Adv. Appl. Prob. **49,** 84–107 (2017)
*doi:*10.1017/apr.2016.80

HAWKES PROCESSES WITH VARIABLE LENGTH MEMORY AND AN INFINITE NUMBER OF COMPONENTS

PIERRE HODARA * ** AND
EVA LÖCHERBACH,* *Université de Cergy-Pontoise*

Abstract

In this paper we propose a model for biological neural nets where the activity of the network is described by Hawkes processes having a variable length memory. The particularity in this paper is that we deal with an infinite number of components. We propose a graphical construction of the process and build, by means of a perfect simulation algorithm, a stationary version of the process. To implement this algorithm, we make use of a Kalikow-type decomposition technique. Two models are described in this paper. In the first model, we associate to each edge of the interaction graph a saturation threshold that controls the influence of a neuron on another. In the second model, we impose a structure on the interaction graph leading to a cascade of spike trains. Such structures, where neurons are divided into layers, can be found in the retina.

Keywords: Point process; multivariate Hawkes process; biological neural net; Kalikow-type decomposition; perfect simulation

2010 Mathematics Subject Classification: Primary 60G55
Secondary 60K35; 90B15

1. Introduction and main results

1.1. Motivation

The aim of this paper is to present a model in continuous time for an infinite system of interacting neurons. Each neuron is represented by its spike train, i.e. the series of events of spiking over time. Since neural activity is continuously recorded in time, a time continuous description is natural. The model considered in this paper is an extension to the continuous-time framework of a model which was recently introduced by Galves and Löcherbach [10] in discrete time.

We consider a countable set of neurons I. The activity of each neuron $i \in I$ is described by a counting process Z^i where, for any $-\infty < s < t < \infty$, $Z^i((s,t])$ records the number of spikes of neuron i during the interval $(s,t]$. Under suitable assumptions, the sequence of counting processes $(Z^i,\ i \in I)$ is characterized by its intensity process $(\lambda_t^i,\ i \in I)$ which is defined through the relation

$$\mathbb{P}(Z^i \text{ has a jump in } (t, t+\mathrm{d}t] \mid \mathcal{F}_t) = \lambda_t^i \,\mathrm{d}t, \qquad i \in I.$$

Here, $\mathcal{F}_t$ is the sigma-field generated by $Z^i((s,u]), s \le u \le t,\ i \in I$.

Received 18 September 2015; revision received 27 May 2016.
* Postal address: CNRS UMR 8088, Département de Mathématiques, Université de Cergy-Pontoise, 2 avenue Adolphe Chauvin, 95302 Cergy-Pontoise Cedex, France.
** Email address: pierre.hodara@u-cergy.fr

Our main motivation is to model neural nets. This naturally leads to the following choice of intensity processes:

$$\lambda_t^i = \psi_i\left(\sum_{j\in I} h_{j\to i}\left(\int_{[L_t^i,t)} g_j(t-s)\,\mathrm{d}Z_s^j\right)\right), \tag{1}$$

where $\psi_i\colon \mathbb{R}\to\mathbb{R}_+$ is the *spiking rate function*, $\{h_{j\to i}\colon \mathbb{R}_+\to\mathbb{R}, i, j\in I\}$ a family of *synaptic weight functions* modeling the influence of neuron j on neuron i, $g_j\colon \mathbb{R}_+\to\mathbb{R}_+$ a nonincreasing *decay function*, and

$$L_t^i = \sup\{s<t\colon Z^i([s])>0\} \tag{2}$$

the last spiking time before time t of neuron i (with the convention $[s]:=[s,s]$).

The form (1) of our intensity process is close to the typical form of the intensity of a multivariate nonlinear Hawkes process. The original papers of Hawkes [12] and Hawkes and Oakes [13], introducing the model, dealt with linear intensity functions. Extensions to the nonlinear case were considered by Brémaud and Massoulié [1]; see also Massoulié [16], who proposed a study of the stability properties of multivariate nonlinear Hawkes processes. Hawkes processes have shown to be important in various fields of applications. To name just a few, Hansen *et al.* [11] and Chevallier [2] are excellent references proving the use of Hawkes processes as models of spike trains in neuroscience. Reynaud-Bouret and Schbath [18] dealt with an application to genome analysis. In a completely different context, Jaisson and Rosenbaum [14] obtained limit theorems for nearly unstable Hawkes processes in view of applications in financial price modeling. For a general introduction to Hawkes processes and their basic properties we refer the reader to Daley and Vere-Jones [4].

Our form of the intensity (1) differs from the classical Hawkes setting by its *variable memory structure* introduced through the term L_t^i. Hence, the spiking intensity of a neuron only depends on its history up to its last spike time, which is a biologically very plausible assumption on the memory structure of the process. Therefore, our model can be seen as a nonlinear multivariate Hawkes process where the number of components is infinite with a variable memory structure. The interactions between neurons are encoded through the synaptic weight functions $h_{j\to i}$ that we are going to specify below.

1.2. The setting

We work on a filtered measurable space $(\Omega,\mathcal{A},\mathbb{F})$ which we define as follows. We write $\mathbb{M}$ for the canonical path space of simple point processes given by

$$\mathbb{M} := \Big\{\mathtt{m}=(t_n)_{n\in\mathbb{Z}}\colon t_0\le 0<t_1,\ t_n\le t_{n+1},\ t_n<t_{n+1} \text{ if } t_n<+\infty \text{ or } t_{n+1}>-\infty;\ \lim_{n\to+\infty} t_n=+\infty,\ \lim_{n\to-\infty} t_n=-\infty\Big\}.$$

For any $\mathtt{m}\in\mathbb{M}$, any $n\in\mathbb{Z}$, let $T_n(\mathtt{m})=t_n$. We identify $\mathtt{m}\in\mathbb{M}$ with the associated point measure $\mu=\sum_n \delta_{T_n(\mathtt{m})}$ and set $\mathcal{M}_t := \sigma\{\mu(A)\colon A\in\mathcal{B}(\mathbb{R}), A\subset(-\infty,t]\}$, $\mathcal{M}=\mathcal{M}_\infty$. We will also systematically identify $\mathtt{m}$ with the associated counting process $\alpha(\mathtt{m})$, defined by $\alpha_0(\mathtt{m})=0$,

$$\alpha_t(\mathtt{m}) = \begin{cases}\mu((0,t]) & \text{if } t\ge 0,\\ -\mu((t,0]) & \text{if } t\le 0.\end{cases}$$

Finally, we set $(\Omega,\mathcal{A},\mathbb{F}) := (\mathbb{M},\mathcal{M},(\mathcal{M}_t)_{t\in\mathbb{R}})^I$. We write $(Z^i,\ i\in I)$ for the canonical multivariate point measure defined on Ω.

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

We specify the following parameters: a family of firing rate functions $\psi = \{\psi_i : \mathbb{R} \to \mathbb{R}_+,\ i \in I\}$, a family of synaptic weight functions $h = \{h_{j\to i} : \mathbb{R} \to \mathbb{R},\ i, j \in I\}$, a family of functions $g = \{g_j : \mathbb{R}_+ \to \mathbb{R}_+,\ j \in I\}$, which are nonincreasing. Recall the definition of the last spiking time of neuron i before time t, given in (2).

Definition 1. A Hawkes process with variable length memory and an infinite number of interacting components with parameters (ψ, h, g) is a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A}, \mathbb{F})$ such that

- $\mathbb{P}$-almost surely, for all $i \neq j$, Z^i and Z^j never jump simultaneously,
- for all $i \in I$, the compensator of Z^i is given by $\nu^i(\mathrm{d}t) = \lambda_t^i \,\mathrm{d}t$, where

$$\lambda_t^i = \psi_i\left(\sum_{j\in I} h_{j\to i}\left(\int_{[L_t^i,t)} g_j(t-s)\,\mathrm{d}Z_s^j\right)\right).$$

2. Main results

In the case where I is a finite set, under suitable assumptions on the parameters of the process, the existence and construction of $(Z^i,\ i \in I)$ is standard (see [1] and [5]). In our case, however, the number of interacting components defining the process is infinite. In such a framework, Delattre *et al.* [5] proved pathwise existence and uniqueness of the processes, however without giving an explicit construction of the process. In this paper we show that – under suitable assumptions – a *graphical construction* is possible. This graphical construction does not only imply the existence but also the possibility of a perfect simulation of a stationary version of the process (i.e. a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A}, \mathbb{F})$, such that under $\mathbb{P}$, for all $u \in \mathbb{R}$ and all $i \in I$, the processes $Z^i(]u, u+t])$, $t \in \mathbb{R}$ are stationary). These results are achieved via a *Kalikow-type decomposition* in two types of models. The first model is a system containing a *saturation threshold* for any directed edge $i \to j$ in the interaction graph defined by the synaptic weights. The second model deals with a *cascade of spike trains*.

Kalikow-type decompositions are now largely used in the literature for perfect simulation issues and similar scopes. They were considered first by Ferrari *et al.* [8] and Comets *et al.* [3]. This type of technique was then studied in a series of papers for perfect simulation issues. See Galves and Löcherbach [10] for an application in the context of neural biological nets in discrete time. The decomposition that we use in this paper is a nontrivial extension of the previous considerations to the framework of continuous-time neural nets. In the case of our first model, it has to be achieved in a random environment.

To the best of the authors' knowledge, the perfect simulation algorithm constructed in this paper is the first result in this direction obtained for Hawkes processes with nonlinear intensity functions. The well known work of Møller and Rasmussen [17] on perfect simulation of Hawkes processes dealt with linear intensity functions and exploited very heavily the underling branching structure. The precise form of our perfect simulation algorithm is given in Section 6.2.

2.1. Assumptions and notation

Throughout this paper we suppose that the firing rate functions $\psi_i : \mathbb{R} \to \mathbb{R}_+$ are nondecreasing and bounded by a real number Λ_i. Introducing

$$\phi_i = \frac{\psi_i}{\Lambda_i}, \tag{3}$$

we make the following assumption.

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

Assumption 1. *The family of functions $\{\phi_i\}_{i\in I}$, is equicontinuous, i.e. there exists a modulus of continuity $\omega\colon \mathbb{R}_+ \to \mathbb{R}_+$, satisfying $\omega(0) = 0$ and $\lim_{x\to 0}\omega(x) = 0$, such that for all $x, x' \in \mathbb{R}$, $i \in I$,*

$$|\phi_i(x) - \phi_i(x')| \le \omega(|x - x'|). \tag{4}$$

Note that we can always choose the function ω to be subadditive.

The interactions between neurons are coded via the synaptic weight functions $h_{j\to i}$. For each neuron i, we define

$$\mathcal{V}_{i\to\cdot} = \{j \in I,\ j \neq i\colon h_{i\to j} \neq 0\},$$

where 0 denotes the constant function 0. As a consequence, $\mathcal{V}_{i\to\cdot}$ is the set of all neurons that are directly influenced by neuron i. In the same spirit, we set

$$\mathcal{V}_{\cdot\to i} = \{j \in I,\ j \neq i\colon h_{j\to i} \neq 0\}, \tag{5}$$

which is the set of neurons that have a direct influence on i. These sets may be finite or infinite.

In the following we introduce the two main types of models that we consider, firstly *models with saturation thresholds* and secondly *models with a cascade of spike trains*.

3. Models with saturation threshold

3.1. Models with saturation thresholds

We suppose that to each directed edge $j \to i$ is associated a saturation threshold $K_{j\to i} > 0$ representing the maximal number of spikes that the synapse $j \to i$ is able to support. We suppose that

$$h_{j\to i}(x) = W_{j\to i}(x \wedge K_{j\to i}), \tag{6}$$

where $W_{j\to i} \in \mathbb{R}$ is called the *synaptic weight* of neuron j on i. Moreover, we suppose that $g_j \equiv 1$ for all $j \in I$ and write for short $g = \mathbf{1}$. Hence, we can write

$$\lambda_t^i = \psi_i\left(\sum_{j\in I} W_{j\to i}(Z^j((L_t^i, t)) \wedge K_{j\to i})\right).$$

Introduce, for any $i \in I$, a nondecreasing sequence $(V_i(k))_{k\ge 0}$ of finite subsets of I such that $V_i(0) = \varnothing$, $V_i(1) = \{i\}$, $V_i(k) \subset V_i(k+1)$, $V_i(k) \neq V_i(k+1)$ if $V_i(k) \neq \mathcal{V}_{\cdot\to i} \cup \{i\}$, and $\bigcup_k V_i(k) = \mathcal{V}_{\cdot\to i} \cup \{i\}$ (recall (5)).

We define, for all $k \ge 0$, $\partial V_i(k) := V_i(k+1) \setminus V_i(k)$, the border of the set $V_i(k)$ and make the following assumption.

Assumption 2. *For all $i \in I$, $W_{i\to i} = 0$, and*

$$\sup_{i\in I}\sum_j |W_{j\to i}|K_{j\to i} < +\infty. \tag{7}$$

The following theorem states that if any neuron has a sufficiently high spontaneous firing activity, then a unique stationary version of the Hawkes process with saturation threshold exists.

Theorem 1. *Grant Assumptions 1 and 2 and suppose that, for all $i \in I$,*

$$\psi_i \ge \delta_i \quad \textit{for some } \delta_i > 0. \tag{8}$$

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

Suppose, moreover, that

$$\sup_{i\in I}\left(\sum_{k\geq 1}\left[\left(\sum_{j\in V_i(k)}\frac{\Lambda_j-\delta_j}{\delta_i}\right)\bar{\mu}_i(k)\right]\right)<1, \tag{9}$$

where

$$\bar{\mu}_i(k):=\omega\left(\sum_{j\in\partial V_i(k-1)}|W_{j\to i}|K_{j\to i}\,\mathbf{1}_{\{W_{j\to i}<0\}}\right)+\omega\left(\sum_{j\in\partial V_i(k-1)}|W_{j\to i}|K_{j\to i}\,\mathbf{1}_{\{W_{j\to i}>0\}}\right), \tag{10}$$

with the indicator denoted by $\mathbf{1}_{\{\cdot\}}$. *Then there exists a unique probability measure* $\mathbb{P}$ *on* $(\Omega,\mathcal{A})$ *such that under* $\mathbb{P}$, *the canonical process* $(Z^i,\ i\in I)$ *on* Ω *is a stationary nonlinear Hawkes process with variable length memory and an infinite number of interacting components, with parameters* (ψ,h,g), *where* h *is given by (6) and* $g=\mathbf{1}$.

Remark 1. (i) By subadditivity of ω, the condition

$$\sup_{i\in I}\left(\sum_{k\geq 1}\left[\left(\sum_{j\in V_i(k)}\frac{\Lambda_j-\delta_j}{\delta_i}\right)\left(\sum_{j\in\partial V_i(k-1)}\omega(|W_{j\to i}|K_{j\to i})\right)\right]\right)<1$$

implies condition (9).

(ii) The following additional assumptions on the parameters allow us to work with a condition simpler than (9). Suppose that $\delta_i=\delta>0$ for all $i\in I$ and that a stronger summability than (7) holds, i.e.

$$\sup_i\Lambda_i\sum_{k\geq 1}|V_i(k)|\sum_{j\in\partial V_i(k-1)}\omega(|W_{j\to i}|K_{j\to i})<\infty. \tag{11}$$

By the subadditivity of ω, condition (11) implies (9) for sufficiently large δ.

(iii) In the above model, there is a saturation threshold for each directed edge between two neurons. A different model would be a system where the saturation threshold concerns only the global input received by each neuron. This would amount to considering the following intensity:

$$\lambda_t^i=\psi_i\left(K_i^-\vee\left(\left(\sum_{j\in I}W_{j\to i}Z^j((L_t^i,t))\right)\wedge K_i^+\right)\right),$$

where K_i^- and K_i^+ are global saturation thresholds, respectively, for inhibition and stimulation. Under obvious changes of condition (9), Theorem 1 continues to hold in this framework.

(iv) In [16], the author obtained similar results in a slightly different context dealing with truly infinite memory models. The fact that we consider variable length memory Hawkes processes constitutes, of course, the main difference with respect to [16]. As a matter of fact, due to this difference, instead of (8) and (9), [16] required that $\sum_{i\in I}\sum_{j\in I}W_{j\to i}K_{j\to i}<\infty$, which is a stronger assumption than (7). Finally, we stress that using our approach we obtain more than the existence of a stationary solution alone. It actually gives a graphical construction of the stationary measure, which is not the case in [16].

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

3.2. A Markovian description in terms of a coupled 'house of cards' process

As in the case of classical Hawkes processes with exponential loss functions, under suitable assumptions, an alternative description of $(Z^i,\ i \in I)$ via its intensity processes yields a Markovian description of the process. Throughout this subsection, we impose the summability assumption (7). Moreover, we suppose that

$$\sup_i \Lambda_i |\mathcal{V}|_{i\to\cdot} < \infty, \qquad \inf_i \delta_i = \delta > 0 \tag{12}$$

and that

$$\mathcal{V}_i := \mathcal{V}_{\cdot\to i} \cup \mathcal{V}_{i\to\cdot} \quad \text{are finite sets for all } i \in I. \tag{13}$$

Denote by $\mathcal{V}_i$ the set of neurons that directly influence neuron i or that are directly influenced by it, i.e. the set of neurons j such that $W_{j\to i} \neq 0$ or $W_{i\to j} \neq 0$.

It is convenient to adopt a description of a process living on the set of directed edges $\mathcal{E} = \{j \to i,\ j \in \mathcal{V}_{\cdot\to i},\ i \in I\}$. We write $e = j \to i \in \mathcal{E}$ for a directed edge and introduce, for any such e, the process $U_t(e)$ defined by

$$U_t(e) = Z^j((L_t^i, t)), \qquad t \in \mathbb{R}.$$

With this point of view, the neural network is described by the process $(U_t(e),\ e \in \mathcal{E})_{t\in\mathbb{R}}$, taking values in $S := \mathbb{N}^{\mathcal{E}}$. Its dynamic is described by its generator defined by

$$\mathcal{G}f(\eta) = \sum_{i\in I} \psi_i\Big(\sum_j W_{j\to i}(\eta(j\to i) \wedge K_{j\to i})\Big)[f(\eta + \Delta_i\eta) - f(\eta)],$$

where

$$(\Delta_i\eta)(k\to l) = \begin{cases} -\eta(k\to l) & \text{if } l = i,\ k \in \mathcal{V}_{\cdot\to i}, \\ 1 & \text{if } k = i,\ l \in \mathcal{V}_{i\to\cdot}, \\ 0 & \text{otherwise,} \end{cases}$$

and where $f \in \mathcal{D}(\mathcal{G}) = \{f\colon |||f||| := \sum_{e\in\mathcal{E}} \Delta_f(e) < \infty\}$ with $\Delta_f(e) = \sup\{|f(\eta) - f(\zeta)|\colon \eta, \zeta \in S,\ \eta(e') = \zeta(e') \text{ for all } e' \neq e\}$.

Remark 2. (i) Note that a spike of neuron i does not only affect all neurons $j \in \mathcal{V}_{i\to\cdot}$, which receive an additional potential, but also all neurons $j \in \mathcal{V}_{\cdot\to i}$, since all $\eta(j\to i)$ are reset to 0 when a spike of neuron i occurs. It is for this reason that we call the above process a coupled 'house of cards' process.

(ii) We could also work with $\tilde{S} := \{\eta \in S\colon \eta(j\to i) \le K_{j\to i} \text{ for all } e = (j\to i) \in \mathcal{E}\}$, which is the state space of relevant configurations of the process. This would imply us redefining $U_t(j\to i) := Z^j((L_t^i, t)) \wedge K_{j\to i}$.

We introduce $T_i := \{e = j_1 \to j_2 \in \mathcal{E}\colon j_1 = i \text{ or } j_2 = i\}$.

By Theorem 3.9 of Liggett [15, Chapter 1], (11), together with (12) and (13), implies that $\mathcal{G}$ is the generator of a Feller process $(U_t(e),\ e \in \mathcal{E})_{t\in\mathbb{R}}$ on S.

Proof. Under the above conditions, the generator $\mathcal{G}$ can be written as

$$\mathcal{G}f(\eta) = \sum_i c_{T_i}(\eta, \mathrm{d}\zeta)[f(\eta_i^\zeta) - f(\eta)],$$

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

where

$$\eta_i^{\zeta}(e) = \begin{cases} \eta(e) & \text{if } e \notin T_i, \\ \zeta(e) & \text{if } e \in T_i, \end{cases}$$

and where

$$c_{T_i}(\eta, \mathrm{d}\zeta) = \psi_i\Big(\sum_j W_{j\to i}(\eta(j \to i) \wedge K_{j\to i})\Big)\delta_{\eta+\Delta_i\eta}(\mathrm{d}\zeta).$$

A straightforward calculation shows that the quantity $c_{T_i} = \sup_\eta c_{T_i}(\eta, S)$ defined by Equation (3.3) in Liggett [15, Chapter 1] can be upper bounded by $c_{T_i} \le \Lambda_i$ and that

$$c_{T_i}(e) := \sup\{\|c_{T_i}(\eta, \cdot) - c_{T_i}(\zeta, \cdot)\|_{\mathrm{TV}} \colon \eta(e') = \zeta(e') \text{ for all } e' \neq e\},$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation distance, can be controlled by

$$c_{T_i}(e) \le \begin{cases} \Lambda_i \omega(|W_e|K_e) & \text{if } e = k \to i, \\ \Lambda_i & \text{if } e = i \to l, \\ 0 & \text{otherwise.} \end{cases}$$

Condition (11) together with (12) and (13) thus imply that M defined in Equation (3.8) of Liggett [15, Chapter 1] can be controlled as

$$M = \sup_e \sum_{i\colon e\in T_i} \sum_{u\neq e} c_{T_i}(u) \le 2 \sup_i \Lambda_i \Big(\sum_k \omega(|W_{k\to i}|K_{k\to i}) + |\mathcal{V}|_{i\to}\Big) < \infty,$$

and then Theorem 3.9 of Liggett [15, Chapter 1] allows us to conclude. □

As a consequence, we can reformulate Theorem 1 in the following way.

Theorem 2. *Grant the assumptions of Theorem 1 and suppose, moreover, that (7) together with (11), (12) and (13) are satisfied. Then the process* $(U_t^e,\ e \in \mathcal{E})_{t\in\mathbb{R}}$ *is ergodic.*

Proof. The proof follows immediately from Theorem 1. □

Remark 3. We compare the above result to the $M < \varepsilon$-criterion of Theorem 4.1 of Liggett [15, Chapter 1]. The quantity M has already been introduced above. Moreover, ε is defined by

$$\varepsilon := \inf_{e\in\mathcal{E}} \inf_{\{\eta_1,\eta_2\in S\colon\, \eta_1(e')=\eta_2(e') \text{ for all } e'\neq e\}} \sum_{\{k\colon\, e\in T_k\}} [c_{T_k}(\eta_1, \{\zeta : \zeta(e) = \eta_2(e)\}) + c_{T_k}(\eta_2, \{\zeta : \zeta(e) = \eta_1(e)\})].$$

The sufficient condition for ergodicity in Theorem 4.1 of Liggett [15] is $M < \varepsilon$. Note that in our particular case we have $\varepsilon = 0$ which can easily be seen by considering η_1 and η_2 with $\eta_1(e) = 1$ and $\eta_2(e) = 3$ for some $e \in \mathcal{E}$. Consequently, the sufficient condition $M < \varepsilon$ is not satisfied. However, Theorem 2 implies the ergodicity of the process without the condition $M < \varepsilon$.

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

4. A cascade of spike trains

We now describe the second model that we consider in this paper, a *cascade of spike trains*. We suppose that for each $j \in I$, the function $g_j : \mathbb{R}_+ \to \mathbb{R}_+$ is measurable and nonincreasing such that $\int_0^{+\infty} g_j(x)\,\mathrm{d}x < +\infty$. The function g_j models a leak phenomenon. We suppose, moreover, that

$$h_{j\to i}(x) = W_{j\to i} \cdot x \tag{14}$$

for a family of *synaptic weights* $\{W_{j\to i} \in \mathbb{R}\}$ satisfying the summability condition

$$\sup_{i\in I} \sum_j |W_{j\to i}| < \infty. \tag{15}$$

Neurons j such that $W_{j\to i} > 0$ are called *excitatory* for i, if $W_{j\to i} < 0$, then j is called *inhibitory* for i. Finally, we impose the following condition on the structure of interactions.

Assumption 3. *The set I of the neurons is divided into layers $(I_n)_{n\in\mathbb{Z}}$ such that we have the partition $I = \bigsqcup_{n\in\mathbb{Z}} I_n$. For each $n \in \mathbb{Z}$ and for each $i \in I_n$, we suppose that*

$$\mathcal{V}_{\cdot\to i} \subset I_{n-1}. \tag{16}$$

Therefore, a neuron only receives information from neurons in the layer just above itself. This assumption does not apply to the brain's structure, which is far too complicated. But such a structure can be found in simpler nervous tissues such as the retina. The nonlinear Hawkes process that we consider in this section is defined by its intensity given by

$$\lambda_t^i = \psi_i\Big(\sum_{j\in I} W_{j\to i} \int_{[L_t^i,t)} g_j(t-s)\,\mathrm{d}Z_s^j\Big), \tag{17}$$

together with the assumption (16) on the structure of the interactions.

In order to state our result for this model, we need to strengthen Assumption 1.

Assumption 4. *The family $(\phi_i)_{i\in I}$ is uniformly Lipschitz continuous, i.e. there exists $\gamma > 0$ such that, for all $x, x' \in \mathbb{R}$, $i \in I$,*

$$|\phi_i(x) - \phi_i(x')| \le \gamma |x - x'|. \tag{18}$$

Theorem 3. *Grant Assumptions 3 and 4 and suppose, moreover, that condition (15) is satisfied. If*

$$\begin{aligned}\sup_{i\in I}\Bigg(\sum_{k\ge 1}\Bigg[\Bigg(k\Bigg(\sum_{j\in V_i(k)} \Lambda_j\Bigg)+1\Bigg) &\times \Bigg(\sum_{j\in V_i(k-1)} |W_{j\to i}|\Lambda_j \int_{k-1}^{k} g_j(s)\,\mathrm{d}s + \sum_{j\in \partial V_i(k-1)} |W_{j\to i}|\Lambda_j \int_0^k g_j(s)\,\mathrm{d}s\Bigg)\Bigg]\Bigg)\\ &< \frac{1}{\gamma},\end{aligned} \tag{19}$$

where γ is given in Assumption 4, then there exists a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A})$ such that under $\mathbb{P}$, the canonical process $(Z^i,\ i \in I)$ on Ω is a stationary nonlinear Hawkes process with variable length memory and an infinite number of interacting components, with parameters (ψ, h, g), where h is given by (14).

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

Remark 4. (i) A similar result still holds imposing only the weaker Assumption 1 instead of Assumption 4. However, in this case, the summability condition (19) has to be restated in terms of the modulus of continuity of the ϕ_i and turns out to be much more complicated than (19) above. In particular, it cannot be expressed directly in terms of the parameters of the model. See Remark 8 following the proof of Theorem 3 for more details.

(ii) We once more compare this result with Theorem 3 of [16]. Instead of our conditions (19) and Assumption 3 on the structure of interactions, [16] requires that

$$\sum_{i\in I}\sum_{j\in I} W_{j\to i}\int_0^{+\infty} t g_j(t)\,\mathrm{d}t < \infty.$$

A direct comparison with (19) is difficult, but (19) seems to be slightly less restrictive.

(iii) We can restate our result in the framework of one-dimensional Hawkes processes with the following assumptions: $I = \{i\}$, $W_{i\to i} \neq 0$, $\Lambda_i = \Lambda \in \mathbb{R}_+$ and allowing for a time dependence where L_t^i is replaced by $-\infty$ for all $t \in \mathbb{R}$. In this framework our condition (19) can be expressed as

$$\sum_{k\geq 1}\left[(k\Lambda + 1)\int_{k-1}^{k} g(s)\,\mathrm{d}s\right] < \frac{1}{\Lambda\gamma}.$$

This assumption is stronger than the following condition of Equation (4) in [1]:

$$\int_0^{\infty} s g(s)\,\mathrm{d}s < +\infty,$$

since Equation (4) of [1] does not impose a condition on the value of the L^1-norm of g and on the Lipschitz constant γ. However, our method gives a graphical construction of the stationary measure by the mean of a perfect simulation algorithm. To the best of the authors' knowledge this has been obtained in the literature only for linear Hawkes processes as in [17], whereas our process has a nonlinear intensity.

We give an example in which condition (19) is satisfied in order to illustrate our result. For the sake of simplicity, we will assume that, for all j, $\Lambda_j = \Lambda$ and $g_j = g$. We will also choose a simple structure for the network; each layer of neurons is identical and indexed by $\mathbb{Z}$ so that we can take $I = \mathbb{Z}^2$ with indexes $i = (i_1, i_2) \in \mathbb{Z}^2$, where i_1 denotes the 'site' of the neuron and i_2 the 'height' of its layer.

Example 1. Suppose that $g(s) = \mathrm{e}^{-as}$ for some $a > 0$. We assume that each neuron is influenced by an infinite number of neurons and that

$$W_{(j_1,j_2)\to(i_1,i_2)} = \frac{W}{|j_1 - i_1|^{\beta}}\,\mathbf{1}_{\{j_2 = i_2 - 1\}}\,\mathbf{1}_{\{j_1 \neq i_1\}}$$

for some $\beta > 0$ and $W > 0$, where $\mathbf{1}_{\{\cdot\}}$ is the indicator. Putting $V_{(i_1,i_2)}(k) := \{(j_1, i_2 - 1);\ 1 \leq |j_1 - i_1| \leq k\}$, we have, for all neurons $j \in \partial V_i(k-1)$, $W_{j\to i} = 1/k^{\beta}$. Condition (19) can now be written as

$$\sum_{k\geq 2}\left[(k\Lambda(2k+1)+1)\left(\sum_{l=1}^{k-1}\frac{\Lambda}{l^{\beta}}\mathrm{e}^{-ak}(\mathrm{e}^{a}-1) + \frac{\Lambda}{k^{\beta}}(1-\mathrm{e}^{-ak})\right)\right] < \frac{a}{2W\gamma}.$$

This sum is finite if and only if $\beta > 3$. In particular, for fixed a, Λ, γ, and $\beta > 3$ there exists W^* such that $W < W^*$ implies (19).

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

Remark 5. In the model of Section 3 the presence of spontaneous spikes guarantees that the neural network is always active. In the frame of cascades of spike trains, we do not impose the presence of spontaneous spikes. Thus, we have to study the nonextinction of the process. Two cases have to be considered. Firstly, if for all $i \in I$, $\psi_i(0) = 0$, then $Z^i \equiv \mathbf{0}$ for all $i \in I$ is a possible stationary version of the Hawkes process with intensity (17). Since we are dealing with the uniqueness regime here, Theorem 3 implies that this is the unique stationary solution in this case. This situation is of course not of interest for us.

The second case to be considered is that of $\psi_i(0) > 0$ for at least one i. In this situation, the process does not go extinct as we show in the following lemma. As a consequence, in this case, the only invariant measure of the process is not the trivial one.

Lemma 1. *If there exists $i \in I$ such that $\psi_i(0) > 0$, then, for all $t \in \mathbb{R}$, we have almost surely,*

$$\sum_{j \in \{i\} \cup \mathcal{V}_{\cdot \to i}} Z^j([t, +\infty)) > 0.$$

Proof. Here $\psi_i(0) > 0$ means that if the neuron i receives no information from the other neurons, it has a positive rate of fire. In other words, the only way for the neuron i to stay silent is to be regularly inhibited. In both situations we will have an activity in the network, either of neuron i itself or of its inhibitors. The idea of the proof is then to suppose that for all $j \in \mathcal{V}_{\cdot \to i}$, $Z^j([t, +\infty)) = 0$, and to prove that in this case, almost surely $Z^i([t, +\infty)) > 0$ using the hypothesis $\psi_i(0) > 0$ together with the continuity of ψ_i. □

Remark 6. (*Some remarks on the extinction problem.*) We continue the discussion of Remark 5 concerning the extinction probability of the system in the $\psi_i(0) = 0$ case for all $i \in I$. As pointed out above, in this case, the only invariant measure of the system is the trivial one δ_0.

This situation might be different when considering the process from a macroscopic point of view, reminiscent of what is called 'hydrodynamical limit behavior' in statistical physics. To be more precise, suppose that we observe a set of N neurons, $I = I^N = \{1, \ldots, N\}$, that $W_{j \to i} = 1/N$ for all $i \neq j$, $g_j(t) = \mathrm{e}^{-\alpha t}$ for some $\alpha > 0$, and that $\psi_i(x) = \psi(x)$ for all i. We suppose, moreover, that $\psi(0) = 0$ and that $\psi'(0) = \beta$. In this case, for fixed N, the process goes extinct almost surely, as has been shown in Theorem 2.3 of [7]. We now consider the macroscopic behavior as $N \to \infty$. It has been proven (see, e.g. [9]) that the system possesses the so-called 'propagation of chaos' property. This means that in the large population limit, the neurons converge in law to independent and identically distributed (i.i.d.) copies of the same limit law. In this macroscopic limit, the longtime behavior of a typical neuron can be different from the one of a typical neuron in a finite-size system. In particular, in Theorem 12 of [6] it was shown that under the assumption that $\beta > \alpha$, the macroscopic system does not go extinct almost surely in the long run, although $\psi(0) = 0$. In other words, the trivial invariant measure is unstable in this case.

However, the point of view adopted in this paper is a different one. We are looking at infinite systems of neurons, but at a microscopic level at which each neuron is observed at scale 1. In this situation, the only way of preventing the system from going extinct is summarized in the above Lemma 1.

5. A dominating Poisson random measure

Recall that the firing rate functions ψ_i considered in this paper are bounded by a constant Λ_i for each $i \in I$. We will use this assumption to introduce a Poisson random measure (PRM)

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

$N(\mathrm{d}t, \mathrm{d}i, \mathrm{d}z)$ on $\mathbb{R} \times I \times [0, 1]$ with intensity $\mathrm{d}t\,(\sum_{i \in I} \Lambda_i \delta_i)\,\mathrm{d}z$ on $\mathbb{R} \times I \times [0, 1]$ dominating the process $(Z^i,\ i \in I)$. This allows us to select the jump times of Z^i among those of N^i according to probabilities driven by the function ϕ_i (recall (3)). This leads to the following definition.

Definition 2. A family $(Z^i,\ i \in I)$ of random point measures defined on $(\Omega, \mathcal{A}, \mathbb{F})$ is said to be a Hawkes process with variable length memory and an infinite number of interacting components with parameters (ψ, h, g) if almost surely, for all $i \in I$ and $C \in \mathcal{B}(\mathbb{R})$,

$$Z^i(C) = \int_C \int_{\{i\}} \int_{[0,1]} \mathbf{1}_{\{z \le (1/\Lambda_i)\psi_i(\sum_j h_{j\to i}(\int_{[L_t^i, t)} g_j(t-s)\,\mathrm{d}Z_s^j))\}}\, N(\mathrm{d}t, \mathrm{d}i, \mathrm{d}z). \tag{20}$$

According to Brémaud and Massoulié [1] (see also Proposition 3 of Delattre *et al.* [5]), a Hawkes process according to Definition 2 is a Hawkes process according to Definition 1 and vice versa.

Equation (20) implies that we can construct the process $(Z^i,\ i \in I)$ by a thinning procedure applied to the *a priori* family of dominating PRMs $N^i(\mathrm{d}t) = N(\mathrm{d}t, \{i\}, [0, 1])$ having intensity $\Lambda_i\,\mathrm{d}t$ each. Since Z^i is a simple point measure, it is enough to define it through the jump times of its atoms. Each atom of Z^i must also be an atom of N^i since $Z^i \ll N^i$. In other words, the associated counting process $\alpha(Z^i)$ can only jump when the counting process $\alpha(N^i)$ jumps. We write $T_n^i,\ n \in \mathbb{Z}$, for the jump times of N^i. Fix a given jump $t = T_n^i$ of N^i. Then, conditionally on N, the probability that this time is also a jump time of Z^i is given by

$$\mathbb{P}(Z^i(\{t\}) = 1 \mid \mathcal{F}_t) = \phi_i\left(\sum_j W_{j\to i}\left(\int_{[L_t^i, t)} g_j(t-s)\,\mathrm{d}Z_s^j\right)\right) =: p_{(i,t)}(1 \mid \mathcal{F}_t). \tag{21}$$

In other words, given that t is a jump time of N^i, $p_{(i,t)}(1 \mid \mathcal{F}_t)$ is the probability that this jump is also a jump of Z^i. This probability depends on the past before time t of the process. In what follows we propose a decomposition of $p_{(i,t)}(1 \mid \mathcal{F}_t)$ according to growing time-space neighborhoods that explore the *relevant past* needed in order to determine $p_{(i,t)}(1 \mid \mathcal{F}_t)$. This decomposition is a Kalikow-type decomposition as considered first by Ferrari *et al.* [8] and in Comets *et al.* [3]. This type of technique was then considered in a series of papers for perfect simulation issues. See Galves and Löcherbach [10] for an application in the context of neural biological nets in discrete time. The decomposition that we consider here is a nontrivial extension of the previous considerations to the framework of continuous-time neural nets. In the case of Theorem 3, this decomposition has to be achieved in a random environment, where the environment is given by the *a priori* realization of the PRM N. We start with the proof of Theorem 1, which is conceptually simpler.

6. Proof of Theorem 1

6.1. Kalikow-type decomposition

The condition (9) of Theorem 1 allows us to decompose the law of the conditional probability (21) according to space neighborhoods $V_i(k)$. This decomposition will be independent of the realization of the *a priori* PRM N. This will be crucial in the perfect simulation procedure described in the next subsection. We will work with $\tilde{S}$, the state space of relevant configurations of the process defined in Remark 6. For the convenience of the reader, we recall its definition here:

$$\tilde{S} := \{\eta \in S\colon \eta(j \to i) \le K_{j\to i} \quad \text{for all } e = (j \to i) \in \mathcal{E}\}$$

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

and introduce the following notation. First,

$$r_i^{[0]}(1) = \inf_{\eta \in \tilde{S}} \phi_i\left(\sum_j W_{j\to i}\eta(j \to i)\right),$$

which is the minimal probability that neuron i spikes uniformly with respect to all possible configurations. Then we define

$$r_i^{[0]}(1) = \inf_{\eta \in \tilde{S}}\left(1 - \phi_i\left(\sum_j W_{j\to i}\eta(j \to i)\right)\right),$$

which is the minimal probability that neuron i does not spike. Next, for each $k \geq 1$ and each $\zeta \in \tilde{S}$, we define the set $D_i^k(\zeta)$ by

$$D_i^k(\zeta) := \{\eta \in \tilde{S} \colon \text{for all } j \in V_i(k),\ \eta(j \to i) = \zeta(j \to i)\}$$

and set

$$r_i^{[k]}(1 \mid \zeta) = \inf_{\eta \in D_i^k(\zeta)} \phi_i\left(\sum_j W_{j\to i}\eta(j \to i)\right),$$
$$r_i^{[k]}(0 \mid \zeta) = \inf_{\eta \in D_i^k(\zeta)}\left(1 - \phi_i\left(\sum_j W_{j\to i}\eta(j \to i)\right)\right).$$

Finally, we define

$$\alpha_i(0) = r_i^{[0]}(1) + r_i^{[0]}(0), \quad \alpha_i(k) = \inf_{\zeta \in \tilde{S}}(r_i^{[k]}(1 \mid \zeta) + r_i^{[k]}(0 \mid \zeta)) \quad \text{for all } k \geq 1,$$

and let

$$\mu_i(0) = \alpha_i(0), \qquad \mu_i(k) = \alpha_i(k) - \alpha_i(k-1) \quad \text{for all } k \geq 1.$$

Lemma 2. *It holds that* $(\mu_i(k))_{k\geq 0}$ *defines a probability on* $\mathbb{N}$.

Proof. Indeed, since $D_i^k(\zeta) \subset D_i^{k-1}(\zeta)$, we have $\mu_i(k) \geq 0$ for all $k \geq 0$. Therefore, all we have to show is that

$$\sum_{k\geq 0} \mu_i(k) = 1. \tag{22}$$

Note that $\sum_{k\geq 0}\mu_i(k) = \lim_{k\to+\infty}[\inf_{\zeta\in\tilde{S}}(r_i^{[k]}(1 \mid \zeta) + r_i^{[k]}(0 \mid \zeta))]$. Hence, it is sufficient to show that, for all $\zeta \in \tilde{S}$,

$$\lim_{k\to+\infty}(r_i^{[k]}(1 \mid \zeta) + r_i^{[k]}(0 \mid \zeta)) = 1.$$

For all $k \geq 0$, we have

$$r_i^{[k]}(1 \mid \zeta) + r_i^{[k]}(0 \mid \zeta)$$
$$= 1 - \left[\sup_{\eta \in D_i^k(\zeta)} \phi_i\left(\sum_j W_{j\to i}\eta(j \to i)\right) - \inf_{\eta \in D_i^k(\zeta)} \phi_i\left(\sum_j W_{j\to i}\eta(j \to i)\right)\right].$$

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

Using the fact that ϕ_i is increasing and uniformly continuous with modulus ω, we deduce that

$$\begin{aligned}
&\sup_{\eta\in D_i^k(\zeta)} \phi_i\Big(\sum_j W_{j\to i}\eta(j\to i)\Big) - \inf_{\eta\in D_i^k(\zeta)} \phi_i\Big(\sum_j W_{j\to i}\eta(j\to i)\Big)\\
&\quad\le \omega\Big(\sup_{\eta\in D_i^k(\zeta)}\Big(\sum_j W_{j\to i}\eta(j\to i)\Big) - \inf_{\eta\in D_i^k(\zeta)}\Big(\sum_j W_{j\to i}\eta(j\to i)\Big)\Big)\\
&\quad\le \omega\Big(\sum_{j\notin V_i(k)} |W_{j\to i}|K_{j\to i}\Big).
\end{aligned}$$

Now, taking the limit as k tends to $+\infty$ and taking into account condition (7), we obtain (22) as desired. □

We return to the conditional probability $p_{(i,t)}(1 \mid \mathcal{F}_t)$ introduced in (21). The history is realized only through the effected choices of acceptance or rejection of jumps of the *a priori* PRM N. Therefore, we introduce the time grid $\mathcal{G} = \{(i, T_n^i),\ i \in I\}$. Any realization of the Hawkes process, conditionally with respect to the PRM N, can be identified with an element of $X := \{0,1\}^{\mathcal{G}}$. We write $x = (x^i)_{i\in I}$ for elements of X, where $x^i = (x^i(T_n^i))_{n\in\mathbb{Z}}$. Elements $x \in X$ can be interpreted as point measures. The object of our study is

$$p_{(i,t)}(1 \mid x) = \phi_i\Big(\sum_j W_{j\to i}(x^j([L_t^i(x), t)) \wedge K_{j\to i})\Big).$$

In the following proposition we establish a Kalikow-type decomposition for $p_{(i,t)}(\cdot \mid x)$ with respect to growing neighborhoods of $\mathcal{V}_{\cdot\to i}$.

Proposition 1. *Grant Assumption 1 and assume that (7) is satisfied. Fix $t = T_n^i$ for some $n \in \mathbb{Z}$ and $i \in I$. Then there exists a family of conditional probabilities $(p_{(i,t)}^{[k]}(\cdot \mid x))_{k\ge 0}$ on $\{0,1\}$ satisfying the following properties.*

- *For all $a \in \{0,1\}$, $p_{(i,t)}^{[0]}(a \mid x) := r_i^{[0]}(a)/\mu_i(0)$ does not depend on the configuration x.*
- *For all $a \in \{0,1\}$, $k \ge 1$, $X \ni x \mapsto p_{(i,t)}^{[k]}(a \mid x)$ depends only on the variables $x^j : j \in V_i(k)$.*
- *For all $x \in X$, $k \ge 1$, $p_{(i,t)}^{[k]}(1 \mid x) \ge 0$ and $p_{(i,t)}^{[k]}(1 \mid x) + p_{(i,t)}^{[k]}(0 \mid x) = 1$.*
- *For all $x \in X$, we have the following convex decomposition:*

$$p_{(i,t)}(a \mid x) = \sum_{k\ge 0} \mu_i(k)\, p_{(i,t)}^{[k]}(a \mid x).$$

Proof. We identify a configuration $x \in X$ with an element x_t of $\tilde{S}$ for any $t \in \mathbb{R}$ by introducing

$$x_t(j \to i) = x^j([L_t^i, t)) \wedge K_{j\to i},$$

where x^j is interpreted as a point measure. Note that $p_{(i,t)}(a \mid x)$ only depends on x_t. We have

$$\begin{aligned}
p_{(i,t)}(a \mid x) &= p_{(i,t)}(a \mid x_t)\\
&= r_i^{[0]}(a) + \sum_{k=1}^{N} \Delta_i^{[k]}(a \mid x_t) + (p_{(i,t)}(a \mid x_t) - r_i^{[N]}(a \mid x_t)), \qquad (23)
\end{aligned}$$

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

where $\Delta_i^{[k]}(a \mid x_t) := r_i^{[k]}(a \mid x_t) - r_i^{[k-1]}(a \mid x_t)$. We start by showing that the last term in (23) tends to 0 as $N \to \infty$. Indeed, rewriting the definitions, we have

$$|p_{(i,t)}(a \mid x_t) - r_i^{[N]}(a \mid x_t)| = \sup_{z \in D_i^N(x_t)} \left| \phi_i\left(\sum_j W_{j\to i} x_t(j \to i)\right) - \phi_i\left(\sum_j W_{j\to i} z_{j\to i}\right)\right|.$$

Using the uniform continuity of ϕ_i and the definition of $D_i^N(x_t)$, we obtain

$$\begin{aligned}
|p_{(i,t)}(a \mid x_t) - r_i^{[N]}(a \mid x_t)| &\le \sup_{z \in D_i^N(x_t)} \left[\omega\left(\sum_j |W_{j\to i}||x_t(j \to i) - z_{j\to i}|\right)\right] \\
&\le \sup_{z \in \tilde{S}} \left[\omega\left(\sum_{j \notin V_i(N)} |W_{j\to i}||x_t(j \to i) - z_{j\to i}|\right)\right] \\
&\le \omega\left(\sum_{j \notin V_i(N)} |W_{j\to i}| K_{j\to i}\right) \to 0 \quad \text{as } N \to +\infty.
\end{aligned}$$

Taking the limit when N tends to $+\infty$ in (23), we obtain

$$p_{(i,t)}(a \mid x_t) = r_i^{[0]}(a) + \left(\sum_{k \ge 1} \Delta_i^{[k]}(a \mid x_t)\right).$$

Now we set, for $k \ge 1$,

$$\tilde{\mu}_i(k, x_t) := \sum_a \Delta_i^{[k]}(a \mid x_t) \quad \text{and} \quad \tilde{p}_{(i,t)}^{[k]}(a \mid x) = \frac{\Delta_i^{[k]}(a \mid x_t)}{\tilde{\mu}_i(k, x_t)},$$

where we define $\tilde{p}_{(i,t)}^{[k]}(a \mid x_t)$ in an arbitrary way if $\tilde{\mu}_i(k, x_t) = 0$. In this way we can write $\Delta_i^{[k]}(a \mid x_t) = \tilde{\mu}_i(k, x_t)\tilde{p}_{(i,t)}^{[k]}(a \mid x_t)$ and, therefore,

$$p_{(i,t)}(a \mid x_t) = \mu_i(0) p_{(i,t)}^{[0]}(a) + \sum_{k=1}^{\infty} \tilde{\mu}_i(k, x_t)\tilde{p}_{(i,t)}^{[k]}(a \mid x_t).$$

This decomposition is not yet the one announced in the proposition since the $\tilde{\mu}_i(k, x_t)$ depend on the configuration x_t. The weights $\mu_{(i,t)}(k)$ are already defined and they do not depend on the configuration. So we have to define new probabilities $p_{(i,t)}^{[k]}$, based on the previously defined $\tilde{p}_{(i,t)}^{[k]}$.

To define the new probabilities $p_{(i,t)}^{[k]}$, we introduce $\alpha_i(k, x_t) := \sum_{l=1}^{k} \tilde{\mu}_i(l, x_t)$. For each $k \ge 0$ let l' and l be indexes such that

$$\alpha_i(l' - 1, x_t) < \alpha_i(k - 1) \le \alpha_i(l', x_t) < \cdots < \alpha_i(l, x_t) < \alpha_i(k) \le \alpha_i(l + 1, x_t).$$

We then decompose the interval $(\alpha_i(k - 1), \alpha_i(k)]$ in the following way:

$$\begin{aligned}
&(\alpha_i(k - 1), \alpha_i(k)] \\
&\quad = (\alpha_i(k - 1), \alpha_i(l', x_t)] \cup \left(\bigcup_{m=l'+1}^{l} (\alpha_i(m - 1, x_t), \alpha_i(m, x_t))]\right) \cup (\alpha_i(l, x_t), \alpha_i(k)].
\end{aligned}$$

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

We define $p_i^{[k]}$ on each of the intervals of this partition. On the first interval $(\alpha_i(k-1), \alpha_i(l', x_t)]$, we use $\tilde{p}_i^{[l']}$, on each interval $(\alpha_i(m-1, x_t), \alpha_i(m, x_t)]$, we use $\tilde{p}_i^{[m]}$, and on $(\alpha_i(l, x_t), \alpha_i(k)]$, we use $\tilde{p}_i^{[l+1]}$. This leads, for each $k \geq 0$, to the following definition:

$$p_{(i,t)}^{[k]}(a \mid x_t) = \sum_{-1=l' \leq l}^{k-1} \mathbf{1}_{\{\alpha_i(l'-1,x_t) < \alpha_i(k-1) \leq \alpha_i(l',x_t)\}} \mathbf{1}_{\{\alpha_i(l,x_t) < \alpha_i(k) \leq \alpha_i(l+1,x_t)\}}$$
$$\times \left[\frac{\alpha_i(l', x_t) - \alpha_i(k-1)}{\mu_i(k)} \tilde{p}_{(i,t)}^{[l']}(a \mid x) + \sum_{m=l'+1}^{l} \frac{\tilde{\mu}_i(m, x_t)}{\mu_i(k)} \tilde{p}_{(i,t)}^{[m]}(a \mid x) \right.$$
$$\left. + \frac{\alpha_i(k) - \alpha_i(l, x_t)}{\mu_i(k)} \tilde{p}_{(i,t)}^{[l+1]}(a \mid x_t) \right].$$

It can easily be verified that, with this definition, we obtain the announced decomposition. □

6.2. Perfect simulation

In this section we show how to construct the stationary nonlinear Hawkes process with saturation threshold by a *perfect simulation procedure*, based on an *a priori* realization of the processes $(N^i,\ i \in I)$. Condition (8) allows us to decompose the Poisson process N^i of intensity Λ_i as

$$N^i = \hat{N}^i + \tilde{N}^i,$$

where $\hat{N}^i$ and $\tilde{N}^i$ are independent Poisson processes with respective intensities δ_i and $\Lambda_i - \delta_i$. Conditionally on these processes, we can characterize the process Z^i by the element $x \in X = \{0, 1\}^{\mathcal{G}}$ recording the times and the indices of the neuron for which we have accepted a jump. All jumps of $\hat{N}^i$ are also jumps of Z^i; we call them the *spontaneous spikes*. They appear at a jump time T_n^i of N^i with probability $d_i := \delta_i / \Lambda_i$. Moreover, any jump T_n^i of $\tilde{N}^i$ will be a jump of Z^i with probability

$$\frac{1}{1-d_i} \left[\phi_i \left(\sum_j W_{j \to i}(Z^j([L_t^i, t)) \wedge K_{j \to i}) \right) - d_i \right],$$

and we have to decide, for each $i \in I$ and each time $\tilde{T}_n^i$, whether this jump is accepted or not. This acceptance/rejection procedure will be achieved by means of the Kalikow-type decomposition and gives rise to a perfect simulation algorithm that we now introduce.

Fix $i \in I$ and $t = \tilde{T}_n^i$. In this algorithm we describe the random space-time subset of neurons and their associated spiking times that can possibly have an influence on the acceptance or rejection of a spike of neuron i at time t.

(0) In the following steps we will simulate $\hat{N}^j$ and $\tilde{N}^j$ for some sites $j \in I$ on bounded intervals of time. We introduce, for each $j \in I$, the sets $\hat{S}^j$ and $\tilde{S}^j$ that will contain the intervals of time where the processes $\hat{N}^j$ and $\tilde{N}^j$ have already been simulated at the current step of the algorithm. We initialize these sets so that, for all $j \in I$, $\hat{S}^j = \tilde{S}^j = \varnothing$.

(i) We simulate a random variable $K^{(i,t)} \in \mathbb{N}$ such that

$$\mathbb{P}(K^{(i,t)} = k) = \mu_j(k) \quad \text{for all } k \geq 0.$$

(ii) We determine $\hat{L}_t^i = \hat{T}^i_{\hat{N}^i_{t-}}$ in the following way. We simulate a random variable τ following an exponential law of parameter δ_i and introduce $\hat{S}_{\max}^{(i,t)} := \sup\{s \in \hat{S}^i \cap]-\infty, t]\}$ with the convention $\sup(\varnothing) = -\infty$.

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

- If $\tau \le t - \hat{S}^{(i,t)}_{\max}$ we set $\hat{L}^i_t = t - \tau$.
- If $\tau > t - \hat{S}^{(i,t)}_{\max}$ we choose for $\hat{L}^i_t$ the biggest jump time of $\hat{N}^i$ in $\hat{S}^i \cap (-\infty, t]$.

In both cases, we update $\hat{S}^i$, which in an algorithmic way can be written $\hat{S}^i \leftarrow \hat{S}^i \cup [\hat{L}^i_t, t)$.

(iii) For each $j \in V_i(K^{(i,t)})$, we simulate $\tilde{N}^j$ on the time interval $[\hat{L}^i_t, t) \setminus \tilde{S}^j$ and update $\tilde{S}^j$ such that $\tilde{S}^j \leftarrow \tilde{S}^j \cup [\hat{L}^i_t, t)$.

(iv) We introduce the set

$$C^{(i,t)}_1 := \{(j, \tilde{T}^j_l) \in I \times]-\infty, t) \colon j \in V_i(K^{(i,t)}),\ \tilde{T}^j_l \in [\hat{L}^i_t, t)\},$$

which we call 'first generation of the clan of ancestors' of element (i,t). Here, by convention, $C^{(i,t)}_1 = \varnothing$ if $K^{(i,t)} = 0$.

(v) If $C^{(i,t)}_1 \neq \varnothing$ then we iterate the above procedure and simulate for each element of $C^{(i,t)}_1$ a new clan of ancestors with steps (i) to (iv). So we set, for any $n \ge 2$,

$$C^{(i,t)}_n = \Bigg(\bigcup_{(j,s) \in C^{(i,t)}_{n-1}} C^{(j,s)}_1 \Bigg) \setminus (C^{(i,t)}_1 \cup \cdots \cup C^{(i,t)}_{n-1}), \tag{24}$$

which is the 'nth generation of the clan of ancestors' of (i,t).

Note that if $K^{(i,t)} = 0$ in step (i) then the algorithm stops immediately. Also if, for all $j \in V_i(k)$, for $k = K^{(i,t)}$, we have $\tilde{N}^j([\hat{L}^i_t, t)) = 0$, then $C^{(i,t)}_1 = \varnothing$, and the algorithm stops after one step. Otherwise, we introduce

$$N^{\text{stop}} := \min\{n \colon C^{(i,t)}_n = \varnothing\}$$

the number of steps of the algorithm, where $\min \varnothing := \infty$. The set $\mathcal{C}^{(i,t)} := \bigcup_{n=1}^{N^{\text{stop}}} C^{(i,t)}_n$ contains all nonspontaneous spikes which have to be accepted or not and whose possible presence has an influence on the acceptance/rejection decision of (i,t) itself. We will show below that, under the conditions of Theorem 1, $N^{\text{stop}} < \infty$ almost surely (a.s.).

Once the clans of ancestors are determined, we can realize the acceptance/rejection procedure of the elements in these clans in a second algorithm which is a forward procedure going from the past to the present. We start with the sites for which this decision can be made independently from anything else. During the algorithm the set of all sites for which a decision has already been achieved will then progressively be updated.

(0) At the initial stage of the algorithm, the set of sites for which the acceptance/rejection decision can be achieved is initialized by

$$D^{(i,t)} := \{(j,s) \in \mathcal{C}^{(i,t)},\ C^{(j,s)}_1 = \varnothing\}.$$

The sites within this set are precisely those for which the decision can be made independently from anything else.

(i) For each $(j,s) \in D^{(i,t)}$, we simulate, according to the probabilities

$$\frac{1}{1-d_j}(p^{[0]}_{(j,s)}(a) - d_j),$$

the state of this site.

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

(ii) For any $(j,s) \in \mathcal{C}^{(i,t)}$ with $C_1^{(j,s)} \subset D^{(i,t)}$, we then decide, according to the probabilities $(1/(1-d_j))(p_{(j,s)}^{[k]}(a \mid x) - d_j)$ with $k = K^{(j,s)}$, to accept or to reject the presence of a spike of neuron j at time s. This is possible since $p_{(j,s)}^{[k]}(a \mid x)$ depends on the configuration x only through the sites in $C_1^{(j,s)}$ whose states have already been determined since $C_1^{(j,s)} \subset D^{(i,t)}$. Then we update $D^{(i,t)}$ in the following way:

$$D^{(i,t)} \longleftarrow D^{(i,t)} \cup \{(j,s) \in \mathcal{C}^{(i,t)},\ C_1^{(j,s)} \subset D^{(i,t)}\}.$$

(iii) The update of $D^{(i,t)}$ allows us to repeat the previous step until $(i,t) \in D^{(i,t)}$.

Once we have assigned a decision to the element (i,t) itself, our *perfect simulation* algorithm stops. Of course, the whole procedure makes sense only if $N^{\text{stop}} < +\infty$ a.s., which we will prove now. For that sake, we define $C_1^{(i,t)}(k)$ to be the clan of ancestors of element (i,t) of size k, i.e.

$$C_1^{(i,t)}(k) := \{(j,s) \in I \times \mathbb{R}\colon\ j \in V_i(k), \text{there exists } n \in \mathbb{Z} \text{ such that } s = \tilde{T}_n^i,\ s \in [\hat{L}_t^i; t)\},$$

with the convention $C_1^{(i,t)}(0) := \varnothing$ and put $M^{(i,t)} := \sum_{k\geq 1} |C_1^{(i,t)}(k)|\mu_i(k)$, which is the conditional expectation, conditionally on the realization of the PRM N, of $|C_1^{(i,t)}|$. In order to prove that $N^{\text{stop}} < +\infty$ a.s., we compare the process $|C_n^{(i,t)}|$ with a branching process of reproduction mean M defined by $M := \sup_{i\in I,\, t\in\mathbb{R}} \mathbb{E}(M^{(i,t)})$. We will prove that the parameters of this branching process are such that $M < 1$, i.e. it goes extinct a.s. implying that $N^{\text{stop}} < \infty$ a.s.

Writing $\mathbb{E}^N$ for the conditional expectation with respect to N, we obtain the following recurrence:

$$\mathbb{E}^N(|C_n^{(i,t)}||C_{n-1}^{(i,t)}) \leq \sum_{(j,s)\in C_{n-1}^{(i,t)}} \sum_{k=1}^{+\infty} \#(C^{(j,s)}(k) \setminus C_{n-1}^{(i,t)})\mu_i(k).$$

Here we use $\sum_{k=1}^{+\infty} \#(C^{(j,s)}(k) \setminus C_{n-1}^{(i,t)})\mu_i(k)$ instead of $M^{(j,s)}$ in order to use the independence of $\mathbf{1}_{\{(j,s)\in C_{n-1}^{(i,t)}\}}$ and $\sum_{k=1}^{+\infty} \#(C^{(j,s)}(k) \setminus C_{n-1}^{(i,t)})\mu_i(k)$. This independence is due to the properties of the PRM associated with N and the fact that $C_{n-1}^{(i,t)}$ and $C_{(j,s)}(k) \setminus C_{n-1}^{(i,t)}$ are disjoint. It will allow us to claim that

$$\mathbb{E}\Bigg(\sum_{(j,s)} \mathbf{1}_{\{(j,s)\in C_{n-1}^{(i,t)}\}} \sum_{k=1}^{+\infty} \#(C^{(j,s)}(k) \setminus C_{n-1}^{(i,t)})\Bigg)\mu_i(k)) \leq \sum_{(j,s)} \mathbb{E}(\mathbf{1}_{\{(j,s)\in C_{n-1}^{(i,t)}\}})\mathbb{E}(M^{(j,s)}). \quad (25)$$

In order to prove the above inequality in a rigorous way, we study the transition operator governing the evolution (24). This leads to the definition of the following operator. We fix a neuron $i \in I$, a time $t = \tilde{T}_n^i$ and set

$$\begin{aligned} &Q((i,t),\cdot) \\ &\quad = \sum_{k\geq 0} \mu_i(k) \int_0^{+\infty} \mathrm{d}t_1 (\delta_i \mathrm{e}^{-\delta_i t_1}) \\ &\qquad \times \Bigg[\sum_{j\in V_i(k)} \Bigg(\sum_{n_j\geq 0} \mathrm{e}^{-(\Lambda_j-\delta_j)t_1} \frac{(\Lambda_j-\delta_j)^{n_j}}{n_j!} \int_{[t-t_1,t]^{n_j}} \mathrm{d}s_1^j \cdots \mathrm{d}s_{n_j}^j \delta_{\{(j,s_l^j)\colon l=1,\ldots,n_j\}}\Bigg)\Bigg], \end{aligned}$$

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

where $Q((i,t),\mathrm{d}C_1)$ is the law of $C_1^{(i,t)}$. In the above definition, k is the 'size' of the neighborhood $V_i(k)$ which is simulated according to the probabilities $(\mu_i(k))_{k\in\mathbb{N}}$; t_1 is the time between t and $\hat{L}_t^i$ which is an exponential random variable of parameter δ_i; n_j is the Poisson random variable $\tilde{N}^j([\hat{L}_t^i,t])$ of parameter $(\Lambda_j-\delta_j)t_1$; and $(s_l^j)_{l\in\{1,\dots,n_j\}}$ is the family of jump times in $[\hat{L}_t^i,t]$ of $\tilde{N}^j$, which, conditionally on the event $\{\tilde{N}^j([\hat{L}_t^i,t])=n_j\}$, are uniform random variables on $[\hat{L}_t^i,t]=[t-t_1,t]$.

The definition of the transition kernel $Q(C_1,\mathrm{d}C_2)$ defining the law of $C_2^{(i,t)}$ knowing $C_1^{(i,t)}$ is more complicated since C_2 is not the result of independent simulations for each element of C_1. (Indeed, the jump times of a process $\tilde{N}^j$ simulated for the clan of ancestors of an element (i_1,t_1) of C_1 have to be simulated once and for all, and reused for the determination of the clan of ancestors of an element (i_2,t_2) with $j\in V_{i_1}(k_1)\cap V_{i_2}(k_2)$. Similarly, the jump times of $\hat{N}^i$ have to be simulated once and for all in order to determine all the last spontaneous spiking times for all $(i,t)\in C_1$.) But $Q(C_1,\mathrm{d}C_2)$ can be upper bounded (in the sense of inclusion of the simulated sets) by

$$\prod_{(j,s)\in C_1} Q((j,s),\mathrm{d}C^{(j,s)})\delta_{(\bigcup_{(j,s)\in C_1}C^{(j,s)})}(\mathrm{d}C_2).$$

This upper bound simulates more random variables than necessary leading to bigger clans of ancestors. Since we are only interested in obtaining upper bounds on the number of elements in the clans of ancestors, we will therefore work with this upper bound and obtain

$$\begin{aligned}\mathbb{E}\Bigg(\sum_{(j,s)} 1_{(j,l)\in C_{n-1}^{(i,t)}}\sum_{k=1}^{+\infty}\#(C^{(j,s)}(k)\setminus C_{n-1}^{(i,t)})\mu_i(k)\Bigg)\\ \le\int_{(i,t)^c}Q((i,t),\mathrm{d}C_1)\int_{C_1^c}Q(C_1,\mathrm{d}C_2)\cdots\int_{(\bigcup_{\kappa=1}^{n-2}C_\kappa)^c}Q(C_{n-2},\mathrm{d}C_{n-1})\\ \times\Bigg[\sum_{(j,s)\in C_{n-1}}\int_{(\bigcup_{\kappa=1}^{n-1}C_\kappa)^c}Q((j,s),\mathrm{d}C)|C|\Bigg].\end{aligned}$$

Observe that

$$\int_{(\bigcup_{\kappa=1}^{n-1}C_\kappa)^c}Q((j,s),\mathrm{d}C)|C|\le\sum_{k\ge0}\mu_j(k)\int_0^{+\infty}\mathrm{d}t_1\Bigg[(\delta_j\mathrm{e}^{-\delta_j t_1})\sum_{a\in V_j(k)}((\Lambda_a-\delta_a)t_1)\Bigg].$$

The right-hand term of the above inequality being exactly $\mathbb{E}(M^{(j,s)})$ allows us to claim (25).

As a consequence, we can compare the process $|C_n^{(i,m)}|$ with a branching process. It remains to verify that this branching process becomes extinct a.s. in finite time. We will prove that actually

$$M=\sup_{i\in I,\,t\in\mathbb{R}}\mathbb{E}(M^{(i,t)})<1.$$

Proposition 2. *For all $i\in I$ and all $k\ge1$, we have*

$$\mu_i(k)\le\bar{\mu}_i(k). \tag{26}$$

We recall here the definition of $\bar{\mu}_i(k)$ in (10), i.e.

$$\bar{\mu}_i(k):=\omega\Bigg(\sum_{j\in\partial V_i(k-1)}|W_{j\to i}|K_{j\to i}\,\mathbf{1}_{\{W_{j\to i}<0\}}\Bigg)+\omega\Bigg(\sum_{j\in\partial V_i(k-1)}|W_{j\to i}|K_{j\to i}\,\mathbf{1}_{\{W_{j\to i}>0\}}\Bigg).$$

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

Proof. The proof follows directly from the fact that ϕ_i is increasing and uniformly continuous with modulus of continuity ω. For more details, we refer the reader to the proof of Proposition 4 below, which is similar (yet more difficult due to the fact that the decomposition considered there takes place in a random environment). □

We are now able to complete the proof of Theorem 3. We have the following first upper bound for $M^{(i,t)}$ thanks to the inequality (26): $M^{(i,t)} \leq \sum_{k\geq 1} |C^{(i,t)}(k)|\bar{\mu}_i(k)$. Consequently, by the definition of $C^{(i,t)}(k)$,

$$\mathbb{E}(M_{(i,t)}) \leq \sum_{k\geq 1} \mathbb{E}\Big(\sum_{j\in V_i(k)} \tilde{N}^j([\hat{L}^i_t, t])\Big)\bar{\mu}_i(k) \leq \sum_{k\geq 1}\Big[\Big(\sum_{j\in V_i(k)} \frac{\Lambda_j - \delta_j}{\delta_i}\Big)\bar{\mu}_i(k)\Big].$$

Assumption (9) of Theorem 1 implies that $M = \sup_i \mathbb{E}(M_{(i,t)}) < 1$. Consequently,

$$\mathbb{E}(|C^{(i,t)}_n|) \leq M^n \to 0 \quad \text{when } n \to \infty,$$

implying that the process $(C^{(i,t)}_n)_{n\in\mathbb{N}}$ goes extinct a.s. in finite time. As a consequence, $N^{\text{stop}} < \infty$ a.s. and the perfect simulation algorithm stops after a finite number of steps. This achieves the proof of the construction of the process. □

7. Proof of Theorem 3

7.1. Some comments

We now give the proof of Theorem 3. As in the proof of Theorem 1 we use a Kalikow-type decomposition of the transition probabilities (21). We will use the same notation for objects with slightly different definitions but playing the same role in the proof, in order to simplify the notation.

The main difference between the two models is that in the first one, no leakage term is present, whereas the second model contains a leakage term through the functions g_j. Moreover, in the first model, the presence of thresholds allows us to obtain a Kalikow-type decomposition according to space (and not to time) with probabilities $\mu_i(k)$ that are deterministic and do not depend on the realization of the PRM N. This is crucial because it gives us an independence argument in order to obtain statement (25). However, the statement of Theorem 1 is at the cost of two assumptions. The first one is the presence of spontaneous spikes due to condition (8). The second assumption is the existence of the thresholds $K_{j\to i}$.

In the second model, we will introduce a space-time decomposition in random environment with probabilities $\mu_{(i,t)}(k)$ that are $\sigma(N)$-measurable random variables. The independence argument leading to (25) will here be ensured by the condition (16) that we impose on the structure of the neural network.

7.2. Kalikow-type decomposition

In this section, the Kalikow-type decomposition will take place in a random environment depending on the realization of the PRM N. We will consider space-time neighborhoods $\mathbb{V}^i_t(k) := V_i(k) \times [t-k, t)$.

We work with the state pace $X = \{0,1\}^{\mathcal{G}}$, where $\mathcal{G}$ is the time grid $\{(i, T^i_n),\ (i,n) \in I \times \mathbb{Z}\}$. Note that each element x of X can be interpreted as a discrete measure dominated by $\mathrm{d}N$. We fix a time $t = T^i_n$ and we introduce the following notation:

$$r^{[0]}_{(i,t)}(1) = \inf_{x\in X} \phi_i\Big(\sum_j W_{j\to i} \int_{L^i_t(x)}^t g_j(t-s)\,\mathrm{d}x_j(s)\Big),$$

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

where $\mathrm{d}x_j = \sum_{m\in\mathbb{Z}} x(j, T_m^j)\,\mathrm{d}\delta_{T_m^j}$, $L_t^i(x) = \sup\{T_m^i < t\colon x(i, T_m^i) = 1\}$, and

$$r_{(i,t)}^{[0]}(0) = \inf_{x\in\mathcal{S}}\left(1 - \phi_i\left(\sum_j W_{j\to i}\int_{L_t^i(x)}^t g_j(t-s)\,\mathrm{d}x_j(s)\right)\right).$$

Now fix $x \in X$ and define, for each $k \geq 1$, the set $D_{(i,t)}^k(x)$ by $D_{(i,t)}^k(x) := \{z \in \mathcal{S}\colon z(\mathbb{V}_t^i(k)) = x(\mathbb{V}_t^i(k))\}$ and set

$$r_{(i,t)}^{[k]}(1 \mid x) = \inf_{z\in D_{(i,t)}^k(x)} \phi_i\left(\sum_j W_{j\to i}\int_{L_t^i(z)}^t g_j(t-s)\,\mathrm{d}z_j(s)\right),$$

$$r_{(i,t)}^{[k]}(0 \mid x) = \inf_{z\in D_{(i,t)}^k(x)} \left(1 - \phi_i\left(\sum_j W_{j\to i}\int_{L_t^i(z)}^t g_j(t-s)\,\mathrm{d}z_j(s)\right)\right).$$

We define $\alpha_{(i,t)}(0) := \mu_{(i,t)}(0) := r_{(i,t)}^{[0]}(1) + r_{(i,t)}^{[0]}(0)$, and, for all $k \geq 1$,

$$\alpha_{(i,t)}^{[k]} := \inf_{x\in\mathcal{S}}(r_{(i,t)}^{[k]}(1 \mid x) + r_{(i,t)}^{[k]}(0 \mid x)), \qquad \mu_{(i,t)}(k) := \alpha_{(i,t)}^{[k]} - \alpha_{(i,t)}^{[k-1]}.$$

These definitions are similar to those of Section 4.1 and play the same role in the proof. Note that in this section we have a dependence on time denoted by the index t.

Lemma 3. *It holds that $(\mu_{(i,t)}(k))_{k\geq 0}$ defines a probability on $\mathbb{N}$.*

Proof. The proof is similar to the proof of Lemma 2. □

Remark 7. Instead of condition (19), we can work with the following condition:

$$\lim_{k\to+\infty} \sum_{j\in V_i(k)} |W_{j\to i}|\Lambda_j \int_k^{+\infty} |g_j(s)|\,\mathrm{d}s + \sum_{j\notin V_i(k)} |W_{j\to i}|\Lambda_j \int_0^{+\infty} |g_j(s)|\,\mathrm{d}s = 0, \quad (27)$$

which is a necessary but not sufficient condition for (19).

Using these probabilities $(\mu_{(i,t)}(k))_{k\geq 0}$, we obtain a Kalikow-type decomposition for

$$p_{(i,t)}(1 \mid x) = \phi_i\left(\sum_j W_{j\to i}\int_{L_t^i(x)}^t g_j(t-s)\,\mathrm{d}x_j(s)\right).$$

Recall that $\mathcal{F}_t$ is the sigma-field generated by $Z^i((s,u])$, $s \leq u \leq t$, $i \in I$, and that we work on the probability space $(\Omega, \mathcal{A}, \mathbb{F})$ defined in Section 1.2.

Proposition 3. *Assume that conditions (18) and (27) are satisfied. Then there exists a family of conditional probabilities $(p_{(i,t)}^{[k]}(a \mid x))_{k\geq 0}$ satisfying the following properties.*

- *For all $a \in \{0,1\}$, $p_{(i,t)}^{[0]}(a \mid x) := r_{(i,t)}^{[0]}(a)/\mu_{(i,t)}(0)$ does not depend on the configuration x.*
- *For all $a \in \{0,1\}$, $k \geq 1$, $\mathcal{S} \ni x \mapsto p_{(i,t)}^{[k]}(a \mid x)$ depends only on the variables $(x(j, T_n^j)\colon (j, T_n^j) \in \mathbb{V}_t^i(k))$.*
- *For all $x \in \mathcal{S}$, $k \geq 1$, $p_{(i,t)}^{[k]}(1 \mid x) \in [0,1]$, $p_{(i,t)}^{[k]}(1 \mid x) + p_{(i,t)}^{[k]}(0 \mid x) = 1$.*
- *For all $a \in \{0,1\}$, $x \in \mathcal{S}$, $p_{(i,t)}^{[k]}(a \mid x)$ and $\mu_{(i,t)}(k)$ are $\mathcal{F}_t$-measurable random variables.*

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

- *For all $x \in \mathcal{S}$, we have the following convex decomposition:*

$$p_{(i,t)}(a \mid x) = \sum_{k \geq 0} \mu_{(i,t)}(k)\, p^{[k]}_{(i,t)}(a \mid x).$$

Proof. The proof follows the lines of the proof of Proposition 1. □

7.3. Perfect simulation

The idea for the perfect simulation algorithm is the same as in Section 6.2, but here we use a decomposition in space and time. We work conditionally on the realization of the PRM N and consider, for each $k \geq 1$, the space time neighborhood $\mathbb{V}^i_t(k)$. Define the clan of ancestors of element (i, t) of size k by

$$C_1^{(i,t)}(k) := \mathcal{G} \cap \mathbb{V}^i_t(k),$$

where $\mathcal{G}$ is the time grid $\{(i, T^i_n), (i, n) \in I \times \mathbb{Z}\}$ and where, by convention, $C_1^{(i,t)}(0) := \varnothing$. We choose as before i.i.d. random variables $K^{(i,t)} \in \mathbb{N}$ which are attached to each site $(i, t) \in \mathcal{G}$, chosen according to

$$\mathbb{P}(K^{(i,t)} = k) = \mu_{(i,t)}(k) \quad \text{for all } k \geq 0.$$

These random variables allow us to define the clans of ancestors $(C_n^{(i,t)})_n \subset I \times (-\infty, t)$ as follows. We set $C_1^{(i,t)} := C_1^{(i,t)}(K^{(i,t)})$ and

$$C_n^{(i,t)} := \Bigg(\bigcup_{(j,s) \in C_{n-1}^{(i,t)}} C_1^{(j,s)} \Bigg) \setminus (C_1^{(i,t)} \cup \cdots \cup C_{n-1}^{(i,t)}).$$

As before, we have to prove that the process $|C_n^{(i,t)}|$ converges a.s. to 0 as n tends to $+\infty$. For this sake, we compare the process $|C_n^{(i,t)}|$ with a branching process of reproduction mean (depending on space and time)

$$M^{(i,t)} := \sum_{k \geq 1} |C_1^{(i,t)}(k)| \mu_{(i,t)}(k).$$

Recall that $\mathbb{E}^N$ denotes the conditional expectation with respect to N; clearly,

$$\mathbb{E}^N(|C_n^{(i,t)}||C_{n-1}^{(i,t)}) \leq \sum_{(j,s) \in C_{n-1}^{(i,t)}} M^{(j,s)}.$$

Now, the structure that we imposed on the neural network in (16) implies that, for neurons $i \in I_l$ and $j \in I_m$, the event $\{(j, s) \in C_n^{(i,t)}\}$ is empty if $l \neq m + n$. Moreover, this event depends only on realizations for neurons in layers I_p with $m \leq p < l$, whereas $M^{(j,s)}$ depends on realizations for neurons in the layer I_{m-1}. Consequently, $\mathbf{1}_{\{(j,s) \in C_n^{(i,t)}\}}$ is independent of $M^{(j,s)}$, which allows us to write

$$\mathbb{E}\Bigg(\sum_{(j,s)} \mathbf{1}_{\{(j,s) \in C_n^{(i,t)}\}} M^{(j,s)} \Bigg) = \sum_{(j,s)} \mathbb{E}(\mathbf{1}_{\{(j,s) \in C_n^{(i,t)}\}}) \mathbb{E}(M^{(j,s)}).$$

The next step in the proof is to show that

$$\sup_{i \in I} \mathbb{E}^N(M^{(i,t)}) < 1.$$

As in Section 6.2, we start with a proposition that gives an upper bound for the probabilities $(\mu_{(i,t)}(k))_{k \geq 1}$.

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

Proposition 4. *We have*

$$\mu_{(i,t)}(k) \le \gamma\Bigg(\sum_{j\in V_i(k-1)} |W_{j\to i}| \int_{t-k}^{t-k+1} g_j(t-s)\,\mathrm{d}N_s^j + \sum_{j\in \partial V_i(k-1)} |W_{j\to i}| \int_{t-k}^{t} g_j(t-s)\,\mathrm{d}N_s^j \Bigg). \tag{28}$$

Proof. Using the definition of $\mu_{(i,t)}(k)$,

$$\mu_{(i,t)}(k) = \inf_{x\in\mathcal{S}} (r_{(i,t)}^{[k]}(1 \mid x) + r_{(i,t)}^{[k]}(0 \mid x)) - \inf_{x\in\mathcal{S}} (r_{(i,t)}^{[k-1]}(1 \mid x) + r_{(i,t)}^{[k-1]}(0 \mid x)).$$

Fix $\varepsilon > 0$ and let $u \in \mathcal{S}$ be such that

$$r_{(i,t)}^{[k-1]}(1 \mid u) + r_{(i,t)}^{[k-1]}(0 \mid u) \le \inf_{x\in\mathcal{S}} (r_{(i,t)}^{[k-1]}(1 \mid x) + r_{(i,t)}^{[k-1]}(0 \mid x)) + \varepsilon.$$

Then

$$\mu_{(i,t)}(k) \le (r_{(i,t)}^{[k]}(1 \mid u) + r_{(i,t)}^{[k]}(0 \mid u)) - (r_{(i,t)}^{[k-1]}(1 \mid u) + r_{(i,t)}^{[k-1]}(0 \mid u)) + \varepsilon.$$

Here we can assume, without loss of generality, that $L_t^i(u) = -\infty$. Indeed, if $L_t^i(u) > -\infty$, let u' be such that $L_t^i(u') = -\infty$, $u'((-\infty; L_t^i(u)]) = 0$ and $u'((L_t^i(u); t]) = u((L_t^i(u); t])$, then u' and u are two equivalent configurations in terms of acceptance/rejection decision of the site (i,t). Then

$$\begin{aligned}
\mu_{(i,t)}(k) - \varepsilon \le & \inf_{z\in D_{(i,t)}^k(u)} \phi_i\Bigg(\sum_j W_{j\to i} \int_{L_t^i(z)}^t g_j(t-s)\,\mathrm{d}z_j(s)\Bigg) \\
& - \inf_{z\in D_{(i,t)}^{k-1}(u)} \phi_i\Bigg(\sum_j W_{j\to i} \int_{L_t^i(z)}^t g_j(t-s)\,\mathrm{d}z_j(s)\Bigg) \\
& + \sup_{z\in D_{(i,t)}^{k-1}(u)} \phi_i\Bigg(\sum_j W_{j\to i} \int_{L_t^i(z)}^t g_j(t-s)\,\mathrm{d}z_j(s)\Bigg) \\
& - \sup_{z\in D_{(i,t)}^{k}(u)} \phi_i\Bigg(\sum_j W_{j\to i} \int_{L_t^i(z)}^t g_j(t-s)\,\mathrm{d}z_j(s)\Bigg).
\end{aligned}$$

Using condition (4) and the fact that ϕ_i is nondecreasing, we obtain

$$\begin{aligned}
\mu_{(i,t)}(k) - \varepsilon \le & \gamma\Bigg(\inf_{z\in D_{(i,t)}^k(u)} \sum_j W_{j\to i} \int_{L_t^i(z)}^t g_j(t-s)\,\mathrm{d}z_j(s) \\
& - \inf_{z\in D_{(i,t)}^{k-1}(u)} \sum_j W_{j\to i} \int_{L_t^i(z)}^t g_j(t-s)\,\mathrm{d}z_j(s)\Bigg) \\
& + \gamma\Bigg(\sup_{z\in D_{(i,t)}^{k-1}(u)} \sum_j W_{j\to i} \int_{L_t^i(z)}^t g_j(t-s)\,\mathrm{d}z_j(s) \\
& - \sup_{z\in D_{(i,t)}^{k}(u)} \sum_j W_{j\to i} \int_{L_t^i(z)}^t g_j(t-s)\,\mathrm{d}z_j(s)\Bigg).
\end{aligned}$$

Now we will simplify this expression detailing the configurations that realize the extrema. In order to reach a lower bound, we have to fix z such that $L_t^i(z) = -\infty$ (we can do this

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

since $L_t^i(u) = -\infty$) and whenever we have the choice for z we also have to fix $z = 1$ if the corresponding $W_{j\to i}$ is negative, otherwise we have to fix $z = 0$. We do the opposite choice in order to reach an upper bound. After factorization, we obtain the announced upper bound (28), i.e.

$$\mu_{(i,t)}(k) \le \gamma\Bigg(\sum_{j\in V_i(k-1)} |W_{j\to i}| \int_{t-k}^{t-k+1} g_j(t-s)\,\mathrm{d}N_s^j + \sum_{j\in\partial V_i(k-1)} |W_{j\to i}| \int_{t-k}^{t} g_j(\bar{t}-s)\,\mathrm{d}N_s^j\Bigg),$$

thanks to the simplification $\sum_j W_{j\to i}\,\mathbf{1}_{\{W_{j\to i}>0\}} - \sum_j W_{j\to i}\,\mathbf{1}_{\{W_{j\to i}<0\}} = \sum_j |W_{j\to i}|$.

Now, we will use the stationarity of the PRM; this will allow us to omit the dependence in time (by fixing $t = 0$) since we are only interested in the expectation. We have the following first upper bound for $M^{(i,t)}$ thanks to (28):

$$\frac{M^{(i,t)}}{\gamma} \le \sum_{k\ge1} |C_{(i,t)}(k)|\Bigg(\sum_{j\in V_i(k-1)} |W_{j\to i}| \int_{k-1}^{k} g_j(s)\,\mathrm{d}N_j(s) + \sum_{j\in\partial V_i(k-1)} |W_{j\to i}| \int_{0}^{k} g_j(s)\,\mathrm{d}N_j(s)\Bigg). \tag{29}$$

The difficulty in calculating the expectation of this upper bound is that N_s^j is present in each term of the product so that these terms are not independent. However, we have independence whenever the indexes denoting the neurons are different or when the intervals of time that we consider are disjoint. We will therefore decompose the sums in the previous expression in order to isolate the products of nonindependent terms. Then we calculate separately the expectations of these terms. For example, we have

$$\mathbb{E}\Bigg[\Bigg(\int_0^k \mathrm{d}N_s^j\Bigg)\Bigg(\int_0^k g_j(s)\,\mathrm{d}N_s^j\Bigg)\Bigg] = \Lambda_j(k\Lambda_j + 1)\Bigg(\int_0^k g_j(s)\,\mathrm{d}s\Bigg). \tag{30}$$

This result allows us to decompose the above expectation as the sum of the covariance and the product of expectations. We can make such a decomposition for each product of nonindependent terms in (29). Consequently, the expectation of the upper bound in (29) can be written as the sum of the covariances of the nonindependent terms and the products of expectations of all the terms. After factorization, we finally obtain

$$\frac{\mathbb{E}(M^{(i,t)})}{\gamma} \le \sum_{k\ge1}\Bigg[\Bigg(k\Bigg(\sum_{j\in V_i(k)} \Lambda_j\Bigg) + 1\Bigg)\Bigg(\sum_{j\in V_i(k-1)} |W_{j\to i}|\Lambda_j \int_{k-1}^{k} g_j(s)\,\mathrm{d}s + \sum_{j\in\partial V_i(k-1)} |W_{j\to i}|\Lambda_j \int_{0}^{k} g_j(s)\,\mathrm{d}s\Bigg)\Bigg].$$

Now we can use the assumption (19) of Theorem 3 to deduce that $M := \sup_i \mathbb{E}(M^{(i,t)}) < 1$. Consequently, we have

$$\mathbb{E}(|C_n^{(i,t)}|) \le M^n \to 0 \quad \text{when } n \to \infty.$$

This ensures that the process $(C_n^{(i,t)})_{n\in\mathbb{N}}$ goes extinct a.s., or in other words that $N^{\mathrm{stop}} < \infty$ a.s. Consequently the perfect simulation algorithm ends in a finite time and this achieves the proof of the construction of the process. □

https://doi.org/10.1017/apr.2016.80 Published online by Cambridge University Press

Remark 8. As already mentioned in Remark 4(i), Theorem 3 holds under the weaker Assumption 1 imposing only equicontinuity instead of Lipschitz continuity imposed in Assumption 4. It is easy to adapt the proofs of Lemma 3 and Propositions 3 and 4 working with Assumption 1. However, the form of the upper bound in Proposition 4, under Assumption 4, allows us to compute an upper bound for $\mathbb{E}(M^{(i,t)})$ explicitly in terms of the parameters, as explained in (30). This is not possible imposing only Assumption 1.

This completes the proof of Theorem 3. □

Acknowledgements

We thank an anonymous referee for helpful comments and suggestions. Moreover, we thank A. Galves and P. A. Ferrari for many useful discussions on the subject.

This research has been conducted as part of the project Labex MME-DII (grant number ANR11-LBX-0023-01) and as part of the activities of FAPESP Research, Dissemination and Innovation Center for Neuromathematics (S. Paulo Research Foundation, grant number 2013/07699-0).

References

[1] Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Prob.* **24,** 1563–1588.
[2] Chevallier, J. (2015). Mean-field limit of generalized Hawkes processes. Preprint. Available at https://arxiv.org/abs/1510.05620v1.
[3] Comets, F., Fernández, R. and Ferrari, P. A. (2002). Processes with long memory: regenerative construction and perfect simulation. *Ann. Appl. Prob.* **12,** 921–943.
[4] Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes: Elementary Theory and Methods*, Vol. I, 2nd edn. Springer, New York.
[5] Delattre, S., Fournier, N. and Hoffmann, M. (2016). Hawkes processes on large networks. *Ann. App. Prob.* **26,** 216–261.
[6] Duarte, A. (2015). Stochastic models in neurobiology: from a multiunitary regime to EEG data. Doctoral Thesis, University of Sao Paulo.
[7] Duarte, A. and Ost, G. (2016). A model for neural activity in the absence of external stimuli. *Markov Process. Relat. Fields* **22,** 37–52.
[8] Ferrari, P. A., Maass, A., Martínez, S. and Ney, P. (2000). Cesàro mean distribution of group automata starting from measures with summable decay. *Ergodic Theory Dynam. Systems* **20,** 1657–1670.
[9] Fournier, N. and Löcherbach, E. (2016). On a toy model for interacting neurons. *Ann. Inst. H. Poincaré Prob. Statist.* **52,** 1844–1876.
[10] Galves, A. and Löcherbach, E. (2013). Infinite systems of interacting chains with memory of variable length—a stochastic model for biological neural nets. *J. Statist. Phys.* **151,** 896–921.
[11] Hansen, N. R., Reynaud-Bouret, P. and Rivoirard, V. (2015). Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli* **21,** 83–143.
[12] Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58,** 83–90.
[13] Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Prob.* **11,** 93–503.
[14] Jaisson, T. and Rosenbaum, M. (2015). Limit theorems for nearly unstable Hawkes processes. *Ann. Appl. Prob.* **25,** 600–631.
[15] Liggett, T. M. (1985). *Interacting Particle Systems*. Springer, Berlin.
[16] Massoulié, L. (1998). Stability results for a general class of interacting point processes, and applications. *Stoch. Process. Appl.* **75,** 1–30.
[17] Møller, J. and Rasmussen, J. G. (2005). Perfect simulation of Hawkes processes. *Adv. Appl. Prob.* **37,** 629–646.
[18] Reynaud-Bouret, P. and Schbath, S. (2010). Adaptive estimation for Hawkes processes: application to genome analysis. *Ann. Statist.* **38,** 2781–2822.