

Infinite Systems of Interacting Chains with Memory of Variable Length—A Stochastic Model for Biological Neural Nets

A. Galves · E. Löcherbach

Received: 21 December 2012 / Accepted: 8 March 2013 / Published online: 21 March 2013
© Springer Science+Business Media New York 2013

Abstract We consider a new class of non Markovian processes with a countable number of interacting components. At each time unit, each component can take two values, indicating if it has a spike or not at this precise moment. The system evolves as follows. For each component, the probability of having a spike at the next time unit depends on the entire time evolution of the system after the last spike time of the component. This class of systems extends in a non trivial way both the interacting particle systems, which are Markovian (Spitzer in *Adv. Math.* 5:246–290, 1970) and the stochastic chains with memory of variable length which have finite state space (Rissanen in *IEEE Trans. Inf. Theory* 29(5):656–664, 1983). These features make it suitable to describe the time evolution of biological neural systems. We construct a stationary version of the process by using a probabilistic tool which is a Kalikow-type decomposition either in random environment or in space-time. This construction implies uniqueness of the stationary process. Finally we consider the case where the interactions between components are given by a critical directed Erdős-Rényi-type random graph with a large but finite number of components. In this framework we obtain an explicit upper-bound for the correlation between successive inter-spike intervals which is compatible with previous empirical findings.

Keywords Biological neural nets · Interacting particle systems · Chains of infinite memory · Chains of variable length memory · Hawkes process · Kalikow-decomposition

Dedicated to Errico Presutti, frateddu e mastru.

A. Galves

Instituto de Matemática e Estatística, Universidade de Sao Paulo, Caixa Postal 66281, 05315-970 Sao Paulo, Brazil

E. Löcherbach (✉)

AGM, CNRS-UMR 8088, Université de Cergy-Pontoise, 95000 Cergy-Pontoise, France
e-mail: eva.loecherbach@u-cergy.fr

1 Introduction

A biological neural system has the following characteristics. It is a system with a huge (about 10^{11}) number of interacting components, the neurons. This system evolves in time, and its time evolution is not described by a Markov process [6]. In particular, the times between successive spikes of a single neuron are not exponentially distributed (see, for instance, [5]).

This is the motivation for the introduction of the class of models that we consider in the present paper. To cope with the problem of the large number of components it seems natural to consider infinite systems with a countable number of components. In this new class of stochastic systems, each component depends on a variable length portion of the history. Namely, the spiking probability of a given neuron depends on the accumulated activity of the system after its last spike time. This implies that the system is not Markovian. The time evolution of each single neuron looks like a stochastic chain with memory of variable length, even if the influence from the past is actually of infinite order. This class of systems represents a non trivial extension of the class of interacting particle systems introduced in 1970 by Spitzer [25]. It is also a non trivial extension of the class of stochastic chains with memory of variable length introduced in 1983 by Rissanen [24].

The particular type of dependence from the past considered here is motivated both by empirical as well as theoretical considerations.

From a theoretical point of view, Cessac [6] suggested the same kind of dependence from the past. In the framework of leaky integrate and fire models, he considers a system with a finite number of membrane potential processes. The image of this process in which only the spike times are recorded is a stochastic chain of infinite order where each neuron has to look back into the past until its last spike time. Cessac's process is a finite dimensional version of the model considered here.

Finite systems of point processes in discrete or continuous time aiming to describe biological neural systems have a long history whose starting points are probably Hawkes [19] from a probabilistic point of view and Brillinger [5] from a statistical point of view, see also the interesting paper by Krumin et al. [21] for a review of the statistical aspects. For non-linear Hawkes processes, but in the frame of a finite number of components, Brémaud and Massoulié [3] address the problem of existence, uniqueness and stability. Møller and coauthors propose a perfect simulation algorithm in the linear case, see [22]. In spite of the great interest in Hawkes processes during the last years, especially in association with modeling problems in finance and biology, all the studies are reduced to the case of systems with a finite number of components. Here we propose a new approach which enables us to deal also with infinite systems with a countable number of components, without any assumption of the type linearity or attractiveness.

This paper is organized as follows. In Sect. 2 we state two theorems proving the existence and uniqueness of infinite systems of interacting chains with memory of variable length, under suitable conditions. Our main technical tool is a Kalikow-type decomposition of the infinite order transition probabilities which is a non trivial extension of previous results of the authors in the case of Markovian systems, cf. [15]. The decomposition considered here has two major differences with respect to what has been done before. Firstly this is due to the non-Markovian nature of the system. Secondly, and most importantly, the structure of the transition laws leads to the need of either considering a decomposition depending on a random environment or considering a space-time decomposition. Using the Kalikow-type decomposition we prove the existence, the uniqueness as well as a property of loss of memory of the stationary process.

In Sect. 3 we study the correlation between successive inter-spike intervals (ISI). This aims at explaining empirical results presented in the neuroscientific literature. Gerstner and

Kistler [16], quoting Goldberg et al. [17], observe that in many experimental setups the empirical correlation between successive inter-spike intervals is very small “indicating that a description of spiking as a stationary renewal process is a good approximation”. However, Nawrot et al. [23] find statistical evidence that neighboring inter-spike intervals are correlated, having negative correlation. We show that we can account for these apparently contradictory facts within our model. This requires the introduction of a new setup in which the synaptic weights define a critical directed Erdős-Rényi random graph with a large but finite number of components. We obtain in Theorem 3 an explicit upper bound for the correlations involving the number of components of the system, as well as the typical length of one inter-spike interval. For a system having a large number of components, our result is compatible with the discussion in [16]. Gerstner and Kistler [16] deduce from this that spiking can be described by a renewal process. However, for systems with a small number of components, the correlation might as well be quite big, as reported by Nawrot et al. (2007) who show that neighboring inter-spike intervals are negatively correlated. Therefore, both features are captured by our model, depending on the scale we are working in.

The proofs of all the results are presented in Sects. 4, 5 and 6.

2 Systems of Interacting Chains with Memory of Variable Length: Existence, Uniqueness and Loss of Memory

We consider a stochastic chain $(X_t)_{t \in \mathbb{Z}}$ taking values in $\{0, 1\}^I$ for some countable set of neurons I , defined on a suitable probability space (Ω, \mathcal{A}, P) . For each neuron i at each time $t \in \mathbb{Z}$, $X_t(i) = 1$ if neuron i has a spike at that time t , and $X_t(i) = 0$ otherwise. The global configuration of neurons at time t is denoted $X_t = (X_t(i), i \in I)$. We define the filtration

$$\mathcal{F}_t = \sigma(X_s, s \in \mathbb{Z}, s \leq t), \quad t \in \mathbb{Z}.$$

For each neuron $i \in I$ and each time $t \in \mathbb{Z}$ let

$$L_t^i = \sup\{s < t : X_s(i) = 1\} \tag{2.1}$$

be the last spike time of neuron i strictly before time t . We introduce a family of “synaptic” weights $W_{j \rightarrow i} \in \mathbb{R}$, for $j \neq i$, $W_{j \rightarrow j} = 0$ for all j . $W_{j \rightarrow i}$ is the “synaptic weight of neuron j on neuron i ”. We suppose that the synaptic weights have the following property of uniform summability

$$\sup_{i \in I} \sum_j |W_{j \rightarrow i}| < \infty. \tag{2.2}$$

Now we are ready to introduce the dynamics of our process. At each time t , conditionally on the whole past, sites update independently. This means that for any finite subset $J \subset I$, $a_i \in \{0, 1\}$, $i \in J$, we have

$$P(X_t(i) = a_i, i \in J | \mathcal{F}_{t-1}) = \prod_{i \in J} P(X_t(i) = a_i | \mathcal{F}_{t-1}). \tag{2.3}$$

Moreover, the probability of having a spike in neuron i at time t is given by

$$P(X_t(i) = 1 | \mathcal{F}_{t-1}) = \phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_t^j}^{t-1} g_j(t-s) X_s(j), t - L_t^i \right), \tag{2.4}$$

where $\phi_i : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1]$ and $g_j : \mathbb{N} \rightarrow \mathbb{R}_+$ are measurable functions for all $i \in I, j \in I$. We assume that ϕ_i is uniformly Lipschitz continuous, i.e. there exists a positive constant γ such that for all $s, s' \in \mathbb{R}, n \in \mathbb{N}, i \in I$,

$$|\phi_i(s, n) - \phi_i(s', n)| \leq \gamma |s - s'|. \tag{2.5}$$

Observe that in the case where the function ϕ_i is increasing with respect to the first coordinate, the contribution of components j is either excitatory or inhibitory, depending on the sign of $W_{j \rightarrow i}$. This is reminiscent of the situation in biological neural nets in which neurons can either stimulate or inhibit the expression of other neurons.

It is natural to ask if there exists at least one stationary chain which is consistent with the above dynamics, and if so, if this process is unique. In what follows we shall construct a probability measure P on the configuration space $\Omega = \{0, 1\}^{I \times \mathbb{Z}}$ of all space-time configurations of spike trains, equipped with its natural sigma algebra \mathcal{A} . On this probability space, we consider the canonical chain $(X_t)_{t \in \mathbb{Z}}$ where for each neuron i and each time t , $X_t(i)(\omega) = \omega_t(i)$ is the projection of ω onto the (i, t) coordinate of ω .

For each neuron i , we introduce

$$\mathcal{V}_{\rightarrow i} = \{j \in I, j \neq i : W_{j \rightarrow i} \neq 0\},$$

the set of all neurons that have a direct influence on neuron i . Notice that in our model, $\mathcal{V}_{\rightarrow i}$ can be both finite or infinite. We fix a growing sequence $(V_i(k))_{k \geq -1}$ of subsets of I such that $V_i(-1) = \emptyset, V_i(0) = \{i\}, V_i(k) \subset V_i(k+1), V_i(k) \neq V_i(k+1)$ if $V_i(k) \neq \mathcal{V}_{\rightarrow i} \cup \{i\}$ and $\bigcup_k V_i(k) = \mathcal{V}_{\rightarrow i} \cup \{i\}$.

We consider two types of systems. The first system incorporates spontaneous spike times, see condition (2.6) below. These spontaneous spikes can be interpreted as external stimulus or, alternatively, as autonomous activity of the brain. The existence and uniqueness of this class is granted in our first theorem.

Theorem 1 (Existence and uniqueness in systems with spontaneous spikes) *Grant conditions (2.2) and (2.5). Assume that the functions ϕ_i and g_j satisfy moreover the following assumptions:*

1. *There exists $\delta > 0$ such that for all $i \in I, s \in \mathbb{R}, n \in \mathbb{N}$,*

$$\phi_i(s, n) \geq \delta. \tag{2.6}$$

2. *We have that*

$$G(1) + \sum_{n=2}^{\infty} (1 - \delta)^{n-2} n^2 G(n) < \infty, \tag{2.7}$$

where $G(n) = \sup_i \sum_{m=1}^n g_i(m)$ and where δ is as in condition 1.

3. *We have fast decay of the synaptic weights, i.e.*

$$\sup_i \sum_{k \geq 1} |V_i(k)| \left(\sum_{j \notin V_i(k-1)} |W_{j \rightarrow i}| \right) < \infty. \tag{2.8}$$

Then under these conditions (2.6)–(2.8), there exists a critical parameter $\delta_ \in]0, 1[$ such that for any $\delta > \delta_*$, there exists a unique probability measure P on $\{0, 1\}^{I \times \mathbb{Z}}$, under which the canonical chain satisfies (2.3) and (2.4).*

Remark 1 The stochastic chain $(X_t)_{t \in \mathbb{Z}}$ introduced in Theorem 1 is a chain having memory of infinite order (cf. [1, 4, 8, 9, 18, 20]). The setup we consider here extends what has

been done in the above cited papers. First of all, the chain we consider takes values in the infinite state space $\{0, 1\}^I$. Moreover, in Theorem 1 no summability assumption is imposed on the functions g_j . In particular, the choice $g_j(t) \equiv 1$ is possible. This implies that the specification of the chain is not continuous. More precisely, introducing

$$p_{(i,t)}(1|x) = \phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_t^i(x)}^{t-1} g_j(t-s)x_j(j), t - L_t^i \right),$$

where $L_t^i(x) = \sup\{s < t : x_s(i) = 1\}$, we have that

$$\sup_{x,y: x=y \text{ on } V_i(k) \times [t-k, t-1]} |p_{(i,t)}(1|x) - p_{(i,t)}(1|y)| \not\rightarrow 0 \text{ as } k \rightarrow \infty$$

in the case $g_j(t) \equiv 1$ for all j , which can be seen by taking configurations x and y such that $L_t^i(x) < t - k$ and $L_t^i(y) < t - k$. A similar type of discontinuity has been considered in [14] for stochastic chains with memory of variable length taking values in a finite alphabet.

As an illustration of Theorem 1 we give the following example of a system with interactions of infinite range.

Example 1 We give an example of a system satisfying the assumptions of Theorem 1. Take $I = \mathbb{Z}^d$, $g_j(s) = 1$ for all j, s , and $W_{i \rightarrow j} = \frac{1}{\|j-i\|_1^{2d+\alpha}}$ for some fixed $\alpha > 1$, where $\|\cdot\|_1$ is the L^1 -norm of \mathbb{Z}^d . In this case, if we choose $V_i(k) = \{j \in \mathbb{Z}^d : \|j - i\|_1 \leq k\}$, we have $|V_i(k)| = (k + 1)^d$, and condition (2.8) is satisfied, since

$$\begin{aligned} & \sum_{k \geq 1} |V_i(k)| \left(\sum_{j \notin V_i(k-1)} |W_{j \rightarrow i}| \right) \\ &= \sum_{k \geq 1} (k + 1)^d \sum_{l=k}^{\infty} \text{card}\{j : \|j - i\|_1 = l\} \frac{1}{l^{2d+\alpha}} \\ &\leq C(d) \sum_{k \geq 1} (k + 1)^d \sum_{l=k}^{\infty} \frac{l^{d-1}}{l^{2d+\alpha}} \leq \frac{C(d)}{d + \alpha} \sum_{k \geq 1} \frac{(k + 1)^d}{(k - 1)^{d+\alpha}} < \infty, \end{aligned}$$

as $\alpha > 1$.

The next theorem deals with the second type of system. Now we don't assume a minimal spiking rate. But additionally to the fast decay of the synaptic weights we also assume a sufficiently fast decay of the aging factor g_j , see condition (2.9) below. This additional assumption implies that the specification of the chain is continuous. This is the main difference with the setup of Theorem 1.

Theorem 2 (Existence and uniqueness in systems with uniformly summable memory) *Suppose that $\phi_i(s, n) = \phi_i(s)$ does not depend on n . Assume conditions (2.2) and (2.5) and suppose moreover that*

$$\sup_i \sum_{k \geq 0} (k + 1) \cdot |V_i(k)| \left(\sum_{j \notin V_i(k-1)} |W_{j \rightarrow i}| \sum_{n=1}^{\infty} g_j(n) + \sum_{j \in V_i(k-1)} |W_{j \rightarrow i}| \sum_{n=k \vee 1}^{\infty} g_j(n) \right) < \frac{1}{\gamma}, \tag{2.9}$$

where γ is given in (2.5).

Then there exists a unique probability measure P on $\{0, 1\}^{I \times \mathbb{Z}}$ such that under P , the canonical chain satisfies (2.3) and (2.4).

Now, for any $s < t \in \mathbb{Z}$, let $X_s^t(i) = (X_s(i), X_{s+1}(i), \dots, X_t(i))$ the trajectory of $X(i)$ between times s and t . As a byproduct of the proof of Theorems 1 and 2 we obtain the following loss of memory property.

Corollary 1

1. Under the assumptions of either Theorem 1 or Theorem 2, there exists a non increasing function $\varphi : \mathbb{N} \rightarrow \mathbb{R}_+$, such that for any $0 < s < t \in \mathbb{N}$ the following holds. For all $i \in I$, for all bounded measurable functions $f : \{0, 1\}^{[s,t]} \rightarrow \mathbb{R}_+$,

$$|E[f(X_s^t(i)) | \mathcal{F}_0] - E[f(X_s^t(i))]| \leq (t - s + 1) \|f\|_\infty \varphi(s). \tag{2.10}$$

Moreover, $\varphi(n) \leq C \frac{1}{n-1}$ for some fixed constant C .

2. Under the assumptions of Theorem 2, suppose moreover that there exists a constant $C > 0$ such that

$$g_j(n) \leq C e^{-\beta n} \quad \text{and} \quad \sup_i \sum_{j \notin V_i(n)} |W_{i \rightarrow j}| \leq C e^{-\beta n}, \tag{2.11}$$

for all $j \in I, n \in \mathbb{N}$, for some $\beta > 0$.

Then there exists a critical parameter β_* such that if $\beta > \beta_*$, (2.10) holds with

$$\varphi(s) = C \varrho^s \quad \text{for some } \varrho \in]0, 1[\text{ depending only on } \beta. \tag{2.12}$$

The proof of Theorem 1 is given in Sect. 4 below. It is based on a conditional Kalikow-type decomposition of the transition probabilities ϕ_i , where we decompose with respect to all possible interaction neighborhoods of site i . A main ingredient is the construction of an associated branching process in random environment. The setup of Theorem 2 is conceptually less difficult, since in this case the transition probabilities are continuous. This follows from the summability of the aging factors g_j . The proof of Theorem 2 relies on a space-time Kalikow-type decomposition presented in Sect. 5.

Remark 2 The proofs of both Theorem 1 and Theorem 2 imply the existence of a perfect simulation algorithm of the stochastic chain $(X_t)_{t \in \mathbb{Z}}$. By a perfect simulation algorithm we mean a simulation which samples in a finite space-time window precisely from the stationary law P . In the setup of Theorem 2 the simulation can be implemented analogously to what is presented in Galves et al. [15]. The setup of Theorem 1 requires a conditional approach, conditionally on the realization of the spontaneous spike times.

3 Correlations Between Inter-Spike Intervals in the Critical Directed Erdős-Rényi Random Graph

We consider a finite system consisting of a large number of N neurons with random synaptic weights $W_{i \rightarrow j}, i \neq j$. The sequence $W_{i \rightarrow j}, i \neq j$, is a sequence of i.i.d. Bernoulli random variables defined on some probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ with parameter $p = p_N$, i.e.

$$\tilde{P}(W_{i \rightarrow j} = 1) = 1 - \tilde{P}(W_{i \rightarrow j} = 0) = p_N,$$

where

$$p_N = \lambda/N \quad \text{and} \quad \lambda = 1 + \vartheta/N \quad \text{for some } 0 < \vartheta < \infty. \tag{3.13}$$

If we represent this as a directed graph where the directed link $i \rightarrow j$ is present if and only if $W_{i \rightarrow j} = 1$, we obtain what is called a “critical directed Erdős-Rényi random graph”. For a general reference on random graphs we refer the reader to the classical book by Bollobás [2].

We put $W_{j \rightarrow j} = 0$ for all j . Notice that the synaptic weights $W_{i \rightarrow j}$ and $W_{j \rightarrow i}$ are distinct and independent random variables. Conditionally on the choice of the connectivities $W = (W_{i \rightarrow j}, i \neq j)$, the dynamics of the chain are then given by

$$P^W(X_t(i) = 1 | \mathcal{F}_{t-1}) = \phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_t^i}^{t-1} g_j(t-s) X_s(j) \right).$$

Here we suppose that ϕ_i is a function only of the accumulated and weighted number of spikes coming from interacting neurons, but does not depend directly on the time elapsed since the last spike.

P^W denotes the conditional law of the process, conditioned on the choice of W . We write \mathbb{P} for the annealed law where we also average with respect to the random weights, i.e. $\mathbb{P} = \tilde{E}[P^W(\cdot)]$, where \tilde{E} denotes the expectation with respect to \tilde{P} .

Fix a neuron i and consider its associated sequence of successive spike times

$$\dots < S_{-n}^i < \dots < S_0^i \leq 0 < S_1^i < S_2^i < \dots < S_n^i < \dots, \tag{3.14}$$

where

$$S_1^i = \inf\{t \geq 1 : X_t(i) = 1\}, \dots, S_n^i = \inf\{t > S_{n-1}^i : X_t(i) = 1\}, \quad n \geq 2,$$

and

$$S_0^i = \sup\{t \leq 0 : X_t(i) = 1\}, \dots, S_{-n}^i = \sup\{t < S_{-n+1}^i : X_t(i) = 1\}, \quad n \geq 1.$$

Let us fix W . We are interested in the covariance between successive inter-spike intervals

$$\begin{aligned} Cov^W(S_{k+1}^i - S_k^i, S_k^i - S_{k-1}^i) &= E^W[(S_{k+1}^i - S_k^i)(S_k^i - S_{k-1}^i)] \\ &\quad - E^W(S_{k+1}^i - S_k^i)E^W(S_k^i - S_{k-1}^i), \end{aligned}$$

for any $k \neq 0, 1$. Since the process is stationary, the above covariance does not depend on the particular choice of k . The next theorem shows that neighboring inter-spike intervals are asymptotically uncorrelated as the number of neurons N tends to infinity.

Theorem 3 *Assume that (2.5), (2.6) and (2.7) are satisfied. Then there exists a measurable subset $A \in \tilde{\mathcal{A}}$, such that on A ,*

$$|Cov^W(S_3^i - S_2^i, S_2^i - S_1^i)| \leq \frac{3}{\delta^2} N(1 - \delta)^{\sqrt{N}},$$

where δ is the lower bound appearing in condition (2.6). Moreover,

$$\mathbb{P}(A^c) \leq e^{2\vartheta} N^{-1/2}.$$

For large N , if the graph of synaptic weights belongs to the “good” set A , the above result is compatible with the discussion in [16]. Gerstner and Kistler [16] deduce from this that spiking can be described by a renewal process. However, for small N or on A^c , the correlation might as well be quite big, as reported by Nawrot et al. (2007) who show that neighboring inter-spike intervals are negatively correlated. Therefore, both features are captured by our model, depending on the scale we are working in.

The proof of the above theorem is given in Sect. 6 below.

4 Conditional Kalikow-Type Decomposition and Proof of Theorem 1

In order to prove existence and uniqueness of a chain consistent with (2.3) and (2.4), we introduce a Kalikow-type decomposition of the infinite order transition probabilities. This type of decomposition was considered in the papers by Ferrari, Maass, Martínez and Ney [13], Comets, Fernández and Ferrari [7] and Galves et al. [15]. All these papers deal with the case in which the transition probabilities are continuous. This is not the case here. We are dealing with a more general case in which the transition probabilities might as well be discontinuous, see the discussion in Remark 1. This makes our approach new and interesting by itself. The new ingredient is a construction of a decomposition depending on a random environment. This random environment is given by the realization of the spontaneous spikes.

More precisely, condition (2.6) allows to introduce a sequence of i.i.d. Bernoulli random variables $(\xi_t(i), i \in I, t \in \mathbb{Z})$ of parameter δ , such that positions and times (i, t) with $\xi_t(i) = 1$ are spike times for any realization of the chain. We call these times “spontaneous spike times”. We work conditionally on the choice of $\xi_t(i), t \in \mathbb{Z}, i \in I$. In particular we will restrict everything to the state space

$$\mathcal{S}^\xi = \{x \in \{0, 1\}^{I \times \mathbb{Z}} : x_t(i) \geq \xi_t(i), \forall i \in I, t \in \mathbb{Z}\}$$

which is the space of all configurations compatible with ξ , i.e. all neural systems x such that every spike time of ξ is also a spike time of x . We write

$$R_t^i = \sup\{s < t : \xi_s(i) = 1\} \tag{4.15}$$

for the last spontaneous spike time of neuron i before time t . Moreover, for $x \in \mathcal{S}^\xi$, we put $L_t^i = L_t^i(x) = \sup\{s < t : x_s(i) = 1\}$.

Consider a couple (i, t) with $\xi_t(i) = 0$. In order to prove Theorem 1 we need to introduce the following quantities which depend on the realization of ξ , namely

$$r_{(i,t)}^{[-1]}(1) = \inf_{x \in \mathcal{S}^\xi} \phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_t^i(x)}^{t-1} g_j(t-s) x_s(j), t - L_t^i(x) \right), \tag{4.16}$$

which is the minimal probability that neuron i spikes at time t , uniformly with respect to all configurations, and

$$r_{(i,t)}^{[-1]}(0) = \inf_{x \in \mathcal{S}^\xi} \left[1 - \phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_t^i(x)}^{t-1} g_j(t-s) x_s(j), t - L_t^i(x) \right) \right], \tag{4.17}$$

which is the minimal probability that neuron i does not spike at time t .

Notice that for all $x \in \mathcal{S}^\xi$, $L_t^i(x) \geq R_t^i$. Hence in the above formulas, only a finite time window of the configuration x is used, and uniformly in $x \in \mathcal{S}^\xi$, this time window is contained in $(x_s(j), R_t^i \leq s \leq t - 1, j \in I)$. In particular the above quantities are well-defined.

Now fix $x \in \mathcal{S}^\xi$. For any $k \geq 0$, we write for short $x_{L_t^i}^{t-1}(V_i(k))$ for the space-time configuration

$$x_{L_t^i}^{t-1}(V_i(k)) = (x_s(j) : L_t^i \leq s \leq t - 1, j \in V_i(k))$$

and put

$$\begin{aligned} & r_{(i,t)}^{[k]}(1 | x_{L_t^i}^{t-1}(V_i(k))) \\ &= \inf_{z \in \mathcal{S}^\xi : z(V_i(k)) = x(V_i(k))} \phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_t^i(x)}^{t-1} g_j(t-s) z_s(j), t - L_t^i(x) \right), \end{aligned} \tag{4.18}$$

$$\begin{aligned}
 & r_{(i,t)}^{[k]}(0|x_{L_i^i}^{t-1}(V_i(k))) \\
 &= \inf_{z \in S^\xi: z(V_i(k))=x(V_i(k))} \left[1 - \phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_i^i(x)}^{t-1} g_j(t-s)z_s(j), t - L_i^i(x) \right) \right]. \tag{4.19}
 \end{aligned}$$

In what follows and whenever there is no danger of ambiguity, we will write for short $x(V_i(k))$ and $r_{(i,t)}^{[k]}(a|x(V_i(k)))$ or $r_{(i,t)}^{[k]}(a|x)$ instead of $x_{L_i^i}^{t-1}(V_i(k))$ and $r_{(i,t)}^{[k]}(a|x_{L_i^i}^{t-1}(V_i(k)))$.

We put

$$\alpha_{(i,t)}(-1) = \lambda_{(i,t)}(-1) = \sum_{a=0}^1 r_{(i,t)}^{[-1]}(a), \tag{4.20}$$

$$\alpha_{(i,t)}(k) = \inf_{x \in S^\xi} \left(\sum_{a=0}^1 r_{(i,t)}^{[k]}(a|x(V_i(k))) \right), \quad k \geq 0, \tag{4.21}$$

and

$$\lambda_{(i,t)}(k) = \alpha_{(i,t)}(k) - \alpha_{(i,t)}(k-1), \quad k \geq 0. \tag{4.22}$$

Note that $\lambda_{(i,t)}(k) \in [0, 1]$ and that $\sum_{k \geq -1} \lambda_{(i,t)}(k) = 1$ almost surely with respect to the realization of $(\xi_i(i), i \in I, t \in \mathbb{Z})$.

Remark 3 The $\lambda_{(i,t)}(k), k \geq 0$, are random variables that depend only on the realization of $(\xi_i(i), i \in I, t \in \mathbb{Z})$. More precisely, for any i, t and $k, \lambda_{(i,t)}(k)$ is measurable with respect to the sigma-algebra $\sigma(\xi_s(j), R_t^i \leq s < t, j \in I)$. We write $\lambda_{(i,t)}^\xi(k), k \geq 0$, in order to emphasize the dependence on the external field ξ .

We introduce the short hand notation

$$p_{(i,t)}(1|x) = \phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_i^i(x)}^{t-1} g_j(t-s)x_s(j), t - L_i^i(x) \right) \tag{4.23}$$

and

$$p_{(i,t)}(0|x) = 1 - p_{(i,t)}(1|x).$$

The proof of Theorem 1 is based on the following proposition.

Proposition 1 *Given ξ , for all (i, t) with $\xi_i(i) = 0$, there exists a family of conditional probabilities $(p_{(i,t)}^{[k],\xi}(a|x))_{k \geq 0}$ satisfying the following properties.*

1. *For all $a, k \geq 0, S^\xi \ni x \mapsto p_{(i,t)}^{[k],\xi}(a|x)$ depends only on the variables $(x_s(j) : L_t^i \leq s \leq t-1, j \in V_i(k))$.*
2. *For all $x \in S^\xi, k \geq 0, p_{(i,t)}^{[k],\xi}(1|x) \in [0, 1], p_{(i,t)}^{[k],\xi}(0|x) + p_{(i,t)}^{[k],\xi}(1|x) = 1$.*
3. *For all $a, x, k \geq 0, p_{(i,t)}^{[k],\xi}(a|x)$ is a $\sigma(\xi_s(j), R_t^i \leq s \leq t-1, j \in I)$ -measurable random variable.*
4. *For all $x \in S^\xi$, we have the following convex decomposition.*

$$p_{(i,t)}(a|x) = \lambda_{(i,t)}(-1)p_{(i,t)}^{[-1],\xi}(a) + \sum_{k \geq 0} \lambda_{(i,t)}(k)p_{(i,t)}^{[k],\xi}(a|x(V_i(k))), \tag{4.24}$$

where

$$p_{(i,t)}^{[-1],\xi}(a) = \frac{r_{(i,t)}^{[-1]}(a)}{\lambda_{(i,t)}(-1)}.$$

From now on, we shall omit the subscript ξ whenever there is no danger of ambiguity and write $p_{(i,t)}^{[k]}$ instead of $p_{(i,t)}^{[k],\xi}$.

Remark 4 The decomposition (4.24) of the transition probability $p_{(i,t)}(\cdot|x)$ can be interpreted as follows. In a first step, we choose a random spatial interaction range $k \in \{-1, 0, 1, \dots\}$ according to the probability distribution $\{\lambda_{(i,t)}(k), k \geq -1\}$. Once the range of the spatial interaction is fixed, we then perform a transition according to $p_{(i,t)}^{[k]}$ which depends only on the finite space-time configuration $x_{L_t^i}^{t-1}(V_i(k))$. A comprehensive introduction to this technique can be found in the lecture notes of Fernández et al. [10].

Example 2 Suppose that all interactions are excitatory, i.e. $W_{j \rightarrow i} \geq 0$ for all $i \neq j$. Suppose further that $g_i(s) = 1$ for all $s \in \mathbb{N}, i \in I$, and that $\Phi_i(s, n) = \Phi_i(s)$ does not depend on n and is non-decreasing in s . Then

$$r_{(i,t)}^{[-1]}(1) = \Phi_i \left(\sum_j W_{j \rightarrow i} \xi_{t-1}(j) \right) \quad \text{and} \quad r_{(i,t)}^{[-1]}(0) = 1 - \Phi_i \left(\sum_j W_{j \rightarrow i} (t - R_t^i) \right).$$

Hence,

$$\lambda_{(i,t)}(-1) = 1 + \Phi_i \left(\sum_j W_{j \rightarrow i} \xi_{t-1}(j) \right) - \Phi_i \left(\sum_j W_{j \rightarrow i} (t - R_t^i) \right).$$

Similarly,

$$r_{(i,t)}^{[k]}(1|x) = \Phi_i \left(\sum_{j \notin V_i(k)} W_{j \rightarrow i} \sum_{m=L_t^i(x)}^{t-1} \xi_s(j) + \sum_{j \in V_i(k)} W_{j \rightarrow i} \sum_{s=L_t^i(x)}^{t-1} x_s(j) \right)$$

and

$$r_{(i,t)}^{[k]}(0|x) = 1 - \Phi_i \left(\sum_{j \notin V_i(k)} W_{j \rightarrow i} (t - L_t^i(x)) + \sum_{j \in V_i(k)} W_{j \rightarrow i} \sum_{s=L_t^i(x)}^{t-1} x_s(j) \right).$$

Proof of Proposition 1 We have for any $N \geq 1, a \in \{0, 1\}$ and $x \in \mathcal{S}^\xi$,

$$p_{(i,t)}(a|x) = r_{(i,t)}^{[-1]}(a) + \left(\sum_{k=0}^N \Delta_{(i,t)}^{[k]}(a|x(V_i(k))) \right) + (p_{(i,t)}(a|x) - r_{(i,t)}^{[N]}(a|x)),$$

where

$$\Delta_{(i,t)}^{[k]}(a|x(V_i(k))) = r_{(i,t)}^{[k]}(a|x(V_i(k))) - r_{(i,t)}^{[k-1]}(a|x(V_i(k-1))).$$

Now, due to condition (2.5),

$$|p_{(i,t)}(a|x) - r_{(i,t)}^{[N]}(a|x)| \leq \gamma \left(\sum_{s=R_t^i}^{t-1} \sup_j g_j(t-s) \right) \sum_{j \notin V_i(N)} |W_{j \rightarrow i}| \rightarrow 0$$

as $N \rightarrow \infty$ due to (2.2). In the above upper bound we used that

$$\sum_{s=L_t^i}^{t-1} g_j(t-s) |z_s(j) - x_s(j)| \leq \sum_{s=R_t^i}^{t-1} \sup_j g_j(t-s) < \infty$$

almost surely, which is a consequence of (2.7). Therefore we obtain the following decomposition.

$$p_{(i,t)}(a|x) = r_{(\tilde{\lambda}_{(i,t)})}^{[-1]}(a) + \sum_{k=0}^{\infty} \Delta_{(\tilde{\lambda}_{(i,t)})}^{[k]}(a|x(V_i(k))), \quad a \in \{0, 1\}, \quad \text{for all } x \in \mathcal{S}^{\mathbb{Z}}. \quad (4.25)$$

Now let

$$p_{(i,t)}^{[-1]}(a) = \frac{r_{(\tilde{\lambda}_{(i,t)})}^{[-1]}(a)}{\lambda_{(\tilde{\lambda}_{(i,t)})}(-1)}.$$

Moreover, for $k \geq 0$, put

$$\tilde{\lambda}_{(i,t)}(k, x(V_i(k))) = \sum_a \Delta_{(\tilde{\lambda}_{(i,t)})}^{[k]}(a|x(V_i(k))), \quad (4.26)$$

and for any $(i, t), k$ such that $\tilde{\lambda}_{(i,t)}(k, x(V_i(k))) > 0$, we define

$$\tilde{p}_{(i,t)}^{[k]}(a|x(V_i(k))) = \frac{\Delta_{(\tilde{\lambda}_{(i,t)})}^{[k]}(a|x(V_i(k)))}{\tilde{\lambda}_{(i,t)}(k, x(V_i(k)))}.$$

For $(i, t), k$ such that $\tilde{\lambda}_{(i,t)}(k, x(V_i(k))) = 0$, define $\tilde{p}_{(i,t)}^{[k]}(a|x(V_i(k)))$ in an arbitrary fixed way. Hence

$$p_{(i,t)}(a|x) = \lambda_{(\tilde{\lambda}_{(i,t)})}(-1)p_{(\tilde{\lambda}_{(i,t)})}^{[-1]}(a) + \sum_{k=0}^{\infty} \tilde{\lambda}_{(i,t)}(k, x(V_i(k)))\tilde{p}_{(i,t)}^{[k]}(a|x(V_i(k))). \quad (4.27)$$

In (4.27) the factors $\tilde{\lambda}_{(i,t)}(k, x(V_i(k))), k \geq 0$, still depend on $x_{L_i}^{l-1}(V_i(k))$. To obtain the desired decomposition, we must rewrite it as follows.

For any (i, t) , take the sequences $\alpha_{(i,t)}(k), \lambda_{(i,t)}(k), k \geq -1$, as defined in (4.21) and (4.22), respectively. Define the new quantities

$$\alpha_{(i,t)}(k, x(V_i(k))) = \sum_{l \leq k} \tilde{\lambda}_{(i,t)}(l, x(V_i(l)))$$

and notice that

$$\alpha_{(i,t)}(k, x(V_i(k))) = \sum_a r_{(\tilde{\lambda}_{(i,t)})}^{[k]}(a, x(V_i(k)))$$

is the total mass associated to $r_{(\tilde{\lambda}_{(i,t)})}^{[k]}(\cdot, x(V_i(k)))$. From now on and for the rest of the proof we shall write for short $\alpha_{(i,t)}(k, x)$ instead of writing $\alpha_{(i,t)}(k, x(V_i(k)))$.

Reading (4.27) again, this means that for any $k \geq 0$, we have to use the transition probability $\tilde{p}_{(i,t)}^{[k]}$ on the interval $]\alpha_{(i,t)}(k - 1, x), \alpha_{(i,t)}(k, x)[$.

By definition of $\alpha_{(i,t)}(k)$ in (4.21), $\alpha_{(i,t)}(k)$ is the smallest total mass associated to $r_{(\tilde{\lambda}_{(i,t)})}^{[k]}$, uniformly with respect to all possible neighborhoods $x(V_i(k))$. Hence, in order to get the decomposition (4.24) with weights $\lambda_{(i,t)}(k)$ not depending on the configuration, we have to define a partition of the interval $[0, \alpha_{(i,t)}(k, x)]$ according to the values of $\alpha_{(i,t)}(k)$ and we have to define probabilities $p_{(i,t)}^{[k]}$ working on the intervals $]\alpha_{(i,t)}(k - 1), \alpha_{(i,t)}(k)[$. This can be done as follows.

Fix $k \geq 0$ and suppose that for some $l' \leq l \leq k - 1$,

$$\alpha_{(i,t)}(l' - 1, x) < \alpha_{(i,t)}(k - 1) \leq \alpha_{(i,t)}(l', x) < \dots < \alpha_{(i,t)}(l, x) < \alpha_{(i,t)}(k) \leq \alpha_{(i,t)}(l + 1, x),$$

and therefore

$$\begin{aligned}]\alpha_{(i,t)}(k-1), \alpha_{(i,t)}(k)] &=]\alpha_{(i,t)}(k-1), \alpha_{(i,t)}(l', x)] \\ &\cup \left(\bigcup_{m=l'+1}^l]\alpha_{(i,t)}(m-1, x), \alpha_{(i,t)}(m, x)] \right) \\ &\cup]\alpha_{(i,t)}(l, x), \alpha_{(i,t)}(k)]. \end{aligned}$$

Hence the probability $p_{(i,t)}^{[k]}$ that has to be defined on the interval $]\alpha_{(i,t)}(k-1), \alpha_{(i,t)}(k)]$ has to be decomposed according to the above decomposition into sub-intervals. On the first interval $]\alpha_{(i,t)}(k-1), \alpha_{(i,t)}(l', x)]$, we have to use the original probability $\tilde{p}_{(i,t)}^{[l']}$, on each of the intervals $]\alpha_{(i,t)}(m-1, x), \alpha_{(i,t)}(m, x)]$, we have to use $\tilde{p}_{(i,t)}^{[m]}$, and finally on $]\alpha_{(i,t)}(l, x), \alpha_{(i,t)}(k)]$, we use $\tilde{p}_{(i,t)}^{[l+1]}$.

This yields, for any $k \geq 0$, the following definition of the conditional finite range probability densities.

$$\begin{aligned} p_{(i,t)}^{[k]}(a|x(V_i(k))) &= \sum_{-1=l' \leq l}^{k-1} \mathbf{1}_{\{\alpha_{(i,t)}(l'-1, x) < \alpha_{(i,t)}(k-1) \leq \alpha_{(i,t)}(l', x)\}} \mathbf{1}_{\{\alpha_{(i,t)}(l, x) < \alpha_{(i,t)}(k) \leq \alpha_{(i,t)}(l+1, x)\}} \\ &\times \left[\frac{\alpha_{(i,t)}(l', x) - \alpha_{(i,t)}(k-1)}{\lambda_{(i,t)}(k)} \tilde{p}_{(i,t)}^{[l']}(a|x(V_i(l'))) \right. \\ &+ \sum_{m=l'+1}^l \frac{\tilde{\lambda}_{(i,t)}(m, x(V_i(m)))}{\lambda_{(i,t)}(k)} \tilde{p}_{(i,t)}^{[m]}(a|x(V_i(m))) \\ &\left. + \frac{\alpha_{(i,t)}(k) - \alpha_{(i,t)}(l, x)}{\lambda_{(i,t)}(k)} \tilde{p}_{(i,t)}^{[l+1]}(a|x(V_i(l+1))) \right]. \end{aligned} \tag{4.28}$$

Note that by construction, the above defined probability $p_{(i,t)}^{[k]}$ depends only on the configuration $x(V_i(l+1))$, hence at most on $x(V_i(k))$, since $l \leq k-1$. Multiplying the above formula with $\lambda_{(i,t)}(k)$ and summing over all k shows that the decomposition (4.27) implies the desired decomposition (4.24). This finishes our proof. \square

Thanks to condition (2.5), the following estimate holds. It will be useful in the sequel.

Lemma 1 *We have for all $k \geq 1$,*

$$\begin{aligned} \lambda_{(i,t)}^\xi(k) &\leq \gamma \sum_{j \notin V_i(k-1)} |W_{j \rightarrow i}| \sum_{s=R_i^j}^{t-1} g_j(t-s)(1-\xi_s(j)) \\ &\leq \gamma \left(\sum_{s=R_i^j}^{t-1} \sup_j g_j(t-s) \right) \sum_{j \notin V_i(k-1)} |W_{j \rightarrow i}|. \end{aligned} \tag{4.29}$$

We use this estimate in

Proof of Theorem 1 We work conditionally on the realization of the process $\xi_t(i), t \in \mathbb{Z}, i \in I$. Take a couple $(i, t), i \in I, t \in \mathbb{Z}$, such that $\xi_t(i) = 0$. We have to show that it is possible

to decide in a uniquely determined manner if t will be a spike time or not. In order to achieve this decision, we have to calculate $p_{(i,t)}(\cdot|x)$, where x is the unknown history of the process.

We will construct a sequence of sets $(C_n^{(i,t)})_n, C_n^{(i,t)} \subset I \times]-\infty, t - 1]$, which contain the sets of sites and anterior spike times that have an influence on the appearance of a spike at time t for neuron i . The choice of these sets is based on the decomposition (4.24).

First, for any couple (j, s) with $\xi_s(j) = 0$, we choose, independently from anything else, an interaction neighborhood $\mathcal{V}_{(j,s)} = V_j(k), k \geq -1$, with probability $\lambda_{(j,s)}^\xi(k)$. Here, we put $V_j(-1) = \emptyset$. A choice $k = -1$ for a couple (j, s) implies that we can immediately decide to accept a spike at time s for neuron j with probability

$$\frac{r_{(j,s)}^{[-1]}(1) - \delta}{1 - \delta}$$

and to reject it with probability

$$\frac{r_{(j,s)}^{[-1]}(0)}{1 - \delta}.$$

We suppose that the choice of all $\mathcal{V}_{(j,s)}$ is fixed. Then by (4.24), the decision concerning the couple (i, t) depends on the configuration of the past $x_{L_i^{t-1}}(\mathcal{V}_{(i,t)})$. Since we do not know L_i^t , we use the a priori estimate on L_i^t which is given by R_i^t . Thus we consider the worst case in which we have to evaluate the past up to time R_i^t .

In order to decide about the length $t - L_i^t$, we have to assign values 0 or 1 to all couples $(i, s), R_i^t < s \leq t$. All these couples are influenced by their associated interaction region $\mathcal{V}_{(i,s)}$. So we consider

$$C_1^{(i,t)} = \bigcup_{s=R_i^t+1}^t \bigcup_{j \in \mathcal{V}_{(i,s)} \setminus \{i\}} \bigcup_{u=R_i^t}^{s-1} \{(j, u), \xi_j(u) = 0\}. \tag{4.30}$$

It is clear that if we know the values of all couples belonging to $C_1^{(i,t)}$ then we are able to assign a value to any $(i, s), R_i^t < s \leq t$. Therefore, we call $C_1^{(i,t)}$ the set of ancestors of generation 1 of (i, t) . Continuing this procedure, any couple $(j, s) \in C_1^{(i,t)}$ itself has to be replaced by the set of its ancestors $C_1^{(j,s)}$. We put

$$C_2^{(i,t)} = \left(\bigcup_{(j,s) \in C_1^{(i,t)}} C_1^{(j,s)} \right) \setminus C_1^{(i,t)}, \tag{4.31}$$

where $C_1^{(j,s)}$ is defined as in (4.30), and then recursively,

$$C_n^{(i,t)} = \left(\bigcup_{(j,s) \in C_{n-1}^{(i,t)}} C_1^{(j,s)} \right) \setminus (C_1^{(i,t)} \cup \dots \cup C_{n-1}^{(i,t)}). \tag{4.32}$$

We will show below that almost surely there exists a first time $n < \infty$ such that $C_n^{(i,t)} = \emptyset$. Then necessarily $C_{n-1}^{(i,t)}$ consists only of couples (j, s) which chose an interaction neighborhood of range -1 or which interact only with couples representing spontaneous spiking. Thus we can decide to accept or to reject a spike for any of them independently of anything else. Once the values associated to all elements of $C_{n-1}^{(i,t)}$ are known, we can realize all decisions needed in order to attribute values to the elements of $C_{n-2}^{(i,t)}$. In this way, we will be able to assign values in a recursive way to all ancestor sets up to the first set, $C_1^{(i,t)}$, which allows us finally to assign a value for neuron i at time t . We call this procedure the forward

coloring procedure and we call $Y_t(i)$ the value of neuron t at time t obtained at the end of this procedure.

1. We first show that the probability constructed above is the law of $X_t(i)$ under the unique invariant measure P . Put $C_\infty^{(i,t)} = \bigcup_{n \geq 1} C_n^{(i,t)}$ and let

$$T_{STOP}^{(i,t)} = \inf\{s : C_\infty^{(i,t)} \subset I \times [t - s, t - 1]\}. \tag{4.33}$$

Fix some initial space-time configuration $\eta \in \{0, 1\}^{I \times \mathbb{N}}$ such that $\eta_0(i) = 1$ for all $i \in I$. Let X_t^η be the chain that evolves according to (2.3) and (2.4), conditionally on $X_{-\infty}^0 = \eta$. Recall that $Y_t(i)$ is the value obtained at the end of the above described forward coloring procedure. Then for any $f : \{0, 1\} \rightarrow \mathbb{R}_+$,

$$\begin{aligned} E(f(X_t^\eta(i))) &= E(f(X_t^\eta(i)), T_{STOP}^{(i,t)} < t) + E(f(X_t^\eta(i)), T_{STOP}^{(i,t)} \geq t) \\ &= E(f(Y_t(i)), T_{STOP}^{(i,t)} < t) + E(f(X_t^\eta(i)), T_{STOP}^{(i,t)} \geq t). \end{aligned} \tag{4.34}$$

But

$$E(f(X_t^\eta(i)), T_{STOP}^{(i,t)} \geq t) \leq \|f\|_\infty P(T_{STOP}^{(i,t)} \geq t) \rightarrow 0 \quad \text{as } t \rightarrow \infty.$$

Hence we obtain that

$$\lim_{t \rightarrow \infty} E(f(X_t^\eta(i))) = E(f(Y_t(i))),$$

since $1_{\{T_{STOP}^{(i,t)} < t\}} \rightarrow 1$ almost surely.

This implies that P is the unique invariant measure of the process.

2. We now show that the above procedure stops after a finite time almost surely. For that sake we put

$$N_{STOP}^{(i,t)} = \min\{n : C_n^{(i,t)} = \emptyset\}. \tag{4.35}$$

Our goal is to show that $N_{STOP}^{(i,t)} < \infty$ almost surely. Notice that this implies that $T_{STOP}^{(i,t)} < \infty$ as well. In order to do so, let

$$|C_n^{(i,t)}|, \quad n \geq 1,$$

be the cardinal of the set of ancestors after n steps of the above procedure. Then

$$P(N_{STOP}^{(i,t)} > n) \leq E(|C_n^{(i,t)}|). \tag{4.36}$$

As a consequence, it is sufficient to show that $E(|C_n^{(i,t)}|) \rightarrow 0$ as $n \rightarrow \infty$. For that sake we compare $(|C_n^{(i,t)}|)_n$ to a branching process in random environment, where the environment is given by the i.i.d. field ξ .

Recall the definition of $G(n) = \sum_{m=1}^n \sup_i g_i(m)$ and the upper bound (4.29). Thus for any $k \geq 1$,

$$\lambda_{(i,t)}(k) \leq \gamma \left(\sum_{s=R_t^i}^{t-1} \sup_j g_j(t-s) \right) \sum_{j \notin V_i(k-1)} |W_{j \rightarrow i}| = \gamma G(t - R_t^i) \sum_{j \notin V_i(k-1)} |W_{j \rightarrow i}|.$$

Let

$$\bar{\lambda}_{(i,t)}(k) = \gamma G(t - R_t^i) \sum_{j \notin V_i(k-1)} |W_{j \rightarrow i}|. \tag{4.37}$$

Notice that $\bar{\lambda}_{(i,t)}(k)$ depends only on the realization of the i.i.d. field ξ through the value of $t - R_t^i$.

Since $\lambda_{(i,t)}(k) \leq \bar{\lambda}_{(i,t)}(k)$ for all $k \geq 1$, it is possible, by standard branching process coupling arguments, to construct a sequence $(\bar{C}_n^{(i,t)})_n$ coupled with the original sequence $(C_n^{(i,t)})_n$ satisfying the following properties.

1. $\bar{C}_n^{(i,t)}$ is defined through (4.30) and (4.31) by using the family $(\bar{\lambda}_{(i,t)}(k))$ instead of $(\lambda_{(i,t)}(k))$.
2. For all $n, |\bar{C}_n^{(i,t)}| \leq |\bar{C}_n^{(i,t)}|$.

Hence it is sufficient to show that $E(|\bar{C}_n^{(i,t)}|)$ tends to 0 as $n \rightarrow \infty$. In order to do so, we will control the reproduction mean depending on the environment ξ . Here, reproduction mean stands for the mean number of sites belonging to $\bar{C}_1^{(j,s)}$, for any fixed couple (j, s) , where the expectation is taken with respect to the choices of the interaction neighborhoods $\bar{V}_{(i,s)}$, conditionally on the realization of ξ . More precisely, given $\bar{C}_n^{(i,t)} = c$, the reproduction mean of any couple (j, s) belonging to $\bar{C}_n^{(i,t)}$ is given by

$$E^\xi(|\bar{C}_1^{(j,s)}| \mid (j, s) \in \bar{C}_n^{(i,t)} = c) \leq \sum_{\tilde{s}=R_s^j+1}^s \sum_{k \geq 1} \bar{\lambda}_{(j,\tilde{s})}(k) \left(\sum_{l \in V_j(k), l \neq j} \sum_{u=R_s^j}^{\tilde{s}-1} [1 - \xi_u(l)] 1_{\{(l,u) \notin c\}} \right),$$

where E^ξ denotes expectation conditionally on ξ .

In what follows we upper bound the above expression. We first use that, by definition (4.37), for $\tilde{s} \leq s, \bar{\lambda}_{(j,\tilde{s})}(k) \leq \bar{\lambda}_{(j,s)}(k)$. Therefore,

$$E^\xi(|\bar{C}_1^{(j,s)}| \mid (j, s) \in \bar{C}_n^{(i,t)} = c) \leq (s - R_s^j) \sum_{k \geq 1} \bar{\lambda}_{(j,s)}(k) \left(\sum_{l \in V_j(k), l \neq j} \sum_{u=R_s^j}^{s-1} [1 - \xi_u(l)] 1_{\{(l,u) \notin c\}} \right).$$

Recalling the explicit form of $\bar{\lambda}_{(j,s)}(k)$ in (4.37) and using moreover the upper bound

$$\sum_{u=R_s^j}^{s-1} [1 - \xi_u(l)] 1_{\{(l,u) \notin c\}} \leq s - R_s^j,$$

on the event $R_s^j < s - 1$, we obtain

$$\begin{aligned} E^\xi(|\bar{C}_1^{(j,s)}| \mid (j, s) \in \bar{C}_n^{(i,t)} = c) &\leq 2\gamma \left(1_{\{R_s^j < s-1\}} (s - R_s^j)^2 G(s - R_s^j) \sum_{k \geq 1} |V_j(k)| \left(\sum_{l \notin V_j(k-1)} |W_{l \rightarrow j}| \right) \right. \\ &\quad \left. + 1_{\{R_s^j = s-1\}} G(1) \sum_{k \geq 1} \left(\sum_{l \notin V_j(k-1)} |W_{l \rightarrow j}| \right) \sum_{l \in V_j(k), l \neq j} (1 - \xi_{s-1}(l)) 1_{\{(l,s-1) \notin c\}} \right) \\ &=: \bar{m}_{(j,s)}^c. \end{aligned} \tag{4.38}$$

Observe that $\bar{m}_{(j,s)}^c$ depends on ξ , but to avoid too cumbersome notation, we omit the superscript ξ .

Taking conditional expectation, conditionally with respect to ξ , we get

$$E^\xi(|\bar{C}_n^{(i,t)}| \mid \bar{C}_{n-1}^{(i,t)}) \leq \sum_{(j,s) \in \bar{C}_{n-1}^{(i,t)}} \bar{m}_{(j,s)}^{\bar{C}_{n-1}^{(i,t)}}. \tag{4.39}$$

Therefore, taking expectation with respect to ξ ,

$$E(|\bar{C}_n^{(i,t)}|) \leq \sum_{(j,s)} E(1_{(j,s) \in \bar{C}_{n-1}^{(i,t)}} \bar{m}_{(j,s)}^{\bar{C}_{n-1}^{(i,t)}}). \tag{4.40}$$

Now note that the event

$$\{(j, s) \in \bar{C}_{n-1}^{(i,t)}\} = \{\exists(k, \tilde{s}) \in \bar{C}_{n-2}^{(i,t)} : \tilde{s} > s, k \neq j, j \in \bar{V}_{(k,\tilde{s})}\}$$

depends only on R_s^k . Hence, recalling (4.38), the above event is independent of $\bar{m}_{(j,s)}^{\bar{C}_{n-1}^{(i,t)}}$ which depends only on $s - R_s^j$. As a consequence,

$$E(1_{(j,s) \in \bar{C}_{n-1}^{(i,t)}} \bar{m}_{(j,s)}^{\bar{C}_{n-1}^{(i,t)}}) = P((j, s) \in \bar{C}_{n-1}^{(i,t)}) E(\bar{m}_{(j,s)}^{\bar{C}_{n-1}^{(i,t)}}).$$

But

$$E(\bar{m}_{(j,s)}^{\bar{C}_{n-1}^{(i,t)}}) \leq 2\gamma \left[\sum_{n=2}^{\infty} \delta(1 - \delta)^{n-1} n^2 G(n) \sum_{k \geq 1} |V_j(k)| \left(\sum_{l \notin V_j(k-1)} |W_{l \rightarrow j}| \right) + \delta G(1) \sum_{k \geq 1} \left(\sum_{l \notin V_j(k-1)} |W_{l \rightarrow j}| \right) |V_j(k)| (1 - \delta) \right]. \tag{4.41}$$

We put

$$C_\gamma := 2\gamma \sup_j \sum_{k \geq 1} |V_j(k)| \left(\sum_{l \notin V_j(k-1)} |W_{l \rightarrow j}| \right) \tag{4.42}$$

and define

$$E(G, \delta) = G(1) + \sum_{n=2}^{\infty} (1 - \delta)^{n-2} n^2 G(n).$$

Then

$$E[\bar{m}_{(j,s)}^{\bar{C}_{n-1}^{(i,t)}}] \leq C_\gamma (1 - \delta) E(G, \delta) =: e(\delta). \tag{4.43}$$

Note that $e(\delta)$ is decreasing as a function of δ and tends to 0 for $\delta \rightarrow 1$. In particular, if we put

$$\delta^* = \inf\{\delta \in]0, 1[: e(\delta) \leq 1\}, \tag{4.44}$$

then $e(\delta) < 1$ for all $\delta \geq \delta_*$. Now we may iterate (4.40) and (4.43) and obtain

$$E(|\bar{C}_n^{(i,t)}|) \leq \sum_{(j,s)} E(1_{\{(j,s) \in \bar{C}_{n-1}^{(i,t)}\}}) e(\delta) = E(|\bar{C}_{n-1}^{(i,t)}|) e(\delta) \leq e(\delta)^n. \tag{4.45}$$

Since $e(\delta)^n \rightarrow 0$ as $n \rightarrow \infty$ for $\delta \geq \delta_*$, this implies our result. □

Remark 5 Note that the value of δ^* given in (4.44) can be explicitly calculated depending on the specific structure of the aging functions $g_j(s)$. For instance, if $g_j(s) = 1$ for all j, s , then $G(n) = n$ for all n , and the value of δ^* follows from standard evaluations of the third moment of a geometrical distribution.

Remark 6 The above proof uses a non-trivial extension of the so-called Clan of ancestors method employed by Fernández et al. [11, 12]. In these papers the authors study the Clan of ancestors of a given vertex (or object) in a random system. They prove that this set is almost surely finite and deduce a perfect simulation algorithm.

Proof of Corollary 1, part 1 We prove the first part of Corollary 1. We keep the notation of the proof of Theorem 1. Put $C_\infty^{(i,t)} = \bigcup_n C_n^{(i,t)}$. Then the random variable $T_{STOP}^{(i,t)}$ defined in (4.33) can be trivially upper bounded by

$$T_{STOP}^{(i,t)} \leq |C_\infty^{(i,t)}|, \tag{4.46}$$

which is the total number of elements appearing in the ancestor process. This is a very rough upper bound on the number of steps that we have to look back into the past in order to choose a value for $X_i(t)$.

By construction, the value $X_i(t)$ depends on all choices of interaction regions $\mathcal{V}_{(j,s)}$ such that $(j, s) \in C_\infty^{(i,t)}$ and on the values of the i.i.d. Bernoulli field ξ_u for $t - T_{STOP}^{(i,t)} \leq u \leq t$. Moreover, for any $(j, s) \in C_\infty^{(i,t)}$, a random decision has to be made whether to associate the value +1 or 0 to this couple. These decisions can be realized based on a sequence of i.i.d. uniform random variables $(U_i(i), i \in I, t \in \mathbb{Z})$, which are uniform on $[0, 1]$. As a consequence, $X_i(t)$ is measurable with respect to the sigma-algebra

$$\sigma\{\mathcal{V}_{(j,s)}, U_s(j) : (j, s) \in C_\infty^{(i,t)}, \xi_u : t - T_{STOP}^{(i,t)} \leq u \leq t\}.$$

Writing

$$R = \inf_{u \in [s,t]} (u - T_{STOP}^{(i,u)}), \tag{4.47}$$

we deduce that

$$X'_s(i) \text{ is } \sigma\{\mathcal{V}_{(j,u)}, U_u(j), \xi_u(j) : j \in I, R \leq u \leq t\}\text{-measurable.} \tag{4.48}$$

Let $A \in \mathcal{F}_0$. We write E^ξ for conditional expectation, conditionally with respect to ξ . Then we have

$$\begin{aligned} E[f(X'_s(i))1_A] &= E[f(X'_s(i))1_A 1_{\{R>0\}}] + E[f(X'_s(i))1_A 1_{\{R\leq 0\}}] \\ &= E[E^\xi[f(X'_s(i))1_A 1_{\{R>0\}}]] + E[E^\xi[f(X'_s(i))1_A 1_{\{R\leq 0\}}]]. \end{aligned}$$

Clearly, $f(X'_s(i))1_{\{R>0\}}$ and A are independent, which follows from (4.48). Hence

$$E^\xi[f(X'_s(i))1_A] = E^\xi[f(X'_s(i))1_{\{R>0\}}]P^\xi(A) + E^\xi[f(X'_s(i))1_A 1_{\{R\leq 0\}}].$$

We take expectation with respect to ξ and use the fact that $E^\xi[f(X'_s(i))1_{\{R>0\}}]$ is measurable with respect to $\sigma\{\xi_u : 1 \leq u \leq t\}$ and $P^\xi(A)$ is measurable with respect to $\sigma\{\xi_u : u \leq 0\}$. Hence by independence

$$E[f(X'_s(i))1_A] = E[f(X'_s(i))1_{\{R>0\}}]P(A) + E[f(X'_s(i))1_A 1_{\{R\leq 0\}}],$$

and therefore

$$\begin{aligned} &E[f(X'_s(i))1_A] - E[f(X'_s(i))]P(A) \\ &= E[f(X'_s(i))1_A 1_{\{R\leq 0\}}] - E[f(X'_s(i))1_{\{R\leq 0\}}]P(A). \end{aligned}$$

As a consequence,

$$\begin{aligned} |E[f(X'_s(i))1_A] - E[f(X'_s(i))]P(A)| &\leq \|f\|_\infty P(A)[P(R \leq 0|A) + P(R \leq 0)] \\ &\leq \|f\|_\infty P(A)[P(R \leq 1|A) + P(R \leq 0)] \\ &\leq 2\|f\|_\infty P(A)P(R \leq 1). \end{aligned}$$

Here we have used that $\{R \leq 1\}$ is independent of A . Indeed, the event $\{R \leq 1\}$ does only depend on choices of $\mathcal{V}_{(j,s)}$ and $\xi_s(j)$ having time component $s \geq 1$. Now we conclude as follows. By definition of R ,

$$P(R \leq 1) \leq \sum_{u \in [s,t]} P(T_{STOP}^{(i,u)} \geq s - 1).$$

Using (4.46), we obtain

$$P(T_{STOP}^{(i,u)} \geq s - 1) \leq P(|C_\infty^{(i,u)}| \geq s - 1) \leq E(|C_\infty^{(i,u)}|) \frac{1}{s - 1} \leq \frac{1}{1 - e(\delta)} \frac{1}{s - 1},$$

since

$$E(|C_\infty^{(i,u)}|) \leq \sum_{n=1}^\infty E(|C_n^{(i,u)}|) \leq \frac{1}{1 - e(\delta)},$$

where we have used (4.45). This implies the result. □

5 A Space-Time Kalikow-Type Decomposition and Proof of Theorem 2

The additional condition (2.9) allows to introduce a space-time Kalikow-type decomposition without the assumption of existence of spontaneous spikes. We will decompose with respect to increasing space-time neighborhoods $V_i(k) \times [-k - 1, -1]$, $k \geq -1$, where $V_i(-1) = \emptyset$. Write $\mathcal{S} = \{0, 1\}^{I \times \mathbb{Z}}$ for the state space of the process and introduce, analogously to (4.16)–(4.19),

$$r_i^{[-1]}(1) = \inf_{x \in \mathcal{S}} \Phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_0^i(x)}^{-1} g_j(-s)x_s(j) \right), \tag{5.49}$$

$$r_i^{[-1]}(0) = \inf_{x \in \mathcal{S}} \left(1 - \Phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_0^i(x)}^{-1} g_j(-s)x_s(j) \right) \right), \tag{5.50}$$

$$r_i^{[0]}(1|x) = \inf_{z \in \mathcal{S}: z_i(-1)=x_i(-1)} \Phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_0^i(z)}^{-1} g_j(-s)z_s(j) \right), \tag{5.51}$$

$$r_i^{[0]}(0|x) = \inf_{z \in \mathcal{S}: z_i(-1)=x_i(-1)} \left(1 - \Phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_0^i(z)}^{-1} g_j(-s)z_s(j) \right) \right), \tag{5.52}$$

and then for any $k \geq 1$,

$$r_i^{[k]}(1|x) = \inf_{z \in \mathcal{S}: z(V_i(k) \times [-k-1, -1])=x(V_i(k) \times [-k-1, -1])} \Phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_0^i(z)}^{-1} g_j(-s)z_s(j) \right), \tag{5.53}$$

$$r_i^{[k]}(0|x) = \inf_{z \in \mathcal{S}: z(V_i(k) \times [-k-1, -1])=x(V_i(k) \times [-k-1, -1])} \left(1 - \Phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_0^i(z)}^{-1} g_j(-s)z_s(j) \right) \right). \tag{5.54}$$

Putting

$$\alpha_i(-1) = \lambda_i(-1) = r_i^{[-1]}(1) + r_i^{[-1]}(0),$$

$$\alpha_i(k) = \inf_x (r_i^{[k]}(1|x) + r_i^{[k]}(0|x)), \quad \lambda_i(k) = \alpha_i(k) - \alpha_i(k-1), \quad k \geq 0, \quad (5.55)$$

we obtain the following space-time Kalikow-type decomposition for

$$p_{(i,t)}(1|x) = \Phi_i \left(\sum_j W_{j \rightarrow i} \sum_{s=L_i^j(x)}^{t-1} g_j(t-s)x_s(j) \right).$$

Proposition 2 *Under the conditions (2.2), (2.5) and (2.9), for any $i \in I$, the above defined quantities $\lambda_i(k)$ take values in $[0, 1]$ and*

$$\sum_{k \geq -1} \lambda_i(k) = 1.$$

Moreover, there exists a family of conditional probabilities $(p_{(i,t)}^{[k]}(a|x))_{k \geq 0}$ satisfying the following properties.

1. For all a , $\mathcal{S} \ni x \mapsto p_{(i,t)}^{[0]}(a|x)$ depends only on the variable $x_{t-1}(i)$.
2. For all a , $k \geq 1$, $\mathcal{S} \ni x \mapsto p_{(i,t)}^{[k]}(a|x)$ depends only on the variables $(x_s(j) : t-k-1 \leq s \leq t-1, j \in V_i(k))$.
3. For all $x \in \mathcal{S}$, $k \geq 0$, $p_{(i,t)}^{[k]}(1|x) \in [0, 1]$, $p_{(i,t)}^{[k]}(0|x) + p_{(i,t)}^{[k]}(1|x) = 1$.
4. For all $x \in \mathcal{S}$, we have the following convex decomposition

$$p_{(i,t)}(a|x) = \lambda_i(-1)p_{(i,t)}^{[-1]}(a) + \sum_{k \geq 0} \lambda_i(k)p_{(i,t)}^{[k]}(a|x), \quad (5.56)$$

where

$$p_{(i,t)}^{[-1]}(a) = \frac{r_i^{[-1]}(a)}{\lambda_i(-1)}.$$

The proof of this proposition follows the lines of the proof of Proposition 1. Moreover, the following estimates hold.

$$\lambda_i(0) \leq \gamma \sum_j |W_{j \rightarrow i}| \sum_{n=1}^{\infty} g_j(n), \quad (5.57)$$

and for all $k \geq 1$,

$$\lambda_i(k) \leq \gamma \left(\sum_{j \notin V_i(k-1)} |W_{j \rightarrow i}| \sum_{n=1}^{\infty} g_j(n) + \sum_{j \in V_i(k-1)} |W_{j \rightarrow i}| \sum_{n=k}^{\infty} g_j(n) \right). \quad (5.58)$$

Proof of Theorem 2 We construct again a sequence of sets $(C_n^{(i,t)})_n \subset I \times]-\infty, t-1]$ which contain the sets of sites and anterior spike times that have an influence on the appearance of a spike at time t for neuron i . The choice of these sets is based on the decomposition (5.56).

First, we choose for any couple (j, s) , independently from anything else, a space-time interaction neighborhood $\mathcal{O}_{(j,s)} \subset I \times]-\infty, s-1]$

$$\mathcal{O}_{(j,s)} = \begin{cases} V_j(k) \times [s-k-1, s-1] & \text{with probability } \lambda_j(k), k \geq 1, \\ \{j\} \times \{s-1\} & \text{with probability } \lambda_j(0), \\ \emptyset & \text{with probability } \lambda_j(-1). \end{cases}$$

Then we put

$$C_1^{(i,t)} = \mathcal{O}_{(i,t)}, \quad C_n^{(i,t)} = \left(\bigcup_{(j,s) \in C_{n-1}^{(i,t)}} \mathcal{O}_{(j,s)} \right) \setminus (C_1^{(i,t)} \cup \dots \cup C_{n-1}^{(i,t)}). \tag{5.59}$$

The process $|C_n^{(i,t)}|$ can be compared to a classical multi-type branching process with reproduction mean

$$m_i = \lambda_i(0) + \sum_{k \geq 1} (k+1) |V_i(k)| \lambda_i(k).$$

Now, condition (2.9) together with the estimates (5.57) and (5.58) shows that $m := \sup_i m_i < 1$. As a consequence,

$$P(N_{STOP}^{(i,t)} > n) \leq E(|C_n^{(i,t)}|) \leq m^n \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and this finishes the proof. □

Proof of Corollary 1 The proof of the first part of corollary 1 is analogous to the proof under the conditions of Theorem 1. We give the proof of the second part of the corollary. We keep the notation of the proof of Theorem 2. We define the projection on the time coordinate

$$\begin{aligned} T(\mathcal{O}_{(j,s)}) &:= s - k - 1 \quad \text{if } \mathcal{O}_{(j,s)} = V_j(k) \times [s - k - 1, s - 1], \quad k \geq 0, \\ T(\mathcal{O}_{(j,s)}) &:= s \quad \text{else.} \end{aligned} \tag{5.60}$$

Define recursively

$$T_1^{(i,t)} = T(\mathcal{O}_{(i,t)}), \quad T_n^{(i,t)} = \min\{T(\mathcal{O}_{(j,s)}) : (j,s) \in C_{n-1}^{(i,t)}\}, \quad n \geq 1,$$

and let finally

$$T_{STOP}^{(i,t)} := T_{N_{STOP}^{(i,t)}-1}^{(i,t)}.$$

Then as in the proof of the first part of the corollary,

$$|E[f(X_s^t(i))1_A] - E[f(X_s^t(i))]P(A)| \leq 2\|f\|_\infty P(A)P(R \leq 1),$$

where

$$R = \inf_{u \in [s,t]} T_{STOP}^{(i,u)}.$$

We have

$$P(R \leq 1) \leq \sum_{u \in [s,t]} P(T_{STOP}^{(i,u)} \leq 1),$$

where

$$P(T_{STOP}^{(i,u)} \leq 1) = P(u - T_{STOP}^{(i,u)} \geq u - 1).$$

In what follows, C will denote a constant that might change from line to line, but that does not depend on β .

We wish to compare $u - T_{STOP}^{(i,u)}$ to the total offspring of a classical Galton-Watson branching process. In order to do so, notice first that using (5.57) and (5.58), we get for any $k \geq 1$,

$$\lambda_i(k) \leq C \frac{e^{-\beta(k-1)}}{1 - e^{-\beta}} =: \bar{\lambda}(k),$$

where we have used (2.11). Moreover,

$$\lambda_i(0) \leq C \frac{e^{-\beta}}{1 - e^{-\beta}} =: \bar{\lambda}(0).$$

It is immediate to see that for β sufficiently large, $\sum_{k \geq 0} \bar{\lambda}(k) < 1$, since

$$\sum_{k \geq 0} \bar{\lambda}(k) = \frac{C}{1 - e^{-\beta}} \left[e^{-\beta} + \frac{e^{-\beta}}{1 - e^{-\beta}} \right] \rightarrow 0$$

as $\beta \rightarrow \infty$. Therefore we can couple $u - T_{STOP}^{(i,u)}$ with the total offspring \mathcal{T} of a classical Galton-Watson process, where each particle has $k + 1$ offspring, $k \geq 0$, with probability $\bar{\lambda}(k)$, and offspring 0 with probability $1 - \sum_{k \geq 0} \bar{\lambda}(k)$. This coupling can be done such that $u - T_{STOP}^{(i,u)} \leq \mathcal{T}$. Thus it is sufficient to evaluate the law of the total offspring \mathcal{T} . For that sake let W_n be a random walk starting from 0 at time 0, with step size distribution η , where

$$\eta = \begin{cases} -1 & \text{with probability } 1 - \sum_{k \geq 0} \bar{\lambda}(k), \\ k & \text{with probability } \bar{\lambda}(k), k \geq 0. \end{cases}$$

Then $\mathcal{T} \stackrel{c}{=} T_{-1} = \inf\{n : W_n = -1\}$. Hence for any $\lambda \in]0, \beta[$, since $u \geq s$,

$$\begin{aligned} P(u - T_{STOP}^{(i,u)} \geq u - 1) &\leq P(T_{-1} \geq u - 1) \leq P(T_{-1} \geq s - 1) = P(T_{-1} > s - 2) \\ &\leq P(e^{\lambda W_{s-2}} \geq 1) \leq \varphi(\lambda)^{s-2}, \end{aligned}$$

where

$$\varphi(\lambda) = E(e^{\lambda \eta}) = e^{-\lambda} \left[1 - \sum_{k \geq 0} \bar{\lambda}(k) \right] + \frac{C e^{-\beta}}{1 - e^{-\beta}} + \frac{C}{1 - e^{-\beta}} \frac{e^{\lambda - \beta}}{1 - e^{\lambda - \beta}}. \tag{5.61}$$

Now, fix $\lambda = 1$, then there exists β_* such that for all $\beta \geq \beta_*$, $\varphi(1) < 1$. Putting $\varrho = \varphi(1)$ and $C = \varrho^{-2}$ yields the desired result. □

6 Proof of Theorem 3

In order to prove Theorem 3, we introduce the sequence of sets

$$\mathcal{V}_{i \rightarrow \cdot}^1 = \{j : W_{i \rightarrow j} = 1\}, \dots, \mathcal{V}_{i \rightarrow \cdot}^n = \{j : \exists k \in \mathcal{V}_{i \rightarrow \cdot}^{n-1} : W_{k \rightarrow j} = 1\}, \quad n \geq 2.$$

Note that $j \in \mathcal{V}_{i \rightarrow \cdot}$ if and only if neuron i has a direct influence on the spiking behavior of neuron j . We put

$$\tau^i = \inf\{n : i \in \mathcal{V}_{i \rightarrow \cdot}^n\}.$$

This is the first time that an information emitted by neuron i can return to neuron i itself.

Recall that $\lambda = 1 + \vartheta/N$ and define $\mu = \frac{N-2}{N}\lambda$. We have the following lower bound.

Proposition 3 *For any k the following inequality holds.*

$$\tilde{P}(\tau^i \leq k) \leq \frac{k-1}{N} \exp\left(\vartheta \frac{k}{N}\right).$$

Proof Put

$$\tilde{\mathcal{V}}_{i \rightarrow \cdot}^n = \{j \neq i : \exists k \in \tilde{\mathcal{V}}_{i \rightarrow \cdot}^{n-1} : W_{k \rightarrow j} = 1\}, \quad n \geq 2, \quad \tilde{\mathcal{V}}_{i \rightarrow \cdot}^1 = \mathcal{V}_{i \rightarrow \cdot}^1.$$

The sequence of sets $\tilde{\mathcal{V}}_{i \rightarrow \cdot}^n, n \geq 1$, equals the original sequence $\mathcal{V}_{i \rightarrow \cdot}^n, n \geq 1$, except that we excluded the choice of i itself. On $\{\tau^i > k\}$, clearly $\bigcup_{n \leq k-1} \mathcal{V}_{i \rightarrow \cdot}^n = \bigcup_{n \leq k-1} \tilde{\mathcal{V}}_{i \rightarrow \cdot}^n$, and we can write

$$\tilde{P}(\tau^i > k) = \tilde{P}\left(W_{j \rightarrow i} = 0 \forall j \in \bigcup_{n \leq k-1} \tilde{\mathcal{V}}_{i \rightarrow \cdot}^n\right).$$

Since in the definition of $\tilde{\mathcal{V}}_{i \rightarrow \cdot}^n$, no choice $W_{\cdot \rightarrow i}$ has been made, we can condition with respect to $\bigcup_{n \leq k-1} \tilde{\mathcal{V}}_{i \rightarrow \cdot}^n$, use the fact that for any $j \in \bigcup_{n \leq k-1} \tilde{\mathcal{V}}_{i \rightarrow \cdot}^n$, the random variable $W_{j \rightarrow i}$ is independent of $\bigcup_{n \leq k-1} \tilde{\mathcal{V}}_{i \rightarrow \cdot}^n$, and obtain the following equality

$$\tilde{P}(\tau^i > k) = \tilde{E}\left[(1 - p_N)^{|\bigcup_{1 \leq n \leq k-1} \tilde{\mathcal{V}}_{i \rightarrow \cdot}^n|}\right].$$

We conclude as follows. We can couple the process $|\tilde{\mathcal{V}}_{i \rightarrow \cdot}^n|, n \geq 2$, with a classical Galton-Watson process $Z_n, n \geq 2$, starting from $Z_1 = \mathcal{V}_{i \rightarrow \cdot}^1$, such that $|\tilde{\mathcal{V}}_{i \rightarrow \cdot}^n| \leq Z_n$ for all $n \geq 2$. The Galton-Watson process has offspring mean $\mu = (N - 2)\frac{\lambda}{N}$. Here, the factor $N - 2$ comes from the fact that any j has $N - 2$ choices of choosing arrows $W_{j \rightarrow \cdot}$, since j itself and i are excluded.

Therefore,

$$\tilde{P}(\tau^i > k) = \tilde{E}\left[(1 - p_N)^{|\bigcup_{1 \leq n \leq k-1} \tilde{\mathcal{V}}_{i \rightarrow \cdot}^n|}\right] \geq \tilde{E}\left[(1 - p_N)^{\sum_{n=1}^{k-1} Z_n}\right].$$

Write $\Sigma_{k-1} = Z_1 + \dots + Z_{k-1}$ and let $\tilde{E}(s^{\Sigma_{k-1}}), s \leq 1$, be its moment generating function. Using the convexity of the moment generating function, we have that

$$\tilde{E}(s^{\Sigma_{k-1}}) \geq 1 + \tilde{E}(\Sigma_{k-1})(s - 1).$$

Using that $\tilde{E}(Z_1) = \frac{N-1}{N}\lambda$ and that the offspring mean equals μ , the claim follows from

$$\tilde{E}(\Sigma_{k-1}) = \frac{N-1}{N}\lambda[1 + \mu + \dots + \mu^{k-2}] \leq \lambda + \dots + \lambda^{k-1} \leq (k-1)\lambda^{k-1},$$

since $\mu \leq \lambda$ and $\lambda \geq 1$. Hence, evaluating the above lower bound in $s = 1 - p_N$, we obtain

$$\tilde{P}(\tau^i > k) \geq 1 - p_N(k-1)\lambda^{k-1},$$

and therefore,

$$\tilde{P}(\tau^i \leq k) \leq p_N(k-1)\lambda^{k-1} = \frac{k-1}{N}\lambda^k,$$

since $p_N = \lambda/N$. Using that $\lambda = 1 + \vartheta/N$, we obtain the assertion. □

In what follows, a_{-k}^{-1} denotes the finite sequence (a_{-k}, \dots, a_{-1}) . In particular, the notation $a_{-l}^{-1}10^{k-1}$ denotes the sequence given by $(a_{-l}, \dots, a_{-1}, 1, 0, \dots, 0)$. We write for short

$$p^{(W,i)}(a|a_{-k}^{-1}) = P^W(X_k(i) = a | X_0^{k-1}(i) = a_{-k}^{-1})$$

for the transition probability of neuron i , given a fixed choice of synaptic weights W . However, conditionings will be read from the left to the right. In particular, we write

$$\begin{aligned} & p^{(W,i)}(a|0^{k-1}1a_{-l}^{-1}) \\ &= P^W(X_k(i) = a | X_{k-1}(i) = \dots = X_1(i) = 0, X_0(i) = 1, X_{-l}^{-1}(i) = a_{-l}^{-1}). \end{aligned}$$

The following proposition shows that on the event $\{\tau^i > k + l\}$, the two transition probabilities $p^{(W,i)}(1|0^{k-1}1a_{-l}^{-1})$ and $p^{(W,i)}(1|0^{k-1}1)$ necessarily coincide.

Proposition 4 For any $k \geq 1, l \geq 1$,

$$\{p^{(W,i)}(1|0^{k-1}1a_{-l}^{-1}) \neq p^{(W,i)}(1|0^{k-1}1)\} \subset \{\tau^i \leq k + l\}.$$

Proof Let W be fixed. From now on, since we will work for this fixed choice of W , we will omit the superscript W and write for short $p^i(a|a_{-k}^{-1})$ instead of $p^{(W,i)}(a|a_{-k}^{-1})$. We have

$$\begin{aligned} & P^W(X_k(i) = 1, X_1^{k-1}(i) = 0^{k-1}, X_0(i) = 1, X_{-l}^{-1}(i) = a_{-l}^{-1}) \\ &= \sum_{j \in \mathcal{V}_i} \sum_{z_0^{k-1}(j) \in \{0,1\}^k} P^W(X_k(i) = 1, X_1^{k-1}(i) = 0^{k-1}, X_0(i) = 1, \\ & \quad X_{-l}^{-1}(i) = a_{-l}^{-1}, X_0^{k-1}(j) = z_0^{k-1}(j), \forall j \in \mathcal{V}_i) \\ &= \sum_{j \in \mathcal{V}_i} \sum_{z_0^{k-1}(j) \in \{0,1\}^k} \phi_i \left(\sum_{j \in \mathcal{V}_i} \sum_{s=0}^{k-1} g_j(k-s)z_s(j) \right) \\ & \quad \times P^W(X_0^{k-1}(j) = z_0^{k-1}(j), \forall j \in \mathcal{V}_i, X_{-l}^{k-1}(i) = a_{-l}^{-1}10^{k-1}). \end{aligned}$$

Thus,

$$\begin{aligned} p^i(1|0^{k-1}1a_{-l}^{-1}) &= \sum_{j \in \mathcal{V}_i} \sum_{z_0^{k-1}(j) \in \{0,1\}^k} \phi_i \left(\sum_{j \in \mathcal{V}_i} \sum_{s=0}^{k-1} g_j(k-s)z_s(j) \right) \\ & \quad \times P^W(X_0^{k-1}(j) = z_0^{k-1}(j), \forall j \in \mathcal{V}_i | X_{-l}^{k-1} = a_{-l}^{-1}10^{k-1}). \end{aligned}$$

The same calculus shows that

$$\begin{aligned} p^i(1|0^{k-1}1) &= \sum_{j \in \mathcal{V}_i} \sum_{z_0^{k-1}(j) \in \{0,1\}^k} \phi_i \left(\sum_{j \in \mathcal{V}_i} \sum_{s=0}^{k-1} g_j(k-s)z_s(j) \right) \\ & \quad \times P^W(X_0^{k-1}(j) = z_0^{k-1}(j), \forall j \in \mathcal{V}_i | X_0^{k-1}(i) = 10^{k-1}). \end{aligned}$$

This shows that in order to ensure that $p^i(1|0^{k-1}1a_{-l}^{-1}) = p^i(1|0^{k-1}1)$, it is sufficient to have

$$\begin{aligned} & P^W(X_0^{k-1}(j) = z_0^{k-1}(j), \forall j \in \mathcal{V}_i | X_0^{k-1}(i) = 10^{k-1}) \\ &= P^W(X_0^{k-1}(j) = z_0^{k-1}(j), \forall j \in \mathcal{V}_i | X_{-l}^{k-1}(i) = a_{-l}^{-1}10^{k-1}), \end{aligned} \tag{6.62}$$

for all possible choices of $z_0^{k-1}(j), j \in \mathcal{V}_i$, which is implied by $\tau^i > k + l$. □

Proof of Theorem 3 In this proof, without loss of generality and to simplify the presentation, we suppose that $\tilde{\Omega}$ is the canonical state space of W . We will use the spontaneous spike times $\{n \in \mathbb{Z} : \xi_n(i) = 1\}$ introduced in the proof of Theorem 1, in Sect. 4 above. We recall that these are independent Bernoulli random variables with $P(\xi_n(i) = 1) = \delta$ for all $i \in \{1, \dots, N\}$, for all $n \in \mathbb{Z}$. Write $l = \sup\{n < S_2^i : \xi_n(i) = 1\}$ and $r = \inf\{n > S_2^i : \xi_n(i) = 1\}$. Put

$$A = \{\tau^i > 2k(N)\},$$

where $k(N)$ is such that $k(N) \rightarrow \infty$ as $N \rightarrow \infty$ and $k(N) \leq N$. We will fix the choice of $k(N)$ later. We have for any realization of $W \in A$,

$$\begin{aligned}
 & E^W [(S_3^i - S_2^i)(S_2^i - S_1^i)] \\
 & \leq E^W [(r - S_2^i)(S_2^i - l)1_{\{l < S_2^i - k(N)\} \cup \{r > S_2^i + k(N)\}}] \\
 & \quad + E^W [(S_3^i - S_2^i)(S_2^i - S_1^i)1_{\{l \geq S_2^i - k(N); r \leq S_2^i + k(N)\}}]. \tag{6.63}
 \end{aligned}$$

Using that conditionally on S_2^i , $r - S_2^i$ and $S_2^i - l$ are independent and geometrically distributed, we obtain a first upper bound

$$E^W [(r - S_2^i)(S_2^i - l)1_{\{l < S_2^i - k(N)\} \cup \{r > S_2^i + k(N)\}}] \leq \frac{1}{\delta^2} (k(N) + 2)(1 - \delta)^{k(N)}. \tag{6.64}$$

We now consider the second term and use that $\tau^i > 2k(N)$. We have

$$\begin{aligned}
 & E^W [(S_3^i - S_2^i)(S_2^i - S_1^i)1_{\{l \geq S_2^i - k(N); r \leq S_2^i + k(N)\}}] \\
 & = \sum_t E^W [(S_3^i - t)(t - S_1^i)1_{\{l \geq t - k(N); r \leq t + k(N)\}}1_{\{S_2^i = t\}}] \\
 & = \sum_t E^W [(t - S_1^i)1_{\{l \geq t - k(N)\}}1_{\{S_2^i = t\}}E^W [(S_3^i - t)1_{\{r \leq t + k(N)\}} | \mathcal{G}_{t-k(N)}^t]], \tag{6.65}
 \end{aligned}$$

where

$$\mathcal{G}_{t-k(N)}^t = \sigma \{X_s(i) : t - k(N) \leq s \leq t\}.$$

Now, since $S_3^i \leq r$,

$$\begin{aligned}
 E^W [(S_3^i - t)1_{\{r \leq t + k(N)\}} | \mathcal{G}_{t-k(N)}^t] & = \sum_{n=1}^{k(N)} n \times P^W (S_3^i - t = n; r \leq t + k(N) | \mathcal{G}_{t-k(N)}^t) \\
 & \leq \sum_{n=1}^{k(N)} n \times P^W (S_3^i - t = n | \mathcal{G}_{t-k(N)}^t).
 \end{aligned}$$

Notice that

$$\begin{aligned}
 & P^W (S_3^i - t = n | \mathcal{G}_{t-k(N)}^t) \\
 & = p^i (0 | 1X_{t-k(N)}^{t-1}) p^i (0 | 01X_{t-k(N)}^{t-1}) \times \dots \times p^i (0 | 0^{n-2}1X_{t-k(N)}^{t-1}) p^i (1 | 0^{n-1}1X_{t-k(N)}^{t-1}).
 \end{aligned}$$

Now we use Proposition 4. Since we are working on $\{\tau^i > 2k(N)\}$, we have

$$p^i (0 | 1X_{t-k(N)}^{t-1}) = p^i (0 | 1), \dots, p^i (1 | 0^{n-1}1X_{t-k(N)}^{t-1}) = p^i (1 | 0^{n-1}1),$$

for all $n \leq k(N)$. Therefore,

$$\begin{aligned}
 & E^W [(S_3^i - t)1_{\{r \leq t + k(N)\}} | \mathcal{G}_{t-k(N)}^t] \\
 & \leq \sum_{n=1}^{k(N)} n \times p^i (0 | 1) p^i (0 | 01) \times \dots \times p^i (0 | 0^{n-2}1) p^i (1 | 0^{n-1}1) \\
 & \leq \sum_{n=1}^{\infty} n \times p^i (0 | 1) p^i (0 | 01) \times \dots \times p^i (0 | 0^{n-2}1) p^i (1 | 0^{n-1}1) \\
 & = E^W (S_3^i - S_2^i). \tag{6.66}
 \end{aligned}$$

We conclude that on A , using successively (6.63)–(6.66),

$$E^W [(S_3^i - S_2^i)(S_2^i - S_1^i)] \leq \frac{1}{\delta^2} (k(N) + 2)(1 - \delta)^{k(N)} + E^W (S_3^i - S_2^i) E^W (S_2^i - S_1^i).$$

In a second step, we are seeking for lower bounds. We start with

$$E^W [(S_3^i - S_2^i)(S_2^i - S_1^i)] \geq E^W [(S_3^i - S_2^i)(S_2^i - S_1^i)1_{\{t \geq S_2^i - k(N); r \leq S_2^i + k(N)\}}]. \tag{6.67}$$

Then on $\{S_2^i = t\}$,

$$\begin{aligned} E^W [(S_3^i - t)1_{\{r \leq t + k(N)\}} | \mathcal{G}_{t-k(N)}^i] &= \sum_{n=1}^{k(N)} n \times P^W (S_3^i - t = n; r \leq t + k(N) | \mathcal{G}_{t-k(N)}^i) \\ &\geq \left(\sum_{n=1}^{k(N)} n \times P^W (S_3^i - t = n | \mathcal{G}_{t-k(N)}^i) \right) - k(N)^2 P^W (r > t + k(N) | \mathcal{G}_{t-k(N)}^i) \\ &= \left(\sum_{n=1}^{k(N)} n \times P^W (S_3^i - t = n | \mathcal{G}_{t-k(N)}^i) \right) - k(N)^2 (1 - \delta)^{k(N)}. \end{aligned}$$

Now, on $\{S_2^i = t\}$,

$$\begin{aligned} &\sum_{n=1}^{k(N)} n \times P^W (S_3^i - t = n | \mathcal{G}_{t-k(N)}^i) \\ &= E^W (S_3^i - S_2^i; S_3^i - S_2^i \leq k(N)) = E^W (S_3^i - S_2^i) - E^W (S_3^i - S_2^i; S_3^i - S_2^i > k(N)) \\ &\geq E^W (S_3^i - S_2^i) - E^W (r - S_2^i; r - S_2^i > k(N)) \\ &\geq E^W (S_3^i - S_2^i) - \frac{1}{\delta} (k(N) + 2) (1 - \delta)^{k(N)}. \end{aligned}$$

Therefore, for any realization $W \in A$,

$$\begin{aligned} E^W [(S_3^i - S_2^i)(S_2^i - S_1^i)] &\geq E^W (S_3^i - S_2^i) E^W (S_2^i - S_1^i) \\ &\quad - \left[\frac{2}{\delta^2} (k(N) + 2) + k(N)^2 \right] (1 - \delta)^{k(N)}. \end{aligned}$$

Putting things together and supposing that $k(N) + 2 \leq k(N)^2$, we obtain finally

$$|E^W [(S_3^i - S_2^i)(S_2^i - S_1^i)] - E^W (S_3^i - S_2^i) E^W (S_2^i - S_1^i)| \leq \frac{3}{\delta^2} k(N)^2 (1 - \delta)^{k(N)}.$$

It remains to find an upper bound for $\mathbb{P}(A^c)$. Clearly, applying Proposition 3, since $k(N) \leq N$, we have

$$\mathbb{P}(A^c) \leq e^{2\vartheta} \frac{k(N)}{N}.$$

It is enough to choose $k(N) = \sqrt{N}$ to conclude the proof. □

Acknowledgements We thank D. Brillinger, B. Cessac, S. Ditlevsen, M. Jara, M. Kelbert, Y. Kohayakawa, C. Landim, R.I. Oliveira, S. Ribeiro, L. Triolo, C. Vargas and N. Vasconcelos for many discussions on Hawkes processes, random graphs and neural nets at the beginning of this project.

This work is part of USP project “Mathematics, computation, language and the brain”, FAPESP project “NeuroMat” (grant 2011/51350-6), USP/COFECUB project “Stochastic systems with interactions of variable range” and CNPq project “Stochastic modeling of the brain activity” (grant 480108/2012-9). A.G. is partially supported by a CNPq fellowship (grant 309501/2011-3), A.G. and E.L. have been partially supported by the MathAmSud project “Stochastic structures of large interacting systems” (grant 009/10). E.L. thanks Numec, USP, for hospitality and support.

References

1. Berbee, H., Chains with infinite connections: uniqueness and Markov representation. *Probab. Theory Relat. Fields* **76**, 243–253 (1987)
2. Bollobás, B.: *Random Graphs*, 2nd edn. Cambridge Studies in Advanced Mathematics, vol. 73. Cambridge University Press, Cambridge (2001)
3. Brémaud, P., Massoulié, L.: Stability of nonlinear Hawkes processes. *Ann. Probab.* **24**(3), 1563–1588 (1996)
4. Bressaud, X., Fernández, R., Galves, A.: Decay of correlations for non-Hölderian dynamics. A coupling approach. *Elect. J. Probab.* **4**, 1–19 (1999)
5. Brillinger, D.: Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol. Cybern.* **59**(3), 189–200 (1988)
6. Cessac, B.: A discrete time neural network model with spiking neurons: II: Dynamics with noise. *J. Math. Biol.* **62**, 863–900 (2011)
7. Comets, F., Fernández, R., Ferrari, P.A.: Processes with long memory: regenerative construction and perfect simulation. *Ann. Appl. Probab.* **12**(3), 921–943 (2002)
8. Doebelin, W., Fortet, R.: Sur les chaînes à liaisons complètes. *Bull. Soc. Math. Fr.* **65**, 132–148 (1937)
9. Fernández, R., Maillard, G.: Chains with complete connections and one-dimensional Gibbs measures. *Electron. J. Probab.* **9**(6), 145–176 (2004)
10. Fernández, R., Ferrari, P.A., Galves, A.: Coupling, renewal and perfect simulation of chains of infinite order. Notes for a minicourse given at the Vth Brazilian School of Probability. Manuscript (2001). <http://www.univ-rouen.fr/LMRS/Persopage/Fernandez/resucoup.html>
11. Fernández, R., Ferrari, P.A., Garcia, N.L.: Loss network representation of Peierls contours. *Ann. Probab.* **29**(2), 902–937 (2001)
12. Fernández, R., Ferrari, P.A., Garcia, N.L.: Perfect simulation for interacting point processes, loss networks and Ising models. *Stoch. Process. Appl.* **102**(1), 63–88 (2002)
13. Ferrari, P.A., Maass, A., Martínez S., Ney, P.: Cesàro mean distribution of group automata starting from measures with summable decay. *Ergod. Theory Dyn. Syst.* **20**(6), 1657–1670 (2000)
14. Gallo, S.: Chains with unbounded variable length memory: perfect simulation and visible regeneration scheme. *Adv. Appl. Probab.* **43**, 735–759 (2011)
15. Galves, A., Garcia, N., Löcherbach, E., Orlandi, E.: Kalikow-type decomposition for multicolor infinite range particle systems (2013)
16. Gerstner, W., Kistler, W.M.: *Spiking Neuron Models. Single Neurons, Populations, Plasticity*. Cambridge University Press, Cambridge (2002)
17. Goldberg, J.M., Adrian, H.O., Smith, F.D.: Response of neurons of the superior olivary complex of the cat to acoustic stimuli of long duration. *J. Neurophysiol.* **27**, 706–749 (1964)
18. Harris, T.E.: On chains of infinite order. *Pac. J. Math.* **5**, 707–724 (1955)
19. Hawkes, A.G.: Point spectra of some mutually exciting point processes. *J. R. Stat. Soc. B* **33**, 438–443 (1971)
20. Johansson, A., Öberg, A.: Square summability of variations of g -functions and uniqueness of g -measures. *Math. Res. Lett.* **10**(5–6), 587–601 (2003)
21. Krumin, M., Reutsky, I., Shoham, S.: Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input. *Front Comput Neurosci.* **4**, 147 (2010). doi:10.3389/fncom.2010.00147
22. Möller, J., Rasmussen, J.G.: Perfect simulation of Hawkes processes. *Adv. Appl. Probab.* **37**(3), 629–646 (2005)
23. Nawrot, M.P., Bousein, C., Rodriguez-Molina, V., Aertsen, A., Grün, S., Rotter, S.: Serial interval statistics of spontaneous activity in cortical neurons in vivo and in vitro. *Neurocomputing* **70**, 1717–1722 (2007)
24. Rissanen, J.: A universal data compression system. *IEEE Trans. Inf. Theory* **29**(5), 656–664 (1983)
25. Spitzer, F.: Interaction of Markov processes. *Adv. Math.* **5**(2), 246–290 (1970)