Working Memory and Metastability in a System of Spiking Neurons with Synaptic Plasticity

Christophe Pouzat
IRMA, Strasbourg University, CNRS
christophe.pouzat@math.unistra.fr
Joint work with Morgan André (UNICAMP)

BioHasard, June 3 2021

Outline

Contents

1 Introduction: delayed responses, working memory, persistent activity and all that 2

2 Definition of the model 11

3 Empirical results 13

4 Mean-field analysis 18

5 Conclusion and perspectives 21

1 Introduction: delayed responses, working memory, persistent activity and all that

It starts with Fuster in 1973

> 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.

Reference: Fuster J. (1973) *Unit Activity in Prefrontal Cortex During Delayed-Response Performance: Neuronal Correlates of Transient Memory. J. Neurophys.* **36**: 61-78.

Fuster's paradigm

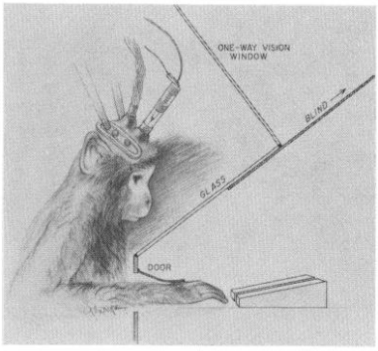


FIG. 1. Diagram of an experimental animal in the testing apparatus.

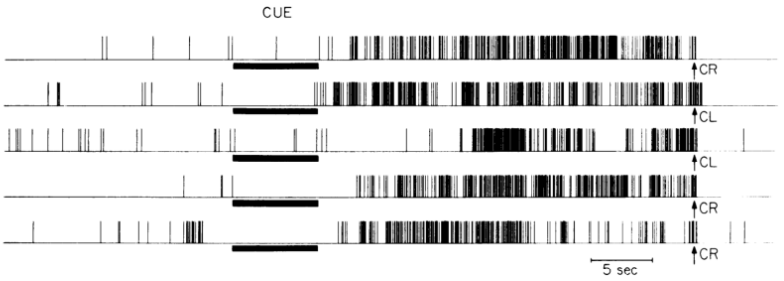


FIG. 4. Unit of type D during five delayed-response trials with 32-sec delay. Spikes are represented by vertical lines in a graphic display obtained by computer method. The notation next to the arrow at the end of each trial's delay refers to the accuracy (C, correct; I, incorrect) and side (R, right; L, left) of the response. The series of single-trial records in this figure—as in subsequent figures—is made of records from consecutive trials.

Figures 1 and 4 of Fuster (1973).

Other delayed activities are observed

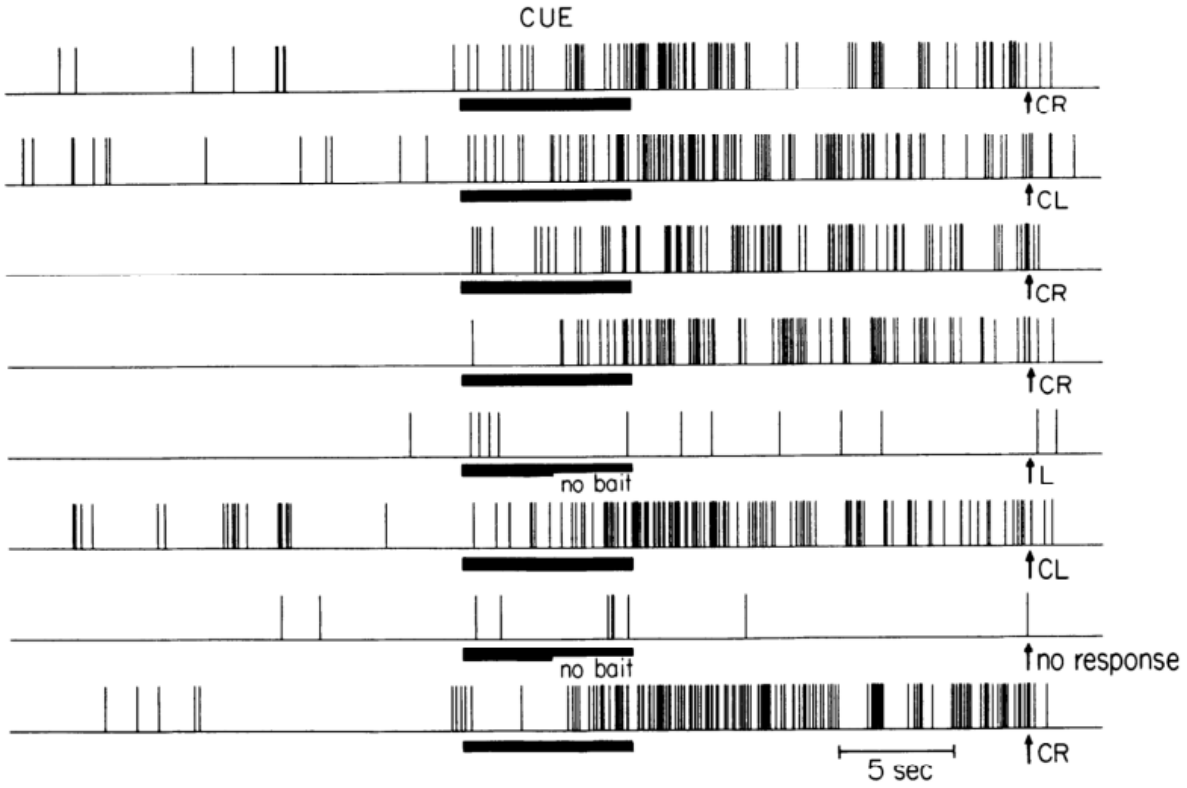
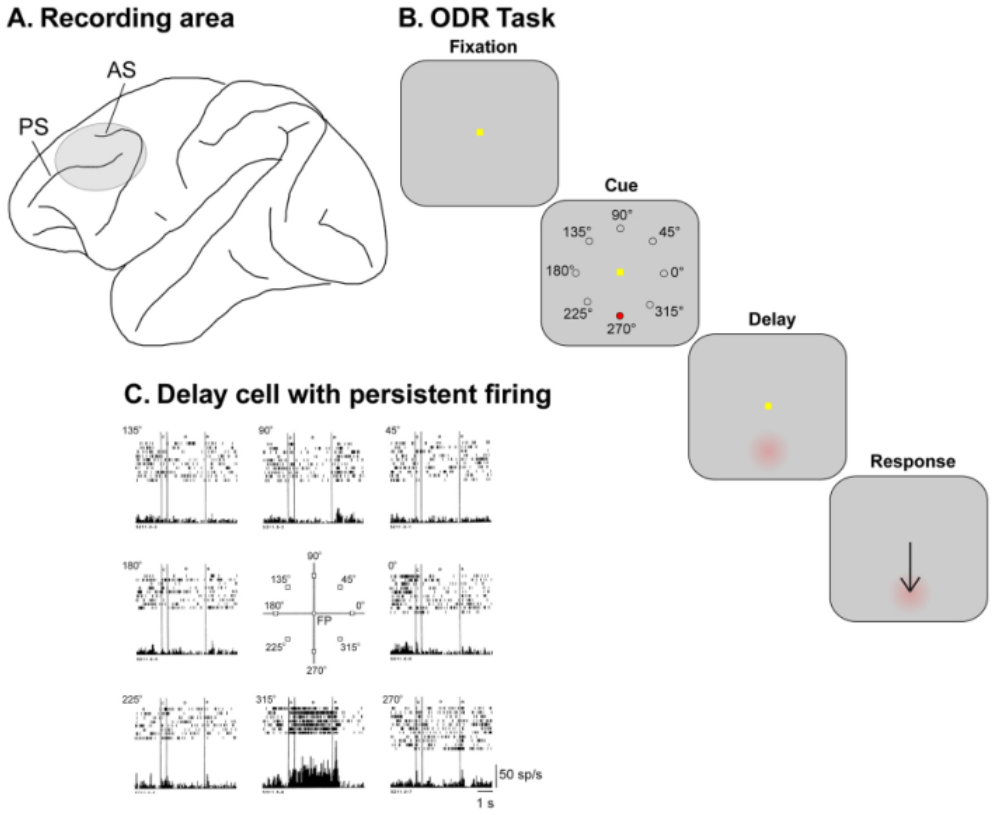


FIG. 6. Unit of type C. Note absence of sustained activation on dry-run trials (fifth and seventh).

Figure 6 of Fuster (1973).

A "modern" version of Fuster's paradigm

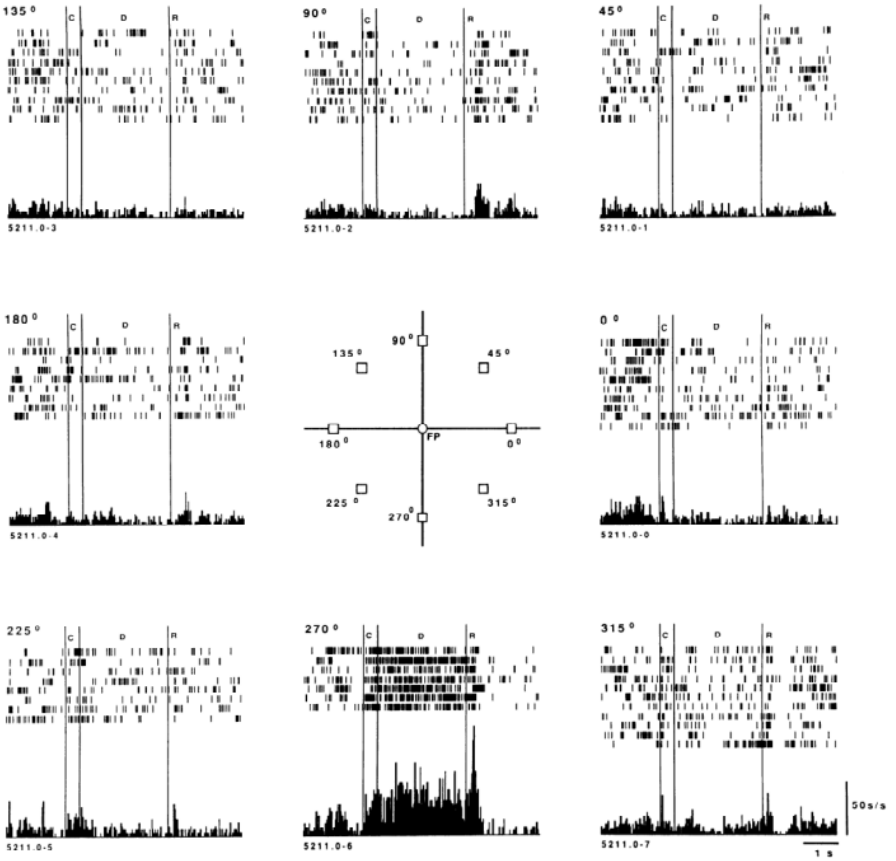


Adaptation of figures from Funahashi et al (1989) by Constantinidis et al (2018).

References:

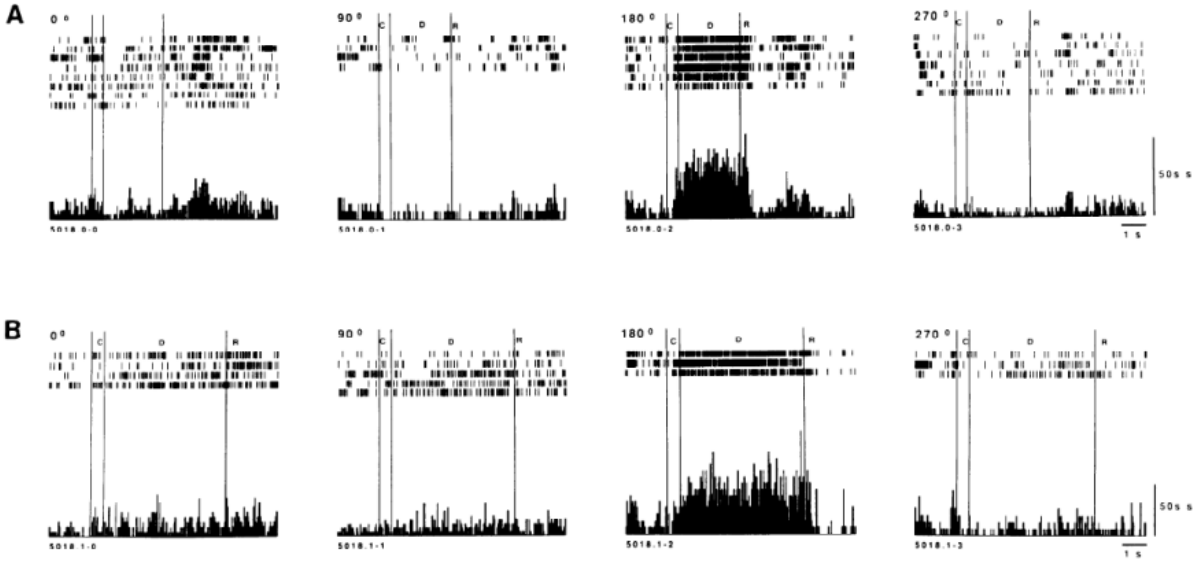
- S. Funahashi, C. J. Bruce, and P. S. Goldman-Rakic (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex . *J. Neurophys.* **61**: 341-349.
- Christos Constantinidis, Shintaro Funahashi, Daeyeol Lee, John D. Murray, Xue-Lian Qi, Min Wang and Amy F.T. Arnsten (2018) Persistent Spiking Activity Underlies Working Memory *Journal of Neuroscience* **38 (32)**: 7020-7028.

A better view of the rasters



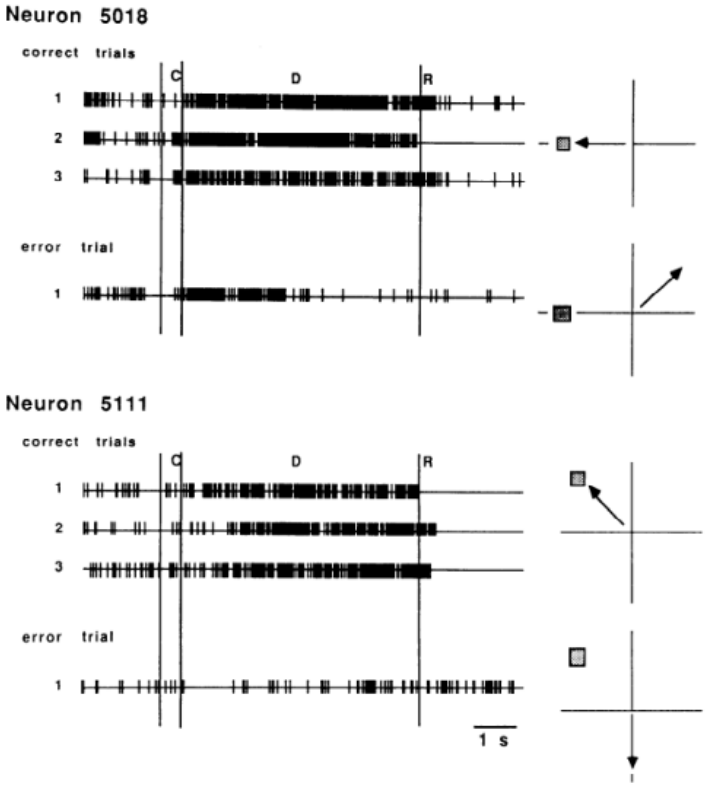
Funahashi et al (1989) Figure 3.

Changing the delay



Funahashi et al (1989) Figure 11.

What happens when mistakes are made?

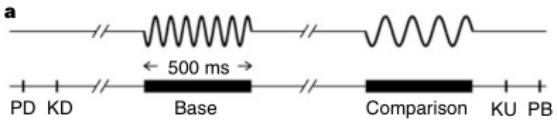


Funahashi et al (1989) Figure 13.

Neuronal correlates of parametric working memory in the prefrontal cortex

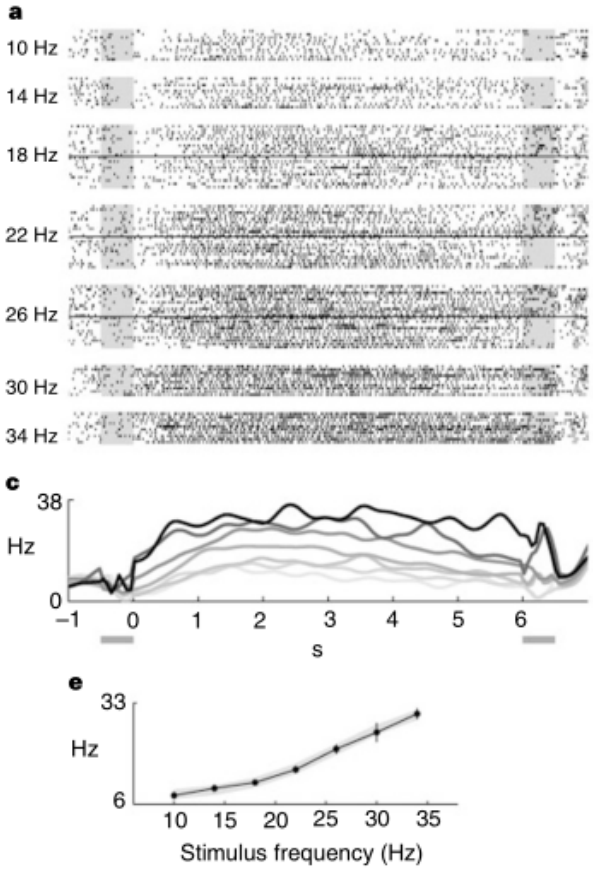
Ranulfo Romo, Carlos D. Brody, Adrián Hernández & Luis Lemus

Instituto de Fisiología Celular, Universidad Nacional Autonóma de México, México D.F. 04510, México



Romo et al (1999) title and figure 1a.

Reference: Romo, R., Brody, C., Hernández, A. et al. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* **399**, 470–473 (1999). https://doi.org/10.1038/20939.



Part of Romo et al (1999) figure 2.

First modelling efforts

Synaptic reverberation underlying mnemonic persistent activity

Xiao-Jing Wang

Stimulus-specific persistent neural activity is the neural process underlying active (working) memory. Since its discovery 30 years ago, mnemonic activity has been hypothesized to be sustained by synaptic reverberation in a recurrent circuit. Recently, experimental and modeling work has begun to test the reverberation hypothesis at the cellular level. Moreover, theory has been developed to describe memory storage of an analog stimulus (such as spatial location or eye position), in terms of continuous 'bump attractors' and 'line attractors'. This review summarizes new studies, and discusses insights and predictions from biophysically based models. The stability of a working memory network is recognized as a serious problem; stability can be achieved if reverberation is largely mediated by NMDA receptors at recurrent synapses.

persistent activity to subserve working memory, it must be stimulus-selective, and therefore information-specific. Moreover, it must be able to be turned on and switched off rapidly (≈100 ms) by transient inputs.

For 30 years, persistent activity in the cortex has been documented by numerous unit recordings from behaving monkeys during working memory tasks (Box 1). How does stimulus-selective persistent activity arise in a neural network? Can we explain persistent activity in terms of the biophysics of neurons and synapses, and circuit connectivity?

Reference: Wang XJ. Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.* 2001 Aug;**24(8)**:455-63. doi: 10.1016/s0166-2236(00)01868-3.

Cellular substrate

Heterogeneity in the pyramidal network of the medial prefrontal cortex

Yun Wang[1], Henry Markram[2], Philip H Goodman[3], Thomas K Berger[2], Junying Ma[1] & Patricia S Goldman-Rakic[4,5]

The prefrontal cortex is specially adapted to generate persistent activity that outlasts stimuli and is resistant to distractors, presumed to be the basis of working memory. The pyramidal network that supports this activity is unknown. Multineuron patch-clamp recordings in the ferret medial prefrontal cortex showed a heterogeneity of synapses interconnecting distinct subnetworks of different pyramidal cells. One subnetwork was similar to the pyramidal network commonly found in primary sensory areas, consisting of accommodating pyramidal cells interconnected with depressing synapses. The other subnetwork contained complex pyramidal cells with dual apical dendrites displaying nonaccommodating discharge patterns; these cells were hyper-reciprocally connected with facilitating synapses displaying pronounced synaptic augmentation and post-tetanic potentiation. These cellular, synaptic and network properties could amplify recurrent interactions between pyramidal neurons and support persistent activity in the prefrontal cortex.

Reference: Wang, Y., Markram, H., Goodman, P. H., Berger, T. K., Ma, J., & Goldman-Rakic, P. S. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nature Neuroscience*, **9(4)**, 534–542. doi:10.1038/nn1670.

This is not the whole story!

- NMDA receptors are also involved: Min Wang, Yang Yang, Ching-Jung Wang, Nao J. Gamo, Lu E. Jin, James A. Mazer, John H. Morrison, Xiao-Jing Wang, Amy F.T. Arnsten (2013) NMDA Receptors Subserve Persistent Neuronal Firing during Working Memory in Dorsolateral Prefrontal Cortex. *Neuron*, **77 (4)**: 736-749.
- Dopamine also plays a key role: MIN WANG, SUSHEEL VIJAYRAGHAVAN, PATRICIA S. GOLDMAN-RAKIC (2004) Selective D2 Receptor Actions on the Functional Circuitry of Working Memory. *SCIENCE*, **303**: 853-856

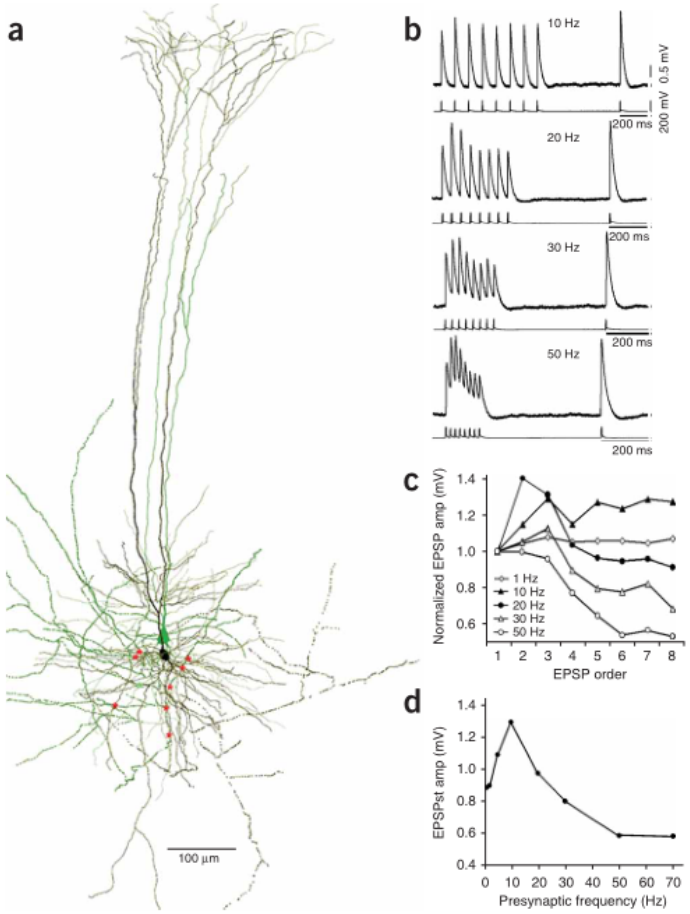


Figure 1 of Wang et al (2006).

Models with short term facilitation

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Neurobiology

ELSEVIER

Working models of working memory

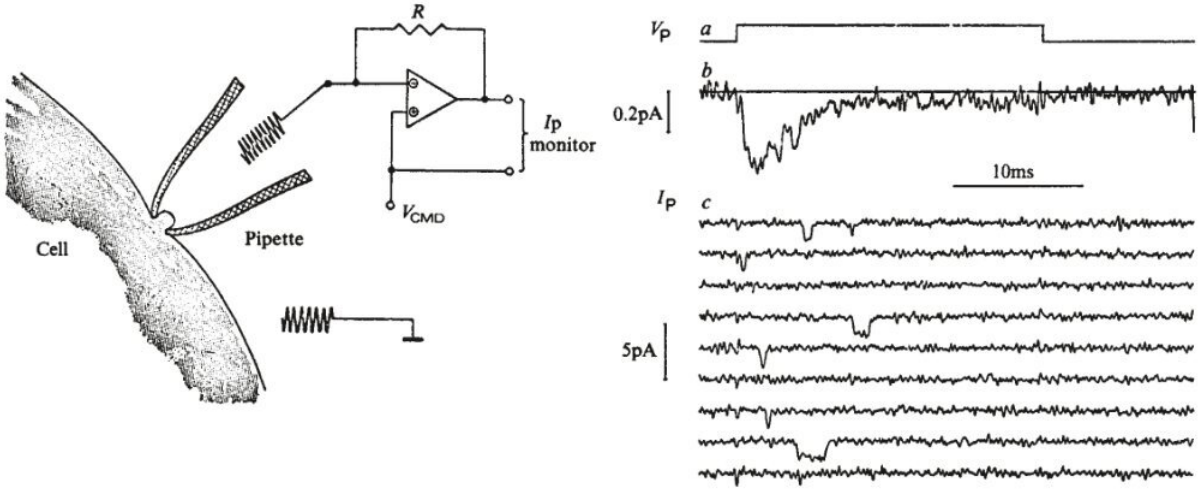
Omri Barak[1] and Misha Tsodyks[2]

Working memory is a system that maintains and manipulates information for several seconds during the planning and execution of many cognitive tasks. Traditionally, it was believed that the neuronal underpinning of working memory is stationary persistent firing of selective neuronal populations. Recent advances introduced new ideas regarding possible mechanisms of working memory, such as short-term synaptic facilitation, precise tuning of recurrent excitation and inhibition, and intrinsic network dynamics. These ideas are motivated by computational considerations and careful analysis of experimental data. Taken together, they may indicate the plethora of different processes underlying working memory in the brain.

activity related to storing a fixed item is not stationary, and there is a large heterogeneity in the firing profiles of different neurons [3,4,5•,6]. From the computational side, the network activity representing a memorized item should exhibit a sufficient degree of stability to ensure memory retainment. This requirement is especially challenging for storing continuous variables, such as orientation or spatial position of a visual cue, because of an inevitable drift along the variable's representation. Furthermore, integrating the various data-driven challenges in a self-consistent manner is often a non-trivial computational problem.

Reference: Omri Barak, Misha Tsodyks (2014) Working models of working memory, *Current Opinion in Neurobiology*, **25**: 20-24.

Membrane conductances (ion channels) generate fluctuations



Figures 1 and 2 of Sigworth and Neher (1980). Reference: Sigworth, F. J., & Neher, E. (1980). Single Na+ channel currents observed in cultured rat muscle cells. *Nature*, **287**: 447-449.

Synapses generate even more fluctuations

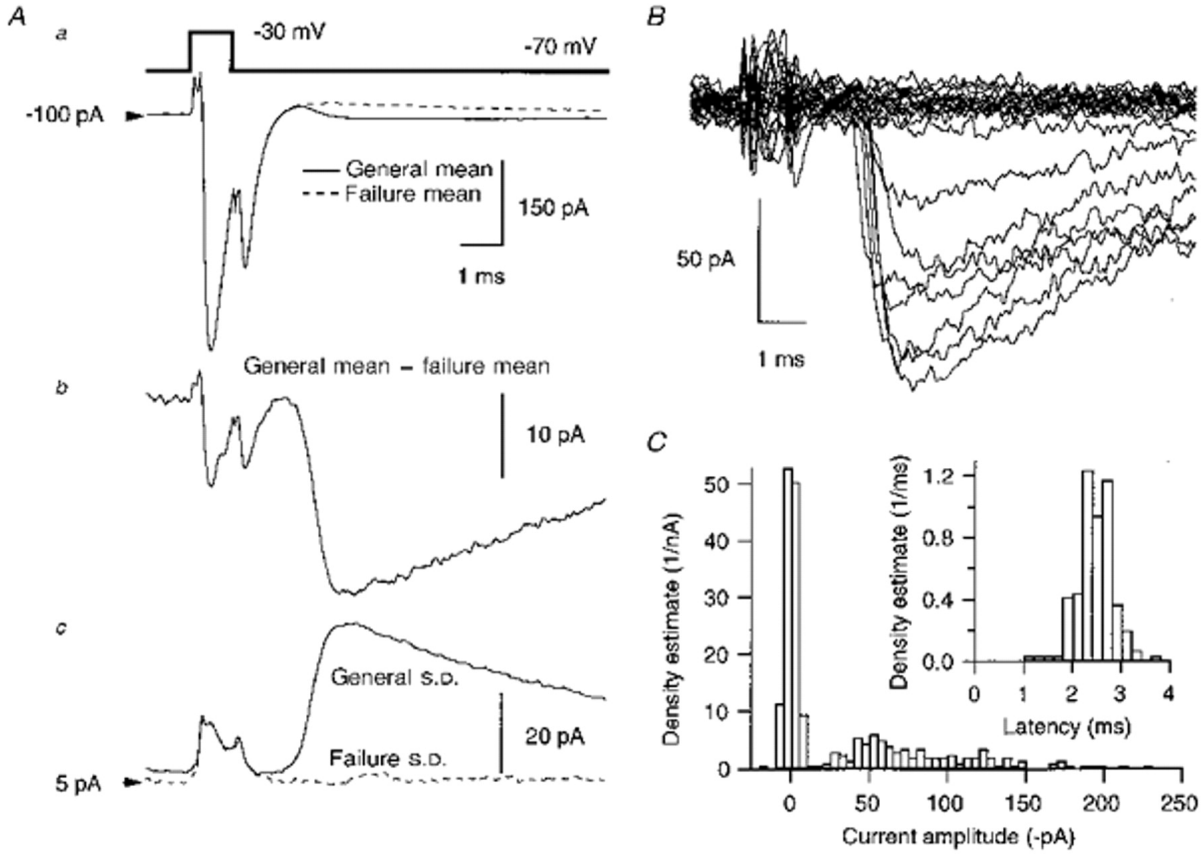


Figure 1 of Pouzat and Marty (1998).

Reference: Pouzat, C., & Marty, A. (1998). Autaptic inhibitory currents recorded from interneurones in rat cerebellar slices. *The Journal of Physiology*, **509(Pt 3)**, 777.

2 Definition of the model

Definition

- The system consists in a finite set of N identical neurons.

- Each neuron is synaptically connected to all the others.

- Each neuron $i \in \{1, \dots N\}$ is associated with a membrane potential denoted $(U_i(t))_{t\geq 0}$, taking value in $\mathbb{N}$.

- There is a threshold $\theta \in \mathbb{N}$. If $U_i(t) < \theta$ neuron i cannot spike, while if $U_i(t) \geq \theta$ it spikes at rate β.

- When a neuron spikes its membrane potential is reset to zero. That's the only way the membrane potential can decrease.

- Each neuron i has a facilitation state evolving with t, we denote it $(F_i(t))_{t\geq 0}$ and it takes value in $\{0, 1\}$.

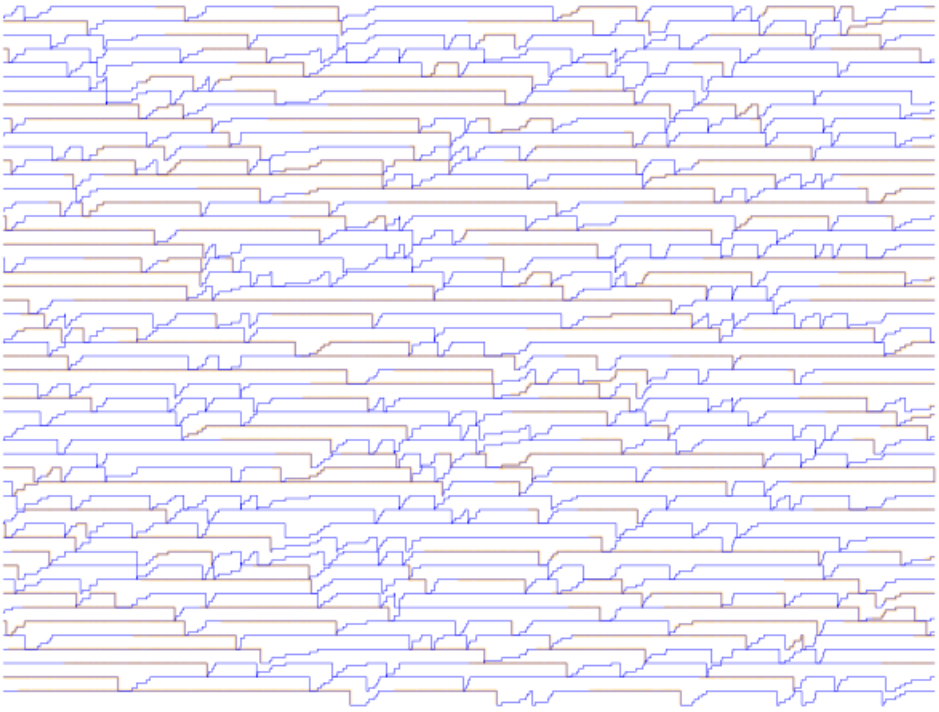
- If $F_i(t) = 1$ and a spike occurs at time t for neuron i, then the membrane potential of every neuron is incremented by 1.

- If $F_i(t) = 0$ the spike has no post-synaptic effect.

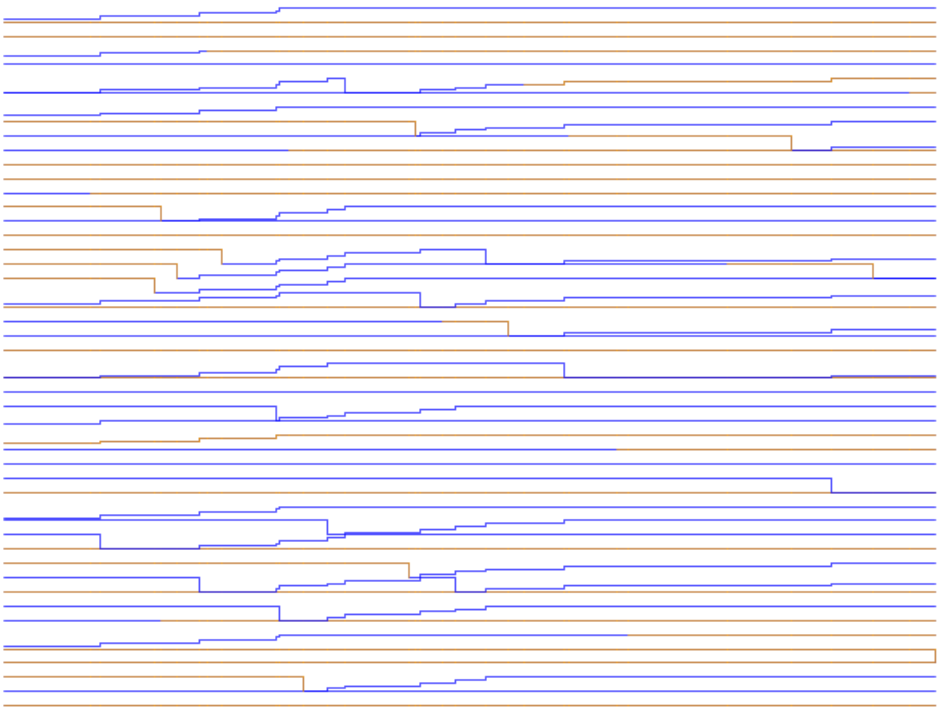
- The facilitation state of a given neuron is set to 1 immediately after a spike has been emitted by this neuron, then the facilitation is lost at rate λ.

- We are here modelling the sub-network of strongly interconnected pyramidal cells with facilitating synapses described by Wang et al (2006) in the prefrontal cortex.

In picture



Simulation with $N = 50$, $\beta = 10$, $\lambda = 6.7$ and $\theta = 5$ between time 1 and 2.



Zoom between time 1.20 and 1.25.

3 Empirical results

Simulations outline

Simulations are easily performed since the "global" network rate is constant between two successive events (spike or facilitation loss). Our `C` code writes to disk:

```
# Simulation of a networks with 50 neurons
# Xoroshiro128+ PRNG seeds set at 18710305 and 1857075
# The initial max membrane potential was set to 50
# The initial probability for a synapse to be active was set to 0.750000
# Parameter theta = 5.000000
# Parameter beta = 10.000000
# Parameter lambda = 6.700000
# Simulation duration = 50.000000

# Spike time  Total nb of spikes  Neuron of origin  Neurons >= theta  N synapse active  f=1 at spike
0.0018467869                   1                28                45                38             1
0.0051172237                   2                49                44                39             0
0.0078398923                   3                41                44                37             1
0.0132602453                   4                47                43                35             1
0.0140281557                   5                16                44                35             1
...
```

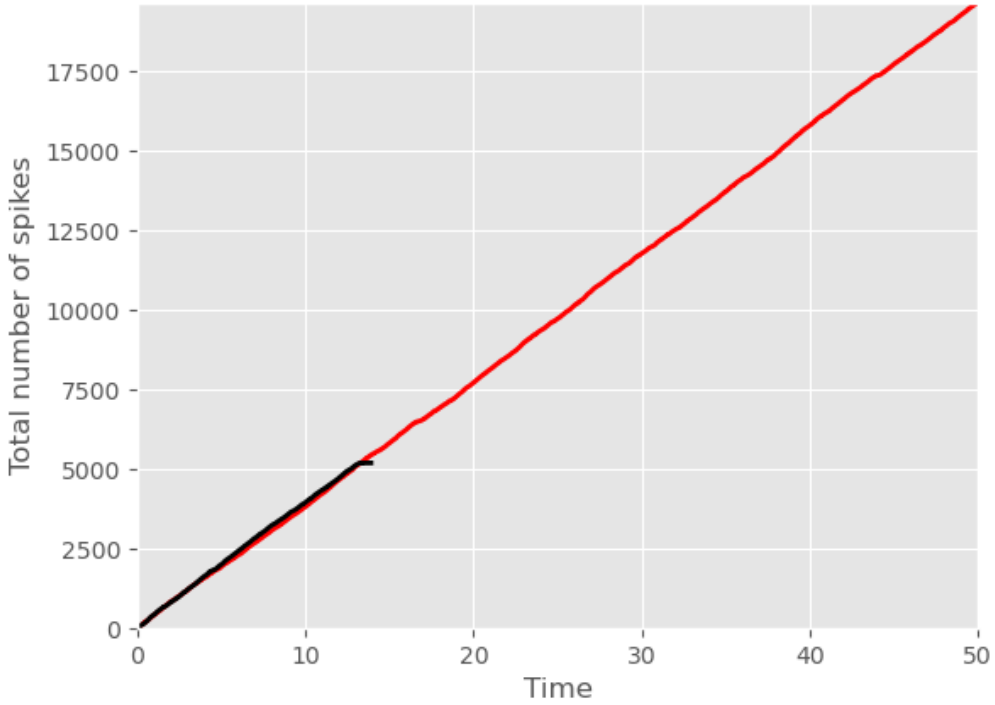
Tiny network example

Raster plots of 50 neurons network, with $\lambda = 6.7$, $\beta = 10$ and $\theta = 5$. The initial probability for the synapses to be active was 0.75, the initial membrane potentials were drawn uniformly on $\{0, 1, \ldots, 49\}$. Left, from time 0 to 14; right from time 12 to 14.

Same network different seed

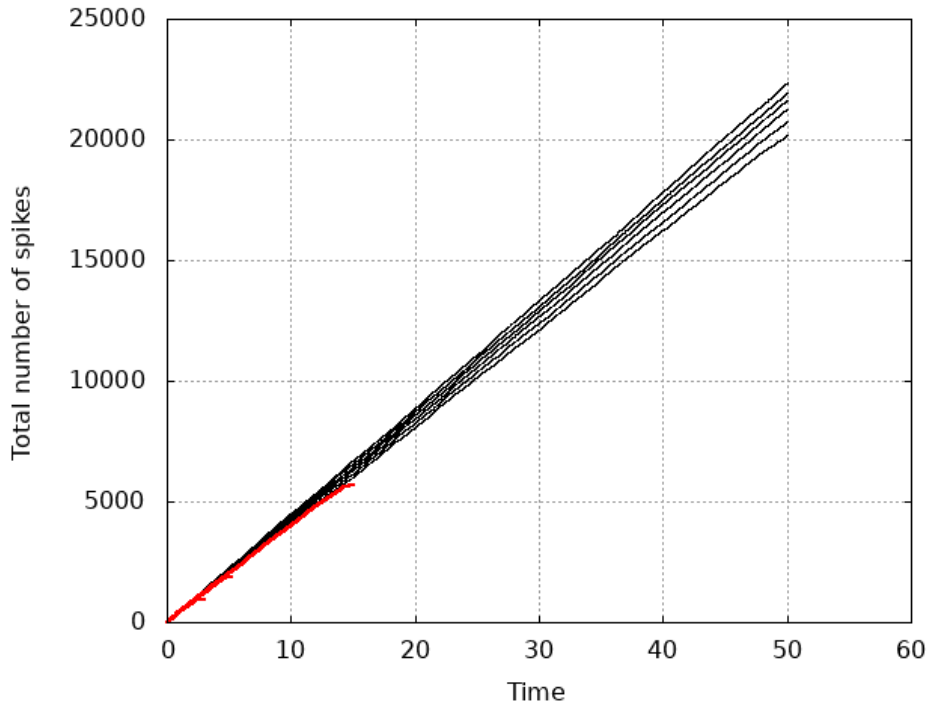
The scale bar is drawn between time 10 and time 15.

The counting process representation



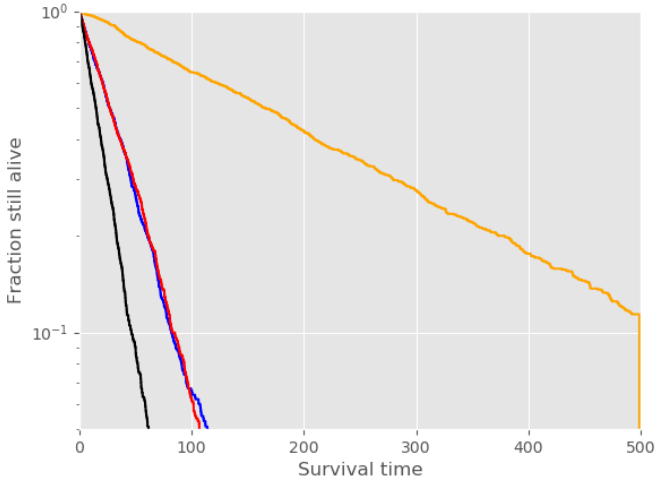
The two previous simulations, the first in black, the second in red.

Increasing λ



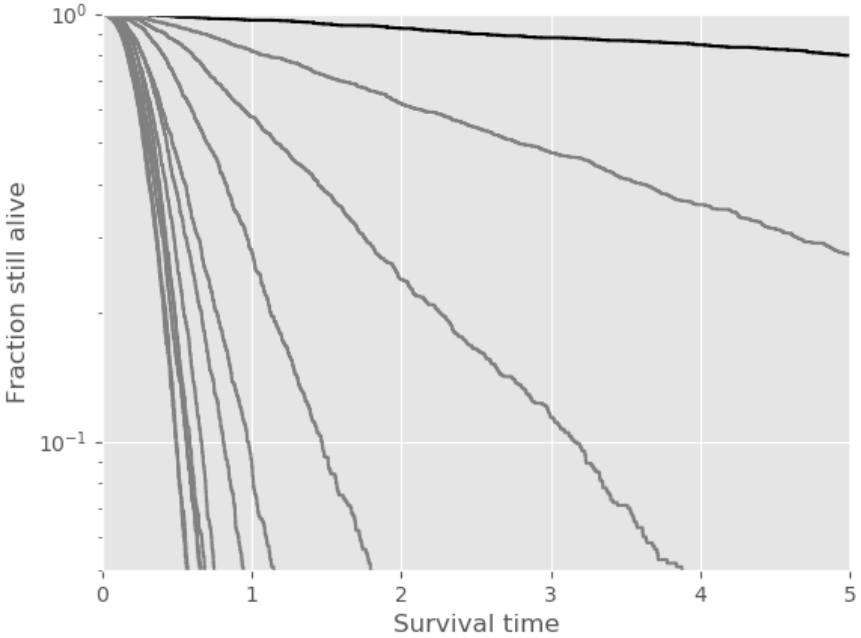
Observed counting processes of a network made of 50 neurons with increasing values of λ from 1 to 9. In black, "top to bottom", $\lambda \in \{1, 2, \ldots, 6\}$; in red, $\lambda > 6$.

Survival time distribution



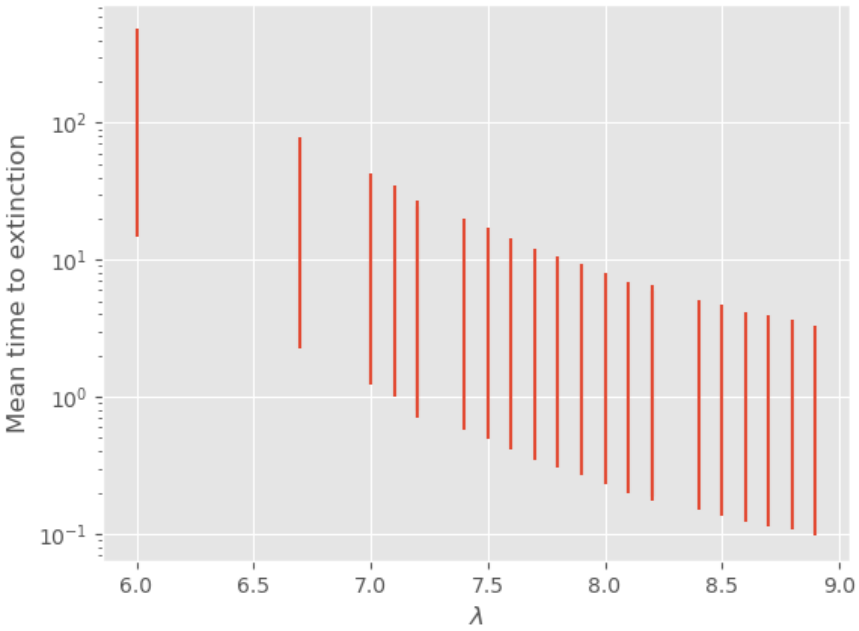
Empirical survival functions obtained from 1000 replicates with $\theta = 5$, $\lambda = 6.7$ (red and blue), $\lambda = 7$ (black) and $\lambda = 6$ (orange), $\beta = 10$ and a network with 50 neurons. The initial probability for the synapses to be active was 0.75, the initial membrane potentials were drawn uniformly on $\{0, 1, \ldots, 49\}$. All simulations start from *the same* random initial state except the red and blue ones. **A log scale is used for the ordinate**.

Survival time when λ is "too" large



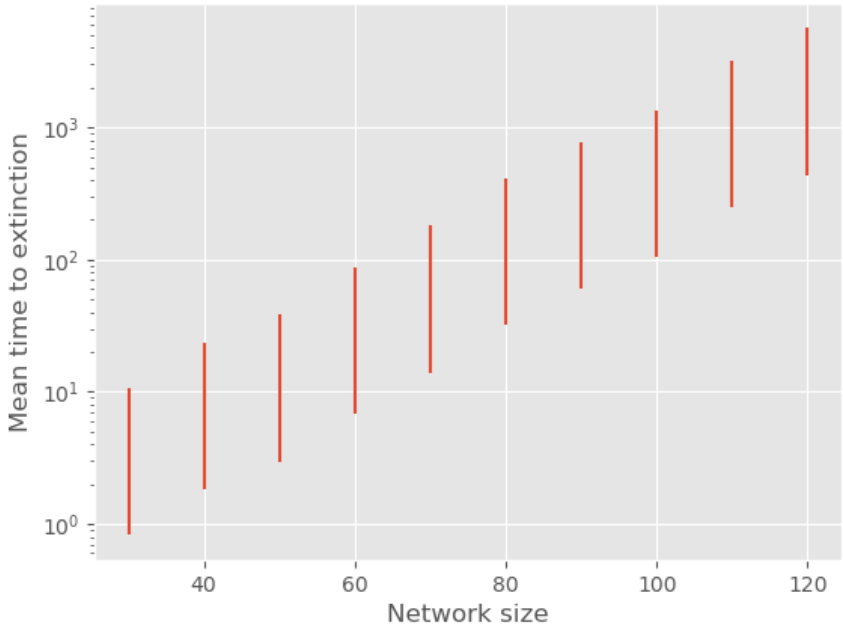
Same as before with $\lambda = 7$ (black) and $\lambda = 8, \ldots, 18$ (grey).

Survival time *vs* λ



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**.

Survival time *vs* Network size



95 % CI of the mean time to extinction 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**.

4 Mean-field analysis

What can we do, what do we want?

- We cannot yet prove that the metastable state exists.
- We will therefore postulate that it does: that's what the simulations show.
- We will use the intrinsic symmetry of the model: the neurons are all equivalent.
- We will try to get network properties in the metastable state:
  - network firing rate
  - number of neurons in each state
  - number of facilitated synapses
  - ...

  from the 4 network parameters: N, θ, β, λ.

Notations and remarks

- We have $(U_i(t))_{t\geq 0} \in \mathbb{N}$, but from the network dynamics what matters is to know whether $U_i(t) \geq \theta$ or not.
- We then have to consider $\theta + 1$ different *states* for $U_i(t)$: $\{0, 1, \ldots, \theta - 1, \geq \theta\}$, that is, θ states below threshold and 1 state above.
- Let us write
  - $N_i(t)$ for $i \in \{0, 1, \ldots, \theta - 1\}$ the number of neurons whose membrane potential equals i
  - $N_\theta(t)$ the number of neurons whose membrane potential is $\geq \theta$

  at time t.
- We obviously have: $\sum_{i=0}^{\theta} N_i(t) = N$ at all times.
- *Then under our assumption of quasi-stationarity, the expectations of the N_i should be almost constant in the metastable phase.*
- Thus we let $\mu_0, \mu_1, \ldots \mu_\theta$ be the constants such that $\mathbb{E}(N_0(t)) \approx \mu_0$, $\ldots$, $\mathbb{E}(N_\theta(t)) \approx \mu_\theta$, where t is any time before the extinction of the system.

Another key quantity

- If we manage to compute μ_θ, we know the approximate network rate at anytime (before extinction): $\nu_N = \mu_\theta \beta$.
- In our model, when neuron j spikes at time s we have $F_j(s+) = 1$, the question is:
- *if the next spike of j happens at time $s + \tau$, do we still have $F_j(s + \tau) = 1$?*
- By our model definition *and our quasi-stationarity assumption* we have: $\mathbb{E}\left[F_j(s + \tau)|\tau\right] = e^{-\lambda\tau}$.
- We introduce now our second "key" quantity:

$$\mu_E = \mathbb{E}\left(e^{-\lambda\tau}\right),$$

  where the expectation is taken with respect to the unknown distribution of the conditioning *rv* T whose realization is τ.
- μ_E is the "mean probability" that the synapse is still facilitated when the neuron spikes.

Circulation among U states

- Remark that μ_E allows us to define the rate of "effective" spikes (spikes that have a post-synaptic effect): $\mu_\theta \beta \mu_E$.
- Stationarity means that the rate at which neurons leave membrane potential state $i \in \{0, 1, \ldots, \theta - 1, \geq \theta\}$ must equal the rate at which neurons enter that state.
- For $i \in \{1, \ldots, \theta - 1\}$ this translates into:

$$(\mu_\theta \beta \mu_E)\mu_i = (\mu_\theta \beta \mu_E)\mu_{i-1}\,,$$

that is:

$$\mu_0 = \mu_1 = \cdots = \mu_{\theta-1}\,.$$

- For the two extrem states, we have:

$$(\mu_\theta \beta \mu_E)\mu_0 = \mu_\theta \beta\,,$$

leading to

$$\mu_0 = 1/\mu_E\,.$$

- But we have:

$$\sum_{i=0}^{\theta-1} \mu_i + \mu_\theta = N\,.$$

- Using the equality of the μ_i for $i < \theta$ and our last equality ($\mu_0 = 1/\mu_E$), yields:

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

- We see that is μ_E increases, so does μ_θ and therefore $\nu_N = \mu_\theta \beta$, the network spike rate.
- We can also obtain a new expression for the rate of "effective" spikes:

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

Getting an implicit equation for μ_E

- In the metastable state, a neuron leaves a membrane potential state below threshold at rate: $\beta(\mu_E N - \theta)$.
- That neuron must go through a succession of θ states to reach threshold, the distribution of the time to reach threshold is therefore an Erlang distribution with parameters θ and $\beta(\mu_E N - \theta)$ and its mean value is:

$$\frac{\theta}{\beta(\mu_E N - \theta)}\,.$$

- Once threshold has been reach, the rate at which a spike is generated is β so the interval between two successive spikes of a given neuron is approximately

$$T \approx \frac{\theta}{\beta(\mu_E N - \theta)} + Y\,,$$

where Y is an exponential random variable with rate parameter β.

- Remember that $\mu_E = \mathbb{E}\left[\exp(-\lambda T)\right]$.
- We therefore have:

$$\mu_E \approx \int_0^\infty \exp\left[-\lambda\left(\frac{\theta}{\beta(\mu_E N-\theta)}+y\right)\right]\beta\exp(-\beta y)dy\,,$$

that is

$$\mu_E \approx \left[\exp\left(-\frac{\lambda\theta}{\beta(\mu_E N-\theta)}\right)\right]\int_0^\infty \beta\exp\left(-(\lambda+\beta)y\right)dy\,.$$

- Leading to:

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

- *This is an implicit equation we must solve for μ_E.*

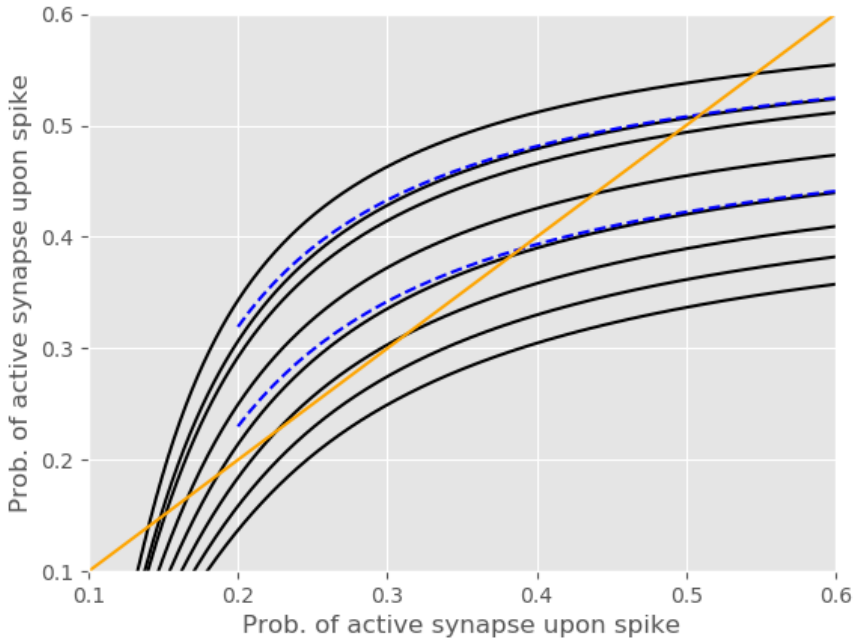
Remarks

- We can do better than that and work with the distribution of the Erlang random variable–giving the time spent below threshold—instead of the mean of the latter as we just did.
- This requires a numerical integration whose precision we can check.
- Looking at:

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

we see that the right hand side is a decreasing function of λ, so if λ is too large the equation could have no solution implying that there is no metastable state as we saw in the simulations.

Graphical solution of the implicit equation



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".

Comparison between mean-field solution and simulations

The implicit equation solution gives:

```
With N=500, beta=10.0, lambda=6.0, ceil_theta=50 we get:
  [...]
  nu_E       = 4085.11 (network spiking rate),
  mu_theta   = 408.51 (mean nb of neurons at or above threshold),
  mu_A       = 308.76 (mean nb of active synapses),
  mu_E       = 0.547 (prob of active synapse upon spike).
```

One numerical simulation gives:

```
Dealing with sim_n500_u50_f0p75_b10_l6_sim1_neuron:
[...]
*** Network level statistics ****
Ignoring 10 time unit(s) at both ends we get:
  nu_E     = 4079.86 [4069,4091] (empirical network spiking rate and 95% CI)
  mu_theta = 408.18 (empirical mean nb of neurons at or above threshold)
  mu_A     = 309.31 (empirical mean nb of active synapses)
  mu_E     = 0.547 (fraction of active synapse upon spike).
```

5 Conclusion and perspectives

Conclusion and perspectives

- The serious work just begins: *we must prove the existence of the metastable sate.*
- The potential tuning of short-term facilitation (changing our λ) does not seem to have been studied by experimentalists; so we will try to convince some of doing so.
- The exponential loss of memory in delayed response paradigms implied by our metastable state hypothesis could be tested with psychological experiments on humans.

Thank you all for listening!

The End