define $n^* = \min_n Q_n = 0$ and then the size of the burst is $\sum_{i=0}^{n^*} Q_i$. So we have

$$\begin{aligned}
\mathbb{P}(Q_i = q_i, i = 1, \dots, n^*) &= \mathbb{P}(Q_1 = q_1) \prod_{i=2}^{n^*} \mathbb{P}(Q_i = q_i | Q_{i-1} = q_{i-1}) \\
&\lesssim \frac{\gamma^{q_1} e^{-\gamma}}{q_1!} \prod_{i=2}^{n^*} \frac{(\gamma q_{i-1})^{q_i}}{q_i!} e^{-\gamma q_{i-1}} = \frac{\gamma^{\sum q_i} e^{-\gamma(1 + \sum q_i)}}{q_1! q_2! \dots q_{n^*}!} \prod_{i=2}^{n^*} q_{i-1}^{q_i}
\end{aligned}$$

and of course

$$\mathbb{P}\left(\sum_{i=1}^{n^*} Q_i = B - 1\right) = \sum_{q_1+q_2+\dots+q_{n^*}=B-1} \mathbb{P}(Q_i = q_i, i = 1, \dots, n^*).$$

Recall that by definition $Q_0 = 1$, so this last expression is exactly the probability of the burst size being B . Thus we have

$$\mathbb{P}(\text{burst} = B) = \gamma^{B-1} e^{-\gamma B} \sum_{q_1+q_2+\dots+q_{n^*}=B-1} \frac{\prod_{i=2}^{n^*} q_{i-1}^{q_i}}{q_1! q_2! \dots q_{n^*}!}.$$

For any vector of integers \mathbf{q} , define

$$C_{\mathbf{q}} = \frac{\prod_{i=2}^{n^*} q_{i-1}^{q_i}}{q_1! q_2! \dots q_{n^*}!},$$

and we are left with

$$\mathbb{P}(\text{burst} = B) = \gamma^{B-1} e^{-\gamma B} \sum_{|\mathbf{q}|=B-1} C_{\mathbf{q}}.$$

Using the fact that

$$\sum_{|\mathbf{q}|=B-1} C_{\mathbf{q}} = \frac{B^{B-2}}{(B-1)!}, \quad (\text{A1})$$

which we prove in Lemma 8 below, gives (4.3). Now we apply Stirling's approximation to the $(B+1)$ st term in the sum, namely

$$\begin{aligned} \mathbb{P}(\text{burst} = B + 1) &= \frac{\gamma^B e^{-\gamma(B+1)} (B+1)^{B-1}}{B!} \approx \frac{\gamma^B e^{-\gamma(B+1)} (B+1)^{B-1} e^B}{\sqrt{2\pi B} B^B} \\ &= \frac{e^{-\gamma}}{\sqrt{2\pi}} B^{-3/2} \left(\frac{B+1}{B}\right)^{B-1} \gamma^B e^{B(1-\gamma)} \approx \frac{e^{1-\gamma}}{\sqrt{2\pi}} (\gamma e^{1-\gamma})^B B^{-3/2}. \end{aligned}$$

Define $\phi(\alpha, B) = \alpha^B B^{-3/2}$. Using the Stirling approximation that

$$B! = \sqrt{2\pi B} B^B e^{-B} e^{\lambda_B}$$

where $(12B+1)^{-1} \leq \lambda_B \leq (12B)^{-1}$ for all B , it is straightforward to show that

$$\mathbb{P}(\text{burst} = B + 1) \in \left(\frac{e^{-1/12}}{\sqrt{2\pi}} \phi(B), 1 + \frac{e}{\sqrt{2\pi}} \phi(B) \right) \approx (0.367\phi(B), 1 + 1.08\phi(B)).$$

Define

$$\Psi(\alpha) = \sum_{B=1}^{\infty} \alpha^B B^{-1/2}$$

and we have

$$\mu(\gamma) = \mathbb{E}(B^{(N)}(x)) \in (0.367\Psi(\alpha), 1 + 1.08\Psi(\alpha)),$$

where $\alpha = \gamma e^{1-\gamma}$. In particular, this implies that $\mu(\gamma) < \infty$ for $\gamma < 1$.

To prove that $\mu(\gamma) = (1 - \gamma)^{-1}$, we will show that

$$\mu(\gamma) = 1 + \gamma\mu(\gamma).$$

Let us now define $\mu(\gamma|k)$ as the mean size of a burst with $Q_0 = k$ (as opposed to the normal burst where Q_0 is always 1). Notice that if we can show

$$\mu(\gamma|k) = k\mu(\gamma|1) = k\mu(\gamma), \tag{A2}$$

then we are done, since by the law of total probability,

$$\mu(\gamma) = 1 + \sum_k \mu(\gamma|k)\mathbb{P}(Q_1 = k) = 1 + \sum_k k\mathbb{P}(Q_1 = k)\mu(\gamma) = 1 + \gamma\mu(\gamma).$$

To establish (A2), we define $\mu(\gamma|k)$ as above. Now, let us also consider the auxiliary stochastic process where we compute k bursts B_1, \dots, B_k and define

$$\tilde{\mu}(\gamma|k) = \mathbb{E} \left(\sum_{i=1}^k B_i \right).$$

It is easy to see that $\tilde{\mu}(\gamma|k) \geq \mu(\gamma|k)$, and, in fact, the only time the two stochastic processes differ is when, in the second case, two different bursts B_i happen to promote the same neuron. Given that a finite number of neurons are promoted in each case, the probability that the same neuron is promoted in different B_i must scale like N^{-2} , but since there are at most N neurons the total probability of this ever happening is $O(N^{-1})$. Since each random variable has a finite mean, the difference in expectations can be bounded above by $O(N^{-1})$. □

Proof of Lemma 3. From Lemma 2 we know that for $\gamma < 1$,

$$\sum_{b=1}^{\infty} \mathbb{P}(B^{(N)}(x) = b) = 1.$$

We want to show that this quantity is less than 1 when $\gamma > 1$. Define

$$A(\gamma) := \sum_{b=1}^{\infty} \frac{b^{b-2}\gamma^{b-1}e^{-\gamma b}}{(b-1)!},$$

and we compute that

$$A(\gamma) = \gamma^{-1} \sum_{b=1}^{\infty} \frac{(\gamma e^{-\gamma})^b b^{b-1}}{(b-1)!}.$$

Define $A(\gamma) = \gamma^{-1}\tilde{A}(\gamma)$ and notice that $\tilde{A}(\gamma)$ depends on γ *only through the quantity* $\gamma e^{-\gamma}$. Thus, for any $\gamma < 1$ (resp. $\gamma > 1$) we define $\gamma' > 1$ (resp. $\gamma' < 1$) to satisfy $\gamma e^{-\gamma} = \gamma' e^{-\gamma'}$, and we have $\tilde{A}(\gamma') = \tilde{A}(\gamma)$. Thus whenever $\gamma > 1$,

$$A(\gamma) = \frac{\tilde{A}(\gamma)}{\gamma} = \frac{\tilde{A}(\gamma')}{\gamma} = \frac{\gamma'}{\gamma},$$

since $\gamma' < 1$.

Finally, to establish (4.9) is straightforward using the Central Limit Theorem. \square

To prove Lemma 4, we will use Kurtz' theorem for continuous time Markov processes [31, 32] (for an excellent exposition see [40, p.76]), which we state without proof:

Theorem 7 (Kurtz' Theorem for Markov chains). *Consider a Markov process with generator*

$$L^{(N)}f(x) = \sum_{j=0}^J N\lambda_j(x)(f(x + v_j/N) - f(x)), \quad (\text{A3})$$

where $\lambda_j(x)$ are Lipschitz continuous and bounded. Let $X_t^{(N)}$ be a realization of (A3). If we define $x^\infty(t)$ by the solution to the ODE

$$\frac{d}{dt}x^\infty(t) = \sum_{j=0}^J \lambda_j(x)v_j, \quad x^\infty(0) = X_0^{(N)}, \quad (\text{A4})$$

then for any $\epsilon > 0, T > 0$, there exist $C_1, C_2 > 0$ such that for every N ,

$$\mathbb{P} \left(\sup_{t \in [0, T]} |x^\infty(t) - X_t^{(N)}| > \epsilon \right) \leq C_1(\epsilon, T) \exp(-C_2(\epsilon, T)N/\bar{v}^2\bar{\lambda}J),$$

where

$$\bar{\lambda} = \sup_{x \in \mathbb{R}^k} \sup_{j=1}^J \lambda_j(x), \quad \bar{v} = \sup_{j=1}^J |v_j|,$$

and $\lim_{\epsilon \rightarrow 0} \epsilon^{-2}C_2(\epsilon, T) \in (0, \infty)$ for all T .

We do not prove this theorem; we have followed almost exactly the formulation of Theorem 5.3 of [40], with the exception that we have made the dependence on $J, \bar{v}, \bar{\lambda}$ explicit. The need for this restatement will be made clear below: if we could assume that J is fixed then each of the terms $J, \bar{v}, \bar{\lambda}$ are *a priori* bounded and then can be subsumed into the definition of $C_2(\epsilon)$. However, we will need to take $J = J(N)$ and thus the dependence of the constant on these quantities will play a role below.

In short, what Theorem 7 says is that the stochastic process stays close to its mean with probability exponentially close to one on any finite time interval.

Proof of Lemma 4. We first give a sketch of the main points of the proof of the lemma. We proceed as follows:

1. Choose $B^* > 0$ and $\epsilon > 0$. We consider only those realizations of $X_t^{(N)}$ which do not have a burst of size greater than B^* for $t \in [0, \tau_1 - \epsilon]$. Define this set of realizations to be Ω_{B^*} . From Lemma 2, it follows that there are $C_3(\epsilon), C_4(\epsilon)$ so that

$$\mathbb{P}(\Omega_{B^*}) \geq 1 - C_3(\epsilon)N \exp(-C_4(\epsilon)B^*).$$

Notice the presence of the prefactor N ; this appears since we have $O(N)$ opportunities to have an untimely big burst.

2. On Ω_{B^*} , we can replace the generator of the process $X_t^{(N)}$ with the one which is the same but which explicitly caps all bursts at size B^* . We will write down this generator in the form of

$$L^{(N)}f = \sum_{j=0}^{J(B^*)} N\lambda_j(x)(f(x + v_j/N) - f(x)),$$

(notice that J now depends on B^*), and we will show that the mean-field equation as defined in (A4) gives the same ODE as in (3.5) for $\gamma < 1$.

3. We now let $B^* \rightarrow \infty$ as $B^* = N^p$ for some $p \in (0, 1)$. We will show that there is a C_2 so that

$$J(B^*) \leq (B^*)^K + K, \quad \bar{v} \leq KB^*, \quad \bar{\lambda} \leq C_2, \quad (\text{A5})$$

and from this we have

$$\mathbb{P} \left(\Omega_{B^*} \cap \left\{ \omega : \sup_{t \in [0, \tau_1 \wedge T_1^{(N)} - \epsilon]} |x^\infty(t) - X_t^{(N)}(\omega)| > \epsilon \right\} \right) \leq C_1(\epsilon) \exp(-C_2(\epsilon)N^{1-p(K+2)}),$$

and this combined with step 1 above gives, for any $\epsilon > 0$,

$$\begin{aligned} \mathbb{P} \left(\sup_{t \in [0, \tau_1 \wedge T_1^{(N)} - \epsilon]} |x^\infty(t) - X_t^{(N)}(\omega)| > \epsilon \right) \\ \leq C_1(\epsilon) \exp(-C_2(\epsilon)N^{1-p(K+2)}) + C_3(\epsilon)N \exp(-C_4(\epsilon)N^p). \end{aligned}$$

One error term is improved by increasing p (if we take B^* larger then we are discarding fewer realizations in step 1) and the other is worsened (if we take B^* larger then there are more ways to have a fluctuation away from the mean). The best (asymptotic) estimate is obtained if we choose p to make the powers equal in the two terms, which is if we choose $p = \psi(K) := (K + 3)^{-1}$. Then we have

$$\mathbb{P} \left(\sup_{t \in [0, \tau_1 \wedge T_1^{(N)} - \epsilon]} |x^\infty(t) - X_t^{(N)}(\omega)| > \epsilon \right) \leq C_4 \exp(-C_5 N^{\psi(K)}).$$

This establishes(4.11).

4. We have already shown half of (4.10), namely that $T_1^{(N)} > \tau_1 - \epsilon$ with exponentially large probability. We now need to show that once $t > \tau_1 - \epsilon$, then a big burst occurs soon, and its size is close to $t^*(\xi(\tau_1))$. Two things can occur here: we have a slightly early big burst with $\gamma \leq 1$, or we have no big burst until $\gamma > 1$.

If the latter, it is straightforward, using Lemma 3: once $\gamma > 1$, there is a nonzero $O(1)$ probability of a big burst every time something changes in $X_t^{(N)}$. Moreover, in any $O(1)$ time, the process $X_t^{(N)}$ will undergo $O(N)$ changes and the probability of having a big burst is $O(1)$, so that it is exponentially unlikely that a big burst not take place in this time interval.

If we have a burst while γ is still less than or equal to one, we can use the alternate statement at the bottom of Lemma 3. Let us consider the case where the mean-field has a big burst with $\xi_{K-2}\beta > 1$, $\xi_{K-1}\beta = 1$ (the cases where $\xi_{K-2}\beta = 1$, but $\xi_{K-3}\beta > 1$, etc. are similar). By definition $\chi_{\beta,K}(\xi, 0)$ has zero first derivative, but positive second derivative, at zero. Let us now assume that the big burst happens early and that we have a burst larger than αN but not close to $t^*(\xi(\tau_1))N$. Choose $\epsilon > 0$, and by (4.11) we can guarantee that this happens with $\gamma > 1 - \epsilon$. Let $\tilde{\xi}$ be a vector obtained by decreasing ξ_{K-1} by $O(\epsilon)$ and changing all of the other coefficients of ξ by no more than ϵ . Then $\chi(\tilde{\xi}, t)$ becomes negative for small t , but it has a root in a $O(\epsilon)$ neighborhood of the origin (consider its Taylor series), and thus becomes positive after its first root, which is $O(\epsilon)$. By choosing ϵ small enough, we can guarantee that this root is less than α , and by the last statement of Lemma 3, if the burst is larger than αN then it is near $\tilde{t}_\alpha(\tilde{\xi})N$ with probability exponentially close to one. Moreover, notice that $\tilde{t}_\alpha(\tilde{\xi})$ will be within $O(\epsilon)$ of $t^*(\tilde{\xi})$, since $\chi_{\beta,K}(\xi, t)$ and $\chi_{\beta,K}(\tilde{\xi}, t)$ are uniformly within $O(\epsilon)$ of each other. Therefore, even if the big burst occurs early, it will be within $O(\epsilon)$ of $t^*(\tau_1)$. This establishes (4.10) and (4.12).

5. Finally, we need to show is that if we consider the generator of the process where we cap the bursts at size B^* , then the mean-field ODE is the correct one, and moreover that (A5) holds. In fact, what we will show is even simpler: we can compute x^∞ formally even without the restriction of making B^* finite.

To show these two, let us first consider a simpler process, namely the process where we have no bursting whatsoever (i.e. setting $p = 0$). It is easy to see that this can be written as a continuous-time Markov chain (CTMC) with generator

$$Lf(x) = \sum_{j=0}^J \lambda_j(x) N (f(x + N^{-1}v_j) - f(x)), \quad (\text{A6})$$

where $\lambda_j(x) = x_k$, $v_k = e^{(k+1)} - e^{(k)}$ where $e^{(k)}$ is the k th standard basis function. The reason why this is true is that neurons are being promoted as a Poisson process, and we choose a neuron in state S_k with probability x_k , and if this neuron is chosen we increment S_{k+1} and decrement S_k . Applying (A4) gives

$$\frac{d}{dt}x_k^\infty = x_{k-1}^\infty - x_k^\infty,$$

which is the same as (3.5) with $\gamma = 0$, which is what we expect.

Now, let us consider the network with $p > 0$. Let us say that the network undergoes a burst of size B . The effect of this on all neurons not in state S_{K-1} is to present each of them with B chances to be promoted, each with probability $p = \beta/N$. The effect of this, on average, will be to additionally promote, on average, $x_i\beta B$ neurons from state S_i to state S_{i+1} . To be more precise, if we have $x_i N$ neurons in state S_i , and we have B neurons fire in a given burst, then we will promote k neurons out of state S_i with probability

$$b(k; BNx_i, \beta/N) + O(N^{-1}) = (\beta B x_i)^k \exp(-\beta B x_i) + O(N^{-1}).$$

For any burst, let us define the random vector $\mathbf{B} = (B_0, B_1, \dots, B_{K-1})$ as the number of neurons promoted from state S_i to state S_{i+1} . The effect of such a change is to add the vector

$$\sum_{k=0}^{K-1} B_k v_k$$

to the state of the system. Now, consider any vector $\mathbf{b} \in \mathbb{Z}^K$ and define

$$p(\mathbf{b}) = \mathbb{P}(\mathbf{B} = \mathbf{b}).$$

Our stochastic process is literally defined as follows: at any given time, $K - 1$ “simple” events can occur; for $k = 0, \dots, K - 2$, we can add the vector v_k to the system, and do so with rate $x_k N$, or we can have a burst, which occurs at rate $x_{K-1} N$, at which time we add the vector $\sum_{k=0}^{K-1} b_k v_k$ with probability $p(\mathbf{b})$. However, since all rates are constant, it is equivalent to instead consider the process where we add the vector v_k with rate $x_k N$ for $k = 0, \dots, K - 2$ and add the vector $\sum b_k v_k$ with rate $x_{K-1} N p(\mathbf{b})$.

From this, we have that

$$\begin{aligned} \frac{d}{dt} x^\infty &= \sum_{k=0}^{K-2} x_k^\infty v_k + \sum_{\mathbf{b} \in \mathbb{Z}^K} x_{K-1}^\infty p(\mathbf{b}) \sum_{k=0}^{K-1} b_k v_k \\ &= \sum_{k=0}^{K-2} \left(x_k^\infty + \sum_{\mathbf{b} \in \mathbb{Z}^K} x_{K-1}^\infty p(\mathbf{b}) b_k \right) v_k + x_{K-1}^\infty \sum_{\mathbf{b} \in \mathbb{Z}^K} p(\mathbf{b}) b_{K-1} v_{K-1}. \end{aligned}$$

The sum over \mathbf{b} in the last term is the expected size of a burst, namely, $\mu(\gamma)$. The sum over \mathbf{b} in the first term is the expected number of neurons promoted from state S_k during a burst. We compute:

$$\begin{aligned} \mathbb{E}(B_k) &= \sum_{\mathbf{b}} \mathbb{P}(B_k = b_k) b_k = \sum_{b_k} \sum_q b_k \mathbb{P}(B_k = b_k | B_{K-1} = q) \mathbb{P}(B_{K-1} = q) \\ &= \sum_q \mathbb{P}(B_{K-1} = q) \sum_{b_k} b_k \frac{(x_k \beta q)^{b_k}}{b_k!} \exp(-x_k \beta q) \\ &= \sum_q \mathbb{P}(B_{K-1} = q) x_k \beta q = x_k \beta \sum_q \mathbb{P}(B_{K-1} = q) q = x_k \beta \mu(\gamma). \end{aligned}$$

So we have

$$\frac{d}{dt}x^\infty = \sum_{k=0}^{K-2} x_k^\infty (1 + \gamma\mu(\gamma))v_k + x_{K-1}^\infty \mu(\gamma)v_{K-1}.$$

As we proved in Lemma 2, $\mu(\gamma) = 1 + \gamma\mu(\gamma)$, and this gives (3.5). Finally, to establish (A5): every rate is bounded above by β , with a cap on burst size of B^* , the total number of changes which can occur are bounded above by $(B^*)^K + K$ (the number of $\mathbf{b} \in \mathbb{Z}^K$ with $b_i \leq B^*$, plus the $K - 1$ non-burst steps, and finally $|v_j| \leq KB^*$ (each entry is at most B^* and there are K coordinates)).

□

Remark. We can see from the proof (specifically, part 4) why the critical, or near-critical, graph does not affect the mean-field. One might think that as $\gamma \rightarrow 1$ — as the graph approaches criticality — something untoward might happen in the limit. However, any burst of size $o(N)$ does not appear in the limit, and, moreover, the conditional estimate of Lemma 3 shows that even if we do have a (rare) event of $O(N)$ when the graph is near-critical, this event will be very close to the expected super-critical event which would have occurred soon anyway. Basically, when the process is near-critical, either nothing happens, or if it does, it is close to what is expected, and in any case, the process is far away from criticality after the event. Moreover, the system only spends a finite time near criticality, so if no big bursts occur for a long enough time, the process moves through criticality to supercriticality. It will of course be interesting to understand this critical behavior for several reasons, one of which is that it surely plays a role in understanding the most likely path of switching in the case where we have multiple attractors.

All that remains is to show:

Lemma 8.

$$\sum_{|\mathbf{q}|=B-1} C_{\mathbf{q}} = \frac{B^{B-2}}{(B-1)!}.$$

Proof. We will use an induction argument. Now, let us consider $\mathbf{q} = (q_1, \dots, q_k)$ and $\tilde{\mathbf{q}} = (\gamma, q_1, \dots, q_k)$ with $\gamma \in \mathbb{N}$. Then

$$C_{\mathbf{q}} = \frac{q_1^{q_2} \dots r_{q-1}^{q_k}}{q_1! \dots q_k!}, \quad C_{\tilde{\mathbf{q}}} = \frac{\gamma^{q_1} q_1^{q_2} \dots q_{k-1}^{q_k}}{\gamma! q_1! \dots q_k!}.$$

In short, prepending an integer to a vector effectively multiplies the associated constant by $\gamma_1^q / \gamma!$, where q_1 is the starting constant of the vector to which we are prepending. Define

$$C_{n,k} = \sum_{|\mathbf{q}|=n, q_1=k} C_{\mathbf{q}}.$$

We first want to compute $C_{n,1}$. Now, take any partition \mathbf{q} with $|\mathbf{q}| = n$ and $q_1 = 1$. Then, after removing the leading 1, we have a partition of $n - 1$ with leading digit q_2 . Then

$$C_{(q_1, \dots, q_n)} = 1^{q_2} C_{(q_2, \dots, q_n)} = C_{(q_2, \dots, q_n)}.$$

Thus it is easy to see that

$$C_{n,1} = \sum_{j=1}^{n-1} C_{n-1,j}.$$

Now, consider $k = 2$, so we have $q_1 = 2$. After removing the leading two, we have a partition of $n - 2$ with leading digit q_2 . Then we have

$$C_{(q_1, \dots, q_n)} = \frac{2^{q_2}}{2!} C_{(q_2, \dots, q_n)}.$$

From this we have

$$C_{n,2} = \sum_{j=1}^{n-2} \frac{2^j}{2} C_{n-2,j}.$$

In general, by the same argument we have

$$C_{n,k} = \sum_{j=1}^{n-k} \frac{k^j}{k!} C_{n-k,j}.$$

We now assert that for $k \leq n$,

$$C_{n,k} = \frac{n^{n-k-1}}{(k-1)!(n-k)!}. \quad (\text{A7})$$

First check that $C_{1,1} = 1$. Now assume that the claim holds for all $n \leq n^*$. Then we compute

$$\begin{aligned} C_{n^*,k} &= \sum_{j=1}^{n^*-k} \frac{k^j}{k!} C_{n^*-k,j} = \sum_{j=1}^{n^*-k} \frac{k^j}{k!} \frac{(n^* - k)^{n^*-k-j-1}}{(j-1)!(n^* - k - j)!} \\ &= \frac{1}{k!} \sum_{l=0}^{n^*-k-1} \frac{k^{l+1} (n^* - k)^{n^*-k-l-2}}{l!(n^* - k - l - 1)!} = \frac{k}{(n^* - k)k!} \sum_{l=0}^{n^*-k-1} \frac{k^l (n^* - k)^{n^*-k-l-1}}{l!(n^* - k - l - 1)!} \\ &= \frac{1}{(k-1)!(n^* - k)!} \sum_{l=0}^{n^*-k-1} \binom{n^* - k - 1}{l} k^l (n^* - k)^{n^*-k-l-1} \\ &= \frac{1}{(k-1)!(n^* - k)!} (n^* - k + k)^{n^*-k-1}. \end{aligned}$$

This establishes (A7). Now, finally notice that

$$\sum_{|\mathbf{q}|=B-1} C_{\mathbf{q}} = C_{B,1} = \sum_{j=1}^{B-1} C_{B-1,j}.$$

Expanding, we obtain

$$\begin{aligned}
B^{B-2} &= \sum_{j=0}^{B-2} \frac{(B-2)!}{j!(B-j-2)!} (B-1)^{B-j-2} \\
&= \sum_{l=1}^{B-1} \frac{(B-2)!}{(l-1)!(B-l-1)!} (B-1)^{B-l-1} \\
&= (B-1)! \sum_{j=1}^{B-1} C_{B-1,j},
\end{aligned}$$

and we are done. \square

B Proof of Theorem 5

We first give a quick recap of the mean-field process and establish many of the calculational shortcuts we will use later in the proof.

B1. Preliminary remarks

Here we reconsider both the flow ϕ and the map $G_{\beta,K}$ and give a simplifying description for the dynamics of each, and then we show the Gaussian approximations we use in the proof.

Simplified description of ϕ . Consider (3.5). Up to a change in time variable, $\phi(x, t) = e^{tL_P}x$, where L_P is the $K \times K$ matrix with

$$(L_P)_{ij} = \begin{cases} -1, & i = j, \\ 1, & i \equiv j + 1 \pmod{K}, \\ 0 & \text{else.} \end{cases}$$

Since the trajectories of an ODE are not affected by a time change, the question of whether a trajectory enters a set or not is independent of this change. Because of this, we will simply replace $\phi(x, t)$ with e^{tL_P} throughout this proof without further comment.

From the definition of Poisson RVs, if we consider the initial condition $x(0) = e^{(0)}$, then we have

$$x_i(t) = \mathbb{P}(Po(t) \equiv i \pmod{K}).$$

Moreover, by linearity, if we choose $x \in \mathbb{R}^K$, $x = (x_0, \dots, x_{K-1})$, then

$$x(t) = \sum_{i=0}^{K-1} x_i x^{(i)}(t),$$

where $x^{(i)}(t)$ is the solution to $x' = L_P x$ with initial condition $e^{(i)}$. Of course, we also have

$$x_j^{(i)}(t) = \mathbb{P}(Po(t) \equiv (j-i) \pmod{K}),$$

which then gives

$$x(t)_j = \sum_{i=0}^{K-1} x_i \mathbb{P}(Po(t) \equiv (j-i) \pmod{K}). \quad (\text{B1})$$

Simplified description of $G_{\beta,K}$. The map $G_{\beta,K}$ and $t^*(x)$ are defined above in (3.9, 3.11); however, there exists an alternate description of the map $G_{\beta,K}$ (and t^*) motivated by the queue arguments of Section 3.3., which we now describe.

Consider $x \in D(G_{\beta,K})$. Now consider $y \in \mathbb{R}^{K+1}$ (we now let the index go from $0, \dots, K$, where y_K corresponds to the queue) where we set $y_i = x_i$ for $i = 1, \dots, K-1$ and $y_K = 0$. Let it flow under the linear flow $\dot{y} = \beta L_B y$, with L_B defined in (3.10). $y_K(t)$ will be the (rescaled) number of neurons which have been pushed into the queue by time t , and since neurons are flowing out at rate 1, we continue this process until $y_K(t) < t$. (Note that such a time is guaranteed to occur since $y_K(t) < 1$ for all t .) Call this time t^* , and then define $G_{\beta,K}(x)$ as the vector with

$$G_{\beta,K}(x)_i = \begin{cases} y_i(t^*), & i = 1, \dots, K-1, \\ y_0(t^*) + y_K(t^*), & i = 0. \end{cases}$$

We can go further: if we now replace the flow L_B with the ODE on the set of variables indexed by \mathbb{N} where we define $\dot{z} = \beta L'_B z$ with

$$(L'_B)_{i,j} = \begin{cases} -1, & i = j, \\ 1, & i = j + 1, \\ 0, & \text{else} \end{cases}$$

then it is easy to see that $y_K(t) = \sum_{i>K} z_i(t)$. In other words, we can think of the ‘‘burst flow’’ as a flow on the whole real line where we count how much mass has passed the point K , or we can consider the finite-dimensional flow which absorbs all of the mass at point K . We will use both viewpoints below as convenient. In particular, this allows us to note that the exact solution to the L_B flow gives us, for all $t > 0$,

$$y_K(t_0 + t) = \sum_{i=1}^K y_{K-i}(t_0) \mathbb{P}(Po(\beta t) \geq i), \quad (\text{B2})$$

and

$$y_i(t_0 + t) = \sum_{i=0}^j y_i(t_0) \mathbb{P}(Po(\beta t) = j - i). \quad (\text{B3})$$

Compare (B1) with (B2, B3); these formulas are basically the same, and the only difference is a factor of β in the time variable or how boundary conditions are handled.

Gaussian approximation. We can use the standard Gaussian approximation for the Poisson RV: for any $\alpha > 0$, if we choose $t > \alpha K$, then

$$\mathbb{P}(Po(t) = j) = \frac{1}{\sqrt{2\pi t}} e^{-(t-j)^2/2t} + o(K^{-10}). \quad (\text{B4})$$

(We choose the power 10 for concreteness in what follows; of course, we could use any polynomial decay here.) It also follows that

$$\mathbb{P}(Po(t) \equiv \alpha \pmod{K}) = \frac{1}{\sqrt{2\pi t}} e^{-(j-t)^2/2t} + o(K^{-10}), \quad (\text{B5})$$

where j is the member of the equivalence class $[\alpha]_K$ closest to t . In particular, if this j has the further property that $|t - j| > K^p$ for some $p > 1/2$, then

$$\mathbb{P}(Po(t) \equiv \alpha \pmod{K}) = o(K^{-10}). \quad (\text{B6})$$

These approximations will be of direct use for understanding the L_P flow; of course we need only replace t with βt when approximating the L_B flow.

B2. Proof of Theorem 5

Proof of Theorem 5. We break up the problem into several steps. In Step 1, we estimate $s^*(x)$ for $x \in N_{\theta K, K}$. In Step 2, we estimate $t^*(F_{\beta, K}(x))$ for $x \in N_{\theta K, K}$. Finally, in Step 3, we put it all together and show that $G_{\beta, K}(F_{\beta, K}(x)) \in N_{\theta K, K}$ for any $x \in N_{\theta K, K}$. Throughout the proof, we will use C as a generic constant which is independent of K .

Step 1. Estimating s^* . Here we show that if $x \in N_{\beta, K}$, then there is a $C > 0$ with $s^*(x) = K - C\sqrt{K \log(K)} + o(\sqrt{K \log(K)})$. To establish this, we have to both estimate the time at which $x_{K-1} = \beta^{-1}$, and show that at this time, $x_{K-2} > x_{K-1}$.

Consider any $s < 0.99K$. By the definition of $N_{\beta, K}$, the contribution to $x_{K-1}(s)$ from the modes in R is at most $(10\beta)^{-1}$, and the contribution from the modes in L is at most K^{-10} . However, from (B6), the contribution from $x_0(0)$ can be at most K^{-10} as well. Therefore $x_{K-1}(s)$ is certainly less than β^{-1} for all $x < 0.99K$ and thus $s^*(x) > 0.99K$.

Now consider $s \in [0.99K, 1.01K]$. By the arguments of the last paragraph, the total contribution from L and R can be bounded above by $K^{-10} + (10\beta)^{-1} = O(K^{-1})$. The contribution from $x_0(0)$ is

$$\mathbb{P}(Po(s) = K - 1) + o(K^{-10}).$$

However, it is easy to see that $\mathbb{P}(Po(s) = K - 1)$ achieves its maximum at $s = K - 1$ and this maximum is $\Theta(K^{-1/2})$. Therefore $s^* < K - 1$. In fact, we can get an even better estimate of s^* : use the Gaussian approximation (B5), plug in $s = K - \phi(K)$ and consider large K asymptotics, matching scales shows that we have $x_{K-1} = \beta^{-1}$ at $s^* = K - (1 + o(1))\sqrt{K \log K}$. Thus we can choose and fix $C_1 > 1, C_2 < 1$ and for K sufficiently large,

$$K - C_1\sqrt{K \log(K)} \leq s^* \leq K - C_2\sqrt{K \log(K)}. \quad (\text{B7})$$

We have established the time s^* when $(e^{tL_P} x_0)_{K-1} = \beta^{-1}$, but we also need to show that

$$(e^{tL_P} x_0)_{K-2} > \beta^{-1},$$

or a burst will not occur.

Choose any $p > 1/2$. By (B7, B6), we know that (up to $O(K^{-10})$), the contributions to $x_{K-2}(s^*)$, $x_{K-1}(s^*)$ must have come from levels within K^p of zero, on one side or the other. By the definition of $N_{\theta K, K}$, these all came from R , since there is only $O(K^{-10})$ total mass in L . Thus we know

$$\begin{aligned} x_{K-1}(s^*) &= x_0 \mathbb{P}(Po(s^*) = K-1) + \sum_{i=1}^{K^p} x_{K-i} \mathbb{P}(Po(s^*) = i-1+K) + O(K^{-10}), \\ x_{K-2}(s^*) &= x_0 \mathbb{P}(Po(s^*) = K-2) + \sum_{i=1}^{K^p} x_{K-i} \mathbb{P}(Po(s^*) = i-2+K) + O(K^{-10}). \end{aligned} \quad (\text{B8})$$

We want to show $x_{K-2}(s^*) > x_{K-1}(s^*)$, thus we will need to estimate quotients of the form

$$\frac{\mathbb{P}(Po(s^*) = K+i-1)}{\mathbb{P}(Po(s^*) = K+i-2)} = \frac{s^*}{K+i-1} \geq \frac{K - C\sqrt{K \log(K)}}{K+i-1} \geq 1 - C\sqrt{\frac{\log(K)}{K}},$$

for any $i = o(K)$. But then comparing the equations in (B8) term-by-term gives

$$x_{K-2}(s^*) \geq (1 - CK^{-1/2}(\log(K))^{-1/2})x_{K-1}(s^*) > \beta^{-1},$$

so that a burst does occur at $s = s^*$.

Step 2. Estimating t^* . Define y as above, namely we set $y_k(s^*) = x_k(s^*)$ for $k = 0, 1, \dots, K-1$, $y_K(s^*) = 0$, and let y flow under $y' = \beta L_B y$.

We will show $t^* > 0.99$. We break this up into three intervals:

$$t \in [0, 4/K], [4/K, 2C_1\sqrt{K \log K}/\beta], [2C_1\sqrt{K \log K}/\beta, 0.99],$$

which we denote as Stages 1–3. The constant C_1 is the one which appears in the estimate for s^* in (B7). We will show that t^* cannot lie in any of these three intervals.

Stage 1. $t \in [0, 4/K]$. Here we only consider the last two entries in y :

$$y_K(s^* + t) \geq y_{K-1}(s^*) \mathbb{P}(Po(\beta t) \geq 1) + y_{K-2}(s^*) \mathbb{P}(Po(\beta t) \geq 2).$$

Writing out the condition $y_K(s^* + t) > t$ gives

$$((1 - e^{-t\beta} + (1 - e^{-t\beta} - t\beta e^{-t\beta}))\beta^{-1} - t) > 0,$$

or

$$2(1 - e^{-t\beta}) + t\beta e^{-t\beta} - t\beta > 0.$$

It is easy to check that the function

$$2(1 - e^{-\alpha}) + \alpha e^{-\alpha} - \alpha$$

is positive for all $\alpha \in (0, 2)$, so that as long as $t\beta < 2$, or $t < 4/K$, then these inequalities all hold.

Stage 2. $t \in [4/K, 2C_1\sqrt{K \log(K)}/\beta]$. Pick any $p > 1/2$. From (B5, B6), we know

$$y_{K-1}(s^*) = x_0 \mathbb{P}(Po(s^*) = K-1) + \sum_{i=1}^{K^p} x_{K-i} \mathbb{P}(Po(s^*) = i-1+K) + o(K^{-10}),$$

$$y_{K-1}(s^* + t) = x_0 \mathbb{P}(Po(s^* + t) = K-1) + \sum_{i=1}^{K^p} x_{K-i} \mathbb{P}(Po(s^* + t) = i-1+K) + o(K^{-10}),$$

for all $t \in [4/K, 2C_1\sqrt{K \log(K)}/\beta]$. But now choose $p = 0.6$, and for all $i = 0, \dots, K^p$, we can show that

$$\frac{\mathbb{P}(Po(s^* + t) = i-1+K)}{\mathbb{P}(Po(s^*) = i-1+K)} \geq 1 + \frac{1}{K^2}. \quad (\text{B9})$$

To see this, notice that this quotient is

$$\frac{s^*}{s^* + t} \exp \left[\frac{(K-1-s^*)^2}{2s^*} - \frac{(K-1-(s^*+t))^2}{2(s^*+t)} \right].$$

We simply need to observe that the argument to the exponential is positive for $0 < t < (K-1)^2 - (s^*)^2/s^*$ and scales like K , and (B9) follows. Thus we have

$$y_{K-1}(s^* + t) \geq y_{K-1}(s^*)(1 + K^{-2}) + o(K^{-10}) > \beta^{-1},$$

and $y'_K(s^* + t) > 1$ for all t in Stage 2. Therefore $y_K(s^* + t) - t > 0$ for all t in Stage 2.

Stage 3. $t \in [2C_1\sqrt{K \log(K)}/\beta, 0.99K]$. Consider $y_K(s^* + t)$ at $t = 2C_1\sqrt{K \log(K)}/\beta$. We can get the lower bound

$$y_K(s^* + t) \geq x_0(0) (\mathbb{P}(Po(s^* + \beta t) \geq K) - \mathbb{P}(Po(s^*) \geq K)) + O(K^{-1})$$

from the following observation. Imagine that we had the flow on the real line generated by L'_B , as described above. Then $z_k(t)$ would simply be $\mathbb{P}(Po(t) = k)$. However, the way in which we modify this flow is that until time s^* , we wrap the flow around periodically, and after time s^* , we absorb everything into level K . This guarantees that there is at least as much at level K for our system as would have passed level K between the times of s^* and $s^* + t$. But from (B7),

$$\mathbb{P}(Po(s^*) \geq K) \leq o(K^{-1}).$$

We also have

$$\mathbb{P}(Po(s^* + \beta t) \geq K) = \frac{1}{\sqrt{s^* + \beta t}} \int_K^\infty e^{-(x-(s^*+t))^2/(2(s^*+t))} dx + O(K^{-10}).$$

This is clearly an increasing function of t . In Stage 3, we have

$$s^* + t \geq K + C_1\sqrt{K \log(K)},$$

so we replace $s^* + t$ with its smallest value and change variables to obtain

$$\mathbb{P}(Po(s^* + \beta t) \geq K) \geq \frac{1}{\sqrt{2\pi(K + \sqrt{K \log(K)}/2)}} \int_0^\infty e^{-(y - \sqrt{K \log(K)}/2)^2 / (K + \sqrt{K \log(K)}/2)} dy.$$

This integral is in the form

$$\frac{1}{\sqrt{2\pi\sigma^2(K)}} \int_0^\infty e^{-(y - \mu(K))^2 / 2\sigma^2(K)} dy = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{\mu(K)}{\sqrt{2}\sigma(K)} \right) \right).$$

We have $\mu(K)/\sigma(K) \rightarrow \infty$ (notice that this is one place where the $\sqrt{\log(K)}$ term is crucial, without it we would have $\mu/\sigma = O(1)!$), and using the standard asymptotic expansion (see, for example, [2, Chapter 7]) for $\operatorname{erf}(x)$ for large x gives

$$\mathbb{P}(Po(s^* + \beta t) \geq K) \geq x_0(0)(1 + 3K^{-1}) \geq 1 - (\delta + 3K^{-1}) + o(K^{-1}),$$

and by the definition of δ we can make this as close to one as we like, and in particular we can make it larger than 0.99. But then this makes $y_K(s^* + t) - t > 0$ for $t < 0.99$, and therefore $t^* > 0.99$.

Step 3. $G_{\beta,K}(F_{\beta,K}(x))$ lies in $N_{\theta K,K}$. We denote $z = G_{\beta,K}(F_{\beta,K}(x))$. Now, for any $j \neq 0$, we have

$$z_j = \sum_{i=0}^j y_i(s^*) \mathbb{P}(Po(\beta t^*) = j - i).$$

Consider $j \in L$, so that the sum is only over $i \in L$, and, in particular, $j - i \leq K/2$. But since $t^* > 0.99$, $\beta t^* = O(K)$ and this probability is exponentially small as $K \rightarrow \infty$. So we have $\sum_{j \in L} z_j < K^{-10}$, which satisfies the first criterion of $N_{\theta K,K}$.

To get an estimate on z_j for $j \in R$, we need to control how much of the mass “wrapped around” between $s = 0$ and $s = s^*$. So, let us define the set LE (the “left edge”) as $LE = \{0, 1, \dots, \lfloor K/100 \rfloor\}$ (notice this time we do include zero!). From our estimate on s^* above, we know that the entire mass which contributes to $\sum_{i \in LE} y_i$ had to come from modes within $2C_1 \sqrt{K \log(K)}$ of LE . By the assumption on $N_{\beta,K}$, we know that the terms in the left did not contribute, so the contribution had to have been from the x_0 and the modes within $2C_1 \sqrt{K \log(K)}$ of the right, so we have

$$\sum_{i \in LE} y_i(s^*) \leq x_0 \mathbb{P}(Po(s^*) \geq K) + \sum_{j=1}^{2C_1 \sqrt{K \log(K)}} x_{K-j} \mathbb{P}(Po(s^*) \geq K + j).$$

On the other hand, we can get the upper bound

$$y_{K-1}(s^*) \geq x_0 \mathbb{P}(Po(s^*) = K - 1) + \sum_{j=1}^{2C_1 \sqrt{K \log(K)}} x_{K-j} \mathbb{P}(Po(s^*) = j - 1 + K) + O(K^{-10}),$$

but, by definition, $y_{K-1}(s^*) = \beta^{-1}$. To compare these two, we will need to estimate quotients of the form

$$\frac{\mathbb{P}(Po(s^*) \geq K + j)}{\mathbb{P}(Po(s^*) = K + j - 1)}$$

Use the Gaussian approximation and see that the quotient becomes

$$\begin{aligned} & \int_{K+j}^{\infty} \exp\left(\frac{-(x - s^*)^2 + (K + j - 1 - s^*)^2}{2s^*}\right) dx \\ &= e^{(K+j-1-s^*)/2s^*} \sqrt{\frac{\pi s^*}{2}} \operatorname{erfc}\left(\frac{K + j - 1 - s^*}{2s^*}\right) \end{aligned}$$

Using the standard asymptotics for $\operatorname{erfc}(x)$ gives

$$\frac{\mathbb{P}(Po(s^*) \geq K + j)}{\mathbb{P}(Po(s^*) = K + j - 1)} \leq C \frac{s^*}{K + j - 1 - s^*} \leq C \sqrt{\frac{K}{\log(K)}}.$$

The last inequality is true for all $j \geq 0$, and in fact gets better as j gets larger. Thus we have

$$\sum_{i \in LE} y_i(s^*) \leq C \sqrt{\frac{K}{\log(K)}} y_{K-1}(s^*) + O(K^{-10}) \leq \frac{C}{\sqrt{K \log(K)}}$$

(recall that $\beta > \theta K$). This gives an upper bound on all of the mass contained in LE , and since we know the rest of L can only contain $O(K^{-10})$, this means the bulk of the mass is contained in R .

However, note that, up to $O(K^{-10})$, all of the mass within $\theta K + 2\sqrt{K \log(K)}$ of K will be absorbed into level K during the burst — in particular, this includes all mass contained in R at the beginning of the burst. Therefore, the total mass which can escape moving to y_K during the burst is contained in LE at the start of the burst, thus giving

$$\sum_{j \in R} z_j \leq C(K \log(K))^{-1/2}, \tag{B10}$$

which satisfies the second condition on $N_{\theta K, K}$.

Finally, we need to show that $P_R z \in O_{\theta K, K}$, that is, that $\phi(P_R z, t)$ never gets more than $(10\beta)^{-1}$ in the $(K - 1)$ st level. Note that the bound in (B10) is not good enough, since $\beta = O(K)$. We've already noted that all of the mass in $P_R z$ had to have come from $y_i(s^*)$ with $i \in LE$, and in fact this is an upper bound (some of the mass will have gotten to K during the burst) up to $O(K^{-10})$. Therefore, if we consider the flow e^{tL_P} with the initial condition $P_{LE} y(s^*)$, this (again up to $O(K^{-10})$) gives an upper bound on $P_R z$. However, notice that if we choose $t^+ > 0.99$, and any i , then

$$\mathbb{P}(Po(\beta t^+) \equiv i \pmod{k}) \leq CK^{-1/2} \tag{B11}$$

since $\beta = O(K)$. Therefore every coordinate of $e^{t^+ L_P} P_{LE}(y(s^*))$ is bounded above by (B11) plus the total mass in $P_{LE}(y(s^*))$, and from (B10) this means that every coordinate of $e^{t^+ L_P} P_{LE}(y(s^*))$ is bounded above by $CK^{-1}(\log(K))^{-1/2}$. This establishes the third condition for $N_{\theta K, K}$ (notice that yet again the $\sqrt{\log(K)}$ factor plays a critical role!) and we are done. \square

References

- [1] L. F. Abbott, C. van Vreeswijk. *Asynchronous states in networks of pulse-coupled oscillators*. Phys. Rev. E, 48 (1993), No. 2, 1483–1490.
- [2] M. Abramowitz, I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, National Bureau of Standards Applied Mathematics Series, Vol. 55, Washington, D.C., 1964.
- [3] N. Alon, J. H. Spencer. *The probabilistic method*. Wiley & Sons Inc., Hoboken, NJ, 2008.
- [4] F. Apfaltrer, C. Ly, D. Tranchina. *Population density methods for stochastic neurons with realistic synaptic kinetics: Firing rate dynamics and fast computational methods*. Network-computation in Neural Systems, 17 (2006), No. 4, 373–418.
- [5] M. Bennett, M. F. Schatz, H. Rockwood, K. Wiesenfeld. *Huygens’s clocks*. R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci., 458 (2002), No. 2019, 563–579.
- [6] Béla Bollobás. *Random graphs*. Cambridge Studies in Advanced Mathematics, Vol. 73, Cambridge University Press, Cambridge, 2001.
- [7] P. C. Bressloff, S. Coombes. *Desynchronization, mode locking, and bursting in strongly coupled integrate-and-fire oscillators*. Phys. Rev. Lett., 81 (1998), No. 10, 2168–2171.
- [8] N. Brunel, V. Hakim. *Fast global oscillations in networks of integrate-and-fire neurons with low firing rates*. Neural Comp., 11 (1999), No. 7, 1621–1671.
- [9] J. Buck, E. Buck. *Mechanism of rhythmic synchronous flashing of fireflies*, Science 159 (1968), No. 3821, 1319–1327.
- [10] D. Cai, L. Tao, A. V. Rangan, D. W. McLaughlin. *Kinetic theory for neuronal network dynamics*. Comm. Math. Sci., 4 (2006), No. 1, 97–127.
- [11] D. Cai, L. Tao, M. Shelley, D. W. McLaughlin. *An effective kinetic representation of fluctuation-driven neuronal networks with application to simple and complex cells in visual cortex*. Proc. Nat. Acad. Sci. USA, 101 (2004), No. 20, 7757–7762.
- [12] S. R. Campbell, D. L. L. Wang, C. Jayaprakash. *Synchrony and desynchrony in integrate-and-fire oscillators*. Neur. Comp., 11 (1999), No. 7, 1595–1619.
- [13] J. H. E. Cartwright, V. M. Eguíluz, E. Hernández-García, O. Piro. *Dynamics of elastic excitable media*. Int. J. Bif. Chaos, 9 (1999), No. 11, 2197–2202.
- [14] C. A. Czeisler, E. Weitzman, M. C. Moore-Ede, J. C. Zimmerman, R. S. Knauer. *Human sleep: its duration and organization depend on its circadian phase*. Science 210 (1980), No. 4475, 1264–1267.

- [15] M. de Sousa Vieira. *Chaos and synchronized chaos in an earthquake model*. Phys. Rev. Lett., 82 (1999), No. 1, 201–204.
- [16] R. E. L. DeVille, C. S. Peskin. *Synchrony and asynchrony in a fully stochastic neural network*. Bull. Math. Bio., 70 (2008), No. 6, 1608–1633.
- [17] B. Doiron, J. Rinzel, A. Reyes. *Stochastic synchronization in finite size spiking networks*. Phys. Rev. E (3), 74 (2006), No. 3, 030903, 4 pages.
- [18] P. Erdős, A. Rényi. *On random graphs. I*. Publ. Math. Debrecen, 6 (1959), 290–297.
- [19] P. Erdős, A. Rényi. *On the evolution of random graphs*. Magyar Tud. Akad. Mat. Kutató Int. Közl., 5 (1960), 17–61.
- [20] G. B. Ermentrout, J. Rinzel. *Reflected waves in an inhomogeneous excitable medium*. SIAM J. Appl. Math., 56 (1996), No. 4, 1107–1128.
- [21] W. Gerstner, J. L. van Hemmen. *Coherence and incoherence in a globally-coupled ensemble of pulse-emitting units*. Phys. Rev. Lett., 71 (1993), No. 3, 312–315.
- [22] L. Glass, A. L. Goldberger, M. Courtemanche, A. Shrier. *Nonlinear dynamics, chaos and complex cardiac arrhythmias*. Proc. Roy. Soc. London Ser. A, 413 (1987), No. 1844, 9–26.
- [23] M. R. Guevara, L. Glass. *Phase locking, period doubling bifurcations and chaos in a mathematical model of a periodically driven oscillator: A theory for the entrainment of biological oscillators and the generation of cardiac dysrhythmias*. J. Math. Bio. 14 (1982), No. 1, 1–23.
- [24] D. Hansel, H. Sompolinsky. *Synchronization and computation in a chaotic neural network*. Phys. Rev. Lett., 68 (1992), No. 5, 718–721.
- [25] E. Haskell, D. Q. Nykamp, 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 (2001), No. 2, 141–174.
- [26] C. Huygens. *Horologioium oscilatorium*. Parisiis, Paris, 1673.
- [27] R. Kapral, K. Showalter (eds.). *Chemical waves and patterns*. Springer, 1994.
- [28] B. W. Knight. *Dynamics of encoding in a population of neurons*. J. Gen. Phys., 59 (1972), No. 6, 734–766.
- [29] Y. Kuramoto. *Chemical oscillations, waves, and turbulence*. Springer Series in Synergetics, Vol. 19, Springer-Verlag, Berlin, 1984.
- [30] Y. Kuramoto. *Collective synchronization of pulse-coupled oscillators and excitable units*. Phys. D, 50 (1991), No. 1, 15–30.

- [31] T. G. Kurtz. *Relationship between stochastic and deterministic models for chemical reactions*. J. Chem. Phys., 57 (1972), No. 7, 2976–2978.
- [32] T. G. Kurtz. *Strong approximation theorems for density dependent Markov chains*. Stoch. Proc. Appl., 6 (1977/78), No. 3, 223–240.
- [33] Z.-H. Liu, P.M. Hui. *Collective signaling behavior in a networked-oscillator model*. Phys. A, 383 (2007), No. 2, 714 – 724.
- [34] R. E. Mirollo, S. H. Strogatz. *Synchronization of pulse-coupled biological oscillators*. SIAM J. Appl. Math., 50 (1990), No. 6, 1645–1662.
- [35] Z. Olami, H. J. S. Feder, K. Christensen. *Self-organized criticality in a continuous, non-conservative cellular automaton modeling earthquakes*. Phys. Rev. Lett., 68 (1992), No. 8, 1244–1247.
- [36] K. Pakdaman, D. Mestivier. *Noise induced synchronization in a neuronal oscillator*. Phys. D, 192 (2004), No. 1-2, 123–137.
- [37] C. S. Peskin. *Mathematical aspects of heart physiology*. Courant Institute, New York University, New York, 1975.
- [38] A. Pikovsky, M. Rosenblum, J. Kurths. *Synchronization: A universal concept in nonlinear sciences*. Cambridge University Press, 2003.
- [39] W. Senn, R. Urbanczik. *Similar nonleaky integrate-and-fire neurons with instantaneous couplings always synchronize*. SIAM J. Appl. Math., 61 (2000/01), No. 4, 1143–1155 (electronic).
- [40] A. Shwartz, A. Weiss. *Large deviations for performance analysis*. Chapman & Hall, London, 1995.
- [41] L. Sirovich. *Dynamics of neuronal populations: eigenfunction theory; some solvable cases*. Network-computation in Neural Systems, 14 (2003), No. 2, 249–272.
- [42] L. Sirovich, A. Omrta, B. W. Knight. *Dynamics of neuronal populations: The equilibrium solution*. SIAM J. Appl. Math., 60 (2000), No. 6, 2009–2028.
- [43] S. Strogatz. *Sync: The emerging science of spontaneous order*. Hyperion, 2003.
- [44] C. Sulem, P.-L. Sulem. *The nonlinear Schrödinger equation*. Applied Mathematical Sciences, Vol. 139, Springer-Verlag, New York, 1999.
- [45] R. Temam, A. Miranville. *Mathematical modeling in continuum mechanics*. Cambridge University Press, Cambridge, 2005.
- [46] D. Terman, N. Kopell, A. Bose. *Dynamics of two mutually coupled slow inhibitory neurons*. Phys. D, 117 (1998), No. 1-4, 241–275.

- [47] M. Tsodyks, I. Mitkov, H. Sompolinsky. *Pattern of synchrony in inhomogeneous networks of oscillators with pulse interactions*. Phys. Rev. Lett., 71 (1993), No. 8, 1280–1283.
- [48] J. J. Tyson, C. I. Hong, C. D. Thron, B. Novak. *A Simple Model of Circadian Rhythms Based on Dimerization and Proteolysis of PER and TIM*. Biophys. J., 77 (1999), No. 5, 2411–2417.
- [49] J. J. Tyson, J. P. Keener. *Singular perturbation theory of traveling waves in excitable media (a review)*. Phys. D, 32 (1988), No. 3, 327–361.
- [50] C. van Vreeswijk, L. Abbott, G. Ermentrout. *When inhibition not excitation synchronizes neural firing*. J. Comp. Neurosci., 1 (1994), No. 4, 313–322.
- [51] C. van Vreeswijk, H. Sompolinsky. *Chaotic balance state in a model of cortical circuits*. Neur. Comp., 10 (1998), No. 6, 1321–1372.
- [52] A. T. Winfree. *The geometry of biological time*. Interdisciplinary Applied Mathematics, Vol. 12, Springer-Verlag, New York, 2001.