Metastability in a System of Spiking Neurons with Synaptic Plasticity

Morgan André, Christophe Pouzat

Metastability in a System of Spiking Neurons with Synaptic Plasticity

M. André[1] C. Pouzat[2]

July 19, 2021

Affiliations

[1] Instituto de Computação-Unicamp, State University of Campinas
[2] Strasbourg University and CNRS UMR 7501

Correspondance should be addressed to:
C. Pouzat
IRMA
7 rue René-Descartes
67084 Strasbourg Cedex
France
christophe.pouzat@math.unistra.fr

Abstract

After reviewing the behavioral studies of working memory and of its cellular substrate, we argue that metastable states constitute candidates for the type of transient information storage required by working memory. We then present a simple neural network model made of stochastic units whose synapses exhibit short-term facilitation. This model was specifically designed to be analytically tractable, simple to simulate numerically and to exhibit a strong form of metastability. We present both numerical results: the existence of metastable states able to represent scalar quantities; and analytical ones: the properties of the metastable states that can be fully deduced from an implicit equation once the 4 model parameters have been specified.

Acknowledgments

This article was produced as part of the activities of FAPESP Research, Innovation and Dissemination Center for Neuromathematics (grant number 2013/07699-0, S.Paulo Research Foundation, Brazil), and Morgan André was supported by FAPESP, Brazil scholarships (grant number 2020/12708-1).

1. Introduction

1.1. The neurobiological horizon

As described by Fuster in the introduction of his 1973 article [16]: *A delayed-response trial typically consists of the presentation of one of two possible visual cues, an ensuing period of enforced delay and, at the end of it, a choice of motor response in accord with the cue. The temporal separation between cue and response is the principal element making the delayed response procedure a test of an operationally defined short-term memory function.* In that article Fuster described, in the monkey prefrontal cortex, neurons that switch between no activity and a sustained activity at constant rate *during the delay period* when the animal had to perform a delayed-response task. He showed moreover, using distracting stimuli that interrupted the sustained activity of these neurons, that the monkey errors at the end of the delay period were positively correlated with the interruption of sustained activity. Since then, many experimental investigations reviewed in [15, Chap. VII] and [10] have confirmed this basic finding and showed that some of the 'sustained activity neurons' are insensitive to the type of cue (color, shape, location, sound) and seem to encode the 'abstract' notion of remembering 'any' cue until the expiration of a delay, while others, especially outside the prefrontal cortex, are sensitive to the type of cue. These 'sustained activity neurons' are relatively easy to record from *implying that they are fairly abundant* [22]. This sustained activity has been intriguing modelers for a long time, leading them to explore first network models with subgroups of strongly reciprocally coupled excitatory neurons [37, 2]. The sustained activity has then been interpreted as a local attractor of some dynamical system (reviewed in [34]). Stability issues when the transiently memorized item is a continuous quantity–like an angle–, lead them to include some 'slow' and 'use dependent' coupling, initially in the form of NMDA receptors [9, 34]–for a review of basic neurophysiology, see [24]. But [35] described a subclass of pyramidal (and therefore excitatory) cells in the prefrontal cortex that are strongly interconnected and whose synapses are unusual since they exhibit a marked short-term facilitation–synapses between these cell types exhibit most of the time short-term depression, as illustrated in the same article by recordings from the visual cortex. This has lead to several studies giving a more or less central role to short-term facilitation in sustained activity generation or stabilization, *e.g.* [5, 19, 18]–or even proposing a working memory mechanism without sustained activity [26]–, reviewed in [6]. But the secondary status of the 'noise' in these studies, where variability comes into play mostly at the neurons input level, is at odd with basic empirical observations. It is indeed well known [36, 24] that neurons depend on ion channels that are randomly going back and forth between closed and opened states both for the action potential generation [33] and the synaptic transmission [21]; that (chemical) synaptic transmission involves the release of a variable number of transmitter packets / quanta [12, 11] giving rise to the rather noisy membrane voltage trajectories that are actually observed. These considerations strongly suggest an alternative model construction strategy: working with stochastic units / neurons instead of deterministic ones. Continuing and simplifying [17], we therefore developed a minimal model of the sub-network of reciprocally coupled pyramidal cells with facilitating synapses [35]; model that is both amenable to analytical solutions and

that can be easily simulated. This model is made of stochastic neurons that accumulate their inputs until a threshold is reached. The synapses between the neuron exhibit short-term facilitation enabling the sub-network to exhibit a transient *memoryless* sustained activity–that is, genuine metastability–, reminiscent of what is observed in working memory experiments.

1.2. Definition of the model

We consider a stochastic system of interacting spiking neurons. The system consists in a finite set of N neurons, each neuron $i \in \{1, \dots N\}$ being associated with a membrane potential denoted $(U_i(t))_{t\geq 0}$ which takes value in the set of non-negative integers. The spiking activity of the neurons depends on a threshold value $\theta \in \mathbb{Z}^+$. When $U_i(t) < \theta$ neuron i cannot spike, and we say that it is *quiescent*, while when $U_i(t) \geq \theta$ we say that neuron i is *active*, and it spikes at rate β : i.e. it waits a time Δ_t distributed as an exponential random variable of parameter β and then spike at time $t + \Delta_t$. At any time the synaptic connection between the neuron and the rest of the network can be either facilitated or not, meaning that if a spike occurs, it will be transmitted to the system or not. The facilitation state for neuron i is denoted $(F_i(t))_{t\geq 0}$, and it is a stochastic process taking value in $\{0, 1\}$. Whenever $F_i(t) = 1$ we say that the synapse of neuron i is facilitated at time t, and the synapse looses its facilitation, i.e. goes back to the state $F_i(t + \Delta_t) = 0$, at a given rate λ. When a neuron spikes, first the membrane potential of the said neuron is systematically reset to 0, secondly its synapse becomes facilitated if it wasn't already, and thirdly if its synapse was facilitated at the moment of the spike then the membrane potential of all the other neurons in the system increases by one unit, while nothing happens for these neurons if the synapse wasn't facilitated at the moment of the spike. All exponential random variables involved are supposed to be independent.

1.3. Metastability

Our system clearly has a class of absorbent states, namely the class of the states in which $U_i(t) < \theta$ for all $i \in \{1, \dots N\}$. When an element of this class is reached then the system will never spike again. This quiescent state is always reachable with a positive (but typically small) probability. It will be reached for example if all active neurons at any given time loose their synaptic facilitation before they spike, so that if we wait long enough this state will always be reached at some point. This quiescent state is therefore the only equilibrium.

However many stochastic systems with a similar setting exhibit some kind of second-order equilibrium. In these systems the activity has a tendency to stay a long and unpredictable time in a seemingly stable phase before falling into the actual equilibrium because of an unusually important deviation. Such systems are said to be *metastable*. The large class of metastable stochastic processes includes for example the contact process [8, 32, 27], the curie-weiss model [8], the Ising model [7], and more recently stochastic neural networks models reminiscent of the model considered here but without any synaptic plasticity [3, 4,

25]. More precisely metastable stochastic systems are characterized by the following two properties. For a suitable choice of the values of the parameters:

1. the time it takes to reach the quiescent state is asymptotically exponentially distributed as the number of components in the system grows,

2. before reaching the quiescent state the system behave in an almost stationary manner.

The purpose of this paper is to give evidence that the behavior described above indeed holds for our system by both simulation and heuristic reasoning.

For the first point, related to the time of extinction of the system, we shall give some explanation. O. Penrose and J. L. Lebowitz proposed in a seminal paper [28] to characterize a metastable state by requiring that *if the system start in such a state then it is likely to take a long time to get out*. It is only a few years latter that M. Cassandro et al. [8] refined this point as they realized that an important property of metastable stochastic systems is that *the exit time from the metastable phase is not only long but also unpredictable*. Formally what it means is that this exit time is asymptotically memoryless, or in other words exponentially distributed–as the exponential law is the only continuous distribution satisfying this memorylessness. In our specific setting, if we denote by σ_N the time of extinction of the system with a number N of neurons, then the assumption we would like to test is whether or not, for suitable values of β and λ, the following holds

$$\frac{\sigma_N}{\mathbb{E}(\sigma_N)} \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{E}(1),$$

where $\mathbb{E}$ denotes the mathematical expectation, $\mathcal{E}(1)$ denotes an exponential random variable of parameter 1, and the superscript $\mathcal{D}$ denotes a convergence in distribution.

This property of convergence of the renormalized time of extinction toward an exponential random variable was studied both rigorously and numerically for a different but similar stochastic system of spiking neurons in [3], [4] and [29]. An interesting difference between our model and the model studied in these papers, which was originally introduced in [13], is that there the membrane potential of each neuron was subject to a leakage effect, which was the direct mechanism by which the system was pushed toward its quiescent equilibrium, while in our model this role is assumed by the loss of synaptic facilitation, as there is no leakage.

For the second point we adopt a mean field approach, assuming that before reaching the quiescent equilibrium all the elements of the system have roughly the same stationary behavior. We characterize the system by establishing an implicit equation linking all the parameters, and after solving this equation we compute meaningful values. We then compare our predictions with estimates we get from simulations in order to validate them.

1.4. A remark

If we *fix* the facilitation state of each synapse to 1 ($F_i(t) = 1$ for all i and $t \geq 0$) and if at least θ neurons are above threshold at some point, then the network activity never stops since a neuron whose membrane potential is 0 needs θ synaptic inputs to reach threshold (θ).

2. Methods

In this section we describe the algorithm used to simulate our model.

2.1. Informal description of the procedure used to simulate the system

At any given time $t \geq 0$ we write $U(t)$ the number of active neurons and $F(t)$ the number of facilitated synapses in the system. That is:

$$U(t) = \sum_{i=1}^{N} \mathbb{1}_{U_i(t) \geq \theta} \qquad \text{and} \qquad F(t) = \sum_{i=1}^{N} \mathbb{1}_{F_i(t)=1}.$$

The general rate of the next event is then given by $\nu(t) = \beta U(t) + \lambda F(t)$. For the sake of efficiency we first generate the time of the next event using this general rate, by taking

$$\Delta_t = \frac{-\log(v)}{\nu(t)},$$

where v is sampled from a uniform law on $]0, 1[$. We decide only then if this event shall be a spike or a synaptic inactivation by sampling a Bernoulli random variable of parameter

$$p = \frac{\beta U(t)}{\nu(t)}.$$

2.2. Pseudo-code of the algorithm

The general pseudo-algorithm giving the time and type of the next event is as follows:

1: N ▷ Networks size (number of neurons)
2: θ ▷ Threshold for entering the "spiking state"
3: β ▷ Spike rate when potential is above θ
4: λ ▷ Rate of synaptic de-facilitation
5: t_{now} ▷ "Present" time, a spike was just generated
6: $(u_1(t_{now}), u_2(t_{now}), \ldots, u_N(t_{now}))$ ▷ Known membrane potential of each neuron
7: $(f_1(t_{now}), f_2(t_{now}), \ldots, f_N(t_{now}))$ ▷ Known synaptic state of each neuron
8: $U \leftarrow 0$ ▷ Holds the number of neurons above threshold
9: $F \leftarrow 0$ ▷ Holds the global synaptic facilitation
10: **for** $i \leftarrow 0, N-1$ **do**
11: $\quad u_i \leftarrow u_i(t_{now})$ ▷ Definition and initialization of variables

$f_i \leftarrow f_i(t_{now})$ ▷ Definition and initialization of variables
$U \leftarrow U + 1$ if $u_i \geq \theta$
$F \leftarrow F + 1$ if $f_i = 1$
end for
if F = 0 **then**
Abort, network dead or about to die
end if
$\nu \leftarrow \beta U + \lambda F$ ▷ The global events' rate
$v \leftarrow \mathcal{U}(0,1)$ ▷ Draw from a uniform distribution on (0,1)
$\Delta t \leftarrow -\frac{\log v}{\nu}$
$t_{now} \leftarrow t_{now} + \Delta t$
$v \leftarrow \mathcal{U}(0,1)$
if $\beta U/\nu > v$ **then** ▷ The event is a spike
event_type $\leftarrow 1$ ▷ entent_type is 1 for a spike
$n \leftarrow \mathcal{M}\left(\mathbb{1}_{u_0 \geq \theta}/U, \ldots, \mathbb{1}_{u_{N-1} \geq \theta}/U\right)$ ▷ Draw neuron from multinomial dist.
if $f_n = 1$ **then** ▷ The neuron that spiked has a facilitated synapse
for $i \leftarrow 0, N-1$ **do**
$u_i \leftarrow u_i + 1$
if $u_i - \theta \leq 1$ and $u_i \geq \theta$ and $i \neq n$ **then** ▷ Neuron i just crossed threshold
$U \leftarrow U + 1$
end if
end for
end if
$u_n \leftarrow 0$ ▷ Reset potential of neuron that spiked
$U \leftarrow U - 1$ ▷ Neuron n was above θ and is now below it
if $f_n = 0$ **then** ▷ The synapse was in the resting state
$F \leftarrow F + 1$
end if
$f_n \leftarrow 1$ ▷ The synapse is facilitated
else ▷ The event is a synaptic inactivation
event_type $\leftarrow 0$ ▷ entent_type is 0 for a synaptic inactivation
$n \leftarrow \mathcal{M}\left(\mathbb{1}_{f_0=1}/R, \ldots, \mathbb{1}_{f_{n-1}=1}/R\right)$ ▷ Draw neuron from multinomial dist.
$f_n \leftarrow 0$ ▷ Inactivate synapse of neuron n
$F \leftarrow F - 1$
end if
Return t_{now}, event_type and n

2.3. Simulations initialization

Our simulations were initialized by drawing the membrane potential of each neuron independently on the set $\{0, 1, \ldots, N-1\}$ (N is the network size) and the synapse facilitation state from a Bernoully distribution with a success probability of 0.75 ("success" means that the synapse is facilitated). These parameters can be changed by the users of our

codes.

2.4. Some remarks

The four model parameters, N, θ, β and λ can be reduced to three. All that matters is the ratio λ/β has inspection of lines 19 and 24 of the above algorithm makes clear.

We have chosen fixed increments of 1 upon synaptic inputs in order to have membrane potentials with integer values; this is mostly motivated by numerical convenience since it leads to codes requiring less memory. In the sequel, when we study the effect of network size on various network quantities, we will always set the threshold θ as a fraction of N (10% unless otherwise stated). This is equivalent to the traditional scaling in $1/N$ of the synaptic weights used when studying the asymptotic limit. This allows us to keep integer valued membrane potentials.

2.5. Programs accessibility and results replication

All programs used in this article, as well as the details of their implementation, the commands required to replicate the simulations and figures are available on `GitLab`[1]. The simulation and numerical integration (Sec. A.1) programs are written in `C` and the analysis and plotting programs / scripts are written in `Python`.

3. Empirical results

In this section we present simulations of the model. We are interested in observing how our neural system behave in general and, more specifically, if it exhibits the kind of "strong" metastability described in the introduction.

3.1. Basic model features: individual neuron level

Figure 1 shows the trajectories of the "membrane potentials" of all the (50) neurons of a simulation ($N = 50$, $\theta = 5$, $\beta = 10$, $\lambda = 6.7$) during one time unit.

When the membrane potential of a neuron is at or above threshold, that is, when the neuron is in the active state, the displayed membrane potential value is the threshold value (θ). From the dynamics viewpoint, it only matters to know that the neuron has reached threshold, not the actual membrane potential value when the latter is above threshold. The membrane potential traces are drawn in blue when the synapse of the corresponding neuron is facilitated and orange otherwise. *This figure displays therefore the complete state of the network*. Notice that at any given time, most of the neurons are in the active state (their membrane potential is $\geq \theta$). Notice also that the membrane potentials of the neurons that have not yet reached θ evolve in parallel. Spike are emitted when the membrane potential of one neuron goes from θ to zero, this is the only way the membrane potential can decrease.

A finer time display is proposed on Fig. 2.

[1] https://gitlab.com/c_pouzat/metastability-in-a-system-of-spiking-neurons-with-synaptic-plasticity

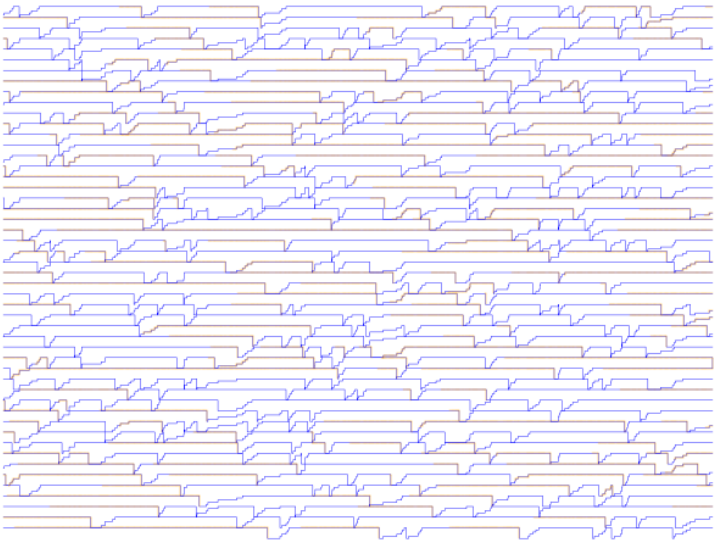

Figure 1: Trajectories between time units 1 and 2 of the membrane potentials of the fifty neurons of a simulated network. The traces are blue when the synapse of the neuron is facilitated and orange otherwise.

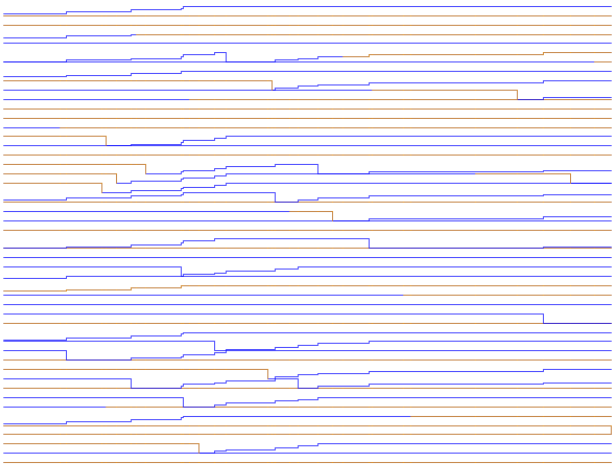

Figure 2: Enlarge display between times 1.20 and 1.25 of the data shown of Fig. 1.

The features of the model are clearly visible:

- When a neuron with an un-facilitated synapse spikes (the trace is orange when the membrane potential is at the threshold level just before dropping to 0):

  - its synapse gets facilitated (the trace turns blue) immediately after the spike,
  - the membrane potential of all the other neurons remains the same.

- When a neuron with a facilitated synapse spikes (the trace is blue when the membrane potential is at the threshold level just before dropping to 0):
  - its synapse remains facilitated (the trace stays blue) immediately after the spike,
  - the membrane potential of *all* the other neurons that are below threshold increases by 1.

3.2. Basic model features: it is the same but it is not the same

We now turn to the key property our model was designed to exhibit. We show next spike trains displayed as raster plots (every spike is represented by a dot) of the same network of 50 neurons started from the same initial state but using two different sequences of (pseudo) random numbers. Fig. 3 shows an abrupt disappearance of the activity (after 13 time units).

Figure 3: Raster plots of a 50 neurons network, with $\lambda = 6.7$, $\beta = 10$ and $\theta = 5$. Left, from time 0 to 14; right from time 12 to 14. Dots are blue when the synapse is active and orange otherwise

The dots color is blue when the synapse is active and orange otherwise. We see on the right side of Fig. 3 that the last spikes occurring before the "network death" are all with an un-facilitated synapse. Fig. 4 shows the same network as Fig. 3, starting from the same state and remaining active for whole simulation (50 time units).

Judging from the dots pattern, the activity looks regular with a constant ratio of blue dots over orange dots. But a better way to graphically asses the network activity (network spiking frequency) is provided by the observed counting process (a step function that increases by

Figure 4: Same as Fig. 3 but different random numbers sequence. The scale bar is drawn between time 10 and time 15.

one every time an event occurs) as shown on Fig. 5 for the two simulations of Fig. 3 and 4. Extracting the slope by eye, we see that the network generates roughly 375 events per

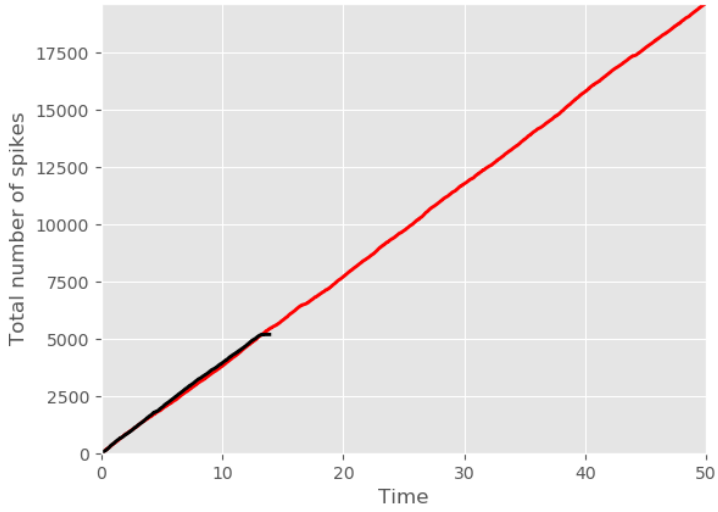

Figure 5: Observed counting processes for the simulations of Fig. 3 (black) and Fig. 4 (red).

time unit (before it reaches the quiescent state in the case of the black trace). *We have, qualitatively at least, the behavior we are interested in: the activity seems "stationary"*

until it abruptly vanishes.

3.3. Survival time

Figure 6 shows the empirical survival function obtained from 1000 repetitions of the simulation of the model, for various values of the parameter λ (6, 6.7 and 7) with $N = 50$, $\theta = 5$ and $\beta = 10$. Each simulation with a given value of λ starts from the exact same initial state except for the sets of simulations corresponding to the red and blue lines. The latter illustrate the fact that the precise initial state from which the simulations are started is irrelevant as long as these states are drawn from the same distribution.

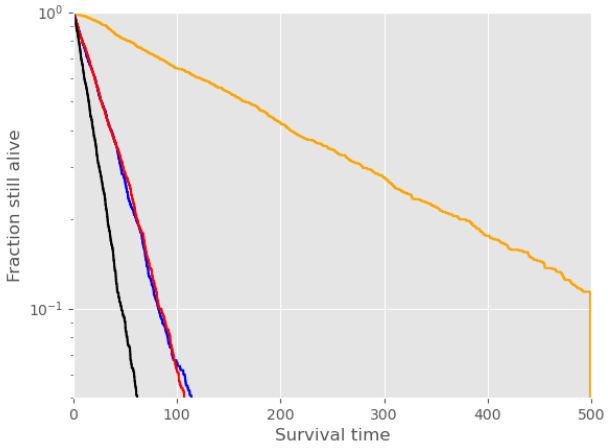

Figure 6: Empirical survival functions obtained from 1000 replicates with $\theta = 5$, $\lambda = 6.7$ (red and blue), $\lambda = 7$ (black) and $\lambda = 6$ (orange), $\beta = 10$ and a network with 50 neurons. All simulations start from *the same* random initial state except the red and blue ones. **A log scale is used for the ordinate**.

The survival function gives the fraction of the 1000 simulations that are still "active" as a function of the observation duration. Notice that a *log scale* is used on the ordinate of this (and the following) figure. The rational for that choice follows [23]. If we observe a sample $t_1, t_2, \ldots, t_{1000}$ drawn from $T_1, T_2, \ldots, T_{1000}$ independently and identically distributed exponential random variables with parameter γ, the probability density function (PDF) is given by: $f_T(t) = \exp(-t/\gamma)/\gamma$, the cumulative distribution function (CDF) is: $F_T(t) = 1 - \exp(-t/\gamma)$ and the *survival function* is $S_T(t) = 1 - F_T(t) = \exp(-t/\gamma)$. The expected value of T is $\mathbb{E}T = \gamma$ and a graph of $S_T(t)$ with a log scale on the ordinate is a straight line with slope $-1/\gamma$. The estimator $\widehat{S}_T(t) = \sum_{i=1}^{1000} \mathbb{1}_{\geq t}(t_i)/1000$ of $S_T(t)$ should therefore be close to a straight line on a log scale for exponential distributions. That is precisely what we see here. The exponential distribution of the survival times (or *times to extinction*) is the key feature of the metastable states as we defined them in the

introduction. Not surprisingly, the mean survival time (the opposite of the inverse of the slopes on Fig. 6) decreases as λ increases.

If we keep increasing λ as shown on Fig. 7, the mean survival time keeps decreasing until the exponential (straight line) behavior is lost (for $\lambda \approx 9$). The empirical survival functions start exhibiting a concavity (on the log scale) for small values of the survival time. This is a clear deviation from what is expected from an exponential distribution and allows us to rule out the adequacy of the latter for these large values of λ [23]. Figure 8 shows

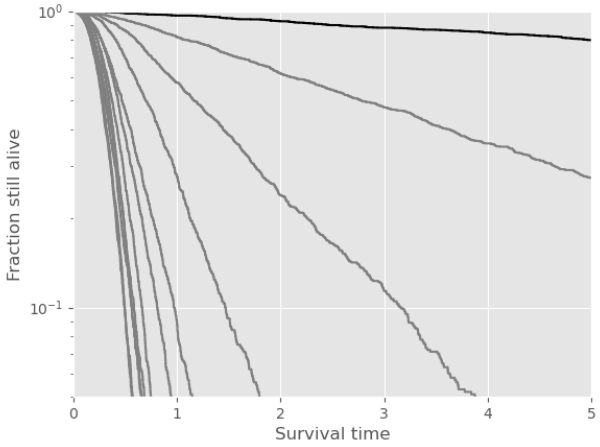

Figure 7: Same as before with $\lambda = 7$ (black) and $\lambda = 8, \ldots, 18$ (grey). Notice that the domain covered by the abscissa is 100 smaller than on Fig. 6.

the 95% confidence intervals of the mean survival time as a function of $\lambda < 9$ ($N = 50$, $\theta = 5$, $\beta = 10$) obtained from 1000 simulations at each λ value. A mono-exponential decay was fitted to the observed survival times by maximizing the likelihood with a model including right censoring. Some simulations are indeed still in the "metastable" state at the time they are stopped (*e.g.*, slightly more than 10% for $\lambda = 6$, orange line on Fig. 6), so we do not know their exact survival time, but a lower bound of it. The parameter of this exponential decay is the mean survival time. The confidence intervals were computed with a likelihood ratio method [20, pp. 32-36]. We see that for small values of λ (< 8), the mean survival time decays exponentially and when λ approaches the "critical value" at which the exponential distribution of the survival time is lost, the decay of the mean survival time slows down.

Fig. 9 shows the dependence of the mean survival time on network size in a setting where the threshold $\theta = N/10$. The simulation strategy of the previous figure was used but only 100 replicates were generated (to save time) and a relatively large λ value ($\lambda = 7$) was used in order to observe enough "network deaths" in a relatively short time. We see that the growth of the mean survival time or mean time to extinction as a function of network size (with $\theta = N/10$) is compatible with an exponential, suggesting that by making the

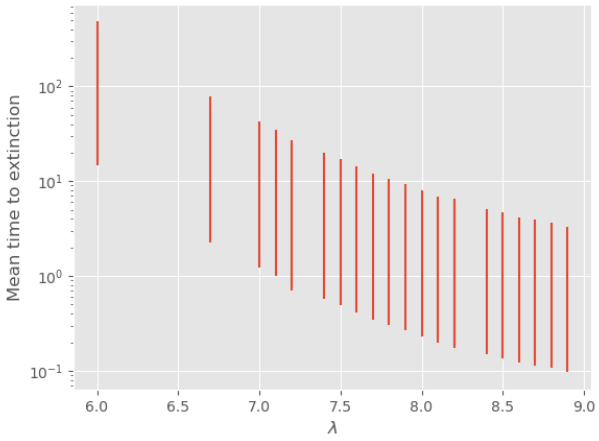

Figure 8: 95 % CI of the mean time to extinction as a function of λ. From 1000 simulations for each λ and $\beta = 10$ and a network with 50 neurons. **A log scale is used for the ordinate**.

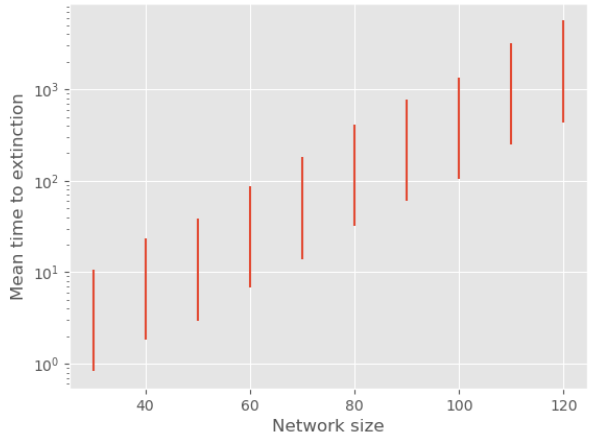

Figure 9: 95 % CI of the mean survival as a function of N. From 100 simulations, for each N: $\lambda = 7$, $\beta = 10$ and $\theta = N/10$. **A log scale is used for the ordinate**.

network grow we could obtain a time of residency in the active regime of any order of magnitude.

3.4. Empirical properties: summary and tentative conclusions

Our empirical investigation strongly suggests (Fig. 6) that for λ small enough—more precisely for λ/β small enough—our model exhibits a "strong form" of metastablity in which the residency time in the "active" state is memoryless (exponentially distributed). There seems to be a "critical" λ value above which metastability is lost (Fig. 7). For λ small enough, the mean survival time seems to increase exponentially as λ approaches 0 (Fig. 8), implying that the time of residency in the active state grows toward infinity as λ goes to zero. In other words if facilitation stops being a transient phenomenon to become a permanent one, the active state becomes permanent (as we already argued upon model specification). When the network size increases (together with the threshold), the mean survival time undergoes an exponential growth (Fig. 9), becoming closer and closer to a permanent state as well.

4. A heuristic approach

In this section we study our model by heuristic reasoning checked with simulations. Our objective is to give a complete characterization of the way our system behave before extinction. Assuming stationarity in the pre-extinction period, we aim at computing meaningful values, such as the proportion of active synapses, the global spiking rate of the system, the mean inter-spike interval and so on. This is done by establishing an implicit equation linking all parameters of the model (β, λ, θ and N) as well as the probability for a given neuron to have its synapse facilitated at the moment of its next spike. Solving this last equation for this probability then allows us to obtain all the other quantities. Then, using simulations, we can get estimates of the quantities we just mentioned and compare them to our predictions.

4.1. Mean-field approach

We let $N \in \mathbb{Z}^+$ be the number of neurons in the system. While the membrane potential of a given neuron can take value in the whole set $\mathbb{Z}^+$, the only thing we really care about is whether or not it is at or above the threshold θ. We will therefore identify the set $\mathbb{S} = \{\theta, \theta+1, \ldots\}$ as one single state in which the neuron is susceptible or active. For any $i \in \{0, 1, \ldots, \theta-1\}$ we define $N_i(t)$ to be the number of neurons whose membrane potential is equal to i at time t, that is

$$N_i(t) = \sum_{j \in V} \mathbb{1}_{U_j(t)=i}.$$

We define $N_\theta(t)$ as well, the number of neurons which membrane potential is greater than or equal to θ at time t:

$$N_\theta(t) = \sum_{j \in V} \mathbb{1}_{U_j(t)\geq\theta}.$$

At any time t we obviously have

$$\sum_{i=0}^{\theta} N_i(t) = N. \tag{4.1}$$

We also define the total number of facilitated synapses in the network at time t, which we denote $F(t)$, given by

$$F(t) = \sum_{i \in V} F_i(t).$$

Under our assumption of quasi-stationarity, the expectations of the quantities defined above should be almost constant in the metastable phase. Thus we let $\mu_0, \mu_1, \dots \mu_\theta$, and μ_F be the constants such that

$$\mathbb{E}(N_0(t)) \approx \mu_0, \ \dots \ \mathbb{E}(N_\theta(t)) \approx \mu_\theta,$$

and

$$\mathbb{E}(F(t)) \approx \mu_F,$$

where t is any time before the extinction of the system.

To carry out our calculations we will assume that the random variables above are close to their expectations and simply replace them in practice by the constants $\mu_0, \dots, \mu_\theta$, and μ_F. The global spiking rate of the system in the metastable phase should then be well approximated by

$$\nu_N \stackrel{\text{def}}{=} \mu_\theta \beta, \tag{4.2}$$

that is, the spiking rate for a single neuron multiplied by the expected number of susceptible neurons.

4.1.1. The effective spiking rate

While the global spiking rate of the network is an interesting quantity, it would be even more interesting to know the rate of effective spikes; that is, the rate of the spikes that are actually propagated via the synapse. In order to get this rate we need to compute the probability that a neuron that just spiked will still have its synapse facilitated at the moment of the next spike. To fix our ideas we suppose that the system is in any metastable state at time 0 and we consider neuron 1 (which neuron you look at doesn't matter in our mean field approach). We assume that $U_1(0) = 0$ and that $F_1(0) = 1$. Let τ be the random variable corresponding to the time of the next spike. The next spike will be effective if and only if the synapse is still facilitated at the time of the spike, so that we need to consider the following random variable

$$E \stackrel{\text{def}}{=} F_1(\tau).$$

We have

$$\mathbb{E}\Big(F_1(\tau) \mid \tau\Big) = e^{-\lambda\tau},$$

so that

$$\mu_E \overset{\text{def}}{=} \mathbb{E}(E) = \mathbb{E}\Big[\mathbb{E}\Big(F_1(\tau) \mid \tau\Big)\Big] = \mathbb{E}\left(e^{-\lambda\tau}\right). \tag{4.3}$$

Of course we don't know the actual distribution of τ so that the exact value of μ_E is still unknown to us, but we can nonetheless use it to define the rate of effective spikes we were looking for:

$$\nu_E \overset{\text{def}}{=} \mu_E \nu_N. \tag{4.4}$$

Figure 10 shows a simple diagram of the evolution of the membrane potential for a single neuron when the system is in the metastable phase.

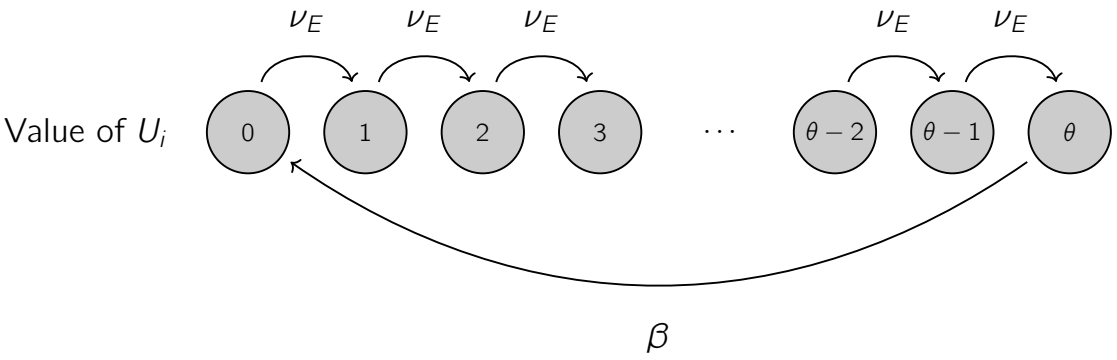

Figure 10: Schematic representation of the way the membrane potential behave for one single neuron in the metastable state.

A rough, first order approximation for μ_E would be to replace τ by its expectation in (4.3). Nonetheless we can do better. In our mean-field approach we go from a value below threshold to the next one with rate ν_E, thus the time τ_b spent below the threshold follows an Erlang distribution with shape parameter θ and rate ν_E, and we have the following expression for its mean

$$\mu_{\tau_b} = \frac{\theta}{\nu_E}. \tag{4.5}$$

Once the threshold has been reached we need to wait an additional exponentially distributed time with rate β, and when the size of the network is big the time spent below the threshold should be small, so that we might replace τ_b by its mean and use the following approximation:

$$\tau \approx \mu_{\tau_b} + \epsilon,$$

where $\epsilon \sim \text{Exponential}(\beta)$. From (4.3) we then get the following approximation for μ_E

$$\mu_E \approx \int_0^\infty \exp\left(-\lambda(\mu_{\tau_b} + t)\right) \beta \exp\left(\beta t\right) dt,$$

which leads to

$$\mu_E \approx \frac{\beta}{\beta+\lambda}\exp(-\lambda\mu_{\tau_b}). \tag{4.6}$$

4.1.2. Membrane potential and facilitation state

Now consider the membrane potential of any neuron in the metastable phase. As long as the extinction time as not been reached, it goes in cycle, from 0 to 1, then from 1 to 2 etc. up to the susceptible state $\mathbb{S}$, it then spikes and starts over from 0. We might look at the flux entering and leaving each of these states.

- At state 0, the entering flux is $\beta\mu_\theta$ while the leaving flux is $\mu_0\nu_E$. Because of our hypothesis of stationarity they should be equal, leading to $\mu_0 = \frac{1}{\mu_E}$ (using Eq. 4.2 and 4.4).
- At state $i \in \{1, \dots \theta - 1\}$, the entering flux is $\mu_{i-1}\nu_E$ and the leaving flux is $\mu_i\nu_E$, leading to $\mu_i = \mu_{i-1}$.

From this two points we get

$$\mu_0 = \dots = \mu_{\theta-1} = \frac{1}{\mu_E}.$$

From equation (4.1) it follows that

$$\mu_\theta = N - \frac{\theta}{\mu_E}. \tag{4.7}$$

In order to get an equation for μ_F we can write the following differential equation, which should be satisfied in the metastable phase,

$$\frac{d\mathbb{E}(F(t))}{dt} = -\lambda\mu_F + \beta\mu_\theta(1-\mu_E).$$

Setting the derivative to 0 we get

$$\mu_F = \frac{\beta}{\lambda}\mu_\theta(1-\mu_E) = \frac{\beta}{\lambda}\left(N - \frac{\theta}{\mu_E}\right)(1-\mu_E).$$

4.1.3. The implicit equation

Equation (4.7), together with (4.5) and the definitions of ν_E (Eq. 4.4) and ν_N (Eq. 4.2), gives the following formula for μ_{τ_b}:

$$\mu_{\tau_b} = \frac{\theta}{\beta(N\mu_E - \theta)}.$$

Substituting this μ_{τ_b} expression in Eq. 4.6, we obtain the following implicit equation for μ_E:

$$\mu_E \approx \frac{\beta}{\beta+\lambda} \exp\left(\frac{-\lambda\theta}{\beta(N\mu_E-\theta)}\right). \tag{4.8}$$

This equation is important as, besides μ_E, it depends only on the parameters of the model. Once μ_E has been obtained from this equation—or a more sophisticated version of it described in Sec. A.1—by numerical methods, all the other quantities will immediately follow using the previously established equations. Eq. 4.8 also makes clear that the ratio λ/β is the only quantity that maters, as opposed to the precise values of λ and β. If we moreover choose, as we did in our simulations, $\theta = \alpha N$ $(0 < \alpha < 1)$, the network size disappears from the equation.

4.1.4. Illustration

Fig. 11 shows how Eq. 4.8 can be solved graphically. The typical parameter values of Sec. 3, $N = 50$, $\theta = 5$, $\beta = 10$, are used. The right-hand side of Eq. 4.8 is drawn in black for different values of λ ($\lambda = 6, 6.7, 7, 8, 9, 10, 11, 12$). Two additional curves (dashed blue) for $\lambda = 6.7$ and 9 are also drawn using a more precise estimate of the right hand-side of Eq. 4.8 described in Sec. A.1. The straight line in orange has slope 1 and the solutions of Eq. 4.8 correspond to the intersection points of that line with the black (or dashed blue) curves.

We see that if λ is too large (here slightly larger than 10) there is no intersection point and therefore no solution of Eq. 4.8, meaning no metastable state. This is what we inferred from our simulations (Fig. 7) although the critical point looked slightly smaller there. When λ decreases, the upper right intersection point moves farther up-right meaning that μ_θ is a decreasing function of λ. Since the network spiking rate is $\nu_N = \mu_\theta\beta = (N - \theta/\mu_E)\beta$ (Eq. 4.2 and 4.7), this implies that the network rate is also a decreasing function of λ.

4.2. How does our computations compare to the simulations?

The presence of two intersection points between the diagonal and the right hand side of Eq. 4.8 suggests there could be two metastable states. But we were unable to observe the "lower" one, suggesting that such a state is not metastable but just unstable.

We pointed out in our description of Fig. 11 that the network rate is a decreasing function of λ and this is precisely what we observe on Fig. 12 (the smaller λ, the steeper the slope).

We now set the model parameters, solve our implicit equation for μ_E and deduce from that ν_E, μ_θ and μ_F. We then compare these predicted values with empirical measures from simulations of the corresponding network. This is what is reported in the next two tables were 5 simulations were done in each of two settings that only differ by the value of θ.

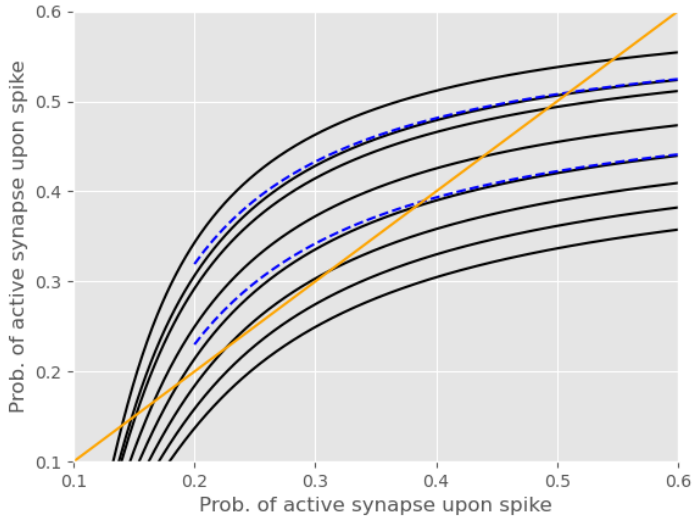

Figure 11: Examples with $N = 50$, $\theta = 5$, $\beta = 10$, $\lambda = 6, 6.7, 7, 8, 9, 10, 11, 12$ (top to bottom). Dashed blue lines are obtained in two cases by "numerical integration" (see next section).

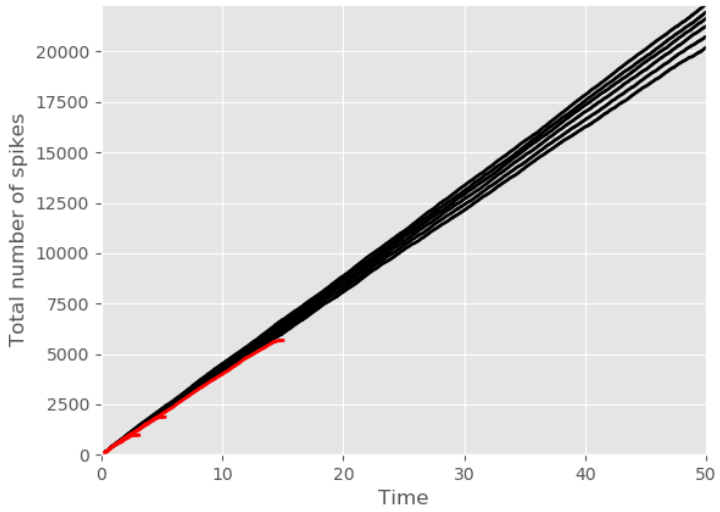

Figure 12: Observed counting processes of a network made of 50 neurons with $\theta = 5$, $\beta = 10$ and increasing values of λ from 1 to 9. In black, "top to bottom", $\lambda \in \{1, 2, \ldots, 6\}$; in red, $\lambda > 6$.

	1	2	3	4	5	analytical
ν_E	4080	4076	4072	4078	4081	4085
μ_θ	408.2	408.1	407.8	407.9	408.0	408.5
μ_F	309.3	309.6	308.8	308.6	309.2	308.8
μ_E	0.547	0.546	0.545	0.545	0.546	0.547

Table 1: Observed and predicted network properties in the metastable state of a network made of 500 neurons with $\theta = 50$, $\beta = 10$ and $\lambda = 6$. Columns 1 to 5 contain the empirical mean values obtained from 5 independent replicates, column "analytical" contain the predicted values obtained by solving the implict equation.

We see a rather precise match, despite of the fact that with $\mu_\theta \approx 408$, $N = 500$ and $\theta = 50$ we have only typically 82 neurons to populate the 50 states below threshold. So not even two neurons per state below θ. It is then rather optimistic to postulate that $\mathbb{E}(N_0(t)) \approx \mu_0, \ldots$ as we did.

	1	2	3	4	5	analytical
ν_E	4666	4661	4666	4665	4668	4666
μ_θ	466.6	466.5	466.5	466.5	466.5	466.6
μ_F	312.4	313.0	311.8	311.9	313.1	312.0
μ_E	0.600	0.599	0.598	0.598	0.599	0.599

Table 2: Observed and predicted network properties in the metastable state of a network made of 500 neurons with $\theta = 20$, $\beta = 10$ and $\lambda = 6$. Columns 1 to 5 contain the empirical mean values obtained from 5 independent replicates, column "analytical" contain the predicted values obtained by solving the implict equation.

Here the match is even better.

5. Conclusion and perspectives

We have suggested in this article that the sustained activity observed during delayed-response task experiments could very well be interpreted through the notion of metastability as initially rigorously characterized in the statistical physics context. To do so we have proposed a simple stochastic model of spiking neurons featuring a short-term synaptic plasticity mechanism, and we have shown by simulations and semi-rigorous reasoning that this model indeed presents a metastable behavior for suitably chosen parameters values. Nonetheless a rigorous mathematical proof remains to be obtained. As mentioned in the introduction a proof has been given already for the first of the two properties characterizing metastable systems–the asymptotic memorylessness of the time to extinction–in [3] and [4] for a slightly simpler model that did not include short-term synaptic plasticity. The general proof strategy, used in both articles, is to show that[2]:

[2]Remember that σ_N denotes the time of extinction of a system with N neurons.

$$\lim_{N\to\infty}\left|\mathbb{P}\left(\frac{\sigma_N}{\mathbb{E}(\sigma_N)}>s+t\right)-\mathbb{P}\left(\frac{\sigma_N}{\mathbb{E}(\sigma_N)}>s\right)\mathbb{P}\left(\frac{\sigma_N}{\mathbb{E}(\sigma_N)}>t\right)\right|=0. \tag{5.9}$$

From there the convergence in law toward and exponential random variable of mean 1 follows from standard arguments.

One of the problems is that the way (5.9) is obtained relies heavily on the Markovianity of the processes considered in [3, 4]. Indeed, even when considered as processes taking value in $\{0,1\}^N$–that is, the state space in which we only indicate whether a given neuron is active (susceptible to spike) or quiescent (not susceptible to spike)–the models considered in [3, 4] remain Markovian, which is not the case for the process of the present article. The process considered here is Markovian only if we consider it as a process living in the extended state space $(\mathbb{N}\times\{0,1\})^N$, that gives both the exact membrane potential and the facilitation state for each neuron[3]. Unfortunately this makes our process far less tractable than the previously considered ones.

Furthermore, the question of how to give a precise mathematical formulation of the second point in the metastability characterization remains to be answered. If we denote by $\xi_N(t)$ the state of some stochastic system at time t, taking value in some state space X^N, where N corresponds to the number of components in the system, a general approach [8, 32] consists in proving that there exists some non-trivial measure μ on $X^{\mathbb{N}}$ corresponding to the weak limit of the process $(\xi_N(t))_{t\geq 0}$ when N diverge, and to prove that before extinction the system behave as if it were described by this measure, restricted to the interval $[\![-N,N]\!]$. In other words the finite system behave like its infinite counterpart in invariant regime, as it would look like if we were observing it through some finite window. This property is sometimes called *thermalization* (see [32]) and it captures nicely the "pseudo-stationarity" of metastable systems that we would like to formalize.

The difficulty here comes from the underlying structure of the considered network. In some cases–such as the one studied in [3], where the interaction graph is a lattice–it is possible to define an increasing sequence of finite graphs such that the infinite counterpart (the union of all the finite graphs) has the same local structure as each of the graphs in the sequence. In such cases the existence of a non-trivial limit measure μ, as well as the possibility of establishing a thermalization result is generally not a problem. With complete interaction, as in the model considered here (as well as in the model from [4]), the average degree of the graph grows toward infinity when N diverges, the local structure isn't preserved and the establishment of such result is compromised.

Nonetheless models on complete graphs are generally of high interest as the specific topology of neural networks isn't very well known, making mean-field type approaches a good choice; furthermore the complete graph setting is analytically more manageable than other graphs as it doesn't carry any specific spatial structure (every neuron is virtually of the same importance in the network). Obtaining a well-defined mathematical approach of the pseudo-stationarity aspect of metastability for systems with complete interaction is therefore of primary interest.

[3] Actually, as it was already mentioned, $([\![0,\theta]\!]\times\{0,1\})^N$ is sufficient as whenever the membrane potential is bigger than θ the spiking rate remains unchanged until the next spike.

From a neurobiological viewpoint, if delayed-response tasks do indeed rely on the form of metastability discussed here, the probability of making a mistake (like in [14, Fig. 13]) should increase exponentially with delay duration. This could well be tested at the behavioral level on human subjects. It is also worth noticing that the continuous network activity modulation (Fig. 12) is reminiscent of the parametric working memory studied by [30]. That is, if it is possible to change the ratio λ/β, it is also possible to represent a continuously varying stimulus.

A. Appendix

A.1. Inter-spike interval distribution

We consider in this section the inter-spike interval (ISI) distribution, that is, the distribution of τ. Let X and Y be two random variables with $X \sim \text{Erlang}(\theta, \nu_E)$ and $Y \sim \text{Exponential}(\beta)$, and write $T = X + Y$. The distribution of the sum of an Erlang random variable and an exponential random variable is called an Hypoexponential distribution. One could give an explicit expression for its cumulative distribution function by using for example the results from [31] and [1], but the said expression is uselessly complicated so that we will prefer a numerical integration. Let G_T denote the cumulative distribution function of T, G_X the cumulative distribution function of X and f_Y the density function of Y. We have

$$G_T(t) = \int_0^t dy f_Y(y) \int_0^{t-y} dG_X(x),$$

which simplifies to

$$G_T(t) = \int_0^t dy f_Y(y) G_X(t-y) \tag{A.10}$$

Once our numerical integration is done we will be able to control the precision by comparing

$$\mu_\tau = \frac{1}{\beta}\left(\frac{\theta}{N\mu_E - \theta} + 1\right) \tag{A.11}$$

with the expectation of T, given by

$$\mathbb{E}(T) = \int_0^\infty dt(1 - G_T(t)). \tag{A.12}$$

Note also that (A.11) can be written

$$\mu_E = \int_0^\infty \exp(-\lambda t)\, dG_T(t),$$

which by an integration by part can be rewritten as follows

$$\mu_E = \int_0^\infty dt\lambda \exp(-\lambda t)\, G_T(t),$$

so that our numerical integration gives us another way of getting a value for μ_E. As illustrated on Fig. 11 (dotted blue curves), the difference is not huge compared to the result of the simpler approach of Sec. 4.1.3. Since our `C` implementation is fast, the results computed by our programs are nevertheless based on the numerical integration of Eq. A.10.

References

[1] S.V. Amari and R.B. Misra. Closed-form expressions for distribution of sum of exponential random variables. *IEEE Transactions on Reliability*, 46(4):519–522, 1997.

[2] Daniel Amit and Nicolas Brunel. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, 7(3):237–252, Apr 1997.

[3] Morgan André. A result of metastability for an infinite system of spiking neurons. *Journal of Statistical Physics*, 177(5):984–1008, Oct 2019.

[4] Morgan André and Léo Planche. The effect of graph connectivity on metastability in a stochastic system of spiking neurons. *Stochastic Processes and their Applications*, 131:292–310, Jan 2021.

[5] Omri Barak and Misha Tsodyks. Persistent activity in neural networks with dynamic synapses. *PLOS Computational Biology*, 3(2):1–1, 02 2007.

[6] Omri Barak and Misha Tsodyks. Working models of working memory. *Current Opinion in Neurobiology*, 25:20–24, Apr 2014.

[7] D Capocaccia, M Cassandro, and E Olivieri. A study of metastability in the ising model. *Commun.Math.Phys.*, 39:185–205, 1974.

[8] Marzio Cassandro, Antonio Galves, Enzo Olivieri, and Maria Eulália Vares. Metastable behavior of stochastic dynamics: A pathwise approach. *Journal of Statistical Physics*, 35(5-6):603–634, Jun 1984.

[9] Albert Compte, Nicolas Brunel, Patricia S. Goldman-Rakic, and Xiao-Jing Wang. Synaptic Mechanisms and Network Dynamics Underlying Spatial Working Memory in a Cortical Network Model. *Cerebral Cortex*, 10(9):910–923, 09 2000.

[10] Christos Constantinidis, Shintaro Funahashi, Daeyeol Lee, John D. Murray, Xue-Lian Qi, Min Wang, and Amy F.T. Arnsten. Persistent spiking activity underlies working memory. *The Journal of Neuroscience*, 38(32):7020–7028, Aug 2018.

[11] J. del Castillo and B. Katz. Quantal components of the end-plate potential. *The Journal of Physiology*, 124(3):560–573, Jun 1954.

[12] P. Fatt and B. Katz. Spontaneous subthreshold activity at motor nerve endings. *The Journal of Physiology*, 117(1):109–128, 1952.

[13] P. A. Ferrari, A. Galves, I. Grigorescu, and E. Löcherbach. Phase transition for infinite systems of spiking neurons. *Journal of Statistical Physics*, 172(6):1564–1575, Jul 2018.

[14] S. Funahashi, C. J. Bruce, and P. S. Goldman-Rakic. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61(2):331–349, 1989.

[15] J. Fuster. *The Prefrontal Cortex*. Elsevier Science, 2015.

[16] J M Fuster. Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. *Journal of Neurophysiology*, 36(1):61–78, Jan 1973.

[17] A. Galves, E. Löcherbach, C. Pouzat, and E. Presutti. A system of interacting neurons with short term synaptic facilitation. *Journal of Statistical Physics*, 178(4):869–892, Dec 2019.

[18] D. Hansel and G. Mato. Short-term plasticity explains irregular persistent activity in working memory tasks. *Journal of Neuroscience*, 33(1):133–149, Jan 2013.

[19] Vladimir Itskov, David Hansel, and Misha Tsodyks. Short-term facilitation may stabilize parametric working memory trace. *Frontiers in Computational Neuroscience*, 5:40, 2011.

[20] J. G. Kalbfleisch. *Probability and Statistical Inference. Volume 2: Statistical Inference*, volume 2. Springer-Verlag, 1985.

[21] B. KATZ and R. MILEDI. Membrane noise produced by acetylcholine. *Nature*, 226(5249):962–963, Jun 1970.

[22] Matthew L. Leavitt, Diego Mendoza-Halliday, and Julio C. Martinez-Trujillo. Sustained activity encoding working memories: Not fully distributed. *Trends in Neurosciences*, 40(6):328 – 346, 2017.

[23] J. K. Lindsey. *Statistical Analysis of Stochastic Processes in Time*. Cambridge University Press, Aug 2004.

[24] Liqun Luo. *Principles of Neurobiology*. Garland Science, Jul 2015.

[25] Eva Löcherbach and Pierre Monmarché. Metastability for systems of interacting neurons. arXiv:2004.13353v2, 2020.

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

[27] T.S. MOUNTFORD. A metastable result for the finite multidimensional contact process. *Canad. Math. Bull.*, 36:216–226, 1993.

[28] O. Penrose and J. L. Lebowitz. Rigorous treatment of metastable states in the van der waals-maxwell theory. *Journal of Statistical Physics*, 3(2):211–236, 1971.

[29] Cecilia Romaro, Fernando Araujo Najman, and Morgan André. A Numerical Study of the Time of Extinction in a Class of Systems of Spiking Neurons. arXiv:1911.02609v1, 2019.

[30] Ranulfo Romo, Carlos D. Brody, Adrián Hernández, and Luis Lemus. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, 399(6735):470–473, Jun 1999.

[31] E.M. Scheuer. Reliability of an m-out of-n system when component failure induces higher failure rates in survivors. *IEEE Transactions on Reliability*, 37(1):73–74, Apr 1988.

[32] R.H. SCHONMANN. Metastability for the contact process. *Journal of Statistical Physics*, 41:445–464, 1985.

[33] A.A. Verveen and H.E. Derksen. Fluctuation phenomena in nerve membrane. *Proceedings of the IEEE*, 56(6):906–916, 1968.

[34] X J Wang. Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci*, 24(8):455–63, Aug 2001.

[35] Yun Wang, Henry Markram, Philip H Goodman, Thomas K Berger, Junying Ma, and Patricia S Goldman-Rakic. Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nature Neuroscience*, 9(4):534–542, Mar 2006.

[36] Yosef Yarom and Jorn Hounsgaard. Voltage fluctuations in neurons: Signal or noise? *Physiological Reviews*, 91(3):917–929, Jul 2011.

[37] D Zipser, B Kehoe, G Littlewort, and J Fuster. A spiking network model of short-term active memory. *The Journal of Neuroscience*, 13(8):3406–3420, Aug 1993.