Taming neuronal noise with large networks

Valentin Marc SCHMUTZ

Abstract

 How does reliable computation emerge from networks of noisy neurons? While individual neurons are intrinsically noisy, the collective dynamics of populations of neurons taken as a whole can be almost deterministic, supporting the hypothesis that, in the brain, computation takes place at the level of neuronal populations.

Mathematical models of networks of noisy spiking neurons allow us to study the effects of neuronal noise on the dynamics of large networks. Classical mean-field models, i.e., models where all neurons are identical and where each neuron receives the average spike activity of the other neurons, offer toy examples where neuronal noise is absorbed in large networks, that is, large networks behave like deterministic systems. In particular, the dynamics of these large networks can be described by deterministic neuronal population equations.

In this thesis, I first generalize classical mean-field limit proofs to a broad class of spiking neuron models that can exhibit spike-frequency adaptation and short-term synaptic plasticity, in addition to refractoriness. The mean-field limit can be exactly described by a multidimensional partial differential equation; the long time behavior of which can be rigorously studied using deterministic methods.

Then, we show that there is a conceptual link between mean-field models for networks of spiking neurons and latent variable models used for the analysis of multi-neuronal recordings. More specifically, we use a recently proposed finite-size neuronal population equation, which we first mathematically clarify, to design a tractable Expectation-Maximization-type algorithm capable of inferring the latent population activities of multi-population spiking neural networks from the spike activity of a few visible neurons only, illustrating the idea that latent variable models can be seen as partially observed mean-field models.

In classical mean-field models, neurons in large networks behave like independent, identically distributed processes driven by the average population activity – a deterministic quantity, by the law of large numbers. The fact the neurons are identically distributed processes implies a form of redundancy that has not been observed in the cortex and which seems biologically implausible. To show, numerically, that the redundancy present in classical mean-field models is unnecessary for neuronal noise absorption in large networks, I construct a disordered network model where networks of spiking neurons behave like deterministic rate networks, despite the absence of redundancy.

This last result suggests that the concentration of measure phenomenon, which generalizes

the "law of large numbers" of classical mean-field models, might be an instrumental principle for understanding the emergence of noise-robust population dynamics in large networks of noisy neurons.

Résumé

Comment est-ce des computations fiables peuvent-elles émerger de réseaux de neurones aléatoires ? Bien que les neurones, pris individuellement, soient aléatoires, la dynamique collective de populations de neurones peut être quasi déterministe, ce qui soutient l'hypothèse que dans le cerveau, les computations ont lieu à l'échelle de populations neuronales.

Les modèles mathématiques de réseaux de neurones aléatoires nous permettent d'étudier l'effet du bruit neuronal sur la dynamique de grands réseaux. Les modèles classiques de champ moyen, où tous les neurones sont identiques et où chaque neurone reçoit la moyenne de l'activité de tous les autres neurones, nous offrent des exemples simples où le bruit neuronal est absorbé dans les grands réseaux, c'est-à-dire où les grands réseaux se comportent comme des systèmes déterministes. En particulier, la dynamique de grands réseaux peut être décrite par des équations de population neuronale.

Dans cette thèse, je commence par généraliser les résultats classiques de limite de champ moyen à une large classe de modèles de neurone à impulsion pouvant manifester de l'adaptation ou de la plasticité synaptique à court terme, en plus d'une période réfractaire. La limite de champ moyen peut être exactement décrite par une équation aux dérivées partielles multidimensionnelle dont nous étudions le comportement en temps long en employant des méthodes déterministe.

Ensuite, nous montrons qu'il existe un lien conceptuel entre les modèles de champ moyen pour les réseaux de neurones à impulsion et les modèles à variable latente utilisés pour l'analyse d'enregistrements multi-neuronaux. Plus spécifiquement, nous nous appuyons sur une équation de population neuronale pour population finie, récemment proposée, que nous commençons par clarifier mathématiquement, pour concevoir un algorithme de type espérance-maximisation capable d'inférer l'activité populationnelle de réseaux multi-population de neurones à impulsion, à partir de l'activité de quelques neurones visibles seulement, illustrant l'idée que les modèles à variable latente peuvent être vus comme des modèles de champ moyen partiellement observables.

Dans les modèles classiques de champ moyen, les neurones se comportent comme des processus indépendents et identiquement distribués, entraînés par l'activité moyenne de la population, une quantité déterministe par la loi des grands nombres. Le fait que les neurones soient identiquement distribués implique une forme de redondance qui n'a pas été observée expérimentalement et qui semble peu plausible. Pour montrer numériquement que la

redondance des modèles classiques de champ moyen n'est pas nécessaire à l'absorption du bruit neuronal dans les grands réseaux, je construis un exemple de réseau désordonné où les réseaux de neurones à impulsion se comportent comme des réseau de neurones à taux de décharge déterministes, en dépit de l'absence de redondance.

Ce dernier résultat suggère que la concentration de la mesure, qui peut être vue comme une généralisation de la "loi des grands nombres" des modèles classiques de champ moyen, peut être un principe clé pour comprendre l'émergence de dynamique de population robuste dans les grands réseaux de neurones aléatoires.

Mots clés : neurone à impulsion, processus de Hawkes nonlinéaire, approximation de champ moyen, adaptation, plasticité synaptique à court terme, équation de transport nonlocale, flucutations, modèle à variable latente, système désordonné, concentration de la mesure

Acknowledgements

My first thanks goes to my thesis director Prof. Wulfram Gerstner, without whom the present thesis would not have seen the daylight, for a list of reasons too long to be enumerated here. Throughout this thesis, his advice has been a never-failing compass. His support has been both scientific and practical, effective and selfless. I am particularly grateful for the fact that at every step of the thesis, Prof. Gerstner has always taken all necessary actions for things to move swiftly forward – those who have gone through a PhD know how invaluable this is.

My co-director Prof. Eva Löcherbach has played an equally important scientific role in this thesis for she has taught me how to do mathematically rigorous research as well as most of what I know in mathematical neuroscience. I warmly thank her for her many explanations, her availability, and her timely encouragements. Maybe more important for my morale, she has been for me an example of fulfilled, dedicated, and enthusiastic mathematician.

Although not an official co-director, Prof. Tilo Schwalger has almost acted as one. He has been a mentor and collaborator since 2016 and he is really the person who introduced me to the kind of models this thesis is concerned with.

I would like to thank Claudia Fonte for bringing her PDE expertise to this thesis and for her collaboration. Our intellectually stimulating Zoom calls made the lockdown much more tolerable. I also thank her thesis director Prof. Stéphane Mischler for having arranged this collaboration.

In the last year of my thesis, I had the great pleasure of supervising the Master thesis of Shuqi Wang; our joint work led to an important chapter of this thesis that I would certainly not have been able to produce on my own.

I also have to thank senior collaborators Dr. Guillaume Bellec and Dr. Johanni Brea for their wise advice, some pivotal discussions, and their contributions to the text of the last two result chapters of this thesis.

All the members of the Laboratory of Computational Neuroscience at EPFL have contributed to the successful completion of this thesis in one way or another. I express my warm thanks to my friends and collaborators on ongoing and past projects Berfin Simşek, Louis Pezon, and Christos Sourmpis. A very special thanks goes to my friend Alireza Modirshanechi without whom these last four years would not have been the same at all. Our countless discussions (including a marathon of animated debates on a Venice terrace with jugs of memorable wine) have had a significant impact on the content of this thesis; one of its chapters might not have

Contents

1 Introduction

1.1 Reliable computation in networks of unreliable neurons

On tour, a concert pianist can perform a dozen of recitals with the same program – the same sequence of pieces – in different towns over the course of several weeks. A recital contains around one and a half hour of music; the total number of notes, that is, the sum of the number of notes of each piece, can amount to tens of thousands. In each town, the sound of the piano, the weight of its keys, and the acoustics of the hall will differ but the experienced musician will manage to adjust his touch to these variations to produce the best possible sound. The fact that he can reliably play the same recital, by heart, in a dozen of towns – memory slips, while possible, are rare events – is an undeniable example of reproducible, complex behavior. While each performance is unique, they are far from being random, quite the contrary. At the beginning of a concert, we can all predict what is going to happen on stage over the next hour or two: the musician will follow the program.

The neuroscientist in the audience can not help but be astonished by the reproducibility of the behavior she is witnessing. She knows that if she buys a ticket for the next concert in the next town, it will be the same program, the same sequence of notes, and maybe even the same phrasing, the same dynamic contrasts. The reproducibility of this complex behavior implies reliable and precise brain computations. The neuroscientist is puzzled because this manifestation of reliable and precise computation *by* the brain seems incompatible with the messy, chaotic-looking, hardly reproducible neuronal activity observed *in* the brain. Where is the order – the crystalline performance – in the apparent disorder of brain activity?

Neurons in the brain communicate with each other by sending short pulses of electrical activity called spikes. A neuron integrates the signals it receives from other neurons and, depending on the received signals and its own past spike activity, it can either emit a spike (send a signal to other neurons) or stay silent. The chaotic aspect of neuronal activity reflects, in part, the fact that the neurons are intrinsically random. Even if we know exactly the signals a neuron receives, we can not perfectly predict the signals (the spikes) it will emit. This is not due to a lack of understanding of how neurons work but to their "imperfect" biological

machinery. How can the brain perform reliable computation with its unreliable wetware is the general question motivating this – mostly mathematical – thesis. Mathematics can help to address this question by articulating the "how" of hypothetical answers to the question: *how* can large networks of noisy neurons perform reliable and complex computation. The "how", whatever it is, can probably not hold in a few sentences or a simple schematic diagram.

In this thesis, I construct and analyse mathematical models of networks of neurons where the randomness of single neurons is "averaged out" at the network level, continuing a long line of work in theoretical and mathematical neuroscience. The models I consider neglect some important biological details and are therefore far from being perfect. However, they offer enlightening examples where the "how" can be rigorously reduced to principles of probability theory.

1.1.1 The mechanics of computation in the brain

Information processing in the brain is not centralized; there is no "central processing unit" (CPU). Instead, information is represented, transformed and stored in a highly distributed fashion over large numbers of neurons, even larger numbers of synaptic connections, and across multiple brain regions. The seminal papers of Hopfield (Hopfield 1982, 1984) on associative memory in neural networks provide an important conceptual example of how the brain could perform computation in a decentralized fashion. Without going into the details of the original Hopfield model (Hopfield 1982, 1984), I briefly recall here the general point of view on brain computation that this model inspired (Amit 1989). At each time point, the brain is in a certain state, a network state; the evolution over time of the network state depends on neuronal characteristics and on the synaptic connections between neurons; sensory inputs can influence the state and the state determines motor actions. This dynamical systems point of view may sound innocent, but it is what is absent from it that makes it nontrivial: there is no need for a central processing unit; in particular, there is no need for a homunculus – a fallacious "little man" inside the brain, reading neuronal activity and making decisions. Hopfield showed that decentralized computation was possible in dynamical systems representing networks of (abstract) neurons, and argued that this type of computation could be understood as an emergent behavior (Hopfield 1982, 1984), an idea that led to major theoretical results in statistical physics (Amit, Gutfreund, and Sompolinsky 1985a,b, 1987; Gardner 1988; Gardner and Derrida 1988).

Following the general principles of the Hopfield model (see also Amari (1977) and Cohen and Grossberg (1983)) and adding biological realism, theorists have proposed models of networks of *spiking* neurons, i.e., where neurons communicate via discrete spikes, capable of approximating continuous dynamical systems performing computations related to associative memory (Gerstner and van Hemmen 1992) and perceptual decision making (Wang 2002; Wong and Wang 2006) (see the forth part of Gerstner, Kistler, et al. (2014) for an overview). In parallel, researchers analyzing *in vivo* multi-neuronal recordings have provided converging evidence

that populations of neurons approximate computation-performing, continuous dynamical systems (Mante et al. 2013; Shenoy, Sahani, Churchland, et al. 2013; Vyas et al. 2020). The motivating question of this thesis is: How can noise-robust dynamics supporting reliable computation emerge from networks of noisy neurons? As I explain in the following, neurons are intrinsically noisy and neuronal noise has to be tamed, or "averaged out," at the network level.

1.1.2   The problem of spike time variability

Before focusing on neuronal noise, we first need to discuss the broader question of spike time variability. To do so, let us consider a famous example given by Shadlen and Newsome (1994). When the spike activity of a neuron of the visual cortex (area MT) is recorded over repeated presentations of the exact same time-varying visual stimulus (moving dots), the spike times of the neuron varies from one repetition to the other, although the average spike activity over repetitions, the peristimulus time histogram (PSTH), clearly follows the time-varying stimulus. The authors therefore argue that the spike times of a cortical neuron are not reproducible. Adopting a dynamical systems point of view, we can distinguish three possible sources of spike time variability.

First, although the stimulus is exactly the same across repetitions, the brain can be in a different state at the beginning of each repetition, meaning that the "initial conditions" of brain dynamics can vary across repetitions (Arieli et al. 1996). Second, inputs coming from cortical regions unrelated to vision can vary across repetitions; repeating the same visual stimulus is not sufficient to fully control the inputs received by a neuron in the visual cortex (Stringer et al. 2019). Finally, the biophysics of the neuron may be intrinsically noisy. Neurons in the mammalian brain are indeed intrinsically noisy (Faisal, Selen, and Wolpert 2008) and this represents a source of noise the brain needs to manage, one way or another, in order to guarantee reliable computation.

Neuronal noise

To isolate neuronal noise – the intrinsic noise due to neuronal biophysics – from the other sources of variability, an experiment similar to that described above can be performed *in vitro*, replacing the time-varying visual stimulus by the injection of a time-varying current directly into the soma of the neuron, thereby fully controlling the input received by the neuron. While these *in vitro* experiments show reduced spike time variability compared to the *in vivo* experiment mentioned above, spike times are still not fully reproducible (Bryant and Segundo 1976; Harsch and Robinson 2000; Mainen and Sejnowski 1995; Schreiber et al. 2004), indicating the presence of neuronal noise. The main source of neuronal noise, at least in these *in vitro* experiments where the input is fully controlled, is channel noise: the stochastic opening and closing of voltage-gated ion channels. These channels maintain the resting membrane potential of neurons (the membrane potential between spikes) and, as was famously shown
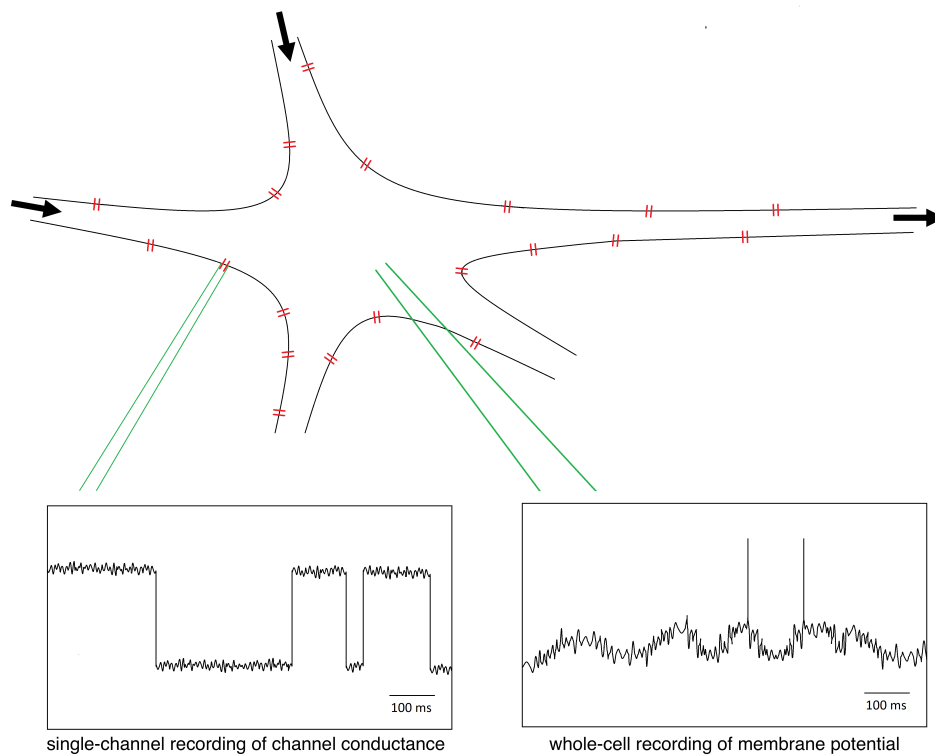
Figure 1.1: **Schematic drawing of neuronal noise.** Voltage-gated ion channels (red) on the cell membrane (black) of a neuron. Arrows indicate the flow of information in a neuron, arriving at the dendrites (incoming arrows) to sent out through the axon (outgoing arrow). (Bottom-left panel) Stochastic opening and closing of a voltage-gated ion channel (channel noise) seen in single-channel recordings. (Bottom-right panel) Membrane potential fluctuations seen in whole-cell recordings.

by Hodgkin and Huxley, the kinetics of these channels explains the shape of spikes (Hodgkin and Huxley 1952). Later, single-channel recordings (Sakmann and Neher 1984) revealed the stochastic nature of these channels; they randomly switch between both open and closed states, the transition probabilities between these two states being modulated by the membrane potential (White, Rubinstein, and Kay 2000) (Fig. 1.1, bottom-left panel). In contrast with the giant axon of the squid studied by Hodgkin and and Huxley where the effect of channel noise could be neglected because of the large number of channels, in the mammalian brain, neurons are smaller and have therefore fewer channels (White, Klink, et al. 1998), leading to finite-size fluctuations of the membrane potential significant enough to trigger spikes (Chow and White 1996; Schneidman, Freedman, and Segev 1998; Strassberg and DeFelice 1993) (Fig. 1.1, bottom-right pannel). Given the nonlinearity of single-neuron membrane potential dynamics – numerical simulations of the deterministic Hodgkin-Huxley model show chaotic membrane dynamics (Aihara, Matsumoto, and Ikegaya 1984; Chay and Rinzel 1985; Guckenheimer and Oliva 2002) – one can easily understand how even a small amount of channel-noise-induced membrane potential fluctuations can make the spike times of a neuron unreproducible.

Channel noise is most likely a serious biophysical constraint shaping the design of neural circuits. For example, channel noise imposes a theoretical lower bound on the diameter of unmyelinated axons below which reliable information transmission becomes impossible (Faisal, White, and Laughlin 2005). While the effect of channel noise on the membrane can theoretically vanish when the number of channels tends to infinity (Pakdaman, Thieullen, and Wainrib 2010), the number of channels per neuron is probably constrained by their metabolic cost: the more channels, the more ions cross the membrane, which then need to be transported back across the membrane via energy-consuming sodium-potassium exchanging pumps (Laughlin, Ruyter van Steveninck, and Anderson 1998). In addition to channel noise, there are other important sources of noise in the brain (Faisal, Selen, and Wolpert 2008), like synaptic noise, but they will not be considered in this thesis, the important point being that neurons are, to a certain extent, unreliable because their intrinsically noisy biophysics. In this thesis, we will not work with detailed biophysical models of neuronal noise; instead, we will consider simpler, phenomenological models which are much easier to manipulate in the context of our main question: how can neuronal noise be tamed in large networks?

1.1.3 Noise-robust dynamics as an emergent behavior

Our working hypothesis is that reliable computation, supported by noise-robust dynamics, is an emergent behavior in networks of noisy neurons. To understand how noise-robust dynamics can emerge from noisy networks, our strategy is to propose tractable mathematical models of networks of neurons where this idea of emergent noise-robust dynamics can be clearly articulated.

1.2 Mathematical framework

Deterministic biophysical models of single-neuron membrane dynamics, like the Hodgkin-Huxley model, can be approximated by deterministic threshold models (Brette and Gerstner 2005; Jolivet, Lewis, and Gerstner 2004; Kistler, Gerstner, and van Hemmen 1997) (see also Gerstner, Kistler, et al. (2014, Chapters 4 and 5)), which are mathematically more tractable. Roughly speaking, in deterministic threshold models, e.g., the leaky integrate-and-fire neuron model (Knight 1972a,b; Stein 1965), a neuron emits a spike, whose shape is approximated by a Dirac pulse, when its membrane potential reaches a threshold; just after the spike, the membrane potential can be reset to another value, introducing a discontinuous jump (see Gerstner, Kistler, et al. (2014, Chapter 5)). There are then two approaches for adding neuronal noise stemming from channel noise to these deterministic models:

- (Membrane noise) Noise can be added to the membrane potential dynamics, in the form of diffusive noise for example (see Tuckwell (1988)).

- ("Escape noise") The *hard* threshold of the neuron can be replaced by a *soft* threshold: spike emission is probabilistic and the instantaneous probability for a neuron to emit a

spike at a given time (its "escape rate") is given by a monotonically increasing function of its membrane potential, this function representing the "soft threshold" (see Gerstner, Kistler, et al. (2014, Chapter 9)).

Membrane noise should not be confused with "input noise," i.e., the membrane potential fluctuations caused by the (stochastic) inputs the neuron receives, which can also be modeled as diffusive noise (Brunel 2000; Brunel and Hakim 1999; Tuckwell 1988) but is not *intrinsic* neuronal noise (see also Gerstner, Kistler, et al. (2014, Chapter 8)). The diffusive noise of membrane (or input) noise can be mapped to "escape noise," but this involves solving difficult first-passage-time problems (see Plesser and Gerstner (2000), Sacerdote and Giraudo (2013), and Schwalger (2021) and references therein). In this thesis, all the models considered will directly use escape noise which is mathematically more tractable and which will allow us to focus our efforts on the effects of neuronal noise in large networks and avoid first-passage-time problems.

Spiking neuron models with escape noise are nothing but history-dependent, intensity-based point processes, as I will explain next. These neuron models have different names in the literature for they are at the crossroads of several branches of neuroscience. In theoretical neuroscience, spiking neuron models with escape noise are accurate yet tractable models of *in vitro* single-neuron spike activity (Jolivet, Rauch, et al. 2006; Pozzorini et al. 2015). As such, they can be seen as effective simplifications of biophysical models (see Gerstner and Kistler (2002) and Gerstner, Kistler, et al. (2014)). In statistical neuroscience, the same models where proposed, under the name of Generalized Linear Model-Point Processes, as *statistical* models of neuronal spike activity *in vivo* (Paninski 2004; J. W. Pillow et al. 2008; Truccolo et al. 2005). Hence, these models are both mechanistic models of single-neuron dynamics *and* convenient statistical models of multi-neuronal spike activity. Finally, in mathematical neuroscience, these models have rigorous probabilistic constructions, which greatly facilitates their mathematical analysis, as I will review below.

1.2.1   Modeling neuronal noise with Poisson random measures

Poisson random measures (Kingman 1992) are somewhat abstract mathematical objects – little-known in the computational neuroscience community – which turn out to be very convenient for formulating and studying models of networks of spiking neurons with escape noise. Because common confusions about spiking neurons with escape noise can be avoided once they are formulated in terms of stochastic processes driven by Poisson random measures, I give here a brief and very informal introduction to Poisson random measure for non-probabilists.

A Poisson process on $\mathbb{R}^2$ (the two-dimensional Euclidean space) with unit intensity can be informally defined as a collection of random points in $\mathbb{R}^2$ satisfying the following properties. Let π be a "function" which takes, as inputs, rectangles in $\mathbb{R}^2$ (see S and S' in Fig. 1.2) and returns the number of points contained in the rectangle, for any realization of the Poisson process. A collection of random points in $\mathbb{R}^2$ is a Poisson process with unit intensity if
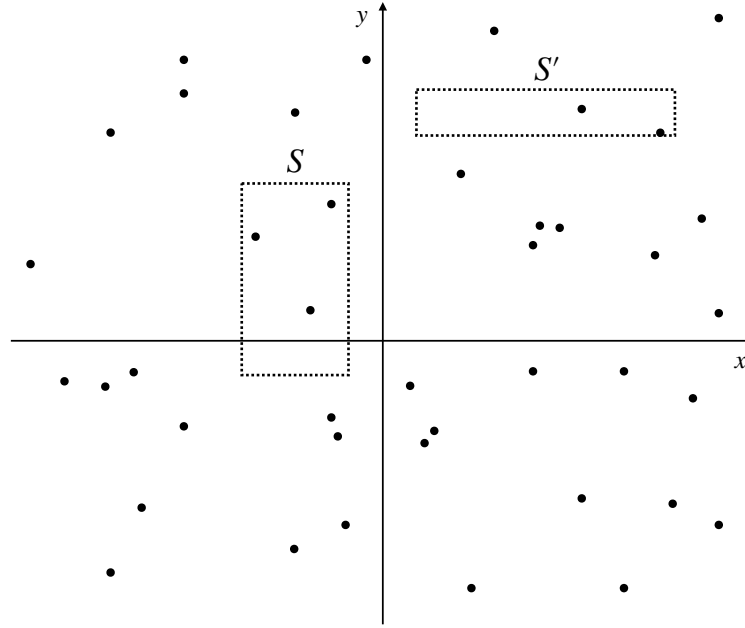
Figure 1.2: **A realization of a Poisson process in $\mathbb{R}^2$.**

1. For any rectangle S in $\mathbb{R}^2$, the number of points contained in the rectangle, $\pi(S)$, has the same law as a Poisson random variable with mean given by the area of the rectangle S, i.e.,

   $$\pi(S) \sim \text{Pois}(\text{Area}(S)).$$

2. For any finite collection of pairwise disjoint rectangles $S, S', \ldots$ in $\mathbb{R}^2$, the random variables $\pi(S), \pi(S'), \ldots$ are independent.

This two-dimensional Poisson process allows us to define a Poisson random measure on $\mathbb{R}^2$ (with unit intensity) which is nothing but π. The Poisson random measure π is said to have unit intensity (or Lebesgue intensity measure) because, for any rectangle $S \in \mathbb{R}^2$,

$$\mathbb{E}[\pi(S)] = \int_{\mathbb{R}\times\mathbb{R}} \mathbb{1}_{(x,y)\in S} \, dx dy.$$

The formal definition of Poisson random measures, which can be found in (Kingman 1992), involves some measure theory jargon, which is probably the reason why this probabilistic object is almost unknown in computational neuroscience. The informal definition given above has the advantage of avoiding jargon, and is sufficient, at least at the intuitive level, to understand the formulation of the network models we will consider. In the following, I will show how Poisson random measures enable us to write down equations defining spiking neuron models with escape noise.

As previously mentioned, spiking neuron models with escape noise have different names in different fields, which complicates the integration of theoretical results. To address this nomenclature problem and to highlight the equivalence of models behind different names, in the titles of the following subsections, I first write the theoretical neuroscience name, followed, in parenthesis, by the mathematical name of the model. Since this thesis is interested in large networks of neurons, I will directly present the neuron models as elements of recurrent networks.

Linear-Nonlinear-Poisson neurons (or *nonlinear Hawkes processes*)

Let us consider a network of N interacting Linear-Nonlinear-Poisson neurons (Chichilnisky 2001; Ostojic and Brunel 2011; Simoncelli et al. 2004) or, equivalently, nonlinear Hawkes processes (Brémaud and Massoulié 1996). The model is parametrized by two sets of functions. First, each neuron $i = 1, \dots, N$ has an intensity function $\phi_i : \mathbb{R} \to \mathbb{R}_+$, a positive-valued, monotonically increasing, nonlinear function.[I] Second, the effect of a spike of neuron $j \neq i$ on neuron i is captured by the interaction function $h_{ij} : \mathbb{R}_+ \to \mathbb{R}$.[II] Let $\{\pi^i\}_i$ be a collection of independent Poisson random measures with unit intensity as defined above. Then, we can describe the stochastic dynamics of the recurrent network as the solution of a system of stochastic integral equations: for all $i = 1, \dots, N$,

$$X_t^i = \int_0^t \sum_{j \neq i} h_{ij}(t - s) dZ_s^j, \tag{1.1a}$$

$$Z_t^i = \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \leq \phi_i(X_{s-}^i)} \pi^i(ds, dz). \tag{1.1b}$$

In Eq. (1.1b), X_{s-}^i is a shorthand for the left limit, $\lim_{\substack{u \to s \\ u < s}} X_u^i$. The compact expression (1.1) contains all the information necessary to define the network model. The variable Z_t^i counts the number of spikes of neuron i in the time interval $]0, t]$ and the (instantaneous) probability for neuron i to emit a spike in the infinitesimal interval $]t, t + dt]$ is $\phi_i(X_t^i) dt$, i.e,

$$\mathbb{P}(Z_{t+dt}^i > Z_t^i \,|\, \text{past until time } t) = \phi_i(X_t^i) dt.$$

In theoretical neuroscience, it is common to represent the spike activity of a neuron by a sum of Dirac δ-functions, called the spike train (see (Gerstner, Kistler, et al. 2014, Chapter 1)). The spike train of neuron i is written $S_i(t) = \sum_k \delta(t - t_i^k)$ where the $\{t_i^k\}_k$ are the spike times of neuron i. Then, formally, we have that

$$S_i(t) = \frac{dZ_t^i}{dt},$$

which links the "probabilist" notations of Eq. (1.1) to the "physicist" notations of theoretical neuroscience. For physicists, the meaning of the integral on the right hand side of Eq. (1.1a)

[I]which we can assume to be bounded and Lipschitz continuous

[II]which we can assume to be continuous and integrable

should become clear once it is replaced by $\int_0^t \sum_{j \neq i} h_{ij}(t-s)S_i(s)ds$. One important advantage of the probabilist notations (1.1) over the traditional physicist notations is that it allows to formulate recurrent networks models in terms of simple stochastic equations, which facilitates the mathematical analysis of the models.

The name "Linear-Nonlinear-Poisson neuron" comes from the fact that in Eq. (1.1), X_t^i is a *linear* function of spike activity of the other neurons, X_t^i defines the firing rate (the intensity) of the neuron through the *nonlinear* function ϕ_i, and $\phi_i(X_t^i)$ is the time-varying rate (intensity) of an inhomogeneous, one-dimensional *Poisson* process. Networks of Linear-Nonlinear-Poisson neurons are probably the simplest models for studying the effect of neuronal noise on the dynamics of large networks. However, they lack realistic single-neuron dynamics. For example, they neglect the effect of a neuron's past spike history on its own spike activity. In particular, real neurons exhibit refractoriness: after spiking, a neuron has to recover before it can emit another spike. Below, I present two closely related models which include refractoriness, extending the Linear-Nonlinear-Poisson model.

Spike Response Model-0 neurons (or *age-dependent nonlinear Hawkes processes*)

The effect of refractoriness can be included if, for each neuron, we keep track of its *age*, i.e., the time-elapsed since its last spike. The age variable A_t^i of neuron i can then influence the neuron's firing rate (intensity), e.g., the firing rate of neuron i can be given by the positive-valued function $f_i(A_t^i, X_t^i) = \phi_i(\eta_i(A_t^i) + X_t^i)$, where the function $\eta_i : \mathbb{R}_+ \to \mathbb{R}$ defines the refractoriness of neuron i (and ϕ_i is defined as in Sec. 1.2.1). Using the same formalism as in Sec. 1.2.1, we get: for all $i = 1, \dots, N$,

$$dA_t^i = dt - A_{t-}^i dZ_t^i, \tag{1.2a}$$

$$X_t^i = \sum_{j \neq i} \int_0^t h_{ij}(t-s)dZ_s^j, \tag{1.2b}$$

$$Z_t^i = \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \leq f_i(A_{s-}^i, X_{s-}^i)} \pi^i(ds, dz). \tag{1.2c}$$

Equation (1.2) defines a network of Spike Response Model-0 neurons (Gerstner 2000), also called age-dependent Hawkes processes in the probability theory literature (Chevallier 2017; Raad, Ditlevsen, and Löcherbach 2020). Equation (1.2a) simply means that the age variable A_t^i grows linearly with time and is reset to 0 at each spike of neuron i. The sum $\eta_i(A_t^i) + X_t^i$ in $\phi_i(\eta_i(A_t^i) + X_t^i)$ can now be interpreted as the membrane potential of neuron i (Gerstner 1995) and the nonlinear function ϕ_i can be accurately estimated from *in vitro*, single-neuron experiments (Jolivet, Rauch, et al. 2006). The Spike Response Model-0 is closely related, but not exactly equivalent, to integrate-and-fire models, as shown below.

Leaky Integrate-and-Fire neurons with escape noise (or *Galves-Löcherbach neurons*)

Inspired by the early work of Lapique (1907), leaky integrate-and-fire models were introduced by Stein (1965) and Knight (1972a,b) (see Brunel and Van Rossum (2007) for a short historical note) and escape noise was added to these models by Gerstner (1995, 2000). Motivated by the goal of giving a rigorous probabilistic construction of a network model with an infinite number of stochastic spiking neurons, Galves and Löcherbach (2013) proposed a infinite-dimensional, discrete-time model, later generalized to continuous time (Galves and Löcherbach 2016), which can be seen as an infinite network of leaky integrate-and-fire neurons with escape noise.

In networks of Leaky Integrate-and-Fire neurons with escape noise, the membrane potential (or voltage) U_t^i of neuron i integrates, with some leakage, the inputs it receives from other neurons until it emits a spike; after each spike of neuron i, the potential U_t^i is reset to some fixed value, forgetting all previously integrated inputs. The firing rate (the intensity) of neuron i is again defined by the membrane potential U_t^i via the nonlinear intensity function ϕ_i. Each time neuron $j \neq i$ emits a spike, the membrane potential U_t^i of neuron i makes a jump of height J_{ij}, which can either be positive (for an excitatory synapse) or negative (for an inhibitory synapse). For a finite network of N neurons, the model reads: for all $i = 1, \ldots, N$,

$$dU_t^i = \frac{\mu_i - U_t^i}{\tau_i} dt - U_{t-}^i dZ_t^i + \sum_{j \neq i} J_{ij} dZ_t^j, \tag{1.3a}$$

$$Z_t^i = \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \leq \phi_i(U_{s-}^i)} \pi^i(ds, dz). \tag{1.3b}$$

The parameters $\mu_i \in \mathbb{R}$ and $\tau_i > 0$ are, respectively, the resting potential and the membrane time constant of neuron i. On the right hand side of Eq. (1.3a), the term $\frac{\mu_i - U_t^i}{\tau_i} dt$ is the leaky part of the membrane dynamics and term $-U_{t-}^i dZ_t^i$ resets the membrane potential to 0 after each spike.

Note that the voltage-dependent model Eq. (1.3a) can be translated into an age-dependent model (Gerstner 1995, 2000). Indeed, solving the differential equation Eq. (1.3a), we get: for all $i = 1, \ldots, N$,

$$dA_t^i = dt - A_{t-}^i dZ_t^i, \tag{1.4a}$$

$$U_t^i = \mu \left(1 - e^{-\frac{A_t^i}{\tau_i}} \right) + \sum_{j \neq i} \int_{t - A_t^i}^{t} e^{-\frac{t-s}{\tau_i}} J_{ij} dZ_s^j, \tag{1.4b}$$

$$Z_t^i = \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \leq \phi_i(U_{s-}^i)} \pi^i(ds, dz). \tag{1.4c}$$

From the single-neuron modeling point of view, the main difference between the Spike Response Model-0 (1.2) and the Leaky Integrate-and-Fire model with escape noise (1.4) is that in the latter, a neuron forgets all the spike activity of the other neurons which has occurred before its last spike time (see the integral on the right hand side of Eq. (1.4b)).

1.2.2 Neuronal population equations

To study the effect of noise in large networks, the stochastic network models of the previous sections are particularly interesting because they all have *exact* mean-field limits where the dynamics of large networks can be described by deterministic neuronal population equations. Moreover, the convergence of the dynamics of large networks to neuronal population equation can be rigorously proven using techniques from interacting particle systems (Sznitman 1991). These exact mean-field limit results were conjectured in theoretical neuroscience (Gerstner 1995, 2000; Gerstner and van Hemmen 1992; Wilson and Cowan 1972) and were later proven in a series mathematical papers (Chevallier 2017; Chevallier, Duarte, et al. 2019; De Masi et al. 2015; Delattre, Fournier, and Hoffmann 2016; Fournier and Löcherbach 2016; Cristóbal Quiñinao 2016).

Networks of ***Linear-Nonlinear-Poisson neurons*** have the simplest mean-field limits. For example, let us assume that all the neurons in Eq. (1.1) are identical, i.e., $\phi_i = \phi$ for all i, and that the interactions are homogeneous and scaled by $1/N$, i.e., $h_{ij} = \frac{1}{N}h$ for all $i \neq j$. Then, when N tends to infinity, all the variables X_t^i concentrate around a single value $x(t)$ which solves the simple neuronal population equation:

$$x(t) = \int_0^t h(t-s)\phi(x(s))\,ds. \tag{1.5}$$

Importantly, Eq. (1.5) is deterministic which means, in a certain sense, that the dynamics of the microscopic system (1.1), taken as a whole becomes deterministic in the mean-field limit $N \to \infty$. This constitutes an elementary example of how neuronal noise can be tamed in large networks. Equation (1.5) is what is called a *neural mass equation*. These equations have a long history in theoretical neuroscience (Jansen and Rit 1995; Wilson and Cowan 1972) and are widely used in neuroimaging to model the dynamics of populations of neurons (Breakspear 2017; Coombes 2010; Deco et al. 2008). The convergence of the microscopic system (1.1) to the mean-field limit (1.5) was proved by Delattre, Fournier, and Hoffmann (2016), who showed how the coupling method of Sznitman (1991) could be applied to systems of interacting point processes. Building on this rigorous result, Chevallier, Duarte, et al. (2019) proved that when the network Eq. (1.1) is spatially structured, it converges, in the mean-field limit, to the neural field equation of Wilson and Cowan (1973) (see (Chevallier and Ost 2020) for the analysis of the limit fluctuations).

In the case of networks of ***Spike Response Model-0 neurons***, the mean-field limit neuronal population equation is richer than the neural mass equation (1.5). Again, let us assume that all the neurons in Eq. (1.2) are identical, i.e., $f_i = f$ and for all i, and that the interactions are homogeneous and scaled by $1/N$, i.e., $h_{ij} = \frac{1}{N}h$ for all $i \neq j$. As in the previous example, when N tends to infinity, all the variables X_t^i concentrate around a single value $x(t)$. By contrast, the age variables A_t^i do not concentrate and have to be described with a density ρ_t (on $\mathbb{R}_+$). In the

mean-field limit $N \to \infty$, the system Eq. (1.2) can be described by the age-structured equation

$$\partial_t \rho_t(a) + \partial_a \rho_t(a) = -f(a, x(t)) \rho_t(a) + r(t) \delta(a), \tag{1.6a}$$

$$r(t) = \int_0^\infty f(a, x(t)) \rho_t(a) \, da, \tag{1.6b}$$

$$x(t) = \int_0^t h(t-s) r(s) \, ds, \tag{1.6c}$$

$$\rho_0(a) = v_0(a), \tag{1.6d}$$

where v_0 is the initial age density at time 0. The partial differential equation (1.6) is known under the names of refractory density equation (Dumont, Pérez-Cervera, and Gutkin 2022; Gerstner, Kistler, et al. 2014; Schwalger and Chizhov 2019) and time-elapsed neuron network model (Pakdaman, Perthame, and Salort 2010, 2013, 2014). The convergence of the microscopic system (1.2) to the mean-field limit (1.6) was proved in Chevallier (2017) and Cristóbal Quiñinao (2016) (see Chevallier et al. (2017) for the analysis of the limit fluctuations). The long time behavior of this nonlinear equation has been rigorously analysed in Cañizo and Yoldaş (2019), Mischler, Cristobal Quiñinao, and Weng (2018), Mischler and Weng (2018), and Pakdaman, Perthame, and Salort (2010, 2013) (see also Torres, Perthame, and Salort (2022) for a recent generalization of the model Eq. (1.6)).

In the case of networks of ***Leaky Integrate-and-Fire neurons with escape noise***, we obtain, in the mean-field limit, a neuronal population equation which is similar to Eq. (1.6). Again, if all the neurons in Eq. (1.3) are identical, i.e., $\phi_i = \phi$ and $\tau_i = \tau$ for all i, and if the interactions are homogeneous and scaled by $1/N$, i.e., $J_{ij} = \frac{1}{N} J$ for all $i \neq j$, then, in the mean-field limit $N \to \infty$, the voltage variables U_t^i can be summarized by a density ρ_t (on $\mathbb{R}$) solving the voltage-structured equation

$$\partial_t \rho_t(u) + \partial_u \left(\left(\frac{\mu - u}{\tau} + J r(t) \right) \rho_t(u) \right) = -\phi(u) \rho_t(u) + r(t) \delta(u), \tag{1.7a}$$

$$r(t) = \int_{\mathbb{R}} \phi(u) \rho_t(u) \, du, \tag{1.7b}$$

$$\rho_0(u) = v_0(u), \tag{1.7c}$$

where v_0 is the initial voltage density at time 0. The convergence of the microscopic system (1.3) to the mean-field limit (1.7) was proved in De Masi et al. (2015) and Fournier and Löcherbach (2016) (see Löcherbach (2022) for the analysis of the limit fluctuations). The long time behavior of this nonlinear equation, including the emergence of oscillations, has been rigorously studied in Cormier (2020) and Cormier, Tanré, and Veltz (2020, 2021).

While Eq. (1.6) is age-structured whereas Eq. (1.7) is a voltage-structured, these two nonlinear partial differential equations are relatively similar, as the voltage-to-age transformation, Eq. (1.4) shows. Actually, both the age-structured equation (1.6) and the voltage-structured equation (1.7) can be rewritten as a nonlinear integral equation for the population activity

(i.e., average firing rate of the population) $r(t)$ (Gerstner 2000):

$$r(t) = H^r(t) + \int_0^t \lambda^r(t|s) S^r(t|s) r(s) ds, \qquad \text{with } S^r(t|s) = \exp\left(-\int_s^t \lambda^r(s'|s) ds'\right), \qquad (1.8)$$

where, in the case of Spike Response Model-0 neurons, Eq. (1.6), we have

$$H^r(t) = \int_0^\infty f(a+t, x(t)) \, e^{-\int_0^t f(a+s, x(s)) ds} \, v_0(a) da,$$

$$\lambda^r(t|s) = f(t-s, x(s)),$$

$$x(t) = \int_0^t h(t-s) r(s) ds,$$

whereas, in the case of Leaky Integrate-and-Fire neurons with escape noise, Eq. (1.7), we have,

$$H^r(t) = \int_{\mathbb{R}} \phi\left(u e^{-\frac{t}{\tau}} + \mathrm{u}(t|0)\right) e^{-\int_0^t \phi\left(u e^{-\frac{s}{\tau}} + \mathrm{u}(s|0)\right) ds} \, v_0(u) du,$$

$$\lambda^r(t|s) = \phi(\mathrm{u}(t|s)),$$

$$\mathrm{u}(t|s) = \mu\left(1 - e^{-\frac{t-s}{\tau}}\right) + J \int_s^t e^{-\frac{t-s'}{\tau}} r(s') ds'.$$

The integral equation (1.8) clearly reveals the renewal-type structure (Cox 1962) of the both the age-structured model (1.6) and the voltage-structured model (1.7).

1.2.3 Alternative approach: membrane noise

At the beginning of Sec. 1.2, I have mentioned two different ways to model the effects of channel noise on spike time variability: membrane noise and escape noise. For *fully connected* networks of Leaky Integrate-and-Fire neurons with diffusive membrane noise, there are also exact mean-field neuronal population equations (see Delarue et al. (2015b) and Inglis and Talay (2015) for the proof of the mean-field limit), which are special cases of the Nonlinear Noisy Leaky Integrate-and-Fire model (NNLIF) (Caceres and Perthame 2014; Cáceres, Carrillo, and Perthame 2011; Carrillo, González, et al. 2013; Carrillo, Perthame, et al. 2015; Delarue et al. 2015a), which itself comes from the heuristic theory of Brunel and Hakim (1999) and Brunel (2000) for *sparsely connected* networks of deterministic neurons. There is a potential source of confusion here since the NNLIF population equation can be either seen as an exact mean-field limit or a heuristic approximation, depending on what microscopic system it is supposed to model: for fully connected networks of neurons with membrane noise (*intrinsic* neuronal noise), the NNLIF population equation is exact in the mean-field limit (Delarue et al. 2015b; Inglis and Talay 2015); on the contrary, for sparsely connected networks of deterministic neurons, the NNLIF population equation is only a heuristic approximation (Brunel 2000; Brunel and Hakim 1999).[III]

[III]Adding somewhat to the confusion, the fully connected network model studied in Delarue et al. (2015b) and Inglis and Talay (2015) was originally proposed by Ostojic, Brunel, and Hakim (2009) as a model for electrically coupled neurons via gap junctions.

1.2.4   A note on sparsely connected networks

To my knowledge, exact mean-field limit results for sparsely connected networks, such as the models of Brunel and Hakim (1999) and Brunel (2000), have remained elusive. Here is a likely explanation for the current lack of exact results. In sparsely connected networks, the limit weight of some synapses can be nonzero as $N \to \infty$. As a consequence, a single spike of a presynaptic neuron can have some effect on a postsynaptic neuron, even when $N \to \infty$. This prevents neurons from behaving like independent processes as in large networks (*propagation of chaos*), an essential property for deriving exact neuronal population equations.

1.3   Summary of the thesis

In Chapter 2 (**Mean-field limit of age- and leaky memory-dependent Hawkes processes**), I generalize the mean-field convergence proofs for networks of renewal-type neurons – neurons whose memory of their own past spike activity is restricted to their last spike time (e.g., the Spike Response Model-0, Section 1.2.1, and the Leaky Integrate-and-Fire model with escape noise, Section 1.2.1) – to the case where neurons have a memory of their own spike activity going beyond their last spike. This extended memory is necessary to model the effects of spike-frequency adaptation and short-term synaptic plasticity. I propose a general neuron model, the age- and leaky memory-dependent Hawkes process, which encompasses many existing models in the computational neuroscience literature. I prove that in the mean-field limit, large networks of age- and leaky memory-dependent Hawkes processes can be described by a multidimentional, nonlinear, nonlocal transport equation. This partial differential equation (PDE) generalizes the Time Elapsed Neuron Network Model, Eq. (1.6), of Pakdaman, Perthame, and Salort (2010) (an example a refractory density equation (Gerstner and Kistler 2002; Schwalger and Chizhov 2019)) and it cannot, in general, be reduced to a scalar nonlinear integral equation such as Eq. (1.8).

The proof builds on the work of Chevallier (2017), who used the coupling method of Sznitman (1991) adapted to interacting point processes by Delattre, Fournier, and Hoffmann (2016).

The important message of this chapter is that even when single neurons have a multi-timescale memory of their own activity, in the mean-field limit (when neurons are identical and when the interactions between neurons are homogeneous and scaled by $1/N$), the neuronal noise is absorbed as in the simpler renewal-type models of Section 1.2.2, namely, the dynamics of large networks can be described a deterministic neuronal population equation – the neuronal population equation is simply more complicated.

In Chapter 3 (**Long time behavior of and age- and leaky memory-dependent neuronal population equation**), with Claudia Fonte, we show that, despite its complicated appearance, the neuronal population equation derived in Chapter 2 can be rigorously analyzed from a deterministic, PDE point of view. We focus on the two-dimensional case (each neuron has an age variable and a single leaky memory variable), which is sufficient to model the effect

of spike-frequency adaptation or short-term synaptic depression. We prove that in the weak connectivity regime and for general initial conditions, the neuronal population equation converges to a unique stationary solution, at exponential speed.

The proof combines methods from Cañizo and Yoldaş (2019) and Mischler and Weng (2018), the general strategy being to use recent deterministic versions of Doeblin-Harris/Meyn-Tweedie-type theorems for the stability of Markov processes (Cañizo and Mischler 2021).

While, in the weak connectivity regime, the network is in an asynchronous state, we also show numerically that for stronger connectivity, the equation can exhibit self-sustained population bursts, a long time behavior that remains to be rigorously studied.

In Chapter 4 (**On a finite-size neuronal population equation**), with Eva Löcherbach and Tilo Schwalger, we study the heuristic stochastic neuronal population equation for finite-size networks of renewal-type neurons derived by Schwalger, Deger, and Gerstner (2017). While not exact, this heuristic stochastic equation is very useful for the design of efficient simulation algorithms and, more importantly, data analysis algorithms, as we will show in Chapter 5. A crucial empirical feature of this stochastic equation is that it is stable over long simulation times but the reason for this stability was not theoretically studied in the original paper of Schwalger, Deger, and Gerstner (2017).

After establishing the well-posedness of the stochastic equation, we prove, on a slightly simplified version of the model, that the finite-size neuronal population equation is indeed stable. The proof involves arguments for the stability of Markov processes (taking values in the space of positive measures) and techniques introduced by Brémaud and Massoulié (1996) for the stability analysis of nonlinear Hawkes processes.

In Chapter 5 (**Mesoscopic modeling of hidden spiking neurons**), with Shuqi Wang, we show that mean-field models can be linked to latent variable models (or hidden Markov models) used to infer latent population dynamics from multi-neuronal recordings. In particular, we show, on synthetic data, that the finite-size neuronal population equation of Chapter 4 can be used design a tractable Expectation-Maximization-type algorithm capable of inferring the latent population activities of multi-population spiking neural networks from the spike activity of a few visible neurons only.

A link between mean-field models and latent variable models in neuroscience is, in hindsight, not surprising because both kinds of models are motivated by the hypothesis that the brain performs computation via the collective dynamics of populations of neurons approximating continuous dynamical systems (Gerstner, Kistler, et al. 2014; Vyas et al. 2020). In particular, both kinds of models share the idea that under the *visible* noisiness of single neuron spike activities, there are much less noisy *latent* variables driving the collective dynamics of populations of neurons.

In classical mean-field models, neurons in large networks behave like independent, identically

distributed processes driven by the average population activity – a deterministic quantity, by the law of large numbers. The fact the neurons are identically distributed processes implies a form of redundancy that has not been observed in the cortex and which seems biologically implausible. To show numerically that the redundancy present in classical mean-field models is unnecessary for neuronal noise absorption in large networks, in Chapter 6 (**Convergence of redundancy-free spiking neural networks to rate networks**), I construct a disordered network model where networks of spiking neurons behave like deterministic rate networks, despite the absence of redundancy. An informal argument then suggests that the mechanism responsible for noise absorption in these networks is the concentration of measure phenomenon, which quite naturally generalizes the "law of large numbers" of classical mean-field models. Based on these numerical results and an informal argument, I argue that the dual phenomena of *propagation of independence* (cf. Jabin, Poyato, and Soler (2021)) and *concentration of measure* in large networks might explain neuronal noise absorption, and thereby noise-robust population dynamics, in many situations beyond classical mean-field models.

The writing style and the level of mathematical rigor varies across chapters:
Chapters 2, 3, and 4 present rigorous mathematical results and are written for a mathematics audience; Chapter 5 is written for a machine learning audience; and Chapter 6 is written for a physics audience and contains several claims which are not supported by rigorous proofs.

2 Mean-field limit of age- and leaky memory- dependent Hawkes processes

Single author

Mean-field limit of age and leaky memory dependent Hawkes processes

Valentin Schmutz

Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

Abstract

We propose a mean-field model of interacting point processes where each process has a memory of the time elapsed since its last event (age) and its recent past (leaky memory), generalizing Age-dependent Hawkes processes. The model is motivated by interacting nonlinear Hawkes processes with Markovian self-interaction and networks of spiking neurons with adaptation and short-term synaptic plasticity.

By proving propagation of chaos and using a path integral representation for the law of the limit process, we show that, in the mean-field limit, the empirical measure of the system follows a multidimensional nonlocal transport equation.
© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

MSC: primary 60F05; secondary 35F15; 35F20; 60G55; 92B20

Keywords: Hawkes process; Mean-field approximation; Nonlocal transport equation; Propagation of chaos; Erlang kernel; Short-term synaptic plasticity

1. Introduction

The dynamics of many interacting particle systems can be approximated, when the size of the system tends to infinity, by a partial differential equation (PDE) [26]. This not only links microscopic and macroscopic scales but also stochastic and deterministic models. For mean-field models, one can prove this type of results by exploiting the propagation of chaos phenomenon, i.e. for *i.i.d.* initial conditions, particles become asymptotically independent in the mean-field limit [31,47].

Propagation of chaos arguments have been applied to the study of interacting point processes [9,15,29,30]. This has been particularly important for the field of theoretical neuroscience as it has provided a rigorous footing to the *population density* formalism, where the

dynamics of a population of neurons is described by a PDE (see [22, Part III]). An example of population density equation is the refractory density (or age-structured) equation [18,20–22, 36,37,46], which has recently been proved to be exact in the mean-field limit [4,8,42].

The models considered in [4,8,42] all assume that the point processes are 'renewal' (in some loose sense), i.e. each process has a memory of its past that is restricted to the time elapsed since its last event. The fact that, in the 'renewal' case, the mean-field limit can by characterized by relatively simple deterministic equations has long been recognized in theoretical neuroscience and has led to a large body of work [18,19,46,52]. In contrast, the case where point processes are not 'renewal' is much less understood. In particular, even though some heuristic population density equations have been proposed for the 'non-renewal' case [33,38,48], their exactness in the mean-field limit has not been proved. The aim of this work is therefore to propose a general framework for relating interacting 'non-renewal' point processes with PDEs, in the mean-field limit. This framework relies on the definition of an abstract interacting point process model, which generalizes Age-dependent Hawkes processes [4,43].

1.1. Interacting age and leaky memory dependent Hawkes processes

Consider a system of N interacting point-processes, interacting through a common variable X_t^N. Each point process i is associated with $1 + d$ variables (for d a positive integer): an *age* variable $A_t^{i,N}$ which represents the time elapsed since the last event of process i and a d-dimensional vector of *leaky memory* variables $\mathbf{M}_t^{i,N}$ which models the effect of the recent past of process i. The point process i has stochastic intensity $(f(A_{t-}^{i,N}, \mathbf{M}_{t-}^{i,N}, X_{t-}^N))_{t \in \mathbb{R}_+}$ where $f : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}_+$ is the intensity function. Intuitively, this means that if we write $(Z_t^{i,N})_{t \in \mathbb{R}_+}$ the counting process associated with the point process i, the instantaneous probability for $Z^{i,N}$ to jump in $]t, t + dt]$, given the past $\mathcal{F}_t$, is

$$\mathbb{P}(Z_{t+dt}^{i,N} > Z_t^{i,N} | \mathcal{F}_t) = f(A_t^{i,N}, \mathbf{M}_t^{i,N}, X_t^N) dt.$$

Between events (jumps) of process i, the *age* variable $A_t^{i,N}$ grows linearly with time whereas the *leaky memory* variables $\mathbf{M}_t^{i,N}$ drift following the vector field $b : \mathbb{R}^d \to \mathbb{R}^d$.

At each event of process i, its *age* $A_t^{i,N}$ is reset to 0 and its *leaky memory* $\mathbf{M}_t^{i,N}$ jumps to $\mathbf{M}_t^{i,N} + \Gamma(\mathbf{M}_t^{i,N})$, where $\Gamma : \mathbb{R}^d \to \mathbb{R}^d$ is the *jump function*. The fact that the variables $\mathbf{M}_t^{i,N}$ are not reset to a fixed value at each event allows them to accumulate the effect of successive events.

Finally, the time-dependent effect of an event of point process i on point process j is determined by the interaction function $h : \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ which depends on $A_t^{j,N}$ and $\mathbf{M}_t^{j,N}$. Since the function h is the same for all i and j, the interaction is said to be of mean-field type.

The model can be described by a system of stochastic integral equations: for $i = 1, \dots, N$,

$$A_t^{i,N} = A_0^i + t - \int_0^t A_{s-}^{i,N} dZ_s^{i,N}, \tag{1a}$$

$$\mathbf{M}_t^{i,N} = \mathbf{M}_0^i + \int_0^t b(\mathbf{M}_s^{i,N}) ds + \int_0^t \Gamma(\mathbf{M}_{s-}^{i,N}) dZ_s^{i,N}, \tag{1b}$$

$$Z_t^{i,N} = \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \leq f(A_{s-}^{i,N}, \mathbf{M}_{s-}^{i,N}, X_{s-}^N)} \pi^i(ds, dz), \tag{1c}$$

with

$$X_t^N = \frac{1}{N} \sum_{j=1}^{N} H_t^j + \frac{1}{N} \sum_{j=1}^{N} \int_0^t h(t - s, A_{s-}^{j,N}, \mathbf{M}_{s-}^{j,N}) dZ_s^{j,N}. \tag{1d}$$

The collection $\{\pi^i\}_{i\in\mathbb{N}^*}$ is a sequence of independent Poisson random measures on $\mathbb{R}_+ \times \mathbb{R}_+$ with Lebesgue intensity measure. We work on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbb{P})$ where $\{\pi^i\}_{i\in\mathbb{N}^*}$ is independent of $\mathcal{F}_0$ and $\mathcal{F}_t = \mathcal{F}_0 \cup \sigma\left(\{\pi^i([0,t], B)\}_{i\in\mathbb{N}^*, B\in\mathcal{B}(\mathbb{R}_+)}\right)$. For all $i \in \mathbb{N}^*$, A_0^i and $\mathbf{M}_0^i$ are $\mathcal{F}_0$-measurable random variables taking values in $\mathbb{R}_+$ and $\mathbb{R}^d$ respectively and $(H_t^i)_{t\in\mathbb{R}_+}$ is a $\mathcal{F}_0$-measurable $\mathcal{C}(\mathbb{R}_+)$ random function. The $1/N$ scaling in (1d) will allow us to take the mean-field limit $N \to \infty$.

If f does not depend on the leaky memory variables $\mathbf{M}$ and h does not depend on the age A nor $\mathbf{M}$, (1) reduces to a system of interacting Age-dependent Hawkes processes [4,43]. If, in addition, f does not depend on A, the model further reduces to a mean-field system of interacting nonlinear Hawkes processes (with vanishing self-interaction) [9,10]. The model (1) has two motivations: first, it is general enough to encompass several concrete examples from the theory of nonlinear Hawkes processes and neuroscience (see below); second, its mean-field limit can be characterized by a PDE.

1.2. Motivating examples

Hawkes processes [23] provide a flexible and intuitive model for point processes with dependence on the past. They have found applications in finance [1,24], seismology [35], social systems [6], genomics [45] and neuroscience [17,28,41,44,49,50], among other fields. Neuroscience research has mainly focused on *nonlinear* Hawkes processes [2] since they are closely related to well-established neuron models such as the Spike Response Model [19,21,22,25] and the Recursive Linear-Nonlinear Poisson Model [41], both variations of Generalized Linear Models (see [22] Part II and references therein). However, the models differ from the nonlinear Hawkes processes considered in [9,10] in that, even when N is large, self-interaction (the effect of process i on itself) does not vanish. Self-interaction vanishes in [9,10] because it is scaled by $1/N$. Let us now consider the case where self-interaction $\mathfrak{h}$ can be different from hetero-interaction h (the effect of a process on the other processes) and only hetero-interaction is scaled by $1/N$: for $i = 1, \ldots, N$,

$$Z_t^{i,N} = \int_{[0,t]\times\mathbb{R}_+} \mathbb{1}_{\{z \le f(X_{s-}^{i,N})\}} \pi^i(ds, dz), \tag{2a}$$

$$X_t^{i,N}(i) = \mathfrak{H}_t^i + \frac{1}{N} \sum_{j\neq i} H_t^j + \int_0^t \mathfrak{h}(t-s) dZ_s^{i,N} + \frac{1}{N} \sum_{j\neq i} \int_0^t h(t-s) dZ_s^{j,N}, \tag{2b}$$

where $f : \mathbb{R} \to \mathbb{R}_+$ is monotonically increasing and $\{(\mathfrak{H}_t(i))_{t\in\mathbb{R}_+}\}_{i=1,\ldots,N}$ are $\mathcal{F}_0$-measurable random $\mathcal{C}(\mathbb{R}_+)$ functions. The model (2) is a mean-field system of interacting nonlinear Hawkes processes with non-vanishing self-interaction. In the context of neuroscience, (2) can be seen as a mean-field network of Generalized Linear Model/Spike Response Model neurons.

Let us now assume that $\mathfrak{h}$ is an Erlang kernel, i.e. there exists $d \in \mathbb{N}^*$ such that

$$\mathfrak{h}(t) = ce^{-\alpha t} \frac{t^{d-1}}{(d-1)!},$$

for some $c \in \mathbb{R}$ and $\alpha > 0$. Self-interaction can then be translated into a 'Markovian cascade of memory terms' [10,11]: adjusting the initial conditions, (2) can be equivalently written: for $i = 1, \ldots, N$,

$$Z_t^{i,N} = \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{\{z \leq f(\mathbf{M}_{s-}^{i,N}(1) + X_{s-}^{i,N})\}} \pi^i(ds, dz), \tag{3a}$$

$$\mathbf{M}_t^{i,N} = \mathbf{M}_0^i + \int_0^t \mathfrak{A} \mathbf{M}_s^{i,N} ds + \mathbf{c} Z_t^{i,N}, \tag{3b}$$

$$X_t^{i,N} = \frac{1}{N} \sum_{j \neq i} H_t^j + \frac{1}{N} \sum_{j \neq i} \int_0^t h(t - s) dZ_s^{j,N}, \tag{3c}$$

where $\mathbf{M}_{s-}^{i,N}(1)$ denotes the first element of the vector $\mathbf{M}_{s-}^{i,N}$. The d-by-d matrix $\mathfrak{A}$ has all diagonal terms equal to $-\alpha$, all superdiagonal terms equal to 1, and all other terms equal to 0; the d dimensional vector $\mathbf{c}$ is defined by $\mathbf{c}(k) = \mathbb{1}_{k=d} c$. This is equivalent to setting $b(\mathbf{m}) = \mathfrak{A}\mathbf{m}$ and $\Gamma(\mathbf{m}) = (0, \ldots, 0, c)$ in (1b). The fact that there are N distinct variables $X_t^{i,N}$ instead of a common shared variable X_t^N as in (1) does not affect the mean-field limit since the difference between X_t^N and $X_t^{i,N}$ is of order $1/N$ (see [4]).

The generalization of (3) to the case where $\mathfrak{h}$ is a sum of Erlang kernels is straightforward. Of course, a sum of Erlang kernels can simply be a sum of exponential kernels, which is more common in neuroscience [22,33,34,48]. Notably, taking one exponential kernel with $c < 0$ is enough to model the effects of neuronal refractoriness and spike-frequency adaptation [33].

In the example (3), the leaky memory variables $\mathbf{M}$ influence the intensity function f but does not influence the interaction function h. However, in the general model (1), h can depend on $\mathbf{M}$. In the context of neuronal modeling, this dependence can be used to account for the effects of short-term synaptic plasticity (STP) [53]. Using the notation of the general model (1), we can describe a network of spiking neurons with refractoriness and 'Tsodyks–Markram' STP [51]. The Tsodyks–Markram model [51] captures the interplay between synaptic depression and facilitation and has been used to model working memory [32], chaotic dynamics [5] and learning in hierarchical circuits [39]. Taking $d = 2$, the leaky memory variables (the STP variables of the Tsodyks–Markram model) follow, for initial conditions $\mathbf{M}_0^i$ supported in $[U, 1] \times [0, 1]$ with $U \in]0, 1[$,

$$\mathbf{M}_t^{i,N} = \mathbf{M}_0^i + \int_0^t b(\mathbf{M}_s^{i,N}) ds + \int_0^t \Gamma(\mathbf{M}_{s-}^{i,N}) dZ_s^{i,N}, \tag{4a}$$

with the vector field

$$b\left(\mathbf{m}(1), \mathbf{m}(2)\right) = \left(\frac{U - \mathbf{m}(1)}{\tau_F}, \frac{1 - \mathbf{m}(2)}{\tau_D}\right), \tag{4b}$$

where $\tau_F > 0$ and $\tau_D > 0$ are the facilitation and depression timescales respectively, and with the jump function

$$\Gamma\left(\mathbf{m}(1), \mathbf{m}(2)\right) = \left(U(1 - \mathbf{m}(1)), -\mathbf{m}(1)\mathbf{m}(2)\right). \tag{4c}$$

It is easy to verify that the leaky memory variables $\mathbf{M}_t^{i,N}$ then take values in $[U, 1] \times [0, 1]$. Finally, we take f independent of the leaky memory variables and h of the form $h(t, a, \mathbf{m}) = \mathbf{m}(1)\mathbf{m}(2)\bar{h}(t)$. The model we just described generalizes interacting Age-dependent Hawkes processes [4,43] and is more detailed than the model with purely facilitating synapses and without refractoriness studied in [16].

These two motivating examples are clearly special cases of the general model (1). The fact that in both examples, the variables $\mathbf{M}_t^{i,N}$ relax to some fixed value in the absence of jumps motivates the name 'leaky memory'. Importantly, both examples satisfy the main assumptions we will use in this work (see Section 2).

1.3. Methods and relation to previous work

To prove propagation of chaos in the mean-field limit, we use the method of coupling *à la Sznitman* [47]: to show the convergence of the time-marginals, we follow Fournier and Löcherbach [14] (see also [16]); to show the convergence of the processes, we use the method of Delattre, Fournier and Hoffmann [9] (later used by Chevallier [4] and Ditlevsen and Löcherbach [10]). Our approach for relating the limit process with the limit PDE differs from previous work [4] for it relies on a path integral representation. This representation turns an earlier heuristic method from Naud and Gerstner [34] into a rigorous argument. Contrarily to [4] where PDE solutions in measure space are considered, our method treats PDE solutions in L^1 space and does not involve the semigroup theory results of [3]. More importantly, the path integral method allows us to derive a representation formula for the solution to the PDE. The limit PDE we obtain is a generalization of the Time Elapsed Neuron Network Model of Pakdaman, Perthame and Salort [36] and of the refractory density equation [18,20,22] to the case of neurons with adaptation and short-term synaptic plasticity.

1.4. Plan of the paper

The main results of this work, namely propagation of chaos (Theorem 1) and the characterization of the mean-field limit by a PDE (Theorem 2), are presented in Section 2, together with the assumptions required. The proof of Theorem 1 is presented in Section 3. In Section 4, we show that under more restrictive assumptions, we can get a propagation of chaos result analogous to that of [4,9,10]. Finally, the proof of Theorem 2 is presented in Section 5.

2. Assumptions and main results

General notations. The uniform and Euclidean norms are denoted by $\|\cdot\|_\infty$ and $\|\cdot\|$ respectively. We write $\|\Gamma\|_\infty := \sup_{\mathbf{m}\in\mathbb{R}^d} \|\Gamma(\mathbf{m})\|$. We use C, C_T and $C_{T,0}$ to denote positive constants (that can change from line to line) where the subscript T signals the dependence on time and the subscript 0 the dependence on the law of $\left(A_0^1, \mathbf{M}_0^1, (H_t^1)_{t\in\mathbb{R}_+}\right)$.

In this work, we always assume that the functions f, h, b and Γ satisfy:

Assumption 1.

(i) The functions f, h and Γ are bounded.

(ii) There exists a bounded, strictly increasing and continuously differentiable function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ with $\psi(0) = 0$ and satisfying

$$|\psi'(a) - \psi'(a^*)| \le \kappa|\psi(a) - \psi(a^*)|, \qquad \forall a, a^* \in \mathbb{R}_+,$$

for some $\kappa > 0$, such that, for all $(a, \mathbf{m}, x, a^*, \mathbf{m}^*, x^*) \in (\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R})^2$, and for all $t \in \mathbb{R}_+$,

$$|f(a, \mathbf{m}, x) - f(a^*, \mathbf{m}^*, x^*)| \le L_f(|\psi(a) - \psi(a^*)| + \|\mathbf{m} - \mathbf{m}^*\| + |x - x^*|),$$
$$|h(t, a, \mathbf{m}) - h(t, a^*, \mathbf{m}^*)| \le L_h(|\psi(a) - \psi(a^*)| + \|\mathbf{m} - \mathbf{m}^*\|),$$

for some L_f and $L_h > 0$.

(iii) The vector field b and jump function Γ are Lipschitz continuous.

The fact that f is bounded guarantees the well-posedness of the system (1) and a path-wise unique càdlàg strong solution to (1) can be constructed using a standard thinning procedure. (ii) says that f and h are Lipschitz continuous with respect to a ψ-modified metric on the *age* variable. An example of possible function is $\psi(a) = 1 - \exp(-a\kappa)$. The ψ-modified Lipschitz continuity of f implies that the effect of a on f saturates for large a. Note that in the STP example (4), since the leaky memory variables take values in the compact $[U, 1] \times [0, 1]$, the jump function (4c) is effectively Lipschitz.

To prove propagation of chaos, we need some assumptions on $\{(A_0(i), \mathbf{M}_0(i), (H_t(i))_{t\in\mathbb{R}_+})\}_{i\in\mathbb{N}}$:

Assumption 2.

(i) The 3-tuples $\{(A_0(i), \mathbf{M}_0(i), (H_t(i))_{t\in\mathbb{R}_+})\}_{i\in\mathbb{N}}$ are *i.i.d.*
(ii) The random function $(H_t^1)_{t\in\mathbb{R}_+}$ is such that $(\mathbb{E}[H_t^1])_{t\in\mathbb{R}_+} \in \mathcal{C}(\mathbb{R}_+)$.
(iii) For all $T > 0$, there exists $C_{T,0} > 0$ such that $\sup_{t\in[0,T]} \mathrm{Var}[H_t^1] \leq C_{T,0}$.

A condition similar to (iii) is used in [4]. Note that in [4], Chevallier considers *i.i.d.* random interaction functions instead of a deterministic function h, common to all the point processes. As he proved that, under some square integrability condition, the randomness in the interaction functions averages out in the mean-field limit, we focus here on the fixed h case.

The first main result of this work is a quantified propagation of chaos theorem:

Theorem 1. *Grant Assumptions 1 and 2. For all $T > 0$, there exists $C_{T,0} > 0$ such that*

$$\sup_{t\in[0,T]} \mathbb{E}\left[|\psi(A_t^{1,N}) - \psi(A_t^1)| + \|\mathbf{M}_t^{1,N} - \mathbf{M}_t^1\| + |X_t^N - x_t|\right] \leq C_{T,0}N^{-1/2}, \tag{5}$$

where $(A_t^1, \mathbf{M}_t^1, x_t)_{t\in\mathbb{R}_+}$ (the limit process) is given by the path-wise unique càdlàg strong solution to

$$A_t^1 = A_0^1 + t - \int_0^t A_{s-}^1 dZ_s^1, \tag{6a}$$

$$\mathbf{M}_t^1 = \mathbf{M}_0^1 + \int_0^t b(\mathbf{M}_s^1)ds + \int_0^t \Gamma(\mathbf{M}_{s-}^1)dZ_s^1, \tag{6b}$$

$$Z_t^1 = \int_{[0,t]\times\mathbb{R}_+} \mathbb{1}_{z\leq f(A_{s-}^1, \mathbf{M}_{s-}^1, x_s)}\pi^1(ds, dz), \tag{6c}$$

$$x_t = \mathbb{E}[H_t^1] + \int_0^t \mathbb{E}[h(t - s, A_s^1, \mathbf{M}_s^1)f(A_s^1, \mathbf{M}_s^1, x_s)]ds. \tag{6d}$$

Furthermore, for all $t \in [0, T]$, writing $\mathcal{L}\left(\psi(A_t^1), \mathbf{M}_t^1\right)$ the law of $(\psi(A_t^1), \mathbf{M}_t^1)$, there exists $C'_{T,0} > 0$ such that

$$\sup_{t\in[0,T]} \mathbb{E}\left[W_1\left(\frac{1}{N}\sum_{i=1}^N \delta_{(\psi(A_t^{i,N}), \mathbf{M}_t^{i,N})}, \mathcal{L}\left(\psi(A_t^1), \mathbf{M}_t^1\right)\right)\right] \leq C'_{T,0}N^{-1/2}, \tag{7}$$

where W_1 denotes the 1-Wasserstein distance.

It directly follows from Theorem 1 and the Continuous mapping theorem that for all $t > 0$, the empirical measure of the system (1) at time t converges in probability to the time-marginal

of the law of the limit process:

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{(A_t^{i,N}, \mathbf{M}_t^{i,N}, X^N)} \xrightarrow[N\to\infty]{\mathbb{P}} \mathcal{L}\left(A_t^1, \mathbf{M}_t^1, x_t\right). \tag{8}$$

The second main result relates the time-marginals $\mathcal{L}\left(A_t^1, \mathbf{M}_t^1, x_t\right)$ of the law of the limit process with the solution to a nonlocal transport equation. To formulate the transport equation, we define the *jump mapping* $\gamma(\mathbf{m}) = \mathbf{m} + \Gamma(\mathbf{m})$ and we write $\nabla\cdot$ the divergence operator in $\mathbb{R}^d$. To stay within the standard framework of (mass-conservative) transport equation with solutions in L^1 [40], we need:

Assumption 3.

  (i) The vector field satisfies $b \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^d)$ and $\nabla \cdot b \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^d)$.
  (ii) The jump mapping γ is a proper local $\mathcal{C}^1$-diffeomorphism.

Theorem 2. *Grant Assumptions 1 and 3. Further assume that the law of the initial condition $(A_0^1, \mathbf{M}_0^1)$ is the absolutely continuous probability measure $u_0(a, \mathbf{m})da\,d\mathbf{m}$ and $(\mathbb{E}[H_t])_{t\in\mathbb{R}_+} = (\bar{H}_t^1)_{t\in\mathbb{R}_+} \in \mathcal{C}(\mathbb{R}_+)$. Then, the time-marginals $\rho_t \otimes \delta_{x_t} := \mathcal{L}\left(A_t^1, \mathbf{M}_t^1, x_t\right)$ of the law of the limit process (6) correspond to the unique weak solution to*

$$\partial_t \rho_t(a, \mathbf{m}) + \partial_a \rho_t(a, \mathbf{m}) + \nabla \cdot \left(b(\mathbf{m})\rho_t(a, \mathbf{m})\right) = -f(a, \mathbf{m}, x_t)\rho_t(a, \mathbf{m}), \tag{9a}$$

$$\rho_t(0, \cdot) = \gamma_*\left(\int_{\mathbb{R}_+} f(a, \cdot, x_t)\rho_t(a, \cdot)da\right), \tag{9b}$$

$$x_t = \bar{H}_t + \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}_+} h(t-s, a, \mathbf{m})f(a, \mathbf{m}, x_s)\rho_s(a, \mathbf{m})da\,d\mathbf{m}\,ds, \tag{9c}$$

$$\rho_0(a, \mathbf{m}) = u_0(a, \mathbf{m}), \tag{9d}$$

(where $\nabla\cdot$ denotes the divergence on the variables $\mathbf{m}$ and $\gamma_*(\dots)$ denotes the pushforward measure by γ) in the sense that $(\rho, x) \in \mathcal{C}(\mathbb{R}_+, L^1(\mathbb{R}_+ \times \mathbb{R}^d)) \times \mathcal{C}(\mathbb{R}_+)$ and, for all $G \in \mathcal{C}_c^\infty(\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^d)$,

$$0 = \int_{\mathbb{R}^d} \int_{\mathbb{R}_+} G(0, a, \mathbf{m})u_0(a, \mathbf{m})da\,d\mathbf{m} + \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \int_{\mathbb{R}_+} \left\{[\partial_t + \partial_a + b(\mathbf{m}) \cdot \nabla]G(t, a, \mathbf{m}) \right.$$

$$\left. +(G(t, 0, \gamma(\mathbf{m})) - G(t, a, \mathbf{m}))f(a, \mathbf{m}, x_t)\right\}\rho_t(a, \mathbf{m})da\,d\mathbf{m}\,dt, \tag{10}$$

where ∇ denotes the gradient operator on the variables $\mathbf{m}$.

Assumption 3 *(ii)* guarantees that for all $\mathbf{m} \in \gamma(\mathbb{R}^d)$, the preimage $\gamma^{-1}(\mathbf{m})$ is a finite set of points and (9b) can be more explicitly written

$$\rho_t(0, \mathbf{m}) = \mathbb{1}_{\gamma(\mathbb{R}^d)}(\mathbf{m}) \sum_{\mathbf{m}'\in\gamma^{-1}(\mathbf{m})} \frac{1}{|\det(\mathbf{J}_\gamma(\mathbf{m}'))|} \int_{\mathbb{R}_+} f(a, \mathbf{m}', x_t)\rho_t(a, \mathbf{m}')da, \tag{11}$$

where $\det(\mathbf{J}_\gamma(\mathbf{m}'))$ denotes the determinant of the Jacobian matrix $\mathbf{J}_\gamma(\mathbf{m}')$.

All the results and proofs can be adapted to the simpler case where the system is not age-dependent, as in the Erlang kernel example (3). For this example, the limit PDE (9) becomes

$$\partial_t \rho_t(\mathbf{m}) + \mathfrak{A}\nabla \cdot \left(\mathbf{m}\rho_t(\mathbf{m})\right) = f(\mathbf{m}(1) + x_t)\rho_t(\mathbf{m} - \mathbf{c}) - f(\mathbf{m}(1) + x_t)\rho_t(\mathbf{m}),$$

$$x_t = \bar{H}_t + \int_0^t h(t-s) \int_{\mathbb{R}^d} f(\mathbf{m}(1) + x_s)\rho_s(\mathbf{m})d\mathbf{m}ds,$$

$$\rho_0(\mathbf{m}) = u_0(\mathbf{m}).$$

For the STP example (4), the limit PDE reads

$$\partial_t \rho_t(a, \mathbf{m}) + \partial_a \rho_t(a, \mathbf{m}) + \nabla \cdot \big(b(\mathbf{m})\rho_t(a, \mathbf{m})\big) = -f(a, x_t)\rho_t(a, \mathbf{m}),$$

$$\rho_t(0, \mathbf{m}) = \mathbb{1}_{\gamma(D)}(\mathbf{m}) \frac{1}{1 - \mathbf{m}(1)} \int_{\mathbb{R}_+} f(a, x_t)\rho_t(a, \gamma^{-1}(\mathbf{m}))da,$$

$$x_t = \bar{H}_t + \int_0^t \bar{h}(t-s) \int_{\mathbb{R}^d} \int_{\mathbb{R}_+} \mathbf{m}(1)\mathbf{m}(2)f(a, x_s)\rho_s(a, \mathbf{m})da d\mathbf{m}ds,$$

$$\rho_0(a, \mathbf{m}) = u_0(a, \mathbf{m}),$$

where $D =]U, 1[\times]0, 1[$ and $\gamma(\mathbf{m}) = \big(U + [1 - U]\mathbf{m}(1), [1 - \mathbf{m}(1)]\mathbf{m}(2)\big)$.[1]

Theorems 1 and 2 have two important implications for neuronal modeling: first, they provide a rigorous footing to multidimensional population density equations, which could be simulated using mesh-based methods described in [7,27,33]; second, they confirm that, not only in the simple 'renewal' cases, the PDE point of view can be used to study the nonlinear dynamics of large networks of spiking neurons [12].

3. Proof of Theorem 1 (Propagation of chaos)

The approach here is standard. We use a fixed-point argument to show that the limit process (6) is well-defined. Then, we use the coupling method [47] to prove that a *typical particle* converges to the limit process.

3.1. Well-posedness of the limit process

Proposition 1. *Grant Assumption 1 and assume that $(\mathbb{E}[H_t^1])_{t\in\mathbb{R}_+} \in \mathcal{C}(\mathbb{R}_+)$. There exists a path-wise unique càdlàg strong solution $(A_t^1, \mathbf{M}_t^1, x_t)_{t\in\mathbb{R}_+}$ taking values in $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}$ to (6). Furthermore, $(x_t)_{t\in\mathbb{R}_+} \in \mathcal{C}(\mathbb{R}_+)$.*

Proof. For all $y \in \mathcal{C}(\mathbb{R}_+)$, let us write $(A_t^y, \mathbf{M}_t^y, x_t^y)_{t\in\mathbb{R}_+}$ the càdlàg strong solution to

$$A_t^y = A_0^1 + t - \int_0^t A_{s-}^y dZ_s^y,$$

$$\mathbf{M}_t^y = \mathbf{M}_0^1 + \int_0^t b(\mathbf{M}_s^y)ds + \int_0^t \Gamma(\mathbf{M}_{s-}^y)dZ_s^y,$$

$$Z_t^y = \int_{[0,t]\times\mathbb{R}_+} \mathbb{1}_{z \le f(A_{s-}^y, \mathbf{M}_{s-}^y, y_s)} \pi^1(ds, dz).$$

Then, we set

$$x_t^y = \mathbb{E}[H_t^1] + \int_0^t \mathbb{E}[h(t-s, A_s^y, \mathbf{M}_s^y)f(A_s^y, \mathbf{M}_s^y, y_s)]ds.$$

Since f and h are bounded, by dominated convergence, we have that $(x_t^y)_{t\in\mathbb{R}_+} \in \mathcal{C}(\mathbb{R}_+)$. Thus, for all $T > 0$, we can define the operator

$$\Phi_T : \mathcal{C}([0, T]) \to \mathcal{C}([0, T]), \qquad (y_t)_{t\in[0,T]} \mapsto (x_t^y)_{t\in[0,T]}.$$

---

[1] Using (11), a simple calculation gives $\det(\mathbf{J}_\gamma(\gamma^{-1}(\mathbf{m}))) = 1 - \mathbf{m}(1)$.

The solution $(A_t^y, \mathbf{M}_t^y, x_t^y)_{t \in [0,T]}$ is a solution to (6) on $[0,T]$ if and only if $(y_t)_{t \in [0,T]} = (x_t^y)_{t \in [0,T]}$, or equivalently, if and only if $(y_t)_{t \in [0,T]}$ is a fixed point of Φ_T. We are going to show that for T small enough, Φ_T is a contraction for the uniform norm.

For all y and $y^* \in \mathcal{C}([0,T])$, by triangular inequality and using the Lipschitz continuity and the boundedness of f and h, we have, for all $t \in [0,T]$,

$$\left| \int_0^t \mathbb{E}[h(t-s, A_s^y, \mathbf{M}_s^y) f(A_s^y, \mathbf{M}_s^y, y_s)] ds \right.$$

$$\left. - \int_0^t \mathbb{E}[h(t-s, A_s^{y^*}, \mathbf{M}_s^{y^*}) f(A_s^{y^*}, \mathbf{M}_s^{y^*}, y_s^*)] ds \right| \le C \int_0^t \Delta_s ds,$$

where

$$\Delta_s := \mathbb{E}\left[|\psi(A_s^y) - \psi(A_s^{y^*})| + \|\mathbf{M}_s^y - \mathbf{M}_s^{y^*}\| + |y_s - y_s^*| \right].$$

By Itô's formula for jump processes,

$$\psi(A_t^y) = \psi(A_0^1) + \int_0^t \psi'(A_s^y) ds - \int_{[0,t] \times \mathbb{R}_+} \psi(A_{s-}^y) \mathbb{1}_{z \le f(A_{s-}^y, \mathbf{M}_{s-}^y, y_s)} \pi^1(ds, dz).$$

Whence,

$$\mathbb{E}[|\psi(A_t^y) - \psi(A_t^{y^*})|] \le \mathbb{E}\left[\left| \int_0^t \psi'(A_s^y) - \psi'(A_s^{y^*}) ds \right| \right]$$

$$+ \mathbb{E}\left[\left| \int_{[0,t] \times \mathbb{R}_+} \psi(A_{s-}^y) \mathbb{1}_{z \le f(A_{s-}^y, \mathbf{M}_{s-}^y, y_s)} - \psi(A_{s-}^{y^*}) \mathbb{1}_{z \le f(A_{s-}^{y^*}, \mathbf{M}_{s-}^{y^*}, y_s^*)} \pi^1(ds, dz) \right| \right].$$

Notice that by Assumption 1, $|\psi'(A_s^y) - \psi'(A_s^{y^*})| \le \kappa |\psi(A_s^y) - \psi(A_s^{y^*})|$. Then, by triangular inequality and using the Lipschitz continuity and the boundedness of f and ψ, we easily get

$$\mathbb{E}[|\psi(A_t^y) - \psi(A_t^{y^*})|] \le C \int_0^t \Delta_s ds;$$

similarly, using the Lipschitz continuity of b, Γ and f and the boundedness of Γ and f, we get $\mathbb{E}[\|\mathbf{M}_t^y - \mathbf{M}_t^{y^*}\|] \le C \int_0^t \Delta_s ds$. Thus, for all $t \in [0,T]$,

$$\Delta_t \le C \int_0^t \Delta_s ds + \|y - y^*\|_\infty,$$

and by Grönwall's lemma, $\Delta_t \le \|y - y^*\|_\infty \exp(Ct)$. Whence,

$$\|\Phi_T(y) - \Phi_T(y^*)\|_\infty \le C'T \exp(CT) \|y - y^*\|_\infty.$$

For T small enough, Φ_T is a contraction and has a unique fixed point by Banach's fixed-point theorem. The fixed point gives the unique solution to (6) on $[0,T]$. Since the constants C and C' do not depend on T nor on the law of $(A_0^1, \mathbf{M}_0^1, (H_t^1)_{t \in \mathbb{R}_+})$, we can iterate the argument above on successive time intervals of length T to obtain the solution to (6) on $\mathbb{R}_+$. $\square$

3.2. Convergence

Proof. The proof of the convergence (5) follows the same general strategy as in [9, Theorem 7].

For all $i = 1, \ldots, N$, we define the *coupled* limit process $(A_t^i, \mathbf{M}_t^i, Z_t^i)_{t \in \mathbb{R}_+}$ as the path-wise unique càdlàg strong solution to

$$A_t^i = A_0^i + t - \int_0^t A_{s-}^i dZ_s^i,$$

$$\mathbf{M}_t^i = \mathbf{M}_0^i + \int_0^t b(\mathbf{M}_s^i)ds + \int_0^t \Gamma(\mathbf{M}_{s-}^i)dZ_s^i,$$

$$Z_t^i = \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \leq f(A_{s-}^i, \mathbf{M}_{s-}^i, x_s)} \pi^i(ds, dz),$$

$$x_t = \mathbb{E}[H_t^1] + \int_0^t \mathbb{E}[h(t-s, A_s^1, \mathbf{M}_s^1)f(A_s^1, \mathbf{M}_s^1, x_s)]ds.$$

The process $(A_t^i, \mathbf{M}_t^i, Z_t^i)_{t \in \mathbb{R}_+}$ is coupled to $(A_t^{i,N}, \mathbf{M}_t^{i,N}, Z_t^{i,N})_{t \in \mathbb{R}_+}$ in the sense that it shares the same $(A_0^i, \mathbf{M}_0^i, (H_t^i)_{t \in \mathbb{R}_+})$ and the same Poisson random measure π^i. The variable x_t has no index i as it the same for all i; it can be interpreted as the deterministic time-varying 'mean field' which acts uniformly on all the individual processes. Importantly, the limit processes $\{(A_t^i, \mathbf{M}_t^i, Z_t^i)_{t \in \mathbb{R}_+}\}_{i=1}^N$ are *i.i.d.* For all $t \geq 0$, let us define

$$\Delta_t^{1,N} := \mathbb{E}\left[|\psi(A_t^{1,N}) - \psi(A_t^1)| + \|\mathbf{M}_t^{1,N} - \mathbf{M}_t^1\| + |X_t^N - x_t|\right].$$

Arguing as in the proof of Proposition 1, we get

$$\mathbb{E}\left[|\psi(A_t^{1,N}) - \psi(A_t^1)| + \|\mathbf{M}_t^{1,N} - \mathbf{M}_t^1\|\right] \leq C \int_0^t \Delta_s^{1,N} ds.$$

It remains to control the term $\mathbb{E}\left[|X_t^N - x_t|\right]$:

Fix $T > 0$. For all $t \in [0, T]$, by triangular inequality,

$$\mathbb{E}\left[|X_t^N - x_t|\right] \leq \mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^N H_t^i - \mathbb{E}[H_t^1]\right|\right]$$

$$+ \mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^N \int_{[0,t] \times \mathbb{R}_+} h(t-s, A_{s-}^{i,N}, \mathbf{M}_{s-}^{i,N})\mathbb{1}_{z \leq f(A_{s-}^{i,N}, \mathbf{M}_{s-}^{i,N}, X_{s-}^N)} \pi^i(ds, dz)\right.\right.$$

$$- \frac{1}{N}\sum_{i=1}^N \int_{[0,t] \times \mathbb{R}_+} h(t-s, A_{s-}^i, \mathbf{M}_{s-}^i)$$

$$\left.\left.\times \ \mathbb{1}_{z \leq f(A_{s-}^i, \mathbf{M}_{s-}^i, x_s)} \pi^i(ds, dz)\right|\right]$$

$$+ \mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^N \int_{[0,t] \times \mathbb{R}_+} h(t-s, A_{s-}^i, \mathbf{M}_{s-}^i)\mathbb{1}_{z \leq f(A_{s-}^i, \mathbf{M}_{s-}^i, x_s)} \pi^i(ds, dz)\right.\right.$$

$$- \int_0^t \mathbb{E}[h(t-s, A_s^1, \mathbf{M}_s^1)$$

$$\left.\left.\times \ f(A_s^1, \mathbf{M}_s^1, x_s)]ds\right|\right]$$

$$=: Q_t^N + R_t^N + S_t^N. \tag{12}$$

By Cauchy–Schwarz inequality and Assumption 2,

$$Q_t^N \leq \left(\frac{\mathrm{Var}[H_t^1]}{N} \right)^{1/2} \leq C_{T,0} N^{-1/2}.$$

By exchangeability, triangular inequality and by the Lipschitz continuity and boundedness of f and h,

$$R_t^N \leq C \int_0^t \Delta_s^{1,N} ds.$$

By Cauchy–Schwarz inequality,

$$S_t^N \leq \mathbb{E}\left[\left(\frac{1}{N} \sum_{i=1}^N \int_{[0,t]\times\mathbb{R}_+} h(t-s, A_{s-}^i, \mathbf{M}_{s-}^i) \mathbb{1}_{z \leq f(A_{s-}^i, \mathbf{M}_{s-}^i, x_s)} \pi^i(ds, dz) \right. \right.$$

$$\left. \left. - \int_0^t \mathbb{E}[h(t-s, A_s^1, \mathbf{M}_s^1) \right. \right.$$

$$\left. \left. \times f(A_s^1, \mathbf{M}_s^1, x_s)] ds \right)^2 \right]^{1/2}$$

$$= \mathrm{Var}\left[\int_{[0,t]\times\mathbb{R}_+} h(t-s, A_{s-}^1, \mathbf{M}_{s-}^1) \mathbb{1}_{z \leq f(A_{s-}^1, \mathbf{M}_{s-}^1, x_s)} \pi^1(ds, dz) \right]^{1/2} N^{-1/2}.$$

However, writing $\tilde{\pi}^1(ds, dz) := \pi^1(ds, dz) - ds\, dz$ the compensated Poisson random measure, we have, by Itô isometry for compensated jump processes,

$$\mathrm{Var}\left[\int_{[0,t]\times\mathbb{R}_+} h(t-s, A_{s-}^1, \mathbf{M}_{s-}^1) \mathbb{1}_{z \leq f(A_{s-}^1, \mathbf{M}_{s-}^1, x_s)} \pi^1(ds, dz) \right]$$

$$= \mathbb{E}\left[\left(\int_{[0,t]\times\mathbb{R}_+} h(t-s, A_{s-}^1, \mathbf{M}_{s-}^1) \mathbb{1}_{z \leq f(A_{s-}^1, \mathbf{M}_{s-}^1, x_s)} \tilde{\pi}^1(ds, dz) \right)^2 \right] \leq T \|h\|_\infty^2 \|f\|_\infty.$$

Hence, $S_t^N \leq C_T N^{-1/2}$. Gathering the bounds, we get

$$\mathbb{E}\left[|X_t^N - x_t| \right] \leq C \int_0^t \Delta_s^{1,N} ds + C_{T,0} N^{-1/2}.$$

Finally,

$$\Delta_t^{1,N} \leq C \int_0^t \Delta_s^{1,N} ds + C_{T,0} N^{-1/2}, \qquad \forall t \in [0, T],$$

and by Grönwall's lemma,

$$\Delta_t^{1,N} \leq C_{T,0} \exp(C_T T) N^{-1/2}, \qquad \forall t \in [0, T], \tag{13}$$

which concludes the proof of (5).

By exchangeability,

$$\mathbb{E}\left[W_1 \left(\frac{1}{N} \sum_{i=1}^N \delta_{(\psi(A_t^{i,N}), \mathbf{M}_t^{i,N})}, \mathcal{L}\left(\psi(A_t^1), \mathbf{M}_t^1 \right) \right) \right]$$

$$\leq \mathbb{E}\left[|\psi(A_t^{1,N}) - \psi(A_t^1)| + \|\mathbf{M}_t^{1,N} - \mathbf{M}_t^1\| \right]$$

$$+ \mathbb{E}\left[W_1\left(\frac{1}{N} \sum_{i=1}^{N} \delta_{(\psi(A_t^i), \mathbf{M}_t^i)}, \mathcal{L}\left(\psi(A_t^1), \mathbf{M}_t^1 \right) \right) \right].$$

Then, we simply use (13) and a result on the convergence of the empirical measures in Wasserstein distance [13, Theorem 1] to get (7). □

4. Alternative propagation of chaos result

Theorem 1 guarantees the convergence of the time-marginals (see (8)), which is sufficient for relating the empirical measure of the system (1) with the PDE (9). However, under more restrictive assumptions on the vector field b and the jump mapping γ, it is possible to get the convergence of the processes, as in [4,9,10].

Assumption 4.

    (i) Writing $(B_t)_{t \in \mathbb{R}_+}$ the flow associated with the vector field b, for all $t \geq 0$, B_t is 1-Lipschitz for the Euclidean distance.
    (ii) The jump mapping γ is 1-Lipschitz for the Euclidean distance.

Theorem 3. *Grant Assumptions 1, 2 and 4. For all $T > 0$, there exists $C_{T,0} > 0$ such that*

$$\mathbb{E}\left[\sup_{t \in [0,T]} |\psi(A_t^{1,N}) - \psi(A_t^1)| + \|\mathbf{M}_t^{1,N} - \mathbf{M}_t^1\| \right] \leq C_{T,0} N^{-1/2}, \tag{14}$$

where $(A_t^1, \mathbf{M}_t^1)_{t \in \mathbb{R}_+}$ is given by the path-wise unique strong solution to (6).

Proof. The well-posedness of the limit process $(A_t^1, \mathbf{M}_t^1, x_t)_{t \in \mathbb{R}_+}$ has already been proved in Proposition 1. For the convergence, we follow the same strategy as in [9, Theorem 8] (see also [4, Theorem IV.1] and [10, Theorem 1]).

Let $\{(A_t^i, \mathbf{M}_t^i, Z_t^i)_{t \in \mathbb{R}_+}\}_{i=1}^{N}$ be the same coupled limit process as in Section 3.2. The integral $\int_0^t |d(Z_s^{1,N} - Z_s^1)|$ counts the number of times one counting process jumps whereas the other does not, on the time interval $[0, t]$. We define

$$\delta_t^N := \mathbb{E}\left[\int_0^t |d(Z_s^{1,N} - Z_s^1)| \right] = \int_0^t \mathbb{E}[|f(A_s^{1,N}, \mathbf{M}_s^{1,N}, X_s^N) - f(A_s^1, \mathbf{M}_s^1, x_s)|] ds.$$

The key observation is that Assumption 4 guarantees

$$\mathbb{E}\left[\sup_{s \in [0,t]} |\psi(A_s^{1,N}) - \psi(A_s^1)| + \|\mathbf{M}_s^{1,N} - \mathbf{M}_s^1\| \right] \leq C \delta_t^N : \tag{15}$$

Clearly, $\sup_{s \in [0,t]} |\psi(A_s^{1,N}) - \psi(A_s^1)| \leq \|\psi\|_\infty \cdot \mathbb{1}_{\int_0^t |d(Z_s^{1,N} - Z_s^1)| > 0} \leq \|\psi\|_\infty \int_0^t |d(Z_s^{1,N} - Z_s^1)|$, which implies that $\mathbb{E}[\sup_{s \in [0,t]} |\psi(A_s^{1,N}) - \psi(A_s^1)|] \leq \|\psi\|_\infty \delta_t^N$. On the other hand, by Assumption 4(i), in a time interval with no jumps in $Z^{1,N}$ nor Z^1, $\|\mathbf{M}_t^{1,N} - \mathbf{M}_t^1\|$ cannot increase. If both $Z^{1,N}$ and Z^1 jump at time t, by Assumption 4(ii), $\|\mathbf{M}_t^{1,N} - \mathbf{M}_t^1\| \leq \|\mathbf{M}_{t-}^{1,N} - \mathbf{M}_{t-}^1\|$. Hence, the only way to have $\|\mathbf{M}_t^{1,N} - \mathbf{M}_t^1\| > \|\mathbf{M}_{t-}^{1,N} - \mathbf{M}_{t-}^1\|$ is if $Z^{1,N}$ jumps at time t but not Z^1 or vice versa. However, in these cases, the increase is bounded by $\|\Gamma\|_\infty$. In summary, we have that $\sup_{s \in [0,t]} \|\mathbf{M}_s^{1,N} - \mathbf{M}_s^1\| \leq \|\Gamma\|_\infty \int_0^t |d(Z_s^{1,N} - Z_s^1)|$, which concludes the verification of (15).

We now have to control δ_t^N. Using (15), we have that

$$\delta_t^N \leq L_f \int_0^t \mathbb{E}\left[|\psi(A_s^{1,N}) - \psi(A_s^1)| + \|\mathbf{M}_s^{1,N} - \mathbf{M}_s^1\| + |X_s^N - x_s|\right] ds$$

$$\leq L_f \int_0^t C\delta_s^N + \mathbb{E}[|X_s^N - x_s|]ds.$$

Fix $T > 0$, for all $s \in [0, T]$, we can bound $\mathbb{E}[|X_s^N - x_s|]$ as in the proof of Theorem 1 (see (12)):

$$\mathbb{E}[|X_s^N - x_s|] \leq Q_s^N + R_s^N + S_s^N.$$

with the same variance bounds $Q_s^N + S_s^N \leq C_{T,0}N^{-1/2}$. By exchangeability, triangular inequality and using (15), we have

$$R_s^N \leq \|h\|_\infty \delta_s^N + \|f\|_\infty L_h \int_0^s \mathbb{E}[|\psi(A_u^{1,N}) - \psi(A_u^1)| - \|\mathbf{M}_u^{1,N} - \mathbf{M}_u^1\|]du \leq C_T \delta_s^N.$$

Gathering the bounds, we get

$$\delta_t^N \leq C_T \int_0^t \delta_s^N ds + C_{T,0}N^{-1/2}, \qquad \forall t \in [0, T],$$

and we conclude using Grönwall's lemma. $\square$

By exchangeability, (14) implies that for all $T > 0$, there exists $C'_{T,0} > 0$ such that

$$\mathbb{E}\left[\sup_{t \in [0,T]} |\psi(A_t^{1,N}) - \psi(A_t^1)| + \|\mathbf{M}_t^{1,N} - \mathbf{M}_t^1\| + |\psi(A_t^{2,N}) - \psi(A_t^2)| + \|\mathbf{M}_t^{2,N} - \mathbf{M}_t^2\|\right]$$

$$\leq C'_{T,0}N^{-1/2}.$$

By standard arguments on the Skorokhod metric and the Continuous mapping theorem, we have the weak convergence

$$\left(A_t^{1,N}, \mathbf{M}_t^{1,N}, A_t^{2,N}, \mathbf{M}_t^{2,N}\right)_{t \in \mathbb{R}_+} \xrightarrow[N \to \infty]{w} \left(A_t^1, \mathbf{M}_t^1, A_t^2, \mathbf{M}_t^2\right)_{t \in \mathbb{R}_+}.$$

Since $(A_t^1, \mathbf{M}_t^1)_{t \in \mathbb{R}_+}$ and $(A_t^2, \mathbf{M}_t^2)_{t \in \mathbb{R}_+}$ have the same law, by [47, Proposition 2.2], we have the convergence in probability of the empirical measure of the system (1) to the law of the limit process:

$$\frac{1}{N}\sum_{i=1}^N \delta_{(A_t^{i,N}, \mathbf{M}_t^{i,N})_{t \in \mathbb{R}_+}} \xrightarrow[N \to \infty]{\mathbb{P}} \mathcal{L}\left((A_t^1, \mathbf{M}_t^1)_{t \in \mathbb{R}_+}\right) \quad \text{in } \mathcal{P}(\mathcal{D}(\mathbb{R}_+, \mathbb{R}_+ \times \mathbb{R}^d)), \qquad (16)$$

where $\mathcal{P}(\mathcal{D}(\mathbb{R}_+, \mathbb{R}_+ \times \mathbb{R}^d))$ denotes the space of probability measures on the Skorokhod space $\mathcal{D}(\mathbb{R}_+, \mathbb{R}_+ \times \mathbb{R}^d)$ of càdlàg functions $\mathbb{R}_+ \to \mathbb{R}_+ \times \mathbb{R}^d$ and $\mathcal{L}\left((A_t^1, \mathbf{M}_t^1)_{t \in \mathbb{R}_+}\right)$ denotes the law of the process $(A_t^1, \mathbf{M}_t^1)_{t \in \mathbb{R}_+}$.

The convergence (16) is clearly stronger than the convergence of the time-marginals (7) but it requires the additional Assumption 4, which is somewhat restrictive.

5. Proof of Theorem 2 (Transport equation for the empirical measure)

Here, our aim is to show that if we write $(\rho_t \otimes \delta_{x_t})_{t \in \mathbb{R}_+}$ the time-marginals of the law of the process (6), then $(\rho_t, x_t)_{t \in \mathbb{R}_+}$ is a weak solution to (9). To show this, we use the limit

process to construct a representation formula for ρ_t. The representation formula is obtained by making rigorous the heuristic 'path integral' method in [34]. We then show that the path integral representation gives a weak solution to (9). Finally, we prove that weak solution to (9) is unique.

5.1. Path integral representation for the time-marginals of the law ot the limit process

To formulate the path integral representation, we first need to introduce some notations and definitions.

Let $(x_t)_{t\in\mathbb{R}_+}$ be given by the limit process (6) and let us write $(A_t^*, \mathbf{M}_t^*)_{t\in\mathbb{R}_+}$ the càdlàg process following the dynamics (6a) and (6b) given $(x_t)_{t\in\mathbb{R}_+}$ and the initial condition $(a_0, \mathbf{m}_0) \in \mathbb{R}_+ \times \mathbb{R}^d$. For all $t > 0$, $(A_t^*, \mathbf{M}_t^*)$ is deterministic given the initial condition and the jump times in $[0, t]$. Hence, for all $k \in \mathbb{N}$ (the number of jumps in $[0, t]$) and for all $0 < t_1 < \cdots < t_k \leq t$ (the jump times in $[0, t]$), we can define, recursively, the mappings $\theta_t^k(t_1, \ldots, t_k) : \mathbb{R}^d \to \mathbb{R}^d$, giving $\mathbf{M}_t^*$ as a function of the initial condition $\mathbf{m}_0$:

$$\theta_t^0 := B_t, \tag{17a}$$

$$\forall k \geq 1, \quad \theta_t^k(t_1, \ldots, t_k) := B_{t-t_k} \circ \gamma \circ \theta_{t_k}^{k-1}(t_1, \ldots, t_{k-1}), \tag{17b}$$

where $(B_t)_{t\in\mathbb{R}_+}$ is the flow associated with the vector field b.

For all $k \geq 1$, we can now define the mapping

$$\phi_t^k \left(\begin{pmatrix} t_1 \\ \vdots \\ t_{k-1} \\ t_k \\ \mathbf{m}_0 \end{pmatrix} \right) = \begin{pmatrix} t_1 \\ \vdots \\ t_{k-1} \\ t - t_k \\ \theta_t^k(t_1, \ldots, t_{k-1}, t_k)(\mathbf{m}_0) \end{pmatrix}.$$

If there are k jumps in the time interval $[0, t]$ and if these jumps occur at times $t_1, \ldots, t_k$, then $(A_t^*, \mathbf{M}_t^*) = (t - t_k, \theta_t^k(t_1, \ldots, t_k)(\mathbf{m}_0))$. For $k = 0$, we simply have

$$\phi_t^0 \left(\begin{pmatrix} a_0 \\ \mathbf{m}_0 \end{pmatrix} \right) = \begin{pmatrix} a_0 + t \\ \theta_t^0(\mathbf{m}_0) \end{pmatrix}.$$

If f is bounded (Assumption 1), we can write $\eta^k(t_1, \ldots, t_k; a_0, \mathbf{m}_0)dt_1 \ldots dt_k$ the probability density over the k-first jump times of the process $(A_t^*, M_t^*)_{t\in\mathbb{R}_+}$ having initial condition $(a_0, \mathbf{m}_0)$. We further define the sub-probability density $v_t^k(t_1, \ldots, t_k; a_0, \mathbf{m}_0)dt_1 \ldots dt_k$:

$$v_t^k(t_1, \ldots, t_k; a_0, \mathbf{m}_0) := \mathbb{1}_{t_k \leq t} \int_t^\infty \eta^{k+1}(t_1, \ldots, t_k, t_{k+1}; a_0, \mathbf{m}_0)dt_{k+1}.$$

Note that the mass of v_t^k is the probability of having exactly k jumps in the time interval $[0, t]$. Hence, $v_t^k / \int v_t^k$ can be interpreted as the probability density over the k jump times knowing that there are exactly k jumps in the time interval $[0, t]$.

Lastly, for all $k \geq 1$, we denote by $\Pi_{t_1,\ldots,t_k,\mathbf{m}}^k$ and $\Pi_{a,\mathbf{m}}^k$ the projections

$$\Pi_{t_1,\ldots,t_k,\mathbf{m}_0}^k : (t_1, \ldots, t_k, a_0, \mathbf{m}_0) \mapsto (t_1, \ldots, t_k, \mathbf{m}_0),$$

$$\Pi_{a,\mathbf{m}}^k : (t_1, \ldots, t_{k-1}, a, \mathbf{m}) \mapsto (a, \mathbf{m}).$$

By convention, for $k = 0$, these projections are the identity.

We have the path integral representation:

Lemma 1. *Grant Assumption* 1. *Let* $(A_t^1, \mathbf{M}_t^1, x_t)_{t \in \mathbb{R}_+}$ *denote the limit process* (6) *for the initial condition* $(A_0^1, \mathbf{M}_0^1) \sim u_0$ *and* $(\mathbb{E}[H_t])_{t \in \mathbb{R}_+} \in \mathcal{C}(\mathbb{R}_+)$. *Then, for all* $t > 0$, *the time-marginal* $\rho_t := \mathcal{L}(A_t^1, \mathbf{M}_t^1)$ *is given by the representation formula*

$$\rho_t = \sum_{k=0}^\infty (\Pi_{a,\mathbf{m}}^k \circ \phi_t^k \circ \Pi_{t_1,\dots,t_k,\mathbf{m}_0}^k)_* (\nu_t^k u_0). \tag{18}$$

If we further grant Assumption 3 *and if* u_0 *is absolutely continuous, then* ρ_t *is also absolutely continuous.*

Proof. Let τ_k be the time of the kth jump of $(A_t^1, \mathbf{M}_t^1, x_t)_{t \in \mathbb{R}_+}$. Since $(A_t^1, \mathbf{M}_t^1)$ is a function of the initial conditions $(A_0^1, \mathbf{M}_0^1)$ and the jump times $\{\tau_k\}_{k \in \mathbb{N}^*}$, for any continuous and bounded test function F on $\mathbb{R}_+ \times \mathbb{R}^d$, we can write $\mathbb{E}[F(A_t, \mathbf{M}_t)]$ as a 'path integral':

$$\mathbb{E}[F(A_t^1, \mathbf{M}_t^1)] = \mathbb{E}\left[F(A_t^1, \mathbf{M}_t^1)\mathbb{1}_{\{t < \tau_1\}}\right] + \sum_{k=1}^\infty \mathbb{E}\left[F(A_t^1, \mathbf{M}_t^1)\mathbb{1}_{\{\tau_k \le t < \tau_{k+1}\}}\right]$$

$$= \mathbb{E}\left[F(\phi_t^0(A_0^1, \mathbf{M}_0^1))\mathbb{1}_{\{t < \tau_1\}}\right] + \sum_{k=1}^\infty \mathbb{E}\left[F(\Pi_{a,\mathbf{m}}^k \circ \phi_t^k(\tau_1, \dots, \tau_k, \mathbf{M}_0^1))\mathbb{1}_{\{\tau_k \le t < \tau_{k+1}\}}\right]$$

$$= \int_{\mathbb{R}^d} \int_{\mathbb{R}_+} F(\phi_t^0(a_0, \mathbf{m}_0))\nu_t^0(a_0, \mathbf{m}_0)u_0(da_0, d\mathbf{m}_0)$$

$$+ \sum_{k=1}^\infty \int_{\mathbb{R}^d} \underbrace{\int_0^t \cdots \int_0^t}_{k \text{ times}} F(\Pi_{a,\mathbf{m}}^k \circ \phi_t^k(t_1, \dots, t_k, \mathbf{m}_0))$$

$$\times \underbrace{\int_{\mathbb{R}_+} \nu_t^k(t_1, \dots, t_k; a_0, \mathbf{m}_0)u_0(da_0, d\mathbf{m}_0) \, dt_1 \dots dt_k}_{(\Pi_{t_1,\dots,t_k,\mathbf{m}}^k)_*(\nu_t^k u_0)}$$

$$= \int_{\mathbb{R}^d} \int_{\mathbb{R}_+} F(a, \mathbf{m})\left(\sum_{k=0}^\infty (\Pi_{a,\mathbf{m}}^k \circ \phi_t^k \circ \Pi_{t_1,\dots,t_k,\mathbf{m}_0}^k)_*(\nu_t^k u_0)\right)(da, d\mathbf{m}),$$

whence the representation formula (18).

If u_0 is absolutely continuous, then $\nu_t^k u_0$ is absolutely continuous for all $k \ge 0$. If, in addition, Assumption 3 is granted, then ϕ_t^k is a proper local diffeomorphism and $(\Pi_{a,\mathbf{m}}^k \circ \phi_t^k \circ \Pi_{t_1,\dots,t_k,\mathbf{m}}^k)_*(\nu_t^k u_0)$ is absolutely continuous for all $k \ge 0$. The probability measure ρ_t is therefore absolutely continuous. $\square$

5.2. From the limit process to weak solutions

Proposition 2. *Grant Assumptions* 1 *and* 3. *Further assume that the law of the initial condition* $(A_0^1, \mathbf{M}_0^1)$ *is the absolutely continuous probability measure* $u_0(a, \mathbf{m})da\,d\mathbf{m}$ *and* $(\mathbb{E}[H_t])_{t \in \mathbb{R}_+} = (\bar{H}_t^1)_{t \in \mathbb{R}_+} \in \mathcal{C}(\mathbb{R}_+)$. *Then, the time-marginals* $(\rho_t)_{t \in \mathbb{R}_+}$ *of the law of the process (given by the path integral representation* (18)*) and* $(x_t)_{t \in \mathbb{R}_+}$ *is a weak solution to* (9).

Proof. First, we use the path integral representation (18) to prove that $\rho \in \mathcal{C}(\mathbb{R}_+, L^1(\mathbb{R}_+ \times \mathbb{R}^d))$. We have to show that for all $T > 0$, $\rho \in \mathcal{C}([0, T], L^1(\mathbb{R}_+ \times \mathbb{R}^d))$. Let us take Z_t^1 from (6c), which counts the number of events in the time interval $[0, t]$. For any $t \in [0, T]$ and any $l \in \mathbb{N}$,

$$\sum_{k=l}^{\infty} \left\| \nu_t^k u_0 \right\|_{L^1} = \mathbb{P}(Z_t^1 \geq l) \leq \mathbb{P}(Z_T^1 \geq l).$$

Hence, for all $t', t \in [0, T]$,

$$\left\| \rho_{t'} - \rho_t \right\|_{L^1}$$
$$\leq \sum_{k=0}^{\infty} \left\| (\Pi_{a,\mathbf{m}}^k \circ \phi_{t'}^k \circ \Pi_{t_1,\dots,t_k,\mathbf{m_0}}^k)_*(\nu_{t'}^k u_0) - (\Pi_{a,\mathbf{m}}^k \circ \phi_t^k \circ \Pi_{t_1,\dots,t_k,\mathbf{m_0}}^k)_*(\nu_t^k u_0) \right\|_{L^1}$$
$$\leq \sum_{k=0}^{l} \left\| (\Pi_{a,\mathbf{m}}^k \circ \phi_{t'}^k \circ \Pi_{t_1,\dots,t_k,\mathbf{m_0}}^k)_*(\nu_{t'}^k u_0) - (\Pi_{a,\mathbf{m}}^k \circ \phi_t^k \circ \Pi_{t_1,\dots,t_k,\mathbf{m_0}}^k)_*(\nu_t^k u_0) \right\|_{L^1}$$
$$+ 2\mathbb{P}(Z_T > l).$$

Since $\mathbb{P}(Z_T > l) \to 0$ as $l \to \infty$, to show that $\rho \in \mathcal{C}([0, T], L^1(\mathbb{R}_+ \times \mathbb{R}^d))$, it suffices to show that for all $k \in \mathbb{N}$,

$$\left((\Pi_{a,\mathbf{m}}^k \circ \phi_t^k \circ \Pi_{t_1,\dots,t_k,\mathbf{m_0}}^k)_*(\nu_t^k u_0) \right)_{t \in [0,T]} \in \mathcal{C}([0, T], L^1(\mathbb{R}_+ \times \mathbb{R}^d)).$$

By the density of $\mathcal{C}_c(\mathbb{R}_+ \times \mathbb{R}^d)$ in $L^1(\mathbb{R}_+ \times \mathbb{R}^d)$, for any $\epsilon > 0$, there exists $\widetilde{u_0} \in \mathcal{C}_c(\mathbb{R}_+ \times \mathbb{R}^d)$ such that $\|\widetilde{u_0} - u_0\|_{L^1} < \frac{\epsilon}{3\|f\|_\infty^k}$. For all $t \in [0, T]$,

$$\left\| (\Pi_{a,\mathbf{m}}^k \circ \phi_t^k \circ \Pi_{t_1,\dots,t_k,\mathbf{m_0}}^k)_*(\nu_t^k \widetilde{u_0}) - (\Pi_{a,\mathbf{m}}^k \circ \phi_t^k \circ \Pi_{t_1,\dots,t_k,\mathbf{m_0}}^k)_*(\nu_t^k u_0) \right\|_{L_1}$$
$$= \left\| \nu_t^k (\widetilde{u_0} - u_0) \right\|_{L^1}$$
$$\leq \|f\|_\infty^k \|\widetilde{u_0} - u_0\|_{L^1} \leq \frac{\epsilon}{3}.$$

Hence, by triangular inequality, it only remains to show that for all $\widetilde{u_0} \in \mathcal{C}_c(\mathbb{R}_+ \times \mathbb{R}^d)$,

$$\left((\Pi_{a,\mathbf{m}}^k \circ \phi_t^k \circ \Pi_{t_1,\dots,t_k,\mathbf{m_0}}^k)_*(\nu_t^k \widetilde{u_0}) \right)_{t \in [0,T]} \in \mathcal{C}([0, T], L^1(\mathbb{R}_+ \times \mathbb{R}^d)). \tag{$*$}$$

Since $\widetilde{u_0}$ is compactly supported, there exists $C > 0$ such that $\mathrm{Supp}(\widetilde{u_0}) \subset [0, C] \times [-C, C]^d$. For all $t \in [0, T]$,

$$(\Pi_{a,\mathbf{m}}^k \circ \phi_t^k \circ \Pi_{t_1,\dots,t_k,\mathbf{m_0}}^k)_*(\nu_t^k \widetilde{u_0}) \leq \mathbb{1}_{[0,C+T] \times [-C-k\|\Gamma\|_\infty, C+k\|\Gamma\|_\infty]^d} \|\widetilde{u_0}\|_\infty \in L^1(\mathbb{R}_+ \times \mathbb{R}^d).$$

Therefore, ($*$) is verified by dominated convergence. This achieves the proof that $\rho \in \mathcal{C}(\mathbb{R}_+, L^1(\mathbb{R}_+ \times \mathbb{R}^d))$.

Now, we verify that ρ satisfies (10) for all test functions. For any $G \in \mathcal{C}_c^\infty(\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^d)$ and any $T > 0$, by Itô's formula for jump processes,

$$G(T, A_T^1, \mathbf{M}_T^1) = G(0, A_0^1, \mathbf{M}_0^1) + \int_0^T [\partial_t + \partial_a + b(\mathbf{M}_t^1) \cdot \nabla] G(t, A_t^1, \mathbf{M}_t^1) dt$$
$$+ \int_{[0,T] \times \mathbb{R}_+} \left(G(t, 0, \gamma(\mathbf{M}_{t-}^1)) - G(t, A_{t-}^1, \mathbf{M}_{t-}^1) \right) \mathbb{1}_{z \leq f(A_{t-}^1, \mathbf{M}_{t-}^1, x_t)} \pi^1(dt, dz).$$

Taking the expectation,

$$\mathbb{E}[G(T, A_T^1, \mathbf{M}_T^1)] = \mathbb{E}[G(0, A_0^1, \mathbf{M}_0^1)] + \int_0^T \mathbb{E}\left[[\partial_t + \partial_a + b(\mathbf{M}_t^1) \cdot \nabla]G(t, A_t^1, \mathbf{M}_t^1)\right] dt$$

$$+ \int_0^T \mathbb{E}\left[\left(G(t, 0, \gamma(\mathbf{M}_t^1)) - G(t, A_t^1, \mathbf{M}_t^1)\right) f(A_t^1, \mathbf{M}_t^1, x_t)\right] dt,$$

which is equivalent to

$$\int_{\mathbb{R}_+} \int_{\mathbb{R}^d} G(T, a, \mathbf{m})\rho_t(a, \mathbf{m})da d\mathbf{m}$$

$$= \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} G(0, a, \mathbf{m})u_0(a, \mathbf{m})da d\mathbf{m} + \int_0^T \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \left\{[\partial_t + \partial_a + b(\mathbf{m}) \cdot \nabla]G(t, a, \mathbf{m})\right.$$

$$\left. + \left(G(t, 0, \gamma(\mathbf{m})) - G(t, a, \mathbf{m})\right) f(a, \mathbf{m}, x_t)\right\} \rho_t(a, \mathbf{m})da d\mathbf{m}dt. \tag{19}$$

Since G in compactly supported, the $T \to \infty$ limit of (19) is (10). This concludes the proof. $\square$

5.3. Uniqueness of weak solutions

Proposition 3. *Grant Assumptions 1 and 3. For any $(u_0, \bar{H}) \in L^1(\mathbb{R}_+ \times \mathbb{R}^d, \mathbb{R}_+) \times \mathcal{C}(\mathbb{R}_+)$, the solution to (9) is unique.*

Proof. Let (ρ, x) be a weak solution for some $(u_0, \bar{H}) \in L^1(\mathbb{R}_+ \times \mathbb{R}^d, \mathbb{R}_+) \times \mathcal{C}(\mathbb{R}_+)$. By Assumption 3, the border condition (9b) can be written like (11) and the function

$$(t, \mathbf{m}) \mapsto \mathbb{1}_{\gamma(\mathbb{R}^d)}(\mathbf{m}) \sum_{\mathbf{m}' \in \gamma^{-1}(\mathbf{m})} \frac{1}{|\det(\mathbf{J}_\gamma(\mathbf{m}'))|} \int_{\mathbb{R}_+} f(a, \mathbf{m}', x_t)\rho_t(a, \mathbf{m}')da =: p_t(\mathbf{m}) \tag{20}$$

is in $\mathcal{C}(\mathbb{R}_+, L^1(\mathbb{R}^d))$ since f is bounded and Lipschitz with respect to the third variable.

By the standard theory of transport equations with initial datum in L^1 (see [40]) and treating (20) as a source term, ρ solves

$$\rho_t(a, \mathbf{m})$$

$$= \begin{cases} u_0(a - t, B_t^{-1}(\mathbf{m})) \exp\left(\int_0^t (\nabla \cdot b)(B_{t-s}^{-1}(\mathbf{m}))\right. \\ \qquad \left. - f(a - t + s, B_{t-s}^{-1}(\mathbf{m}), x_s)ds\right) & \text{if } a \geq t, \\ p_{t-a}(B_a^{-1}(\mathbf{m})) \exp\left(\int_{t-a}^t (\nabla \cdot b)(B_{t-s}^{-1}(\mathbf{m}))\right. \\ \qquad \left. - f(a - t + s, B_{t-s}^{-1}(\mathbf{m}), x_s)ds\right) & \text{if } 0 < a < t. \end{cases} \tag{21}$$

Using (20) and (21), we have the rough bound $\|\rho_t\|_{L^1} \leq \|u_0\|_{L^1} \exp(t\|f\|_\infty)$:

$$\|\rho_t\|_{L^1} \leq \|u_0\|_{L^1} + \int_0^t \int_{\mathbb{R}^d} p_{t-a}(\mathbf{m})d\mathbf{m}da = \|u_0\|_{L^1} + \int_0^t \int_{\mathbb{R}^d} p_s(\mathbf{m})d\mathbf{m}ds$$

$$\leq \|u_0\|_{L^1} + \|f\|_\infty \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}_+} \rho_s(a, \mathbf{m})da d\mathbf{m}ds$$

and the bound is obtained using Grönwall's lemma.

Let (ρ^*, x^*) be another weak solution to (9) for the same $(u_0, \bar{H})$. In the following, we derive bounds on the distance $\|\rho_t - \rho_t^*\|_{L^1} + |x_t - x_t^*|$ and apply Grönwall's lemma. This is relatively straightforward since f and h are bounded and f is Lipschitz with respect to the third variable. For all finite time $T > 0$ and for all $t \in [0, T]$,

$$
\begin{aligned}
&\|\rho_t - \rho_t^*\|_{L^1} \\
&\leq \int_{\mathbb{R}^d} \int_t^\infty \bigg| u_0(a - t, B_t^{-1}(\mathbf{m})) \\
&\qquad\qquad \times \exp\left(\int_0^t (\nabla \cdot b)(B_{t-s}^{-1}(\mathbf{m})) - f(a - t + s, B_{t-s}^{-1}(\mathbf{m}), x_s)ds\right) \\
&\qquad - u_0(a - t, B_t^{-1}(\mathbf{m})) \\
&\qquad\qquad \times \exp\left(\int_0^t (\nabla \cdot b)(B_{t-s}^{-1}(\mathbf{m})) - f(a - t + s, B_{t-s}^{-1}(\mathbf{m}), x_s^*)ds\right) \bigg| da\,d\mathbf{m} \\
&\quad + \int_{\mathbb{R}^d} \int_0^t \bigg| p_{t-a}(B_a^{-1}(\mathbf{m})) \\
&\qquad\qquad \times \exp\left(\int_{t-a}^t (\nabla \cdot b)(B_{t-s}^{-1}(\mathbf{m})) - f(a - t + s, B_{t-s}^{-1}(\mathbf{m}), x_s)ds\right) \\
&\qquad - p_{t-a}^*(B_a^{-1}(\mathbf{m})) \\
&\qquad\qquad \times \exp\left(\int_{t-a}^t (\nabla \cdot b)(B_{t-s}^{-1}(\mathbf{m})) - f(a - t + s, B_{t-s}^{-1}(\mathbf{m}), x_s^*)ds\right) \bigg| da\,d\mathbf{m} \\
&=: Q_1 + Q_2.
\end{aligned}
$$

But $Q_1 \leq \|u_0\| L_f \int_0^t |x_s - x_s^*|ds$, and by triangular inequality (using the shorthand $f(x_t) := f(a, \mathbf{m}, x_t)$),

$$
\begin{aligned}
Q_2 &\leq \left(\int_0^t \int_{\mathbb{R}^d} \rho_s(0, \mathbf{m})d\mathbf{m}\,ds\right) L_f \int_0^t |x_s - x_s^*|ds + \int_0^t \int_{\mathbb{R}^d} \left| p_s(\mathbf{m}) - p_s^*(\mathbf{m}) \right| d\mathbf{m}\,ds \\
&= \left(\int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}_+} f(x_t)\rho_s\, da\,d\mathbf{m}\,ds\right) L_f \int_0^t |x_s - x_s^*|ds \\
&\quad + \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}_+} \left| f(x_t)\rho_s - f(x_t^*)\rho_s^* \right| da\,d\mathbf{m}\,ds \\
&\leq \|f\|_\infty \int_0^t \|\rho_s\|_{L^1}\, ds\, L_f \int_0^t |x_s - x_s^*|ds + \sup_{s \in [0,t]} \|\rho_s\|_{L^1} L_f \int_0^t |x_s - x_s^*|ds \\
&\quad + \|f\|_\infty \int_0^t \|\rho_s - \rho_s^*\|\, ds.
\end{aligned}
$$

By the rough bound on $\|\rho_t\|_{L^1}$ established above, we get

$$
\|\rho_t - \rho_t^*\|_{L^1} \leq C_{T,0} \int_0^t \|\rho_s - \rho_s^*\|_{L_1} + |x_s - x_s^*|ds.
$$

On the other hand,

$$
|x_t - x_t^*| \leq \|h\|_\infty \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}_+} \left| f(x_s)\rho_s - f(x_s^*)\rho_s^* \right| da\,d\mathbf{m}\,ds,
$$

which can be bounded as shown above. Whence,

$$\left\| \rho_t - \rho_t^* \right\|_{L^1} + |x_t - x_t^*| \leq C_{T,0} \int_0^t \left\| \rho_s - \rho_s^* \right\|_{L^1} + |x_s - x_s^*| ds.$$

By Grönwall's lemma, $\left\| \rho_t - \rho_t^* \right\|_{L^1} + |x_t - x_t^*| = 0$ for all $t \in [0, T]$. Since this is true for all $T > 0$, $(\rho, x) = (\rho^*, x^*)$, which concludes the proof. $\square$

Declaration of competing interest

Acknowledgments

References

[1] Emmanuel Bacry, Iacopo Mastromatteo, Jean-François Muzy, Hawkes processes in finance, Mark. Microstruct. Liq. 1 (01) (2015) 1550005.

[2] Pierre Brémaud, Laurent Massoulié, Stability of nonlinear Hawkes processes, Ann. Probab. 24 (3) (1996) 1563–1588.

[3] José A. Cañizo, José A. Carrillo, Sílvia Cuadrado, Measure solutions for some models in population dynamics, Acta Appl. Math. 123 (2013) 141–156.

[4] Julien Chevallier, Mean-field limit of generalized hawkes processes, Stochastic Process. Appl. 127 (12) (2017) 3870–3912.

[5] Jesus M. Cortes, Mathieu Desroches, Serafim Rodrigues, Romain Veltz, Miguel A. Muñoz, Terrence J. Sejnowski, Short-term synaptic plasticity in the deterministic Tsodyks–Markram model leads to unpredictable network dynamics, Proc. Natl. Acad. Sci. USA 110 (41) (2013) 16610–16615.

[6] Riley Crane, Didier Sornette, Robust dynamic classes revealed by measuring the response function of a social system, Proc. Natl. Acad. Sci. USA 105 (41) (2008) 15649–15653.

[7] Marc De Kamps, Mikkel Lepperød, Yi Ming Lai, Computational geometry for modeling neural populations: From visualization to simulation, PLoS Comput. Biol. 15 (3) (2019) e1006729.

[8] Anna De Masi, Antonio Galves, Eva Löcherbach, Errico Presutti, Hydrodynamic limit for interacting neurons, J. Stat. Phys. 158 (4) (2015) 866–902.

[9] Sylvain Delattre, Nicolas Fournier, Marc Hoffmann, Hawkes processes on large networks, Ann. Appl. Probab. 26 (1) (2016) 216–261.

[10] Susanne Ditlevsen, Eva Löcherbach, Multi-class oscillating systems of interacting neurons, Stochastic Process. Appl. 127 (6) (2017) 1840–1869.

[11] Aline Duarte, Eva Löcherbach, Guilherme Ost, Stability, convergence to equilibrium and simulation of non-linear Hawkes processes with memory kernels given by the sum of Erlang kernels, ESAIM Probab. Stat. 23 (2019) 770–796.

[12] Claudia Fonte, Valentin Schmutz, Long time behavior of an age and leaky memory-structured neuronal population equation, 2021, arXiv preprint arXiv:2106.11110.

[13] Nicolas Fournier, Arnaud Guillin, On the rate of convergence in Wasserstein distance of the empirical measure, Probab. Theory Related Fields 162 (3–4) (2015) 707–738.

[14] Nicolas Fournier, Eva Löcherbach, On a toy model of interacting neurons, in: Annales de L'Institut Henri Poincaré, Probabilités et Statistiques, Vol. 52, Institut Henri Poincaré, 2016, pp. 1844–1876.

[15] Antonio Galves, Eva Löcherbach, Modeling networks of spiking neurons as interacting processes with memory of variable length, J. SFdS 157 (1) (2016) 17–32.

[16] Antonio Galves, Eva Löcherbach, Christophe Pouzat, Errico Presutti, A system of interacting neurons with short term synaptic facilitation, J. Stat. Phys. 178 (4) (2020) 869–892.

[17] Felipe Gerhard, Tilman Kispersky, Gabrielle J. Gutierrez, Eve Marder, Mark Kramer, Uri Eden, Successful reconstruction of a physiological circuit with known connectivity from spiking activity alone, PLoS Comput. Biol. 9 (7) (2013) e1003138.

[18] Wulfram Gerstner, Time structure of the activity in neural network models, Phys. Rev. E 51 (1) (1995) 738.

[19] Wulfram Gerstner, Population dynamics of spiking neurons: fast transients, asynchronous states, and locking, Neural Comput. 12 (1) (2000) 43–89.

[20] Wulfram Gerstner, J. Leo van Hemmen, Associative memory in a network of 'spiking' neurons, Netw. Comput. Neural Syst. 3 (2) (1992) 139–164.

[21] Wulfram Gerstner, Werner M. Kistler, Spiking Neuron Models: Single Neurons, Populations, Plasticity, Cambridge University Press, 2002.

[22] Wulfram Gerstner, Werner M. Kistler, Richard Naud, Liam Paninski, Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition, Cambridge University Press, 2014.

[23] Alan G. Hawkes, Spectra of some self-exciting and mutually exciting point processes, Biometrika 58 (1971) 83–90.

[24] Alan G. Hawkes, Hawkes processes and their applications to finance: a review, Quant. Finance 18 (2) (2018) 193–198.

[25] Renaud Jolivet, Alexander Rauch, Hans-Rudolf Lüscher, Wulfram Gerstner, Predicting spike timing of neocortical pyramidal neurons by simple threshold models, J. Comput. Neurosci. 21 (1) (2006) 35–49.

[26] Claude Kipnis, Claudio Landim, Scaling Limits of Interacting Particle Systems, in: Fundamental Principles of Mathematical Sciences, vol. 320, Springer-Verlag, Berlin, 1999.

[27] Yi Ming Lai, Marc de Kamps, Population density equations for stochastic processes with memory kernels, Phys. Rev. E 95 (6) (2017) 062125.

[28] Regis C. Lambert, Christine Tuleau-Malot, Thomas Bessaih, Vincent Rivoirard, Yann Bouret, Nathalie Leresche, Patricia Reynaud-Bouret, Reconstructing the functional connectivity of multiple spike trains using Hawkes models, J. Neurosci. Methods 297 (2018) 9–21.

[29] Eva Löcherbach, Spiking neurons: interacting hawkes processes, mean field limits and oscillations, in: Journées MAS 2016 de la SMAI—Phénomènes Complexes et Hétérogènes, in: ESAIM Proc. Surveys, vol. 60, EDP Sci. Les Ulis, 2017, pp. 90–103.

[30] Sima Mehri, Michael Scheutzow, Wilhelm Stannat, Bian Z. Zangeneh, Propagation of chaos for stochastic spatially structured neuronal networks with delay driven by jump diffusions, Ann. Appl. Probab. 30 (1) (2020) 175–207.

[31] Sylvie Méléard, Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models, in: Probabilistic Models for Nonlinear Partial Differential Equations (Montecatini Terme, 1995), in: Lecture Notes in Math, vol. 1627, Springer, Berlin, 1996, pp. 42–95.

[32] Gianluigi Mongillo, Omri Barak, Misha Tsodyks, Synaptic theory of working memory, Science 319 (5869) (2008) 1543–1546.

[33] Eilif Muller, Lars Buesing, Johannes Schemmel, Karlheinz Meier, Spike-frequency adapting neural ensembles: beyond mean adaptation and renewal theories, Neural Comput. 19 (11) (2007) 2958–3010.

[34] Richard Naud, Wulfram Gerstner, Coding and decoding with adapting neurons: a population approach to the peri-stimulus time histogram, PLoS Comput. Biol. 8 (10) (2012) e1002711.

[35] Yosihiko Ogata, Seismicity analysis through point-process modeling: A review, in: Seismicity Patterns, their Statistical Significance and Physical Meaning, Springer, 1999, pp. 471–507.

[36] Khashayar Pakdaman, Benoît Perthame, Delphine Salort, Dynamics of a structured neuron population, Nonlinearity 23 (1) (2010) 55–75.

[37] Khashayar Pakdaman, Benoît Perthame, Delphine Salort, Relaxation and self-sustained oscillations in the time elapsed neuron network model, SIAM J. Appl. Math. 73 (3) (2013) 1260–1279.

[38] Khashayar Pakdaman, Benoît Perthame, Delphine Salort, Adaptation and fatigue model for neuron networks and large time asymptotics in a nonlinear fragmentation equation, J. Math. Neurosci. 4 (2014) 14, 26.

[39] Alexandre Payeur, Jordan Guerguiev, Friedemann Zenke, Blake A. Richards, Richard Naud, Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits, Nat. Neurosci. 24 (7) (2021) 1010–1019.

[40] Benoît Perthame, Transport Equations in Biology, in: Frontiers in Mathematics, Birkhäuser Verlag, Basel, 2007.

[41] Jonathan W. Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M. Litke, E.J. Chichilnisky, Eero P. Simoncelli, Spatio-temporal correlations and visual signalling in a complete neuronal population, Nature 454 (7207) (2008) 995–999.

[42] Cristóbal Quiñinao, A microscopic spiking neuronal network for the age-structured model, Acta Appl. Math. 146 (2016) 29–55.

[43] Mads Bonde Raad, Susanne Ditlevsen, Eva Löcherbach, Stability and mean-field limits of age dependent Hawkes processes, Ann. Inst. Henri Poincaré Probab. Stat. 56 (3) (2020) 1958–1990.

[44] Patricia Reynaud-Bouret, Vincent Rivoirard, Christine Tuleau-Malot, Inference of functional connectivity in neurosciences via Hawkes processes, in: 2013 IEEE Glob. Conf. Signal Inf. Process, IEEE, 2013, pp. 317–320.

[45] Patricia Reynaud-Bouret, Sophie Schbath, Adaptive estimation for Hawkes processes; application to genome analysis, Ann. Statist. 38 (5) (2010) 2781–2822.

[46] Tilo Schwalger, Anton V. Chizhov, Mind the last spike—firing rate models for mesoscopic populations of spiking neurons, Curr. Opin. Neurobiol. 58 (2019) 155–166.

[47] Alain-Sol Sznitman, Topics in propagation of chaos, in: École D'été de Probabilités de Saint-FLour XIX—1989, in: Lecture Notes in Math., vol. 1464, Springer, Berlin, 1991, pp. 165–251.

[48] Taro Toyoizumi, Kamiar R. Rad, Liam Paninski, Mean-field approximations for coupled populations of generalized linear model spiking neurons with Markov refractoriness, Neural Comput. 21 (5) (2009) 1203–1243.

[49] Wilson Truccolo, From point process observations to collective neural dynamics: Nonlinear hawkes process GLMs, low-dimensional dynamics and coarse graining, J. Physiol. Paris 110 (4) (2016) 336–347.

[50] Wilson Truccolo, Uri T. Eden, Matthew R. Fellows, John P. Donoghue, Emery N. Brown, A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects, J. Neurophysiol. 93 (2) (2005) 1074–1089.

[51] Misha Tsodyks, Klaus Pawelzik, Henry Markram, Neural networks with dynamic synapses, Neural Comput. 10 (4) (1998) 821–835.

[52] Hugh R. Wilson, Jack D. Cowan, Excitatory and inhibitory interactions in localized populations of model neurons, Biophys. J. 12 (1) (1972) 1–24.

[53] Robert S. Zucker, Wade G. Regehr, Short-term synaptic plasticity, Annu. Rev. Physiol. 64 (1) (2002) 355–405.

3 Long time behavior of an age- and leaky memory-structured neuronal population equation

Authors: Claudia Fonte and Valentin Schmutz

Contribution: Claudia Fonte (PhD student at Université Paris Daupine-PSL) and I contributed equally to this work.

(reproduced with permission)

LONG TIME BEHAVIOR OF AN AGE- AND LEAKY MEMORY-STRUCTURED NEURONAL POPULATION EQUATION*

CLAUDIA FONTE† AND VALENTIN SCHMUTZ‡

Abstract. We study the asymptotic stability of a two-dimensional mean-field equation, which takes the form of a nonlocal transport equation and generalizes the time-elapsed neuron network model by the inclusion of a leaky memory variable. This additional variable can represent a slow fatigue mechanism, such as spike-frequency adaptation or short-term synaptic depression. Even though two-dimensional models are known to have emergent behaviors, such as population bursts, which are not observed in standard one-dimensional models, we show that in the weak connectivity regime, two-dimensional models behave like one-dimensional models, i.e., they relax to a unique stationary state. The proof is based on an application of Harris's ergodic theorem and a perturbation argument, both adapted to the case of a multidimensional equation with delays.

1. Introduction. Multidimensional mean-field models in theoretical neuroscience are challenging to analyze [41, 48, 2, 32], but their study is a necessary step towards understanding how multiple timescales present at the single-neuron level [40, 45] affect the dynamics of large networks of neurons.

One-dimensional mean-field equations for populations of spiking neurons with deterministic drift and stochastic jumps have been a subject of mathematical studies since the works of Pakdaman, Perthame, and Salort [35, 36, 37], providing rigorous foundations for earlier works in theoretical neuroscience [50, 21, 18, 19]. These population equations correspond to the mean-field limit of large networks of interacting neurons [10, 15, 6]. However, they are derived from spiking neuron models that are of the "renewal" type (with the exception of [37]), which means that, while they capture the effect of neuronal refractoriness, they neglect slower neuronal timescales, such as those of spike-frequency adaptation and short-term synaptic plasticity.

To take into account slow neuronal timescales, state-of-the-art phenomenological spiking neuron models must be multidimensional [28, 45] or kernel-based [46, 38, 39] (see also [20, Chap. 6.4]). In the following, we consider a class of neuron models that characterize neuronal refractoriness by an "age" variable (the time elapsed since the last spike) and effects of spike-frequency adaptation or short-term synaptic plasticity by a "leaky memory" variable. For this class of neuron models, the mean-field limit is characterized by a multidimensional transport equation with a nonlocal boundary

†CEREMADE, Université Paris Dauphine-PSL, 75016 Paris, France (fonte@ceremade.dauphine.fr).

‡Brain Mind Institute, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (valentin.schmutz@epfl.ch).

condition [43]. In this work, we study in the two-dimensional case the long time behavior of the solutions to the equation proposed in [43].

1.1. The age- and leaky memory-structured model. The population model we consider describes the evolution of a density ρ_t over the state-space $(a, m) \in \mathbb{R}_+ \times \mathbb{R}_+^*$, where a and m are the "age" and "leaky memory" variables of the neuron, and $\rho_t(a, m)$ represents the density of neurons in state (a, m) at time t.

The nonlinear evolution problem for the density ρ_t, for the initial datum u_0, reads

$$(1.1a) \qquad \partial_t \rho_t + \nabla \cdot (b\rho_t) = -f(a, m, \varepsilon x_t)\rho_t,$$

$$(1.1b) \qquad \rho_t(0, m) = \mathbb{1}_{m > \gamma(0)} \left| (\gamma^{-1})'(m) \right| \int_0^\infty f(a, \gamma^{-1}(m), \varepsilon x_t)\rho_t(a, \gamma^{-1}(m))da,$$

$$(1.1c) \qquad x_t = \int_0^t \int_0^\infty \int_0^\infty h(t - s, a, m)f(a, m, \varepsilon x_s)\rho_s(a, m)dadmds,$$

$$(1.1d) \qquad \rho_0 = u_0.$$

The dynamics of the model can be decomposed into three elements: (i) the behavior of neurons between spikes, (ii) the spike-triggered jumps, and (iii) the interaction between neurons, which we discuss in turn.

(i) Between spikes, neurons are transported along the vector field $b(a, m) = (1, -\lambda m)$, with $\lambda > 0$ ($\nabla\cdot$ denotes the divergence operator over the state-space).

(ii) Neurons spike at rate $f(a, m, \varepsilon x_t)$, where $f : \mathbb{R}_+ \times \mathbb{R}_+^* \times \mathbb{R} \to \mathbb{R}_+$ is the "firing rate function" corresponding to the stochastic intensity of the spike generation process, and $\varepsilon \in \mathbb{R}$ is the connection strength. When a neuron spikes, its age a is reset to 0, and its leaky memory variable m jumps to $\gamma(m)$, where $\gamma : \mathbb{R}_+ \to \mathbb{R}_+^*$ is the "jump mapping" and is assumed to be a strictly increasing $\mathcal{C}^1$-diffeomorphism. As a consequence, the border condition (1.1b) has a simple interpretation: the density of neurons in state $(0, m)$ at time t is equal to the marginal density of those neurons that have their leaky memory variable in state $\gamma^{-1}(m)$ and spike at time t. The indicator function $\mathbb{1}_{m > \gamma(0)}$ reflects the fact that m is always strictly positive, and the term $\left| (\gamma^{-1})'(m) \right|$ is necessary to guarantee the conservation of the total mass of neurons. Indeed, formally,

$$\partial_t \int \rho_t$$
$$= \int \mathbb{1}_{m > \gamma(0)} \left| (\gamma^{-1})'(m) \right| \int_0^\infty f(a, \gamma^{-1}(m), \varepsilon x_t)\rho_t(a, \gamma^{-1}(m))dadm - \int f(a, m, \varepsilon x_t)\rho_t = 0,$$

by a change of variable.

(iii) Neurons interact through the "total postsynaptic potential" x_t, which integrates the past spiking activity of the population, filtered by the "interaction function" $h : \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+^* \to \mathbb{R}$, and which weighted by the connection strength $\varepsilon \in \mathbb{R}$, influences the firing rate f. If we write $N(t)$ the population activity (the mean firing rate)

$$N(t) := \int_0^\infty \int_0^\infty f(a, m, \varepsilon x_t)\rho_t(a, m)dadm,$$

and if we take h independent of a and m, then x_t takes the form

$$x_t = \int_0^t h(t - s)N(s)ds,$$

where h is now a simple delay kernel, as in [21, 18, 19, 35]. In our formulation, h in (1.1c) allows us to model more general interactions. For example, in subsection 1.2.2, we show that by choosing $h(t, a, m) = \hat{h}(t)(1 - m)$, we can include the effects of a classical short-term synaptic plasticity model [47].

1.2. Motivation. The model (1.1) extends the time-elapsed neuron network model [35] (see also [18, 19]) by the addition of a leaky memory variable which can accumulate over spikes (as opposed to the age variable which is reset to 0 at each spike) and hence introduces a slow timescale in the population dynamics. Such a slow timescale is typically used to account for some form of fatigue mechanism, which can act on the spiking activity (spike-frequency adaptation) or on synaptic transmission (short-term synaptic depression). Slow fatigue at the single-neuron level can lead to nontrivial emergent behaviors at the population level, such as population bursts [49, 22, 17] (see Figure 1), which have not been observed in the age- or voltage-structured models of [35] and [10] (but see [37]). Even though some population equations have been successfully used in the computational neuroscience literature to study emergent behaviors in networks of neurons with fatigue, these population equations were obtained at the cost of a timescale separation approximation [22, 17] or a "mixing" assumption [33, 44], making them inexact. In contrast, the model (1.1) is the exact mean-field limit [43] for spiking neuron models with spike-frequency adaptation or short-term synaptic depression, as we discuss now.

1.2.1. Spike-frequency adaptation. The recent spike history of a neuron can modulate its firing rate f, leading to spike-frequency adaptation [3]. If h is independent of a and m, and if $\gamma(m) = m + \hat{\Gamma}$ for a fixed $\hat{\Gamma} > 0$, (1.1) becomes

$$(1.2a) \qquad \partial_t \rho_t + \nabla \cdot (b\rho_t) = -f(a, m, \varepsilon x_t)\rho_t,$$

$$(1.2b) \qquad \rho_t(0, m) = \mathbb{1}_{m > \hat{\Gamma}} \int_0^\infty f(a, m - \hat{\Gamma}, \varepsilon x_t)\rho_t(a, m - \hat{\Gamma})da,$$

$$(1.2c) \qquad x_t = \int_0^t h(t - s) \int_0^\infty \int_0^\infty f(a, m, \varepsilon x_s)\rho_s(a, m)da\,dm\,ds,$$

$$(1.2d) \qquad \rho_0 = u_0.$$

If $\eta : \mathbb{R}_+ \to \mathbb{R}$ is a bounded function such that $\lim_{a \to +\infty} \eta(a) = 0$ (η is the "refractory kernel" [20, sect. 9.3]), we can define f more explicitly as

$$(1.2e) \qquad f(a, m, \varepsilon x_t) := \hat{f}(\eta(a) - m + \varepsilon x_t),$$

where $\hat{f} : \mathbb{R} \to \mathbb{R}_+$ is typically a nondecreasing function. Since m makes jumps of size $\hat{\Gamma} > 0$ at each spike and decays exponentially at rate λ between spikes, m accumulates over spikes, which decreases the firing rate f (1.2e), leading to spike-frequency adaptation [3]. More specifically, (1.2) is a population equation for adaptive SRM_0 (Spike Response Model) neurons [27, 20].

Populations of spiking neurons with spike-frequency adaptation exhibit self-sustained population bursts when the connectivity strength is sufficiently strong [49, 22, 17]. We call self-sustained population bursts a periodic pattern of activity characterized by an alternation between periods of low population activity and sequences of population spikes (short time intervals where almost all the neurons in the population fire). This definition is borrowed from the following definition of single-neuron bursting [26]: "When neuron activity alternates between a quiescent state and repetitive spiking, the neuron activity is said to be *bursting*." In Figure 1, we show simulations of
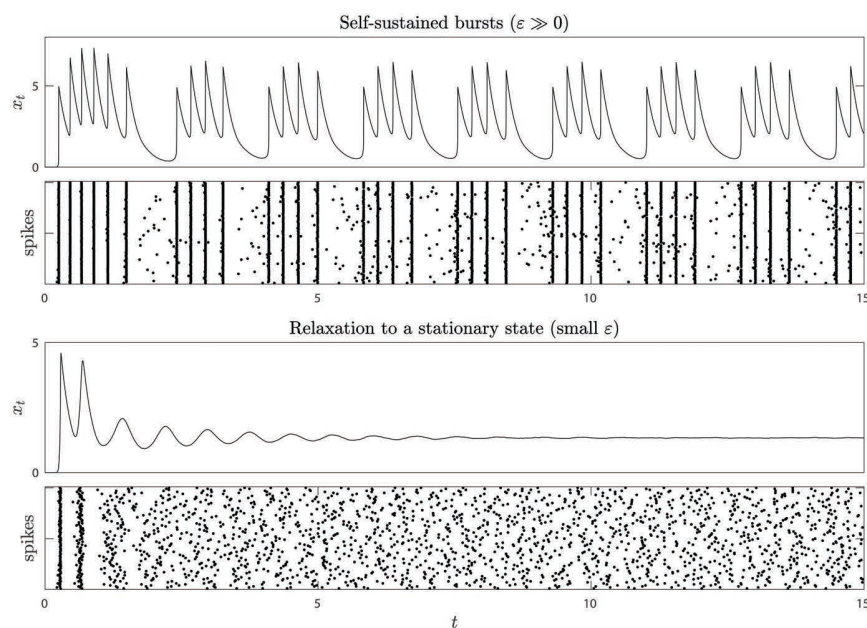
FIG. 1. *Depending on the connectivity strength ε, a population of adaptive SRM_0 neurons can exhibit self-sustained bursts ($\varepsilon \gg 0$) or relaxation to a stationary state (small ε). We show simulations of a network of $5 \cdot 10^5$ adaptive SRM_0 neurons, approximating the mean-field limit (1.2), with identical parameters (except for ε) and identical initial conditions. The raster plots below the plots for the time-evolution of the total postsynaptic potential x_t represent the spikes of 100 randomly selected neurons.*

(1.2) for two different connectivity strengths ε. For large ε, we observe self-sustained bursts, whereas for small ε, we observe relaxation to a stationary state. Note that the neurons considered here are not intrinsically bursting: if an adaptive SRM_0 neuron receives no input (or a constant input), it does not burst. Population bursts are therefore an emergent behavior of the mean-field model (1.2).

For comparison, in Appendix A we show similar simulations for the time-elapsed neuron network model [35], where, as expected, we only observe self-sustained oscillations or relaxation to a stationary state.

1.2.2. Short-term synaptic depression. The recent spike history of a presynaptic neuron can modulate the synaptic transmission, leading to short-term synaptic plasticity [51]. We will consider here the case of depressive synapses and use the model of [47] (with a change of variable for convenience). In this case, the state-space is $(a, m) \in \mathbb{R}_+ \times \,]0, 1[$. Taking f independent of m, and choosing h and γ of the form

$h(t,a,m) := \hat{h}(t)(1-m)$ and $\gamma(m) := 1 - v + vm$ for a fixed $v \in]0,1[$, (1.1) becomes

$$\text{(1.3a)} \qquad \partial_t \rho_t + \nabla \cdot (b\rho_t) = -f(a, \varepsilon x_t)\rho_t,$$

$$\text{(1.3b)} \qquad \rho_t(0,m) = \mathbb{1}_{m > \gamma(0)} \frac{1}{v} \int_0^\infty f(a, \varepsilon x_t)\rho_t(a, \gamma^{-1}(m))da,$$

$$\text{(1.3c)} \qquad x_t = \int_0^t \hat{h}(t-s) \int_0^1 \int_0^\infty (1-m)f(a, \varepsilon x_s)\rho_s(a,m)dadmds,$$

$$\text{(1.3d)} \qquad \rho_0 = u_0.$$

Note that the term $\frac{1}{v}$ on the right-hand side of (1.3b) simply comes from the fact that $|(\gamma^{-1})'(m)| = \frac{1}{v}$ for all $m \in]0,1[$. Here, at each spike, m makes strictly positive jumps whose size tends to 0 as m tends to 1 (since $\gamma(1) = 1$) and decays exponentially at rate λ between spikes. If m is close to 1, synaptic transmission is weak because of the factor $(1-m)$ in (1.3c).

As observed in [42], the stationary state of populations of neurons with short-term synaptic plasticity can be described by a simple formula, which we rederive in subsection 4.3.

1.3. Assumptions and main results. The main result of this work is the exponential stability of (1.1) in the weak connectivity regime (Theorem 1.4)—or, more explicitly, there exists $\varepsilon^{**} > 0$ such that (1.1) is exponentially stable for all connectivity strength $\varepsilon \in]-\varepsilon^{**}, +\varepsilon^{**}[$. Before proving the exponential stability, we first establish the well-posedness of (1.1) in the appropriate function space (Theorem 1.2) and show that stationary solutions exist and are unique for sufficiently weak connectivity (Theorem 1.3).

Here, we study the weak solutions to (1.1) for an initial datum in $L_+^1 := L^1(\mathbb{R}_+ \times \mathbb{R}_+^*, \mathbb{R}_+)$ and write $L_+^1(\mathbb{R}_+^*) := L^1(\mathbb{R}_+^*, \mathbb{R}_+)$.

DEFINITION 1.1 (solutions). $(\rho, x) \in \mathcal{C}(\mathbb{R}_+, L_+^1) \times \mathcal{C}(\mathbb{R}_+)$ is a solution to (1.1), for the initial datum $u_0 \in L_+^1$, if
(1.4a)
$$x_t = \int_0^t \int_0^\infty \int_0^\infty h(t-s,a,m)f(a,m,\varepsilon x_s)\rho_s(a,m)dadmds \qquad \forall t \geq 0$$

and if for all $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+^*)$,

$$\text{(1.4b)} \quad 0 = \int_0^\infty \int_0^\infty u_0(a,m)\varphi(0,a,m)dadm$$
$$+ \int_0^\infty \int_0^\infty \int_0^\infty \rho_t(a,m)\Big\{ [\partial_t + \partial_a - \lambda m \partial_m]\varphi + (\varphi(t,0,\gamma(m))$$
$$- \varphi(t,a,m))f(a,m,\varepsilon x_t)\Big\}dadmdt.$$

To prove the well-posedness of (1.1), we need some simple assumptions on the firing rate function f and the interaction function h.

ASSUMPTION 1. f is bounded and L_f-Lipschitz, i.e.,

$$|f(a,m,x) - f(a^*,m^*,x^*)| \leq L_f(|a - a^*| + |m - m^*| + |x - x^*|),$$

and h is bounded and continuous.

Since we want to apply Harris's theorem, the well-posedness in L^1 (which is treated in [43]) is not enough, and we need the well-posedness in a weighted L^1 space (where the weight satisfies a Lyapunov condition [29]) with a global-in-time estimate in the weighted L^1 norm.

Using the weight function

$$w : \mathbb{R}_+ \times \mathbb{R}_+ \to [1, \infty[, \quad (a, m) \mapsto 1 + m,$$

we define the function space

$$L_+^1(w) := \left\{ g \in L^1(\mathbb{R}_+ \times \mathbb{R}_+^*, \mathbb{R}_+) \,\Big|\, \|g\|_{L^1(w)} := \int_0^\infty \int_0^\infty g(a, m)w(a, m)dadm < \infty \right\}.$$

To obtain a global-in-time estimate in the $L_+^1(w)$ norm, we further require the jump sizes of γ to be bounded.

ASSUMPTION 2. *There exists a bounded function* $\Gamma : \mathbb{R}_+^* \to \mathbb{R}_+^*$ *such that for all* $m \in \mathbb{R}_+^*$, $\gamma(m) = m + \Gamma(m)$.

THEOREM 1.2 (well-posedness). *Grant Assumption 1. For any initial datum* $u_0 \in L_+^1$, *there exists a unique weak solution* (ρ, x) *to* (1.1). *This solution satisfies*
   (i) (L^1-*stability*) $\|\rho_t\|_{L^1} = \|u_0\|_{L^1}$ *for all* $t > 0$.
   (ii) (*Global bound in* $L_+^1(w)$) *If, in addition, Assumption 2 holds and* $u_0 \in L_+^1(w)$, *then*

$$(1.5) \qquad \forall t > 0, \qquad \|\rho_t\|_{L^1(w)} \leq \|u_0\|_{L^1(w)}\, e^{-\alpha t} + \frac{b}{\alpha}(1 - e^{-\alpha t})\|u_0\|_{L^1}$$

for some constants $\alpha > 0$ *and* $b \in \mathbb{R}$.

In contrast to [43], the well-posedness proof presented here does not involve any probabilistic argument. The proof consists of two consecutive applications of Banach's fixed-point theorem, where a first fixed-point gives the unique solution to a linearized version of (1.1) which is then used in a second fixed-point treating the nonlinearity of (1.1).

The second step towards the exponential stability proof is the study of the existence and uniqueness of the stationary solutions to (1.1). For this step, we require the following assumption.

ASSUMPTION 3.
   (i) *There exist* $\Delta_{abs} > 0$ *and* $\sigma > 0$ *such that*

$$f(a, m, x) \geq \sigma \qquad \forall (a, m, x) \in [\Delta_{abs}, +\infty[\times \mathbb{R}_+^* \times \mathbb{R}.$$

   (ii) *There exists* $C_\gamma \in]0, 1]$ *such that* $C_\gamma \leq \gamma' \leq 1$.
   (iii) $\bar{h}(a, m) = \int_0^\infty h(t, a, m)dt$ *is bounded.*

The first point of Assumption 3 sets a lower bound on the firing rate function f for any $a \geq \Delta_{\text{abs}}$ while allowing neurons to have an absolute refractory period $\Delta_{\text{abs}} > 0$, i.e., a period of time following a spike during which f can be 0 (which is an important neurodynamical feature [20, sect. 1.1]). This assumption is also used in [5].

In the second point of Assumption 3, the lower bound $0 < C_\gamma \leq \gamma'$ guarantees that γ is strictly increasing, which reflects the idea that m is a "leaky memory" variable of the past neuronal activity. On the other hand, the upper bound $\gamma' \leq 1$, which can be rewritten in terms of the jump size function Γ as $\Gamma' \leq 0$, prevents the variable m

from growing too fast and allows for a potential saturation of the memory, as in the example with short-term synaptic plasticity (1.3). The third point of Assumption 3 reflects the fact that a single spike has a finite impact on the neuron that receives it.

We emphasize that the two examples shown above—spike-frequency adaptation (1.2) and short-term synaptic depression (1.3)—satisfy Assumption 3.

THEOREM 1.3 (stationary solutions). *Grant Assumptions* 1–3.
(i) *There exists a stationary solution to* (1.1).
(ii) *There exists* $\varepsilon^* > 0$ *such that for all* $\varepsilon \in]-\varepsilon^*, +\varepsilon^*[$, *the stationary solution to* (1.1) *is unique.*

Over the course of this work, we obtained the existence of the stationary solution by two different approaches. The first approach is based on the Doeblin–Harris method [23] and is similar to that of [5]. First, we show that when x_t is fixed and time-invariant in (1.1) (neurons are noninteracting), the system satisfies a Harris condition—this constitutes a key result of this work—and we can use Harris's theorem to get the stationary solution. Then, we use the Lipschitz continuity of the stationary solutions with respect to the fixed x to prove the existence of a stationary solution for arbitrary connectivity strengths ε. Finally, for ε small enough, we also get the uniqueness of the stationary solution by Banach's fixed-point theorem.

The second approach relies on the fact that the stationary solutions solve an integral equation, for which we can show that a solution exists by Schauder's fixed-point theorem. In the process, we get several estimates on the stationary solutions, namely that they are continuous, bounded, and exponentially decaying in m. However, this approach does not give uniqueness.

As mentioned above, the application of Harris's theorem requires us to consider solutions in the weighted space $L^1(w)$. However, in the case where the state-space of the leaky memory variable m is bounded, the situation is simpler: we can use Doeblin's theorem in L^1. The following assumption guarantees that m stays in a bounded state-space.

ASSUMPTION 4. *There exists* $G > 0$ *such that for all* $m \in \mathbb{R}_+^*$, $\gamma(m) < G$.

Note that this assumption is satisfied in the example with short-term synaptic plasticity (1.3), with $G = 1$.

Finally, to study the exponential stability of (1.1), we need an exponential decay on h.

ASSUMPTION 5. *There exists* $\mathfrak{h}, C_h > 0$ *such that* $h(t, a, m) \leq C_h e^{-\mathfrak{h}t}$ *for all* $(t, a, m) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+^*$.

By a perturbation argument similar to that of [30], we obtain our main result.

THEOREM 1.4 (exponential stability in the weak connectivity regime). *Grant Assumptions* 1–3 *and* 5. *For any* $W > 0$, *there exists* $\varepsilon_W^{**} > 0$ *such that for* $\varepsilon \in]-\varepsilon_W^{**}, +\varepsilon_W^{**}[$, *there exist* $C \geq 1$ *and* $c_W > 0$ *such that for all initial data* $u_0 \in L_+^1(w)$ *with* $\|u_0\|_{L^1} = 1$ *and* $\|u_0\|_{L^1(w)} \leq W$, *the solution* (ρ, x) *to* (1.1) *satisfies*

$$(1.6) \qquad \|\rho_t - \rho_\infty\|_{L^1(w)} + |x_t - x_\infty| \leq C e^{-c_W t} \left(\|u_0 - \rho_\infty\|_{L^1(w)} + 1 \right) \qquad \forall t \geq 0,$$

where (ρ_∞, x_∞) *is the unique stationary solution given by Theorem* 1.3(ii).

If, in addition, we grant Assumption 4, *then there exists* $\varepsilon^{**} > 0$ *such that for all* $\varepsilon \in]-\varepsilon^{**}, +\varepsilon^{**}[$, *there exist* $C' \geq 1$ *and* $c > 0$ *such that for all initial data* $u_0 \in L_+^1$

with $\|u_0\|_{L^1} = 1$,

$$(1.7) \qquad \|\rho_t - \rho_\infty\|_{L^1} + |x_t - x_\infty| \le C' e^{-ct} \left(\|u_0 - \rho_\infty\|_{L^1} + 1\right) \qquad \forall t \ge 0.$$

From the neuronal modeling point of view, this result is not surprising: when the connection strength is weak enough, neurons do not synchronize, and the population activity converges to a stationary state. This was already proved for simpler one-dimensional models (see below), and the addition of a leaky memory variable carrying the effect of spike-frequency adaptation or short-term synaptic plasticity does not change this behavior.

1.4. Discussion of the methods. The asymptotic stability of the age-structured model of [35] in the weak connectivity regime has been studied using entropy methods (assuming that f is a step-function) [35, 36], spectral analysis of semigroups in Banach spaces [31, 30], and Doeblin's theorem [5]. For treatments of the strong connectivity regime, we refer the reader to [35, 36, 30].

The asymptotic stability of the closely related voltage-structured model of [10] in the weak connectivity regime has also been studied by Cormier, Tanré, and Veltz [8] using Laplace transform techniques. In addition, the same authors have analyzed the nonlinear stability of the stationary solutions [7] (see also [12]) and proved the existence of periodic solutions [9].

Doeblin's theorem has also been used in [14] in the case of the "threshold crossing" neuronal population equation of [34]. Note that closely related methods have been used by probabilists to study the ergodicity of single-neuron models [25, 13].

Our approach combines strategies from [31] and [5], even though [5] uses Doeblin's instead of Harris's theorem. On the one hand, our proof is based on the application of Harris's theorem for the linear problem, which simplifies the proof of [31]. On the other hand, we use an argument from [31] to deal with delay effects, which are not considered in [5]. Note that our model is two-dimensional (by the addition of the leaky memory variable), whereas the aforementioned works only considered one-dimensional models.

1.5. Plan of the paper. The proof of Theorem 1.2 (well-posedness) is presented in section 2. In section 3, we prove the exponential stability of (1.1) in the noninteracting case $\varepsilon = 0$ using Harris's or Doeblin's theorem. The proof of Theorem 1.3 (stationary solutions) is presented in section 4 which is divided into three parts: in the first part, we present a proof which uses the exponential stability of the noninteracting case; in the second part, we present an alternative proof for the existence of stationary solutions which does not involve the Doeblin–Harris method; and in the last part, we present a proof for the formula of [42] in the case of short-term synaptic plasticity (1.3). Finally, section 5 is dedicated to the proof of Theorem 1.4 (exponential stability in the weak connectivity regime).

2. Well-posedness. This section is dedicated to the proof of Theorem 1.2, which we decompose into several lemmas. First, we verify the a priori L^1-stability of the solutions to (1.1), a technical result we use later in the proof. Then, we introduce a linearized version of (1.1) and show that it is well-posed by an application of Banach's fixed-point theorem. Another application of Banach's fixed-point theorem is used to treat the nonlinearity of (1.1) and concludes the proof of the well-posedness in L^1. Finally, we prove the global bound in $L^1_+(w)$ (Theorem 1.2(ii)), which we will use to apply Harris's theorem in the next sections.

LEMMA 2.1 (a priori L^1-stability). *Grant Assumption 1. If (ρ, x) is a weak solution to (1.1) for the initial datum $u_0 \in L^1_+$, then*

$$\|\rho_t\|_{L^1} = \|u_0\|_{L^1} \qquad \forall t > 0.$$

Proof. By a standard cut-off in time argument, we have that for all $T > 0$ and for all $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+^*)$,

$$\int_0^\infty \int_0^\infty \rho_T(a, m)\varphi(T, a, m)dadm - \int_0^\infty \int_0^\infty u_0(a, m)\varphi(0, a, m)dadm$$
$$= \int_0^T \int_0^\infty \int_0^\infty \rho_t(a, m)\Big\{[\partial_t + \partial_a - \lambda m \partial_m]\varphi + (\varphi(t, 0, \gamma(m))$$
$$- \varphi(t, a, m))f(a, m, \varepsilon x_t)\Big\}dadmdt.$$

Let χ be a function in $\mathcal{C}_c^\infty(\mathbb{R}_+ \times \mathbb{R}_+^*, \mathbb{R}_+)$ such that

$$\chi(a, m) = 1 \quad \forall\, a^2 + m^2 \leq 1.$$

For all $n \in \mathbb{N}^*$, we write $\widetilde{\varphi}^n \in \mathcal{C}^\infty(\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+^*)$, the classical solution to the transport equation

(2.1a) $$\partial_t \widetilde{\varphi}^n(t, a, m) + \partial_a \widetilde{\varphi}^n(t, a, m) - \lambda m \partial_m \widetilde{\varphi}^n(t, a, m) = 0,$$
(2.1b) $$\widetilde{\varphi}^n(0, a, m) = \chi(a/n, m/n).$$

Because of the finite speed of propagation of the transport equation, for all n, there exists a function $\varphi^n \in \mathcal{C}_c^\infty(\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+^*)$ such that $\varphi^n(t, a, m) = \widetilde{\varphi}^n(t, a, m)$ for all $(t, a, m) \in [0, T] \times \mathbb{R}_+ \times \mathbb{R}^*$. Hence, for all $n \in \mathbb{N}^*$,

$$\int_0^\infty \int_0^\infty \varphi^n(T, a, m)\rho_T(a, m)dadm - \int_0^\infty \int_0^\infty \varphi^n(0, a, m)u_0(a, m)dadm$$
$$= \int_0^T \int_0^\infty \int_0^\infty \Big\{\partial_t \varphi^n + \partial_a \varphi^n - \lambda m \partial_m \varphi^n + \big(\varphi^n(t, 0, \gamma(m))$$
$$- \varphi^n(t, a, m)\big)f(a, m, \varepsilon x_t)\Big\}\rho_t(a, m)dadmdt.$$

As φ_n is a solution to (2.1a) on time $[0, T]$, we get

$$\int_0^\infty \int_0^\infty \varphi^n(T, a, m)\rho_T(a, m)dadm - \int_0^\infty \int_0^\infty \varphi^n(0, a, m)u_0(a, m)dadm$$
$$= \int_0^T \int_0^\infty \int_0^\infty \Big\{ \big(\varphi^n(t, 0, \gamma(m)) - \varphi^n(t, a, m)\big)f(a, m, \varepsilon x_t)\Big\}\rho_t(a, m)dadmdt.$$

For all $(t, a, m) \in [0, T] \times \mathbb{R}_+ \times \mathbb{R}_+^*$, $\varphi^n(t, a, m) \xrightarrow[n \to \infty]{} 1$, since the initial datum tends to 1 as $n \to \infty$ (2.1b) and by finite speed of propagation. Thus, by dominated convergence, we get

(2.2) $$\int_0^\infty \int_0^\infty \rho_T(a, m)dadm - \int_0^\infty \int_0^\infty u_0(a, m)dadm = 0.$$

Since ρ is nonnegative, this concludes the proof. $\qquad\square$

Lemma 2.1 will allow us to prove the well-posedness of (1.1) by means of fixed-point arguments. Let us first introduce a linearized version of (1.1): for all $x \in \mathcal{C}(\mathbb{R}_+)$, we consider the linear evolution problem

$$(2.3a) \qquad \partial_t \rho_t + \nabla \cdot (b \rho_t) = -f(a, m, \varepsilon x_t) \rho_t,$$

$$(2.3b) \qquad \rho_t(0, m) = \mathbb{1}_{m > \gamma(0)} \left| (\gamma^{-1})'(m) \right| \int_0^\infty f(a, \gamma^{-1}(m), \varepsilon x_t) \rho_t(a, \gamma^{-1}(m)) da,$$

$$(2.3c) \qquad \rho_0 = u_0.$$

We can see (2.3) as the Kolmogorov forward equation of a time-dependent Markov process. Indeed, we can rewrite (2.3a) and (2.3b) as

$$(2.4) \qquad\qquad\qquad\qquad \partial \rho_t = \mathcal{L}_t \rho_t,$$

where, for all suitable test functions $\phi : \mathbb{R}_+ \times \mathbb{R}_+^* \to \mathbb{R}$,

$$(2.5) \qquad \mathcal{L}_t^* \phi(a, m) = b(a, m) \cdot \nabla \phi(a, m) + [\phi(0, \gamma(m)) - \phi(a, m)] f(a, m, \varepsilon x_t).$$

$\mathcal{L}_t^*$ is the time-dependent generator of a piecewise deterministic Markov process with degenerate jumps.

The linearized equation (2.3) will play a special role in the following sections and therefore deserves its own proposition.

PROPOSITION 2.2 (well-posedness of the linearized equation (2.3)). *Grant Assumption 1. For any initial datum $u_0 \in L_+^1$ and any $x \in \mathcal{C}(\mathbb{R}_+)$, there exists a unique weak solution $\rho^x \in \mathcal{C}(\mathbb{R}_+, L_+^1)$ to (2.3). Furthermore, ρ^x satisfies the following:*
(i) *For all $t > 0$ and for all $m \in \mathbb{R}_+^*$,*

$$\rho_t^x(0, m) = \mathbb{1}_{m > \gamma(0)} \left| (\gamma^{-1})'(m) \right| \int_0^\infty f(a, \gamma^{-1}(m), \varepsilon x) \rho_t^x(a, \gamma^{-1}(m)) da,$$

$$\rho_t^x(a, m) =$$
$$\begin{cases} u_0(a - t, e^{\lambda t} m) \exp \left(\lambda t - \int_0^t f(a - t + s, e^{\lambda(t-s)} m, \varepsilon x) ds \right) \text{ if } a \geq t, \\ \rho_{t-a}^x(0, e^{\lambda a} m) \exp \left(\lambda a - \int_{t-a}^t f(a - t + s, e^{\lambda(t-s)} m, \varepsilon x) ds \right) \text{ if } 0 < a < t. \end{cases}$$

(ii) *For all $t > 0$ and for all $\phi \in \mathcal{C}_c^\infty(\mathbb{R}_+ \times \mathbb{R}_+^*)$,*

$$(2.6) \qquad\qquad \langle \rho_t^x, \phi \rangle = \langle u_0, \phi \rangle + \int_0^t \langle \rho_s^x, \mathcal{L}_s^* \phi \rangle ds.$$

Proof. Fix $x \in \mathcal{C}(\mathbb{R}_+)$. For all $p \in \mathcal{C}(\mathbb{R}_+, L_+^1(\mathbb{R}_+^*))$ and $u_0 \in L_+^1$, we know from the standard theory of transport equations that there is a unique weak solution to

$$\partial_t \rho_t + \nabla \cdot (b \rho_t) = -f(a, m, \varepsilon x_t) \rho_t,$$
$$\rho_t(0, m) = p_t(m),$$
$$\rho_0 = u_0,$$

which we denote by $\rho^{x,p}$, and it is given by the representation formula,

$$\rho_t^{x,p}(a, m) := \begin{cases} u_0(a - t, e^{\lambda t} m) \exp \left(\lambda t - \int_0^t f(a - t + s, e^{\lambda(t-s)} m, \varepsilon x_s) ds \right) \text{ if } a \geq t, \\ p_{t-a}(e^{\lambda a} m) \exp \left(\lambda a - \int_{t-a}^t f(a - t + s, e^{\lambda(t-s)} m, \varepsilon x_s) ds \right) \text{ if } 0 < a < t. \end{cases}$$

The solution $\rho^{x,p}$ is in $\mathcal{C}(\mathbb{R}_+, L^1)$ since

$$\forall t \in \mathbb{R}_+, \qquad \left\| \rho_t^{x,p} \right\|_{L^1} \leq \|u_0\|_{L^1} + \int_0^t \|p_s\|_{L^1} \, ds.$$

We have

$$\left(\mathbb{1}_{m > \gamma(0)} \left| (\gamma^{-1})'(m) \right| \int_0^\infty f(a, \gamma^{-1}(m), x) \rho_t^{x,p}(a, \gamma^{-1}(m)) da \right)_{(t,m) \in \mathbb{R}_+ \times \mathbb{R}_+^*}$$
$$\in \mathcal{C}(\mathbb{R}_+, L_+^1(\mathbb{R}_+^*))$$

since

$$\forall t \in \mathbb{R}_+, \qquad \int_{\gamma(0)}^\infty \left| (\gamma^{-1})'(m) \right| \int_0^\infty f(a, \gamma^{-1}(m), x) \rho_t^{x,p}(a, \gamma^{-1}(m)) da \, dm$$
$$\leq \|f\|_\infty \left\| \rho_t^{x,p} \right\|_{L^1} \leq \|f\|_\infty \left(\|u_0\|_{L^1} + \int_0^t \|p_s\|_{L^1} \, ds \right).$$

Hence, we can define, for any $T > 0$, the operator Φ_T^x:

$$\mathcal{C}([0,T], L_+^1(\mathbb{R}_+^*)) \to \mathcal{C}([0,T], L_+^1(\mathbb{R}_+^*))$$
$$p \mapsto \left(\mathbb{1}_{m > \gamma(0)} \left| (\gamma^{-1})'(m) \right| \int_0^\infty f(a, \gamma^{-1}(m), x) \rho_t^{x,p}(a, \gamma^{-1}(m)) da \right)_{(t,m) \in [0,T] \times \mathbb{R}_+^*}.$$

For any $p, q \in \mathcal{C}([0,T], L_+^1(\mathbb{R}_+^*))$,

$$\left\| \Phi_T^x(p) - \Phi_T^x(q) \right\|_{\mathcal{C}([0,T], L^1)} \leq \|f\|_\infty \sup_{t \in [0,T]} \left\| \rho_t^{x,p} - \rho_t^{x,q} \right\|_{L^1}$$
$$\leq \|f\|_\infty \int_0^T \|p_s - q_s\|_{L^1} \, ds$$
$$\leq T \|f\|_\infty \|p - q\|_{\mathcal{C}([0,T], L^1)}.$$

Therefore, if $0 < T < \|f\|_\infty^{-1}$, Φ_T^x is a contraction. By Banach's fixed-point theorem, there exists a unique $\rho^x \in \mathcal{C}([0,T], L_+^1)$ solving (2.3). Since the choice of the contracting T does not depend on the initial datum, we can iterate the above argument on successive time intervals of length T and conclude that there exists a unique $\rho^x \in \mathcal{C}(\mathbb{R}_+, L_+^1)$ solving (2.3), for which formula (i) is satisfied. Then, (ii) follows from a standard cut-off-in-time argument. $\qquad \square$

Now, we can prove the existence and uniqueness of a solution to the nonlinear problem (1.1) by means of a second application of Banach's fixed-point theorem.

Proof of the well-posedness of (1.1) *in* L^1. For any $x \in \mathcal{C}(\mathbb{R}_+)$, we take the ρ^x given by Proposition 2.2. We have

$$\left(\int_0^t \int_{\mathbb{R}_+ \times \mathbb{R}_+^*} h(t-s) f(\varepsilon x_s) \rho_s^x \, da \, dm \, ds \right)_{t \in \mathbb{R}_+} \in \mathcal{C}(\mathbb{R}_+)$$

since

$$\forall t \in \mathbb{R}_+, \qquad \left| \int_0^t \int_{\mathbb{R}_+ \times \mathbb{R}_+^*} h(t-s) f(\varepsilon x_s) \rho_s^x \, da \, dm \, ds \right| \leq \|h\|_\infty \|f\|_\infty \int_0^t \|\rho_s^x\|_{L^1} \, ds.$$

Hence, for any $T > 0$, we can define the operator

$$\Psi_T : \mathcal{C}([0,T]) \to \mathcal{C}([0,T])$$

$$x \mapsto \left(\int_0^t \int_{\mathbb{R}_+ \times \mathbb{R}_+^*} h(t-s) f(\varepsilon x_s) \rho_s^x \, da \, dm \, ds \right)_{t \in [0,T]}.$$

For any $x, y \in \mathcal{C}([0,T])$, we have

$$\left\| \Psi_T(x) - \Psi_T(y) \right\|_{\mathcal{C}([0,T])} \leq T \|h\|_\infty \sup_{t \in [0,T]} \int_{\mathbb{R}_+ \times \mathbb{R}_+^*} |f(\varepsilon x_t) \rho_t^x - f(\varepsilon y_t) \rho_t^y| \, da \, dm$$

$$\leq T \|h\|_\infty \sup_{t \in [0,T]} \left(\varepsilon L_f |x_t - y_t| \left\| \rho_t^x \right\|_{L^1} + \|f\|_\infty \left\| \rho_t^x - \rho_t^y \right\|_{L^1} \right).$$

By Grönwall's lemma, $\|\rho_t^x\|_{L^1} \leq \|u_0\|_{L^1} \exp(\|f\|_\infty t)$ since

$$\forall t \in [0,T], \qquad \|\rho_t^x\|_{L^1} \leq \|u_0\|_{L^1} + \|f\|_\infty \int_0^t \|\rho_s^x\|_{L^1} \, ds.$$

On the other hand, we have, for all $t \in [0,T]$,

$$\left\| \rho_t^x - \rho_t^y \right\|_{L^1} \leq \int_0^t \int_0^\infty \left| \rho_s^x(0,m) \exp \left(-\int_s^t f(u-s, e^{-\lambda(u-s)} \gamma(m), \varepsilon x_u) du \right) \right.$$

$$\left. - \rho_s^y(0,m) \exp \left(-\int_s^t f(u-s, e^{-\lambda(u-s)} \gamma(m), \varepsilon y_u) du \right) \right| dm \, ds$$

$$\leq \|f\|_\infty \int_0^t \|\rho_s^x - \rho_s^y\|_{L^1} \, ds + t\varepsilon \|f\|_\infty L_f \|x - y\|_{\mathcal{C}([0,T])} \int_0^t \|\rho_s^x\|_{L^1} \, ds.$$

Hence, by Grönwall's lemma, for all $t \in [0,T]$,

$$\left\| \rho_t^x - \rho_t^y \right\|_{L^1} \leq \varepsilon L_f \|u_0\|_{L^1} \frac{\left(\exp(\|f\|_\infty t) - 1 \right)^2}{\|f\|_\infty} \|x - y\|_{\mathcal{C}([0,T])}.$$

Gathering the bounds, we get

$$\left\| \Psi_T(x) - \Psi_T(y) \right\|_{\mathcal{C}([0,T])}$$
$$\leq T\varepsilon \|h\|_\infty L_f \|u_0\|_{L^1} \exp(\|f\|_\infty T) \left[1 + \exp(\|f\|_\infty T) \right] \|x - y\|_{\mathcal{C}([0,T])}.$$

For T small enough, Ψ_T is a contraction and, by Banach's fixed-point theorem, has a unique fixed-point. Thus, there exists a unique solution $(\rho, x) \in \mathcal{C}([0,T], L_+^1)$. Since, by Lemma 2.1, $\|\rho_T\|_{L^1} = \|u_0\|_{L^1}$, we can iterate this argument on successive time intervals of length T and conclude that there exists a unique solution in $\mathcal{C}(\mathbb{R}_+, L_+^1)$. □

To conclude the proof of Theorem 1.2, it remains to show the estimate (1.5). Under Assumption 2, the weight function

$$w : \mathbb{R}_+ \times \mathbb{R}_+ \to [1, \infty[, \qquad (a,m) \mapsto 1 + m$$

satisfies $w(a,m) \to \infty$ when $m \to \infty$ and the Lyapunov condition on m:

$$(2.7) \qquad \exists \alpha > 0, b \geq 0 \quad \text{such that} \quad \mathcal{L}_t^* w \leq -\alpha w + b.$$

Indeed, for all $(t, a, m) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+^*$,

$$\mathcal{L}_t^* w(a, m) = -\lambda m + \Gamma(m) f(a, m, \varepsilon x_t) \leq -\lambda w(a, m) + \lambda + \|\Gamma\|_\infty \|f\|_\infty \,.$$

Importantly, the constants α and b do not depend on x.

LEMMA 2.3 (global bound in $L_+^1(w)$). *Grant Assumptions 1 and 2. If the initial datum u_0 is in $L_+^1(w)$, then $\rho_t \in L_+^1(w)$ for all $t \geq 0$. Moreover,*

$$(2.8) \qquad \forall t > 0, \qquad \|\rho_t\|_{L^1(w)} \leq \|u_0\|_{L^1(w)} \, e^{-\alpha t} + \frac{b}{\alpha}(1 - e^{-\alpha t})\|u_0\|_{L^1} \,,$$

where the constants α and b are taken from the Lyapunov condition (2.7).

Proof. We divide the proof into two steps. First, we prove that the solution is stable in $L_+^1(w)$ with a weaker and time-dependent bound; then we use this first bound to apply the dominated convergence theorem and obtain (2.8) by Grönwall's lemma.

Step 1. Fix any $T > 0$. Let $\chi \in \mathcal{C}_c^\infty(\mathbb{R}_+, \mathbb{R}_+)$ be a nonincreasing function such that $\chi(x) = 1$ if $0 \leq x \leq 1$ and $\chi(x) = 0$ if $x > 2$. For all $n \in \mathbb{N}^*$, let us write $\varphi_k(a)\chi_n(m) := \chi(a/k)\chi(m/n)$. We also consider $g_M(w)$ a smooth approximation of $w \wedge M$, such that $\|g'\|_\infty \leq 1$ and $M\mathbb{1}_{w \geq M} \leq g(w) \leq M$. For all n, k, and M, $g_M(w)\chi_n\varphi_k \in \mathcal{C}_c^\infty(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R}_+)$. Hence (by Proposition 2.2(ii)) the solution (ρ, x) to (1.1) satisfies

$$\forall n \in \mathbb{N}^*, \qquad \langle \rho_T, g_M(w)\chi_n\varphi_k \rangle = \langle u_0, g_M(w)\chi_n\varphi_k \rangle + \int_0^T \langle \rho_t, \mathcal{L}_x^*(g_M(w)\chi_n\varphi_k) \rangle dt,$$

where

$$\begin{aligned}
\mathcal{L}_x^*(g_M(w)\chi_n\varphi_k) &= \partial_a(g_M(w)\chi_n\varphi_k) - \lambda m \partial_m(g_M(w)\chi_n\varphi_k) \\
&\quad + \big(g_M(w(\gamma(m)))\chi_n(\gamma(m))\varphi_k(0) - g_M(w)\chi_n\varphi_k\big) f \\
&= g_M(w)\chi_n \frac{1}{k}\chi'(a/k) - \lambda m g_M(w)\varphi_k \frac{1}{n}\chi'(m/n) - \lambda m g_M'(w)\chi_n\varphi_k \\
&\quad + \big(g_M(w(\gamma(m)))\chi_n(\gamma(m))\varphi_k(0) - g_M(w)\chi_n\varphi_k\big) f.
\end{aligned}$$

From the L^1-stability and the fact that both $g_M(w)\partial_m\chi_n$ and $g_M(w)\chi_n$ are bounded and have compact support, we can go to the limit in k by dominated convergence:

$$(2.9) \quad \langle \rho_T, g_M(w)\chi_n \rangle = \langle u_0, g_M(w)\chi_n \rangle$$

$$+ \int_0^T \left\langle \rho_t, -\lambda m g_M(w)\frac{1}{n}\chi'(m/n) - \lambda m g_M'(w)\chi_n \right\rangle dt$$

$$+ \int_0^T \left\langle \rho_t, (g_M(w(\gamma(m)))\chi_n(\gamma(m)) - g_M(w)\chi_n)f \right\rangle dt.$$

On the other hand, from the properties of χ and g_M, we have

$$\left| \lambda m g_M(w)\frac{1}{n}\chi'(m/n) \right| \leq \lambda g_M(w)\frac{2n}{n}\|\chi'\|_\infty \leq 2\lambda M\|\chi'\|_\infty$$

and

$$|\lambda m g_M'(w)\chi_n| \leq \lambda g_M(w) \leq \lambda M,$$

whence

$$\langle \rho_T, g_M(w)\chi_n \rangle \leq \langle u_0, g_M(w)\chi_n \rangle$$
$$+ \int_0^T \langle \rho_t, \lambda g_M(w)\|\chi'\|_\infty + \lambda g_M(w) \rangle dt$$
$$+ \int_0^T \langle \rho_t, (g_M(w(\gamma(m)))\chi_n(\gamma(m)) - g_M(w)\chi_n)f \rangle dt,$$

and we can take the limit in n by dominated convergence:

$$\langle \rho_T, g_M(w) \rangle \leq \langle u_0, g_M(w) \rangle$$
$$+ \int_0^T \langle \rho_t, \lambda g_M(w)\|\chi'\|_\infty + \lambda g_M(w) \rangle dt$$
$$+ \int_0^T \langle \rho_t, (g_M(w(\gamma(m))) - g_M(w))f \rangle dt.$$

From the properties of γ, we get

$$w(0, \gamma(m)) \leq w(0, m + \|\Gamma\|_\infty) \leq (1 + \|\Gamma\|_\infty)w(a, m)$$

and

$$g_M(w(0, \gamma(m))) \leq (1 + \|\Gamma\|_\infty)g_M(w(m)).$$

This, together with the fact that f is bounded, shows that there exists a constant C, which does not depend on M, such that

$$\langle \rho_T, g_M(w) \rangle \leq \langle u_0, g_M(w) \rangle + C \int_0^T \langle \rho_t, g_M(w) \rangle dt,$$

and we can apply Grönwall's lemma to obtain

$$\langle \rho_T, g_M(w) \rangle \leq \langle u_0, g_M(w) \rangle e^{Ct}.$$

Finally, it follows from Fatou lemma that $\rho_T \in L^1_+(w)$.

Step 2. To improve the previous estimate, we come back to (2.9) and use dominated convergence in n and M (domination being guaranteed by Step 1) to show

$$\langle \rho_T, w \rangle = \langle u_0, w \rangle + \int_0^T \langle \rho_t, \mathcal{L}_x^* w \rangle dt.$$

By the Lyapunov condition (2.7),

$$\|\rho_T\|_{L^1(w)} \leq \|u_0\|_{L^1(w)} - \alpha \int_0^T \|\rho_t\|_{L^1(w)} dt + Tb,$$

and by Grönwall's lemma, we have, for all $t \in [0, T]$,

$$\|\rho_T\|_{L^1(w)} \leq \|u_0\|_{L^1(w)} e^{-\alpha t} + \frac{b}{\alpha}(1 - e^{-\alpha t}).$$

Since T can be chosen arbitrarily large, this achieves the proof. $\quad\square$

Remark 2.4. Following the same steps as in the proof above, we can show that the bound (2.8) also holds for the linearized equation (2.3) and does not depend on x or the constants α and b.

3. Exponential stability in the noninteracting case. If $x \in \mathcal{C}(\mathbb{R}_+)$ in the linearized equation (2.3) is time-invariant, i.e., $x \equiv \tilde{x}$ for some $\tilde{x} \in \mathbb{R}$, then (2.3) can be seen as the dynamics of a noninteracting population of neurons. In this section, we prove the exponential stability in the noninteracting case using Harris's or Doeblin's theorem. This is the key result of this work and will allow us to prove the existence and uniqueness of the stationary solution to (1.1) (section 4) and the exponential convergence towards it (section 5).

For $\tilde{x} \in \mathbb{R}$, $u_0 \in L^1$, we denote by $\rho_t^{\tilde{x}}$ the unique solution to (2.3) for the initial datum u_0 and for $x \equiv \tilde{x}$ given by Proposition 2.2. We write, using the semigroup notation,

$$(3.1) \qquad S_t^{\tilde{x}} u_0 := \rho_t^{\tilde{x}} \qquad \forall t \geq 0.$$

To show that the semigroup (3.1) is exponentially stable, we will use Harris's theorem in the general case or Doeblin's theorem if Assumption 4 is granted. The original theorems of Doeblin [11] and Harris [24] have since been refined and extended—see the well-known works of Meyn and Tweedie [29] and Hairer and Mattingly [23]. More recently, these theorems have been generalized to stochastic semigroups [16, 5, 1, 4]. Below, we give general statements of Doeblin's and Harris's theorems. For completeness, a short yet enlightening proof of Doeblin's theorem is presented in Appendix B. A proof of Harris's theorem can be found in the recent work of Cañizo and Mischler [4].

Let Ω denote a general state-space, and let $(S_t)_{t \geq 0}$ be a stochastic semigroup; i.e., for all $t \geq 0$, S_t is a mass and positivity preserving linear operator on $L^1(\Omega)$; S_0 is the identity operator; and for all $t, s \geq 0$, $S_t S_s = S_{s+t}$. We say that $\rho_\infty \in L^1_+(\Omega)$ is an *invariant probability measure* of the semigroup $(S_t)_{t \geq 0}$ if $\|\rho_\infty\|_{L^1} = 1$ and if, for all $t \geq 0$, $S_t \rho_\infty = \rho_\infty$.

THEOREM 3.1 (Doeblin). *If there exist $T > 0$ and a nonzero $\nu \in L^1_+$ such that*

$$(3.2) \qquad S_T u_0 \geq \nu \|u_0\|_{L^1} \quad \forall u_0 \in L^1_+,$$

then there exists a unique invariant probability measure ρ_∞, and for all initial data $u_0 \in L^1_+$ with $\|u_0\|_{L^1} = 1$,

$$(3.3) \qquad \|S_t u_0 - \rho_\infty\|_{L^1} \leq K e^{-\mathfrak{a} t} \|u_0 - \rho_\infty\|_{L^1} \quad \forall t \geq 0,$$

with

$$K = \frac{1}{1 - \|\nu\|_{L^1}}; \quad \mathfrak{a} = -\frac{\log(1 - \|\nu\|_{L^1})}{T} > 0.$$

We call (3.2) the Doeblin minoration condition. Very loosely speaking, the Doeblin minoration condition is best suited for compact state-spaces (but see [16, 5] for examples on $\mathbb{R}_+$). In the case of unbounded state-spaces, Harris's theorem tells us that the Doeblin minoration condition can be relaxed to a more local form if there is a Lyapunov-type localization condition.

THEOREM 3.2 (Harris). *Let $w : \Omega \to [1, +\infty)$ be a measurable weight function. If there exists $T > 0$ such that*
   (i) *(operator Lyapunov condition) there exist $A \in]0, 1[$ and $B \geq 0$ such that*

$$(3.4) \qquad \|S_T u_0\|_{L^1(w)} \leq A \|u_0\|_{L^1(w)} + B \|u_0\|_{L^1} \quad \forall u_0 \in L^1_+(w);$$

(ii) *(Harris minoration condition) there exist a nonzero $\nu \in L^1_+$ and $R > 0$ such that*

$$(3.5) \qquad S_T u_0 \geq \nu \int_{\mathfrak{C}} u_0 \quad \forall u_0 \in L^1_+, \quad \text{with } \mathfrak{C} = \{x \in \Omega \mid w(x) \leq R\},$$

then there exists a unique invariant probability measure ρ_∞ such that $\rho_\infty \in L^1_+(w)$,[1] and there exist $K \geq 1$ and $\mathfrak{a} > 0$ such that for all initial data $u_0 \in L^1_+(w)$ with $\|u_0\|_{L^1} = 1$,

$$(3.6) \qquad \|S_t u_0 - \rho_\infty\|_{L^1(w)} \leq K e^{-\mathfrak{a}t} \|u_0 - \rho_\infty\|_{L^1(w)} \quad \forall t \geq 0.$$

For the model (1.1) considered in the present work, Assumption 3 will be necessary to show the Doeblin or Harris minoration conditions. We then have to distinguish two cases: either Assumption 2 holds, and we can use Harris's theorem since the Lyapunov condition (2.7) implies the operator Lyapunov condition (3.4) by Lemma 2.3 (the constants are then $A = e^{-\alpha T}$ and $B = \frac{b}{\alpha}$); or Assumption 4 holds, and we can simply use Doeblin's theorem. The main technical difficulty is to verify the minoration conditions, as the jumps of the process described by (2.5) are degenerate, and the model is two-dimensional. The rest of the section is devoted to the verification of the minoration condition.

LEMMA 3.3 (minoration condition). *Grant Assumptions 1 and 3. Fix any $x \in \mathbb{R}$. For all $R > 0$, there exist $T > 0$ and a nonzero $\nu \in L^1_+$ such that*

$$(3.7) \qquad S^{\tilde{x}}_T u_0 \geq \nu \int_{\mathbb{R}_+ \times \,]0,R]} u_0 \, da \, dm \quad \forall u_0 \in L^1_+.$$

If, in addition, Assumption 4 holds, then there exist $T > 0$ and a nonzero $\nu \in L^1_+$ such that

$$(3.8) \qquad S^{\tilde{x}}_T u_0 \geq \nu \|u_0\|_{L^1} \quad \forall u_0 \in L^1_+.$$

Proof. We proceed in two steps. First (Step 1), we choose a time $T > 0$ and a rectangle $[0, \bar{a}] \times [\underline{m}, \overline{m}] \subset \mathbb{R}_+ \times \mathbb{R}^*_+$ (with nonzero Lebesgue measure) and show that the density $S^{\tilde{x}}_T u_0 \in L^1$ has a lower bound on $[0, \bar{a}] \times [\underline{m}, \overline{m}]$ which depends on a Lebesgue integral in $\mathbb{R}^2_+$ involving u_0. Then (Step 2), we perform a change of variable to express this lower bound in terms of $\int_{\mathbb{R}_+ \times \,]0,R]} u_0 \, da \, dm$. The proof only relies on the expression of $S^{\tilde{x}}_t u_0$ given by the method of characteristics (see Proposition 2.2), and this allows treating a typically probabilistic question—the Doeblin/Harris minoration condition—from a transport point of view. This is possible because $S^{\tilde{x}}_t$ is the stochastic (mass-conservative) semigroup of a piecewise deterministic Markov process.

The constants Δ_{abs}, σ, and C_γ are taken from Assumption 3.

Step 1. Fix $R > 0$. Since $\gamma(e^{-\lambda \Delta_{\mathrm{abs}}} \gamma(0)) > \gamma(0)$ and $\gamma(e^{-\lambda t} \gamma(e^{-\lambda \Delta_{\mathrm{abs}}} R)) \to \gamma(0)$ as $t \to \infty$, there exist $\bar{a} > 0$ and $T > \bar{a} + \Delta_{\mathrm{abs}}$ such that

$$(3.9) \qquad \underline{m} =: \gamma(e^{-\lambda(T - \bar{a} - \Delta_{\mathrm{abs}})} \gamma(e^{-\lambda \Delta_{\mathrm{abs}}} R)) < e^{-\lambda \bar{a}} \gamma(e^{-\lambda \Delta_{\mathrm{abs}}} \gamma(0)) =: \overline{m}.$$

Equation (3.9) has the following heuristic interpretation: if we see $S^{\tilde{x}}_t$ as the stochastic semigroup of the piecewise deterministic Markov process defined by the generator

---

[1]Note that Harris's theorem does not exclude the existence of an invariant probability measure with infinite $L^1(w)$ norm.

(2.5), for any initial point $(a_0, m_0) \in \mathbb{R}_+ \times \,]0, R]$ and any landing point $(a, m) \in [0, \bar{a}] \times [\underline{m}, \overline{m}]$ at time T, there is a "possible" trajectory going from (a_0, m_0) to (a, m), with exactly two jumps (spikes). Since the trajectories of the process are determined by the jump times, we will exploit the fact that these "possible" trajectories correspond to jump times with strictly positive probability density. Below, we take a transport point of view on this probabilistic argument.

For all $(a, m) \in [0, \bar{a}] \times [\underline{m}, \overline{m}]$,

$$(S_T^{\tilde{x}} u_0)(a, m)$$

$$\geq \mathbb{1}_{\{a < T\}} (S_{T-a}^{\tilde{x}} u_0)(0, e^{\lambda a} m) \exp\left(\lambda a - \int_{T-a}^{T} f(a - T + s, e^{\lambda(T-s)} m, \tilde{x}) ds\right)$$

$$\geq \mathbb{1}_{\{a < T\}} e^{-\|f\|_\infty T} e^{\lambda a} (S_{T-a}^{\tilde{x}} u_0)(0, e^{\lambda a} m)$$

$$\geq \mathbb{1}_{\{a < T\}} e^{-\|f\|_\infty T} \sigma e^{\lambda a} \left|(\gamma^{-1})'(e^{\lambda a} m)\right| \int_{\Delta_{\text{abs}}}^{\infty} (S_{T-a}^{\tilde{x}} u_0)(a', \gamma^{-1}(e^{\lambda a} m)) da'$$

$$= \mathbb{1}_{\{a < T\}} e^{-\|f\|_\infty T} \sigma \left|\frac{d}{dm} \gamma^{-1}(e^{\lambda a} m)\right| \int_{\Delta_{\text{abs}}}^{\infty} \underbrace{(S_{T-a}^{\tilde{x}} u_0)(a', \gamma^{-1}(e^{\lambda a} m))}_{(\star)} da'.$$

Above, we went back in time to the last jump time $T-a$. Let us note that $\gamma^{-1}(e^{\lambda a} m) \geq \gamma^{-1}(e^{\lambda a} \underline{m}) > 0$. We can therefore define

$$a_{a,m}^* := \frac{1}{\lambda} \left(\log \gamma(0) - \log \gamma^{-1}(e^{\lambda a} m)\right).$$

Note that $a_{a,m}^*$ satisfies $\gamma^{-1}(e^{\lambda a_{a,m}^*} \gamma^{-1}(e^{\lambda a} m)) = 0$. In other words, $a_{a,m}^*$ is the minimal time between the last and second-to-last jumps for a trajectory landing at (a, m) at time T. We can easily verify that, by our choice of $\{T, \bar{a}, \underline{m}, \overline{m}\}$, $\Delta_{\text{abs}} \leq a_{a,m}^* < T - a - \Delta_{\text{abs}}$. This guarantees that it is possible to make two jumps in $[0, T]$ and land at (a, m) at time T while respecting the absolute refractoriness of the neuron (i.e., there needs to be a time interval $\geq \Delta_{\text{abs}}$ between jumps). This allows us to go further back in time to the second-to-last jump as follows: for all $a' \in [a_{a,m}^*, T - a - \Delta_{\text{abs}}]$,

$$(\star) \geq \mathbb{1}_{\{a' < T-a\}} e^{-\|f\|_\infty T} \sigma \left|(\gamma^{-1})'(e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m))\right| e^{\lambda a'}$$

$$\times \int_{\Delta_{\text{abs}}}^{\infty} \underbrace{(S_{T-a-a'}^{\tilde{x}} u_0)(a'', \gamma^{-1}(e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m)))}_{(\star\star)} da''.$$

Then, we can go further back to time 0 to get u_0:

$$(\star\star) \geq \mathbb{1}_{\{a'' \geq T-a-a'\}} e^{-\|f\|_\infty T} e^{\lambda(T-a-a')}$$

$$\times u_0(a'' - (T - a - a'), e^{\lambda(T-a-a')} \gamma^{-1}(e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m))).$$

Putting all the lower bounds together, we get

$$(S_T^{\tilde{x}} u_0)(a, m) \geq \mathbb{1}_{\{a < T\}} e^{-3\|f\|_\infty T} \sigma^2$$

$$\int_{a_{a,m}^*}^{T-a-\Delta_{\text{abs}}} \int_{T-a-a'}^{\infty} \left|\frac{d}{dm} e^{\lambda(T-a-a')} \gamma^{-1}(e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m))\right|$$

$$u_0(a'' - (T - a - a'), e^{\lambda(T-a-a')} \gamma^{-1}(e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m))) da'' da'.$$

Since $\gamma' \leq 1$ (Assumption 3),

$$\left| \frac{d}{dm} e^{\lambda(T-a-a')} \gamma^{-1}(e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m)) \right| \geq e^{\lambda T}.$$

Thus,

$$(S_T^{\tilde{x}} u_0)(a, m) \geq \mathbb{1}_{\{a<T\}} e^{(\lambda - 3\|f\|_\infty)T} \sigma^2$$

(3.10)
$$\times \int_{a_{a,m}^*}^{T-a-\Delta_{\mathrm{abs}}} \int_0^\infty u_0(a_0, e^{\lambda(T-a-a')} \gamma^{-1}(e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m))) da_0 da'.$$

We have obtained that on $[0, \bar{a}] \times [\underline{m}, \overline{m}]$, the density $(S_T^{\tilde{x}} u_0)$ is lower bounded by a constant depending on a Lebesgue integral on $\mathbb{R}_+^2$ involving u_0.

Step 2. Now, we want to express the lower bound (3.10) in terms of

$$\int_{\mathbb{R}_+ \times]0,R]} u_0 \, da \, dm$$

by a change of variable. Let us define the function $\psi_{a,m}^T$:

$$\psi_{a,m}^T : [a_{a,m}^*, T-a-\Delta_{\mathrm{abs}}] \to \mathbb{R}_+, \quad a' \mapsto e^{\lambda(T-a-a')} \gamma^{-1}(e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m)).$$

We verify that $(\psi_{a,m}^T)' > 0$ as follows: for all $a' \in [a_{a,m}^*, T-a]$,

(3.11)
$$(\psi_{a,m}^T)'(a')$$
$$= \lambda e^{\lambda(T-a-a')} \left\{ (\gamma^{-1})'(e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m)) e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m) - \gamma^{-1}(e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m)) \right\}.$$

As $\Gamma > 0$ and $\gamma' \leq 1$ (Assumption 3), we have

$$(\psi_{a,m}^T)'(a') > \lambda e^{\lambda(T-a-a')} \left\{ (\gamma^{-1})'(e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m)) e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m) - e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m) \right\}$$

$$= \lambda e^{\lambda(T-a)} \gamma^{-1}(e^{\lambda a} m) \left\{ \underbrace{(\gamma^{-1})'(e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m))}_{\geq 1} - 1 \right\} \geq 0.$$

Therefore, $\psi_{a,m}^T$ is a strictly increasing $\mathcal{C}^1$-diffeomorphism from $[a_{a,m}^*, T-a-\Delta_{\mathrm{abs}}]$ to $[\psi_{a,m}^T(a_{a,m}^*), \psi_{a,m}^T(T-a-\Delta_{\mathrm{abs}})]$. We can now rewrite (3.10) as

$$(S_T^{\tilde{x}} u_0)(a, m) \geq e^{(\lambda - 3\|f\|_\infty)T} \sigma^2 \int_{a_{a,m}^*}^{T-a-\Delta_{\mathrm{abs}}} \int_0^\infty u_0(a_0, \psi_{a,m}^T(a')) da_0 da'$$

$$= e^{(\lambda - 3\|f\|_\infty)T} \sigma^2 \int_{\psi_{a,m}^T(a_{a,m}^*)}^{\psi_{a,m}^T(T-a-\Delta_{\mathrm{abs}})} \int_0^\infty u_0(a_0, m_0) \left| ((\psi_{a,m}^T)^{-1})'(m_0) \right| da_0 dm_0.$$

Going back to (3.11) and using the fact that there exists C_γ such that $C_\gamma \leq \gamma' \leq 1$ (Assumption 3), we have, for all $a' \in [a_{a,m}^*, T-a-\Delta_{\mathrm{abs}}]$,

$$(\psi_{a,m}^T)'(a') \leq \lambda e^{\lambda(T-a-a')} C_\gamma^{-1} e^{\lambda a'} \gamma^{-1}(e^{\lambda a} m) \leq \lambda e^{\lambda T} C_\gamma^{-1} \overline{m}.$$

Hence,

$$(S_T^{\tilde{x}}u_0)(a,m) \geq \frac{e^{-3\|f\|_\infty T}\sigma^2 C_\gamma}{\lambda\overline{m}} \int_{\psi_{a,m}^T(a_{a,m}^*)}^{\psi_{a,m}^T(T-a-\Delta_{\mathrm{abs}})} \int_0^\infty u_0(a_0,m_0)da_0 dm_0.$$

In addition, by our choice of $\{T,\bar{a},\underline{m},\overline{m}\}$, we have

$$\psi_{a,m}^T(a_{a,m}^*) = 0,$$
$$\psi_{a,m}^T(T-a-\Delta_{\mathrm{abs}}) = e^{\lambda\Delta_{\mathrm{abs}}}\gamma^{-1}(e^{\lambda(T-a-\Delta_{\mathrm{abs}})}\gamma^{-1}(e^{\lambda a}m)) > R.$$

Therefore,

$$(S_T^{\tilde{x}}u_0)(a,m) \geq \frac{e^{-3\|f\|_\infty T}\sigma^2 C_\gamma}{\lambda\overline{m}} \int_0^R \int_0^\infty u_0(a_0,m_0)da_0 dm_0.$$

Since we have assumed that $(a,m) \in [0,\bar{a}] \times [\underline{m},\overline{m}]$, this concludes the proof of (3.7).

In the case where Assumption 4 also holds, the proof of (3.8) is similar except that we can simply take $R = +\infty$ and $\underline{m} =: \gamma(e^{-\lambda(T-\bar{a}-\Delta_{\mathrm{abs}})}G) < e^{-\lambda\bar{a}}\gamma(e^{-\lambda\Delta_{\mathrm{abs}}}\gamma(0)) =: \overline{m}$. $\square$

In summary, by Harris's theorem, we have the following.

THEOREM 3.4. *Grant Assumptions 1–3. For all $\tilde{x} \in \mathbb{R}$, there exists a unique $\rho_\infty^{\tilde{x}} \in L_+^1(w)$ with $\|\rho_\infty^{\tilde{x}}\|_{L^1} = 1$ such that $S_t^{\tilde{x}}\rho_\infty^{\tilde{x}} = \rho_\infty^{\tilde{x}}$ for all $t \geq 0$, and there exist $K \geq 1$ and $\mathfrak{a} > 0$ such that for all initial data $u_0 \in L_+^1(w)$ with $\|u_0\|_{L^1} = 1$,*

$$(3.12) \qquad \left\|S_t^{\tilde{x}}u_0 - \rho_\infty^{\tilde{x}}\right\|_{L^1(w)} \leq Ke^{-\mathfrak{a}t}\left\|u_0 - \rho_\infty^{\tilde{x}}\right\|_{L^1(w)} \qquad \forall t \geq 0.$$

Furthermore, by Lemma 2.3, we have that $\|\rho_\infty^{\tilde{x}}\|_{L^1(w)} \leq \frac{b}{\alpha}$, where the constants α and b are taken from the Lyapunov condition (2.7).

If Assumption 2 is replaced by Assumption 4, we can simply apply Doeblin's theorem.

THEOREM 3.5. *Grant Assumptions 1, 3, and 4. For all $\tilde{x} \in \mathbb{R}$, there exists a unique $\rho_\infty^{\tilde{x}} \in L_+^1$ with $\|\rho_\infty^{\tilde{x}}\|_{L^1} = 1$ such that $S_t^{\tilde{x}}\rho_\infty^{\tilde{x}} = \rho_\infty^{\tilde{x}}$ for all $t \geq 0$, and there exist $K \geq 1$ and $\mathfrak{a} > 0$ such that for all initial data $u_0 \in L_+^1$ with $\|u_0\|_{L^1} = 1$,*

$$(3.13) \qquad \left\|S_t^{\tilde{x}}u_0 - \rho_\infty^{\tilde{x}}\right\|_{L^1} \leq Ke^{-\mathfrak{a}t}\left\|u_0 - \rho_\infty^{\tilde{x}}\right\|_{L^1} \qquad \forall t \geq 0.$$

Note that both theorems imply the following.

COROLLARY 3.6. *Grant Assumptions 1–3 (or Assumptions 1, 3, and 4). For all $\tilde{x} \in \mathbb{R}$, there exists a unique $\rho_\infty^{\tilde{x}} \in L_+^1(w)$ (or $\in L_+^1$) with $\|\rho_\infty^{\tilde{x}}\|_{L^1} = 1$ solving*

$$(3.14a) \qquad \partial_a \rho_\infty^{\tilde{x}}(a,m) - \lambda\partial_m(m\rho_\infty^{\tilde{x}}(a,m)) = -f(a,m,\tilde{x})\rho_\infty^{\tilde{x}}(a,m),$$

$$(3.14b) \qquad \rho_\infty^{\tilde{x}}(0,m) = \mathbb{1}_{m>\gamma(0)}\left|(\gamma^{-1})'(m)\right|\int_0^\infty f(a,\gamma^{-1}(m),\tilde{x})\rho_\infty^{\tilde{x}}(a,\gamma^{-1}(m))da$$

in the weak sense. Furthermore, we have that $\rho_\infty^{\tilde{x}} \in \mathcal{C}(\mathbb{R}_+, L_+^1(\mathbb{R}_+^*)) \cap L^\infty(\mathbb{R}_+, L_+^1(\mathbb{R}_+^*))$.

4. Stationary solutions for arbitrary connectivity strength. In this section, we study the stationary solutions to (1.1), namely the solution to

$$
(4.1a) \quad \partial_a \rho_\infty(a,m) - \lambda \partial_m(m\rho_\infty(a,m)) = -f(a,m,\varepsilon x_\infty)\rho_\infty(a,m),
$$

$$
(4.1b) \quad \rho_\infty(0,m) = \mathbb{1}_{m>\gamma(0)} \left|(\gamma^{-1})'(m)\right| \int_0^\infty f(a,\gamma^{-1}(m),\varepsilon x_\infty)\rho_\infty(a,\gamma^{-1}(m))da,
$$

$$
(4.1c) \quad x_\infty = \int_0^\infty \int_0^\infty \bar{h}(a,m)f(a,m,\varepsilon x_\infty)\rho_\infty(a,m)dadm.
$$

DEFINITION 4.1. $(\rho_\infty, x_\infty) \in L^1_+(w) \cap \mathcal{C}(\mathbb{R}_+, L^1_+(\mathbb{R}^*_+)) \cap L^\infty(\mathbb{R}_+, L^1_+(\mathbb{R}^*_+)) \times \mathbb{R}_+$
is a stationary solution to (1.1) if $\|\rho_\infty\|_{L^1} = 1$ and if it solves (4.1) in the weak sense.

4.1. Existence and uniqueness using the Doeblin–Harris method. We present two Lipschitz continuity results, which will allow us to prove the existence (and the uniqueness when ε is small) of stationary solutions. The following lemma plays the same role as Theorem 4.5 in [5].

LEMMA 4.2 (Lipschitz continuity at finite T). *Grant Assumptions 1–3. For all initial data $u_0 \in L^1_+(w)$ and for all $T > 0$, there exists a constant $C_{T,\|u_0\|_{L^1(w)}} > 0$ such that*

$$
(4.2) \quad \forall \widetilde{x_1}, \widetilde{x_2} \in \mathbb{R}, \quad \left\|S_T^{\widetilde{x_1}}u_0 - S_T^{\widetilde{x_2}}u_0\right\|_{L^1(w)} \le C_{T,\|u_0\|_{L^1(w)}} |\widetilde{x_1} - \widetilde{x_2}|.
$$

Proof. For all $t > 0$,

$$
\left\|S_t^{\widetilde{x_1}}u_0 - S_t^{\widetilde{x_2}}u_0\right\|_{L^1(w)}
$$
$$
= \int_0^\infty \int_t^\infty \left| u_0(a-t, e^{\lambda t}m)\exp\left(\lambda t - \int_0^t f(a-t+s, e^{\lambda(t-s)}m, \widetilde{x_1})ds\right)\right.
$$
$$
\left. - u_0(a-t, e^{\lambda t}m)\exp\left(\lambda t - \int_0^t f(a-t+s, e^{\lambda(t-s)}m, \widetilde{x_2})ds\right)\right| w(a,m)dadm
$$
$$
+ \int_0^\infty \int_0^t \left| \rho_{t-a}^{\widetilde{x_1}}(0, e^{\lambda a}m)\exp\left(\lambda a - \int_{t-a}^t f(a-t+s, e^{\lambda(t-s)}m, \widetilde{x_1})ds\right)\right.
$$
$$
\left. - \rho_{t-a}^{\widetilde{x_2}}(0, e^{\lambda a}m)\exp\left(\lambda a - \int_{t-a}^t f(a-t+s, e^{\lambda(t-s)}m, \widetilde{x_2})ds\right)\right| w(a,m)dadm
$$
$$
=: Q_1 + Q_2.
$$

$$Q_1 = \int_0^\infty \int_0^\infty u_0(a,m) \left| \exp\left(-\int_0^t f(a+s, e^{-\lambda s}m, \widetilde{x_1}) ds \right) \right.$$

$$\left. - \exp\left(-\int_0^t f(a+s, e^{-\lambda s}m, \widetilde{x_2}) ds \right) \right| w(a+t, e^{-\lambda t}m) da dm$$

$$\leq \int_0^\infty \int_0^\infty u_0(a,m) \left(\int_0^t \left| f(a+s, e^{-\lambda s}m, \widetilde{x_1}) \right. \right.$$

$$\left. \left. - f(a+s, e^{-\lambda s}m, \widetilde{x_2}) \right| ds \right) w(a+t, e^{-\lambda t}m) da dm$$

$$\leq t L_f |\widetilde{x_1} - \widetilde{x_2}| \int_0^\infty \int_0^\infty u_0(a,m) w(a+t, e^{-\lambda t}m) da dm$$

$$\leq t L_f \|u_0\|_{L^1(w)} |\widetilde{x_1} - \widetilde{x_2}|,$$

where, in the last inequality, we used

(4.3)
$$w(a+t, e^{-\lambda t}m) \leq w(a,m) \quad \forall a \geq 0, m \geq 0.$$

$$Q_2 = \int_0^\infty \int_0^t \left| \rho_{t-a}^{\widetilde{x_1}}(0,m) \exp\left(-\int_{t-a}^t f(a-t+s, e^{\lambda(t-s-a)}m, \widetilde{x_1}) ds \right) \right.$$

$$\left. - \rho_{t-a}^{\widetilde{x_2}}(0,m) \exp\left(-\int_{t-a}^t f(a-t+s, e^{\lambda(t-s-a)}m, \widetilde{x_2}) ds \right) \right| w(a, e^{-\lambda a}m) da dm.$$

By changes of variables,

$$Q_2 = \int_0^\infty \int_0^t \left| \rho_s^{\widetilde{x_1}}(0,m) \exp\left(-\int_0^{t-s} f(u, e^{-\lambda u}m, \widetilde{x_1}) du \right) \right.$$

$$\left. - \rho_s^{\widetilde{x_2}}(0,m) \exp\left(-\int_0^{t-s} f(u, e^{-\lambda u}m, \widetilde{x_2}) du \right) \right| w(t-s, e^{-\lambda(t-s)}m) ds dm$$

$$\leq \int_0^\infty \int_0^t \rho_s^{\widetilde{x_1}}(0,m) \left| \exp\left(-\int_0^{t-s} f(u, e^{-\lambda u}m, \widetilde{x_1}) du \right) \right.$$

$$- \exp\left(-\int_0^{t-s} f(u, e^{-\lambda u}m, \widetilde{x_2}) du \right) \left| w(t-s, e^{-\lambda(t-s)}m) ds dm \right.$$

$$+ \int_0^\infty \int_0^t \left| \rho_s^{\widetilde{x_1}}(0,m) - \rho_s^{\widetilde{x_2}}(0,m) \right| w(t-s, e^{-\lambda(t-s)}m) ds dm$$

$$=: Q_{2,1} + Q_{2,2}.$$

$$Q_{2,1} \leq t\|f\|_\infty L_f |\widetilde{x_1} - \widetilde{x_2}| \int_0^t \int_0^\infty \int_0^\infty \left| (\gamma^{-1})'(m) \right| \rho_s^{\widetilde{x_1}}(a, \gamma^{-1}(m)) w(t,m) da dm ds$$

$$\leq t\|f\|_\infty L_f |\widetilde{x_1} - \widetilde{x_2}| \int_0^t \int_0^\infty \int_0^\infty \rho_s^{\widetilde{x_1}}(a,m) w(t, m + \|\Gamma\|_\infty) da dm ds$$

$$\leq t(1 + \|\Gamma\|_\infty) \|f\|_\infty L_f |\widetilde{x_1} - \widetilde{x_2}| \int_0^t \left\| \rho_s^{\widetilde{x_1}} \right\|_{L^1(w)} ds,$$

where, in the last inequality, we used

(4.4) $\quad w(t, m + \|\Gamma\|_\infty) = 1 + m + \|\Gamma\|_\infty \leq (1 + \|\Gamma\|_\infty)w(a, m) \quad \forall a \geq 0, m \geq 0.$

By Lemma 2.3,

$$Q_{2,1} \leq t^2(1 + \|\Gamma\|_\infty)\|f\|_\infty L_f\left(\|u_0\|_{L^1(w)} + \frac{b}{\alpha}\right)|\widetilde{x_1} - \widetilde{x_2}|.$$

$$\begin{aligned}
Q_{2,2} &\leq \|f\|_\infty \int_0^t \int_0^\infty \int_0^\infty \left|(\gamma^{-1})'(m)\right|\left|\rho_s^{\widetilde{x_1}}(a, \gamma^{-1}(m)) - \rho_s^{\widetilde{x_2}}(a, \gamma^{-1}(m))\right| w(t, m) da\, dm\, ds \\
&\leq \|f\|_\infty \int_0^t \int_0^\infty \int_0^\infty \left|\rho_s^{\widetilde{x_1}}(a, m) - \rho_s^{\widetilde{x_2}}(a, m)\right| w(t, m + \|\Gamma\|_\infty) da\, dm\, ds \\
&\leq (1 + \|\Gamma\|_\infty)\|f\|_\infty \int_0^t \left\|S_s^{\widetilde{x_1}}u_0 - S_s^{\widetilde{x_2}}u_0\right\|_{L^1(w)} ds,
\end{aligned}$$

where again, in the last inequality, we used (4.4). Fix $T > 0$. Gathering the bounds for Q_1, $Q_{2,1}$, and $Q_{2,2}$ we see that there exist constants $C > 0$ and $C'_{T,\|u_0\|_{L^1(w)}} > 0$ such that, for all $t \in [0, T]$,

$$\left\|S_t^{\widetilde{x_1}}u_0 - S_t^{\widetilde{x_2}}u_0\right\|_{L^1(w)} \leq C \int_0^t \left\|S_s^{\widetilde{x_1}}u_0 - S_s^{\widetilde{x_2}}u_0\right\|_{L^1(w)} ds + t C'_{T,\|u_0\|_{L^1(w)}}|\widetilde{x_1} - \widetilde{x_2}|.$$

By Grönwall's lemma, for all $t \in [0, T]$,

(4.5) $\quad \left\|S_t^{\widetilde{x_1}}u_0 - S_t^{\widetilde{x_2}}u_0\right\|_{L^1(w)} \leq \frac{C'_{T,\|u_0\|_{L^1(w)}}|\widetilde{x_1} - \widetilde{x_2}|}{C}\left(\exp(Ct) - 1\right).$

Since (4.5) holds for all $t \in [0, T]$, this achieves the proof. $\qquad\square$

LEMMA 4.3 (Lipschitz continuity at $T = \infty$). *Grant Assumptions 1–3. Writing* $\rho_\infty^{\tilde{x}} \in L^1_+(w)$, *the invariant probability measure given by Theorem 3.2 for any* $\tilde{x} \in \mathbb{R}$, *we see that the function*

$$\Upsilon : \mathbb{R}_+ \to \mathbb{R}_+, \qquad \Upsilon(x) = \int_0^\infty \int_0^\infty \bar{h}(a, m)f(a, m, \varepsilon x)\rho_\infty^{\varepsilon x}(a, m)da\, dm$$

is Lipschitz, and there exists $C > 0$ such that

$$\forall x_1, x_2 \in \mathbb{R}_+, \qquad |\Upsilon(x_1) - \Upsilon(x_2)| \leq |\varepsilon|C|x_1 - x_2|.$$

Proof. Since f is Lipschitz in x, we have, for any $x_1, x_2 \in \mathbb{R}_+$,

$$\begin{aligned}
|\Upsilon(x_1) - \Upsilon(x_2)| &\leq \left\|\bar{h}\right\|_\infty \left\{\|f\|_\infty \|\rho_\infty^{\varepsilon x_1} - \rho_\infty^{\varepsilon x_2}\|_{L^1} + L_f|\varepsilon||x_1 - x_2|\right\} \\
&\leq \left\|\bar{h}\right\|_\infty \left\{\|f\|_\infty \|\rho_\infty^{\varepsilon x_1} - \rho_\infty^{\varepsilon x_2}\|_{L^1(w)} + L_f|\varepsilon||x_1 - x_2|\right\},
\end{aligned}$$

from where we only need to bound the first term on the right-hand side. We can use Theorem 3.2 and Lemma 4.2 as follows: for any $T \in \mathbb{R}_+$,

$$\begin{aligned}
\|\rho_\infty^{\varepsilon x_1} - \rho_\infty^{\varepsilon x_2}\|_{L^1(w)} &= \|S_T^{\varepsilon x_1}\rho_\infty^{\varepsilon x_1} - S_T^{\varepsilon x_1}\rho_\infty^{\varepsilon x_2} + S_T^{\varepsilon x_1}\rho_\infty^{\varepsilon x_2} - S_T^{\varepsilon x_2}\rho_\infty^{\varepsilon x_2}\|_{L^1(w)} \\
&\leq Ke^{-\mathfrak{a}T}\|\rho_\infty^{\varepsilon x_1} - \rho_\infty^{\varepsilon x_2}\|_{L^1(w)} + C_T|\varepsilon||x_1 - x_2|,
\end{aligned}$$

where K and $\mathfrak{a}$ are the exponential stability constants of Theorem 3.2. Choosing T such that $Ke^{-\mathfrak{a}T} = 1/2$, we get

$$\|\rho_\infty^{\varepsilon x_1} - \rho_\infty^{\varepsilon x_2}\|_{L^1(w)} \leq 2C_T|\varepsilon||x_1 - x_2|.$$

Gathering the bounds concludes the proof. $\qquad\square$

THEOREM 4.4 (stationary solutions). *Grant Assumptions 1–3. We have the following:*
  (i) *There exists a stationary solution to (1.1).*
  (ii) *There exists $\varepsilon^* > 0$ such that for all $\varepsilon \in\ ]-\varepsilon^*, +\varepsilon^*[$, the stationary solution to (1.1) is unique.*

Proof. For all $\tilde{x} \in \mathbb{R}$, let us write $\rho_\infty^{\tilde{x}} \in L_+^1(w)$ the unique invariant measure given by Theorem 3.2, and let us also take the function Υ from Lemma 4.3. By Corollary 3.6, $(\rho_\infty, x_\infty) \in L_+^1(w) \cap \mathcal{C}(\mathbb{R}_+, L_+^1(\mathbb{R}_+^*)) \cap L^\infty(\mathbb{R}_+, L_+^1(\mathbb{R}_+^*)) \times \mathbb{R}_+$ is a weak solution to (4.1) if and only if $\rho_\infty = \rho_\infty^{\varepsilon x_\infty}$ and x_∞ is a fixed-point of Υ. Hence, the study of the existence and uniqueness of stationary solutions is reduced to the study of the existence and uniqueness of the fixed-point of Υ.

Since for all $x \in \mathbb{R}_+, \|\rho_\infty^{\varepsilon x}\|_{L^1} = 1$, we have that for all $x \in \mathbb{R}_+$, $\Upsilon(x) \leq \|\bar{h}\|_\infty \|f\|_\infty$. Therefore, the set $[0, \|\bar{h}\|_\infty \|f\|_\infty]$ (which is compact and convex) is stable by Υ. Then, the continuity of Υ guarantees the existence of a fixed-point, which proves (i).

To obtain (ii), we observe that the Lipschitz constant of Υ is $|\varepsilon|C$: if we take $|\varepsilon| < \varepsilon^* := C^{-1}$, then Υ is a contraction, and we can apply Banach's fixed-point theorem to conclude the proof. $\qquad\square$

4.2. Alternative proof for the existence using Schauder's fixed-point theorem. We include here an alternative proof for the existence of a stationary solution, which is interesting for two reasons: on the one hand, it does not rely on the Doeblin–Harris method, and on the other hand, it provides some estimates on the stationary solutions.

For any $(\tilde{u}, \tilde{x}) \in L_+^1(]\gamma(0), +\infty[) \times \mathbb{R}$, consider the transport equation

$$\partial_a \varrho(a, m) - \lambda \partial_m(m\varrho(a, m)) = -f(a, m, \tilde{x})\varrho(a, m),$$
$$\varrho(0, m) = \tilde{u}(m).$$

It has a unique weak solution $\rho_\infty^{\tilde{u}, \tilde{x}} \in \mathcal{C}(\mathbb{R}_+, L_+^1(\mathbb{R}_+^*)) \cap L^\infty(\mathbb{R}_+, L_+^1(\mathbb{R}_+^*))$ given by the method of characteristics, i.e., for all $(a, m) \in \mathbb{R}_+ \times \mathbb{R}_+^*$,

$$(4.6) \qquad \rho_\infty^{\tilde{u}, \tilde{x}}(a, m) = \tilde{u}(e^{\lambda a}m) \exp\left(\lambda a - \int_0^a f(s, e^{\lambda(a-s)}m, \tilde{x})ds\right).$$

We can now define the operator $\Phi := (\Phi_1, \Phi_2)$ on $L_+^1(]\gamma(0), +\infty[) \times \mathbb{R}$, where, for all $(\tilde{u}, \tilde{x}) \in L_+^1(]\gamma(0), +\infty[) \times \mathbb{R}$,

$$(4.7a) \quad \Phi_1(\tilde{u}, \tilde{x})(m) := \mathbb{1}_{m > \gamma(0)}\left|(\gamma^{-1})'(m)\right| \int_0^\infty f(a, \gamma^{-1}(m), \tilde{x})\rho_\infty^{(\tilde{u}, \tilde{x})}(a, \gamma^{-1}(m))da,$$

$$(4.7b) \qquad \Phi_2(\tilde{u}, \tilde{x}) := \int_0^\infty \int_0^\infty \bar{h}(a, m)f(a, m, \tilde{x})\rho_\infty^{\tilde{u}, \tilde{x}}(a, m)da\,dm.$$

A stationary solution (ρ_∞, x_∞) is a fixed-point of Φ, and vice versa. Therefore, we have the following a priori estimates.

LEMMA 4.5. *Grant Assumptions* 1 *and* 3. *There exists* $\theta \in]0,1[$ *such that for all* $(\tilde{u}, \tilde{x}) \in L^1_+(]\gamma(0), +\infty[) \times \mathbb{R}$, *we have the following:*

(i) $\left\| \Phi_1(\tilde{u}, \tilde{x}) \right\|_{L^1} = \|\tilde{u}\|_{L^1}$.

(ii) *For all* $m \in \mathbb{R}^*_+$, $|\Phi_1(\tilde{u}, \tilde{x})(m)| \leq \mathbb{1}_{m > \gamma(0)} \frac{\|f\|_\infty}{\lambda \gamma^{-1}(m)} \|\tilde{u}\|_{L^1}$.

(iii)

$$\int_0^\infty \Phi_1(\tilde{u}, \tilde{x})(m) m \, dm \leq \max\left(\int_0^\infty \tilde{u}(m) m \, dm, \frac{\gamma(0)}{1-\theta} \|\tilde{u}\|_{L^1} \right).$$

(iv) *For all* $\beta \in]0, \frac{\min(f)}{\lambda}[$,

$$\int_{\gamma(0)}^\infty \frac{\Phi_1(\tilde{u}, \tilde{x})(m)}{\gamma^{-1}(m)^\beta} dm \leq \frac{\|f\|_\infty}{\lambda \gamma(0)^\beta} \left(\frac{\min(f)}{\lambda} - \beta \right) \|\tilde{u}\|_{L^1}.$$

(v) $\Phi_2(\tilde{u}, \tilde{x}) \leq \left\| \bar{h} \right\|_\infty \|\tilde{u}\|_{L^1}$.

Proof. (i) By changes of variables on m,

$$\left\| \Phi_1(\tilde{u}, \tilde{x}) \right\|_{L^1} = \int_0^\infty \int_0^\infty f(a, m, \tilde{x}) \tilde{u}(e^{\lambda a} m) \exp\left(\lambda a - \int_0^a f(s, e^{\lambda(a-s)} m, \tilde{x}) ds \right) da \, dm$$

$$= \int_0^\infty \tilde{u}(m) \underbrace{\int_0^\infty f(a, e^{-\lambda a} m, \tilde{x}) \exp\left(-\int_0^a f(s, e^{-\lambda s} m, \tilde{x}) ds \right) da}_{=1 \quad \text{(by Assumption 3(i))}} dm.$$

(ii)

$$|\Phi_1(\tilde{u}, \tilde{x})(m)| \leq \mathbb{1}_{m > \gamma(0)} \|f\|_\infty \int_0^\infty \tilde{u}(e^{\lambda a} \gamma^{-1}(m)) \exp(\lambda a) da$$

$$= \mathbb{1}_{m > \gamma(0)} \frac{\|f\|_\infty}{\lambda \gamma^{-1}(m)} \int_0^\infty \tilde{u}(e^{\lambda a} \gamma^{-1}(m)) \gamma^{-1}(m) \lambda \exp(\lambda a) da$$

$$= \mathbb{1}_{m > \gamma(0)} \frac{\|f\|_\infty}{\lambda \gamma^{-1}(m)} \underbrace{\int_{\gamma^{-1}(m)}^\infty \tilde{u}(y) dy}_{\leq \|\tilde{u}\|_{L^1}},$$

where, for the last equality, we used the change of variable $y = e^{\lambda a} \gamma^{-1}(m)$.

(iii) Performing the same change of variable as for (i) and using the fact that $\gamma(m) \leq \gamma(0) + m$ for all $m \in \mathbb{R}_+$ (since $\gamma' \leq 1$), we have

$$\int_0^\infty \Phi_1(\tilde{u}, \tilde{x})(m) m \, dm$$

$$= \int_0^\infty \tilde{u}(m) \int_0^\infty \gamma(e^{-\lambda a} m) f(a, e^{-\lambda a} m, \tilde{x}) \exp\left(-\int_0^a f(s, e^{-\lambda s} m, \tilde{x}) ds \right) da \, dm$$

$$\leq \int_0^\infty \tilde{u}(m) m \underbrace{\int_0^\infty e^{-\lambda a} f(a, e^{-\lambda a} m, \tilde{x}) \exp\left(-\int_0^a f(s, e^{-\lambda s} m, \tilde{x}) ds \right) da}_{=: \vartheta(m)} dm + \gamma(0) \|\tilde{u}\|_{L^1}.$$

There exists $\theta \in]0,1[$ such that for all $m \in \mathbb{R}^*_+$, $\vartheta(m) < 1$, as we show in the

following. Fix $\epsilon > 0$.

$$\vartheta(m) \leq \int_0^\epsilon f(a, e^{-\lambda a}m, \tilde{x}) \exp\left(-\int_0^a f(s, e^{-\lambda s}m, \tilde{x})ds\right) da$$

$$+ \int_\epsilon^\infty e^{-\lambda\epsilon} f(a, e^{-\lambda a}m, \tilde{x}) \exp\left(-\int_0^a f(s, e^{-\lambda s}m, \tilde{x})ds\right) da$$

$$= 1 - (1 - e^{-\lambda\epsilon}) \int_\epsilon^\infty f(a, e^{-\lambda a}m, \tilde{x}) \exp\left(-\int_0^a f(s, e^{-\lambda s}m, \tilde{x})ds\right) da$$

$$= 1 - (1 - e^{-\lambda\epsilon}) \exp\left(-\int_0^\epsilon f(s, e^{-\lambda s}m, \tilde{x})ds\right)$$

$$\leq 1 - (1 - e^{-\lambda\epsilon}) \exp(-\|f\|_\infty \epsilon) =: \theta < 1.$$

Therefore,

$$\int_0^\infty \Phi_1(\tilde{u}, \tilde{x})(m)m\,dm \leq \theta \int_0^\infty \tilde{u}(m)m\,dm + \gamma(0)\|\tilde{u}\|_{L^1}.$$

To see that

$$\int_0^\infty \Phi_1(\tilde{u}, \tilde{x})(m)m\,dm \leq \max\left(\int_0^\infty \tilde{u}(m)m\,dm, \frac{\gamma(0)}{1-\theta}\|\tilde{u}\|_{L^1}\right),$$

we can distinguish three cases: if $\int_0^\infty \tilde{u}(m)m\,dm = \infty$, the inequality is trivial; if $\frac{\gamma(0)}{1-\theta}\|\tilde{u}\|_{L^1} \leq \int_0^\infty \tilde{u}(m)m\,dm < +\infty$, then

$$\int_0^\infty \Phi_1(\tilde{u}, \tilde{x})(m)m\,dm \leq \int_0^\infty \tilde{u}(m)m\,dm - (1-\theta)\int_0^\infty \tilde{u}(m)m\,dm + \gamma(0)\|\tilde{u}\|_{L^1}$$

$$\leq \int_0^\infty \tilde{u}(m)m\,dm;$$

and, finally, if $\int_0^\infty \tilde{u}(m)m\,dm < \frac{\gamma(0)}{1-\theta}\|\tilde{u}\|_{L^1}$, then

$$\int_0^\infty \Phi_1(\tilde{u}, \tilde{x})(m)m\,dm \leq \theta\frac{\gamma(0)}{1-\theta}\|\tilde{u}\|_{L^1} + \gamma(0)\|\tilde{u}\|_{L^1} = \frac{\gamma(0)}{1-\theta}\|\tilde{u}\|_{L^1}.$$

(iv)

$$\int_{\gamma(0)}^\infty \frac{\Phi_1(\tilde{u}, \tilde{x})(m)}{\gamma^{-1}(m)^\beta}dm$$

$$= \int_0^\infty \int_0^\infty \frac{1}{m}f(a, m, \tilde{x})\tilde{u}(e^{\lambda a}m)\exp\left(\lambda a - \int_0^a f(s, e^{\lambda(a-s)}m, \tilde{x})ds\right)da\,dm$$

$$\leq \|f\|_\infty \int_0^\infty \int_0^\infty \frac{1}{m^\beta}\tilde{u}(e^{\lambda a}m)\exp\left(\lambda a - \min(f)a\right)da\,dm.$$

Making the change of variable $y = e^{\lambda a}m$, we get

$$= \|f\|_\infty \int_0^\infty \int_m^\infty \frac{1}{\lambda m^{1+\beta}}\tilde{u}(y)\exp\left(-\min(f)\frac{1}{\lambda}\ln\left(\frac{y}{m}\right)\right)dy\,dm$$

$$= \frac{\|f\|_\infty}{\lambda}\int_0^\infty \int_m^\infty m^{\min(f)/\lambda - 1 - \beta}\tilde{u}(y)y^{-\min(f)/\lambda}dy\,dm,$$

and, using Fubini's theorem and the fact that $\min(f)/\lambda - \beta > 0$, we get

$$= \frac{\|f\|_\infty}{\lambda} \int_0^\infty \tilde{u}(y) y^{-\min(f)/\lambda} \underbrace{\int_0^y m^{\min(f)/\lambda - 1 - \beta} dm}_{= \frac{y^{\min(f)/\lambda - \beta}}{\min(f)/\lambda - \beta}} dy$$

$$= \frac{\|f\|_\infty}{\lambda} \left(\frac{\min(f)}{\lambda} - \beta \right) \int_0^\infty \tilde{u}(y) y^{-\beta} dy.$$

Finally, it is easy to check that $\int_0^\infty \tilde{u}(y) y^{-\beta} dy \leq \gamma(0)^{-\beta} \|\tilde{u}\|_{L^1}$.

(v) Use (4.7b) and see the proof of (i). $\qquad\square$

By these estimates, we see that there exists $\beta, C_1, \ldots, C_4 > 0$ such that the set $\mathscr{C} \times B \subset L^1(]\gamma(0), +\infty[) \times \mathbb{R}$, where

$$\mathscr{C} := \left\{ u \in L^1_+(]\gamma(0), +\infty[) \,\middle|\, \|u\|_{L^1} \leq 1; \right.$$

$$\left. u \leq \frac{C_1}{\gamma^{-1}(\cdot)} \text{ a.e.}; \int_0^\infty u(m) m\, dm \leq C_2; \int_{\gamma(0)}^\infty \frac{u(m)}{\gamma^{-1}(m)^\beta} dm \leq C_3 \right\},$$

and $B := [-C_4, +C_4]$ is stable by the operator Φ.

In order to apply Schauder's fixed-point theorem, we will need the next lemma.

LEMMA 4.6. *Grant Assumptions* 1 *and* 3. $\mathscr{C}$ *is convex, closed, and compact for the weak topology* $\sigma(L^1, L^\infty)$.

Proof. It is easy to verify that $\mathscr{C}$ is convex. Since $\mathscr{C}$ is convex, it suffices to show that it is strongly closed to show that it is weakly closed. Let u_n be a sequence of elements of $\mathscr{C}$ which converge strongly to $u \in L^1(]\gamma(0), +\infty[)$. By the strong convergence, $\|u\|_{L^1} \leq 1$. We can extract a subsequence u_{n_k} such that u_{n_k} converges to u a.e. Taking the pointwise limit, we have that $u \leq \frac{C_1}{\gamma^{-1}(\cdot)}$ a.e. Furthermore, by Fatou lemma,

$$\int_{\gamma(0)}^\infty u(m) m\, dm \leq \liminf_{k \to +\infty} \int_{\gamma(0)}^\infty u_{n_k}(m) m\, dm \leq C_2$$

and

$$\int_{\gamma(0)}^\infty \frac{u(m)}{\gamma^{-1}(m)^\beta} dm \leq \liminf_{k \to +\infty} \int_{\gamma(0)}^\infty \frac{u_{n_k}(m)}{\gamma^{-1}(m)^\beta} dm \leq C_3.$$

Hence, $\mathscr{C}$ is strongly closed.

To show that $\mathscr{C}$ is weakly compact, we will show the following:

(a) $\sup_{u \in \mathscr{C}} \|u\|_{L^1} < \infty$.

(b) For all $\epsilon > 0$, there exists $R > 0$ such that $\int_R^\infty u(m) dm < \epsilon$ for all $u \in \mathscr{C}$.

(c) $\mathscr{C}$ is equi-integrable, i.e., for all $\epsilon > 0$, there exists $\delta > 0$ such that for all Borel set $A \subset \mathbb{R}_+$ with $|A| \leq \delta$ and for all $u \in \mathscr{C}$, $\int_A u(m) dm \leq \epsilon$.

Then use the Dunford–Pettis theorem. (a) is clearly verified. (b) is also verified since for all $R > 0$, $\int_R^\infty u(m) dm \leq \frac{1}{R} \int_0^\infty u(m) m\, dm \leq \frac{C_2}{R}$. To show (c), let us first observe that for all $\delta_1 > 0$,

$$\int_{\gamma(0)}^{\gamma(0)+\delta_1} u(m) dm \leq \gamma^{-1}(\gamma(0) + \delta_1)^\beta \int_{\gamma(0)}^\infty \frac{u(m)}{\gamma^{-1}(m)^\beta} dm \leq \gamma^{-1}(\gamma(0) + \delta_1)^\beta C_3.$$

For any $\epsilon > 0$, let us choose $\delta_1 > 0$ such that $\gamma^{-1}(\gamma(0) + \delta_1)^\beta C_3 \leq \frac{\epsilon}{2}$. Then, for all Borel set $A \subset \mathbb{R}_+$ with $|A| \leq \delta$,

$$\int_A u(m)dm \leq \int_{\gamma(0)}^{\gamma(0)+\delta_1} u(m)dm + \int_{A \setminus [0, \gamma(0)+\delta_1]} u(m)dm \leq \frac{\epsilon}{2} + \delta \frac{C_1}{\gamma^{-1}(\gamma(0) + \delta_1)}.$$

Hence, we can choose $\delta = \min(\delta_1, \frac{\epsilon \gamma^{-1}(\gamma(0)+\delta_1)}{2C_1})$, and (c) is verified. By the Dunford–Pettis theorem, $\mathscr{C}$ is weakly relatively compact. Finally, since $\mathscr{C}$ is weakly closed, $\mathscr{C}$ is weakly compact. $\qquad \square$

We can now give an alternative proof of the existence of stationary solutions to (1.1) for arbitrary connectivity strength ε.

Proof of Theorem 1.3(i). We verify that the operator Φ is weakly continuous: for any sequence $(u_n, x_n) \to (u, x)$ in $\mathscr{C} \times \mathbb{R}$ and for any $\varphi \in L^\infty(\mathbb{R}_+)$,

$$\left| \int (\Phi_1(u_n, x_n) - \Phi_1(u, x))\varphi(m)dm \right| \leq Q_1^n + Q_2^n + Q_3^n,$$

where

$$Q_1^n := \left| \int_0^\infty \int_0^\infty (u_n(e^{\lambda a}m) - u(e^{\lambda a}m))\varphi(\gamma(m))e^{\lambda a}f(a, m, x)e^{-\int_0^a f(\tau, e^{\lambda(a-\tau)}m, x)d\tau} da\,dm \right|,$$

$$Q_2^n := \|\varphi\|_\infty \int_0^\infty \int_0^\infty u_n(e^{\lambda a}m)e^{\lambda a}|f(a, m, x) - f(a, m, x_n)|e^{-\int_0^a f(\tau, e^{\lambda(a-\tau)}m, x)d\tau} da\,dm,$$

$$Q_3^n := \|\varphi\|_\infty \int_0^\infty \int_0^\infty u_n(e^{\lambda a}m)e^{\lambda a}f(a, m, x_n) \left| e^{-\int_0^a f(\tau, e^{\lambda(a-\tau)}m, x)d\tau} \right.$$
$$\left. -e^{-\int_0^a f(\tau, e^{\lambda(a-\tau)}m, x_n)d\tau} \right| da\,dm.$$

Making the change of variable $y\,dy = e^{\lambda a}m\,dm$ in Q_1, we get

$$Q_1^n = \left| \int_0^\infty (u_n(y) - u(y)) \int_0^\infty \varphi(\gamma(ye^{-\lambda a}))f(a, ye^{-\lambda a}, x)e^{-\int_0^a f(\tau, e^{-\lambda \tau}y, x)} da\,dm \right|.$$

Since u_n converges to u in $\sigma(L^1, L^\infty)$ and

$$\int_0^\infty \varphi(\gamma(ye^{-\lambda a}))f(a, ye^{-\lambda a}, x)e^{-\int_0^a f(\tau, e^{-\lambda \tau}y, x)d\tau} da$$
$$\leq \|\varphi\|_\infty \int_0^\infty f(a, ye^{-\lambda a}, x)e^{-\int_0^a f(\tau, e^{-\lambda \tau}y, x)d\tau} da = \|\varphi\|_\infty,$$

Q_1^n converges to 0. On the other hand, since f is bounded and Lipschitz, $Q_2^n, Q_3^n \leq \|u_n\|_{L^1}C|x_n - x| \leq C|x_n - x|$. Therefore, Φ_1 is a continuous operator with respect to the weak topology $\sigma(L^1, L^\infty)$.

The continuity of Φ_2 is shown analogously, taking $\varphi = h$ (h is bounded).

Since $\mathscr{C}$ is stable by Φ, convex, and weakly compact (Lemma 4.6), we can apply Schauder's fixed-point theorem to obtain the existence of a fixed-point, which gives the existence of a stationary solution. $\qquad \square$

COROLLARY 4.7. *Grant Assumptions* 1 *and* 3. *If f is of class $\mathcal{C}^k$, then $u(m)$ is a function of class $\mathcal{C}^k$ for all $m > \gamma(0)$. Consequently, the stationary solutions of (1.1) are of class $\mathcal{C}^k$.*

Proof. If $(u, \tilde{x})$ is a fixed-point of Φ, then

$$(4.8) \quad u(m) = \mathbb{1}_{m > \gamma(0)} \left| (\gamma^{-1})'(m) \right| \int_0^\infty f(a, \gamma^{-1}(m), \tilde{x}) u(e^{\lambda a} \gamma^{-1}(m))$$
$$\times \exp \left(\lambda a - \int_0^a f(s, e^{\lambda(a-s)} \gamma^{-1}(m), \tilde{x}) ds \right) da.$$

Making the change of variable $y = e^{\lambda a} \gamma^{-1}(m)$ in a, as in estimate (ii) of Lemma 4.5, we obtain

$$(4.9) \quad u(m) = \mathbb{1}_{m > \gamma(0)} \frac{\left| (\gamma^{-1})'(m) \right|}{\lambda \gamma^{-1}(m)} \int_{\gamma^{-1}(m)}^\infty f(g(y,m), y, \tilde{x}) u(y)$$
$$\times \exp \left(- \int_0^{g(y,m)} f(s, e^s y, \tilde{x}) ds \right) dy,$$

where $g(y, m) = \ln \frac{y}{\lambda(\gamma^{-1}(m))}$. We conclude with a bootstrap argument: if u is L^1, then the right-hand side of (4.9) is a continuous function of m, meaning that u is continuous. But if u is continuous, then the right-hand side is of class $\mathcal{C}^1$, etc. $\qquad\square$

4.3. Formula in the case of short-term synaptic depression. In general, there is no explicit formula for the invariant probability measure solving (3.14). However, in the case of short-term synaptic depression (1.3), we can derive an explicit expression for the total postsynaptic potential

$$X(\tilde{x}) := \int_0^\infty \hat{h}(t) \int_0^1 \int_0^\infty (1 - m) f(a, \tilde{x}) \rho_\infty^{\tilde{x}}(a, m) da\, dm\, dt$$

for any $\tilde{x} \in \mathbb{R}$. This fact has been reported in the theoretical neuroscience literature [42]; we provide here a rigorous and analytic justification for it.

For all $\tilde{x} \in \mathbb{R}$, let us introduce the quantities

$$I^{\tilde{x}} := \int_0^\infty a f(a, \tilde{x}) \exp \left(- \int_0^a f(s, \tilde{x}) ds \right) da = \int_0^\infty \exp \left(- \int_0^a f(s, \tilde{x}) ds \right) da,$$
$$P^{\tilde{x}}(\lambda) := \int_0^\infty e^{-\lambda a} f(a, \tilde{x}) \exp \left(- \int_0^a f(s, \tilde{x}) ds \right) da.$$

The value $I^{\tilde{x}}$ can be interpreted as the mean interspike interval of a neuron receiving a constant input $\tilde{x}$. The value $P^{\tilde{x}}(\lambda)$ can be seen as the Laplace transform of the interspike interval distribution of that neuron evaluated in λ.

PROPOSITION 4.8. *Grant Assumptions 1 and 3. For all $\tilde{x} \in \mathbb{R}$,*

$$X(\tilde{x}) = \int_0^\infty \hat{h}(t) dt \, \frac{1}{I^{\tilde{x}}} \left\{ \frac{1 - P^{\tilde{x}}(\lambda)}{1 - \upsilon P^{\tilde{x}}(\lambda)} \right\}.$$

Proof. Using the method of characteristics (i.e., combining (4.6) and (3.14b)), we

have

$$1 = \int_0^1 \int_0^\infty \rho_\infty^{\tilde{x}}(a,m) da dm$$

$$= \int_0^1 \int_0^\infty \mathbb{1}_{e^{\lambda a} m < 1} \rho_\infty^{\tilde{x}}(0, e^{\lambda a} m) \exp\left(\lambda a - \int_0^a f(s, \tilde{x}) ds\right) da dm$$

$$= \int_0^1 \int_0^\infty \rho_\infty^{\tilde{x}}(0, m) \exp\left(-\int_0^a f(s, \tilde{x}) ds\right) da dm = I^{\tilde{x}} \int_0^1 \rho_\infty^{\tilde{x}}(0, m) dm.$$

Therefore,

$$\int_0^1 \int_0^\infty f(a, \tilde{x}) \rho_\infty^{\tilde{x}}(a, m) da dm = \int_0^1 \rho_\infty^{\tilde{x}}(0, m) dm = \frac{1}{I^{\tilde{x}}}.$$

On the other hand,

$$\int_0^1 \int_0^\infty m f(a, \tilde{x}) \rho_\infty^{\tilde{x}}(a, m) da dm$$

$$= \int_0^1 \int_0^\infty \mathbb{1}_{e^{\lambda a} m < 1} m f(a, \tilde{x}) \rho_\infty^{\tilde{x}}(0, e^{\lambda a} m) \exp\left(\lambda a - \int_0^a f(s, \tilde{x}) ds\right) da dm$$

$$= \int_0^1 \int_0^\infty e^{-\lambda a} m f(a, \tilde{x}) \rho_\infty^{\tilde{x}}(0, m) \exp\left(-\int_0^a f(s, \tilde{x}) ds\right) da dm$$

$$= P^{\tilde{x}}(\lambda) \int_0^1 m \rho_\infty^{\tilde{x}}(0, m) dm$$

and

$$\int_0^1 m \rho_\infty^{\tilde{x}}(0, m) dm = \int_0^1 m \mathbb{1}_{m > 1 - \upsilon} \frac{1}{\upsilon} \int_0^\infty f(a, \tilde{x}) \rho_\infty^{\tilde{x}}\left(a, 1 - \frac{1-m}{\upsilon}\right) da dm$$

$$= \int_0^1 (1 - \upsilon + \upsilon m) \int_0^\infty f(a, \tilde{x}) \rho_\infty^{\tilde{x}}(a, m) da dm$$

$$= \frac{1-\upsilon}{I^{\tilde{x}}} + \upsilon P^{\tilde{x}}(\lambda) \int_0^1 m \rho_\infty^{\tilde{x}}(0, m) dm.$$

Therefore,

$$\int_0^1 m \rho_\infty^{\tilde{x}}(0, m) dm = \frac{1-\upsilon}{I^{\tilde{x}}(1 - \upsilon P^{\tilde{x}}(\lambda))}$$

and

$$\int_0^1 \int_0^\infty m f(a, \tilde{x}) \rho_\infty^{\tilde{x}}(a, m) da dm = \frac{P^{\tilde{x}}(\lambda)(1-\upsilon)}{I^{\tilde{x}}(1 - \upsilon P^{\tilde{x}}(\lambda))}.$$

Finally, we have

$$X(\tilde{x}) = \int_0^\infty \hat{h}(t) dt \left\{ \int_0^1 \int_0^\infty f(a, \tilde{x}) \rho_\infty^{\tilde{x}}(a, m) da dm - \int_0^1 \int_0^\infty m f(a, \tilde{x}) \rho_\infty^{\tilde{x}}(a, m) da dm \right\}$$

$$= \int_0^\infty \hat{h}(t) dt \frac{1}{I^{\tilde{x}}} \left\{ 1 - \frac{P^{\tilde{x}}(\lambda)(1-\upsilon)}{(1 - \upsilon P^{\tilde{x}}(\lambda))} \right\} = \int_0^\infty \hat{h}(t) dt \frac{1}{I^{\tilde{x}}} \left\{ \frac{1 - P^{\tilde{x}}(\lambda)}{1 - \upsilon P^{\tilde{x}}(\lambda)} \right\}. \qquad \square$$

5. Exponential stability in the weak connectivity regime. To study the long time behavior (1.1) in the weak connectivity regime, we perturb the noninteracting case (3.1), taking $\tilde{x} = \varepsilon x_\infty$, where x_∞ is given by the unique stationary solution to (1.1) when $\varepsilon \in\,]-\varepsilon^*, +\varepsilon^*[$ (ε^* is taken from Theorem 1.3(ii)). In this section, we keep the small ε fixed, and we work under Assumptions 1–3 and 5. We roughly follow the same line of argument as in [30, sect. 5].

For convenience, we first rewrite (1.1) in a more formal and compact form,

$$(5.1a) \qquad \partial_t \rho_t = -\partial_a \rho_t + \lambda \partial_m (m\rho_t) - f(\varepsilon x_t)\rho_t + \delta_0^a (\gamma \circ \Pi)_* \left(f(\varepsilon x_t)\rho_t \right),$$

$$(5.1b) \qquad x_t = \int_0^t \int h(t-s) f(\varepsilon x_s)\rho_s \, da dm \, ds,$$

$$(5.1c) \qquad \rho_0 = u_0,$$

where δ_0^a indicates that (singular) mass enters in $a = 0$,[2] $\Pi : (a, m) \mapsto m$ is the projection on m, and $_*$ denotes the pushforward measure. To write (5.1) as an evolution equation, we introduce an auxiliary transport equation on $\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+^*$,

$$\partial_t \zeta_t = -\partial_s \zeta_t + \delta_0^s f(\varepsilon x_t)\rho_t,$$
$$\zeta_0 = 0,$$

which solution is given by the method of characteristics,

$$\zeta_t(s) = \mathbb{1}_{s \le t} f(\varepsilon x_{t-s})\rho_{t-s} \qquad \forall (t, s) \in \mathbb{R}_+^* \times \mathbb{R}_+.$$

Using the auxiliary equation, (5.1) is equivalent to

$$(5.2a)$$
$$\partial_t(\rho_t, \zeta_t)$$
$$= \left(-\partial_a \rho_t + \lambda \partial_m (m\rho_t) - f(\varepsilon x_t)\rho_t + \delta_0^a (\gamma \circ \Pi)_* \left(f(\varepsilon x_t)\rho_t \right), -\partial_s \zeta_t + \delta_0^s f(\varepsilon x_t)\rho_t \right),$$

$$(5.2b)$$
$$(\rho_0, \zeta_0) = (u_0, 0),$$

where $x_t := \int_0^\infty \int h(s)\zeta_t(s) \, da dm \, ds$.

By Theorem 1.3, for all $\varepsilon \in\,]-\varepsilon^*, +\varepsilon^*[$, there exists a unique stationary solution (ρ_∞, x_∞), and we have

$$(5.3) \qquad -\partial_a \rho_\infty + \lambda \partial_m (m\rho_\infty) - f(\varepsilon x_\infty)\rho_\infty + \delta_0^a (\gamma \circ \Pi)_* \left(f(\varepsilon x_\infty)\rho_\infty \right) = 0.$$

Now, we write (5.2) as the sum of a linear equation and a perturbation,

$$(5.4a) \qquad \partial_t(\rho_t, \zeta_t) = \Lambda(\rho_t, \zeta_t) + (Z_t^{(1)}, Z_t^{(2)}),$$
$$(5.4b) \qquad (\rho_0, \zeta_0) = (u_0, 0),$$

---

[2]δ_0^a should not be confused with the Dirac distribution $\delta_{0=a}$. Using $\delta_{0=a}$, by integration by parts of weak solutions, (5.1a) should read

$$\partial_t \rho_t = -\partial_a \rho_t + \lambda \partial_m (m\rho_t) - f(\varepsilon x_t)\rho_t + \delta_{0=a} \left\{ (\gamma \circ \Pi)_* \left(f(\varepsilon x_t)\rho_t \right) - \rho_t(0, \cdot) \right\}.$$

where

$$\Lambda(\rho_t, \zeta_t)$$
$$:= \left(-\partial_a \rho_t + \lambda \partial_m (m\rho_t) - f(\varepsilon x_\infty)\rho_t + \delta_0^a (\gamma \circ \Pi)_* \left(f(\varepsilon x_\infty)\rho_t\right), -\partial_s \zeta_t + \delta_0^s f(\varepsilon x_\infty)\rho_t\right),$$
$$Z_t^{(1)} := [f(\varepsilon x_\infty) - f(\varepsilon x_t)]\rho_t + \delta_0^a (\gamma \circ \Pi)_* ([f(\varepsilon x_t) - f(\varepsilon x_\infty)]\rho_t),$$
$$Z_t^{(2)} := \delta_0^s [f(\varepsilon x_t) - f(\varepsilon x_\infty)]\rho_t.$$

Let us put $\zeta_\infty(s) := f(\varepsilon x_\infty)\rho_\infty$ for all $s \in \mathbb{R}_+$. Then, using (5.3), by the linearity of the operator Λ and writing $\bar{\rho}_t := \rho_t - \rho_\infty$ and $\bar{\zeta}_t := \zeta_t - \zeta_\infty$, we get

(5.5a) $$\partial_t(\bar{\rho}_t, \bar{\zeta}_t) = \Lambda(\bar{\rho}_t, \bar{\zeta}_t) + (Z_t^{(1)}, Z_t^{(2)}),$$

(5.5b) $$(\bar{\rho}_0, \bar{\zeta}_0) = (u_0 - \rho_\infty, -\zeta_\infty).$$

Writing $(S_t^\Lambda)_{t \in \mathbb{R}_+}$ the semigroup associated with the operator Λ, we have, by Duhamel's formula,

(5.6) $$(\bar{\rho}_t, \bar{\zeta}_t) = S_t^\Lambda(\bar{\rho}_0, \bar{\zeta}_0) + \int_0^t S_{t-s}^\Lambda(Z_s^{(1)}, Z_s^{(2)})ds \qquad \forall t \geq 0.$$

Let us define the weighted space

$$L_+^1(\mu) := \left\{ \zeta \in L^1(\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+^*, \mathbb{R}_+) \, \Big| \, \int_0^\infty \left\| \zeta(s) \right\|_{L^1} \|h\|_\infty \, e^{-\mathfrak{h}s} ds < \infty \right\}.$$

Note that, for all $t \geq 0$,

$$|x_t - x_\infty| = \left| \int_0^\infty \int h(s)\zeta_t(s) \, dadm \, ds - \int_0^\infty \int h(s)\zeta_\infty(s) \, dadm \, ds \right|$$
$$\leq \int_0^\infty \|h\|_\infty \, e^{-\mathfrak{h}s} \left\| \zeta_t(s) - \zeta_\infty(s) \right\|_{L^1} ds = \left\| \bar{\zeta}_t \right\|_{L^1(\mu)}.$$

Also, we have, for all $t \geq 0$,

(5.7a) $$\left\| Z_t^{(1)} \right\|_{L^1} \leq |\varepsilon| 2L_f \|\rho_t\|_{L^1} |x_t - x_\infty| \leq |\varepsilon| 2L_f \left\| \bar{\zeta}_t \right\|_{L^1(\mu)},$$

(5.7b) $$\left\| Z_t^{(1)} \right\|_{L^1(w)} \leq |\varepsilon| 2L_f \|\rho_t\|_{L^1(w)} |x_t - x_\infty| \leq |\varepsilon| 2L_f \left(\|u_0\|_{L^1(w) + \frac{b}{\alpha}} \right) \left\| \bar{\zeta}_t \right\|_{L^1(\mu)},$$

(5.7c) $$\left\| Z_t^{(2)} \right\|_{L^1(\mu)} \leq |\varepsilon| \|h\|_\infty L_f \|\rho_t\|_{L^1} |x_t - x_\infty| \leq |\varepsilon| \|h\|_\infty L_f \left\| \bar{\zeta}_t \right\|_{L^1(\mu)},$$

where we have used Theorem 1.2(ii) in the first line and Lemma 2.1 in the second.

LEMMA 5.1. *Grant 1–3 and 5 and take $(\bar{\rho}_0, \bar{\zeta}_0)$ as in (5.5). There exist $K_1 \geq 1$ and $\mathfrak{a}_1 > 0$ such that, for all initial data $u_0 \in L_+^1(w)$ with $\|u_0\|_{L^1} = 1$,*

(5.8) $$\left\| S_t^\Lambda(\bar{\rho}_0, \bar{\zeta}_0) \right\|_{L^1(w) \times L^1(\mu)} \leq K_1 e^{-\mathfrak{a}_1 t} \left\| (\bar{\rho}_0, \bar{\zeta}_0) \right\|_{L^1(w) \times L^1(\mu)} \qquad \forall t \geq 0.$$

If in addition we grant Assumption 4, then there exist $K_2 \geq 1$ and $\mathfrak{a}_2 > 0$ such that, for all initial data $u_0 \in L_+^1$ with $\|u_0\|_{L^1} = 1$,

(5.9) $$\left\| S_t^\Lambda(\bar{\rho}_0, \bar{\zeta}_0) \right\|_{L^1 \times L^1(\mu)} \leq K_2 e^{-\mathfrak{a}_2 t} \left\| (\bar{\rho}_0, \bar{\zeta}_0) \right\|_{L^1 \times L^1(\mu)} \qquad \forall t \geq 0.$$

Proof. We write $(S_t^\Lambda(\bar\rho_0, \bar\zeta_0)^{(1)}, S_t^\Lambda(\bar\rho_0, \bar\zeta_0)^{(2)}) := S_t^\Lambda(\bar\rho_0, \bar\zeta_0)$ the first and second components of $S_t^\Lambda(\bar\rho_0, \bar\zeta_0)$.

By Theorem 3.2, there exist $K \geq 0$ and $\mathfrak{a} > 0$ such that

$$\left\| S_t^\Lambda(\bar\rho_0, \bar\zeta_0)^{(1)} \right\|_{L^1(w)} \leq K e^{-\mathfrak{a} t} \|\bar\rho_0\|_{L^1(w)} \qquad \forall t \geq 0.$$

Then,

$$\left\| S_t^\Lambda(\bar\rho_0, \bar\zeta_0)^{(2)} \right\|_{L^1(\mu)} = \int_0^t \left\| f(\varepsilon x_\infty) S_{t-s}^\Lambda(\bar\rho_0, \bar\zeta_0)^{(1)} \right\|_{L^1} C_h e^{-\mathfrak{h} s} ds$$
$$+ \int_t^\infty \left\| \bar\zeta_0(s) \right\|_{L^1} C_h e^{-\mathfrak{h} s} ds$$
$$\leq C_h \left\{ \|f\|_\infty K \int_0^t e^{-\mathfrak{a}(t-s)} e^{-\mathfrak{h} s} ds \|\bar\rho_0\|_{L^1(w)} + e^{-\mathfrak{h} t} \|\bar\zeta_0\|_{L^1(\mu)} \right\}.$$

Gathering the bounds on the two components and observing that the function $t \mapsto \int_0^t e^{-\mathfrak{a}(t-s)} e^{-\mathfrak{h} s} ds$ decays exponentially, we conclude that there exist $K_1 \geq 1$ and $\mathfrak{a}_1 > 0$ such that (5.8) holds.

For (5.9), we use Theorem 3.1 and follow the same argument. $\quad\square$

We can now prove our main result.

Proof of Theorem 1.4. By Duhamel's formula (5.6), (5.8) in Lemma 5.1, and the bounds (5.7), for all $t \geq 0$,

$$\left\| (\bar\rho_t, \bar\zeta_t) \right\|_{L^1(w) \times L^1(\mu)}$$
$$\leq \left\| S_t^\Lambda(\bar\rho_0, \bar\zeta_0) \right\|_{L^1(w) \times L^1(\mu)} + \int_0^t \left\| S_{t-s}^\Lambda(Z_s^{(1)}, Z_s^{(2)}) \right\|_{L^1(w) \times L^1(\mu)} ds$$
$$\leq K_1 e^{-\mathfrak{a}_1 t} \left\| (\bar\rho_0, \bar\zeta_0) \right\|_{L^1(w) \times L^1(\mu)} + K_1 \int_0^t e^{-\mathfrak{a}_1(t-s)} \left\| (Z_s^{(1)}, Z_s^{(2)}) \right\|_{L^1(w) \times L^1(\mu)} ds$$
$$\leq K_1 e^{-\mathfrak{a}_1 t} \left\| (\bar\rho_0, \bar\zeta_0) \right\|_{L^1(w) \times L^1(\mu)} + |\varepsilon| \tilde{C}_W \int_0^t e^{-\mathfrak{a}_1(t-s)} \left\| (\bar\rho_s, \bar\zeta_s) \right\|_{L^1(w) \times L^1(\mu)} ds$$
$$=: \mathcal{Q}(t),$$

where $\tilde{C}_K$ is a constant depending on W. We have, for all $t \geq 0$,

$$\frac{d}{dt} \mathcal{Q}(t) = -\mathfrak{a}_1 \mathcal{Q}(t) + |\varepsilon| \tilde{C}_W \left\| (\bar\rho_t, \bar\zeta_t) \right\|_{L^1(w) \times L^1(\mu)}$$
$$\leq \left(-\mathfrak{a}_1 + |\varepsilon| \tilde{C}_W \right) \mathcal{Q}(t),$$

whence, by Grönwall's lemma,

$$\forall t \geq 0, \qquad \mathcal{Q}(t) \leq K_1 \left\| (\bar\rho_0, \bar\zeta_0) \right\|_{L^1(w) \times L^1(\mu)} \exp\left(\left(-\mathfrak{a}_1 + |\varepsilon| \tilde{C}_W \right) t \right).$$

For all $t \geq 0$, we have

$$\|\rho_t - \rho_\infty\|_{L^1(w)} + |x_t - x_\infty| \leq \left\| (\bar\rho_t, \bar\zeta_t) \right\|_{L^1(w) \times L^1(\mu)} \leq \mathcal{Q}(t)$$

and

$$\left\|\bar{\zeta}_0\right\|_{L^1(\mu)} \leq \int_0^\infty \left\|f(\varepsilon x_\infty)\rho_\infty\right\|_{L^1} C_h e^{-\mathfrak{h}s} ds \leq \frac{\|f\|_\infty C_h}{\mathfrak{h}}.$$

Therefore, choosing $\varepsilon_W^{**} := \frac{\mathfrak{a}_1}{\bar{C}_W} \wedge \varepsilon^*$, we easily see that there exist $C \geq 1$ and $c_W > 0$ such that (1.6) holds.

For (1.7), we use (5.9) instead of (5.8) and follow the same argument. □

Appendix A. Time-elapsed neuron network model. Here, we compare simulations of (1.2) with simulations of the time-elapsed neuron network model [35].

If the firing rate function f does not depend on m and if we put

$$(A.1) \qquad\qquad f(a, \varepsilon x_t) := \hat{f}(\eta(a) + \varepsilon x_t),$$

then (1.2) reduces to the time-elapsed neuron network model

$$(A.2a) \qquad\quad \partial_t \rho_t(a) + \partial_a \rho_t(a) = -f(a, \varepsilon x_t)\rho_t(a),$$

$$(A.2b) \qquad\quad \rho_t(0) = \int_0^\infty f(a, \varepsilon x_t)\rho_t(a)da,$$

$$(A.2c) \qquad\quad x_t = \int_0^t h(t-s) \int_0^\infty \int_0^\infty f(a, \varepsilon x_s)\rho_s(a)dads,$$

$$(A.2d) \qquad\quad \rho_0(a) = u_0(a).$$

Equation (A.2) is the population equation for nonadaptive SRM$_0$ neurons (or age-dependent nonlinear Hawkes processes) [6]. As reported previously, (A.2) exhibits self-sustained oscillations for large ε or relaxation to a stationary state for small ε (see Figure 2). Note that in the special case where h is a Dirac delta distribution ("instantaneous transmission"), (A.2) can exhibit elaborate periodic patterns [36], but these patterns do not fulfill our definition of self-sustained population bursts.

Appendix B. Proof of Doeblin's theorem (Theorem 3.1). We follow [4]. We first observe that for any $t \geq 0$, S_t is nonexpansive, i.e.,

$$(B.1) \qquad\qquad \|S_t\mu\|_{L^1} \leq \|\mu\|_{L^1} \quad \forall \mu \in L^1.$$

Indeed, writing μ_+ and μ_- the positive and negative parts of μ, respectively ($\mu = \mu_+ - \mu_-$ and $|\mu| = \mu_+ + \mu_-$),

$$|S_t\mu| \leq |S_t\mu_+| + |S_t\mu_-| = S_t\mu_+ + S_t\mu_- = S_t|\mu|,$$

and we get (B.1) by integrating both sides. The result can then be shown in two steps.

Step 1. The Doeblin minorization condition (3.2) implies

$$(B.2) \qquad\qquad \|S_T\mu\|_{L^1} \leq (1 - \|\nu\|_{L^1})\|\mu\|_{L^1} \quad \forall \mu \in L^1 : \int \mu = 0.$$

Indeed, the Doeblin minoration condition (3.2) and the fact that $\|\mu_+\|_{L^1} = \|\mu_-\|_{L^1} = \|\mu\|_{L^1}/2$ imply

$$|S_T\mu| \leq |S_T\mu_+ - \|\mu_+\|_{L^1}\nu| + |S_T\mu_- - \|\mu_-\|_{L^1}\nu| = S_T|\mu| - \|\mu\|_{L^1}\nu,$$
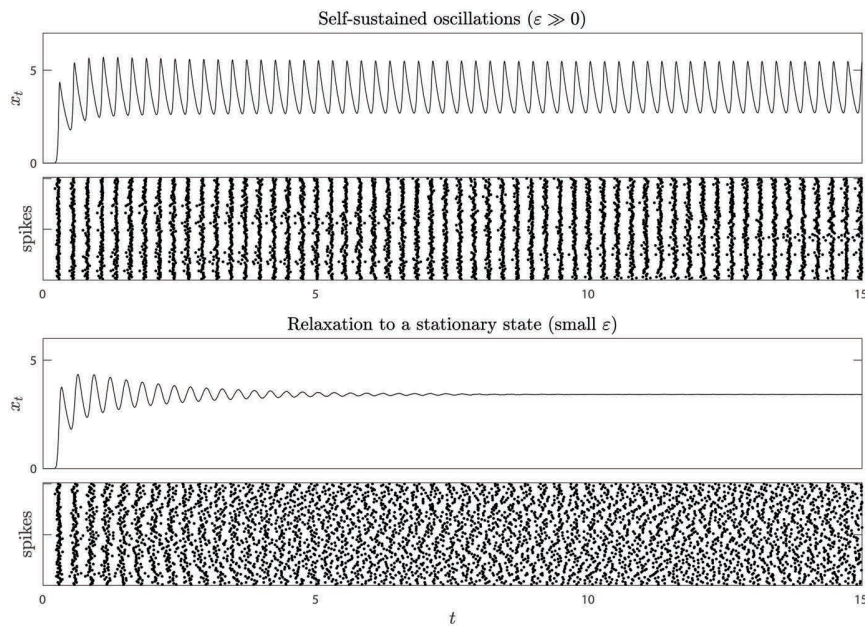
FIG. 2. *Same as Figure 1 but for the time-elapsed neuron network model* (A.2). *Simulations of a network of* $5 \cdot 10^5$ *nonadaptive* SRM_0 *neurons, approximating* (A.2), *with identical parameters (except for* ε) *and identical initial conditions. Neuronal parameters are the same as in Figure* 1, *except that* f *is replaced by* (A.1). *The* ε *have also been adapted.*

and we get (B.2) by integrating both sides.

Then, by the estimate (B.2), there exists a unique $\rho_\infty \in L^1_+$ with $\|\rho_\infty\|_{L^1} = 1$ such that $S_T \rho_\infty = \rho_\infty$. The existence is obtained by taking a $\mu_0 \in L^1$ with $\|\mu_0\|_{L^1} = 1$ and defining $\mu_k = S_T \mu_{k-1}$ for all $k \geq 1$. The estimate (B.2) implies that $\{\mu_k\}_{k\geq 1}$ is a Cauchy sequence, and passing to the limit, we get that $\rho_\infty := \lim_{k\to\infty} \mu_k$ satisfies $S_T \rho_\infty = \rho_\infty$. If $\varrho_\infty \in L^1_+$ with $\|\varrho_\infty\|_{L^1} = 1$ also satisfies $S_T \varrho_\infty = \varrho_\infty$, taking $\mu = \rho_\infty - \varrho_\infty$ in (B.2) implies $\rho_\infty = \varrho_\infty$, whence the uniqueness.

Step 2. By the semigroup property, for any $t > 0$ we have

$$S_t \rho_\infty = S_t S_T \rho_\infty = S_T S_t \rho_\infty,$$

and from the uniqueness of the fixed-point of S_T, we get that $S_t \rho_\infty = \rho_\infty$. Hence, ρ_∞ is the unique invariant probability measure.

The general estimate (3.3) is obtained by taking $\mu = u_0 - \rho_\infty$ and writing $t = \lfloor \frac{t}{T} \rfloor T + r_1$ with $0 \leq r_1 < T$. Indeed, using the semigroup property and nonexpansivity, we get

$$\|S_t \mu\|_{L^1} = \|S_{r_1} S_{\lfloor \frac{t}{T} \rfloor} \mu\|_{L^1} \leq \|S_{\lfloor \frac{t}{T} \rfloor T} \mu\|_{L^1} \leq (1 - \|\nu\|_{L^1})^{\lfloor \frac{t}{T} \rfloor} \|\mu\|_{L^1},$$

which implies (3.3) for the given K and $\mathfrak{a}$.

REFERENCES

[1] V. BANSAYE, B. CLOEZ, AND P. GABRIEL, *Ergodic behavior of non-conservative semigroups via generalized Doeblin's conditions*, Acta Appl. Math., 166 (2020), pp. 29–72.

[2] M. BEIRAN AND S. OSTOJIC, *Contrasting the effects of adaptation and synaptic filtering on the timescales of dynamics in recurrent networks*, PLoS Comput. Biol., 15 (2019), e1006893.

[3] J. BENDA AND A. V. HERZ, *A universal model for spike-frequency adaptation*, Neural Comput., 15 (2003), pp. 2523–2564.

[4] J. A. CAÑIZO AND S. MISCHLER, *Harris-type Results on Geometric and Subgeometric Convergence to Equilibrium for Stochastic Semigroups*, preprint, https://arxiv.org/abs/2110.09650, 2021.

[5] J. A. CAÑIZO AND H. YOLDAŞ, *Asymptotic behaviour of neuron population models structured by elapsed-time*, Nonlinearity, 32 (2019), pp. 464–495.

[6] J. CHEVALLIER, *Mean-field limit of generalized Hawkes processes*, Stochastic Process. Appl., 127 (2017), pp. 3870–3912.

[7] Q. CORMIER, *A Mean-Field Model of Integrate-and-Fire Neurons: Non-linear Stability of the Stationary Solutions*, preprint, https://arxiv.org/abs/2002.08649, 2020.

[8] Q. CORMIER, E. TANRÉ, AND R. VELTZ, *Long time behavior of a mean-field model of interacting neurons*, Stochastic Process. Appl., 130 (2020), pp. 2553–2595.

[9] Q. CORMIER, E. TANRÉ, AND R. VELTZ, *Hopf bifurcation in a mean-field model of spiking neurons*, Electron. J. Probab., 26 (2021), pp. 1–40.

[10] A. DE MASI, A. GALVES, E. LÖCHERBACH, AND E. PRESUTTI, *Hydrodynamic limit for interacting neurons*, J. Stat. Phys., 158 (2015), pp. 866–902.

[11] W. DOEBLIN, *Éléments d'une théorie générale des chaînes simples constantes de Markoff*, Ann. Sci. École Norm. Sup., 57 (1940), pp. 61–111.

[12] A. DROGOUL AND R. VELTZ, *Exponential stability of the stationary distribution of a mean field of spiking neural network*, J. Differential Equations, 270 (2021), pp. 809–842.

[13] A. DUARTE, E. LÖCHERBACH, AND G. OST, *Stability, convergence to equilibrium and simulation of non-linear Hawkes processes with memory kernels given by the sum of Erlang kernels*, ESAIM Probab. Stat., 23 (2019), pp. 770–796.

[14] G. DUMONT AND P. GABRIEL, *The mean-field equation of a leaky integrate-and-fire neural network: Measure solutions and steady states*, Nonlinearity, 33 (2020), pp. 6381–6420.

[15] N. FOURNIER AND E. LÖCHERBACH, *On a toy model of interacting neurons*, Ann. Inst. Henri Poincaré Probab. Stat., 52 (2016), pp. 1844–1876.

[16] P. GABRIEL, *Measure solutions to the conservative renewal equation*, ESAIM Proc. Surveys, 62 (2018), pp. 68–78.

[17] R. GAST, H. SCHMIDT, AND T. R. KNÖSCHE, *A mean-field description of bursting dynamics in spiking neural networks with short-term adaptation*, Neural Comput., 32 (2020), pp. 1615–1634.

[18] W. GERSTNER, *Time structure of the activity in neural network models*, Phys. Rev. E, 51 (1995), pp. 738–758.

[19] W. GERSTNER, *Population dynamics of spiking neurons: Fast transients, asynchronous states, and locking*, Neural Comput., 12 (2000), pp. 43–89.

[20] W. GERSTNER, W. M. KISTLER, R. NAUD, AND L. PANINSKI, *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*, Cambridge University Press, 2014.

[21] W. GERSTNER AND J. L. VAN HEMMEN, *Associative memory in a network of 'spiking' neurons*, Netw. Comput. Neural Syst., 3 (1992), pp. 139–164.

[22] G. GIGANTE, M. MATTIA, AND P. DEL GIUDICE, *Diverse population-bursting modes of adapting spiking neurons*, Phys. Rev. Lett., 98 (2007), 148101.

[23] M. HAIRER AND J. C. MATTINGLY, *Yet another look at Harris' ergodic theorem for Markov chains*, in Seminar on Stochastic Analysis, Random Fields and Applications VI, Springer, 2011, pp. 109–117.

[24] T. HARRIS, *The existence of stationary measures for certain Markov processes*, in Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory, University of California Press, 1956, pp. 113–124.

[25] R. HÖPFNER, E. LÖCHERBACH, AND M. THIEULLEN, *Ergodicity and limit theorems for degenerate diffusions with time periodic drift: Application to a stochastic Hodgkin-Huxley model*, ESAIM Probab. Stat., 20 (2016), pp. 527–554.

[26] E. M. Izhikevich, *Neural excitability, spiking and bursting*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 10 (2000), pp. 1171–1266.

[27] R. Jolivet, A. Rauch, H.-R. Lüscher, and W. Gerstner, *Predicting spike timing of neocortical pyramidal neurons by simple threshold models*, J. Comput. Neurosci., 21 (2006), pp. 35–49.

[28] R. Kobayashi, Y. Tsubo, and S. Shinomoto, *Made-to-order spiking neuron model equipped with a multi-timescale adaptive threshold*, Front. Comput. Neurosci., 3 (2009).

[29] S. P. Meyn and R. L. Tweedie, *Stability of Markovian processes* III*: Foster-Lyapunov criteria for continuous-time processes*, Adv. in Appl. Probab., 25 (1993), pp. 518–548.

[30] S. Mischler, C. Quiñinao, and Q. Weng, *Weak and strong connectivity regimes for a general time elapsed neuron network model*, J. Stat. Phys., 173 (2018), pp. 77–98.

[31] S. Mischler and Q. Weng, *Relaxation in time elapsed neuron network models in the weak connectivity regime*, Acta Appl. Math., 157 (2018), pp. 45–74.

[32] S. P. Muscinelli, W. Gerstner, and T. Schwalger, *How single neuron properties shape chaotic dynamics and signal transmission in random neural networks*, PLoS Comput. Biol., 15 (2019), e1007122.

[33] R. Naud and W. Gerstner, *Coding and decoding with adapting neurons: A population approach to the peri-stimulus time histogram*, PLoS Comput. Biol., 8 (2012), e1002711.

[34] A. Omurtag, B. W. Knight, and L. Sirovich, *On the simulation of large populations of neurons*, J. Comput. Neurosci., 8 (2000), pp. 51–63.

[35] K. Pakdaman, B. Perthame, and D. Salort, *Dynamics of a structured neuron population*, Nonlinearity, 23 (2010), pp. 55–75.

[36] K. Pakdaman, B. Perthame, and D. Salort, *Relaxation and self-sustained oscillations in the time elapsed neuron network model*, SIAM J. Appl. Math., 73 (2013), pp. 1260–1279, https://doi.org/10.1137/110847962.

[37] K. Pakdaman, B. Perthame, and D. Salort, *Adaptation and fatigue model for neuron networks and large time asymptotics in a nonlinear fragmentation equation*, J. Math. Neurosci., 4 (2014), 14.

[38] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. Chichilnisky, and E. P. Simoncelli, *Spatio-temporal correlations and visual signalling in a complete neuronal population*, Nature, 454 (2008), pp. 995–999.

[39] C. Pozzorini, S. Mensi, O. Hagens, R. Naud, C. Koch, and W. Gerstner, *Automated high-throughput characterization of single neurons by means of simplified spiking models*, PLoS Comput. Biol., 11 (2015), e1004275.

[40] C. Pozzorini, R. Naud, S. Mensi, and W. Gerstner, *Temporal whitening by power-law adaptation in neocortical neurons*, Nat. Neurosci., 16 (2013), pp. 942–948.

[41] M. J. Richardson, *Dynamics of populations and networks of neurons with voltage-activated and calcium-activated currents*, Phys. Rev. E, 80 (2009), 021928.

[42] S. Romani, D. J. Amit, and G. Mongillo, *Mean-field analysis of selective persistent activity in presence of short-term synaptic depression*, J. Comput. Neurosci., 20 (2006), pp. 201–217.

[43] V. Schmutz, *Mean-field limit of age and leaky memory dependent Hawkes processes*, Stochastic Process. Appl., 149 (2022), pp. 39–59.

[44] T. Schwalger, M. Deger, and W. Gerstner, *Towards a theory of cortical columns: From spiking neurons to interacting neural populations of finite size*, PLoS Comput. Biol., 13 (2017), e1005507.

[45] C. Teeter, R. Iyer, V. Menon, N. Gouwens, D. Feng, J. Berg, A. Szafer, N. Cain, H. Zeng, M. Hawrylycz, C. Koch, and S. Mihalas, *Generalized leaky integrate-and-fire models classify multiple neuron types*, Nat. Commun., 9 (2018), 709.

[46] W. Truccolo, U. T. Eden, M. R. Fellows, J. P. Donoghue, and E. N. Brown, *A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects*, J. Neurophysiol., 93 (2005), pp. 1074–1089.

[47] M. Tsodyks, K. Pawelzik, and H. Markram, *Neural networks with dynamic synapses*, Neural Comput., 10 (1998), pp. 821–835.

[48] S. Vellmer and B. Lindner, *Theory of spike-train power spectra for multidimensional integrate-and-fire neurons*, Phys. Rev. Res., 1 (2019), 023024.

[49] C. van Vreeswijk and D. Hansel, *Patterns of synchrony in neural networks with spike adaptation*, Neural Comput., 13 (2001), pp. 959–992.

[50] H. R. Wilson and J. D. Cowan, *Excitatory and inhibitory interactions in localized populations of model neurons*, Biophys. J., 12 (1972), pp. 1–24.

[51] R. S. Zucker and W. G. Regehr, *Short-term synaptic plasticity*, Annu. Rev. Physiol., 64 (2002), pp. 355–405.

4 On a finite-size neuronal population equation

Authors: Valentin Schmutz, Eva Löcherbach and Tilo Schwalger

Contribution: All three authors contributed equally to this work.

Article accepted in *SIAM Journal on Applied Dynamical Systems.*

On a finite-size neuronal population equation[*]

Valentin Schmutz[†], Eva Löcherbach[‡], and Tilo Schwalger[§]

Abstract. Population equations for infinitely large networks of spiking neurons have a long tradition in theoretical neuroscience. In this work, we analyze a recent generalization of these equations to populations of finite size, which takes the form of a nonlinear stochastic integral equation. We prove that, in the case of leaky integrate-and-fire (LIF) neurons with escape noise and for a slightly simplified version of the model, the equation is well-posed and stable in the sense of Brémaud-Massoulié. The proof combines methods from Markov processes taking values in the space of positive measures and nonlinear Hawkes processes. For applications, we also provide efficient simulation algorithms.

Key words. Stability, finite-size fluctuations, nonlinear Hawkes processes, piecewise-deterministic Markov processes, Meyn-Tweedie theory, spiking neuron, SPDE's driven by Poisson random measure.

MSC codes. 60G55 (primary) 60H20, 60K35, 92B20 (secondary)

1. Introduction. Neuronal population equations describe the dynamics of large networks of neurons in terms of single neuron parameters [31]. As such, they are useful mathematical abstractions for relating microscopic and large-scale brain signals, and contribute to the biophysical interpretation of the latter [17]. Their motivation is twofold: on the one hand, they enable the theoretical analysis of emergent phenomena, like collective oscillations [7, 30, 14]; on the other hand, from the data analysis point of view, they constitute the basis of 'forward models' of large-scale brain signals [17, 44, 8, 4, 26]. This second motivation requires neuronal population equations to achieve the right balance between accuracy (the equation faithfully captures the dynamics of the population of neurons it represents) and usability (the equation can be efficiently simulated).

An example of such neuronal population equation is the integral equation (or refractory density equation) for a homogeneous network of spiking neurons ("neuronal population") [29, 30, 12, 31, 47]. Contrary to standard neural-mass models [52, 17, 35], the integral equation captures the effect of neuronal refractoriness on the mean population dynamics [12, 31, 47], and is exact in the mean-field limit if neurons are modeled as intensity-based renewal point processes [16, 25, 10]. Specific examples of the integral equation are the time-elapsed neuron network model [39] (or age-structured model [22]) and the voltage-structured model of [16, 25].

Besides capturing the effect of single neuron dynamics (such as post-spike refractory ef-

[†]Brain Mind Institute, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (valentin.schmutz@epfl.ch)

[‡]Statistique, Analyse et Modélisation Multidisciplinaire, EA 4543 et FR FP2M 2036 CNRS, Université Paris 1 Panthéon-Sorbonne, 75013 Paris, France (locherbach70@gmail.com)

[§]Institut für Mathematik, Technische Universität Berlin, 10623 Berlin, Germany; Bernstein Center for Computational Neuroscience Berlin, 10115 Berlin, Germany (schwalger@math.tu-berlin.de)

fects) on the *mean* population dynamics, there is a second challenge for neuronal population equations: the proper account of *fluctuations*. Fluctuations of the average population activity arise in the case of finite population sizes, and vanish in the mean-field limit of infinitely many neurons. From a modeling perspective, an important question arises: Are the relevant neuronal populations large enough so that finite-size fluctuations can be neglected? There is no clear answer to this question but the anatomical and functional organization of the cerebral cortex into different cortical areas, columns and layers each containing different cell classes [32, 41, 45, 2] requires a subdivision of a cortical circuit into many, relatively small populations. For example, at the scale of a cortical column, empirical data from mouse barrel cortex suggests populations of around 50 to 2000 neurons [36]. For these population sizes, finite-size fluctuations are non-negligible and this noise may strongly impact the nonlinear population dynamics [48]. Therefore, modeling cortical circuits at the mesoscopic scale of populations requires a stochastic description, in marked contrast to the deterministic integral equation.

Rigorous extensions of the integral equations to account for finite-size fluctuations are subject to an accuracy/usability trade-off. If neuronal refractoriness is neglected, the population equation reduces to that of [19, 20] and finite-size noise can be added, by the linear-noise approximation [33], or granting some Markov embedding, by the diffusion approximation [20], whose numerical implementation is relatively simple [11]. These approaches fail to reproduce the non-stationary dynamics of the mean population activity and the temporal correlation structure of fluctuations for a population of spiking neurons with refractoriness (Figure 1a). On the other hand, if one does not neglect refractoriness, central limit theorem-based arguments lead to formal SPDE's [9, 23], which are computationally expensive to simulate, or to formal integral equations with colored noise [18], for which a simulation algorithm is unknown.

In [48], a heuristic extension of the integral equation with finite-size fluctuations is derived. It can be easily simulated and takes into account the effects of neuronal refractoriness. While this extension is not exact, its numerical implementation gives an accurate approximation to the dynamics of finite-size networks of spiking neurons, such as the broad class of generalized integrate-and-fire neurons [42, 48] and formal renewal-type neurons [30, 40]. Moreover, since it takes the form of an intensity-based point process, the likelihood of a population spike train can be easily computed, which enables efficient data fitting [43, 51]. The intensity function of this point process exhibits a novel type of nonlinear history dependence that goes beyond nonlinear Hawkes processes and has not been studied mathematically so far. In particular, the stability of the process observed in simulations is poorly understood from the theoretical point of view. Therefore, the aim of this work is to give a rigorous foundation to the model of [48] and prove its stability.

Below, we briefly give a review of some standard population equations. We then present the finite-size model of [48] in a slightly simplified form. Finally, we show that the simplified model, in the case of leaky integrate-and-fire (LIF) neurons with escape noise [30, 27], can be written as a SPDE driven by Poisson noise, which will be the main object of study in this work.

1.1. Neuronal population equations. To give a mathematical introduction to the integral equation formalism, it is useful to consider the special case of LIF neurons with escape noise [30, 27], which is also the main case we will treat in this work. Let us consider a network of

N identical neurons that are all-to-all connected with uniform connection strength J/N for $J \in \mathbb{R}$. Each neuron i has a voltage variable $U^{i,N}$ which evolves according to the system of SDE's: For all $i = 1, \ldots, N$,

$$(1.1a) \qquad dU_t^{i,N} = \frac{\mu_t - U_t^{i,N}}{\tau_{\mathrm{m}}} dt - U_{t^-}^{i,N} dZ_t^{i,N} + \frac{J}{N} \sum_{j=1}^N dZ_t^{j,N},$$

$$(1.1b) \qquad Z_t^{i,N} = \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \leq f(U_{s^-}^{i,N})} \pi^i(ds, dz).$$

Here, $Z_t^{i,N}$ is the spike counting process of the neuron i and has intensity $f(U_{t^-}^{i,N})$, t^- denoting the left limit. Furthermore, μ_t comprises the resting potential and the (possibly time-dependent) external drive, τ_{m} is the membrane time constant, $f : \mathbb{R} \to \mathbb{R}_+$ is the intensity function and $\{\pi^i\}_{i=1,\ldots,N}$ is a collection of independent Poisson random measures on $\mathbb{R}_+ \times \mathbb{R}_+$ with Lebesgue intensity measure.

Equation (1.1) is called a *microscopic model* because the neuronal dynamics is modeled with single-cell resolution (Figure 1a, top). A drastic reduction of the complexity of the model can be achieved by coarse-graining over the population of neurons. To this end, we consider the *empirical population activity*

$$(1.2) \qquad A_{t,\mathfrak{h}}^N = \frac{1}{N} \sum_{i=1}^N \frac{Z_{t+\mathfrak{h}}^{i,N} - Z_t^{i,N}}{\mathfrak{h}},$$

where $\mathfrak{h} > 0$ is a small time interval determining the temporal resolution (Figure 1a, bottom). Neuronal population equations are models of such coarse-grained quantities that describe the neuronal dynamics at the scale of whole populations. If the population is of finite size ($N < \infty$), the dynamics is called a *mesoscopic model*, while the dynamics for an infinitely large population ($N \to \infty$) is referred to as a *macroscopic model*. In [16, 25], the authors proved that in the macroscopic limit $N \to \infty$, if the initial conditions $\{U_0^i\}_{i=1,\ldots,N}$ are *i.i.d.* with law ν_0, the empirical measure of the system (1.1) is characterized by the voltage-structured PDE (with solutions in the sense of measures [14]): For all $u \in \mathbb{R}$ and $t > 0$,

$$(1.3a) \qquad \partial_t \rho(du, t) + \partial_u \left(\left(\frac{\mu_t - u}{\tau_{\mathrm{m}}} + J\rho_t[f] \right) \rho(du, t) \right) = -f(u)\rho(du, t) + \rho_t[f]\delta_0(du),$$

$$(1.3b) \qquad \rho_0 = \nu_0,$$

where $\rho_t := \rho(\cdot, t)$ and $\rho_t[f] := \int_{\mathbb{R}} f(u)\rho(du, t)$.

The latter can be interpreted as the population activity

$$(1.4) \qquad \lim_{\mathfrak{h} \downarrow 0} \lim_{N \to \infty} A_{t,\mathfrak{h}}^N = A(t) := \rho_t[f].$$

Furthermore, $\rho_t[1] = 1$ for all $t > 0$ expressing the fact that the number of neurons is conserved.

We now transform (1.3) into an integral equation. For all continuous functions $a : \mathbb{R}_+ \to \mathbb{R}$, we define the time-dependent vector field $b^a(t, u) := (\mu_t - u)/\tau_{\mathrm{m}} + Ja(t)$ and write, for all

$0 \le s \le t$, $\Phi_{s,t}^a(u)$ the associated flow given by

$$(1.5) \qquad \Phi_{s,t}^a(u) := u e^{-\frac{t-s}{\tau_\mathrm{m}}} + \int_s^t e^{-\frac{t-r}{\tau_\mathrm{m}}} \frac{\mu_r}{\tau_\mathrm{m}}\, dr + J \int_s^t e^{-\frac{t-r}{\tau_\mathrm{m}}} a_r dr, \qquad \forall u \in \mathbb{R}.$$

We can now define, for all $0 \le s \le t$,

$$(1.6) \qquad \lambda^a(t|s) := f(\Phi_{s,t}^a(0)) \qquad \text{and} \qquad S^a(t|s) := \exp\left(-\int_s^t \lambda^a(r|s)\, dr\right).$$

The function $\lambda^a(t|s)$, called hazard rate, gives the intensity at time t (i.e. the instantaneous probability of emitting a spike) as a function of the time of the last spike s and the past population activity $(a(r))_{s \le r \le t}$; the membrane potential dynamics of LIF neurons – leaky integration and spike-triggered reset, (1.1a) – are accounted for in the definition of $\lambda^a(t|s)$. Similarly, the function $S^a(t|s)$, called the survival, gives the probability of not emitting a spike in the time interval $]s, t[$, given that the last spike was emitted at time s. By the method of characteristics, we get that the population activity $A(t)$ solves the integral equation

$$(1.7) \qquad A(t) = H^A(t) + \int_0^t \lambda^A(t|s) S^A(t|s) A(s) ds,$$

where

$$(1.8) \qquad H^A(t) := \int_{\mathbb{R}} f(\Phi_{0,t}^A(u)) e^{-\int_0^t f(\Phi_{0,r}^A(u)) dr} \nu_0(du).$$

Equation (1.7) is the integral equation of [52, 29, 30], see also [13]. Note that, traditionally, the integral equation has no explicit initial condition and therefore requires a normalizing condition [31, Sec. 14.1]. The integral equation (1.7) is normalized such that

$$(1.9) \qquad \widetilde{H}^A(t) + \int_0^t S^A(t|s) A(s)\, ds = 1$$

for all $t > 0$, where we defined

$$(1.10) \qquad \widetilde{H}^A(t) := \int_{\mathbb{R}} e^{-\int_0^t f(\Phi_{0,r}^A(u)) dr} \nu_0(du).$$

The normalization, (1.9), expresses the fact that the number of neurons is conserved.[1] Note that the integral equation (1.7) is simply the time derivative of the normalizing condition (1.9); this fact has been originally used to derive the integral equation [30].

---

[1] The conservation of neuronal mass can be understood as follows: At time t, $\widetilde{H}^A(t)$ represents the fraction of neurons (#neurons divided by N) that had their unique last spike before time 0, while for $s \in [0, t[$ the term $S^A(t|s) A(s) ds$ represents the fraction of neurons that had their unique last spike time in the interval $[s, s+ds[$ (here $A(s)ds$ is the fraction of neurons that fired in that interval and $S^A(t|s)$ is the probability for one neuron of not emetting a spike in $]s, t[$ given a spike at time s). Therefore, $\int_0^t S^A(t|s) A(s)\, ds$ represents the fraction of neurons that had their unique last spike in $[0, t[$. Hence, (1.9) states that the fraction of neurons at time t that had their unique last spike time before time t (either before time 0 or since time 0) is equal to unity. Since this statement holds for all $t > 0$ and each neuron has exactly one last spike time before time t, the total number of neurons must be conserved.

In the case of LIF neurons with escape noise, the voltage-structured equation (1.3) is equivalent to the integral equation (1.7) if $\lambda^A(t|s)$ is defined by (1.6). However, we could have chosen a different definition for the hazard rate $\lambda^A(t|s)$; the integral equation is therefore more general than (1.3). In fact, (1.7) can be seen as a renewal equation that holds for any population of neurons modeled as time-inhomogeneous renewal processes [40]. For example, the Fokker-Planck equation for neuronal networks with diffusive noise (see [31, Ch. 13]) or the time-elapsed neuron network model [39] can also be written as an integral equation with a suitable choice of the hazard rate.

1.2. The finite-size integral equation. In [48], the authors derive a generalization of the integral equation (1.7) which takes into account finite-size noise. For clarity, we will present the equation of [48] in the case of LIF neurons with escape noise. Before presenting the model, we need to extend the definitions (1.6). For all non-decreasing functions $z : \mathbb{R}_+ \ni t \mapsto z_t$ with bounded variation on finite time intervals, we redefine, for all $0 \leq s \leq t$,

$$(1.11) \qquad \Phi_{s,t}^z(u) := u e^{-\frac{t-s}{\tau_{\mathrm{m}}}} + \int_s^t e^{-\frac{t-r}{\tau_{\mathrm{m}}}} \frac{\mu_r}{\tau_{\mathrm{m}}} \, dr + J \int_{]s,t]} e^{-\frac{t-r}{\tau_{\mathrm{m}}}} \, dz_r, \qquad \forall u \in \mathbb{R}.$$

We can now extend the definitions (1.6), (1.8), and (1.10), replacing Φ^A by (1.11).

For a finite number of neurons N, the finite-size integral equation of [48] ("mesoscopic model") can be written as follows: For all $t \geq 0$,

$$(1.12\mathrm{a}) \quad Z_t = \frac{1}{N} \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \leq N \bar{A}_{s-}} \pi(ds, dz),$$

$$(1.12\mathrm{b}) \quad \bar{A}_t = \left[H^Z(t) + \int_{[0,t]} \lambda^Z(t|s) S^Z(t|s) dZ_s + \Lambda_t^Z \left(1 - \widetilde{H}^Z(t) - \int_{[0,t]} S^Z(t|s) dZ_s \right) \right]_+,$$

$$(1.12\mathrm{c}) \quad \Lambda_t^Z = \frac{G^Z(t) + \int_{[0,t]} \lambda^Z(t|s)\{1 - S^Z(t|s)\} S^Z(t|s) dZ_s}{\widetilde{G}^Z(t) + \int_{[0,t]} \{1 - S^Z(t|s)\} S^Z(t|s) dZ_s},$$

where π is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ with Lebesgue intensity measure and $[\,\cdot\,]_+ = \max(0, \cdot)$. The functions G^Z and $\widetilde{G}^Z$ are analogous to H^Z and $\widetilde{H}^Z$:

$$G^Z(t) := \int_{\mathbb{R}} f(\Phi_{0,t}^Z(u)) \left\{ 1 - e^{-\int_0^t f(\Phi_{0,r}^Z(u))dr} \right\} e^{-\int_0^t f(\Phi_{0,r}^Z(u))dr} \nu_0(du),$$

$$\widetilde{G}^Z(t) := \int_{\mathbb{R}} \left\{ 1 - e^{-\int_0^t f(\Phi_{0,r}^Z(u))dr} \right\} e^{-\int_0^t f(\Phi_{0,r}^Z(u))dr} \nu_0(du).$$

The mesoscopic model (1.12) defines a jump process Z_t where jumps of size $1/N$ occur with intensity $N \bar{A}_{t-}$. The derivation of (1.12), explained in detail in [48, pp. 35–43], involves heuristic arguments and approximations. Consequently, this mesoscopic model is inexact (in contrast to the formal SPDE of [9, 23]). However, extensive numerical simulations have shown that the model is highly accurate in many multiscale modeling applications [48] (see also Figure 1b). Moreover, it has the advantage of being an intensity-based and history-dependent point process, and as such, can be efficiently simulated and used for statistical data analysis
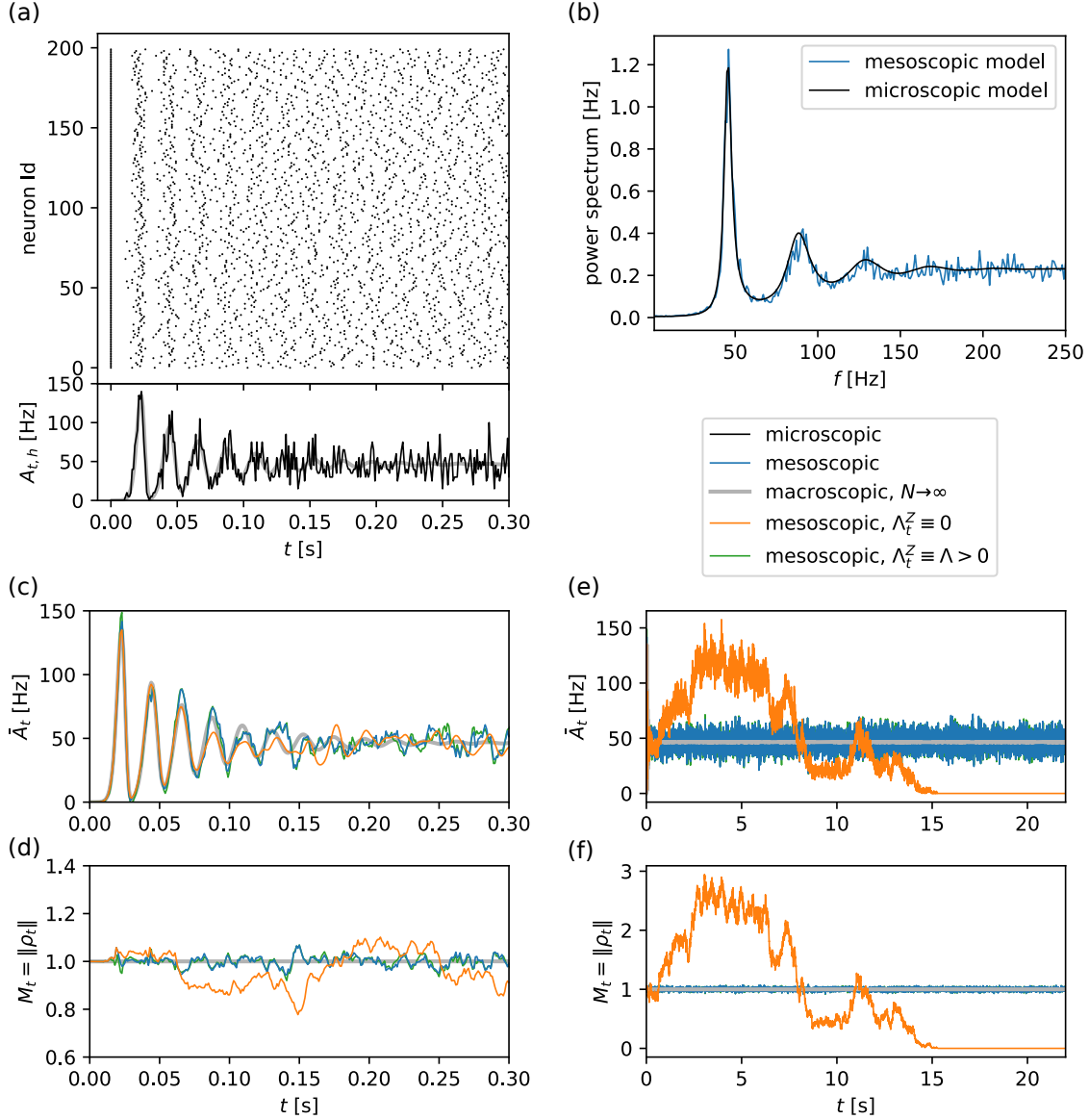
Figure 1. *Mesoscopic population dynamics. (a) Top: Spike-raster plot of a microscopic model of $N = 200$ uncoupled LIF neurons with escape noise, (1.1) with $J = 0$. Neurons were initialized in a synchronized state, i.e. all neurons spiked at time $t = 0$. Bottom: Empirical population activity measured with temporal bin size $\mathfrak{h} = 0.001s$ (black line) and macroscopic population activity predicted by the deterministic integral equation (1.7) for $N \to \infty$ with $\nu_0 = \delta_0$ (gray line). (b) Comparison of the power spectral densities (as defined in Appendix D, see also [48]) of the empirical population activities $A_{t,\mathfrak{h}}(t)$ of the microscopic model (black line, exact theory [31]) and $\hat{A}_{t,\mathfrak{h}}(t)$ of the mesoscopic model (blue line, simulation). (c,d) $\bar{A}_t$ (1.12b) and mass M_t (1.14) for simulations of the mesoscopic model (blue line) and the 'naive' mesoscopic model with $\Lambda_t^Z \equiv 0$ (orange line). For comparison, the macroscopic model and the mesoscopic model with fixed $\Lambda_t^Z \equiv 277$ Hz (corresponding to the temporally averaged Λ_t^Z of the mesoscopic model) are shown by gray and green lines, respectively. (e,f) Same as (c,d) but for a longer simulation time. Parameters: $\tau_m = 0.02$ s, $\mu = 20$ mV, $f(u) = ce^{(u-\vartheta)/\Delta_u}$, $c = 10$ Hz, $\vartheta = 10$ mV, $\Delta_u = 1$ mV.*

[43]. A concise derivation of (1.12) is presented in section 4, where we also show that for some convenient initial condition, the functions H^Z, $\widetilde{H}^Z$, G^Z and $\widetilde{G}^Z$ are trivial.

The finite-size analog of the population activity $A(t)$ for infinitely large populations (1.7) is the distributional derivative of Z_t,

$$\dot{Z}_t = \frac{1}{N} \sum_k \delta(t - t_k),$$

where t_k are the jump times of Z_t and $\delta(\cdot)$ denotes the Dirac delta distribution.[2] We call $\dot{Z}_t$ the population spike train (sum of δ-pulses at spike times t_k). Note that the biologically relevant quantity is the empirical population activity at a finite time resolution, $\hat{A}^N_{t,\mathfrak{h}} :=$ $\mathfrak{h}^{-1} \int_t^{t+\mathfrak{h}^+} \dot{Z}_s \, ds = [Z(t + \mathfrak{h}) - Z(t)]/\mathfrak{h}$, for some small time interval $\mathfrak{h} > 0$. Furthermore, we will often call the finite-size population model, (1.12), *mesoscopic model* in contrast to "macroscopic model" that refers to the case $N \to \infty$. Note that the variables $\bar{A}_t$ and Z_t describe the neuronal activity of the population as a whole, driven by only one single Poisson noise $\pi(dt, [0, N\bar{A}_{t-}])$. A time discretization of the mesoscopic model permits an efficient simulation of the neuronal dynamics directly on the population level, without the need to simulate individual neurons (see section 5 and Algorithm A.1). Importantly, even though the mesoscopic model is an approximation, it accurately captures the statistics of the population activity $A^N_{t,h}$ of the original microscopic model. In particular, the fluctuation statistics of the population activities $A^N_{t,h}$ and $\hat{A}^N_{t,h}$, as expressed by their power spectral density, are well matched (Figure 1b, also see [48] for further examples).

A key difference between the macroscopic model for an infinitely large population (1.7) and the mesoscopic model (1.12) is the 'correction term' $\Lambda^Z_t(1 - \dots)$ in (1.12b) arising due to finite network size, $N < \infty$. This correction term may seem unexpected in light of the following heuristic argument: in (1.7) for infinite N, the fraction of neurons $A(s)ds$ firing in the past, $s < t$, contribute to the current activity $A(t)dt$ with probability $\lambda^A(t|s)S^A(t|s)dt$. For finite N, the corresponding fraction of neurons is dZ_s, and assuming that the probability to fire their next spike at time t is again given by $\lambda^A(t|s)S^A(t|s)dt$, the expected activity should be given by the much simpler expression $\bar{A}_{t,\text{naive}} = H^Z(t) + \int_0^t \lambda^Z(t|s)S^Z(t|s)dZ_s$. This naive finite-size model is obtained by putting $\Lambda^Z_t \equiv 0$, and thus lacks the 'correction term'. Numerical simulations of the naive finite-size model indeed reproduce the transient initial dynamics of the population activity at short times, including damped oscillations caused by refractoriness (Figure 1c, orange curve). However, longer simulations of the naive model reveal that the population rate $\bar{A}_t$ strongly fluctuates and eventually collapses to the silent solution $\bar{A}_t = 0$. In contrast, the mesoscopic model, (1.12) with $\Lambda^Z_t > 0$, reaches a non-silent, stationary state consistent with the microscopic model (1.1) (Figure 1e). A completely open theoretical question is: Why does the 'correction term' in (1.12b) 'stabilize' the finite-size neuronal population dynamics?

To address this question mathematically, we focus our analysis on the case where the modulating factor Λ^Z_t is fixed ($\Lambda^Z_t \equiv \Lambda > 0$). This is a simplified version of the finite-size

---

[2]Formally, $\dot{Z}_t dt := dZ_t$, where dZ is the Lebesgue-Stieltjes measure associated with the counting measure Z.

integral equation (1.12), for which we can prove a rigorous stability result. Note that fixing $\Lambda_t^Z \equiv \Lambda > 0$ is for mathematical tractability only; for practical modeling, Λ_t^Z as defined in (1.12c) should be preferred (a detailed simulation algorithm is presented in section 5).

Before presenting our main stability result in subsection 1.4, we provide some additional insights into the mechanisms of the finite-size integral equation (1.12), in particular, why is the naive model ($\Lambda = 0$) expected to fail. First, in subsection 1.2.1, we show a close relationship between the finite-size integral equation and nonlinear Hawkes processes, for which stability properties are well known. Second, in subsection 1.2.2, we propose a heuristic argument for the stability in terms of neuronal mass conservation and an analogy with the Cox-Ingersoll-Ross process.

1.2.1. Relationship with nonlinear Hawkes processes. If $J = 0$ (neurons do not interact), $\mu_t \equiv \mu$ (the external drive is constant) and $\Lambda_t^Z \equiv \Lambda$, (1.12) reduces to a nonlinear Hawkes process [5]: For all $t \geq 0$,

$$(1.13a) \qquad Z_t = \frac{1}{N} \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \leq N \bar{A}_{s-}} \pi(ds, dz),$$

$$(1.13b) \qquad \bar{A}_t = \left[\Lambda + H^0(t) - \Lambda \widetilde{H}^0(t) + \int_{[0,t]} \underbrace{(\lambda^0(t|s) - \Lambda) S^0(t|s)}_{=:h^\Lambda(t-s)} dZ_s \right]_+,$$

where λ^0, S^0, H^0 and $\tilde{H}^0$ correspond to the definitions (1.6), (1.8), and (1.10) when Φ^Z (1.11) is replaced by $\Phi^0_{s,t}(u) = u e^{-\frac{t-s}{\tau_{\mathrm{m}}}} + \int_s^t e^{-\frac{t-r}{\tau_{\mathrm{m}}}} \frac{\mu_r}{\tau_{\mathrm{m}}} dr$.

The function $h^\Lambda : \mathbb{R}_+ \to \mathbb{R}$ in (1.13b) can be interpreted as the self-interaction kernel of the nonlinear Hawkes process. The model (1.13) is not particularly useful in practice since it only approximates the dynamics of a population of non-interacting neurons with constant external input. Nevertheless it sheds light on the role of Λ on the stability of the mesoscopic model and it helps to see why the theory of nonlinear Hawkes processes [5] will prove to be instrumental in this work. It is easy to verify that $\int_0^\infty h^\Lambda(t) dt = 1$ if $\Lambda = 0$ and $\int_0^\infty h^\Lambda(t) dt < 1$ if $\Lambda > 0$. If $\Lambda = 0$, (1.13) is a critical Hawkes process and has a nontrivial stationary solution only if h^0 is heavy-tailed [6] (which is not the case for the neuron models considered here). On the other hand, if $\Lambda > 0$, (1.13) is a stable nonlinear Hawkes process with a unique stationary solution (Theorem 1 in [5] and see also [15]). Hence, in the time-homogeneous ($\mu_t \equiv \mu$) and non-interacting case ($J = 0$), $\Lambda_t \equiv \Lambda > 0$ is a sufficient condition for the stability of (1.13), in the sense of [5].

To generalize this stability result to the interacting case ($J \neq 0$), we will use a Markov embedding of (1.12) and Meyn-Tweedie theory [37], in addition to standard techniques for nonlinear Hawkes processes [5].

1.2.2. Approximate conservation of neuronal mass. In contrast to the conservation of neuronal mass in the macroscopic model, (1.9), such a strict conservation law does no longer hold for the mesoscopic model, (1.12). However, in analogy to (1.9), we would expect the neuronal "mass"

$$(1.14) \qquad M_t := \widetilde{H}^Z(t) + \int_{[0,t]} S^Z(t|s) dZ_s$$

to stay close to 1. This feature is supported by simulations of the mesoscopic model showing that M_t fluctuates around unity (Figure 1d,f). Indeed, the number of neurons in the system (1.1) being obviously constant, the finite-size population model (1.12) should reflect this mass conservation principle.

Let us consider the first hitting time $\tau^* = \inf\{t > 0 : \bar{A}_t = 0\}$. For $0 < t < \tau^*$, the intensity $\bar{A}_t$ is strictly positive, hence (1.12b) can always be written as

$$\bar{A}_t = H^Z(t) + \int_{[0,t]} \lambda^Z(t|s) S^Z(t|s) dZ_s + \Lambda_t^Z (1 - M_t).$$

By formal differentiation of (1.14), we obtain for $0 < t < \tau^*$

$$(1.15) \qquad dM_t = -H^Z(t)dt + dZ_t - \left(\int_{[0,t]} \lambda^Z(t|s) S^Z(t|s) \, dZ_s \right) dt = \Lambda_t^Z(1 - M_t)dt + d\tilde{Z}_t,$$

where $\tilde{Z}_t := Z_t - \int_0^t \bar{A}_s \, ds$ is the compensated jump process. Equation (1.15) yields some rough insights into the dynamics of the neuronal mass M_t. For simplicity, let us assume $\Lambda_t^Z \equiv \Lambda$ to be constant. First, the conditional mean $\bar{M}_t^c := \mathbb{E}[M_t | \tau^* > t]$ can be obtained by averaging (1.15): $d\bar{M}_t^c = \Lambda(1 - \bar{M}_t^c)dt$. This equation shows that its solution, $\bar{M}_t^c = 1 + (\widetilde{H}^Z(0) - 1)e^{-\Lambda t}$, is attracted to unity if $\Lambda > 0$. Conversely, in the naive model, when $\Lambda = 0$, the conditional mean does not drift towards unity but remains constant, $\bar{M}_t^c = \widetilde{H}^Z(0)$ for all $t > 0$. Second, in the naive model ($\Lambda = 0$), once M_t hits the boundary 0, it sticks to this boundary forever, i.e. $M_t = 0$ for all $t > \tau^*$ (Figure 1f). In fact, if f is upper bounded by $\|f\|_\infty < \infty$, we have $0 \le \bar{A}_t \le \|f\|_\infty M_t + \Lambda(1 - M_t)$. Thus, $M_t = 0$ and $\Lambda = 0$ entails that $\bar{A}_t = 0$, and hence the "noise" $d\tilde{Z}$ in (1.15) vanishes.

Third, if the jumps of $\tilde{Z}_t$ are small and frequent enough and if the increments of $\tilde{Z}_t$ are 'independent' enough, we may replace $d\tilde{Z}_t$ by its diffusion approximation $\sqrt{\bar{A}_t/N} dW_t$, where W_t is a Wiener process. If we further assume that $\bar{A}_t$ and M_t vary roughly in proportion (as suggested by Figure 1e,f for the naive model), we expect that M_t behaves like a Cox-Ingersoll-Ross process, $d\hat{M}_t = \Lambda(1 - \hat{M}_t)dt + \sigma\sqrt{\hat{M}_t} dW_t$ where σ is the volatility parameter. Due to the drift term, this process fluctuates around its mean $\mathbb{E}[\hat{M}_t] = 1$ if $\Lambda > 0$, consistent with simulations of the model (Figure 1d,f). Such drift force is absent in the naive model, $\Lambda = 0$, in which case $d\hat{M}_t = \sigma\sqrt{\hat{M}_t} dW_t$ describes the critical Feller branching diffusion which goes extinct in the long run (and once it hits 0 remains there forever), with extinction probability $P(\hat{M}_t = 0 | \hat{M}_0 = x) = e^{-\frac{x}{\sigma^2 t}}$.

1.3. Markov embedding of the finite-size integral equation. As the voltage-structured equation (1.3) can be transformed into an integral equation, assuming $\Lambda_t^Z \equiv \Lambda$, we can transform the stochastic integral equation (1.12) back into a voltage-structured SPDE driven by Poisson noise. Denoting $\mathcal{M}_+$ the space of nonnegative finite measures on $\mathbb{R}$, for all $\mathcal{M}_+$-valued random variables $\widehat{\nu}_0$, the SPDE formally writes:

For all $t > 0$ and $u \in \mathbb{R}$,

$$(1.16\text{a}) \qquad \partial_t \rho(du, t) + \partial_u \left(\left(\frac{\mu_t - u}{\tau_\mathrm{m}} + J \dot{Z}_t \right) \rho(du, t^-) \right) = -f(u)\rho(du, t) + \dot{Z}_t \delta_0(du),$$

$$(1.16\text{b}) \qquad Z_t = \frac{1}{N} \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \leq N\bar{A}_{s^-}} \pi(ds, dz) \quad \text{with } \bar{A}_t := [\rho_t[f] + \Lambda(1 - \|\rho_t\|)]_+,$$

$$(1.16\text{c}) \qquad \rho_0 = \widehat{\nu}_0,$$

where $\|\cdot\|$ denotes the total variation norm, that is, the total mass of the measure.

We will give a precise meaning to the SPDE (1.16) and show that it is equivalent to the stochastic integral equation (1.12) in section 2 below. The two jump terms $\partial_u(J\dot{Z}_t\rho(du, t^-))$ and $\dot{Z}_t \delta_0(du)$ have the following interpretation. At each jump time of Z_t, the current mass of the solution $\rho(du, t)$ is shifted by J/N and a mass $(1/N)\delta_0$ is added to the current value of the solution (emulating the membrane potential reset of LIF neurons, (1.1a)). Although the jump intensity $N\bar{A}_{t^-}$ of Z_t is not a priori bounded, we shall prove in Lemma 2.4 below that almost surely Z has only a finite number of jumps within each finite time interval such that (1.16) is well-posed as a measure-valued piecewise deterministic Markov process having càdlàg trajectories.

We say that (1.16) is the Markov embedding of the jump process (1.12) (with $\Lambda_t^Z \equiv \Lambda$) and that Z is the jump process associated with the solution ρ.

1.4. Assumptions and main result. The main result of this work concerns the stability of (1.16). We use a notion of stability that is close to that of Brémaud and Massoulié [5] for nonlinear Hawkes processes.

We say that a jump process Z is *stationary* if, for all $\tau > 0$, the time-shifted process $(Z_{t+\tau} - Z_\tau)_{t \geq 0}$ has the same law as $(Z_t - Z_0)_{t \geq 0}$. Then, we say that a solution $\bar{\rho}$ to (1.16) with the $\mathcal{M}_+$-valued random initial condition $\bar{\nu}_0$ is stationary if the associated jump process $\bar{Z}$ is stationary.

Since the noise in (1.16) comes from a Poisson random measure, we can naturally construct a *coupling* of two solutions ρ and $\tilde{\rho}$ to (1.16) (for different, possibly random, initial conditions) on the same probability space, using the same underlying Poisson random measure. Writing Z and $\tilde{Z}$ the jump processes associated with ρ and $\tilde{\rho}$, we define T_c the *coupling time* of Z and $\tilde{Z}$, i.e. the time starting from which Z and $\tilde{Z}$ are identical:

$$(1.17) \qquad T_c := \inf \left\{ \tau \geq 0 : (Z_{t+\tau} - Z_\tau)_{t \geq 0} \equiv (\tilde{Z}_{t+\tau} - \tilde{Z}_\tau)_{t \geq 0} \right\},$$

with the usual convention that $T_c = +\infty$ if Z and $\tilde{Z}$ never couple. In other words, T_c is the time starting from which ρ and $\tilde{\rho}$ have the exact same jump times. By abuse of terminology, we will say that T_c is the coupling time of ρ and $\tilde{\rho}$ although it is in fact the coupling time of the associated jump processes. We can now adapt the definition of stability in variation of [5]:

Definition 1.1 (Stability in variation). *The voltage-structured SPDE* (1.16) *is* stable in variation *if there exists a stationary process* $\{\bar{\rho}, \bar{\nu}_0\}$ *solving* (1.16) *such that for all* $\mathcal{M}_+$*-valued random initial conditions* $\widehat{\nu}_0$*, there exists a coupling of* $\bar{\rho}$ *and* ρ *(a solution to* (1.16) *with initial condition* $\widehat{\nu}_0$*), such that the coupling time* T_c *of* $\bar{\rho}$ *and* ρ *is almost surely finite.*

In modeling terms, the stability in variation implies that for any (random) initial condition $\widehat{\nu}_0$, the population spike train $\dot{Z}_t$ relaxes to a unique stationary process in finite time. More specifically, for any initial condition $\widehat{\nu}_0 \in \mathcal{M}_+$, if we draw $\bar{\nu}_0$ from a stationary distribution and if we simulate the two corresponding processes with the same Poisson noise, they couple in finite time almost surely. In particular, this implies the uniqueness of the stationary distribution.

To prove that (1.16) is stable in variation, we need

Assumption 1.2. $\mu_t \equiv \mu \in \mathbb{R}$.

This just means that the external drive is time-homogeneous and it is a natural assumption to make if we want to show relaxation to a stationary process.

The other important assumption concerns the intensity function f:

Assumption 1.3. f is bounded, i.e. $\|f\|_\infty < \infty$, and $\inf_{u \in \mathbb{R}} f(u) =: f_{\min} > 0$.

A simple example of a function satisfying the assumption is the shifted sigmoid. Note that these bounds do not allow taking an exponential function f (or any unbounded function) nor having an absolute refractory period (short interval of time following a spike during which an neuron cannot spike). In other terms, neurons can not be forced to spike in a finite time interval nor forced to stay silent. Nevertheless, since $\|f\|_\infty$ can be arbitrarily large and $f_{\min}$ can be arbitrarily small, these bounds do not meaningfull alter biological realism.

Finally, to prove that the stationary process exists, we need:

Assumption 1.4. f is differentiable and f' is bounded. Furthermore, there exists a positive constant C such that $|u f'(u)| \leq C$ for all u.

This is a purely technical assumption and is rather innocent since f is anyway bounded.

We can now state our main result:

Theorem 1.5. *Grant Assumptions 1.2 – 1.4. The voltage-structured SPDE* (1.16) *is stable in variation.*

The proof is divided into two parts. In the first part, using Meyn-Tweedie theory [37], we show that the solutions of (1.16) satisfy a certain recurrence property which then allows us to prove that the associated jump processes couple, using methods from [5] for nonlinear Hawkes processes. In the second part, we prove the existence of a non-trivial stationary process solving (1.16).

In simulations, the simplified model with fixed Λ, (1.16), has a qualitatively similar behavior (from the stability point of view) as the original model of [48] where Λ_t^Z has an explicit expression in terms of the past Z (see section 4). Hence, the proof of Theorem 1.5 provides an important understanding of the role of the 'correction term' $\Lambda_t^Z(1 - \dots)$ in the original model (Figure 1c–f).

1.5. Plan of the paper. First, in section 2, we prove the well-posedness of the SPDE (1.16) as a measure-valued piecewise deterministic Markov process. The proof of Theorem 1.5 is then presented in section 3.

In section 4, we present a concise derivation of the finite-size integral equation (1.12) and a simple simulation algorithm is provided in section 5. A general simulation algorithm for

multiple interacting populations of generalized integrate-and-fire neurons can be found in the Appendix.

2. Well-posedness. Although the SPDE (1.16) might look somewhat formal, it can be rigorously formulated in terms of a piecewise deterministic Markov Process (PDMP) taking values in the space $\mathcal{M}_+$ of all positive measures on $\mathbb{R}$. We endow $\mathcal{M}_+$ with the topology of weak convergence, which makes $\mathcal{M}_+$ Polish.

Since Assumptions 1.2 and 1.3 are always imposed in the sequel, we will omit their mention. In particular, we will always assume that f is bounded.

For all $\nu \in \mathcal{M}_+$, let us write $(\mathcal{S}(t)\nu)_{t \geq 0} := (\rho(\cdot,t))_{t \geq 0}$ the solution to the transport equation

$$(2.1) \qquad \partial_t \rho(du,t) - \partial_u \left(\left(\frac{u-\mu}{\tau_\mathrm{m}} \right) \rho(du,t) \right) = -f(u)\rho(du,t), \qquad \forall (u,t) \in \mathbb{R} \times \mathbb{R}_+^*,$$

$$\rho_0 = \nu.$$

With the notation of (1.5), take the flow $\Phi_{s,t}^0$ without exterior input, that is, $a \equiv 0$. Then we have the explicit representation

$$(2.2) \qquad S(t)\nu = \int_\mathbb{R} \delta_{\Phi_{0,t}^0(u)} e^{-\int_0^t f(\Phi_{0,r}^0(u))dr} \nu(du).$$

$(\mathcal{S}(t))_{t \in \mathbb{R}_+}$ can be seen as a sub-stochastic $\mathcal{C}_0$-semigroup of bounded linear operators on $\mathcal{M}_+$. Moreover, we introduce, for any $a \in \mathbb{R}_+$ and any $\nu \in \mathcal{M}_+$, the shifted measure

$$\Delta_a \nu : \mathcal{B}(\mathbb{R}) \ni B \mapsto \nu((B-a)).$$

Putting $\rho_0 = \nu_0$, we can construct a path-wise solution to (1.16) following the procedure:
1. We start from an initial value $\nu_0 \in \mathcal{M}_+$ at time $t = 0$.
2. We consider the counting process

$$Z_t^* = \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \leq N \left[(\mathcal{S}(s)\nu_0)[f] + \Lambda(1 - \|\mathcal{S}(s)\nu_0\|) \right]_+} \pi(ds,dz),$$

   together with its first jump time $\tau^1 := \inf\{t \geq 0 : Z_t^* = 1\}$.
3. We put $\rho_t := \mathcal{S}(t)\nu_0$ for all $t < \tau^1$.
4. At time τ^1, we update

$$(2.3) \qquad \rho_{\tau^1} := \Delta_{\frac{J}{N}} \left(\mathcal{S}(\tau^1)\nu_0 \right) + \frac{1}{N}\delta_0$$

   and we return to step 1. replacing ν_0 by ρ_{τ^1} and time 0 by τ^1.

Remark 2.1. This construction provides indeed a PDMP taking values in $\mathcal{M}_+$; in between the successive jumps of Z_t only the transport equation acts, and we shall show below that only a finite number of jumps occurs within each finite time interval. We have the explicit

representation

$$(2.4a) \qquad \rho_t = \int_{\mathbb{R}} \delta_{\Phi_{0,t}^Z(u)} e^{-\int_0^t f(\Phi_{0,r}^Z(u))dr} \nu_0(du) + \int_{[0,t]} \delta_{\Phi_{s,t}^Z(0)} e^{-\int_s^t f(\Phi_{s,r}^Z(0))dr} dZ_s,$$

$$(2.4b) \qquad Z_t = \frac{1}{N} \int_{[0,t]\times\mathbb{R}_+} \mathbb{1}_{z \leq N[\rho_{t-}[f]+\Lambda(1-\|\rho_{t-}\|)]_+} \pi(ds,dz).$$

In the above formula, the first term on the right hand side of ((2.4)a) corresponds to (2.2), except that we have to replace the null exterior input by Z such that at each jump of Z, the original mass is shifted by J/N, according to the jump term $\Delta_{\frac{J}{N}}$ of (2.3). The second term corresponds to the source term $\frac{1}{N}\delta_0$ which is added at each jump of Z and then transported by $S(t)$.

The above notion of solution is actually equivalent to the notion of a *mild solution* of the SPDE (1.16) driven by Poisson noise (see [24] and [50]). However, since the only underlying noise is Poisson, with finite jump intensity, the notion of a PDMP with values in $\mathcal{M}_+$ seems to be more natural in this context.

Remark 2.2. Using the representation (2.4), we can easily make the link between the SPDE (1.16) and the stochastic integral equation (1.12). Taking the definition of $\bar{A}_t$ in (1.16b), we have

$$\begin{aligned}
\bar{A}_t &= [\rho_t[f] + \Lambda(1-\|\rho_t\|)]_+ \\
&= \Bigg[\int_{\mathbb{R}_+} f(\Phi_{0,t}^Z(u)) e^{-\int_0^t f(\Phi_{0,r}^Z(u))dr} \nu_0(du) + \int_{[0,t]} f(\Phi_{s,t}^Z(0)) e^{-\int_s^t f(\Phi_{s,r}^Z(0))dr} dZ_s \\
&\qquad + \Lambda\left(1 - \int_{\mathbb{R}_+} e^{-\int_0^t f(\Phi_{0,r}^Z(u))dr} \nu_0(du) - \int_{]0,t]} e^{-\int_s^t f(\Phi_{s,r}^Z(0))dr} dZ_s\right)\Bigg]_+
\end{aligned}$$

(using (1.6), (1.8), and (1.10))

$$= \left[H^Z(t) + \int_{[0,t]} \lambda^Z(t|s) S^Z(t|s) dZ_s + \Lambda\left(1 - \widetilde{H}_t^Z - \int_{[0,t]} S^Z(t|s) dZ_s\right)\right]_+,$$

showing that (1.16) and (1.12) are equivalent. Also, since

$$\|\rho_t\| = \int_{\mathbb{R}_+} e^{-\int_0^t f(\Phi_{0,r}^Z(u))dr} \nu_0(du) + \int_{]0,t]} e^{-\int_s^t f(\Phi_{s,r}^Z(0))dr} dZ_s = \widetilde{H}_t^Z + \int_{[0,t]} S^Z(t|s) dZ_s,$$

$\|\rho_t\|$ is equivalent to the neuronal mass M_t defined in (1.14).

In what follows we study *the extended generator* $\mathcal{L}$ of our process, in the sense of Meyn and Tweedie [37]. Extended generators are defined by the pointwise convergence and the fact that a fundamental martingale property reminiscent of the Itô formula is verified. For the convenience of the reader we recall its definition: We set $\mathcal{D}(\mathcal{L})$ the set of all measurable functions $g : \mathcal{M}_+ \to \mathbb{R}$ for which there exists a measurable function $h : \mathcal{M}_+ \to \mathbb{R}$, such that $t \mapsto \mathbb{E}_\nu(h(\rho_t))$ is continuous in 0, and such that $\forall \nu \in \mathcal{M}_+, \forall t \geq 0,$

1. $\mathbb{E}_\nu[g(\rho_t)] - g(\nu) = \mathbb{E}_\nu \int_0^t h(\rho_s) ds$;
2. $\mathbb{E}_\nu[\int_0^t |h(\rho_s)| ds] < \infty$.

In this case, we write $\mathcal{L}g := h$.

On a restricted set of test functions, we can explicitly calculate the extended generator $\mathcal{L}$ of the PDMP described above: For all $\varphi \in \mathcal{C}_b^1(\mathbb{R})$ (bounded and continuously differentiable functions), for all $\nu \in \mathcal{M}_+$ and using the abuse of notation $\varphi(\nu) := \nu[\varphi]$, we have that

$$(2.5) \quad \mathcal{L}\varphi(\nu) = -\int_\mathbb{R} \frac{u - \mu}{\tau_\mathrm{m}} \varphi'(u) \nu(du) - \nu[\varphi f]$$
$$+ N\left[\nu[f] + \Lambda(1 - \|\nu\|)\right]_+ \left(\int_\mathbb{R} \varphi\left(u + \frac{J}{N}\right) \nu(du) + \frac{1}{N}\varphi(0) - \nu[\varphi]\right).$$

We now show that this process is well-defined. For that sake, let us define, for all $K > 0$, the exit time

$$(2.6) \qquad\qquad T^K := \inf\{t \geq 0 : \|\rho_t\| > K\}.$$

Remark 2.3. The T^K are well-defined stopping times since the sets $\{\nu \in \mathcal{M}_+ : \|\nu\| > K\}$ are the pre-image of $]K, +\infty[$ by the linear form $\mathbf{1} : \mathcal{M}_+ \to \mathbb{R}_+$, $\nu \mapsto \nu[\mathbf{1}]$ and we have endowed $\mathcal{M}_+$ with the topology of weak convergence. For a general treatment of the measurability of hitting times, see [1] and in particular Theorem 2.4 of that article.

Up to time T^K, the overall jump intensity of the process is bounded by $\|f\|_\infty K + \Lambda$, such that the procedure described above is well-defined up to the explosion time of the process $\zeta := \lim_{K \to +\infty} T^K$. To show that (1.16) is well-defined on $\mathbb{R}_+$, we need to prove that the PDMP defined above is *non-explosive* in the sense of [37], i.e. $\zeta = +\infty$ a. s. We follow the standard 'drift condition'-based approach of [37]. Writing $V(\nu) := \|\nu\| = \nu[\mathbf{1}]$, $\forall \nu \in \mathcal{M}_+$, we have

Lemma 2.4 (Foster-Lyapunov inequality). *There exist $K^* > 0$, $d > 0$ and $c > 0$ such that*

$$(2.7) \qquad\qquad \forall \nu \in \mathcal{M}_+, \qquad \mathcal{L}V(\nu) \leq d\mathbb{1}_{\|\nu\| \leq K^*} - c(1 + V)(\nu).$$

Proof. Using (2.5) and $V(\nu) = \nu[\mathbf{1}]$, we have $\mathcal{L}V(\nu) = -\nu[f] + \left[\nu[f] + \Lambda(1 - \|\nu\|)\right]_+$.

Two cases arise: either $\left[\nu[f] + \Lambda(1 - \|\nu\|)\right]_+ > 0$, in which case $\mathcal{L}V(\nu) = \Lambda(1 - \|\nu\|) = \Lambda - \Lambda V(\nu)$, or $\left[\nu[f] + \Lambda(1 - \|\nu\|)\right]_+ = 0$, in which case $\mathcal{L}V(\nu) = -\nu[f] \leq -f_{\min} V(\nu)$.

Whence, $\mathcal{L}V(\nu) \leq \Lambda - (f_{\min} \wedge \Lambda)V(\nu)$. We can adapt the constants to obtain (2.7). ∎

Arguing as in Theorem 2.1 of [37], Lemma 2.4 guarantees that the PDMP is non-explosive. Hence, we have proved the well-posedness of (1.16):

Proposition 2.5 (Well-posedness). *For all $\nu_0 \in \mathcal{M}_+$, there exists a $\mathcal{M}_+$-valued path-wise unique solution to (1.16) on $\mathbb{R}_+$.*

3. Stability.

3.1. Coupling. More than non-explosion, the 'drift condition'-based method of [37, 38] allows us to show that the PDMP (1.16) satisfies a certain 'recurrence' property.

For all $K > 0$, let us write the hitting time $t_K := \inf\{t \geq 0 : \|\rho_t\| \leq K\}$ and denote by $\mathbb{E}_{\nu_0}[t_K]$ the expected hitting time of the PDMP (1.16) starting in state $\nu_0 \in \mathcal{M}_+$ at time 0.

Lemma 3.1. *Take the constant K^* of Lemma 2.4. For all $\nu_0 \in \mathcal{M}_+$ such that $\|\nu_0\| > K^*$, $\mathbb{E}_{\nu_0}[t_{K^*}] < +\infty$.*

Proof. The proof is standard but we reproduce it here to highlight the fact that it holds even if the space in which the process evolves is not locally compact.

We use V and the constants of Lemma 2.4. For any $t > 0$ and any $M > K^*$, by Dynkin's formula (see [37]),

$$\mathbb{E}_{\nu_0}[V(\rho_{t \wedge T^M})] = V(\nu_0) + \mathbb{E}_{\nu_0} \int_0^{t \wedge T^M} \mathcal{L}V(\rho_s)ds \leq V(\nu_0) + dt,$$

where T^M is the exit time defined in (2.6) and where d is given in (2.7).

Since $V(\rho_{t \wedge T^M}) \geq M \mathbb{1}_{T_M \leq t}$, this implies

$$\mathbb{P}_{\nu_0}(T_M \leq t) \leq \frac{V(\nu_0) + dt}{M}.$$

Taking $M \to \infty$, by monotone convergence, $\mathbb{P}_{\nu_0}(\zeta \leq t) = 0$, which implies non-explosion.

We now make another use of Dynkin's formula:

$$\mathbb{E}_{\nu_0}[V(\rho_{t \wedge t_{K^*} \wedge T^M})] = V(\nu_0) + \mathbb{E}_{\nu_0} \int_0^{t \wedge t_{K^*} \wedge T^M} \mathcal{L}V(\rho_s)ds$$

$$\leq V(\nu_0) - c\mathbb{E}_{\nu_0} \int_0^{t \wedge t_{K^*} \wedge T^M} (1 + V)(\rho_s)ds.$$

Whence,

$$\mathbb{E}_{\nu_0} \int_0^{t \wedge t_{K^*} \wedge T^M} (1 + V)(\rho_s)ds \leq \frac{V(\nu_0) - \mathbb{E}_{\nu_0}[V(\rho_{t \wedge t_{K^*} \wedge T^M})]}{c} \leq \frac{V(\nu_0) - K^*}{c}.$$

Taking $t, M \to \infty$, we get, by monotone convergence

$$\mathbb{E}_{\nu_0} \int_0^{t_{K^*}} (1 + V)(\rho_s)ds \leq \frac{V(\nu_0) - K^*}{c}.$$

The fact that $\mathbb{E}_{\nu_0}[t_{K^*}] \leq \mathbb{E}_{\nu_0} \int_0^{t_{K^*}} (1 + V)(\rho_s)ds$ concludes the proof. ∎

The definition of stability we use involves the notion of coupling of two processes (see subsection 1.4). For ν_0 and $\tilde{\nu}_0 \in \mathcal{M}_+$, a natural way to couple two processes ρ and $\tilde{\rho}$ following (1.16) with initial condition ν_0 and $\tilde{\nu}_0$ respectively is to construct them with the same Poisson random measure π. With this coupling, the associated jump processes Z and $\tilde{Z}_t$ follow, for

all $t \geq 0$,

$$Z_t := \frac{1}{N} \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \leq N[\rho_s[f] + \Lambda(1 - \|\rho_s\|)]_+} \pi(ds, dz),$$

$$\tilde{Z}_t := \frac{1}{N} \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \leq N[\tilde{\rho}_s[f] + \Lambda(1 - \|\tilde{\rho}_s\|)]_+} \pi(ds, dz).$$

For all $t \geq 0$, we can now introduce the event

$$E_t := \{Z_{t+s} - Z_t = \tilde{Z}_{t+s} - \tilde{Z}_t \text{ for all } s \geq 0\}$$

on which both jump processes couple after time t. With $(\mathcal{F}_t)_{t \geq 0}$ denoting the natural filtration of the coupled process, we have a lower bound on $\mathbb{P}(E_t | \mathcal{F}_t)$:

Lemma 3.2. *For any $K > 0$, there exists a constant $\varepsilon \in]0, 1[$ such that for all $t \geq 0$,*

$$(3.1) \qquad \qquad \mathbb{P}(E_t | \mathcal{F}_t) \geq \varepsilon \mathbb{1}_{\{\|\rho_t\| + \|\tilde{\rho}_t\| \leq K\}}.$$

Proof. We use the shorthand $\bar{A}[\nu] := [\nu[f] + \Lambda(1 - \|\nu\|)]_+$, $\forall \nu \in \mathcal{M}_+$. Fix any $t \geq 0$ such that $\|\rho_t\| + \|\tilde{\rho}_t\| \leq K$. Write $\tau_t^1 := \inf\{s > t : (Z_s - Z_t) + (\tilde{Z}_s - \tilde{Z}_t) \geq 1/N\}$ the next jump after time t. Noticing that for all $t \leq s < \tau_t^1$, $\bar{A}[\rho_s] \vee \bar{A}[\tilde{\rho}_s] \leq \|f\|_\infty K + \Lambda$, we clearly have that $t < \tau_t^1$, that is, there is no accumulation of jumps in finite time.

In what follows, we evaluate the difference $\bar{A}[\rho_s] - \bar{A}[\tilde{\rho}_s]$, for $t \leq s$.

We start by considering the difference $\rho_s[f] - \tilde{\rho}_s[f]$, for all $t \leq s < \tau_t^1$. It is clear that, for all $t \leq s < \tau_t^1$,

$$\rho_s[f] = \int_{\mathbb{R}} \rho_t(du) f(\Phi_{t,s}^0(u)) \exp\left(- \int_t^s f(\Phi_{t,r}^0(u)) dr\right) \leq K \|f\|_\infty e^{-(s-t)f_{\min}},$$

where Φ^0 is the flow of the transport equation (2.1) and where for the inequality, we used the bounds of f given by Assumption 1.3. Consequently, for all $t \leq s < \tau_t^1$, $|\rho_s[f] - \tilde{\rho}_s[f]| \leq 2K \|f\|_\infty e^{-(s-t)f_{\min}}$. Similarly, $|\|\rho_s\| - \|\tilde{\rho}_s\|| \leq 2K e^{-(s-t)f_{\min}}$.

At the jump time τ_t^1, two cases arise:

- τ_t^1 is an <u>asynchronous jump</u>, that is, only one of the two processes, say Z, jumps, in which case ρ is shifted to the right by J/N, and a Dirac mass $\frac{1}{N}\delta_0$ is added (see (2.3)). Then, for all $s \in [\tau_t^1, \tau_t^2[$, where $\tau_t^2 := \inf\{s > \tau_t^1 : (Z_s - Z_{\tau_t^1}) + (\tilde{Z}_s - \tilde{Z}_{\tau_t^1}) \geq 1/N\}$, we have

$$\rho_s[f] = \int_{\mathbb{R}} \rho_{\tau_t^{1-}}(du) f(\Phi_{\tau_t^1, s}^0(u + J/N)) \exp\left(- \int_{\tau_t^1}^s f(\Phi_{\tau_t^1, r}^0(u + J/N)) dr\right)$$
$$+ \frac{1}{N} f(\Phi_{\tau_t^1, s}^0(0)) \exp\left(- \int_{\tau_t^1}^s f(\Phi_{\tau_t^1, r}^0(0)) dr\right),$$

while

$$\tilde{\rho}_s[f] = \int_{\mathbb{R}} \tilde{\rho}_{\tau_t^{1-}}(du) f(\Phi_{\tau_t^1, s}^0(u)) \exp\left(- \int_{\tau_t^1}^s f(\Phi_{\tau_t^1, r}^0(u)) dr\right).$$

As a consequence,

$$|\rho_s[f] - \tilde{\rho}_s[f]| \leq \|f\|_\infty \, e^{-f_{\min}(s-\tau_t^1)}(\|\rho_{\tau_{t-}^1}\| + \|\tilde{\rho}_{\tau_{t-}^1}\|) + \frac{\|f\|_\infty}{N} e^{-f_{\min}(s-\tau_t^1)}$$

$$\leq 2K\|f\|_\infty \, e^{-f_{\min}(s-\tau_t^1)} e^{-f_{\min}(\tau_t^1-t)} + \frac{\|f\|_\infty}{N} e^{-f_{\min}(s-\tau_t^1)}$$

$$= 2K\|f\|_\infty \, e^{-f_{\min}(s-t)} + \frac{\|f\|_\infty}{N} e^{-f_{\min}(s-\tau_t^1)}.$$

- $\underline{\tau_t^1 \text{ is a synchronous jump}}$, in which case we obtain similarly that for all $s \in [\tau_t^1, \tau_t^2[$,

$$|\rho_s[f] - \tilde{\rho}_s[f]| \leq 2K\|f\|_\infty \, e^{-f_{\min}(s-t)}.$$

Similar estimates hold for $|\, \|\rho_s\| - \|\tilde{\rho}_s\| \,|$. Since the function $x \mapsto x_+$ is Lipschitz with Lipschitz constant 1, this implies that

$$|\bar{A}[\rho_s] - \bar{A}[\tilde{\rho}_s]| \leq |\rho_s(f) - \tilde{\rho}_s(f)| + \Lambda |\, \|\rho_s\| - \|\tilde{\rho}_s\| \,|.$$

Working iteratively with respect to the successive jump times $\tau_t^n, n \geq 2$, and using the above arguments, we deduce that for an appropriate constant $C > 0$, for all $t \leq s$,

$$(3.2) \qquad |\bar{A}[\rho_s] - \bar{A}[\tilde{\rho}_s]| \leq C e^{-f_{\min}(s-t)}(\|\rho_t\| + \|\tilde{\rho}_t\|) + C \int_{]t,s]} e^{-f_{\min}(s-r)} dD_r$$

where $(D_s)_{s \geq t}$ is the process counting the asynchronous jumps of Z and $\tilde{Z}$. Notice that $(D_s)_{s \geq t}$ has stochastic intensity $(N|\bar{A}[\rho_s] - \bar{A}[\tilde{\rho}_s]|)_{s \geq t}$. In particular, the above upper bound implies that on $[t, \infty[$, $(D_s)_{s \geq t}$ is stochastically upper bounded by a linear Hawkes process, say $(H_s)_{s \geq t}$, with self-interaction kernel $h(s) = NCe^{-f_{\min}s}$ and with time inhomogeneous baseline rate $s \mapsto NCe^{-f_{\min}(s-t)}(\|\rho_t\| + \|\tilde{\rho}_t\|)$.

The rest of this proof follows the arguments given in the proof of Theorem 2 of [5, p. 1581] together with their Lemma 1. Here are the details of the argument: As a consequence of the above, we obtain the lower bound

$$\mathbb{P}(E_t | \mathcal{F}_t) = \mathbb{P}\left(D([t,\infty[) = 0 \big| \mathcal{F}_t\right) \geq \mathbb{P}\left(H([t,\infty[) = 0 \big| \mathcal{F}_t\right),$$

since D is stochastically upper bounded by N. But by the structure of the Hawkes process,

$$\mathbb{P}\left(H([t,\infty[) = 0 \big| \mathcal{F}_t\right) = \exp\left(-\int_t^\infty NCe^{-f_{\min}(s-t)}(\|\rho_t\| + \|\tilde{\rho}_t\|)ds\right)$$

$$= \exp\left(-NC(\|\rho_t\| + \|\tilde{\rho}_t\|)/f_{\min}\right).$$

Putting $\varepsilon := \exp\left(-2NCK/f_{\min}\right)$ concludes the proof. ∎

Theorem 3.3. *Let ρ and $\tilde{\rho}$ be the coupled processes defined above for initial condition ν_0 and $\tilde{\nu}_0 \in \mathcal{M}_+$, and write $\mathbb{E}_{(\nu_0, \tilde{\nu}_0)}$ for the associated expectation. Then the associated counting processes Z and $\tilde{Z}$ couple a.s. in finite time, i.e.*

$$\mathbb{P}\left(\limsup_{t \to +\infty} \left\{(Z_s)_{s \geq t} \neq (\tilde{Z}_s)_{s \geq t}\right\}\right) = 0.$$

Moreover, the associated coupling time T_c, defined in (1.17) *above, admits exponential moments, that is, there exists a positive constant $\bar{\lambda} > 0$ such that for all initial conditions ν_0 and $\tilde{\nu}_0 \in \mathcal{M}_+$,*

$$(3.3) \qquad \mathbb{E}_{(\nu_0, \tilde{\nu}_0)}[e^{\bar{\lambda} T_c}] < +\infty.$$

Proof. The beginning of the proof of this theorem is similar to the Lemma 5 of [5]. Defining $E_\infty := \cup_{t=0}^{\infty} E_t$, $(\mathbb{E}[\mathbb{1}_{E_\infty}|\mathcal{F}_t])_{t\geq 0}$ is a uniformly integrable martingale and we have $\mathbb{E}[\mathbb{1}_{E_\infty}|\mathcal{F}_t] \to \mathbb{1}_{E_\infty}$ a.s.

However, for all $K > 0$, we have, by Lemma 3.2,

$$\mathbb{E}[\mathbb{1}_{E_\infty}|\mathcal{F}_t] = \mathbb{P}(E_\infty|\mathcal{F}_t) \geq \mathbb{P}(E_t|\mathcal{F}_t) \geq \varepsilon \mathbb{1}_{\{\|\rho_t\|+\|\tilde{\rho}_t\|\leq K\}}, \qquad \forall t \geq 0.$$

We can easily adapt the proofs of Lemma 2.4 and Lemma 3.1 to discrete times $n \in \mathbb{N}$ and show that there exists $K^* > 0$ such that $\mathbb{P}(\limsup_{n\to\infty}\{\|\rho_n\| + \|\tilde{\rho}_n\| \leq K^*\}) = 1$. Hence, $\mathbb{1}_{E_\infty} \geq \varepsilon$ a.s., which in turn implies that $\mathbb{P}(E_\infty) = 1$. Since the event E_∞ is the complement of the event $\limsup_{t\to+\infty}\left\{(Z_s)_{s\geq t} \neq (\tilde{Z}_s)_{s\geq t}\right\}$, this concludes the first part of the proof.

The proof of the existence of exponential moments for the coupling time, which is rather classical, is postponed to Appendix B. ∎

3.2. Existence of the stationary process. We construct a stationary process Z following the lines of [5]. The main idea is to show that a construction on the whole line $\mathbb{R}$, that is, starting from $t = -\infty$ is feasible. If it is so, then intuitively the constructed process is automatically stationary. More precisely, we have the following theorem.

Theorem 3.4. *In addition to the usual assumptions, grant Assumption 1.4. Then there exists a unique stationary process Z solving* (1.16).

Proof. We only need to show that a stationary process Z exists - uniqueness follows then from the coupling property stated in Theorem 3.3 above.

We construct a sequence $Z^{[n]}$ of jump processes in the following way. For any fixed $n \geq 1$, let $(\rho^{[n]}, \tilde{Z}^{[n]})$ be the solution of (1.16) defined on $[-n, \infty[$, starting at time $-n$ from the initial condition $\rho_{-n}^{[n]} = \frac{1}{N}\delta_0$, with

$$\tilde{Z}_t^{[n]} = \frac{1}{N}\int_{[-n,t]\times\mathbb{R}_+} \mathbb{1}_{z\leq N\bar{A}_{s-}^{[n]}} \pi(ds, dz), \quad \text{with } \bar{A}_t^{[n]} := \left[\rho_t^{[n]}[f] + \Lambda(1 - \|\rho_t^{[n]}\|)\right]_+, \quad \forall t \geq -n,$$

and $\tilde{Z}_t^{[n]} \equiv 0$ for all $t \leq -n$.

In order to obtain a standardized sequence of processes, we put

$$Z_t^{[n]} := \tilde{Z}_t^{[n]} - \tilde{Z}_0^{[n]}.$$

In this way, for all n, $Z^{[n]}$ is an element of the Skorokhod space $D(\mathbb{R}, \mathbb{R})$ with $Z_0^{[n]} = 0$. We shall also consider the associated sequence of processes

$$X_s^{[n]} := \rho_s^{[n]}[f] - \Lambda\|\rho_s^{[n]}\|,$$

such that the stochastic intensity of $NZ_s^{[n]}$ is $\lambda^{[n]}(s) := N[X_{s-}^{[n]} + \Lambda]_+$.

Step 1. We first show that the family $(Z^{[n]}, X^{[n]})_{n \geq 1}$ is tight in the Skorokhod space $D(\mathbb{R}, \mathbb{R}^2)$. To do so, we use the criterion of Aldous, see Theorem VI.4.5 of [34]. It is sufficient to prove that

(a) for all $T > 0$, all $\varepsilon > 0$,

$$\lim_{\sigma \downarrow 0} \limsup_{n \to \infty} \sup_{(\tau, \tau') \in P_{\sigma, T}} \mathbb{P}(|Z_{\tau'}^{[n]} - Z_\tau^{[n]}| + |X_{\tau'}^{[n]} - X_\tau^{[n]}| > \varepsilon) = 0,$$

where $P_{\sigma, T}$ is the set of all pairs of stopping times (τ, τ') such that $-T \leq \tau \leq \tau' \leq \tau + \sigma \leq T$ a.s.,

(b) for all $T > 0$, $\lim_{K \uparrow \infty} \sup_n \mathbb{P}(\sup_{-T \leq s \leq T}(|Z_s^{[n]}| + X_s^{[n]}) \geq K) = 0$.

To check (a), observe that,

$$\mathbb{E}[|Z_{\tau'}^{[n]} - Z_\tau^{[n]}|] \leq \frac{1}{N} \mathbb{E} \int_\tau^{\tau + \sigma} \lambda^{[n]}(s) ds \leq \frac{1}{N} \sqrt{2T\sigma} \sqrt{\sup_{-T \leq s \leq T} \mathbb{E}\left[(\lambda^{[n]}(s))^2\right]}.$$

Note that $(\lambda^{[n]}(s))^2 \leq C\|\rho_s^{[n]}\|^2 + C'$, for some constants C, C' independent of n. By similar arguments as in the proof of Lemma 2.4, we have that $W(\nu) := \|\nu\|^2$ satisfies

(3.4) $$\forall \nu \in \mathcal{M}_+, \qquad \mathcal{L}W(\nu) \leq \alpha - \beta W(\nu),$$

for suitable constants $\alpha, \beta > 0$.[2] Then, it is straightforward to show that (3.4) implies

$$\sup_n \sup_{-T \leq s \leq T} \mathbb{E}[W(\rho_s^{[n]})] < \infty,$$

implying (a) for the sequence of processes $Z^{[n]}$.

We now turn to the study of the sequence of processes $X^{[n]}$. We show how to control $\rho^{[n]}[f]$; the control of $\|\rho^{[n]}\|$ is obtained similarly by taking $f \equiv 1$. We fix stopping times $\tau < \tau'$ and consider the increment $\rho_{\tau'}^{[n]}[f] - \rho_\tau^{[n]}[f]$ on the event $Z_{\tau'}^{[n]} - Z_\tau^{[n]} = 0$. On this event,

$$\rho_{\tau'}^{[n]}[f] - \rho_\tau^{[n]}[f] = \int_{\mathbb{R}} \rho_\tau^{[n]}(du) \left(f(\Phi_{\tau, \tau'}^0(u)) \exp\left(-\int_\tau^{\tau'} f(\Phi_{\tau, s}^0(u)) ds\right) - f(u) \right).$$

Then,

$$\left| f(\Phi_{\tau, \tau'}^0(u)) \exp\left(-\int_\tau^{\tau'} f(\Phi_{\tau, s}^0(u)) ds\right) - f(u) \right|$$

$$\leq |f(\Phi_{\tau, \tau'}^0(u)) - f(u)| + \|f\|_\infty \left| \exp\left(-\int_\tau^{\tau'} f(\Phi_{\tau, s}^0(u)) ds\right) - 1 \right|$$

$$\leq |f(\Phi_{\tau, \tau'}^0(u)) - f(u)| + \|f\|_\infty (1 - e^{-\sigma\|f\|_\infty}).$$

---

[2]See Appendix C

Using that $|\Phi^0_{\tau,\tau'}(u) - u| \leq (1 - e^{-\sigma/\tau_{\mathrm{m}}})|u - \mu|$, Taylor's formula implies

$$|f(\Phi^0_{\tau,\tau'}(u)) - f(u)| \leq |f'(\xi)|(1 - e^{-\sigma/\tau_{\mathrm{m}}})|u - \mu|,$$

where $\xi \in [u, \Phi^0_{\tau,\tau'}(u)] \cup [\Phi^0_{\tau,\tau'}(u), u]$.

We first produce an upper bound in the case where $u \geq \mu$ and $\mu \geq 0$. Since $|f'(u)| \leq C/u$ by Assumption 1.4 and since $\xi \geq \Phi^0_{\tau,\tau'}(u)$, we have

(3.5) $$|f(\Phi^0_{\tau,\tau'}(u)) - f(u)| \leq C(1 - e^{-\sigma/\tau_{\mathrm{m}}})C_\sigma,$$

where

$$C_\sigma := \sup_{u \geq \mu} \frac{1}{u e^{-\sigma/\tau_{\mathrm{m}}} + \mu(1 - e^{-\sigma/\tau_{\mathrm{m}}})}(u - \mu).$$

Moreover, it is clear that, for any $\sigma_0 > 0$, $\sup_{\sigma \leq \sigma_0} C_\sigma < \infty$.

If $\mu \leq 0$ and $\mu < u \leq 0$, we use that $f'(\xi)$ is bounded on $[\mu, 0]$ to obtain (3.5). The case $u < \mu$ is treated analogously.

As a consequence, we get the global upper bound (on the event $Z^{[n]}_{\tau'} - Z^{[n]}_\tau = 0$):

$$\left| \rho^{[n]}_{\tau'}[f] - \rho^{[n]}_\tau[f] \right| \leq C(1 - e^{-\kappa\sigma})\|\rho^{[n]}_\tau\|, \quad \text{with } \kappa := \|f\|_\infty \vee 1/\tau_{\mathrm{m}}.$$

We conclude the control of $\rho^{[n]}[f]$, on the event $Z^{[n]}_{\tau'} - Z^{[n]}_\tau = 0$, using the Foster-Lyapunov inequality (Lemma 2.4):

$$\mathbb{E}\|\rho^{[n]}_\tau\| \leq \mathbb{E}\|\rho^{[n]}_0\| + dT, \qquad \text{with } d \text{ from } (2.7),$$

and the fact that $\sup_n \mathbb{E}\|\rho^{[n]}_0\| < \infty$.

To deal with the event $Z^{[n]}_{\tau'} - Z^{[n]}_\tau > 0$, observe that

$$\mathbb{E}\left[\left| \rho^{[n]}_{\tau'}[f] - \rho^{[n]}_\tau[f] \right| \mathbb{1}_{\{Z^{[n]}_{\tau'} - Z^{[n]}_\tau > 0\}} \right] \leq \|f\|_\infty \mathbb{E}\left[\left(\|\rho^{[n]}_{\tau'}\| + \|\rho^{[n]}_\tau\| \right) \mathbb{1}_{\{Z^{[n]}_{\tau'} - Z^{[n]}_\tau > 0\}} \right].$$

Moreover, for any stopping time τ taking values in between $-T$ and T, we have

$$\mathbb{E}\left[\|\rho^{[n]}_\tau\| \mathbb{1}_{\{Z^{[n]}_{\tau'} - Z^{[n]}_\tau > 0\}} \right] \leq \sqrt{\mathbb{E}\|\rho^{[n]}_\tau\|^2}\sqrt{\mathbb{P}(Z^{[n]}_{\tau'} - Z^{[n]}_\tau > 0)}.$$

Using similar arguments as above, but now with the Lyapunov function $W(\nu) = \|\nu\|^2$, we obtain

$$\sup_n \mathbb{E}\|\rho^{[n]}_\tau\|^2 < \infty.$$

Finally, using the already established control over $Z^{[n]}$, we get that

$$\limsup_{\sigma \downarrow 0} \sup_n \mathbb{P}(Z^{[n]}_{\tau'} - Z^{[n]}_\tau > 0) = 0,$$

which concludes the proof of (a).

(b) Let us first observe that $\sup_{-T \leq s \leq T} |Z_s^{[n]}| \leq Z_T^{[n]} - Z_{-T}^{[n]}$, and

$$\sup_{-T \leq s \leq T} |X_s^{[n]}| \leq C \sup_{-T \leq s \leq T} \|\rho_s^{[n]}\| \leq C \left(\|\rho_{-T}^{[n]}\| + Z_T^{[n]} - Z_{-T}^{[n]} \right).$$

We can then conclude using the moment estimates established above.

Step 2. By tightness we can extract a subsequence n_k such that $(Z^{[n_k]}, X^{[n_k]})$ converges, in $D(\mathbb{R}, \mathbb{R}^2)$, to a limit process that we shall denote (Z, X). We now show that Z is necessarily stationary. For that sake, take a test function $\varphi : D(\mathbb{R}, \mathbb{R}) \to \mathbb{R}_+$ which is continuous (with respect to the Skorokhod topology), bounded, and which does only depend on $Z \in D(\mathbb{R}, \mathbb{R})$ within a finite time interval $[a, b] \subset \mathbb{R}_+$. We have to show that for every $t \geq 0$,

$$\mathbb{E}[\varphi(Z)] = \mathbb{E}[\varphi(\theta_t Z)],$$

where $\theta_t Z$ is the shifted counting process defined by $(\theta_t Z)_s = Z_{t+s} - Z_t$, for all $s \geq 0$.

By weak convergence, we have that

$$\mathbb{E}[\varphi(Z)] - \mathbb{E}[\varphi(\theta_t Z)] = \lim_{k \to \infty} \mathbb{E}[\varphi(Z^{[n_k]})] - \mathbb{E}[\varphi(\theta_t Z^{[n_k]})].$$

Now we use the coupling property proven in Theorem 3.3 above. For any fixed k and t we realize $Z^{[n_k]}$ and $\theta_t Z^{[n_k]}$ according to the construction used in the proof of Theorem 3.3.

This means the following. Let $\pi(dt, dz)$ be a Poisson random measure on $\mathbb{R} \times \mathbb{R}_+$ which has intensity $dtdz$ on $\mathbb{R} \times \mathbb{R}_+$. We construct $Z^{[n_k]}$ using the atoms of π within $[-n_k, \infty[\times \mathbb{R}_+$, starting from $\frac{1}{N}\delta_0$ at time $-n_k$. Then we choose, independently of π, a random measure $\tilde{\rho}_{-n_k} \sim \mathcal{L}(\rho_{-n_k+t}^{[n_k]})$. Note that this law does not depend on n_k; it only depends on t. Finally, we realize the process $\theta_t Z^{[n_k]}$ letting it start at time $-n_k$ from the initial condition $\tilde{\rho}_{-n_k}$ and using the same underlying Poisson random measure π. Let $T_{coup}^{n_k}$ be the finite coupling time of the two processes. Notice that once again, $\mathcal{L}(T_{coup}^{n_k})$ does not depend on n_k.

Using this coupling, we obtain

$$\left| \mathbb{E}[\varphi(Z^{[n_k]})] - \mathbb{E}[\varphi(\theta_t Z^{[n_k]})] \right| \leq \|\varphi\|_\infty \mathbb{P}(T_{coup}^{n_k} \geq n_k + a) = \|\varphi\|_\infty \mathbb{P}(T_{coup} > n_k + a) \to 0$$

as $n_k \to \infty$, implying that $\mathbb{E}[\varphi(Z)] - \mathbb{E}[\varphi(\theta_t Z)] = 0$. Since the test functions φ form a separating-class (see Theorem 1.2 in [3, p. 8]), we have that Z and $\theta_t Z$ have the same law, whence stationarity.

Step 3. Now, we verify that the process Z, where Z is taken from the stationary limit process (Z, X) constructed above, is a jump process where jumps of size $1/N$ occur with intensity $\lambda_t := N[X_{t^-} + \Lambda]_+$.

To ease the notation, in what follows, we rename the subsequence n_k by n. Using the Skorokhod representation theorem, we may assume that the above weak convergence is almost sure, for a particular realization of the couples $(Z^{[n]}, X^{[n]})$. Hence, we know that almost surely, $(Z^{[n]}, X^{[n]}) \to (Z, X)$ and $\lambda^{[n]} \to \lambda$. Moreover, let $\bar{Z}$ be the process having intensity λ for the same underlying Poisson random measure as (the realization of) Z. Then, by Fatou's lemma,

for any $t \geq 0$,

$$\mathbb{E}|Z_t - \bar{Z}_t| \leq \liminf_n \mathbb{E}|Z_t^{[n]} - \bar{Z}_t| \leq \frac{1}{N} \liminf_n \mathbb{E} \int_0^t |\lambda^{[n]}(s) - \lambda(s)|ds = 0,$$

where we used the uniform integrability of the $\lambda^{[n]}$, namely that $\sup_n \sup_{s \in [0,t]} \mathbb{E}[\lambda_s^{[n]}] < \infty$. The same argument shows that $\mathbb{E}|Z_t - \bar{Z}_t| = 0$ for all $t \leq 0$. Hence $Z = \bar{Z}$ almost surely, implying that Z has the limit intensity λ.

Step 4. Finally, we show that the limit process Z has the right dynamic, i.e. its intensity λ_t is equal to $\bar{\lambda}_t$ given by

$$(3.6) \quad \bar{\lambda}_t := N \left[\sum_{k:T_k < t} \frac{1}{N} \exp\left(-\int_{T_k}^t f(\Phi_{T_k,s}^Z(0))ds\right) (f(\Phi_{T_k,t}^Z(0)) - \Lambda) + \Lambda \right]_+, \qquad \forall t \in \mathbb{R},$$

where T_k denote the jump times of Z and Φ^Z is given in (1.11).

The goal of this step is to show that $\lambda \equiv \bar{\lambda}$. Fix some time $t \geq 0$ and a truncation level $K > 1$. Since almost surely, Z does not jump at time t nor at time $-K$ for all $K \geq 1$, Proposition VI.2.2.1 of [34] implies that $Z_t^{[n]} - Z_{-K}^{[n]} \to Z_t - Z_{-K}$. Therefore, we may choose n_K be such that $Z_t^{[n]} - Z_{-K}^{[n]} = Z_t - Z_{-K}$ for all $n \geq n_K$. By the continuity properties of the Skorokhod topology, as $n \to \infty$, we have that $T_k^{[n]} \to T_k$ as $n \to \infty$, for all $Z_{-K} \leq k \leq Z_t$ (Proposition VI.2.2.1 of [34]). Hence,

$$\sum_{k:-K \leq T_k^{[n]} < t} \frac{1}{N} \exp\left(-\int_{T_k^{[n]}}^t f\left(\Phi_{T_k^{[n]},s}^{Z^{[n]}}(0)\right) ds\right) \left(f\left(\Phi_{T_k^{[n]},t}^{Z^{[n]}}(0)\right) - \Lambda\right) \to$$

$$\sum_{k:-K \leq T_k < t} \frac{1}{N} \exp\left(-\int_{T_k}^t f(\Phi_{T_k,s}^Z(0))ds\right) (f(\Phi_{T_k,t}^Z(0)) - \Lambda).$$

Notice that the expression on the lhs corresponds to the terms contributing to $X_{t^-}^{[n]}$, issued by jumps happening after time $-K$. Since we know that $X_t^{[n]}$ converges to X_t for almost all t, this implies that for all K,

$$X_{t^-} = \sum_{k:-K \leq T_k < t} \frac{1}{N} \exp\left(-\int_{T_k}^t f(\Phi_{T_k,s}^Z(0))ds\right) (f(\Phi_{T_k,t}^Z(0)) - \Lambda)$$

$$+ \lim_{n \to \infty} \sum_{k:T_k^{[n]} < -K} \frac{1}{N} \exp\left(-\int_{T_k^{[n]}}^t f\left(\Phi_{T_k^{[n]},s}^{Z^{[n]}}(0)\right) ds\right) \left(f\left(\Phi_{T_k^{[n]},t}^{Z^{[n]}}(0)\right) - \Lambda\right),$$

where this last limit is necessarily finite. Letting $K \to \infty$ we deduce that

$$X_{t^-} = \sum_{k:T_k < t} \frac{1}{N} \exp\left(-\int_{T_k}^t f(\Phi_{T_k,s}^Z(0)) ds\right) (f(\Phi_{T_k,t}^Z(0)) - \Lambda)$$
$$+ \lim_{K \to \infty} \lim_{n \to \infty} \sum_{k:T_k^{[n]} < -K} \frac{1}{N} \exp\left(-\int_{T_k^{[n]}}^t f\left(\Phi_{T_k^{[n]},s}^{Z^{[n]}}(0)\right) ds\right) \left(f\left(\Phi_{T_k^{[n]},t}^{Z^{[n]}}(0)\right) - \Lambda\right).$$

Next, we shall prove that

$$(3.7) \quad \lim_{K \to \infty} \lim_{n \to \infty} \sum_{k:T_k^{[n]} < -K} \frac{1}{N} \exp\left(-\int_{T_k^{[n]}}^t f\left(\Phi_{T_k^{[n]},s}^{Z^{[n]}}(0)\right) ds\right) f\left(\Phi_{T_k^{[n]},t}^{Z^{[n]}}(0)\right) = 0 \quad a.s.,$$

a similar argument proving that

$$\lim_{K \to \infty} \lim_{n \to \infty} \sum_{k:T_k^{[n]} < -K} \frac{1}{N} \exp\left(-\int_{T_k^{[n]}}^t f\left(\Phi_{T_k^{[n]},s}^{Z^{[n]}}(0)\right) ds\right) \Lambda = 0 \quad a.s.,$$

to obtain that indeed, $\lambda_t = N[X_{t^-} + \Lambda]_+ = \bar{\lambda}_t$.

Let us now prove (3.7). Using Fatou's lemma, we get

$$(3.8) \quad \mathbb{E} \lim_{K \to \infty} \lim_{n \to \infty} \sum_{k:T_k^{[n]} < -K} \frac{1}{N} \exp\left(-\int_{T_k^{[n]}}^t f\left(\Phi_{T_k^{[n]},s}^{Z^{[n]}}(0)\right) ds\right) f\left(\Phi_{T_k^{[n]},t}^{Z^{[n]}}(0)\right)$$
$$\leq \liminf_{K \to \infty} \liminf_{n \to \infty} \mathbb{E} \sum_{k:T_k^{[n]} < -K} \frac{1}{N} \exp\left(-\int_{T_k^{[n]}}^t f\left(\Phi_{T_k^{[n]},s}^{Z^{[n]}}(0)\right) ds\right) f\left(\Phi_{T_k^{[n]},t}^{Z^{[n]}}(0)\right).$$

Using the same arguments as those leading to (3.2), we have

$$\sum_{k:T_k^{[n]} < -K} \frac{1}{N} \exp\left(-\int_{T_k^{[n]}}^t f\left(\Phi_{T_k^{[n]},s}^{Z^{[n]}}(0)\right) ds\right) f\left(\Phi_{T_k^{[n]},t}^{Z^{[n]}}(0)\right) \leq \|f\|_\infty \|\rho_{-K}^{[n]}\| e^{-\min(f)(t+K)}.$$

Therefore, the rhs of (3.8) is upper bounded by

$$\|f\|_\infty \liminf_{K \to \infty} \liminf_{n \to \infty} \mathbb{E}(\|\rho_{-K}^{[n]}\|) e^{-\min(f)(t+K)} = 0,$$

since $\sup_n \sup_K \mathbb{E}(\|\rho_{-K}^{[n]}\|) < \infty$. This concludes the proof. ∎

Corollary 3.5. *Under the same assumptions as in Theorem 3.4, there exists a unique stationary process $\{\rho, \widehat{\nu}_0\}$ solving (1.16).*

Proof. Taking the process $Z \in D(\mathbb{R}, \mathbb{R})$ constructed in Theorem 3.4 and using the same notations as in (3.6), the stationary process $\{\bar{\rho}, \bar{\nu}_0\}$ corresponding to Z is simply

$$\bar{\nu}_0 = \sum_{T_k \leq 0} \exp\left(-\int_{T_k}^0 f(\Phi_{T_k, s}^Z(0)) ds \right) \frac{1}{N} \delta_{\Phi_{T_k, 0}^Z(0)},$$

and for all $t \geq 0$,

■

$$\bar{\rho}_t = \sum_{T_k \leq t} \exp\left(-\int_{T_k}^t f(\Phi_{T_k, s}^Z(0)) ds \right) \frac{1}{N} \delta_{\Phi_{T_k, t}^Z(0)}.$$

4. Background on the finite-size population equation. In this section, we first present a concise derivation of the stochastic integral equation (1.12), which synthesizes the arguments of the original derivation [48]. Following the integral equation convention [29, 30] and as in [48], we formally put the initial condition at time $-\infty$ and (1.12) reads as follows: for all $t \in \mathbb{R}$,

$$(4.1a) \qquad dZ_t = \frac{1}{N} \pi(dt, [0, N\bar{A}_{t-}]),$$

$$(4.1b) \qquad \bar{A}_t = \left[\int_{]-\infty, t]} \lambda^Z(t|s) S^Z(t|s) dZ_s + \Lambda_t^Z \left(1 - \int_{]-\infty, t]} S^Z(t|s) dZ_s \right) \right]_+,$$

where π is a Poisson random measure on $\mathbb{R} \times \mathbb{R}_+$ having Lebesgue intensity and λ^Z and S^Z are defined by (1.6) with replacement (1.11). Furthermore, the time-dependent modulating factor Λ_t^Z is given by

$$(4.1c) \qquad \Lambda_t^Z = \frac{\int_{]-\infty, t]} \lambda^Z(t|s)\{1 - S^Z(t|s)\} S^Z(t|s) dZ_s}{\int_{]-\infty, t]} \{1 - S^Z(t|s)\} S^Z(t|s) dZ_s}.$$

Note that in the original formulation of the model (see Eqs. (11) and (12) in [48]), the expression for the time-dependent modulating factor Λ_t^Z involved a 'variance function' v. Integrating Eq. (12) in [48] gives $v(t|s) = \{1 - S^Z(t|s)\} S^Z(t|s)\dot{Z}_s$. As a consequence, Eq. (11) in [48] can be written as (4.1c), eliminating v.

To understand the reasoning behind the derivation of (4.1), one needs to keep in mind that the goal is to obtain an intensity-based and history-dependent point process (i.e. that only depends on the past Z) approximating the empirical population activity of the microscopic model (1.1).

Let $(Z_s)_{s<t}$ denote the past population activity. In terms of $(Z_s)_{s<t}$, the stochastic intensity of the empirical population activity (of the microscopic model), at time t, can be expressed as

$$(4.2) \qquad N \int_{]-\infty, t[} \lambda^Z(t|s) \mathfrak{S}^Z(t|s) dZ_s,$$

where, for all past spike times s, $\mathfrak{S}^Z(t|s)$ denotes the 'microscopic survival processes': if there was a spike at time s, $\mathfrak{S}^Z(t|s) = 1$ if the neuron which has fired at time s has not fired a spike in $]s, t[$, and $\mathfrak{S}^Z(t|s) = 0$ if it has. We need to approximate (4.2) by an expression which does not involve the microscopic $\mathfrak{S}^Z(t|s)$, but only the past Z. Writing $\Delta\mathfrak{S}^Z(t|s) := \mathfrak{S}^Z(t|s) - S^Z(t|s)$, we have

$$(4.3) \quad \int_{]-\infty,t[} \lambda^Z(t|s)\mathfrak{S}^Z(t|s)dZ_s = \int_{]-\infty,t[} \lambda^Z(t|s)S^Z(t|s)dZ_s + \int_{]-\infty,t[} \lambda^Z(t|s)\Delta\mathfrak{S}^Z(t|s)dZ_s.$$

Note that since the number of neurons N is strictly preserved (in the microscopic model),

$$(4.4) \quad \int_{]-\infty,t[} \Delta\mathfrak{S}^Z(t|s)dZ_s = 1 - \int_{]-\infty,t[} S^Z(t|s)dZ_s.$$

To replace the microscopic $\Delta\mathfrak{S}^Z(t|s)$ on the RHS of (4.3), we introduce a family of conditionally independent (conditioned on Z) survival processes $\{(\widehat{\mathfrak{S}}^Z(t'|s))_{t'\geq s}\}_s$ – one for each past spike time $s < t$ – defined by

$$\widehat{\mathfrak{S}}^Z(t'|s) = \begin{cases} 1 & \text{if } t' < T_s, \\ 0 & \text{if } t' \geq T_s \end{cases},$$

where $\{T_s\}_{\text{past spike time } s<t}$ are accessory random variables satisfying the following conditions: (i) the variables $\{T_s\}_{\text{past spike time } s<t}$ are conditionally independent given Z and (ii), for all past spike time $s < t$, T_s takes values in $[s, +\infty]$ and satisfies $\mathbb{P}(T_s > t'|Z) = S^Z(t'|s)$, for all $t' \in [s, t[$ (T_s can therefore be interpreted as a 'death' time given by the survival S^Z). Importantly, the processes $\{(\widehat{\mathfrak{S}}^Z(t'|s))_{t'\in[s,t]}\}_s$ are close but not exactly equivalent to the microscopic $\{(\mathfrak{S}^Z(t'|s))_{t'\in[s,t]}\}_s$, e.g. the conservation equation (4.4) does not hold for the processes $\Delta\widehat{\mathfrak{S}}^Z(t|s) := \widehat{\mathfrak{S}}^Z(t|s) - S^Z(t|s)$. However, the conditional independence of the processes $\{(\widehat{\mathfrak{S}}^Z(t'|s))_{t'\in[s,t]}\}_s$ will allow us to close the system of equations (see below) and this is the reason why they are introduced.

We do the approximation

$$(4.5) \quad \int_{]-\infty,t[} \lambda^Z(t|s)\Delta\mathfrak{S}^Z(t|s)dZ_s \approx \Lambda_t^Z \int_{]-\infty,t[} \Delta\mathfrak{S}^Z(t|s)dZ_s,$$

where

$$\Lambda_t^Z := \arg\min_\Lambda \mathbb{E}\left[\left(\int_{]-\infty,t[}\left(\lambda^Z(t|s) - \Lambda\right)\Delta\widehat{\mathfrak{S}}^Z(t|s)dZ_s\right)^2 \middle| Z\right]$$

$$= \arg\min_\Lambda \mathbb{E}\left[\int_{]-\infty,t[}\left(\lambda^Z(t|s) - \Lambda\right)^2 \Delta\widehat{\mathfrak{S}}^Z(t|s)^2 dZ_s \middle| Z\right]$$

$$(4.6) \quad = \arg\min_\Lambda \int_{]-\infty,t[}\left(\lambda^Z(t|s) - \Lambda\right)^2 \mathbb{E}\left[\Delta\widehat{\mathfrak{S}}^Z(t|s)^2 \middle| Z\right]dZ_s.$$

Note that in the definition of Λ_t^Z, (4.6), we have used $\widehat{\mathfrak{S}}^Z(t|s)$ instead of the microscopic $\mathfrak{S}^Z(t|s)$, which would have defined the minimum conditional mean squared error of the approximation (4.5). While this replacement cannot be rigorously justified, it allows us to approximate the conditional mean squared error and, in particular, the position of its minimum. Since, $\mathbb{E}\left[\Delta\widehat{\mathfrak{S}}^Z(t|s)^2\,\middle|\,Z\right] = \{1 - S^Z(t|s)\}S^Z(t|s)$, and taking the derivative with respect to Λ in (4.6), we get

$$\Lambda_t^Z = \frac{\int_{]-\infty,t[}\lambda^Z(t|s)\{1 - S^Z(t|s)\}S^Z(t|s)dZ_s}{\int_{]-\infty,t[}\{1 - S^Z(t|s)\}S^Z(t|s)dZ_s}.$$

We have obtained an approximation of the stochastic intensity (4.2) which only involves the past Z:

$$N\int_{]-\infty,t[}\lambda^Z(t|s)\mathfrak{S}^Z(t|s)dZ_s$$

$$\approx N\left[\int_{]-\infty,t[}\lambda^Z(t|s)S^Z(t|s)dZ_s + \Lambda_t^Z\left(1 - \int_{]-\infty,t[}S^Z(t|s)dZ_s\right)\right]_+.$$

(Taking the positive part on the RHS simply guarantees that the intensity is nonnegative.)

In practice, we can deal with the ill-defined initial condition at time $-\infty$ by assuming that $Z_t = 0$ for all $t < 0$ and $Z_0 = 1$ (all neurons spike at time 0). Consistently, we also put $\Lambda_0^Z = 0$. Then, the model (4.1) can be written

For all $t > 0$,

(4.7a) $$Z_t = 1 + \frac{1}{N}\int_{]0,t]\times\mathbb{R}_+}\mathbb{1}_{z \leq N\bar{A}_{s^-}}\pi(ds, dz),$$

(4.7b) $$\bar{A}_t = \left[\int_{[0,t]}\lambda^Z(t|s)S^Z(t|s)dZ_s + \Lambda_t^Z\left(1 - \int_{[0,t]}S^Z(t|s)dZ_s\right)\right]_+,$$

(4.7c) $$\Lambda_t^Z = \frac{\int_{[0,t]}\lambda^Z(t|s)\{1 - S^Z(t|s)\}S^Z(t|s)dZ_s}{\int_{[0,t]}\{1 - S^Z(t|s)\}S^Z(t|s)dZ_s},$$

with the initial condition $Z_0 = 1$ and $\Lambda_0^Z = 0$. Assuming that the original model (4.1) has the same stability property as the simpler model (1.16), this practical choice of initial condition is acceptable as it will be 'forgotten' after some time.

5. Simulation algorithm. Here, we present a simple simulation algorithm for (4.7). The algorithm presented below can be easily adapted to the more realistic case of multiple interacting populations for generalized integrate-and-fire neurons [48], as we show in Appendix A.

To ease the notation, here, we drop all the superscripts Z. We can rewrite (1.5) and (1.6)

as the solution of a SDE: for any $s > 0$,

(5.1a)
$$\frac{dS(t|s)}{dt} = -\lambda(t|s)S(t|s)$$

(5.1b)
$$du(t|s) = \frac{\mu_t - u(t|s)}{\tau_{\mathrm{m}}}dt + JdZ_t$$

with $\lambda(t|s) = f\big(u(t|s)\big)$ and initial conditions $S(s|s) = 1$ and $u(s|s) = 0$.

Finite history. For all $t \geq 0$, let us define the free membrane potential $h(t)$ as the solution of

(5.2)
$$dh_t = \frac{\mu_t - h_t}{\tau_{\mathrm{m}}}dt + JdZ_t$$

with initial condition $h_0 = 0$ (cf. (5.1b)). It is clear that for fixed $s > 0$, $|u(t|s) - h_t| \to 0$ when $t \to \infty$. In practice, there exists a sufficiently large time $T \gg \tau_{\mathrm{m}}$ such that for $t - s > T$, the initial condition for (5.1b) will be forgotten and the membrane potential $u(t|s)$ with last reset time s can be well approximated by the free membrane potential h_t. We call T the history length. Associated with the free membrane potential is the free hazard rate defined as $\lambda_{free}(t) := f(h_t)$. The free hazard rate can be interpreted as the firing intensity of neurons that have fully recovered from refractoriness because the last spike of those neurons happened before time $t - T$ and thus has been approximately forgotten. For the numerical implementation, it is useful to consider the slightly modified model, in which we use the above approximation, i.e. where $\lambda(t|s)$ is set to $\lambda_{free}(t)$ if $t - s > T$. For the sake of notational simplicity, we will use the same symbols for this approximate model. For $0 < t < T$, there is no difference between the approximate and the original model. Hence, the solution of the approximate model is governed by (4.7) and (5.1). However, for $t > T$, the integrals in (4.7b) and (4.7c) do not need to be evaluated over the whole history from 0 to t but reduce to integrals over $]t - T, t]$:

(5.3a)
$$\bar{A}_t = \left[\int_{]t-T,t]} \lambda(t|s)S(t|s)dZ_s + \lambda_{free}(t)x_t + \Lambda_t \left(1 - \int_{]t-T,t]} S(t|s)dZ_s - x_t \right) \right]_+,$$

(5.3b)
$$\Lambda_t = \frac{\int_{]t-T,t]} \lambda(t|s)\{1 - S(t|s)\}S(t|s)dZ_s + \lambda_{free}(t)z_t}{\int_{]t-T,t]}\{1 - S(t|s)\}S(t|s)dZ_s + z_t}.$$

These expressions depend on the additional variables $x_t := \int_{[0,t-T]} S(t|s)dZ_s$ and $z_t := \int_{[0,t-T]}\{1 - S(t|s)\}S(t|s)dZ_s$ that solve the following SDE's [48]:

(5.4a) $dx_t = -\lambda_{free}(t)x_t dt + S(t|t-T)dZ_{t-T},$ $x_T = 0,$

(5.4b) $dz_t = -2\lambda_{free}(t)z_t dt + \{1 - S(t|t-T)\}S(t|t-T)dZ_{t-T},$ $z_T = 0.$

Time discretization. The model with finite history length (5.3) with the SDE's (5.1) and (5.4) suggests a straightforward update scheme in discrete time. To this end, we consider an equally-spaced partition of the time-axis with mesh Δt and time points $t_{\hat{t}} = \hat{t}\Delta t$, $\hat{t} = 0, 1, 2, \ldots$. Furthermore, we partition the co-moving history frame $]t - T, t]$ in discrete time

points $s_{r,\hat{t}} = (\hat{t} - \mathcal{T} + r)\Delta t$, $r = 1, \ldots, \mathcal{T}$ with $\mathcal{T} = T/\Delta t$. On the discrete time points, we define the following quantities:

$$n_{r,\hat{t}} := Z_{s_{r,\hat{t}}+\Delta t} - Z_{s_{r,\hat{t}}}, \quad S_{r,\hat{t}} := S(\hat{t}\Delta t \,|\, s_{r,\hat{t}}), \quad u_{r,\hat{t}} := u(\hat{t}\Delta t \,|\, s_{r,\hat{t}}),$$

$$P_{r,\hat{t}} := 1 - \exp\left[-\frac{\Delta t}{2}\left(\lambda(\hat{t}\Delta t \,|\, s_{r,\hat{t}}) + \lambda((\hat{t}+1)\Delta t \,|\, s_{r,\hat{t}})\right)\right],$$

$$h_{\hat{t}} := h(\hat{t}\Delta t), \quad x_{\hat{t}} := x(\hat{t}\Delta t), \quad y_{\hat{t}} := y(\hat{t}\Delta t), \quad z_{\hat{t}} := z(\hat{t}\Delta t)$$

$$\bar{P}_{\hat{t}} := 1 - \exp\left[-\frac{\Delta t}{2}\left(\lambda_{free}(\hat{t}\Delta t) + \lambda_{free}((\hat{t}+1)\Delta t)\right)\right].$$

Using these quantities, the mesoscopic model can be simulated with the following update rule [48]: For $r = 1, \ldots, \mathcal{T} - 1$,

(5.5a) $$n_{r,\hat{t}+1} = n_{r+1,\hat{t}}$$

(5.5b) $$S_{r,\hat{t}+1} = \left(1 - P_{r+1,\hat{t}}\right) S_{r+1,\hat{t}}$$

(5.5c) $$u_{r,\hat{t}+1} = u_{r+1,\hat{t}} + \left(\frac{\mu_{\hat{t}\Delta t} - u_{r+1,\hat{t}}}{\tau_{\mathrm{m}}} + J\frac{n_{\mathcal{T},\hat{t}}}{\Delta t}\right)\Delta t$$

(5.5d) $$h_{\hat{t}+1} = h_{\hat{t}} + \left(\frac{\mu_{\hat{t}\Delta t} - h_{\hat{t}}}{\tau_{\mathrm{m}}} + J\frac{n_{\mathcal{T},\hat{t}}}{\Delta t}\right)\Delta t$$

(5.5e) $$x_{\hat{t}+1} = \left(1 - \bar{P}_{\hat{t}}\right) x_{\hat{t}} + S_{1,\hat{t}+1}n_{1,\hat{t}+1}$$

(5.5f) $$z_{\hat{t}+1} = \left(1 - \bar{P}_{\hat{t}}\right)^2 z_{\hat{t}} + \bar{P}_{\hat{t}}x_{\hat{t}} + \left(1 - S_{1,\hat{t}+1}\right) S_{1,\hat{t}+1}n_{1,\hat{t}+1}$$

with boundary conditions $S_{\mathcal{T},\hat{t}} = 1$ and $u_{\mathcal{T},\hat{t}} = 0$ for all $\hat{t} > 0$, and

(5.5g) $$n_{\mathcal{T},\hat{t}+1} = \frac{\xi_{\hat{t}}}{N}, \qquad \xi_{\hat{t}} \sim \mathrm{Binomial}(N, \bar{n}_{\hat{t}}),$$

(5.5h) $$\bar{n}_{\hat{t}} = \bar{P}_{\hat{t}}x_{\hat{t}} + \sum_{r=2}^{\mathcal{T}} P_{r,\hat{t}}S_{r,\hat{t}}n_{r,\hat{t}} + P_{\Lambda,\hat{t}}\left(1 - x_{\hat{t}} - \sum_{r=2}^{\mathcal{T}} S_{r,\hat{t}}n_{r,\hat{t}}\right),$$

(5.5i) $$P_{\Lambda,\hat{t}} = \frac{\bar{P}_{\hat{t}}z_{\hat{t}} + \sum_{r=2}^{\mathcal{T}} P_{r,\hat{t}}\left(1 - S_{r,\hat{t}}\right) S_{r,\hat{t}}n_{r,\hat{t}}}{z_{\hat{t}} + \sum_{r=2}^{\mathcal{T}} \left(1 - S_{r,\hat{t}}\right) S_{r,\hat{t}}n_{r,\hat{t}}}.$$

The independent, identically distributed binomial random variables $\xi_{\hat{t}}^k$ represent the total number of neurons that fire in the time interval $(\hat{t}\Delta t, (\hat{t}+1)\Delta t]$. Therefore, the empirical population activity, (1.2), and the corresponding population rate (intensity) are finally obtained as $A_{\hat{t}\Delta t, \Delta t} = n_{\mathcal{T},\hat{t}+1}/\Delta t$ and $\bar{A}_{\hat{t}\Delta t} = \bar{n}_{\hat{t}}/\Delta t$, respectively. A pseudo-code implementation of the mesoscopic model, (5.5), is given in Algorithm 5.1. A Julia-code implementation of the extended model (Appendix A, Algorithm A.1) is publicly available at the following GitHub link: https://github.com/schwalger/mesodyn-LIF.

Algorithm 5.1 Mesoscopic neuronal population model

Data: External stimulus at grid points $\mu_{\hat{t}\Delta t}$, $\hat{t} = 1, \ldots, t_{sim}$
Result: Population activities $A_{\hat{t}\Delta t, \Delta t}$ and rates $\bar{A}_{\hat{t}\Delta t}$, $\hat{t} = 1, \ldots, t_{sim}$

---

1 $\mathcal{T} = \lfloor 5\tau_{\mathrm{m}}/\Delta t \rfloor + 1$ $x = 0$, $z = 0$, $h = 0$ $n_{\mathcal{T}} = 1, n_{1:\mathcal{T}-1} = 0$ $A_{0,\Delta t} = 1/\Delta t$ $S_{1:\mathcal{T}} = 1$, $u_{1:\mathcal{T}} = 0$
   $\lambda_{free} = f(h)$, $\lambda_{1:\mathcal{T}} = f(h)$;

2 **for** *all times* $\hat{t} = 1, \ldots, t_{sim}$ **do**

3 $\quad$ $h \leftarrow h + [(\mu_{\hat{t}\Delta t} - h)/\tau_{\mathrm{m}} + JA_{(\hat{t}-1)\Delta t, \Delta t}]\Delta t$ $P_\lambda = \texttt{Pfire}(f(h), \lambda_{free})$ $W = P_\lambda x$, $X = x$,
   $\quad$ $Y = P_\lambda z$, $Z = z$ $x \leftarrow x - W$ $z \leftarrow (1 - P_\lambda)^2 z + W$ **for** $r = 2, \ldots, \mathcal{T}$ **do**

4 $\quad\quad$ $u_{r-1} = u_r + [(\mu_{\hat{t}\Delta t} - u_r)/\tau_{\mathrm{m}} + JA_{(\hat{t}-1)\Delta t, \Delta t}]\Delta t$ $P_\lambda, \lambda_{r-1} = \texttt{Pfire}(f(u_{r-1}), \lambda_r)$ $m = S_r n_r$
   $\quad\quad$ $v = (1 - S_r)m$ $W \leftarrow W + P_\lambda m$; $\quad\quad$ // $\mathcal{W} := \int_{[0,t]} \lambda(t|s)S(t|s)dZ_s$

5 $\quad\quad$ $X \leftarrow X + m$; $\quad\quad$ // $\mathcal{X} := \int_{[0,t]} S(t|s)dZ_s$

6 $\quad\quad$ $Y \leftarrow Y + P_\lambda v$; $\quad\quad$ // $\mathcal{Y} := \int_{[0,t]} \lambda(t|s)\{1 - S(t|s)\}S(t|s)dZ_s$

7 $\quad\quad$ $Z \leftarrow Z + v$; $\quad\quad$ // $\mathcal{Z} := \int_{[0,t]}\{1 - S(t|s)\}S(t|s)dZ_s$

8 $\quad\quad$ $S_{r-1} = (1 - P_\lambda)S_r$ $n_{r-1} = n_r$

9 $\quad$ **end**

10 $\quad$ $x \leftarrow x + S_1 n_1$ $z \leftarrow z + (1 - S_1)S_1 n_1$ if $Z > 0$: $P_\Lambda = Y/Z$, else $P_\Lambda = 0$ $\bar{n} = \min(\max(0, W +$
    $\quad$ $P_\Lambda(1 - X)), 1)$; $\quad\quad$ // expected spike count $N\bar{n} = N\bar{A}_t\Delta t$

11 $\quad$ draw $n_{\mathcal{T}} = \text{Binomial}(N, \bar{n})/N$ $\bar{A}_{\hat{t}\Delta t} = \bar{n}/\Delta t$ $A_{\hat{t}\Delta t, \Delta t} = n_{\mathcal{T}}/\Delta t$

12 **end**

Function Pfire(λ, λ_{old})

---

1 $P_\lambda = (\lambda + \lambda_{old})\Delta t/2$;
2 **if** $P_\lambda > 0.01$ **then** $P_\lambda \leftarrow 1 - e^{-P_\lambda}$;
3 **return** P_λ, λ

---

6. Conclusions. We have proven that a simplified version of the model proposed in [48] is well-posed and stable in variation in the sense of Brémaud and Massoulié [5]. The simplified model is a Markov embedding of an intensity-based and history-dependent point process where the history dependence is, loosely speaking, more 'nonlinear' than in nonlinear Hawkes processes (in the sense that the past filtering function is updated at each jump event such that even in the argument of the intensity function $f(\cdot)$, the dependence on the past is not linear any more, that is, not given by convolution over the past events). To deal with this difficulty in the proofs, we combined arguments for Markov processes taking values in the space of positive measures and nonlinear Hawkes processes. From this point of view, the finite-size population equation (1.12) is even more 'nonlinear', which makes its mathematical analysis challenging. The simplified model and the original model of [48] could therefore be seen as examples of general intensity-based and history-dependent point processes, extending nonlinear Hawkes processes. Despite their mathematical complexity, these general point processes are rather practical for applications since they can be efficiently simulated, and, as intensity-based processes, can be easily fitted to empirical data using likelihood-based methods [43]. We hope that this work will stimulate further mathematical research on these general intensity-based

processes, which have already proven to be useful in neuroscience.

Author statement. All authors contributed equally to the present article.

Appendix A. Multi-population model.
The only difference between the neuron model in (1.1) and the Generalized integrate-and-fire model considered in [48] is the addition of a synaptic filtering kernel ϵ and an absolute refractory period $\Delta_{\mathrm{abs}} \geq 0$. Accordingly, (1.1a) is replaced by

$$dU_t^{i,N} = \left[\frac{\mu_t - U_t^{i,N}}{\tau_{\mathrm{m}}} dt - U_{t^-}^{i,N} dZ_t^{i,N} + \left(\frac{J}{N} \sum_{j=1}^{N} \int_{]-\infty,t]} \epsilon(t-s) dZ_s^{j,N} \right) dt \right] \mathbb{1}_{T_t^{i,N} > \Delta_{\mathrm{abs}}},$$

where $T_t^{i,N}$ is an additional "age"-variable defined by the stochastic dynamics $dT_t^{i,N} = dt - T_{t^-}^{i,N} dZ_t^{i,N}$, which clocks the time elapsed since the last spike of neuron i. Then, the definitions for the hazard rate λ and the survival S can be easily adapted replacing Φ in (1.11) by

$$\Phi_{s,t}^z(u) := u e^{-\frac{t-s}{\tau_{\mathrm{m}}}} + \int_s^t e^{-\frac{t-r}{\tau_{\mathrm{m}}}} \left(\frac{\mu_r}{\tau_{\mathrm{m}}} + J \int_{]-\infty,r]} \epsilon(r-s') dz_{s'} \right) dr, \qquad \forall u \in \mathbb{R},$$

and replacing λ in (1.6) by $\lambda^z(t|s) = f(\Phi_{s+\Delta_{\mathrm{abs}},t}^z(0)) \mathbb{1}_{t \geq s+\Delta_{\mathrm{abs}}}$.

As explained in [48], it is straightforward to generalize (4.7) (with the aforementioned extensions) to multiple interacting populations. Importantly, the multi-population model allows to coarse-grain microscopic models of large biological networks of neurons, like a cortical column.

Again, we will henceforth drop the superscripts Z. Let us consider a system of K interacting (homogeneous) populations, each consisting of $N^1, \ldots, N^K$ neurons, with parameters

$$\{N^k, \tau_{\mathrm{m}}^k, \Delta_{\mathrm{abs}}^k, f^k, \epsilon^k, (\mu_t^k)_{t \geq 0}\}_{k=1\ldots,K}$$

and average connectivity matrix $\mathbf{J}$, where J^{kl} is the average connection strength from population l to population k. The multi-population version of (4.7) is

For all $k = 1, \ldots, K$ and $t > 0$,

(A.1a) $$Z_t^k = 1 + \frac{1}{N} \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{z \leq N \bar{A}_{s^-}^k} \pi^k(ds, dz),$$

(A.1b) $$\bar{A}_t^k = \left[\int_{[0,t]} \lambda^k(t|s) S^k(t|s) dZ_s^k + \Lambda_t^k \left(1 - \int_{[0,t]} S^k(t|s) dZ_s^k \right) \right]_+,$$

(A.1c) $$\Lambda_t^k = \frac{\int_{[0,t]} \lambda^k(t|s)\{1 - S^k(t|s)\} S^k(t|s) dZ_s^k}{\int_{[0,t]} \{1 - S^k(t|s)\} S^k(t|s) dZ_s^k},$$

with the initial condition $Z_0^1 = \cdots = Z_0^K = 1$ and $\Lambda_0^1 = \cdots = \Lambda_0^K = 0$, where $\{\pi^k\}_{k=1,\ldots,K}$ are independent Poisson random measures on $\mathbb{R}_+ \times \mathbb{R}_+$ with Lebesgue intensity measure and

(A.2a) $\qquad S^k(t|s) = \exp\left(-\int_s^t \lambda^k(r|s)dr\right),$

$\qquad\qquad \lambda^k(t|s) = f^k(u^k(t|s))\mathbb{1}_{t \geq s+\Delta^k},$

(A.2b) $\qquad u^k(t|s) = \mathbb{1}_{t \geq s+\Delta^k} \int_{s+\Delta^k}^t e^{-\frac{t-r}{\tau_{\mathrm{m}}^k}} \left(\frac{\mu_r^k}{\tau_{\mathrm{m}}^k} + \sum_{l=1}^K J^{kl} \int_{[s,r]} \epsilon^k(r-s')dZ_{s'}^l\right)dr.$

For simplicity, we have presented here a version of the multi-population model without spike-frequency adaptation nor short-term synaptic plasticity but these features can be included [48, 46].

In the following we choose a delayed expontial synaptic filter $\epsilon^k(t) = \frac{1}{\tau_{\mathrm{s}}^k} \exp\left(-\frac{t-d^k}{\tau_{\mathrm{s}}^k}\right)\mathbb{1}_{t \geq d^k}$, where τ_{s}^k is the synaptic decay time constant and $d^k > 0$ denotes the transmission delay associated with the presynaptic population k. This choice allows us to rewrite (A.2a) and (A.2b) as the solution of a SDE (with delay): for any $s > 0$,

$$\frac{dS^k(t|s)}{dt} = -\lambda^k(t|s)S^k(t|s),$$

$$\tau_{\mathrm{m}}^k \frac{du^k(t|s)}{dt} = -u^k(t|s) + \mu_t^k + \tau_{\mathrm{m}}^k \sum_{l=1}^K J^{kl}y_t^l,$$

$$\tau_{\mathrm{s}}^k dy_t^k = -y_t^k dt + dZ_{t-d^k}^k,$$

with initial conditions $S^k(s|s) = 1$, $u^k(s|s) = 0$ and $y_0^k = 0$.

As in the case for a single population (section 5), the infinite history of (A.1) can be approximated by a finite history. The method is completely analogous to that described in section 5 except that now, each population k has its own free membrane potential $h^k(t)$ following

$$\tau_{\mathrm{m}}^k \frac{dh^k(t)}{dt} = -h^k(t) + \mu_t^k + \tau_{\mathrm{m}}^k \sum_{l=1}^K J^{kl}y_t^l,$$

with initial condition $h^k(0) = 0$, and its own history length $T^k \gg \tau_{\mathrm{m}}^k$.

For the discrete time dynamics, being also completely analogous to the single population case, we get the generalized algorithm:

Algorithm A.1 Mesoscopic multi-population model with $\Delta_{abs}^k \geq 0$, $d^k \geq 0$, $\tau_s^k \geq 0$

Data: External stimulus at grid points $\mu_{\hat{t}\Delta t}^k$, $\hat{t} = 1, \ldots, t_{sim}$, $k = 1, \ldots, K$

Result: Population activities $A_{\hat{t}\Delta t, \Delta t}^k$ and rates $\bar{A}_{\hat{t}\Delta t}^k$, $\hat{t} = 1, \ldots, t_{sim}$, $k = 1, \ldots, K$

4 **for** *all populations* $k = 1, \ldots, K$ **do**

5 $\quad$ $\mathcal{T}^k = \lfloor (5\tau_{\mathrm{m}}^k + \Delta_{abs}^k)/\Delta t \rfloor + 1, \hat{\Delta}_{abs}^k = \lfloor \Delta_{abs}^k/\Delta t \rfloor, \hat{d}^k = \lfloor d^k/\Delta t \rfloor$ $x^k = 0$, $y^k = 0$, $z^k = 0$,
$\quad$ $h^k = 0$ $n_{\mathcal{T}^k}^k = 1, n_{1:\mathcal{T}^k-1}^k = 0$ $S_{1:\mathcal{T}^k}^k = 1$, $u_{1:\mathcal{T}^k}^k = 0$ $\lambda_{free}^k = f(h^k)$, $\lambda_{1:\mathcal{T}^k}^k = f(h^k)$;

6 **end**

7 **for** *all times* $\hat{t} = 1, \ldots, t_{sim}$ **do**

8 $\quad$ **for** *all populations* $k = 1, \ldots, K$ **do** $I_{syn}^k = \sum_{l=1}^K J^{kl} y^l$ **for** *all populations* $k = 1, \ldots, K$
$\quad$ **do**

9 $\quad\quad$ $h^k \leftarrow h^k + [(\mu_{\hat{t}\Delta t}^k - h^k)/\tau_{\mathrm{m}}^k + I_{syn}^k]\Delta t$ $P_\lambda, \lambda_{free}^k = \mathtt{Pfire}(f(h^k), \lambda_{free}^k)$ $W = P_\lambda x^k$, $X = x^k$, $Y = P_\lambda z^k$, $Z = z^k$ $x^k \leftarrow x^k - W$ $z^k \leftarrow (1 - P_\lambda)^2 z^k + W$ **for** $r = 2, \ldots, \mathcal{T}^k - \hat{\Delta}_{abs}^k$
$\quad\quad$ **do**

10 $\quad\quad\quad$ $u_{r-1}^k = u_r^k + [(\mu_{\hat{t}\Delta t}^k - u_r^k)/\tau_{\mathrm{m}}^k + I_{syn}^k]\Delta t$ $P_\lambda, \lambda_{r-1}^k = \mathtt{Pfire}(f^k(u_{r-1}^k), \lambda_r^k)$ $m = S_r^k n_r^k$
$\quad\quad\quad$ $v = (1 - S_r^k)m$ $W \leftarrow W + P_\lambda m$; $\quad\quad\quad$ // $\mathcal{W} := \int_{[0,t]} \lambda^k(t|s) S^k(t|s) dZ_s^k$

11 $\quad\quad\quad$ $X \leftarrow X + m$; $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ // $\mathcal{X} := \int_{[0,t]} S^k(t|s) dZ_s^k$

12 $\quad\quad\quad$ $Y \leftarrow Y + P_\lambda v$; $\quad\quad\quad\quad\quad\quad\quad\quad$ // $\mathcal{Y} := \int_{[0,t]} \lambda^k(t|s)\{1 - S^k(t|s)\} S^k(t|s) dZ_s^k$

13 $\quad\quad\quad$ $Z \leftarrow Z + v$; $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ // $\mathcal{Z} := \int_{[0,t]} \{1 - S^k(t|s)\} S^k(t|s) dZ_s^k$

14 $\quad\quad\quad$ $S_{r-1}^k = (1 - P_\lambda)S_r^k$ $n_{r-1}^k = n_r^k$

15 $\quad\quad$ **end**

16 $\quad\quad$ $x^k \leftarrow x^k + S_1^k n_1^k$ $z^k \leftarrow z^k + (1 - S_1^k)S_1^k n_1^k$ **for** *time points in refractory period* $r = \mathcal{T}^k - \hat{\Delta}_{abs}^k + 1, \ldots, \mathcal{T}^k$ **do**

17 $\quad\quad$ $\mid$ $X \leftarrow X + n_r^k$ $n_{r-1}^k = n_r^k$

18 $\quad\quad$ **end**

19 $\quad\quad$ if $Z > 0$: $P_\Lambda = Y/Z$, else $P_\Lambda = 0$ $\bar{n} = \min(\max(0, W + P_\Lambda(1 - X)), 1)$; // expected spike count $N\bar{n} = N\bar{A}_t^k \Delta t$

20 $\quad\quad$ draw $n_{\mathcal{T}^k}^k = \mathrm{Binomial}(N^k, \bar{n})/N^k$ $y^k \leftarrow y^k e^{-\Delta t/\tau_s^k} + \left(1 - e^{-\Delta t/\tau_s^k}\right) n_{\mathcal{T}^k - \hat{d}^k}/\Delta t$ $\bar{A}_t^k = \bar{n}/\Delta t$ $A_t^k = n_{\mathcal{T}^k}^k/\Delta t$

21 $\quad$ **end**

22 **end**

Appendix B. Exponential moments for T_c (end of the proof of Theorem 3.3).

Introducing $\bar{V}(\nu, \tilde{\nu}) := \frac{1}{2}(\|\nu\| + \|\tilde{\nu}\|)$ and $\bar{L}$ the generator of the coupled processes $(\rho_t, \tilde{\rho}_t)$, we obtain as a direct consequence of (2.7) the control

$$\bar{L}\bar{V}(\nu, \tilde{\nu}) \leq \Lambda - (f_{\min} \wedge \Lambda)\bar{V}(\nu, \tilde{\nu}),$$

implying that for any $0 < c < f_{\min} \wedge \Lambda$, there exists a suitable constant K^* such that, with $C := \{\bar{V} \leq K^*\}$,

(B.1) $$\bar{L}\bar{V} \leq -c\bar{V} + \Lambda \mathbb{1}_C.$$

Fix some $\delta > 0$ and introduce the sequence of hitting times

$$T_1(\delta) = \inf\{t \geq \delta : (\rho_t, \tilde{\rho}_t) \in C\}, \quad T_{n+1}(\delta) = \inf\{t \geq T_n(\delta) + \delta : (\rho_t, \tilde{\rho}_t) \in C\}, \quad n \geq 0.$$

Adapting the arguments of Theorem 3.1 of [21] to our frame, we deduce from (B.1) that there exist positive constants $c_1, \bar{\lambda}$ and $c(\delta, \bar{\lambda}), c_2(\delta)$ with

$$\mathbb{E}_{(\nu_0, \tilde{\nu}_0)}[e^{\bar{\lambda} T_1(\delta)}] \leq c_1 \bar{V}(\nu_0, \tilde{\nu}_0) + c_2(\delta)$$

and

$$\mathbb{E}_{(\nu_0, \tilde{\nu}_0)}[e^{\bar{\lambda}(T_{n+1}(\delta) - T_n(\delta))}] \leq c(\delta, \bar{\lambda}) \text{ for all } n \geq 1.$$

Relying on (3.1), we may associate to each $T_n(\delta)$ a Bernoulli random variable $U_n \sim \mathcal{B}(\varepsilon)$, independent of $\mathcal{F}_{T_n(\delta)}$, such that

$$U_n = 1 \text{ implies that at time } T_n(\delta), \text{ the coupling has succeeded.}$$

In particular,

$$T_c \leq \inf\{T_n(\delta) : U_n = 1\}$$

and

$$\mathbb{E}_{(\nu_0, \tilde{\nu}_0)}[e^{\bar{\lambda} T_c}] \leq \sum_{n=1}^{\infty} \mathbb{E}_{(\nu_0, \tilde{\nu}_0)}[e^{\bar{\lambda} T_n(\delta)} \mathbb{1}_{\{U_1 = \ldots = U_{n-1} = 0\}}],$$

for any $\bar{\lambda} > 0$. We are now ready to conclude. Since by monotone convergence,

$$\lim_{\bar{\lambda} \to 0} \mathbb{E}_{(\nu_0, \tilde{\nu}_0)}[e^{\bar{\lambda}(T_{n+1}(\delta) - T_n(\delta))}] = 1,$$

we choose $\lambda_c > 0$ such that for all $0 < \bar{\lambda} < \lambda_c$,

$$\sup_{n \geq 1} \mathbb{E}_{(\nu_0, \tilde{\nu}_0)}[e^{2\bar{\lambda}(T_{n+1}(\delta) - T_n(\delta))}] \cdot (1 - \varepsilon) =: \kappa^2 < 1.$$

Using that, by successive conditioning,

$$\mathbb{E}_{(\nu_0, \tilde{\nu}_0)}[e^{2\bar{\lambda} T_n(\delta)}] \leq \mathbb{E}_{(\nu_0, \tilde{\nu}_0)}[e^{2\bar{\lambda} T_1(\delta)}] \cdot \left(\frac{\kappa^2}{1 - \varepsilon}\right)^{n-1},$$

this implies, using the Cauchy-Schwarz inequality,

$$\mathbb{E}_{(\nu_0, \tilde{\nu}_0)}[e^{\bar{\lambda} T_c}] \leq \sum_{n=1}^{\infty} \mathbb{E}_{(\nu_0, \tilde{\nu}_0)}[e^{\bar{\lambda} T_n(\delta)} \mathbb{1}_{\{U_1 = \ldots = U_{n-1} = 0\}}]$$

$$\leq \sum_{n=1}^{\infty} \sqrt{\mathbb{E}_{(\nu_0, \tilde{\nu}_0)} e^{2\bar{\lambda} T_n(\delta)}} (1 - \varepsilon)^{(n-1)/2} \leq \sqrt{\mathbb{E}_{(\nu_0, \tilde{\nu}_0)} e^{2\bar{\lambda} T_1(\delta)}} \sum_{n=1}^{\infty} \kappa^{n-1} < \infty,$$

which concludes the proof.

Appendix C. Proof of (3.4). Using (2.5), we have $\mathcal{L}W(\nu) = -2\|\nu\|\,\nu[f] + \left[\nu[f] + \Lambda(1 - \|\nu\|)\right]_{+}\left(2\|\nu\| + \frac{1}{N}\right)$. Whenever $\left[\nu[f] + \Lambda(1 - \|\nu\|)\right]_{+} > 0$, this yields, for a suitable constant C,

$$\mathcal{L}W(\nu) \leq -2W(\nu) + C(\|\nu\| + 1),$$

which implies the claim. The easier case $\left[\nu[f] + \Lambda(1 - \|\nu\|)\right]_{+} = 0$ follows simply from the fact that $\nu[f] \geq f_{\min}\|\nu\|$.

Appendix D. Power spectral density.

In Figure 1b, we have characterized the stationary population activity by the power spectral density (PSD) defined for a wide-sense stationary process $X(t)$ and $f > 0$ as [28]

$$(D.1) \qquad \tilde{C}_X(f) := \lim_{T \to \infty} \frac{|\tilde{X}_T(f)|^2}{T}, \qquad \tilde{X}_T(f) := \int_0^T e^{-2\pi i f t} X(t)\,dt.$$

For the mesoscopic model, we estimated the PSD from the simulated, empirical population activity $\hat{A}_{t,\mathfrak{h}}^N(t)$, (1.2) with $\mathfrak{h} = 0.001$ s, using the averaged periodogram (Bartlett's method without windowing). Specifically, for the PSD shown in Figure 1, we segmented a 50 s-long realisation of the empirical population activity (sampled with time step $\mathfrak{h} = 0.001$ s) into 50 non-overlapping segments of length $T = 1$ s, computed the squared absolute values of the fast Fourier transform for each segment, divided the result by T (as in (D.1)) and averaged the resulting periodograms over all 50 segments.

For the microscopic model with $J = 0$ (as in Figure 1), the neuronal population consists of N independent renewal processes generated by the LIF model with escape noise. Therefore, the PSD of $A_{t,\mathfrak{h}}^N(t)$ in the limit $\mathfrak{h} \to 0$ is well-known from the renewal formula [49, 18]

$$(D.2) \qquad \tilde{C}_A(f) = \frac{r}{N}\frac{1 - |\tilde{P}_{ISI}(f)|^2}{|1 - \tilde{P}_{ISI}(f)|^2}.$$

Here, $\tilde{P}_{ISI}(f) = \int_{\mathbb{R}} P_{ISI}(t)e^{-2\pi i f t}\,dt$ is the Fourier transform of the interspike-interval density of single neurons $P_{ISI}(t) = \lambda^0(t|0)S^0(t|0)\mathbb{1}_{t \geq 0}$ and $r = \left[\int_0^\infty S^0(t|0)\,dt\right]^{-1}$ is their firing rate. In Figure 1, these quantities were calculated numerically.

REFERENCES

[1] R. F. BASS, *The measurability of hitting times*, Electron. Commun. Probab., 15 (2010), pp. 99–105.

[2] Y. N. BILLEH, B. CAI, S. L. GRATIY, K. DAI, R. IYER, N. W. GOUWENS, R. ABBASI-ASL, X. JIA, J. H. SIEGLE, S. R. OLSEN, ET AL., *Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex*, Neuron, 106 (2020), pp. 388–403.

[3] P. BILLINGSLEY, *Convergence of probability measures*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, second ed., 1999.

[4] M. BREAKSPEAR, *Dynamic models of large-scale brain activity*, Nat. Neurosci., 20 (2017), pp. 340–352.

[5] P. Brémaud and L. Massoulié, *Stability of nonlinear Hawkes processes*, Ann. Probab., 24 (1996), pp. 1563–1588.

[6] P. Brémaud and L. Massoulié, *Hawkes branching point processes without ancestors*, J. Appl. Probab., 38 (2001), pp. 122–135.

[7] N. Brunel and V. Hakim, *Fast global oscillations in networks of integrate-and-fire neurons with low firing rates*, Neural Comput., 11 (1999), pp. 1621–1671.

[8] N. Cain, R. Iyer, C. Koch, and S. Mihalas, *The computational properties of a simplified cortical column model*, PLoS Comput. Biol., 12 (2016), p. e1005045.

[9] J. Chevallier, *Fluctuations for mean-field interacting age-dependent Hawkes processes*, Electron. J. Probab., 22 (2017), p. 49.

[10] J. Chevallier, *Mean-field limit of generalized Hawkes processes*, Stochastic Process. Appl., 127 (2017), pp. 3870–3912.

[11] J. Chevallier, A. Melnykova, and I. Tubikanec, *Theoretical analysis and simulation methods for Hawkes processes and their diffusion approximation*, arXiv preprint arXiv:2003.10710, (2020).

[12] A. V. Chizhov and L. J. Graham, *Population model of hippocampal pyramidal neurons, linking a refractory density approach to conductance-based neurons*, Phys. Rev. E, 75 (2007), pp. 011924, 14.

[13] Q. Cormier, E. Tanré, and R. Veltz, *Long time behavior of a mean-field model of interacting neurons*, Stochastic Process. Appl., 130 (2020), pp. 2553–2595.

[14] Q. Cormier, E. Tanré, and R. Veltz, *Hopf bifurcation in a mean-field model of spiking neurons*, Electronic Journal of Probability, 26 (2021), pp. 1–40.

[15] M. Costa, C. Graham, L. Marsalle, and V. C. Tran, *Renewal in Hawkes processes with self-excitation and inhibition*, Adv. in Appl. Probab., 52 (2020), pp. 879–915.

[16] A. De Masi, A. Galves, E. Löcherbach, and E. Presutti, *Hydrodynamic limit for interacting neurons*, J. Stat. Phys., 158 (2015), pp. 866–902.

[17] G. Deco, V. K. Jirsa, P. A. Robinson, M. Breakspear, and K. Friston, *The dynamic brain: from spiking neurons to neural masses and cortical fields*, PLoS Comput. Biol., 4 (2008), p. e1000092.

[18] M. Deger, T. Schwalger, R. Naud, and W. Gerstner, *Fluctuations and information filtering in coupled populations of spiking neurons with adaptation*, Phys. Rev. E, 90 (2014), p. 062704.

[19] S. Delattre, N. Fournier, and M. Hoffmann, *Hawkes processes on large networks*, Ann. Appl. Probab., 26 (2016), pp. 216–261.

[20] S. Ditlevsen and E. Löcherbach, *Multi-class oscillating systems of interacting neurons*, Stochastic Process. Appl., 127 (2017), pp. 1840–1869.

[21] R. Douc, G. Fort, and A. Guillin, *Subgeometric rates of convergence of f-ergodic strong Markov processes*, Stochastic Processes and their Applications, 119 (2009), pp. 897–923.

[22] G. Dumont, J. Henry, and C. O. Tarniceriu, *Noisy threshold in neuronal models: connections with the noisy leaky integrate-and-fire model*, J. Math. Biol., 73 (2016), pp. 1413–1436.

[23] G. Dumont, A. Payeur, and A. Longtin, *A stochastic-field description of finite-size spiking neural networks*, PLoS Comput. Biol., 13 (2017), p. e1005691.

[24] N. Fournier, *Malliavin calculus for parabolic SPDEs with jumps*, Stochastic Process. Appl., 87 (2000), pp. 115–147.

[25] N. Fournier and E. Löcherbach, *On a toy model of interacting neurons*, Ann. Inst. Henri Poincaré Probab. Stat., 52 (2016), pp. 1844–1876.

[26] K. Friston, K. H. Preller, C. Mathys, H. Cagnan, J. Heinzle, A. Razi, and P. Zeidman, *Dynamic causal modelling revisited*, Neuroimage, 199 (2019), pp. 730–744.

[27] A. Galves and E. Löcherbach, *Modeling networks of spiking neurons as interacting processes with memory of variable length*, J. SFdS, 157 (2016), pp. 17–32.

[28] C. W. Gardiner, *Handbook of Stochastic Methods*, Springer-Verlag, Berlin, 1985.

[29] W. Gerstner, *Time structure of the activity in neural network models*, Phys. Rev. E, 51 (1995), p. 738.

[30] W. Gerstner, *Population dynamics of spiking neurons: fast transients, asynchronous states, and locking*, Neural Comput., 12 (2000), pp. 43–89.

[31] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal dynamics: From single neurons to networks and models of cognition*, Cambridge University Press, 2014.

[32] K. D. Harris and G. M. Shepherd, *The neocortical circuit: themes and variations*, Nat. Neurosci., 18 (2015), pp. 170–181.

[33] S. HEESEN AND W. STANNAT, *Fluctuation limits for mean-field interacting nonlinear Hawkes processes*, Stochastic Process. Appl., 139 (2021), pp. 280–297.

[34] J. JACOD AND A. N. SHIRYAEV, *Limit theorems for stochastic processes*, vol. 288 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, second ed., 2003.

[35] B. H. JANSEN AND V. G. RIT, *Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns*, Biol. Cybern., 73 (1995), pp. 357–366.

[36] S. LEFORT, C. TOMM, J.-C. FLOYD SARRIA, AND C. C. PETERSEN, *The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex*, Neuron, 61 (2009), pp. 301–316.

[37] S. MEYN AND R. L. TWEEDIE, *Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes*, Adv. in Appl. Probab., 25 (1993), pp. 518–548.

[38] S. MEYN AND R. L. TWEEDIE, *Markov chains and stochastic stability*, Cambridge University Press, Cambridge, second ed., 2009. With a prologue by Peter W. Glynn.

[39] K. PAKDAMAN, B. PERTHAME, AND D. SALORT, *Dynamics of a structured neuron population*, Nonlinearity, 23 (2010), pp. 55–75.

[40] B. PIETRAS, N. GALLICE, AND T. SCHWALGER, *Low-dimensional firing-rate dynamics for populations of renewal-type spiking neurons*, Phys. Rev. E, 102 (2020), pp. 022407, 23.

[41] T. C. POTJANS AND M. DIESMANN, *The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model*, Cereb. Cortex, 24 (2014), pp. 785–806.

[42] C. POZZORINI, S. MENSI, O. HAGENS, R. NAUD, C. KOCH, AND W. GERSTNER, *Automated high-throughput characterization of single neurons by means of simplified spiking models*, PLoS Comput. Biol., 11 (2015), p. e1004275.

[43] A. RENÉ, A. LONGTIN, AND J. H. MACKE, *Inference of a mesoscopic population model from population spike trains*, Neural Comput., 32 (2020), pp. 1448–1498.

[44] P. SANZ-LEON, S. A. KNOCK, A. SPIEGLER, AND V. K. JIRSA, *Mathematical framework for large-scale brain network modeling in the virtual brain*, Neuroimage, 111 (2015), pp. 385–430.

[45] M. SCHMIDT, R. BAKKER, K. SHEN, G. BEZGIN, M. DIESMANN, AND S. J. VAN ALBADA, *A multi-scale layer-resolved spiking network model of resting-state dynamics in macaque visual cortical areas*, PLoS Comput. Biol., 14 (2018), p. e1006359.

[46] V. SCHMUTZ, W. GERSTNER, AND T. SCHWALGER, *Mesoscopic population equations for spiking neural networks with synaptic short-term plasticity*, J. Math. Neurosci., 10 (2020), pp. Paper No. 5, 32.

[47] T. SCHWALGER AND A. V. CHIZHOV, *Mind the last spike—firing rate models for mesoscopic populations of spiking neurons*, Curr. Opin. Neurobiol., 58 (2019), pp. 155–166.

[48] T. SCHWALGER, M. DEGER, AND W. GERSTNER, *Towards a theory of cortical columns: From spiking neurons to interacting neural populations of finite size*, PLoS Comput. Biol., 13 (2017), p. e1005507.

[49] R. L. STRATONOVICH, *Topics in the Theory of Random Noise*, vol. 1, Gordon and Breach, New York, 1967.

[50] J. B. WALSH, *An introduction to stochastic partial differential equations*, in École d'été de probabilités de Saint-Flour, XIV—1984, vol. 1180 of Lecture Notes in Math., Springer, Berlin, 1986, pp. 265–439.

[51] S. WANG, V. SCHMUTZ, G. BELLEC, AND W. GERSTNER, *Mesoscopic modeling of hidden spiking neurons*, arXiv preprint arXiv:2205.13493, (2022).

[52] H. R. WILSON AND J. D. COWAN, *Excitatory and inhibitory interactions in localized populations of model neurons*, Biophys. J., 12 (1972), pp. 1–24.

5 Mesoscopic modeling of hidden spiking neurons

Authors: Shuqi Wang*, Valentin Schmutz*, Guillaume Bellec and Wulfram Gerstner

Contribution: Shuqi Wang and I contributed equally to this work. Shuqi Wang, who completed her Master thesis under my supervision, wrote all the code, did all the numerical simulations and produced all the figures. I played a major role in the conceptualization and in the theoretical aspects of this work; I was also the main writer of the manuscript.

Mesoscopic modeling of hidden spiking neurons

Shuqi Wang*, **Valentin Schmutz***, **Guillaume Bellec, Wulfram Gerstner**
Laboratory of Computational Neuroscience
École polytechnique fédérale de Lausanne (EPFL)
`first.lastname@epfl.ch`

Abstract

Can we use spiking neural networks (SNN) as generative models of multi-neuronal recordings, while taking into account that most neurons are unobserved? Modeling the unobserved neurons with large pools of hidden spiking neurons leads to severely underconstrained problems that are hard to tackle with maximum likelihood estimation. In this work, we use coarse-graining and mean-field approximations to derive a bottom-up, neuronally-grounded latent variable model (neuLVM), where the activity of the unobserved neurons is reduced to a low-dimensional mesoscopic description. In contrast to previous latent variable models, neuLVM can be explicitly mapped to a recurrent, multi-population SNN, giving it a transparent biological interpretation. We show, on synthetic spike trains, that a few observed neurons are sufficient for neuLVM to perform efficient model inversion of large SNNs, in the sense that it can recover connectivity parameters, infer single-trial latent population activity, reproduce ongoing metastable dynamics, and generalize when subjected to perturbations mimicking optogenetic stimulation.

1   Introduction

The progress of large-scale electrophysiological recording techniques [1] begs the following question: can we reverse engineer the probed neural microcircuit from the recorded data? If so, should we try to design large spiking neural networks (SNN), representing the whole microcircuit, capable of generating the recorded spike trains? Such networks would constitute fine-grained mechanistic models and would make *in silico* experiments possible. However appealing this endeavor may appear, it faces a major obstacle – that of unobserved neurons. Indeed, despite the large number of neurons that can be simultaneously recorded, they add up to a tiny fraction of the total number of neurons involved in any given task [2], making the problem largely underdetermined. Training SNNs with large numbers of hidden neurons is challenging because a huge number of possible latent spike patterns result in the same recurrent input to the recorded neurons, making training algorithms nontrivial [3–6].

From the perspective of a single recorded neuron, the spike activity of all the other neurons can be reduced to a single causal variable – the total recurrent input (Figure 1 A). Hence, we argue that fine-grained SNNs are not necessary to model the inputs from hidden neurons but can be replaced by a coarse-grained model of the sea of unobserved neurons. One possible coarse-graining scheme consists in clustering neurons into homogeneous populations with uniform intra- and inter-population connectivity. With the help of mean-field neuronal population equations [7–10], this approach enables the reduction of large SNNs to low-dimensional mesoscopic models composed of neuronal populations interacting with each other [11–13]. Clusters can reflect the presence of different cell-types [11, 14, 15] or groups of highly interconnected excitatory neurons [16–21]. From a computational point of view, coarse-grained SNNs offer biologically plausible implementations of
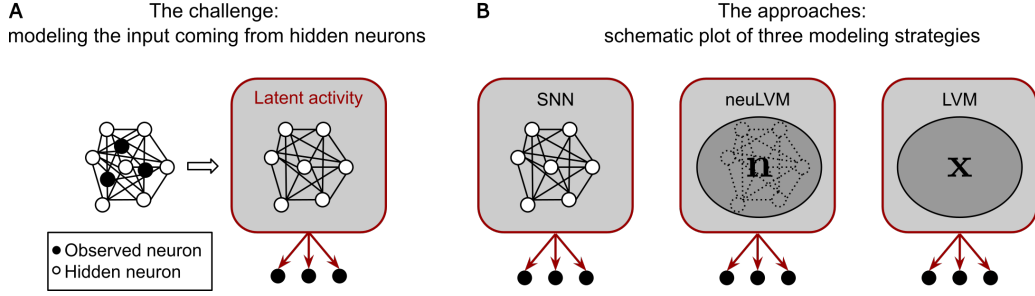
---

*Equal contributions.

Figure 1: **Training SNNs with large number of hidden neurons: challenge and approaches.**
(A) The challenge of modeling the input to the observed neurons (black) coming from the hidden neurons (white) while only a small fraction of neurons is observed. **(B)** Modeling strategies for the input coming from the hidden neurons: SNNs (left) model the fine-grained spike trains of all hidden neurons; neuLVM (middle) uses a mesoscopic description of the population activity, clustering neurons into homogeneous populations; classic LVMs (right) model the latent activity with low-dimensional phenomenological variables (the link to SNNs is lost).

rate coding by ensembles of neurons [22, 23] and 'computation through neural population dynamics' [24].

In this paper, we show that, after clustering the unobserved neurons into several homogeneous populations, the finite-size neuronal population equation of Schwalger et al. [11] can be used to derive a neuronally-grounded latent variable model (neuLVM) where the activity of the unobserved neurons is summarized in a coarse-grained mesoscopic description. The hallmark of neuLVM is its direct correspondence to a multi-population SNN. As a result, both model parameters and latent variables have a transparent biological interpretation: the model is parametrized by single-neuron properties and synaptic connectivity; the latent variables are the summed spiking activities of the neuronal populations. Coarse-graining by clustering therefore turns an underdetermined problem – fitting a SNN with a large number of hidden neurons – into a tractable problem with interpretable solutions.

Switching metastable activity patterns that are not stimulus-locked have attracted a large amount of attention in systems neuroscience [25–27] for their putative role in decision-making [28], attention [29], and sensory processing [30]. Since generative SNN models of these metastable dynamics are available [11, 21, 31, 32], metastable networks constitute ready-to-use testbeds for bottom-up mechanistic latent variable models. Therefore, we propose metastable networks as a new benchmark for mechanistic latent variable models.

2   Relation to prior work

While many latent variable models (LVM), including Poisson Linear Dynamical Systems (PLDS) [33] and Switching Linear Dynamical Systems (SLDS) [34–39], have been designed for inferring low-dimensional population dynamics [40–53], their account of the population activity is a phenomenological one. By contrast, the LVM derived in this work is a true multiscale model, as latent population dynamics directly stems from neuronal dynamics.

Our method builds on René et al. [54], who showed that the mesoscopic model of Schwalger et al. [11] enables the inference of neuronal and connectivity parameters of multi-population SNNs via likelihood-based methods when the mesoscopic population activity is *fully observable*. Here, for the first time, we show that mesoscopic modeling can also be applied to unobserved neurons, relating LVMs to mean-field theories for populations of spiking neurons [7–12]. Our neuLVM approach towards unobserved neurons differs from the Generalized Linear Model (GLM) approach [55–57] (and recent extensions [58–60]), which either neglects unobserved neurons or replaces unobserved neurons by stimulus-locked inputs. Our approach also avoids microscopic simulations of hidden spiking neurons [3, 4, 6], which scale poorly with the number of hidden neurons.

3 Background: mesoscopic modeling of the population activity

Biophysical neuron models can be accurately approximated (neglecting the nonlinearity of dendritic integration) by simple spiking *point* neurons [61–63], which can be efficiently fitted to neural data [64–67]. Stochastic spiking neuron models where the neuron's memory of its own spike history is restricted to the time elapsed since its last spike (its *age*) are said to be of renewal-type.[2] Examples of renewal-type neurons include noisy leaky integrate-and-fire and 'Spike-Response Model 0' neurons [9, 10]. The dynamics of a homogeneous population of interacting renewal-type neurons can be described, in the mean-field limit, by an exact integral equation [7, 9, 10, 12, 69] (see [70–72] for rigorous proofs). In the case of homogeneous but finite populations, Schwalger et al. [11] derived a stochastic integral equation which provides a mesoscopic description (i.e. a description including finite-size fluctuations) of the population dynamics.

For clarity of exposition, in this section and the next, we focus on the case of a single homogeneous population with no external input. All the arguments presented below can be readily generalized to the case of multiple interacting populations with external input (Appendices A B C).

Let us consider a homogeneous SNN of N recurrently connected renewal-type spiking neurons. For T discrete time steps of length Δt, let $\mathbf{y} \in \{0,1\}^{N \times T}$ (a $N \times T$ binary matrix) denote the spike trains generated by the N neurons. The fact that the neurons are of renewal-type implies, by definition, that the probability for neuron i to emit a spike at time t can be written $p(y_t^i = 1|\mathbf{y}_{1:t-1}, \Theta) = \rho_{\theta^i}^{\Delta t}(a^i, \sum_j J^{ij} \mathbf{y}_{1:t-1}^j)$ where a^i is the age of neuron i (i.e. the number of time steps elapsed since the last spike of neuron i), the J^{ij} are the recurrent synaptic weights of the network, and θ^i are the parameters of neuron i. The sum $\sum_j J^{ij} \mathbf{y}_{1:t-1}^j$ represents the past input received by neuron i in all time steps up to $t-1$. The superscript Δt of the function $\rho_{\theta^i}^{\Delta t}$ indicates that we consider here the discrete-time 'escape rate' of the neuron but the transition to continuous time is possible [10]. The explicit expression for $\rho_{\theta^i}^{\Delta t}$ in the case of leaky integrate-and-fire neurons with 'escape noise' (LIF) is given in Appendix A.

A crucial notion in this work is that of 'homogeneous population'. The SNN described above forms a homogeneous population if all the recurrent synaptic weights are identical, that is, $J^{ij} = J/N$ (mean-field approximation) and if all the neurons share the same parameters, that is, $\theta^i = \theta$. In a homogeneous population, all the neurons share the same past input $J\mathbf{n}_{1:t-1}/N$, where $\mathbf{n}_{1:t-1} = (n_1, n_2, \ldots, n_{t-1})$ denotes the total number of spikes in the population in time steps $1, 2, \ldots, t-1$ with $n_{t'} = \sum_{i=1}^N y_{t'}^i$ being the total number of spikes in the population at time t'. Then, for *any* neuron in the population, the probability to emit a spike at time t, given its age a, is

$$p_{t,a}^{\text{fire}} = \rho_\theta^{\Delta t}(a, J\mathbf{n}_{1:t-1}/N). \tag{1}$$

Importantly, Eq. (1) is independent of the identity of the neuron.

In a microscopic description of the spiking activity, the vector $\mathbf{y}_t$ depends nonlinearly on the past $\mathbf{y}_{1:t-1}$. A mesoscopic description aims to reduce the high-dimensional microscopic dynamics to a lower-dimensional dynamical system involving the population activity n_t only (in the case of multiple interacting populations, $\mathbf{n}_t$ is a vector of dimension K equal to the number of populations, Appendix A). While an exact reduction is not possible in general (neuron models being nonlinear), a close approximation in the form of a stochastic integral equation was proposed by Schwalger et al. [11]. In discrete time, the stochastic integral equation reads

$$n_t \sim \text{Binomial}\left(N, \, \bar{n}_t/N\right), \tag{2a}$$

$$\bar{n}_t = \left[\sum_{a \geq 1} p_{t,a}^{\text{fire}} S_{t,a} \, n_{t-a} + \Lambda_t \underbrace{\left(N - \sum_{a \geq 1} S_{t,a} \, n_{t-a} \right)}_{\text{'finite-size correction'}} \right]_+. \tag{2b}$$

The variable $\bar{n}_t$ can be interpreted as the expected number of neurons firing at time t. The survival $S_{t,a} = \prod_{s=0}^{a-1}(1 - p_{t-a+s,s}^{\text{fire}})$ is the probability for a neuron to stay silent between time $t-a$ and

---

[2]Traditional renewal theory in the mathematical literature [68] is restricted to stationary input whereas we use 'renewal-type' in the broader sense that also includes time-dependent input.

3

$t - 1$. The finite-size correction term stabilizes the model by enforcing the approximate neuronal mass conservation $\sum_{a \geq 1} S_{t,a} \, n_{t-a} \approx N$ (see [13] for an in-depth mathematical discussion). The 'modulating factor' Λ_t has an explicit expression [11, 13] in terms of p^{fire}, S and $\mathbf{n}$ (indices are dropped here for simplicity, complete formulas are presented in Appendix A, as well as explanations on how to initialize Eq. (2)). Importantly, for a populations of *interacting* neurons, p^{fire}, S and Λ depend on $\mathbf{n}$, which makes the stochastic equation (2) highly nonlinear. While the mesoscopic model (2) is not mathematically exact, it provides an excellent approximation of the first and second order statistics of the population activity [11], and is much more tractable than the exact 'field' equation [73, 74]. Also, the mesoscopic model (2) can be generalized to the case of non-renewal neurons with spike-frequency adaptation [11] and short-term synaptic plasticity [75].

Formally, Eq. (2) is reminiscent of the Point Process Generalized Linear Model (GLM) [55–57] for single neurons, with the notable difference that Eq. (2) contains additional nonlinearities beyond those of the GLM because p^{fire}, S and Λ all depend on $\mathbf{n}$ (Appendix A). Importantly, Equation (2) readily defines an expression for the probability $p(\mathbf{n}|\Theta)$ [54], where $\Theta = \{J, \theta\}$ denotes the model parameters. Thus, the mesoscopic model (2) allows us to avoid the intractable sum encountered if we naively try to derive $p(\mathbf{n}|\Theta)$ directly from the microscopic description (the intractable sum stems from the fact that the identity of neurons is lost in the observation $n_{t'}$ at each time step t', Figure 1 B).

4 Theoretical result: Neuronally-grounded latent variable model

In this section, we first recall why training SNN with large numbers of hidden neurons via the maximum likelihood estimator is computationally expensive. Then, we show that the mesoscopic description, Eq. (2), allows us to derive a tractable, neuronally-grounded latent variable model, which can be mapped to a multi-population SNNs.

For the sake of simplicity, all the arguments are presented for a single homogeneous population, but the generalization to multiple interacting populations is straightforward (Appendices B and C). Let us assume that we observe, during T time steps, the spike trains of q simultaneously recorded neurons that are part of a homogeneous population of N neurons, with $N > q$. We split the spike trains of the entire population $\mathbf{y} \in \{0, 1\}^{N \times T}$ into the observed spike trains $\mathbf{y^o}$ (q neurons) and hidden spike trains $\mathbf{y^h}$ ($N - q$ neurons). Even for a single population, it is difficult to infer the parameters $\Theta = \{J, \theta\}$ of the SNN from observation $\mathbf{y^o}$ using the maximum likelihood estimator because, in the presence of hidden neurons, the likelihood $\mathcal{L}$ involves a marginalization over the latent spike trains $\mathbf{y^h}$:

$$\mathcal{L} = p(\mathbf{y^o}|\Theta) = \sum_{\mathbf{y^h}} p(\mathbf{y^o}, \mathbf{y^h}|\Theta). \tag{3}$$

While different variants of the Expectation-Maximization (EM) algorithm [76] relying on sampling $\mathbf{y^h}$ have been used to maximize the likelihood [3, 4, 6], these algorithms scale poorly with the number of hidden neurons.

Instead, we exploit the fact that, for a homogeneous population, the fine-grained knowledge of the latent activity $\mathbf{y^h}$ is not necessary since all the observed neurons receive at time t the same input Jn_t, where $n_t = \sum_{i=1}^{N} y_t^i$ is the population activity of Section 3. Hence, we rewrite the likelihood (3) as

$$\mathcal{L} = p(\mathbf{y^o}|\Theta) = \sum_{\mathbf{n}} p(\mathbf{y^o}, \mathbf{n}|\Theta), \tag{4a}$$

where the probability $p(\mathbf{y^o}, \mathbf{n}|\Theta)$ factorizes in T terms of the form

$$p(\mathbf{y_t^o}, n_t|\mathbf{y_{1:t-1}^o}, \mathbf{n}_{1:t-1}, \Theta) = \underbrace{p(\mathbf{y_t^o}|\mathbf{y_{1:t-1}^o}, \mathbf{n}_{1:t-1}, \Theta)}_{\text{given by neuron model, Eq. (1)}} \underbrace{p(n_t|\mathbf{n}_{1:t-1}, \Theta)}_{\text{approx. by meso. model, Eq. (2)}}. \tag{4b}$$

A comparison of Eqs. (3) and (4) shows that the high-dimensional latent activity $\mathbf{y^h}$ has been reduced to a low-dimensional mesoscopic description. Importantly, the q observed spike trains are conditionally independent given the population activity $\mathbf{n}$. While the conditional dependence structure implied by Eq. (4b) is typical of standard latent variable models of multi-neuronal recordings [33, 40, 77, 78], in our approach, the latent variable explicitly represents the population activity

of the generative SNN and the parameters of the model are identical to those of the SNN. As the latent population dynamics directly stems from neuronal dynamics, we call our LVM the neuronally-grounded latent variable model (neuLVM).

The nonlinearity and the non-Markovianity of Eq. (2) prevents us from using previous EM algorithms [33, 40, 77, 78]. Therefore, we fit the neuLVM via the Baum-Viterbi algorithm [79] (also known as Viterbi training or hard EM [80]), which alternates estimation (E) and maximization (M) step

E-step. $\widehat{\mathbf{n}}^n = \mathrm{argmax}_{\mathbf{n}} \log p(\mathbf{y^o}, \mathbf{n}|\widehat{\Theta}^{n-1})$,

M-step. $\widehat{\Theta}^n = \mathrm{argmax}_{\Theta} \log p(\mathbf{y^o}, \widehat{\mathbf{n}}^n|\Theta)$.

The estimated parameters $\widehat{\Theta}$ and the estimated latent population activity $\widehat{\mathbf{n}}$ are the result of many iterations of **E-step** and **M-step** (Appendix C). Note that the computational cost of this algorithm does not depend on the number of hidden neurons (it only depends on the number of populations).[3]

5 Experimental results

5.1 Single homogenous population: SNN with metastable cluster states

Although seemingly simple, homogeneous populations of leaky integrate-and-fire (LIF) neurons without external stimulation are SNNs with a rich repertoire of population dynamics, including asynchronous states, synchronous states, and cluster states [7, 9]. In a m-cluster state (with $m \geq 2$), the population activity oscillates at a frequency m times higher than the average neuronal firing rate: a neuron spikes every m cycles on average; conversely, approximately N/m neurons fire in each cycle (N being to total number of neurons). Cluster states have therefore been described as 'higher harmonics' of the synchronous state (or 1-cluster state) [9, 81–83] where all neurons fire in synchrony.

In this set of experiments, we always consider the same network of 600 LIF neurons (Figure 2 A), where only the connectivity parameter J varies. When initialized at time 0 in the same unstable asynchronous state, the network can spontaneously settle in a m-cluster state, where m depends on the recurrent connectivity parameter J (Figure 2 B): finite-size fluctuations break the symmetry of the asynchronous state and the population splits into m groups of synchronized neurons. The cluster state to which the network converges can be read from the power spectrum of the neuronal spike trains (Figure 2 B) (the fundamental frequency of the m-cluster state is approximately m times lower than that of the 1-cluster state). Generating spike trains for 6 observed neurons (1% of the population), we tested whether neuLVM could recover the connectivity parameter J (neuronal parameters θ were given), for different J's in the 1-, 2-, and 3-cluster states range (Figure 2 C, Table S3). The Pearson correlation between the learned $\widehat{J}$ and the true J was 0.81 with p-value 2.8e−17, showing that, statistically, neuLVM could recover the connectivity parameter of the SNN.

To assess how well neuLVM can infer the latent population activity and how neuLVM compares with the methods assuming full observability (like René et al. [54]), we studied in detail a single trial showing a transition from a metastable 4-cluster state to a 3-cluster state (Figure 2 D,E). To generate this trial, we chose $J = 60.32$ mV and initialized the network in the 4-cluster state. From the spike trains of only two neurons (red stars in Figure 2 D), neuLVM could infer the ground truth population activity $\mathbf{n}^*$ during the 4-cluster state, and during the 3-cluster state, and could approximately detect the transition between the two states (Figure 2 E). While the summed, smoothed spike trains missed two out of four population activity peaks in the 4-cluster state, and one out of three peaks in the 3-cluster state (purple curve in Figure 2 E), the strong inductive biases contained in neuLVM enabled the inference of the 'missing' peaks (blue curve in Figure 2 E). Finally, neuLVM and a method assuming full observability (equivalent to a naive application of René et al. [54]) were compared through their ability to recover the connectivity parameter J, for varying number of observed neurons (Figure 2 F, Table S4). Since a naive application of the method of René et al. [54] does not take into account hidden neurons, it led, as expected, to wildly inaccurate estimate $\widehat{J}$ when the summed spike train was far from the ground truth population activity (which happened when the number of observed neurons was small, see Figure 2 E for an example). In contrast, the neuLVM managed to

---

[3]Our implementation of the algorithm is published online openly. The repository: `TODO`
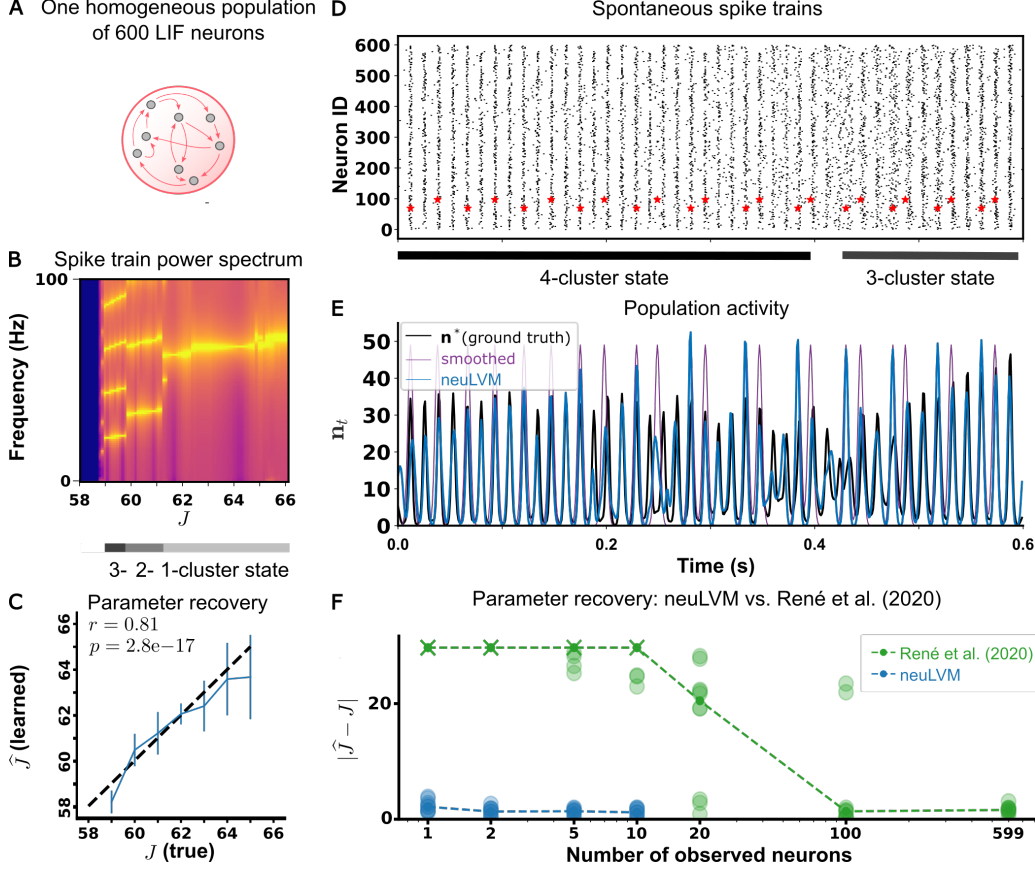
Figure 2: **Single-population SNN with metastable cluster states.** **(A)** Network architecture (for visualization purpose, only a few connections are drawn). **(B)** Spike train power spectrum for different choices of connectivity parameter J. All simulations start from the same unstable asynchronous state. The corresponding cluster states are indicated below. The blue region around $J = 58$ mV indicates the absence of activity. **(C)** Connectivity recovered by the neuLVM $\widehat{J}$ vs ground truth J. The neuLVM was fitted on one-second single-trials recordings of six neurons (1% of the population). For each ground truth J value (seven in total), ten different trials were generated: bars indicate the standard deviations of the recovered $\widehat{J}$. The Pearson correlation coefficient between the recovered $\widehat{J}$ and J is $r = 0.81$ and the associated p-value is $2.8e{-}17$ (see Table S3). **(D)** Spike trains generated by the ground truth SNN for a trial showing a transition from a metastable 4-cluster state to a 3-cluster state. The spike trains of two randomly sampled neurons (red stars) formed the training data (for visualization purpose, only the first 0.6 second of the one-second trial is shown) on which neuLVM was fitted: **(E)** the inferred population activity $\widehat{\mathbf{n}}|\mathbf{y^o}$ is compared to the ground truth $\mathbf{n}^*$ and the summed, smoothed spike trains (Gaussian smoothing window with $\sigma = 1.4$ ms, Appendix D) of the two observed neurons. **(F)** Absolute difference between the recovered $\widehat{J}$ and the ground truth J for the neuLVM algorithm and the method of René et al. (2020) for varying numbers of observed neurons. Using the same trial as in D, for each number of observed neurons, the two methods were tested on 10 different samples of observed neurons (see Table S4). The marker '$\times$' indicates that the difference $|\widehat{J} - J|$ is larger than 30 mV. The median samples are linked with dashed lines to show the trends.

6

recover the connectivity parameter, thanks to the fact that the Baum-Viterbi algorithm of Section 4 also infers the population activity (see Figure 2 E for an example).

5.2 Multiple populations: SNN with metastable point attractors
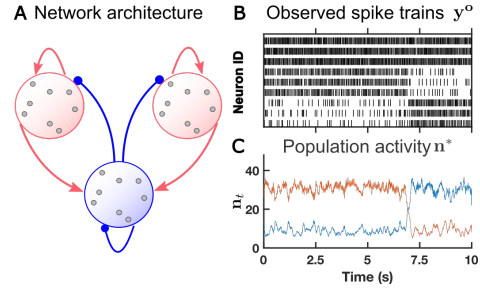
5.2.1 Latent population activity inference and reproduction of metastable dynamics

As a second benchmark, we tested neuLVM on synthetic data generated by three interacting populations with two populations of 400 excitatory LIF neurons and one population of 200 inhibitory neurons (Figure 3 A). Recurrent connections between these population drive winner-take-all dynamics with finite-size fluctuations-induced switches of activity between the two excitatory populations [11, 84] – an example of 'itinerancy between attractor states' [25]. The population activities of this metastable, three-population SNN constitute the ground truth against which different models will be tested.

To build a spiking benchmark dataset, we randomly selected 9 neurons – 3 neurons from each of the three populations – and considered the spike trains of these neurons as the observed data. For simplicity, the correct partitioning of the 9 neurons into 3 groups is given since it can be reliably obtained by k-means clustering [85] using the van-Rossum-Distance [86] between spike trains. The complete dataset the consists of 20 trials of 10 seconds. An example trial is shown in Figure 3 B.

Figure 3: **Network architecture and example trial.** **(A)** Architecture of the three-population, metastable, winner-take-all SNN. **(B)** Example trial from the spiking benchmark dataset: 10 seconds recordings of 9 observed neurons (3 neurons from each of the three populations) and **(C)** corresponding ground truth latent population activity (for the two excitatory populations).

In contrast with the experiments of Section 5.1 where the neuronal parameters were given, here, neuronal and connectivity parameters are not given to neuLVM (see Appendix F). We compared the performance of neuLVM with other generative models of spiking data – PLDS [33], SLDS [39], and GLM [55–57] – on single trials of the spiking benchmark dataset in two ways: (i) we measured the Pearson correlation r between the inferred latent population activity $\widehat{\mathbf{n}}|\mathbf{y^o}$ and the ground truth population activity $\mathbf{n}^*$ (Table 1 first column); (ii) we assessed how well could the fitted models reproduce metastable dynamics by counting the occurrences of stochastic switches in free simulations – or in other words, samples – of the fitted models (Table 1 second column). Tests (i) and (ii) on an example trial are shown in Figure 4.

The Poisson Linear Dynamical Systems approach (PLDS, [33]) assumes that the recorded spikes can be explained by point processes driven by a latent linear dynamical system of low-dimension. The Poisson Switching Linear Dynamical System (SLDS, [34–39]) extends PLDS by allowing the latent variables to switch randomly between two dynamical systems with distinct parameters. We should stress that, in PLDS and SLDS, the latent variables are *phenomenological* representations of neural population activity which have no direct link with the ground truth population activity $\mathbf{n}^*$. In order to still make test (i) possible for PLDS and SLDS, we will consider the best linear transformation of the inferred latent variables which minimizes the mean squared error with the ground truth population activity $\mathbf{n}^*$.

On test (i), neuLVM gave better estimates $\widehat{\mathbf{n}}|\mathbf{y^o}$ of the latent population population activity $\mathbf{n}^*$ (Pearson correlation $r = 0.81$) than the best linear transformation of the latent activity inferred by PLDS and SLDS ($r = 0.69$ and $r = 0.73$ respectively) (Table 1 first column). The GLM approach cannot be included in test (i) since it ignores unobserved neurons. Interestingly, the example trial in Figure 4 A shows the latent population activity $\widehat{\mathbf{n}}|\mathbf{y^o}$ inferred by neuLVM is smoother than the ground truth $\mathbf{n}^*$ before and after the switch (finite-size fluctuations are reduced) but $\widehat{\mathbf{n}}|\mathbf{y^o}$ and $\mathbf{n}^*$ closely match around the time of the switch. In contrast, fluctuations are exaggerated for PLDS and
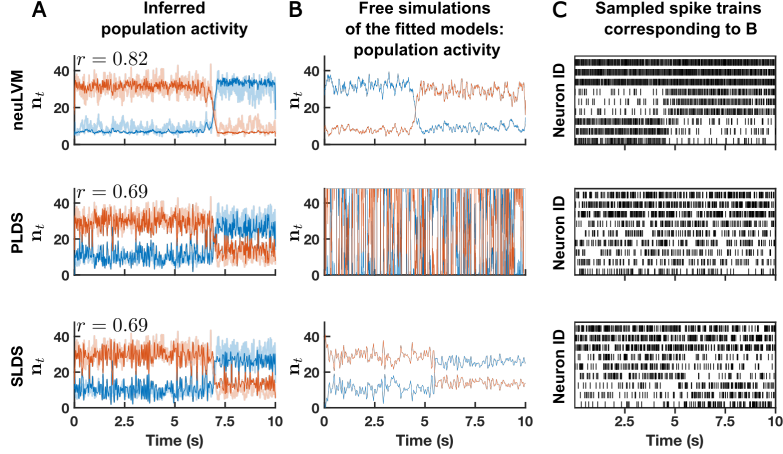
Figure 4: **Three-population SNN with metastable point attractors. (A)** Latent population activity of the two excitatory populations inferred by neuLVM / PLDS / SLDS for one example trial (the same as in Figure 3). The value r is the Pearson correlation coefficient between the inferred $\widehat{n}|\mathbf{y^o}$ and the ground truth $\mathbf{n^*}$ population activities. **(B-C)** Examples of free simulations of the fitted neuLVM / PLDS / SLDS.

Table 1: Model performance summary (corresponding to Figure 4).

Models	Pearson correlation r between $\widehat{n}|\mathbf{y^o}$ and $\mathbf{n^*}$	Number of switches during 100 seconds free simulations of the fitted models (10.3 ± 2.7 for the ground truth SNN)
neuLVM	0.81 ± 0.02	7.8 ± 4.1
PLDS	(0.68 ± 0.11)	not visible
SLDS	(0.73 ± 0.02)	11.9 ± 8.9
GLM	-	not visible

Mean and ($\pm$) standard deviation were computed over 20 different trials. Parentheses for PLDS and SLDS indicate that these results are for the best linear transformation of the inferred latent variables.

SLDS. The population activity estimated by simply summing and smoothing the observed spike trains (Appendix D) is shown in Figure S6.

On test (ii), neuLVM, fitted on a single trial of 10 seconds, was able to reproduce stochastic switches similar to that of the ground truth SNN (Table 1 second column): free simulations of the fitted neuLVM showed 7.8 switches in 100 seconds on average (10.3 switches on average for the ground truth SNN). To make sure that stochastic switches were the result of parameter learning via the Baum-Viterbi algorithm, we verified that, before learning, neuLVM did not show any metastable dynamics (Figure S7). Examples of simulated trial are shown in Figure 4 B. PLDS failed to reproduce stochastic switches, which is not surprising since winner-take-all dynamics are typically nonlinear. SLDS could reproduce stochastic switches at the correct mean frequency (11.9 instead of the ground truth 10.3), but the standard deviation of the simulated switch count, 8.9 (2.7 for the ground truth SNN), indicates that a single 10 seconds trial was probable not sufficient for SLDS to learn switching probabilities reliably. Finally, neuronal stochasticity and small network size (9 neurons) did not allow GLM to produce stochastic switches, even when the training trial was prolonged to 500 seconds.

Taken together, only neuLVM could infer the latent population activity and reliably learn the metastable dynamics on single trials of 10 seconds, demonstrating the effectiveness of its neuronally-grounded inductive biases. Of course, these results do not guarantee that the inductive biases of neuLVM would be effective on real data, since real data is most certainly out-of-distribution for neuLVM. While applications on real data are beyond the scope of this paper, in Appendix F, we show that neuLVM is robust, to a certain extent, to within-population heterogeneity and out-of-distribution data.
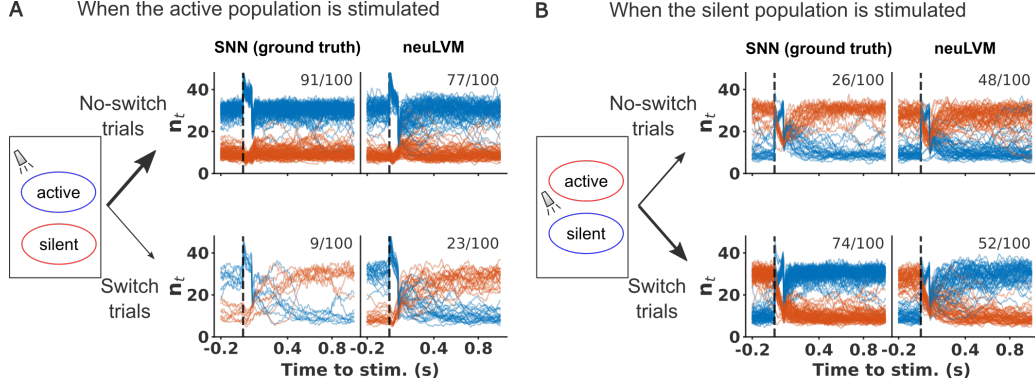
8

Figure 5: **Network responses to perturbations mimicking optogenetic stimulation. (A)** Activities of the excitatory populations when the active population is stimulated (100 trials, ratios indicate the number of No-switch or Switch trials). **(B)** Same as A but the silent population is stimulated.

5.2.2 Generalization: towards experimental predictions with neuronally-grounded modeling

Bottom-up, mechanistic models allow us to perform *in silico* experiments and generate predictions about neural microcircuits, which can then be tested experimentally. So we wondered: can neuLVM, fitted on a single trial of spontaneous activity (like in Section 5.2.1), predict the response of the SNN when an external perturbation is applied? As a preliminary step in that direction, we tested whether an external stimulation of the fitted model would generate the same response as that of the microscopic SNN when subjected to the same perturbation.

Using the same multi-population network as in Section 5.2 (Figure 3 A) and neuLVM fitted on a single trial of spontaneous activity (Figure 3 B), we compared the response of the ground truth SNN with that of neuLVM when one of the populations was stimulated by a current pulse of 4 ms mimicking the stimulation of a optogenetically modified population by a short light pulse. We simulated 100 trials where the momentarily active excitatory population was stimulated, and 100 where the momentarily silent excitatory population was stimulated (Figure 5 A and B respectively). Each stimulation led to two possible outcomes: stimulation could trigger a state switch (Switch trials) or no state switch (No-switch trials). In both the ground truth SNN and the fitted neuLVM, we found that stimulating the silent population triggered more frequent state switches (Figure 5 B) than stimulating the active population (Figure 5 A). Moreover, in both the ground truth and the fitted neuLVM, we could induce 'excitatory rebound' switches by stimulating the active population (Figure 5 A, lower half).

6 Discussion

Understanding the neural dynamics underlying computation in the brain is one of the main goals of latent variable modeling of multi-neuronal recordings [45, 47, 51, 53, 87, 88]. We contribute to this effort by proposing here a bottom-up, mechanistic LVM – the neuronally-grounded latent variable model (neuLVM) – which can be mapped to a multi-population SNN. Using SNN-based generative models, which are more biological than RNN-based models [45], could allow systems neuroscientists to test hypothesis about the architecture of the probed microcircuit, and provide a neuronally-grounded understanding of computation.

While this work shows the potential of the neuLVM approach, the application of neuLVM to real data faces two methodological challenges. First, there is the problem of identifiability: although neuLVM could recover a single unknown connectivity parameter (Section 5.1), our method could not always recover the SNN parameters when many parameters were unknown (Section 5.2). Bayesian inference could circumvent the problem of non-identifiability by estimating the full posterior distribution over model parameters [54, 89]. In addition, perturbing the probed network, with optogenetic stimulation for example, could help model parameter recovery by providing richer data. Second, in the case of real data, choosing the good generative SNN model is a nontrivial task. For example, how many homogeneous populations should the SNN have? Clustering the recorded spike trains could guide the design of possible generative models and Bayesian model comparison, as used in biophysical

modeling of neuroimaging data [90–92], could help in selecting the most likely model among several possible models.

The model proposed here is only one particular example of SNN-based, tractable latent variable model. Whether other such neuronally-grounded models of partially observed spike trains can be formulated and efficiently applied to real data is a question left for future work.

Acknowledgments and Disclosure of Funding

References

[1] Nicholas A Steinmetz, Christof Koch, Kenneth D Harris, and Matteo Carandini. Challenges and opportunities for large-scale electrophysiology with neuropixels probes. *Current opinion in neurobiology*, 50:92–100, 2018.

[2] Peiran Gao and Surya Ganguli. On simplicity and complexity in the brave new world of large-scale neuroscience. *Current opinion in neurobiology*, 32:148–155, 2015.

[3] Jonathan W Pillow and Peter Latham. Neural characterization in partially observed populations of spiking neurons. *Adv Neural Information Processing Systems*, 20(3.5), 2008.

[4] Johanni Brea, Walter Senn, and Jean-Pascal Pfister. Sequence learning with hidden units in spiking neural networks. *Advances in neural information processing systems*, 24, 2011.

[5] Danilo Rezende, Daan Wierstra, and Wulfram Gerstner. Variational learning for recurrent spiking networks. *Advances in neural information processing systems*, 24, 2011.

[6] Guillaume Bellec, Shuqi Wang, Alireza Modirshanechi, Johanni Brea, and Wulfram Gerstner. Fitting summary statistics of neural data with a differentiable spiking network simulator. *Advances in Neural Information Processing Systems*, 34:18552–18563, 2021.

[7] Wulfram Gerstner. Time structure of the activity in neural network models. *Phys. Rev. E*, 51(1): 738, 1995.

[8] Nicolas Brunel and Vincent Hakim. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural computation*, 11(7):1621–1671, 1999.

[9] Wulfram Gerstner. Population dynamics of spiking neurons: fast transients, asynchronous states, and locking. *Neural Comput.*, 12(1):43–89, 2000.

[10] Wulfram Gerstner, Werner M Kistler, Richard Naud, and Liam Paninski. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.

[11] Tilo Schwalger, Moritz Deger, and Wulfram Gerstner. Towards a theory of cortical columns: From spiking neurons to interacting neural populations of finite size. *PLoS computational biology*, 13(4):e1005507, 2017.

[12] Tilo Schwalger and Anton V Chizhov. Mind the last spike—firing rate models for mesoscopic populations of spiking neurons. *Curr. Opin. Neurobiol.*, 58:155–166, 2019.

[13] Valentin Schmutz, Eva Löcherbach, and Tilo Schwalger. On a finite-size neuronal population equation. *arXiv preprint arXiv:2106.14721*, 2021.

[14] Sandrine Lefort, Christian Tomm, J-C Floyd Sarria, and Carl CH Petersen. The excitatory neuronal network of the c2 barrel column in mouse primary somatosensory cortex. *Neuron*, 61 (2):301–316, 2009.

[15] Tobias C Potjans and Markus Diesmann. The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model. *Cerebral cortex*, 24(3):785–806, 2014.

[16] Yumiko Yoshimura, Jami LM Dantzker, and Edward M Callaway. Excitatory cortical neurons form fine-scale functional networks. *Nature*, 433(7028):868–873, 2005.

[17] Sen Song, Per Jesper Sjöström, Markus Reigl, Sacha Nelson, and Dmitri B Chklovskii. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS biology*, 3(3):e68, 2005.

[18] Claudia Clopath, Lars Büsing, Eleni Vasilaki, and Wulfram Gerstner. Connectivity reflects coding: a model of voltage-based stdp with homeostasis. *Nature neuroscience*, 13(3):344–352, 2010.

[19] Rodrigo Perin, Thomas K Berger, and Henry Markram. A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences*, 108(13):5419–5424, 2011.

[20] Ho Ko, Sonja B Hofer, Bruno Pichler, Katherine A Buchanan, P Jesper Sjöström, and Thomas D Mrsic-Flogel. Functional specificity of local synaptic connections in neocortical networks. *Nature*, 473(7345):87–91, 2011.

[21] Ashok Litwin-Kumar and Brent Doiron. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature neuroscience*, 15(11):1498–1505, 2012.

[22] Michael N Shadlen and William T Newsome. Noise, neural codes and cortical organization. *Current opinion in neurobiology*, 4(4):569–579, 1994.

[23] Michael N Shadlen and William T Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of neuroscience*, 18(10):3870–3896, 1998.

[24] Saurabh Vyas, Matthew D Golub, David Sussillo, and Krishna V Shenoy. Computation through neural population dynamics. *Annual Review of Neuroscience*, 43:249–275, 2020.

[25] Paul Miller. Itinerancy between attractor states in neural systems. *Current opinion in neurobiology*, 40:14–22, 2016.

[26] Giancarlo La Camera, Alfredo Fontanini, and Luca Mazzucato. Cortical computations via metastable activity. *Current opinion in neurobiology*, 58:37–45, 2019.

[27] Braden AW Brinkman, Han Yan, Arianna Maffei, Il Memming Park, Alfredo Fontanini, Jin Wang, and Giancarlo La Camera. Metastable dynamics of neural circuits and networks. *Applied Physics Reviews*, 9(1):011313, 2022.

[28] Kenneth W Latimer, Jacob L Yates, Miriam LR Meister, Alexander C Huk, and Jonathan W Pillow. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science*, 349(6244):184–187, 2015.

[29] Tatiana A Engel, Nicholas A Steinmetz, Marc A Gieselmann, Alexander Thiele, Tirin Moore, and Kwabena Boahen. Selective modulation of cortical state during spatial attention. *Science*, 354(6316):1140–1144, 2016.

[30] Luca Mazzucato, Giancarlo La Camera, and Alfredo Fontanini. Expectation-induced modulation of metastable activity underlies faster coding of sensory stimuli. *Nature neuroscience*, 22(5): 787–796, 2019.

[31] Rubén Moreno-Bote, John Rinzel, and Nava Rubin. Noise-induced alternations in an attractor network model of perceptual bistability. *Journal of neurophysiology*, 98(3):1125–1139, 2007.

[32] Luca Mazzucato, Alfredo Fontanini, and Giancarlo La Camera. Dynamics of multistable states during ongoing and evoked cortical activity. *Journal of Neuroscience*, 35(21):8214–8231, 2015.

[33] Jakob H Macke, Lars Buesing, John P Cunningham, Byron M Yu, Krishna V Shenoy, and Maneesh Sahani. Empirical models of spiking in neural populations. In *Advances in Neural Information Processing Systems 24: 25th conference on Neural Information Processing Systems (NIPS 2011)*, pages 1350–1358, 2012.

[34] G Ackerson and K Fu. On state estimation in switching environments. *IEEE transactions on automatic control*, 15(1):10–17, 1970.

[35] Zoubin Ghahramani and Geoffrey E Hinton. Variational learning for switching state-space models. *Neural computation*, 12(4):831–864, 2000.

[36] David Barber. Expectation correction for smoothed inference in switching linear dynamical systems. *Journal of Machine Learning Research*, 7(11), 2006.

[37] Emily Fox, Erik Sudderth, Michael Jordan, and Alan Willsky. Nonparametric bayesian learning of switching linear dynamical systems. *Advances in neural information processing systems*, 21, 2008.

[38] Biljana Petreska, Byron M Yu, John P Cunningham, Gopal Santhanam, Stephen Ryu, Krishna V Shenoy, and Maneesh Sahani. Dynamical segmentation of single trials from population neural data. *Advances in neural information processing systems*, 24, 2011.

[39] Scott Linderman, Matthew Johnson, Andrew Miller, Ryan Adams, David Blei, and Liam Paninski. Bayesian learning and inference in recurrent switching linear dynamical systems. In *Artificial Intelligence and Statistics*, pages 914–922. PMLR, 2017.

[40] Jayant E Kulkarni and Liam Paninski. Common-input models for multiple neural spike-train data. *Network: Computation in Neural Systems*, 18(4):375–407, 2007.

[41] Vernon Lawhern, Wei Wu, Nicholas Hatsopoulos, and Liam Paninski. Population decoding of motor cortical activity using a generalized linear model with hidden states. *Journal of neuroscience methods*, 189(2):267–280, 2010.

[42] Michael Vidne, Yashar Ahmadian, Jonathon Shlens, Jonathan W Pillow, Jayant Kulkarni, Alan M Litke, EJ Chichilnisky, Eero Simoncelli, and Liam Paninski. Modeling the impact of common noise inputs on the network activity of retinal ganglion cells. *Journal of computational neuroscience*, 33(1):97–121, 2012.

[43] Yuanjun Gao, Evan W Archer, Liam Paninski, and John P Cunningham. Linear dynamical neural population models through nonlinear embeddings. *Advances in neural information processing systems*, 29, 2016.

[44] Anqi Wu, Nicholas A Roy, Stephen Keeley, and Jonathan W Pillow. Gaussian process based nonlinear latent structure discovery in multivariate spike train data. *Advances in neural information processing systems*, 30, 2017.

[45] Chethan Pandarinath, Daniel J O'Shea, Jasmine Collins, Rafal Jozefowicz, Sergey D Stavisky, Jonathan C Kao, Eric M Trautmann, Matthew T Kaufman, Stephen I Ryu, Leigh R Hochberg, et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature methods*, 15(10):805–815, 2018.

[46] Lea Duncker and Maneesh Sahani. Temporal alignment and latent gaussian process factor inference in population spike trains. *bioRxiv*, page 331751, 2018.

[47] Lea Duncker, Gergo Bohner, Julien Boussard, and Maneesh Sahani. Learning interpretable continuous-time models of latent stochastic dynamical systems. In *International Conference on Machine Learning*, pages 1726–1734. PMLR, 2019.

[48] J Nassar, SW Linderman, M Bugallo, and IM Park. Tree-structured recurrent switching linear dynamical systems for multi-scale modeling. In *International Conference on Learning Representations (ICLR)*, 2019.

[49] Joshua Glaser, Matthew Whiteway, John P Cunningham, Liam Paninski, and Scott Linderman. Recurrent switching dynamical systems models for multiple interacting neural populations. *Advances in neural information processing systems*, 33:14867–14878, 2020.

[50] David Zoltowski, Jonathan Pillow, and Scott Linderman. A general recurrent state space framework for modeling neural dynamics during decision-making. In *International Conference on Machine Learning*, pages 11680–11691. PMLR, 2020.

[51] Virginia Rutten, Alberto Bernacchia, Maneesh Sahani, and Guillaume Hennequin. Non-reversible gaussian processes for identifying latent dynamical structure in neural data. *Advances in Neural Information Processing Systems*, 2020.

[52] Stephen Keeley, Mikio Aoi, Yiyi Yu, Spencer Smith, and Jonathan W Pillow. Identifying signal and noise structure in neural population activity with gaussian process factor models. *Advances in Neural Information Processing Systems*, 33:13795–13805, 2020.

[53] Timothy D Kim, Thomas Z Luo, Jonathan W Pillow, and Carlos Brody. Inferring latent dynamics underlying neural population activity via neural differential equations. In *International Conference on Machine Learning*, pages 5551–5561. PMLR, 2021.

[54] Alexandre René, André Longtin, and Jakob H Macke. Inference of a mesoscopic population model from population spike trains. *Neural computation*, 32(8):1448–1498, 2020.

[55] Liam Paninski. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15(4):243–262, 2004.

[56] Wilson Truccolo, Uri T Eden, Matthew R Fellows, John P Donoghue, and Emery N Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J. Neurophysiol.*, 93(2):1074–1089, 2005.

[57] Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, EJ Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

[58] David Zoltowski and Jonathan W Pillow. Scaling the poisson glm to massive neural datasets through polynomial approximations. *Advances in neural information processing systems*, 31, 2018.

[59] Regis C Lambert, Christine Tuleau-Malot, Thomas Bessaih, Vincent Rivoirard, Yann Bouret, Nathalie Leresche, and Patricia Reynaud-Bouret. Reconstructing the functional connectivity of multiple spike trains using Hawkes models. *J. Neurosci. Methods*, 297:9–21, 2018.

[60] Ryota Kobayashi, Shuhei Kurita, Anno Kurth, Katsunori Kitano, Kenji Mizuseki, Markus Diesmann, Barry J Richmond, and Shigeru Shinomoto. Reconstructing neuronal circuitry from parallel spike trains. *Nature communications*, 10(1):1–13, 2019.

[61] Werner M Kistler, Wulfram Gerstner, and J Leo van Hemmen. Reduction of the hodgkin-huxley equations to a single-variable threshold model. *Neural computation*, 9(5):1015–1045, 1997.

[62] Renaud Jolivet, Timothy J Lewis, and Wulfram Gerstner. Generalized integrate-and-fire models of neuronal activity approximate spike trains of a detailed model to a high degree of accuracy. *Journal of neurophysiology*, 92(2):959–976, 2004.

[63] Romain Brette and Wulfram Gerstner. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of neurophysiology*, 94(5):3637–3642, 2005.

[64] Renaud Jolivet, Alexander Rauch, Hans-Rudolf Lüscher, and Wulfram Gerstner. Predicting spike timing of neocortical pyramidal neurons by simple threshold models. *Journal of computational neuroscience*, 21(1):35–49, 2006.

[65] Ryota Kobayashi, Yasuhiro Tsubo, and Shigeru Shinomoto. Made-to-order spiking neuron model equipped with a multi-timescale adaptive threshold. *Frontiers in computational neuroscience*, 3:9, 2009.

[66] Christian Pozzorini, Skander Mensi, Olivier Hagens, Richard Naud, Christof Koch, and Wulfram Gerstner. Automated high-throughput characterization of single neurons by means of simplified spiking models. *PLoS Comput Biol*, 11(6):e1004275, 2015.

[67] Corinne Teeter, Ramakrishnan Iyer, Vilas Menon, Nathan Gouwens, David Feng, Jim Berg, Aaron Szafer, Nicholas Cain, Hongkui Zeng, Michael Hawrylycz, et al. Generalized leaky integrate-and-fire models classify multiple neuron types. *Nature communications*, 9(1):1–15, 2018.

[68] David Roxbee Cox. *Renewal theory*. Methuen, 1962.

[69] Hugh R Wilson and Jack D Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.*, 12(1):1–24, 1972.

[70] Anna De Masi, Antonio Galves, Eva Löcherbach, and Errico Presutti. Hydrodynamic limit for interacting neurons. *J. Stat. Phys.*, 158(4):866–902, 2015. ISSN 0022-4715.

[71] Nicolas Fournier and Eva Löcherbach. On a toy model of interacting neurons. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 52, pages 1844–1876. Institut Henri Poincaré, 2016.

[72] Julien Chevallier. Mean-field limit of generalized Hawkes processes. *Stochastic Process. Appl.*, 127(12):3870–3912, 2017. ISSN 0304-4149.

[73] Julien Chevallier et al. Fluctuations for mean-field interacting age-dependent hawkes processes. *Electronic Journal of Probability*, 22, 2017.

[74] Grégory Dumont, Alexandre Payeur, and André Longtin. A stochastic-field description of finite-size spiking neural networks. *PLoS computational biology*, 13(8):e1005691, 2017.

[75] Valentin Schmutz, Wulfram Gerstner, and Tilo Schwalger. Mesoscopic population equations for spiking neural networks with synaptic short-term plasticity. *The Journal of Mathematical Neuroscience*, 10(1):1–32, 2020.

[76] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[77] Anne C Smith and Emery N Brown. Estimating a state-space model from point process observations. *Neural computation*, 15(5):965–991, 2003.

[78] Byron M Yu, John P Cunningham, Gopal Santhanam, Stephen I Ryu, Krishna V Shenoy, and Maneesh Sahani. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Journal of neurophysiology*, 102(1):614–635, 2009.

[79] Yariv Ephraim and Neri Merhav. Hidden markov processes. *IEEE Transactions on information theory*, 48(6):1518–1569, 2002.

[80] Armen Allahverdyan and Aram Galstyan. Comparative analysis of viterbi training and maximum likelihood estimation for hmms. *Advances in Neural Information Processing Systems*, 24, 2011.

[81] Wulfram Gerstner and J Leo van Hemmen. Coherence and incoherence in a globally coupled ensemble of pulse-emitting units. *Physical review letters*, 71(3):312, 1993.

[82] David Golomb and John Rinzel. Clustering in globally coupled inhibitory neurons. *Physica D: Nonlinear Phenomena*, 72(3):259–282, 1994.

[83] U Ernst, K Pawelzik, and T Geisel. Synchronization induced by temporal delays in pulse-coupled oscillators. *Physical review letters*, 74(9):1570, 1995.

[84] Kong-Fatt Wong and Xiao-Jing Wang. A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4):1314–1328, 2006.

[85] SP Lloyd. Least square quantization in pcm. bell telephone laboratories paper. published in journal much later: Lloyd, sp: Least squares quantization in pcm. *IEEE Trans. Inform. Theor.(1957/1982)*, 18:5, 1957.

[86] Mark CW van Rossum. A novel spike distance. *Neural computation*, 13(4):751–763, 2001.

[87] Yuan Zhao and Il Memming Park. Interpretable nonlinear dynamic modeling of neural trajectories. *Advances in neural information processing systems*, 29, 2016.

[88] Matthew R Whiteway and Daniel A Butts. The quest for interpretable models of neural population activity. *Current opinion in neurobiology*, 58:86–93, 2019.

[89] Jan-Matthis Lueckmann, Pedro J Goncalves, Giacomo Bassetto, Kaan Öcal, Marcel Nonnenmacher, and Jakob H Macke. Flexible statistical inference for mechanistic models of neural dynamics. *Advances in neural information processing systems*, 30, 2017.

[90] William D Penny, Klaas E Stephan, Andrea Mechelli, and Karl J Friston. Comparing dynamic causal models. *Neuroimage*, 22(3):1157–1172, 2004.

[91] Will D Penny, Klaas E Stephan, Jean Daunizeau, Maria J Rosa, Karl J Friston, Thomas M Schofield, and Alex P Leff. Comparing families of dynamic causal models. *PLoS computational biology*, 6(3):e1000709, 2010.

[92] Kay H Brodersen, Thomas M Schofield, Alexander P Leff, Cheng Soon Ong, Ekaterina I Lomakina, Joachim M Buhmann, and Klaas E Stephan. Generative embedding for model-based classification of fmri data. *PLoS computational biology*, 7(6):e1002079, 2011.

[93] David Pfau, Eftychios A Pnevmatikakis, and Liam Paninski. Robust learning of low-dimensional dynamics from large neural ensembles. *Advances in neural information processing systems*, 26, 2013.

Appendices of:

Mesoscopic modeling of hidden spiking neurons

A   Mesoscopic model in the case of LIF neurons

In this section, we present in detail the mesoscopic model of Schwalger et al. [11] in the case of multiple interacting populations of LIF neurons, as formulated in [13].

Fine-grained SNN of LIF neurons with escape noise.   Let us consider a general network of N LIF neurons (indexed by $i = 1, \ldots, N$) with escape noise [10]. Neurons are modeled as point processes: the probability for neuron i to emit a spike at time t, given the past network activity $\mathbf{y}_{1:t-1}$, is

$$p(y_t^i = 1 | \mathbf{y}_{1:t-1}, \Theta) = 1 - \exp\left(-\lambda_t^i \Delta t\right), \quad \text{with } \lambda_t^i = \exp\left(V^i(t|\hat{t}^i) - \vartheta^i\right),$$

where the escape rate (or stochastic intensity) λ_t^i depends on the momentary difference between the membrane potential $V^i(t|\hat{t}^i)$ and the firing threshold ϑ^i, via an exponential escape function. The voltage $V^i(t|\hat{t}^i)$ of neuron i at time t depends on its last spike time $\hat{t}^i = t - a^i$ and the inputs received up to time t, which include the inputs coming from the other neurons and the external input $\mathbf{I}_{1:t}^{\text{ext},i}$. Between spikes, for all $t > \hat{t}^i + t_{\text{ref}}^i$ (t_{ref}^i being the absolute refractory period of neuron i), the voltage dynamics follows

$$V^i(t|\hat{t}^i) = V^i(t-1|\hat{t}^i) + \left(\frac{U_{\text{r}}^i + RI_t^{\text{ext},i} - V^i(t-1|\hat{t}^i)}{\tau_{\text{mem}}^i}\right) \Delta t + \sum_{j=1}^N J^{ij}\left(\epsilon^{ij} * \mathbf{y}^j\right)(t),$$

and $V^i(t|\hat{t}^i) = 0$, for all $t \leq \hat{t}^i + t_{\text{ref}}^i$ (which means that the voltage is reset to 0 after each spike and is clamped at 0 for an absolute refractory period $t_{\text{ref}}^i \geq 0$). The parameters $\tau_{\text{mem}}^i > 0$ and $U_{\text{r}}^i > 0$ are the membrane time constant and the resting potential respectively. The neuron i is therefore characterized by the parameters $\theta^i = \{\vartheta^i, U_{\text{r}}^i, \tau_{\text{mem}}^i, t_{\text{ref}}^i\}$. While the escape function is usually parameterized by a rescaled exponential function of the form $f(v) = \frac{1}{\tau_0^i} \exp(\beta^i(v - \tilde{\vartheta}^i))$ [10, Sec 9.1], the parameters τ_0^i, β^i and $\tilde{\vartheta}^i$ can be absorbed in ϑ^i (up to a rescaling of the resting potential U_{r}^i). The resistance $R = 1\,\Omega$ is used here simply for the consistency of physical units. The postsynaptic current induced by a spike of neuron j on neuron i is define by the synaptic weight J^{ij} and the synaptic kernel $\epsilon^{ij} : \mathbb{R}_+ \to \mathbb{R}_+$. In this work, we consider exponential kernels of the form $\epsilon^{ij}(t) = \frac{\mathcal{H}(t - \Delta^{ij})}{\tau_{\text{syn}}^{ij}} \exp\left(-\frac{t - \Delta^{ij}}{\tau_{\text{syn}}^{ij}}\right)$, where τ_{syn}^{ij} is the synaptic time constant, Δ^{ij} is the synaptic delay and $\mathcal{H}$ is the Heaviside function. The symbol $*$ denotes the convolution operator.

Coarse-grained multi-population SNN.   Coarse-graining and mean-field approximations consist in partitioning the N neurons into K homogeneous populations, indexed by $\alpha = 1, \ldots, K$, where (i) all the neurons i in population α share the same neuronal parameters $\theta^i = \theta^\alpha$; (ii) for any neuron j in population β and any neuron i in population α, $J^{ij} = J^{\alpha\beta}/N^\beta$ (N^β being the number of neurons in population β) and $\epsilon^{ij} = \epsilon^{\alpha\beta}$; (iii) all the neurons i in population α share the same external input $\mathbf{I}^{\text{ext},i} = \mathbf{I}^{\text{ext},\alpha}$. In such a coarse-grained K-population SNN, we have, for any neuron i in population α,

$$\sum_{j=1}^N J^{ij}\left(\epsilon^{ij} * \mathbf{y}^j\right)(t) = \sum_{\beta=1}^K J^{\alpha\beta}\left(\epsilon^{\alpha\beta} * \mathbf{n}^\beta\right)(t)/N^\beta,$$

where $n_t^\beta = \sum_{i \in \text{pop. } \beta} y_t^i$ is the total number of spikes in population β at time t. Hence, the probability for *any* neuron i in population α to emit a spike at time t, given its age a and the past

population activity $\mathbf{n}_{1:t-1}$ is

$$p_{t,a}^{\text{fire},\alpha} = 1 - \exp\left(-\lambda_t^\alpha \Delta t\right), \quad \text{with } \lambda_t^\alpha = \exp\left(V^\alpha(t|t-a) - \vartheta^\alpha\right). \tag{5}$$

For all $a > t_{\text{ref}}^\alpha$, we have the update rule

$$V^\alpha(t|t-a) = V^\alpha(t-1|t-a) + \left(\frac{U_{\text{r}}^\alpha + RI_t^{\text{ext},\alpha} - V^\alpha(t-1|t-a)}{\tau_{\text{mem}}^\alpha}\right) \Delta t$$

$$+ \sum_{\beta=1}^{K} J^{\alpha\beta}\left(\epsilon^{\alpha\beta} * \mathbf{n}^\beta\right)(t)/N^\beta,$$

and $V^\alpha(t|t-a) = 0$ for all $a \leq t_{\text{ref}}^\alpha$. This gives the explicit expression for the probability $p_{t,a}^{\text{fire}}$ in Eq. (1). In this work, for simplicity, we will assume that all the synaptic kernels are the same, i.e. $\epsilon^{\alpha\beta} = \epsilon, \forall \alpha, \beta$ (see Table S2).

Mesoscopic description. The K-population SNN described above does not by itself constitute a mesoscopic model because the probability $p_{t,a}^{\text{fire},\alpha}$ still involves the age a of some neuron. To get a mesoscopic model (i.e. a model that does not involve the fine-grained modeling of each individual neuron), Schwalger et al. [11] used the population activity $\mathbf{n}$ to approximate the age density of each population and derived a closed form system of stochastic integral equations: For all $\alpha \in 1, \ldots, K$,

$$n_t^\alpha \sim \text{Binomial}\left(N^\alpha, \ \bar{n}_t^\alpha/N^\alpha\right), \tag{6a}$$

$$\bar{n}_t^\alpha = \left[\sum_{a \geq 1} p_{t,a}^{\text{fire},\alpha} S_{t,a}^\alpha n_{t-a}^\alpha + \Lambda_t^\alpha\left(N^\alpha - \sum_{a \geq 1} S_{t,a}^\alpha n_{t-a}^\alpha\right)\right]_+, \tag{6b}$$

$$\Lambda_t^\alpha = \frac{\sum_{a \geq 1} p_{t,a}^{\text{fire},\alpha}(1 - S_{t,a}^\alpha)S_{t,a}^\alpha n_{t-a}^\alpha}{\sum_{a \geq 1}(1 - S_{t,a}^\alpha)S_{t,a}^\alpha n_{t-a}^\alpha}, \tag{6c}$$

where $S_{t,a}^\alpha = \prod_{s=0}^{a-1}(1 - p_{t-a+s,s}^{\text{fire},\alpha})$ is the survival, i.e. the probability for a neuron in population α to stay silent between time $t - a$ and $t - 1$. A concise version of the derivation of the mesoscopic model (6) is presented in [13]. Note that Eq. (6) is not a one-dimensional stochastic dynamical system: the Markov embedding of the stochastic dynamics (6) is infinite-dimensional [13]. Indeed, Eq. (6) does not only describes the evolution of the population activity n_t^α but it also describes the evolution of the whole age (pseudo) density $\{S_{t,a}^\alpha n_{t-a}\}_{a \geq 1}$ in the population, also called the "refractory density" [12].

Formally, the 'initial condition' of Eq. (6) is defined by the population activity $\mathbf{n}_t$ for all $t \leq 0$ (denoted $\mathbf{n}_{t \leq 0}$). Several practical choices of initial conditions have been discussed in [11, 13, 54]. In this work, if not otherwise specified, $\mathbf{n}_{t \leq 0}$ is taken to be time-invariant, with stationary activities estimated from the observed data (see below).

The size of the discrete time steps Δt does not need to be the same for the fine-grained SNN and for the mesoscopic model (6). Indeed, it can be useful to take longer time steps for the mesoscopic description (time coarse-graining). In the following appendices, when there is an ambiguity, Δt_{meso} will denote the time step length for the mesoscopic model and neuLVM. The length Δt_{meso} will always be smaller or equal to the neuronal absolute refractory periods, so that a neuron can fire at most once in each time step.

B neuLVM for multiple interacting populations

Let us assume that we observe, during T time steps, the spike trains of q simultaneously recorded neurons that are part of a K-population SNN of N neurons, with $N > q$. For each of the population $\alpha = 1, \ldots, K$, $q^\alpha > 0$ neurons are observed ($\sum_{\alpha=1}^{K} q^\alpha = q$) and share the same set of neuronal parameters θ^α, input weights $\{J^{\alpha\beta}/N^\beta\}_{\beta=1}^{K}$, and output weights $\{J^{\beta\alpha}/N^\alpha\}_{\beta=1}^{K}$, where $N^1, \ldots, N^K$ are the numbers of neurons in each population ($\sum_{\alpha=1}^{K} N^\alpha = N$).

The likelihood $\mathcal{L}$ of the observed spike trains. Following the assumptions described above, the likelihood $\mathcal{L}$ of the observed spike trains $\mathbf{y^o}$ (a binary $q \times T$ matrix) can be formally written as $\sum_{\mathbf{n}} p(\mathbf{y^o}, \mathbf{n}|\Theta)$, where $\mathbf{n}$ (an integer-valued $K \times T$ matrix) is the population activity and $\Theta = \{\{J^{\alpha\beta}\}_{1 \leq \alpha, \beta \leq K}, \{\theta^{\alpha}\}_{\alpha=1}^{K}\}$ are the parameters of the K-population SNN. The probability $p(\mathbf{y^o}, \mathbf{n}|\Theta)$ factorizes in T terms of the form

$$p(\mathbf{y}_t^{\mathbf{o}}, \mathbf{n}_t|\mathbf{y}_{1:t-1}^{\mathbf{o}}, \mathbf{n}_{1:t-1}, \Theta) = \underbrace{p(\mathbf{y}_t^{\mathbf{o}}|\mathbf{y}_{1:t-1}^{\mathbf{o}}, \mathbf{n}_{1:t-1}, \Theta)}_{\text{part a}} \underbrace{p(\mathbf{n}_t|\mathbf{n}_{1:t-1}, \Theta)}_{\text{part b}}.$$

The probability (**part a**) of the observed spikes $\mathbf{y}_t^{\mathbf{o}}$ at time t given the past observed spike activity $\mathbf{y}_{1:t-1}^{\mathbf{o}}$ and the past population activity $\mathbf{n}_{1:t-1}$ is

$$p(\mathbf{y}_t^{\mathbf{o}}|\mathbf{y}_{1:t-1}^{\mathbf{o}}, \mathbf{n}_{1:t-1}, \Theta) = \prod_{\alpha=1}^{K} \prod_{i=1}^{q^{\alpha}} p(y_t^{\mathbf{o},\alpha,i}|a^i, \mathbf{n}_{1:t-1}, \Theta) = \prod_{\alpha=1}^{K} \prod_{i=1}^{q^{\alpha}} p_{t,a^i}^{\text{fire},\alpha},$$

where $p_{t,a^i}^{\text{fire},\alpha}$, given by Eq. (5) in Appendix A, is the probability for the recorded neuron i of population α to emit a spike at time t.

The probability (**part b**) of the population activity $\mathbf{n}_t$ at time t given the past population activity $\mathbf{n}_{1:t-1}$ is

$$p(\mathbf{n}_t|\mathbf{n}_{1:t-1}, \Theta) = \prod_{\alpha=1}^{K} p(n_t^{\alpha}|\mathbf{n}_{1:t-1}, \Theta),$$

where $p(n_t^{\alpha}|\mathbf{n}_{1:t-1}, \Theta)$ is approximated by the mesoscopic model (6).

C Fitting algorithm for neuLVM

Baum-Viterbi algorithm. Given the observed spike trains $\mathbf{y^o}$, we optimize the likelihood $\mathcal{L} = \sum_{\mathbf{n}} p(\mathbf{y^o}, \mathbf{n}|\Theta)$ via an EM-like algorithm – the Baum-Viterbi algorithm [79]. Relying on the heuristic that the posterior $p(\mathbf{n}|\mathbf{y^o}, \Theta)$ should be concentrated around its maximum, we approximate the posterior $p(\mathbf{n}|\mathbf{y^o}, \Theta)$ by a point mass δ_{μ}, where $\mu = \arg\max_{\mathbf{n}} \log p(\mathbf{y^o}, \mathbf{n}|\Theta)$. By doing so, the alternating estimation (E) and maximization (M) step of the n-th iteration read

E-step. $\widehat{\mathbf{n}}^n = \text{argmax}_{\mathbf{n}} \log p(\mathbf{y^o}, \mathbf{n}|\widehat{\Theta}^{n-1})$,

M-step. $\widehat{\Theta}^n = \text{argmax}_{\Theta} \log p(\mathbf{y^o}, \widehat{\mathbf{n}}^n|\Theta)$.

Details of the optimization. In the **M-step**, parameters Θ are optimized using the L-BFGS-B algorithm and the optimization stops when either the maximum number of iterations (maxiter$_\text{M}$) is reached, or the objective function improves by less than ftol$_\text{M}$, or the maximum norm of the gradient is less than gtol$_\text{M}$. Hyper-parameters including maxiter$_\text{M}$, ftol$_\text{M}$ and gtol$_\text{M}$ are given in Table S5. In the **E-step**, to carry out gradient ascent, we approximate the discrete Binomial distribution Eq. (6a) by a Gaussian, i.e. $n_t^{\alpha} \sim \mathcal{N}(\bar{n}_t^{\alpha}, \bar{n}_t^{\alpha})$, where $\bar{n}_t^{\alpha}$ is given by the mesoscopic model Eq (6) [11]. With this approximation, the latent population activity $\mathbf{n}$ is optimized with the Adam algorithm with learning rate lr$_\text{E}$ and the optimization stops when either the maximum number of iterations (maxiter$_\text{E}$) is reached, or the objective function stops improving for the last itertol$_\text{E}$ iterations. Hyper-parameters including lr$_\text{E}$, maxiter$_\text{E}$, ftol$_\text{E}$ and itertol$_\text{E}$ are given in Table S5. The estimated parameters $\widehat{\Theta}$ and the estimated latent population activity $\widehat{\mathbf{n}}$ are the result of many iterations of **E-step** and **M-step**. The fitting algorithm ends either when it stops improving the objective function or the maximum number of E-M iterations is reached.

Multiple data-driven initializations. To deal with the fact that the joint probability $p(\mathbf{y^o}, \mathbf{n}|\Theta)$ to optimize is non-convex and high-dimensional ($\mathbf{n}$ has dimension $K \times T$), we perform the Baum-Viterbi algorithm N$_\text{init}$ times with initial parameters $\widehat{\Theta}^0$ uniformly sampled in a certain range given in Appendices E and F. Since the sum over the observed neurons from population α, $\sum_{i=1}^{q^{\alpha}} \mathbf{y}_{1:T}^{\mathbf{o},i}$, already provides a rough estimate of the latent population activity $\mathbf{n}_{1:T}^{\alpha}$, the **E-Step** of the first iteration ($\widehat{\mathbf{n}}^1$) is replaced by an empirical estimation of the population activity $\widehat{\mathbf{n}}_{\sigma}^{\text{sm}}$ from the observed spike trains (see Appendix D).

3

Numerical implementation of the mesoscopic model. To implement the mesoscopic model (6), we approximate the infinite sums $\sum_{a \geq 1}$ in Eq. (6) by finite sums $\sum_{a=1}^{a_{\max}}$, where $a_{\max}$ is chosen to be large enough such that the probability for a neuron to remain silent for a duration longer than $a_{\max}$ is negligible. In a our numerical implementation, the mesoscopic model (6) has therefore a finite memory $a_{\max}$. Note that a more principled way to implement finite memory can be found in [13], where a numerical implementation similar to ours is presented in detail. The hyper-parameter $a_{\max}$ is given in Appendices E and F. If not otherwise specified, the initial condition $\mathbf{n}_{\leq 0}$ of Eq. (6) are chosen to be time-invariant, with stationary activities estimated from the first $a_{\max}$ time steps of the recorded spike trains.

D   Smoothed empirical population activity

A smoothed empirical estimation of the population activity $\widehat{\mathbf{n}}_{\sigma}^{\mathrm{sm}}$ was obtained from the recorded spike trains $\mathbf{y}^{\mathbf{o}}$ by applying a Gaussian smoothing kernel g_{σ} with standard deviation σ. For population $\alpha = 1, \ldots, K$,

$$\widehat{\mathbf{n}}_{\sigma,t}^{\mathrm{sm},\alpha} = \left(\left(\frac{N^{\alpha}}{q^{\alpha}} \sum_{i=1}^{q^{\alpha}} \mathbf{y}_{1:T}^{\mathbf{o},\alpha,i} \right) * g_{\sigma} \right)(t).$$

E   Details of the cluster state example

Values of parameters used in this example are given in Table S2, except if mentioned otherwise.

When the network is initialized in the unstable asynchronous state (Figure 2 B-C). In this case, the network is always initialized, at time 0, in the same unstable asynchronous state with firing rate 20 Hz. The spike train power spectrum (Figure 2 B), for different choices of connectivity parameter J, were computed using 600 non-overlapping segments of 120 s. To measure the goodness of the connectivity recovered by newLVM, for each J in $\{59, 60, 61, 62, 63, 64, 65\}$ mV, we simulated the ground truth SNN (starting from the same unstable asynchronous state mentioned above) for 1 s and further generated 10 different datasets with different samples of six observed neurons (1% of the population).

When the network is initialized in a 4-cluster state (Figure 2 D-F). In this case, we simulated a trial (one second, $J = 60.32$ mV) with a transition from a metastable 4-cluster state to a 3-cluster state (Figure 2 D E). To test how well newLVM work in the regime where only a tiny fraction of the total number of neurons are observed, for each number $\{1, 2, 5, 10\}$ of observed neurons, we generated 10 different datasets with different samples of observed neurons.

Fitting of the neuLVM. The initial parameter $\widehat{J}^0$ was drawn uniformly in $[10, 30] \cup [90, 110]$ mV. The latent population activity was initialized as the smoothed empirical population activity ($\widehat{\mathbf{n}}^1 = \widehat{\mathbf{n}}_{\sigma,t}^{\mathrm{sm}}$, Appendix D) with $\sigma = 1.4$ ms (Figure 2 E). Since the Baum-Viterbi algorithm converged reliably when only J was unknown, N_{init} was set to 1. The hyper-parameter Δt_{meso} was set to 1 ms and $a_{\max}$ was set to 100 ($a_{\max} \Delta t_{\mathrm{meso}} = 100$ ms).

Fitting of René et al. (2020). A naive application of René et al. [54] consists in fitting the model with $\widehat{J} = \mathrm{argmax}_{\Theta} \log p(\widehat{\mathbf{n}}^1 | J)$. The parameter $\widehat{J}^0$ and the latent population activity $\widehat{\mathbf{n}}^1$ were set the same way as for neuLVM, but N_{init} was set to 200. The best performing $\widehat{J}$s were reported in Figure 2 F. The hyper-parameters Δt_{meso} and $a_{\max}$ were the same as for neuLVM.

F   Details of the metastable point attractors example

Values of parameters used in this example are given in Table S2, except if mentioned otherwise. In this example, we simulated a 500 s-long trial and randomly cut out 20 non-overlapping 10 s-segments to generate the training datasets.

Fitting of the neuLVM. The initial parameter $\widehat{\Theta}^0$ (which include the connectivities $J^{\cdot}$, membrane time constants $\tau^{\cdot}_{\mathrm{mem}}$, firing thresholds $\vartheta^{\cdot}$ and resting potentials $U^{\cdot}_r$) were sampled randomly by assuming the uniform prior on the range 0.4 to 2 times the ground truth values. In this example, the connectivity matrix $\mathbf{J}$ was parametrized by $\{J^{e^1} J^{e^2}, J^i\}_{J>0}$: $\mathbf{J} = \begin{pmatrix} J^{e^1} & 0 & 0 \\ 0 & J^{e^2} & 0 \\ 0 & 0 & J^i \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & -1 & -1 \end{pmatrix}$ (see Figure 3 A for the network architecture). The latent population activity was initialized as the smoothed empirical population activity ($\widehat{\mathbf{n}}^1 = \widehat{\mathbf{n}}^{\mathrm{sm},\alpha}_{\sigma,t}$, Appendix D) with $\sigma = 400$ms (Figure S7). Out of 5 fits ($N_{\mathrm{init}} = 5$), the fit with the highest joint likelihood $p(\mathbf{y^o}, \widehat{\mathbf{n}}|\widehat{\Theta})$ was selected. The related hyper-parameter Δt_{meso} was set to 4 ms and a_{max} was set to 250 ($a_{\mathrm{max}}\Delta t_{\mathrm{meso}} = 1000$ ms). When Δt_{meso} was set to a value that was larger than the Δt of the recorded data, the recorded spike trains were downsampled.

PLDS We used code from `https://bitbucket.org/mackelab/pop_spike_dyn/src/master/`. To fit Poisson Linear Dynamical Systems (PLDS) [33] to the three-population example, we initialized the parameters with nuclear norm penalized rate estimation [93] and used the variational EM algorithm of [33]. The dimensionality of the latent states was set to three (the number of populations). The time resolution of the recorded spike trains was downsampled to 4 ms ($\Delta t_{\mathrm{PLDS}} = 4$ ms). Other hyper-parameters were set to default.

SLDS We used code from `https://github.com/lindermanlab/ssm` [39]. To fit Poisson Switching Linear Dynamical Systems (SLDS) [34–38] to the three-population example, we updated the parameters with stochastic variational inference with the posterior approximated by a factorized distribution. The dimensionality of the continuous latent states was chosen to be three (the number of populations) and the dimensionality of the discrete latent states was chosen to be three (corresponding to the number of metastable states plus one for the transition state). We specified the 'emissions model' as 'Poisson_orthog' with the exponential escape function. Other hyper-parameters were set to default. Further, for SLDS to work, discrete time step had to be large enough. Here we downsampled datasets to 40 ms (the smallest Δt_{SLDS} that worked).

An additional test. We were interested to find out whether neuLVM is robust to within-population heterogeneity and slightly out-of-distribution data. To answer this question, we performed an additional test where we introduced within-population heterogeneity in the ground truth winner-take-all (WTA) network (Section 5.2) by adding noise to the connectivity and neuronal parameters as specified in the Table S6 (noise in the neuronal parameters is small to conserve metastable WTA dynamics). Furthermore, we set the N's of the neuLVM to 300, 300, 300 (the N's of the ground truth network are 400, 400, 200). We tested neuLVM on eight 10 s-segments cut out from a 100 s-long trial. The method is only mildly affected by these changes: all fitted neuLVM reproduced metastable WTA dynamics and the Pearson correlation between $\widehat{\mathbf{n}}|\mathbf{y^o}$ and $\mathbf{n}^*$ was 0.76 ± 0.02, which is still higher than the correlations obtained by PLDS and SLDS (see Table 1).
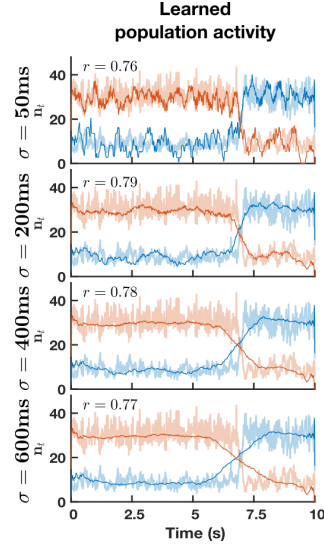
5

Learned population activity

Figure S6: Smoothed empirical estimate $\widehat{\mathbf{n}}_{\sigma,t}^{\mathrm{sm},\alpha}$ (Appendix D) of the latent population activity for one example trial (the same as in Figure 3, two excitatory populations). The value r is the Pearson correlation coefficient between the inferred $\widehat{\mathbf{n}}|\mathbf{y}^{\mathbf{o}}$ and the ground truth $\mathbf{n}^*$ population activities.
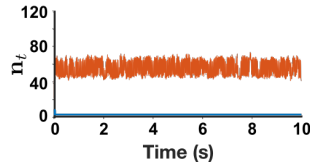

Spontaneous population activity

Figure S7: Spontaneous population activity simulated by the neuLVM before learning. Population activity of one excitatory population (the blue trace) quickly dies out. No visible metastable dynamics.

Table S2: Values of parameters used in simulations. **Boldface** is used to indicate fitted parameters.

	Name	Description	Value	
			Example Section 5.1 Single excit. population	Example Section 5.2 Excitat. (inhib.) populations
	Δt	time step	1 ms	0.2 ms
	N	number of neurons	600	400 (200)
Θ	J	**connectivity**	**60.32 mV**	**9.984 mV (-19.968 mV)***
	ϑ	**firing threshold**	49.7 mV	**3.7 mV (3.7 mV)**
	U_{r}	**resting potential**	26 mV	**14.4 mV (14.4 mV)**
	τ_{mem}	**membrane time constant**	100 ms	**20 ms (20 ms)**
	t_{ref}	absolute refractory period	0 ms	4 ms (4 ms)
ϵ	τ_{syn}	synaptic time constant	4 ms	3 ms (6 ms)
	Δ	synaptic delay	10 ms	0 ms (0 ms)

* i.e. for all population α, $J^{\alpha\beta} = 9.984$ mV if β is an excitatory population and $J^{\alpha\beta} = -19.968$ mV if β is the inhibitory population.

Table S3: Performance summary (ii) when fitting neuLVMs to the single-population example (Section 5.2, Figure 2 C) with m-cluster states. For each ground truth J, 10 different datasets were generated and tested. (6 observed neurons.)

J	59	60	61	62	63	64	65
$\widehat{J}$ (mean)	58.18	60.46	61.20	62.05	62.39	63.58	63.66
$\widehat{J}$ (std)	0.50	0.71	0.93	0.46	1.11	1.59	1.85
Pearson r				0.81 ($p = 2.8\mathrm{e}{-17}$)			

Table S4: Performance summary (i) when fitting neuLVMs to the single-population example (Section 5.2, Figure 2 F) with a transition from a metastable 4-cluster state to a 3-cluster state. For each ground truth J, 10 different datasets were generated and tested. ($J = 60.32$ mV.)

# observed neurons	1	2	5	10	599
$\widehat{J}$ (mean)	60.37	60.14	59.33	59.81	59.10
$\widehat{J}$ (std)	2.43	1.48	0.91	1.25	1.26

Table S5: Hyper-parameters used when fitting neuLVM.

Name	Value	
	Example Section 5.1 Single excit. population	Example Section 5.2 Excitat. (inhib.) populations
$\mathrm{lr_E}$	1e−3	1e−3
$\mathrm{maxiter_E}$	200	200
$\mathrm{itertol_E}$	3	3
$\mathrm{lr_M}$	1e−8 *	1e−8*
$\mathrm{maxiter_M}$	200	200
$\mathrm{ftol_M}$	2e−9*	2e−9*
$\mathrm{gtol_M}$	1e−5*	1e−5*

* Values are default as in scipy.optimize.minimize(method='L-BFGS-B').

Table S6: Within-population heterogeneity introduced in the ground truth winner-take-all (WTA) network (in the additional experiment of Appendix F).

ground truth within-population heterogeneity	μ	σ (normal distribution)
$J^{\mathrm{e}^1} J^{\mathrm{e}^2}, J^{\mathrm{i}}$	9.98 / 9.98 / 19.97	2.00 / 2.00 / 2.00
ϑ	3.70	0.07
U_{r}	14.40	0.29
τ_{mem}	20.00	0.40

6 Convergence of redundancy-free spiking neural networks to rate networks

Authors: Valentin Schmutz, Johanni Brea and Wulfram Gerstner

Contribution: I am the first author of this work.

Manuscript in preparation.

Convergence of redundancy-free spiking neural networks to rate networks

Valentin Schmutz, Johanni Brea, Wulfram Gerstner

Laboratory of Computational Neuroscience,

École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

(Dated: October 12, 2022)

Abstract

Can the dynamics of large Spiking Neural Networks (SNNs) converge to the smooth dynamics of equally large Recurrent Neural Networks (RNNs)? Classical mean-field theory provides a positive answer when the networks are redundant, that is, when each neuron has many "twins" receiving nearly identical inputs. Using a disordered network model which guarantees the absence of redundancy in large networks, we show that redundancy-free, densely connected spiking neural networks converge to RNNs when the ℓ^2 norms of the rows of the connectivity matrix tend to zero.

Neurons in the brain interact via spikes – short and stereotyped membrane potential deflections – commonly modeled as Dirac pulses [1]. SNNs with recurrent connectivity offer simplified models of real networks retaining the essential biological feature of spike-based neuronal communication. On the other hand, traditional RNN are continuous dynamical systems where abstract rate neurons directly transmit their firing rate to other neurons, a type of communication which is not biological. Despite their inferior realism, RNNs continue to play a central role in theoretical neuroscience because they can be trained by modern machine learning methods [2–4], they can be analysed using tools from statistical physics [5–8], and because biological networks are believed to perform computation by implementing continuous dynamical systems [9, 10]. Closing the gap between the more biological SNNs and the more tractable RNNs requires identifying the conditions under which the continuous dynamics of RNNs can be approximated by SNNs [11].

To clearly state the problem, let us consider a SNN composed of N Poisson neurons (linear-nonlinear-Poisson neurons [12] or nonlinear Hawkes processes [13]). For each neuron index i, the spike times $\{t_i^k\}_k$ of neuron i, which define the neuron's spike train $s_i(t) = \sum_k \delta(t - t_i^k)$, are generated by an inhomogeneous Poisson process with instantaneous firing rate $\phi(h_i(t))$, where $h_i(t)$ represents the neuron's potential and ϕ is a positive-valued nonlinear transfer function. The potential $h_i(t)$ is a leaky integrator of the recurrent inputs coming from neurons $j \neq i$ and the external input $I_i^{\text{ext}}(t)$:

$$\tau \frac{d}{dt} h_i(t) = -h_i(t) + \sum_{j=1}^N J_{ij} s_j(t) + I_i^{\text{ext}}(t), \tag{1}$$

where τ is the integration (or membrane) time constant and J_{ij} is the synaptic weight from neuron j to neuron i (by convention, $J_{ii} = 0$). While the spike-based model described here is biologically simplistic, it is mathematically convenient as it has a straightforward

rate-based counterpart. If we replace the spike trains $\{s_j(t)\}_j$ in Eq. (1) by the corresponding instantaneous firing rates $\{\phi(h_j(t))\}_j$ (i.e. neurons communicate their firing rate directly), we get the rate-based dynamics

$$\tau \frac{d}{dt} x_i(t) = -x_i(t) + \sum_{j=1}^{N} J_{ij} \phi(x_j(t)) + I_i^{\text{ext}}(t), \tag{2}$$

which defines a RNN with N rate units. To avoid confusion, we write $h_i(t)$ for the potentials of the SNN (1) and $x_i(t)$ for the potentials of the RNN (2). While the mapping from the SNN to the RNN looks simple at first glance, the spike-based stochastic process Eq. (1) and the rate-based dynamical system Eq. (2) describe very different kinds of systems and the SNN potentials $h_i(t)$ are not guaranteed, in general, to be equal or even close to the RNN potentials $x_i(t)$ even if both networks receive the same external input. Note that if the neurons are uncoupled (i.e. $J_{ij} = 0$ for all i, j), the SNN potentials $h_i(t)$ are trivially equal to the RNN potentials $x_i(t)$. Therefore, comparing the SNN and the RNN is meaningful only if the coupling does not vanish. For nontrivial coupling, there are two known types of scaling limits where the SNN potentials $h_i(t)$ converge to the RNN potentials $x_i(t)$:

(i) *Spatial averaging over many redundant neurons*: Consider networks of increasing size N. If each neuron can be assigned to a point in some fixed space such that two neurons assigned to the same point always share the same recurrent and external input, and if the synaptic weights are scaled by $1/N$, we can take the mean-field limit $N \to \infty$ [14, 15]. The fixed space can be either discrete and finite [16, 17] or continuous and finite-dimensional [18], e.g. a ring. These classical mean-field limits entail redundancy, i.e. the existence of large groups of identical neurons receiving the same recurrent and external input when $N \to \infty$. To our knowledge, this form of redundancy has not been found in the cortex.

(ii) *Temporal averaging over many spikes*: In Eq. (1), we can replace the transfer function ϕ and the synaptic weights $\{J_{ij}\}_{i \neq j}$ by $b \phi$ and $\{J_{ij}/b\}_{i \neq j}$, respectively (for $b > 0$), and take the limit $b \to \infty$ [19]. This limit entails arbitrarily high firing rates in the SNN, which is biological unrealistic since two spikes have to be separeted by at least 1 to 2 milliseconds (the absolute refractory period) [20]. Similarly, we can take the limit $\tau \to \infty$ in both Eq. (1) and (2) while re-scaling the synaptic weights $\{J_{ij}\}_{i \neq j}$ and the external inputs $I_i^{\text{ext}}(t)$ by $1/\tau$. This last limit entails arbitrarily slow network dynamics,

which is incompatible with human visual processing speed (less than 150 milliseconds) [21].

In this letter, we address the following question: can large SNNs, as defined in Eq. (1), converge to equally large RNNs without involving redundancy or temporal averaging.

Avoiding temporal averaging (ii) will be guaranteed by assuming that $\max \phi \leq 1/\tau$. Under this condition, leaky integration by the potential, Eq. (1), is too fast to average out the Poisson noise of individual input spike trains and neither of the two scalings mentioned under (ii) can be applied.

To assess the redundancy of a RNN, we look at the distribution of correlations between pairs of distinct neurons $i \neq j$

$$\rho(x_i, x_j) := \lim_{T \to \infty} \frac{1}{T} \int_0^T \frac{(x_i(t) - \overline{x_i})(x_j(t) - \overline{x_j})}{\sigma_{x_i} \sigma_{x_j}} dt,$$

where $\overline{x_i}$ and $\sigma_{x_i}^2$ are, respectively, the time average and fluctuation of $x_i(t)$:

$$
\begin{aligned}
\overline{x_i} &:= \lim_{T \to \infty} \frac{1}{T} \int_0^T x_i(t) dt, \\
\sigma_{x_i}^2 &:= \lim_{T \to \infty} \frac{1}{T} \int_0^T (x_i(t) - \overline{x_i})^2 dt.
\end{aligned}
\tag{3}
$$

Based on the correlation distribution (which has total mass 1), we say that a network is redundant if the mass of the correlation distribution around 1 (e.g. on some small interval $[\theta, 1]$ with $0 < \theta < 1$) is non-negligible. For example, if the N units of a RNN are uniformly assigned to points on a ring where nearby neurons receive similar recurrent and external (stochastic) inputs, in the mean-field limit $N \to \infty$, the distribution of the correlations $\{\rho(x_i, x_j)\}_{i \neq j}$ converge to a correlation distribution with strictly positive mass around at 1, reflecting the fact that the number of redundant "twins" per unit grows linearly with N (Fig 1A, see the Appendix for the details on the simulated model). This simple ring model illustrates how redundancy builds up in classical mean-field models as $N \to \infty$ (i).

Conversely, we say that large networks of a given model are *redundancy-free* if, for any $0 < \theta < 1$, the mass of the correlation distribution on the interval $[\theta, 1]$ vanishes as $N \to \infty$. In the following, we propose a disordered network model where large networks are redundancy-free.

Input-driven disordered network model.— We construct the connectivity matrix $\mathbf{J} = \{J_{ij}\}_{i,j}$ as a sum of random rank-one matrices (minus self-interaction terms), a construction similar
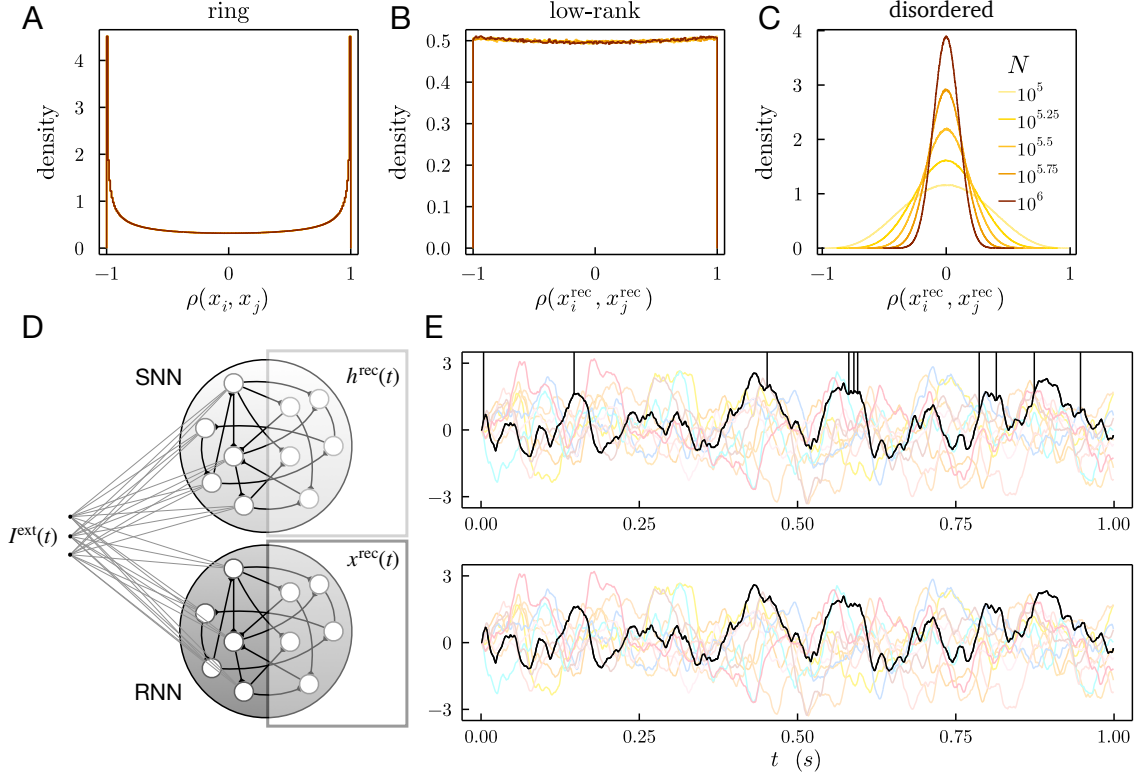
4

FIG. 1. Networks with (**A**,**B**) and without (**C**) redundant neurons. Distributions of correlations $\rho(x_i, x_j)$ in RNNs of increasing size N, for (**A**) a ring model, (**B**) a low-rank model with rank $p = 3$, and (**C**) a disordered network model with load $\alpha = 10^{-4}$. In (B,C), the correlations $\rho(x_i^{\text{rec}}, x_j^{\text{rec}})$ are for pairs of neurons receiving no external input but only recurrent input. In the disordered network model (C), correlations concentrate around 0 as $N \to \infty$. (**D**) SNN and RNN comparison. The potentials of the neurons receiving no external input but only recurrent input, $h^{\text{rec}}(t)$ and $x^{\text{rec}}(t)$, are linear readouts of the recurrent drive. (**E**) Trajectories of single-neuron potentials in the SNN ($h_i^{\text{rec}}(t)$, upper panel) and in the RNN ($x_i^{\text{rec}}(t)$, lower panel) during a one-second simulation of the setup shown in (D). The networks have $N = 10^6$ neurons and load $\alpha = 10^{-4}$, as in (C). The same randomly chosen 11 neurons are recorded in the SNN and in the RNN and each color corresponds to a different neuron index i (the colors in the upper and lower panels correspond). For one neuron index i (the black trace), the spike times of the neuron in the SNN are indicated by vertical bars. The differences between the potentials of the SNN and the RNN are almost imperceptible. (**A**-**E**) Neuronal parameters: $\tau = 10$ ms and $\phi(x) = \frac{1}{2\tau}\left(\tanh(x - \theta) + 1\right)$ with $\theta = 2$. Input noise is $\sigma = 0.5$ in (B-E).

5

to that of Hopfield networks [7, 22, 23]. For any number of units N and any number of patterns p, let $\boldsymbol{\xi}$ be a random $N \times p$-matrix with *i.i.d.*, zero-mean, unit-variance, normally distributed entries $\{\xi_{i\mu}\}_{i,\mu}$. We choose a connectivity matrix given by

$$J_{ij} := \frac{1}{cN} \sum_{\mu=1}^{p} \xi_{i\mu} \left(\phi(\xi_{j\mu}) - a \right) \quad \text{for all } i \neq j, \tag{4}$$

and $J_{ii} := 0$ for all i, where the constants $a := \mathbb{E}[\phi(\xi_{11})]$ and $c := \mathrm{Var}[\phi(\xi_{11})]$ guarantee the normalization

$$\frac{1}{cN} \sum_{i=1}^{N} \left(\phi(\xi_{i1}) - a \right) \phi(\xi_{i1}) \to 1, \quad \text{as } N \to \infty.$$

A well-known feature of this type of connectivity is that, exchanging the order of summation, the dynamics of the RNN (2) can be re-written in terms of p overlap variables $\{m_\mu(t)\}_\mu$ [24–26]: for all $i = 1, \ldots, N$ and for all $\mu = 1, \ldots, p$,

$$\tau \frac{d}{dt} x_i(t) = -x_i(t) + \sum_{\nu=1}^{p} \xi_{i\nu} m_\nu(t) - \gamma_i \phi(x_i(t)) + I_i^{\text{ext}}(t),$$

$$m_\mu(t) = \frac{1}{cN} \sum_{j=1}^{N} \left(\phi(\xi_{j\mu}) - a \right) \phi(x_j(t)),$$

where the $\gamma_i := \frac{1}{cN} \sum_{\nu=1}^{p} \xi_{i\nu} \left(\phi(\xi_{i\nu}) - a \right)$ are virtual self-interaction weights. Analogously, for the SNN, we have

$$\tau \frac{d}{dt} h_i(t) = -h_i(t) + \sum_{\nu=1}^{p} \xi_{i\nu} m_\nu(t) - \gamma_i s_i(t) + I_i^{\text{ext}}(t),$$

$$m_\mu(t) = \frac{1}{cN} \sum_{j=1}^{N} \left(\phi(\xi_{j\mu}) - a \right) s_j(t). \tag{5}$$

This reformulation clearly shows that if $p \ll N$, the γ_i are small and therefore the recurrent drive $\{\sum_{j=1}^{N} J_{ij} \phi(x_j(t))\}_i$ is approximately restricted to the p-dimensional subspace spanned by the p columns of the random matrix $\boldsymbol{\xi}$. To force the recurrent drive to visit all p dimensions homogeneously over time in a single stationary process, we inject the following p-dimensional external input to half of the neurons:

$$I_i^{\text{ext}}(t) = \frac{\sigma}{\sqrt{p}} \sum_{\mu=1}^{p} \xi_{i\mu} \eta_\mu(t) \quad \text{if } i \leq N/2,$$

$$I_i^{\text{ext}}(t) = 0 \qquad\qquad\qquad \text{if } i > N/2, \tag{6}$$

where the $\eta_1(t), \ldots, \eta_p(t)$ are independent Gaussian white noises and $\sigma > 0$ is the input noise parameter. For indexing convenience, the potentials of the $N/2$ neurons receiving external input but only recurrent input will be denoted $x_i^{\text{rec}}(t)$; the $x_i^{\text{rec}}(t)$ can therefore be seen as linear readouts of the recurrent drive; cf. Eq. (2).

If the number of patterns p is kept constant as $N \to \infty$, the limit RNN is a low-rank mean-field model [8]. In such a low-rank model, redundancy also builds up as $N \to \infty$, namely, the distribution of correlations converges to a limit distribution where the density at 1 is strictly positive (Fig. 1B), indicating again that the number of redundant "twins" per unit grows linearly with N. The reason for the accumulation of redundancy is the same as in the ring model except that, here, the fixed space is not a ring but $\mathbb{R}^p$: unit i has coordinate $(\xi_{i1}, \ldots, \xi_{ip})$ and units with similar coordinates receive similar recurrent and external inputs. Therefore, if p is kept constant as $N \to \infty$, we fall again in a case of spatial averaging over redundant neurons (i); cf. [14].

To prevent redundancy from accumulating as $N \to \infty$, we make the number of patterns p grow linearly with N, taking $p = \alpha N$ for some fixed load $\alpha > 0$, as in the Hopfield model [23]. With this choice of scaling, synaptic weights $\{J_{ij}\}_{i,j}$ scale as $\mathcal{O}(1/\sqrt{N})$ (as in random RNNs [6, 27]), whence the name "disordered network" for this model. We will show that for any fixed $\alpha > 0$, large networks are redundancy-free (as defined above). Then, we will show that large SNNs converge to large RNNs, as $\alpha \to 0$, achieving the numerical demonstration that large, redundancy-free SNNs can converge to equally large RNNs. To show the convergence of the SNN (1) to the RNN (2), we will inject the same time-dependent external input (6) in both networks (Fig 1D) and compare the trajectories $h_i^{\text{rec}}(t)$ of the SNN with the trajectories $x_i^{\text{rec}}(t)$ of the RNN (Fig 1E).

We first verify numerically that the SNNs (and the RNNs) converge to the dynamic mean-field limit of a disordered system [6] (not to be confused with classical mean-field limits where there is no disorder [15, 28]). Since the trajectories $h_i^{\text{rec}}(t)$ of the SNN are almost identical to the trajectories $x_i^{\text{rec}}(t)$ of the RNN when α is small ($\alpha = 10^{-4}$ in Fig. 1C,E and Fig 2A-E), we only show numerical results for the SNN; those for the RNN are almost identical. Defining the time average $\overline{h_i^{\text{rec}}}$ and the fluctuation $\sigma_{h_i^{\text{rec}}}$ as in Eq. (3), we see that the distributions of $\overline{h_i^{\text{rec}}}$ concentrate around 0 (Fig 2A) and the distributions of $\sigma_{h_i^{\text{rec}}}$ concentrate around a value slightly larger than 1 (Fig 2B) as $N \to \infty$. Moreover, when N is large, the trajectories $h_i^{\text{rec}}(t)$ (and $x_i^{\text{rec}}(t)$) look like independent realizations of the same
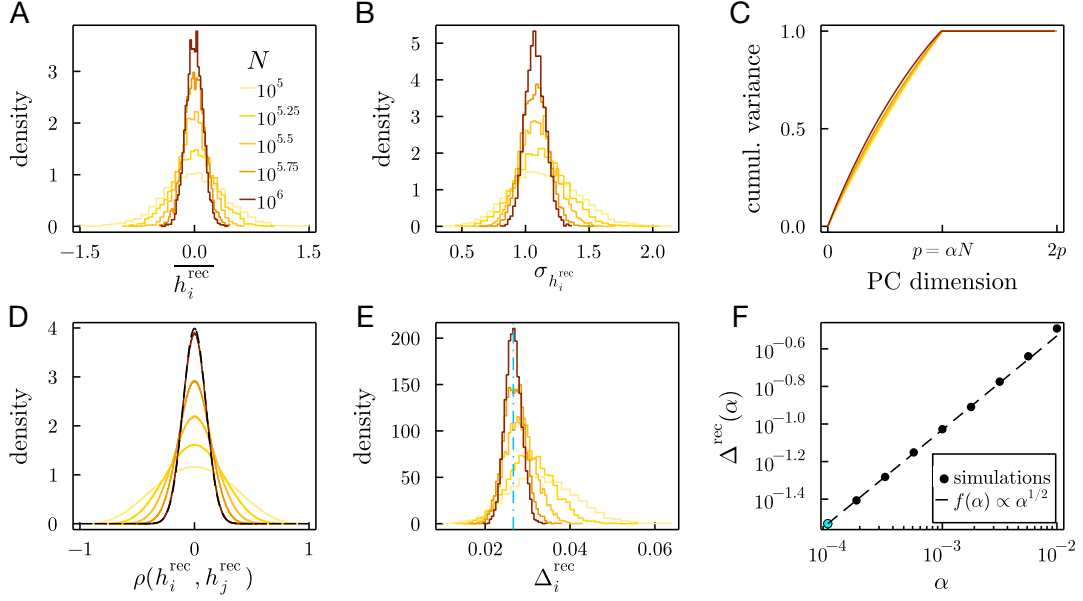
FIG. 2. (**A-E**) Scaling behavior of the disordered SNN as N increases for load $\alpha = 10^{-4}$. (**A**) Distributions the of single-neuron average potentials $\overline{h_i^{\mathrm{rec}}}$ in the SNNs concentrate around 0. (**B**) Distributions the of single-neuron potential fluctuations $\sigma_{h_i^{\mathrm{rec}}}$ concentrate around a value slightly larger than 1. (**C**) Principal component analysis of $h^{\mathrm{rec}}(t)$. The recurrent drive spans the whole $p = \alpha N$-dimensional subspace. (**D**) Distributions of correlations $\rho(h_i^{\mathrm{rec}}, h_j^{\mathrm{rec}})$. The dashed black line represents a centered normal distribution with variance $1/p$, for $N = 10^6$. As in Fig. 1C, correlations concentrate around 0. (**E**) Distributions of single-neuron distances Δ_i^{rec}. Distances concentrate around a limit distance $\Delta(\alpha) \approx 0.03$ (dashed blue line). (**F**) Numerical estimate of the limit distance $\Delta(\alpha)$ (circles) and fitted power law with exponent $1/2$ (dashed line). The value $\Delta^{\mathrm{rec}}(\alpha)$ for $\alpha = 10^{-4}$ is indicated by a blue circle and corresponds to the dashed blue line in (E). (**A-F**) All the quantities are estimated using samples of 5000 neurons. Same neuronal and noise parameters as in Fig. 1B-E.

stochastic process (Fig. 1E), which is reminiscent of the dynamic mean-field theory of random chaotic RNNs [6, 27]. The comprehensive study of this putative dynamic mean-field theory is beyond the scope of this letter.

Large networks are redundancy-free.— To explain why, in the disordered network model described above, correlations concentrate around 0 as $N \to \infty$ (Fig. 1C), we first verify numerically that the recurrent drive of the SNN is of dimension $p = \alpha N$ by performing a principal component analysis (PCA) of the trajectories $h_i^{\mathrm{rec}}(t)$ over a single trial of 100 s. The

result shows that the cumulative explained variance grows approximately linearly with the PC dimension and saturates at p (Fig. 2C), confirming that the recurrent drive homogeneously visits all p dimensions spanned by the columns of $\boldsymbol{\xi}$. This implies that the overlaps, Eq. (5), also homogeneously visit their p dimensions. In the overlap formulation of the dynamics of the SNN, Eq. (5), we see that, when α is small, the recurrent inputs of two distinct neurons i and j are approximately $\sum_{\mu=1}^{p} \xi_{i\mu} m_\mu(t)$ and $\sum_{\mu=1}^{p} \xi_{j\mu} m_\mu(t)$ respectively. Therefore, for large N, we expect the correlations between distinct neurons to approximate the correlations between the corresponding rows of $\boldsymbol{\xi}$:

$$\rho(h_i^{\text{rec}}, h_j^{\text{rec}}) \approx \frac{1}{p} \sum_{\mu=1}^{p} \xi_{i\mu}\xi_{j\mu}. \tag{7}$$

By the Central limit theorem, the approximation Eq. (7) implies that the distribution of correlations converges to a zero-mean normal distribution with variance $1/p$, which we confirm numerically (Fig. 2D). Since $p = \alpha N$, the fact that the normal distribution has variance $1/p$ is sufficient to guarantee the absence of redundancy as $N \to \infty$ (i). Indeed, for any $0 < \theta < 1$, the expected number of pairs of distinct neurons having a correlation greater or equal to θ is the product of the proportion of such pairs and the total number of pairs in the network. This product can be upper-bounded by $e^{-\theta^2 p/2} \cdot N(N-1)/2$ (using a standard Gaussian tail bound) and therefore converges to 0 as $N \to \infty$ (since $p = \alpha N$).

Large SNNs converge to large RNNs as $\alpha \to 0$.— Finally, we study the convergence of the SNN to the RNN as $\alpha \to 0$. Let us first observe that, for a fixed $\alpha > 0$ and as $N \to \infty$, the distributions of the single-neuron distances

$$\Delta_i^{\text{rec}} := \lim_{T \to \infty} \int_0^T \left| h_i^{\text{rec}}(t) - x_i^{\text{rec}}(t) \right| dt.$$

concentrate around a finite value (around 0.03 for $\alpha = 10^{-4}$, Fig. 2E). Thereby, for any $\alpha > 0$, we can define the $N \to \infty$ limit distance between the SNN and the RNN as the time average of the distance between the potential of a *typical* neuron i^* (receiving recurrent input only) in the limit SNN, $\hat{h}_{i^*}^{\text{rec}}(t)$, and the potential of the corresponding rate unit in the limit RNN, $\hat{x}_{i^*}^{\text{rec}}(t)$:

$$\Delta^{\text{rec}}(\alpha) := \lim_{T \to \infty} \frac{1}{T} \int_0^T \left| \hat{h}_{i^*}^{\text{rec}}(t) - \hat{x}_{i^*}^{\text{rec}}(t) \right| dt.$$

Numerical estimates of the limit distance $\Delta^{\text{rec}}(\alpha)$ indeed show that the limit SNN converge to the limit RNN as $\alpha \to 0$ and the power law fit of the estimated $\Delta^{\text{rec}}(\alpha)$ (Fig. 2F) indicates

9

that

$$\Delta^{\text{rec}}(\alpha) = \mathcal{O}(\sqrt{\alpha}), \quad \text{as } \alpha \to 0. \tag{8}$$

We have therefore shown, using the input-driven disordered network model, that large SNNs can converge to equally large RNNs (i) without redundancy and (ii) temporal averaging. In the absence of redundancy, what mechanism allows a SNN to approximate a RNN? Borrowing ideas from an established mean-field convergence proof technique [16, 29] (see also [30]), we present a heuristic argument suggesting that the convergence of the SNNs to the RNNs is related to the ℓ^2 norms (Euclidean norms) of the incoming synaptic weights to each neuron (the rows of the connectivity matrix) $\|\mathbf{J}_i\|_2 = \sqrt{\sum_{j=1}^{N} J_{ij}^2}$.

For simplicity, let us assume that at some initial time t, $h_i(t) = x_i(t)$, for all $i = 1, \dots, N$. Since the SNN (1) and the RNN (2) receive the same external input $I_i^{\text{ext}}(t)$, for a small time step dt, we have, for the expected difference between Eqs. (1) and (2),

$$\mathbb{E}\left[|h_i(t+dt) - x_i(t+dt)|\right] = \frac{1}{\tau} \mathbb{E}\left[\left|\sum_{j=1}^{N} J_{ij} n_j - \sum_{j=1}^{N} J_{ij}\phi(x_j(t))dt\right|\right] + o(dt),$$

where the $\{n_j\}_j$ are independent Poisson-distributed random variables with means $\{\phi(h_j(t))dt\}_j \equiv \{\phi(x_j(t))dt\}_j$. By Jensen's inequality and the independence of the random variables $\{n_j\}_j$, we get the bound

$$\mathbb{E}\left[\left|\sum_{j=1}^{N} J_{ij} n_j - \sum_{j=1}^{N} J_{ij}\phi(x_j(t))dt\right|\right]^2 \leq \sum_{j=1}^{N} J_{ij}^2 \phi(x_j(t))dt.$$

Since the transfer function ϕ is upper-bounded by $1/\tau$, we get

$$\mathbb{E}\left[|h_i(t+dt) - x_i(t+dt)|\right] \leq \frac{1}{\tau}\|\mathbf{J}_i\|_2 \sqrt{\frac{dt}{\tau}} + o(dt). \tag{9}$$

The bound (9) implies that if for all i the ℓ^2 norm of the incoming synaptic weights, $\|\mathbf{J}_i\|_2$, is small, then the distance between the SNN (1) and the RNN (2) remains small (at least over a short time). In our disorder network model, the reason why the distance between the SNN and the RNN does not diverge over time is likely due to the fact that we are in an input-driven regime.

Importantly, the bound (9), which holds for any arbitrary connectivity matrix $\mathbf{J}$, is consistent with the idea that spike noise absorption in large networks – enabling a stochastic

SNN to behave like a deterministic RNN – does not require redundancy (as we have numerically shown with our the disordered network example).

Classical mean-field limit results where synaptic weights are scaled by $1/N$ (i), e.g. for interacting homogeneous populations [14, 15] or spatially structured networks [15], can be rigorously proven using bounds similar to Eq. (9) [16–18, 31, 32]. Indeed, if weights are scaled by $1/N$, then $\|\mathbf{J}_i\|_2^2 = \mathcal{O}(1/N)$ and the bound (9) tends to 0 as $N \to \infty$ (see [16–18] for full mean-field limit convergence proofs).

By contrast, in the case of our disordered network model, the synaptic weights $\{J_{ij}\}_{i \neq j}$ defined in Eq. (4) scale as $\mathcal{O}(1/\sqrt{N})$ and the distribution of the ℓ^2 norms $\|\mathbf{J}_i\|_2$ concentrate around $\sqrt{\alpha/c}$, meaning that for a typical neuron i^* in a large network, $\|\mathbf{J}_{i^*}\|_2 = \sqrt{\alpha/c}$. Hence, as $\alpha \to 0$, the limit distance between a large SNN and a large RNN $\Delta^{\mathrm{rec}}(\alpha)$, Eq. (8), tends to zero at the same speed as the ℓ^2 norm $\|\mathbf{J}_{i^*}\|_2$, which supports the claim that the bound (9) explains (at least heuristically) spike noise absorption in redundancy-free networks. Whereas classical mean-field models rely on redundancy and the law of large numbers to absorb spike noise, the bound (9) suggests that the "independence" of neurons and the concentration of measure phenomenon (which only require variables to be independent [33]) are sufficient for redundancy-free networks to absorb spike noise; this heuristic observation should be confirmed by future mathematical proofs (see [30] for a first step).

More generally, absorbing spike noise by imposing small ℓ^2 norms $\|\mathbf{J}_i\|_2$ (see Eq. (9)) while keeping a nontrivial recurrent drive requires networks to be densely connected (fixed *fraction* of nonzero incoming synaptic weights as $N \to \infty$), as opposed to sparsely connected (fixed *number* of nonzero incoming synaptic weights as $N \to \infty$) [34, 35]. Note however that not all densely connected networks absorb spike noise. For example, in random chaotic networks [6], the ℓ^2 norms $\|\mathbf{J}_i\|_2$ are not small and spike noise is not absorbed [36]. This shows how the ℓ^2 norms of the incoming synaptic weights can help distinguish spike noise-absorbing dense networks, which approximate rate-based dynamics, from dense networks where spike noise plays a significant role in network dynamics.

While the disordered network models considered here are redundancy-free, it can be argued that they are redundant in a weaker sense: the dimensionality $p = \alpha N$ of the recurrent drive has to be small in proportion to the number of neurons N for the SNN to approximate the RNN. Our disordered network model therefore shows a trade-off between spike noise absorption and the dimensionality of noise-robust recurrent dynamics. Whether this trade-off

11

is a general feature of spiking neural networks is an open theoretical question we leave for future work. Such a tradeoff could shed light on how the noisy wetware of the brain [37] constrains the implementation of "computation through neural population dynamics" [10].

————

[1] W. Gerstner and W. M. Kistler, *Spiking neuron models: Single neurons, populations, plasticity* (Cambridge university press, 2002).

[2] P. J. Werbos, Backpropagation through time: what it does and how to do it, Proceedings of the IEEE **78**, 1550 (1990).

[3] D. Sussillo and O. Barak, Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks, Neural computation **25**, 626 (2013).

[4] O. Barak, Recurrent neural networks as versatile tools of neuroscience research, Current opinion in neurobiology **46**, 1 (2017).

[5] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Storing infinite numbers of patterns in a spin-glass model of neural networks, Physical Review Letters **55**, 1530 (1985).

[6] H. Sompolinsky, A. Crisanti, and H.-J. Sommers, Chaos in random neural networks, Physical review letters **61**, 259 (1988).

[7] U. Pereira and N. Brunel, Attractor dynamics in networks with learning rules inferred from in vivo data, Neuron (2018).

[8] F. Mastrogiuseppe and S. Ostojic, Linking connectivity, dynamics, and computations in low-rank recurrent neural networks, Neuron **99**, 609 (2018).

[9] K. V. Shenoy, M. Sahani, M. M. Churchland, *et al.*, Cortical control of arm movements: a dynamical systems perspective, Annu Rev Neurosci **36**, 337 (2013).

[10] S. Vyas, M. D. Golub, D. Sussillo, and K. V. Shenoy, Computation through neural population dynamics, Annual Review of Neuroscience **43**, 249 (2020).

[11] R. Brette, Philosophy of the spike: rate-based vs. spike-based theories of the brain, Frontiers in systems neuroscience **9**, 151 (2015).

[12] E. Chichilnisky, A simple white noise analysis of neuronal light responses, Network: computation in neural systems **12**, 199 (2001).

[13] P. Brémaud and L. Massoulié, Stability of nonlinear Hawkes processes, Ann. Probab. **24**, 1563 (1996).

[14] W. Gerstner and J. L. van Hemmen, Associative memory in a network of 'spiking'neurons, Netw. Comput. Neural Syst. **3**, 139 (1992).

[15] W. Gerstner, Time structure of the activity in neural network models, Phys. Rev. E **51**, 738 (1995).

[16] S. Delattre, N. Fournier, and M. Hoffmann, Hawkes processes on large networks, Ann. Appl. Probab. **26**, 216 (2016).

[17] S. Ditlevsen and E. Löcherbach, Multi-class oscillating systems of interacting neurons, Stochastic Process. Appl. **127**, 1840 (2017).

[18] J. Chevallier, A. Duarte, E. Löcherbach, and G. Ost, Mean field limits for nonlinear spatially extended hawkes processes with exponential memory kernels, Stochastic Processes and their Applications **129**, 1 (2019).

[19] T. G. Kurtz, Limit theorems for sequences of jump markov processes approximating ordinary differential processes, Journal of Applied Probability **8**, 344 (1971).

[20] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. Siegelbaum, A. J. Hudspeth, S. Mack, *et al.*, *Principles of neural science*, Vol. 4 (McGraw-hill New York, 2000).

[21] S. Thorpe, D. Fize, and C. Marlot, Speed of processing in the human visual system, nature **381**, 520 (1996).

[22] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proceedings of the national academy of sciences **79**, 2554 (1982).

[23] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Spin-glass models of neural networks, Physical Review A **32**, 1007 (1985).

[24] D. J. Amit, *Modeling brain function: The world of attractor neural networks* (Cambridge university press, 1989).

[25] J. Hertz, A. Krogh, and R. G. Palmer, Introduction to the theory of neural computation (1991).

[26] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal dynamics: From single neurons to networks and models of cognition* (Cambridge University Press, 2014).

[27] M. Helias and D. Dahmen, *Statistical Field Theory for Neural Networks* (Springer, 2020).

[28] W. Gerstner, Population dynamics of spiking neurons: fast transients, asynchronous states, and locking, Neural Comput. **12**, 43 (2000).

[29] A.-S. Sznitman, Topics in propagation of chaos, in *École d'Été de Probabilités de Saint-Flour XIX—1989*, Lecture Notes in Math., Vol. 1464 (Springer, Berlin, 1991) pp. 165–251.

[30] P.-E. Jabin, D. Poyato, and J. Soler, Mean-field limit of non-exchangeable systems, arXiv preprint arXiv:2112.15406 (2021).

[31] N. Fournier and E. Löcherbach, On a toy model of interacting neurons, in *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, Vol. 52 (Institut Henri Poincaré, 2016) pp. 1844–1876.

[32] J. Chevallier, Mean-field limit of generalized Hawkes processes, Stochastic Process. Appl. **127**, 3870 (2017).

[33] M. Talagrand, A new look at independence, The Annals of probability , 1 (1996).

[34] N. Brunel and V. Hakim, Fast global oscillations in networks of integrate-and-fire neurons with low firing rates, Neural computation **11**, 1621 (1999).

[35] N. Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, Journal of computational neuroscience **8**, 183 (2000).

[36] A. van Meegen and S. J. van Albada, Microscopic theory of intrinsic timescales in spiking neural networks, Physical Review Research **3**, 043077 (2021).

[37] A. A. Faisal, L. P. Selen, and D. M. Wolpert, Noise in the nervous system, Nature reviews neuroscience **9**, 292 (2008).

Ring model

For simplicity, in Fig. 1A, we simulate a ring model with spatially structured stochastic external input but without recurrent connections. For all $i = 1, \ldots, N$,

$$\tau \frac{d}{dt} x_i(t) = -x_i(t) + I_i^{\text{ext}}(t),$$
$$I_i^{\text{ext}}(t) = \cos\left(2\pi \left(\frac{i}{N} + \Theta(t)\right)\right)$$

with

$$\frac{d}{dt} \Theta(t) = \tilde{\sigma} \eta(t),$$

where $\eta(t)$ is a Gaussian white noises and $\tilde{\sigma} = 10^{-3}$.

7 Discussion and conclusion

A well-established working hypothesis in neuroscience posits that computation in the brain rests emergent noise-robust population dynamics in large networks of noisy neurons (Gerstner, Kistler, et al. 2014; Vyas et al. 2020). This hypothesis is shared by theorists, who build bottom-up, mechanistic models of neuronal population dynamics via classical mean-field theory (Gerstner and van Hemmen 1992; Wang 2002; Wong and Wang 2006), and data analysts, who model multi-neuronal recordings using latent variable models (Duncker et al. 2019; Kim et al. 2021; Macke et al. 2012). In Chapter 5, we showed the conceptual link between classical mean-field theory and latent variable models.

Classical mean-field limits for networks of renewal-type neurons provide examples where the dynamics of networks of noisy spiking neurons converge, in the mean-field limit, to a deterministic neuronal population equation (Gerstner 2000; Ostojic, Brunel, and Hakim 2009). In Chapter 2, I showed that mean-field limit proofs for such models can be generalized to a large class of "nonrenewal" neuron models, that I called age- and leaky memory-dependent Hawkes processes. We then analysed the long time behavior of the limit neuronal population equation in Chapter 3.

In classical mean-field models (see Introduction and Chapter 2), large networks absorb neuronal noise by applying the law of large numbers: the noise of independent and identically distributed neurons is averaged out by the dense connectivity of the network. The problem with using the law of large numbers to absorb neuronal noise is that it implies an unrealistic form of redundancy. In Chapter 6, I showed numerically that redundancy was actually unnecessary. Without redundancy and the law of large numbers, how can large networks absorb neuronal noise? As the informal argument at the end of Chapter 6 suggests, *concentration of measure* is sufficient. Studying the effect of neuronal noise in large networks of stochastic spiking neurons through the lens of concentration of measure is an exciting direction for future theoretical work. In a way, this would be the natural direction to take for going beyond classical mean-field models since the law of large number is (roughly speaking) only a special case of the concentration of measure phenomenon. Although concentration of measure was not thoroughly treated in Chapter 6, we can nevertheless extract from the given example a

heuristic strategy for taming neuronal noise in large networks. First, if the direct effect of any single presynaptic neuron on any postsynaptic neuron vanishes in large networks (dense connectivity), we get *propagation of independence* (borrowing the expression of Jabin, Poyato, and Soler (2021)), i.e., neurons behave like independent – but not necessarily identically distributed – processes. Then, since independence is the key ingredient for the concentration of measure, we can search for quantities whose measure (probability distribution) concentrates. Finally, if the interaction between neurons can be reduced to these almost-deterministic quantities, reliable "computation through neural population dynamics" (Vyas et al. 2020) is theoretically possible.

Concentration of measure also appear in deterministic networks of rate units with quenched disorder; landmark results on the statics (Amit, Gutfreund, and Sompolinsky 1985b; Gardner 1988) and dynamics (Sompolinsky, Crisanti, and Sommers 1988) of disordered neural networks have been proven by probability theorists using the idea of concentration of measure (large deviations in particular) (Arous and Guionnet 1995, 1997; Bovier and Picco 2012; Guionnet 1997; Moynot and Samuelides 2002; Talagrand 2010, 2011). The spiking disordered network described in Chapter 6 is an example of model with both quenched disorder and neuronal noise. The rigorous mathematical study of such models would generalize two currently distinct families of mean-field models: mean-field models for networks of stochastic spiking neurons (where synaptic weights are scaled by $1/N$) and mean-field models for disordered networks of rate units (where synaptic weights are scaled by $1/\sqrt{N}$).

Regarding biological realism, all exact mean-field models studied or cited in this thesis share the common limitation of being models of densely connected networks. In densely connected networks, no matter the scaling of the synaptic weights ($1/N$ or $1/\sqrt{N}$), in the mean-field limit $N \to \infty$, the direct effect of any single presynaptic neuron on any postsynaptic neuron vanishes. This feature is incompatible with experiments showing that the stimulation of a single neuron (or a few neurons) can systematically affect other neurons (Carrillo-Reid, Yang, et al. 2016; Jouhanneau, Kremkow, and Poulet 2018; London, Roth, et al. 2010), brain state (Li, Poo, and Dan 2009), and even behavior (Brecht et al. 2004; Carrillo-Reid, Han, et al. 2019; Dalgleish et al. 2020). Single neuron stimulation experiments therefore suggest that neural circuits are, to a certain extent, sparsely connected, that is, a single spike can have a non-negligible effect on some postsynaptic neurons. The problem is that there is no obvious propagation of independence in sparsely connected networks and independence is critical for the concentration of measure. Rigorous mathematical results on the effects of neuronal noise in sparse networks are, to my knowledge, lacking. Such theoretical results would be of interest to many experimentalists currently using two-photon optogenetics for targeted stimulation of single (or few) neurons *in vivo* (Adesnik and Abdeladim 2021; Russell et al. 2022; Yang et al. 2018). Other important biological neuronal features have been neglected in the models considered in this thesis, notably, dendritic dynamics (London and Häusser 2005; Poirazi and Papoutsi 2020) and synaptic noise (Rusakov, Savtchenko, and Latham 2020). Developing mathematically tractable models including these features would contribute to our understanding of how neuronal noise is tamed in the brain.

Starting with the study of classical mean-field models for networks of noisy spiking neurons, this thesis ends with an hypothesis. Noise-robust population dynamics in classical mean-field models – taking the form of deterministic neuronal population equations – is only a very special case of a more general mechanism relying on two probabilistic ingredients: *independence* and *concentration of measure*. Studying the interplay between these two phenomena in mathematical models of biological neural networks could help us understand how neuronal noise constrains the design of brain circuits and how reliable computation emerges from large networks of unreliable neurons.

Bibliography

Adesnik, Hillel and Lamiae Abdeladim (2021). "Probing neural codes with two-photon holographic optogenetics". In: *Nature Neuroscience* 24.10, pp. 1356–1366.

Aihara, Kazuyuki, Gen Matsumoto, and Yuhji Ikegaya (1984). "Periodic and non-periodic responses of a periodically forced Hodgkin-Huxley oscillator". In: *Journal of theoretical biology* 109.2, pp. 249–269.

Amari, Shun-ichi (1977). "Dynamics of pattern formation in lateral-inhibition type neural fields". In: *Biological cybernetics* 27.2, pp. 77–87.

Amit, Daniel J (1989). *Modeling brain function: The world of attractor neural networks*. Cambridge university press.

Amit, Daniel J, Hanoch Gutfreund, and Haim Sompolinsky (1985a). "Spin-glass models of neural networks". In: *Physical Review A* 32.2, p. 1007.

– (1985b). "Storing infinite numbers of patterns in a spin-glass model of neural networks". In: *Physical Review Letters* 55.14, p. 1530.

– (1987). "Statistical mechanics of neural networks near saturation". In: *Annals of physics* 173.1, pp. 30–67.

Arieli, Amos et al. (1996). "Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses". In: *Science* 273.5283, pp. 1868–1871.

Arous, G Ben and Alice Guionnet (1995). "Large deviations for Langevin spin glass dynamics". In: *Probability Theory and Related Fields* 102.4, pp. 455–509.

– (1997). "Symmetric Langevin spin glass dynamics". In: *The Annals of Probability* 25.3, pp. 1367–1422.

Bovier, Anton and Pierre Picco (2012). *Mathematical aspects of spin glasses and neural networks*. Vol. 41. Springer Science & Business Media.

Breakspear, Michael (2017). "Dynamic models of large-scale brain activity". In: *Nature neuroscience* 20.3, pp. 340–352.

Brecht, Michael et al. (2004). "Whisker movements evoked by stimulation of single pyramidal cells in rat motor cortex". In: *Nature* 427.6976, pp. 704–710.

Brémaud, Pierre and Laurent Massoulié (1996). "Stability of nonlinear Hawkes processes". In: *Ann. Probab.* 24.3, pp. 1563–1588. ISSN: 0091-1798. DOI: 10.1214/aop/1065725193. URL: https://doi.org/10.1214/aop/1065725193.

Brette, Romain and Wulfram Gerstner (2005). "Adaptive exponential integrate-and-fire model as an effective description of neuronal activity". In: *Journal of neurophysiology* 94.5, pp. 3637–3642.

Brunel, Nicolas (2000). "Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons". In: *Journal of computational neuroscience* 8.3, pp. 183–208.

Brunel, Nicolas and Vincent Hakim (1999). "Fast global oscillations in networks of integrate-and-fire neurons with low firing rates". In: *Neural computation* 11.7, pp. 1621–1671.

Brunel, Nicolas and Mark CW Van Rossum (2007). "Lapicque's 1907 paper: from frogs to integrate-and-fire". In: *Biological cybernetics* 97.5, pp. 337–339.

Bryant, Hugh L and José P Segundo (1976). "Spike initiation by transmembrane current: a white-noise analysis." In: *The Journal of physiology* 260.2, pp. 279–314.

Caceres, Maria J and Benoît Perthame (2014). "Beyond blow-up in excitatory integrate and fire neuronal networks: refractory period and spontaneous activity". In: *Journal of theoretical biology* 350, pp. 81–89.

Cáceres, Marıa J, José A Carrillo, and Benoît Perthame (2011). "Analysis of nonlinear noisy integrate & fire neuron models: blow-up and steady states". In: *The Journal of Mathematical Neuroscience* 1.1, pp. 1–33.

Cañizo, José A and Stéphane Mischler (2021). "Harris-type results on geometric and subgeometric convergence to equilibrium for stochastic semigroups". In: *arXiv preprint arXiv:2110.09650*.

Cañizo, José A and Havva Yoldaş (2019). "Asymptotic behaviour of neuron population models structured by elapsed-time". In: *Nonlinearity* 32.2, p. 464.

Carrillo, José A, Marıa d M González, et al. (2013). "Classical solutions for a nonlinear Fokker-Planck equation arising in computational neuroscience". In: *Communications in Partial Differential Equations* 38.3, pp. 385–409.

Carrillo, José A, Benoît Perthame, et al. (2015). "Qualitative properties of solutions for the noisy integrate and fire model in computational neuroscience". In: *Nonlinearity* 28.9, p. 3365.

Carrillo-Reid, Luis, Shuting Han, et al. (2019). "Controlling visually guided behavior by holographic recalling of cortical ensembles". In: *Cell* 178.2, pp. 447–457.

Carrillo-Reid, Luis, Weijian Yang, et al. (2016). "Imprinting and recalling cortical ensembles". In: *Science* 353.6300, pp. 691–694.

Chay, Teresa R and John Rinzel (1985). "Bursting, beating, and chaos in an excitable membrane model". In: *Biophysical Journal* 47.3, pp. 357–366.

Chevallier, Julien et al. (2017). "Fluctuations for mean-field interacting age-dependent Hawkes processes". In: *Electronic Journal of Probability* 22.

Chevallier, Julien (2017). "Mean-field limit of generalized Hawkes processes". In: *Stochastic Process. Appl.* 127.12, pp. 3870–3912. ISSN: 0304-4149.

Chevallier, Julien, A Duarte, et al. (2019). "Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels". In: *Stochastic Processes and their Applications* 129.1, pp. 1–27.

Chevallier, Julien and Guilherme Ost (2020). "Fluctuations for spatially extended Hawkes processes". In: *Stochastic Processes and their Applications* 130.9, pp. 5510–5542.

Chichilnisky, EJ (2001). "A simple white noise analysis of neuronal light responses". In: *Network: computation in neural systems* 12.2, p. 199.

Chow, Carson C and John A White (1996). "Spontaneous action potentials due to channel fluctuations". In: *Biophysical journal* 71.6, pp. 3013–3021.

Cohen, Michael A and Stephen Grossberg (1983). "Absolute stability of global pattern formation and parallel memory storage by competitive neural networks". In: *IEEE transactions on systems, man, and cybernetics* 5, pp. 815–826.

Coombes, Stephen (2010). "Large-scale neural dynamics: simple and complex". In: *NeuroImage* 52.3, pp. 731–739.

Cormier, Quentin (2020). "A mean-field model of Integrate-and-Fire neurons: non-linear stability of the stationary solutions". In: *arXiv preprint arXiv:2002.08649*.

Cormier, Quentin, Etienne Tanré, and Romain Veltz (2020). "Long time behavior of a mean-field model of interacting neurons". In: *Stochastic Processes and their Applications* 130.5, pp. 2553–2595.

– (2021). "Hopf bifurcation in a mean-field model of spiking neurons". In: *Electron. J. Probab.* 26, Paper No. 121, 40.

Cox, David Roxbee (1962). *Renewal theory*. Methuen.

Dalgleish, Henry WP et al. (2020). "How many neurons are sufficient for perception of cortical activity?" In: *Elife* 9, e58889.

De Masi, Anna et al. (2015). "Hydrodynamic limit for interacting neurons". In: *J. Stat. Phys.* 158.4, pp. 866–902. ISSN: 0022-4715.

Deco, Gustavo et al. (2008). "The dynamic brain: from spiking neurons to neural masses and cortical fields". In: *PLoS computational biology* 4.8, e1000092.

Delarue, François et al. (2015a). "Global solvability of a networked integrate-and-fire model of McKean–Vlasov type". In: *The Annals of Applied Probability* 25.4, pp. 2096–2133.

– (2015b). "Particle systems with a singular mean-field self-excitation. Application to neuronal networks". In: *Stochastic Processes and their Applications* 125.6, pp. 2451–2492.

Delattre, Sylvain, Nicolas Fournier, and Marc Hoffmann (2016). "Hawkes processes on large networks". In: *Ann. Appl. Probab.* 26.1, pp. 216–261. ISSN: 1050-5164. DOI: 10.1214/14-AAP1089. URL: https://doi.org/10.1214/14-AAP1089.

Dumont, Grégory, Alberto Pérez-Cervera, and Boris Gutkin (2022). "A framework for macroscopic phase-resetting curves for generalised spiking neural networks". In: *PLoS computational biology* 18.8, e1010363.

Duncker, Lea et al. (2019). "Learning interpretable continuous-time models of latent stochastic dynamical systems". In: *International Conference on Machine Learning*. PMLR, pp. 1726–1734.

Faisal, A Aldo, Luc PJ Selen, and Daniel M Wolpert (2008). "Noise in the nervous system". In: *Nature reviews neuroscience* 9.4, pp. 292–303.

Faisal, A Aldo, John A White, and Simon B Laughlin (2005). "Ion-channel noise places limits on the miniaturization of the brain's wiring". In: *Current Biology* 15.12, pp. 1143–1149.

Fournier, Nicolas and Eva Löcherbach (2016). "On a toy model of interacting neurons". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*. Vol. 52. Institut Henri Poincaré, pp. 1844–1876.

Galves, Antonio and Eva Löcherbach (2013). "Infinite systems of interacting chains with memory of variable length—a stochastic model for biological neural nets". In: *Journal of Statistical Physics* 151.5, pp. 896–921.

– (2016). "Modeling networks of spiking neurons as interacting processes with memory of variable length". In: *J. SFdS* 157.1, pp. 17–32.

Gardner, Elizabeth (1988). "The space of interactions in neural network models". In: *Journal of physics A: Mathematical and general* 21.1, p. 257.

Gardner, Elizabeth and Bernard Derrida (1988). "Optimal storage properties of neural network models". In: *Journal of Physics A: Mathematical and general* 21.1, p. 271.

Gerstner, Wulfram (1995). "Time structure of the activity in neural network models". In: *Phys. Rev. E* 51.1, p. 738.

– (2000). "Population dynamics of spiking neurons: fast transients, asynchronous states, and locking". In: *Neural Comput.* 12.1, pp. 43–89.

Gerstner, Wulfram and Werner M Kistler (2002). *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press.

Gerstner, Wulfram, Werner M Kistler, et al. (2014). *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press.

Gerstner, Wulfram and J Leo van Hemmen (1992). "Associative memory in a network of 'spiking' neurons". In: *Netw. Comput. Neural Syst.* 3.2, pp. 139–164.

Guckenheimer, John and Ricardo A Oliva (2002). "Chaos in the Hodgkin–Huxley model". In: *SIAM Journal on Applied Dynamical Systems* 1.1, pp. 105–114.

Guionnet, Alice (1997). "Averaged and quenched propagation of chaos for spin glass dynamics". In: *Probability Theory and Related Fields* 109.2, pp. 183–215.

Harsch, Annette and Hugh PC Robinson (2000). "Postsynaptic variability of firing in rat cortical neurons: the roles of input synchronization and synaptic NMDA receptor conductance". In: *Journal of Neuroscience* 20.16, pp. 6181–6192.

Hodgkin, Alan L and Andrew F Huxley (1952). "A quantitative description of membrane current and its application to conduction and excitation in nerve". In: *The Journal of physiology* 117.4, pp. 500–544.

Hopfield, John J (1982). "Neural networks and physical systems with emergent collective computational abilities". In: *Proceedings of the national academy of sciences* 79.8, pp. 2554–2558.

– (1984). "Neurons with graded response have collective computational properties like those of two-state neurons". In: *Proceedings of the national academy of sciences* 81.10, pp. 3088–3092.

Inglis, James and Denis Talay (2015). "Mean-field limit of a stochastic particle system smoothly interacting through threshold hitting-times and applications to neural networks with dendritic component". In: *SIAM Journal on Mathematical Analysis* 47.5, pp. 3884–3916.

Jabin, Pierre-Emmanuel, David Poyato, and Juan Soler (2021). "Mean-field limit of non-exchangeable systems". In: *arXiv preprint arXiv:2112.15406.*

Jansen, Ben H and Vincent G Rit (1995). "Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns". In: *Biological cybernetics* 73.4, pp. 357–366.

Jolivet, Renaud, Timothy J Lewis, and Wulfram Gerstner (2004). "Generalized integrate-and-fire models of neuronal activity approximate spike trains of a detailed model to a high degree of accuracy". In: *Journal of neurophysiology* 92.2, pp. 959–976.

Jolivet, Renaud, Alexander Rauch, et al. (2006). "Predicting spike timing of neocortical pyramidal neurons by simple threshold models". In: *Journal of computational neuroscience* 21.1, pp. 35–49.

Jouhanneau, Jean-Sébastien, Jens Kremkow, and James FA Poulet (2018). "Single synaptic inputs drive high-precision action potentials in parvalbumin expressing GABA-ergic cortical neurons in vivo". In: *Nature Communications* 9.1, pp. 1–11.

Kim, Timothy D et al. (2021). "Inferring Latent Dynamics Underlying Neural Population Activity via Neural Differential Equations". In: *International Conference on Machine Learning*. PMLR, pp. 5551–5561.

Kingman, John Frank Charles (1992). *Poisson processes*. Vol. 3. Clarendon Press.

Kistler, Werner M, Wulfram Gerstner, and J Leo van Hemmen (1997). "Reduction of the Hodgkin-Huxley equations to a single-variable threshold model". In: *Neural computation* 9.5, pp. 1015–1045.

Knight, Bruce W (1972a). "Dynamics of encoding in a population of neurons". In: *The Journal of general physiology* 59.6, pp. 734–766.

– (1972b). "The relationship between the firing rate of a single neuron and the level of activity in a population of neurons: Experimental evidence for resonant enhancement in the population response". In: *The Journal of general physiology* 59.6, pp. 767–778.

Lapique, Louis (1907). "Recherches quantitatives sur l'excitation electrique des nerfs traitee comme une polarization." In: *Journal of Physiology and Pathololgy* 9, pp. 620–635.

Laughlin, Simon B, Rob R de Ruyter van Steveninck, and John C Anderson (1998). "The metabolic cost of neural information". In: *Nature neuroscience* 1.1, pp. 36–41.

Li, Cheng-yu T, Mu-ming Poo, and Yang Dan (2009). "Burst spiking of a single cortical neuron modifies global brain state". In: *Science* 324.5927, pp. 643–646.

Löcherbach, Eva (2022). "Fluctuations for mean field limits of interacting systems of spiking neurons". In: *arXiv preprint arXiv:2201.09255.*

London, Michael and Michael Häusser (2005). "Dendritic computation". In: *Annu. Rev. Neurosci.* 28, pp. 503–532.

London, Michael, Arnd Roth, et al. (2010). "Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex". In: *Nature* 466.7302, pp. 123–127.

Macke, Jakob H et al. (2012). "Empirical models of spiking in neural populations". In: *Advances in Neural Information Processing Systems 24: 25th conference on Neural Information Processing Systems (NIPS 2011)*, pp. 1350–1358.

Mainen, Zachary F and Terrence J Sejnowski (1995). "Reliability of spike timing in neocortical neurons". In: *Science* 268.5216, pp. 1503–1506.

Mante, Valerio et al. (2013). "Context-dependent computation by recurrent dynamics in prefrontal cortex". In: *nature* 503.7474, pp. 78–84.

Mischler, Stéphane, Cristobal Quiñinao, and Qilong Weng (2018). "Weak and strong connectivity regimes for a general time elapsed neuron network model". In: *Journal of Statistical Physics* 173.1, pp. 77–98.

Mischler, Stéphane and Qilong Weng (2018). "Relaxation in time elapsed neuron network models in the weak connectivity regime". In: *Acta Applicandae Mathematicae* 157.1, pp. 45–74.

Moynot, Olivier and Manuel Samuelides (2002). "Large deviations and mean-field theory for asymmetric random recurrent neural networks". In: *Probability Theory and Related Fields* 123.1, pp. 41–75.

Ostojic, Srdjan and Nicolas Brunel (2011). "From spiking neuron models to linear-nonlinear models". In: *PLoS computational biology* 7.1, e1001056.

Ostojic, Srdjan, Nicolas Brunel, and Vincent Hakim (2009). "Synchronization properties of networks of electrically coupled neurons in the presence of noise and heterogeneities". In: *Journal of computational neuroscience* 26.3, pp. 369–392.

Pakdaman, Khashayar, Benoît Perthame, and Delphine Salort (2010). "Dynamics of a structured neuron population". In: *Nonlinearity* 23.1, pp. 55–75. ISSN: 0951-7715.

– (2013). "Relaxation and self-sustained oscillations in the time elapsed neuron network model". In: *SIAM J. Appl. Math.* 73.3, pp. 1260–1279. ISSN: 0036-1399.

– (2014). "Adaptation and fatigue model for neuron networks and large time asymptotics in a nonlinear fragmentation equation". In: *J. Math. Neurosci.* 4, Art. 14, 26.

Pakdaman, Khashayar, Michele Thieullen, and Gilles Wainrib (2010). "Fluid limit theorems for stochastic hybrid systems with application to neuron models". In: *Advances in Applied Probability* 42.3, pp. 761–794.

Paninski, Liam (2004). "Maximum likelihood estimation of cascade point-process neural encoding models". In: *Network: Computation in Neural Systems* 15.4, pp. 243–262.

Pillow, Jonathan W et al. (2008). "Spatio-temporal correlations and visual signalling in a complete neuronal population". In: *Nature* 454.7207, pp. 995–999.

Plesser, Hans E and Wulfram Gerstner (2000). "Noise in integrate-and-fire neurons: from stochastic input to escape rates". In: *Neural computation* 12.2, pp. 367–384.

Poirazi, Panayiota and Athanasia Papoutsi (2020). "Illuminating dendritic function with computational models". In: *Nature Reviews Neuroscience* 21.6, pp. 303–321.

Pozzorini, Christian et al. (2015). "Automated high-throughput characterization of single neurons by means of simplified spiking models". In: *PLoS Comput Biol* 11.6, e1004275.

Quiñinao, Cristóbal (2016). "A microscopic spiking neuronal network for the age-structured model". In: *Acta Appl. Math.* 146, pp. 29–55. ISSN: 0167-8019. DOI: 10.1007/s10440-016-0056-3. URL: https://doi.org/10.1007/s10440-016-0056-3.

Raad, Mads Bonde, Susanne Ditlevsen, and Eva Löcherbach (2020). "Stability and mean-field limits of age dependent Hawkes processes". In: *Ann. Inst. Henri Poincaré Probab. Stat.* 56.3,

pp. 1958–1990. ISSN: 0246-0203. DOI: 10.1214/19-AIHP1023. URL: https://doi.org/10.1214/19-AIHP1023.

Rusakov, Dmitri A, Leonid P Savtchenko, and Peter E Latham (2020). "Noisy synaptic conductance: Bug or a feature?" In: *Trends in Neurosciences* 43.6, pp. 363–372.

Russell, Lloyd E et al. (2022). "All-optical interrogation of neural circuits in behaving mice". In: *Nature Protocols*, pp. 1–42.

Sacerdote, Laura and Maria Teresa Giraudo (2013). "Stochastic integrate and fire models: a review on mathematical methods and their applications". In: *Stochastic biomathematical models*, pp. 99–148.

Sakmann, Bert and Erwin Neher (1984). "Patch clamp techniques for studying ionic channels in excitable membranes". In: *Annual review of physiology* 46.1, pp. 455–472.

Schneidman, Elad, Barry Freedman, and Idan Segev (1998). "Ion channel stochasticity may be critical in determining the reliability and precision of spike timing". In: *Neural computation* 10.7, pp. 1679–1703.

Schreiber, Susanne et al. (2004). "Influence of ionic conductances on spike timing reliability of cortical neurons for suprathreshold rhythmic inputs". In: *Journal of neurophysiology* 91.1, pp. 194–205.

Schwalger, Tilo (2021). "Mapping input noise to escape noise in integrate-and-fire neurons: a level-crossing approach". In: *Biological Cybernetics* 115.5, pp. 539–562.

Schwalger, Tilo and Anton V Chizhov (2019). "Mind the last spike—firing rate models for mesoscopic populations of spiking neurons". In: *Curr. Opin. Neurobiol.* 58, pp. 155–166.

Schwalger, Tilo, Moritz Deger, and Wulfram Gerstner (2017). "Towards a theory of cortical columns: From spiking neurons to interacting neural populations of finite size". In: *PLoS computational biology* 13.4, e1005507.

Shadlen, Michael N and William T Newsome (1994). "Noise, neural codes and cortical organization". In: *Current opinion in neurobiology* 4.4, pp. 569–579.

Shenoy, Krishna V, Maneesh Sahani, Mark M Churchland, et al. (2013). "Cortical control of arm movements: a dynamical systems perspective". In: *Annu Rev Neurosci* 36.1, pp. 337–359.

Simoncelli, Eero P et al. (2004). "Characterization of neural responses with stochastic stimuli". In: *The cognitive neurosciences* 3.327-338, p. 1.

Sompolinsky, Haim, Andrea Crisanti, and Hans-Jurgen Sommers (1988). "Chaos in random neural networks". In: *Physical review letters* 61.3, p. 259.

Stein, Richard B (1965). "A theoretical analysis of neuronal variability". In: *Biophysical Journal* 5.2, pp. 173–194.

Strassberg, Adam F and Louis J DeFelice (1993). "Limitations of the Hodgkin-Huxley formalism: Effects of single channel kinetics on transmembrane voltage dynamics". In: *Neural computation* 5.6, pp. 843–855.

Stringer, Carsen et al. (2019). "Spontaneous behaviors drive multidimensional, brainwide activity". In: *Science* 364.6437, eaav7893.

Sznitman, Alain-Sol (1991). "Topics in propagation of chaos". In: *École d'Été de Probabilités de Saint-Flour XIX—1989*. Vol. 1464. Lecture Notes in Math. Springer, Berlin, pp. 165–251. DOI: 10.1007/BFb0085169. URL: https://doi.org/10.1007/BFb0085169.

Talagrand, Michel (2010). *Mean field models for spin glasses: Volume I: Basic examples*. Vol. 54. Springer Science & Business Media.

– (2011). *Mean Field Models for Spin Glasses: Volume II: Advanced Replica-Symmetry and Low Temperature*. Vol. 55. Springer Science & Business Media.

Torres, Nicolás, Benoît Perthame, and Delphine Salort (2022). "A multiple time renewal equation for neural assemblies with elapsed time model". In: *Nonlinearity* 35.10, p. 5051.

Truccolo, Wilson et al. (2005). "A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects". In: *J. Neurophysiol.* 93.2, pp. 1074–1089.

Tuckwell, Henry C (1988). *Introduction to theoretical neurobiology: nonlinear and stochastic theories*. Tech. rep.

Vyas, Saurabh et al. (2020). "Computation through neural population dynamics". In: *Annual Review of Neuroscience* 43, pp. 249–275.

Wang, Xiao-Jing (2002). "Probabilistic decision making by slow reverberation in cortical circuits". In: *Neuron* 36.5, pp. 955–968.

White, John A, Ruby Klink, et al. (1998). "Noise from voltage-gated ion channels may influence neuronal dynamics in the entorhinal cortex". In: *Journal of neurophysiology* 80.1, pp. 262–269.

White, John A, Jay T Rubinstein, and Alan R Kay (2000). "Channel noise in neurons". In: *Trends in neurosciences* 23.3, pp. 131–137.

Wilson, Hugh R and Jack D Cowan (1972). "Excitatory and inhibitory interactions in localized populations of model neurons". In: *Biophys. J.* 12.1, pp. 1–24.

– (1973). "A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue". In: *Kybernetik* 13.2, pp. 55–80.

Wong, Kong-Fatt and Xiao-Jing Wang (2006). "A recurrent network mechanism of time integration in perceptual decisions". In: *Journal of Neuroscience* 26.4, pp. 1314–1328.

Yang, Weijian et al. (2018). "Simultaneous two-photon imaging and two-photon optogenetics of cortical circuits in three dimensions". In: *elife* 7, e32671.

VALENTIN SCHMUTZ

PERSONAL INFORMATION

ORCID iD: 0000-0002-0935-6121

Google Scholar ID: I2ihecwAAAAJ

✉ valentin.schmutz@epfl.ch

☎ +41 78 975 28 75

EDUCATION

École Polytechnique Fédérale de Lausanne, Switzerland — Since Aug 2018
PhD in Neuroscience
Director: Prof. Wulfram Gerstner, co-director: Prof. Eva Löcherbach

École normale supérieure Paris, France — Sep 2016 - Jul 2018
1st year MSc (M1) in Mathematics
3rd year BSc (L3) in Mathematics — Oct 2017

Sorbonne Université - UPMC Paris VI, France — Sep 2014 - Jul 2016
2nd year BSc (L2) in Mathematics

Université de Lausanne, Switzerland
Diplôme Fédéral de Médecin (Swiss Federal Examination in Human Medicine) — Oct 2014
Master of Medicine — June 2014
Bachelor of Medicine — Aug 2011

EMPLOYMENT HISTORY

Doctoral Assistant in the Laboratory of Computational Neuroscience — Since Aug 2018
École Polytechnique Fédérale de Lausanne, Switzerland
Supervisor: Prof. Wulfram Gerstner

Médecin assistant (Resident) in Internal Medicine — Apr 2020 - Mai 2020
Centre Hospitalier Universitaire Vaudois, Switzerland — (COVID first wave)
Supervisor: Prof. Peter Vollenweider

Visiting doctoral student in the SAMM Group — Oct 2019 - Jan 2010
University Panthéon-Sorbonne Paris 1, France
Supervisor: Prof. Eva Löcherbach

Visiting doctoral student in the Buenos Aires Probability Group — Jul 2019 - Sep 2019
University of Buenos Aires, Argentina
Supervisor: Prof. Pablo A. Ferrari

Médecin assistant (Resident) in Neurology — Nov 2015 - Apr 2016
Centre Hospitalier Universitaire Vaudois, Switzerland
Supervisor: Prof. Renaud Du Pasquier

Médecin assistant (Resident) in Internal Medicine — Nov 2014 - Oct 2015
Hôpital du Jura Bernois – Moutier, Switzerland
Supervisor: Dr. Pierre A. Triverio

Research intern in the Modelling of Dynamic Action and Perception Group — Sep 2014 - Nov 2014
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
Supervisor: Prof. Stephan Kiebel

MAIN PUBLICATIONS

On a finite-size neuronal population equation.
Schmutz V.*, Löcherbach E.*, Schwalger T*.
Accepted in *SIAM Journal on Applied Dynamical Systems*

Mesoscopic modeling of hidden spiking neurons.
Wang S.*, **Schmutz V.***, Bellec G., Gerstner W.
Accepted in *Advances in Neural Information Processing Systems*

Long Time Behavior of an Age- and Leaky Memory-Structured Neuronal Population Equation.   2022
Fonte C.*, **Schmutz V.***
SIAM Journal on Mathematical Analysis 54 (2022), no. 4, 4721–4756, DOI:10.1137/21M1428571.

Mean-field limit of age and leaky memory dependent Hawkes processes.   2022
Schmutz V.
Stochastic Processes and their Applications 149 (2022), 39–59, DOI:10.1016/j.spa.2022.03.006.

Mesoscopic population equations for neural networks of spiking neurons with synaptic short-term plasticity.   2020
Schmutz V., Gerstner W., Schwalger T.
The Journal of Mathematical Neuroscience 10 (2020), no. 5, 32 pp., DOI:10.1186/s13408-020-00082-z.

*: equal contribution

PREPRINTS

Mesoscopic description of hippocampal replay and metastability in spiking neural networks with short-term plasticity.
Pietras B., **Schmutz V.**, Schwalger T.
arXiv:2204.01675. (Under review)

OTHER PUBLICATIONS

Pre-existing astrocytes form functional perisynaptic processes on neurons generated in the adult hippocampus.   2014
Krzisch M., Temprana S. G., Mongiat L. A., Armida J., **Schmutz V.**, et al.
Brain Structure and Function 220.4 (2015), 2027-2042, DOI:10.1007/s00429-014-0768-y.

Visuospatial viewpoint manipulation during full-body illusion modulates subjective first-person perspective.   2014
Pfeiffer C., **Schmutz V.**, Blanke O.
Experimental Brain Research 232.12 (2014), 4021-4033, DOI:10.1007/s00221-014-4080-0.

Multisensory origin of the first-person perspective: visual, tactile, and vestibular mechanisms.   2013
Pfeiffer C., Lopez C., **Schmutz V.**, Duenas J. A., Martuzzi R., Blanke O.
PLoS One, 8.4 (2013), e61751, DOI:10.1371/journal.pone.0061751.

TALKS AT INTERNATIONAL CONFERENCES

International Conference on Mathematical Neuroscience (invited)   2020
▶ *Mean-field limit for large networks of multidimensional spiking neurons*

Bernoulli-IMS One World Symposium (contributed)   2020

TEACHING ACTIVITIES

Teaching assistant at École Polytechnique Fédérale de Lausanne for the courses:
- Biological modeling of neural networks   2020 - 2021
- Introduction to machine learning for bioengineers   2020
- Analysis IV   2019

Teaching assistant at Université de Lausanne for the course:
- Pratical work: human anatomy dissection   2010 - 2012

STUDENT SUPERVISION

Shuqi Wang (EPFL)	Intern & Master thesis student in Data Science	Sep 2021 - Jul 2022
Louis Pezon (EPFL)	Intern	Dec 2021 - Apr 2022
Joachim Koerfer (EPFL)	Master thesis student in Computational Science and Engineering	Feb 2021 - Jul 2021

PERSONAL SKILLS

Languages: French (mother tongue), English (fluent), Mandarin (intermediate), German (intermediate).

Programming languages: Julia, Matlab, Python.