



HAL
open science

Lifetime and stability in line attractor networks of short-term memory

Pedro Goncalves, Christian Machens

► **To cite this version:**

Pedro Goncalves, Christian Machens. Lifetime and stability in line attractor networks of short-term memory. Deuxième conférence française de Neurosciences Computationnelles, "Neurocomp08", Oct 2008, Marseille, France. hal-00331609

HAL Id: hal-00331609

<https://hal.science/hal-00331609>

Submitted on 17 Oct 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LIFETIME AND STABILITY IN LINE ATTRACTOR NETWORKS OF SHORT-TERM MEMORY

Pedro Gonçalves
 Group for Neural Theory
 Ecole Normale Supérieure
 Paris, France
 email: pedro.goncalves@ens.fr

Christian Machens
 Group for Neural Theory
 Ecole Normale Supérieure
 Paris, France
 email: christian.machens@ens.fr

ABSTRACT

Line attractor networks have long served as the standard model of short-term memory systems for analogue variables. In this study, we investigate the stability of attractor states for a line attractor with monotonic tuning curves. We furthermore quantify the stability of network states against noise and show how the lifetime of short-term memory states depends on the level of neural noise.

KEY WORDS

short-term memory, continuous attractor, robustness

1 Introduction

Short-term memory is the ability to store information for a short period of time (typically ~ 10 seconds). This ability is often reflected in the persistent activity of neurons in various regions of the brain (famous examples include the oculomotor system and the prefrontal cortex) [1][2]. When short-term memory is concerned with analogue values, then neurons will usually sustain activity for a whole range of firing rates, a phenomenon that has been dubbed “graded” persistent activity.

A typical example of graded persistent activity is the representation of eye position in the oculomotor system of the goldfish (Fig. 1). As shown in Fig. 1, the firing rate of this neuron persists around a value that depends on the eye position. As eye position changes, so does the neuron’s firing rate. Note that for certain eye positions, the firing rates are approximately zero, suggesting a threshold below which the neuron does not fire. Above this threshold, the firing rate grows approximately linear with eye position. Similar types of persistent activity feature have been found in working memory tasks with different animal models [3].

Graded persistent activity has generated a lot of theoretical interest and several pieces of evidence suggest that it is a network phenomenon [2][3][4]. Since the neural firing rates change gradually with the stored memory variable, the respective networks seem to have a *continuum* of stable and stationary network states (a line attractor), so that each memory corresponds to exactly one stable point.

Such line attractor networks are most easily generated in networks of linear neurons; with biophysically more realistic, non-linear neurons, however, one can only construct

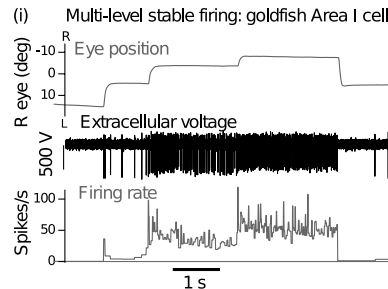


Figure 1. From [2]. Multi-level stable persistent firing in an oculomotor neural integrator cell in a goldfish. Top: horizontal eye position, measured in the dark. Rapid saccades alternate with stable fixations. Middle: extracellularly recorded action potentials. Bottom: instantaneous firing rate.

a network featuring multiple point attractors that are ordered along a line [5]. Such non-linear line attractor networks face several (known) problems, among these a lack of robustness of the memory against small perturbations in various biophysical parameters and a lack of robustness of the memory against noise in the system.

In this study, we investigate the lifetime of stable states and the stability against perturbations, as a way to understand more deeply these working memory systems.

2 The non linear model

In [5] and [6], the authors have developed a firing-rate model of a line attractor network which differs from previous (linear) models by the introduction of sigmoidal input-output functions g for the neurons. The precise shape of these functions does not matter for further theoretical development, it can be matched to the threshold and saturation features found on experimental data, for instance. In this section, we will summarize the main features of the model.

The firing rate of the i -th neuron, x_i , is given by:

$$\tau \dot{x}_i(t) = -x_i(t) + g\left(\sum_{j=1}^n T_{ij} x_j(t) + h_i\right), \quad i = 1, \dots, n, \quad (1)$$

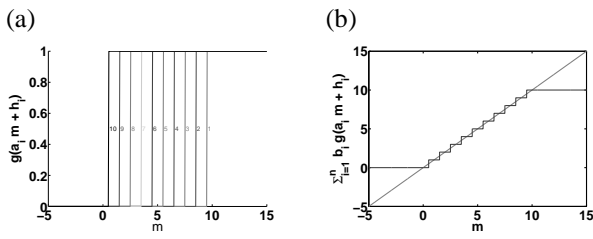


Figure 2. (a) Sigmoidal input-output functions for a network of 10 neurons. For simplicity, we use heaviside functions. (b) The weighted sum of the 10 heaviside functions versus m . This sum oscillates around the identity line.

where g describes the sigmoidal input-output function of the neuron, h_i the external input to neuron i , and T_{ij} the connectivity matrix of the network.

The connectivity matrix is assumed to be a rank-1 matrix,

$$T_{ij} = a_i b_j, \quad (2)$$

which in turn leads to a projection of the network dynamics onto a one-dimensional memory variable,

$$m(t) = \sum_{i=1}^n b_i x_i(t) \quad (3)$$

This variable is then sufficient to describe the essentials of the network dynamics and, using equations (1), (2) and (3), we obtain a reduced, one-dimensional description of the network

$$\tau \dot{m}(t) = -m(t) + \sum_{i=1}^n b_i g(a_i m(t) + h_i) \quad (4)$$

To look into the possible attractor states of the network, we will study the stationary solutions of this equation, i.e., $\dot{m}(t) = 0$, so that

$$m_s = \sum_{i=1}^n b_i g(a_i m_s + h_i) \quad (5)$$

Accordingly, the stationary states correspond to the points of perfect match between a sum of sigmoidal functions and a line. Due to the non-linear shape of the input-output functions, this match can only hold for a discrete set of points, but not for a continuum of points. In other words, we can only approximate a line attractor through a series of fixed points. Several approximations are possible, a particular easy one is to choose $a_i = 1$ and $b_i = 1$ and span the range of m -s by choosing different h_i for each neuron. Next, we will look into that possibility.

In Fig. 2 we use a network of 10 neurons, and show the input-output functions of the neurons as well as their sum (with weights $b_i = 1$).

A perfect match between the weighted sum and the identity line would mean every value of m is a stationary

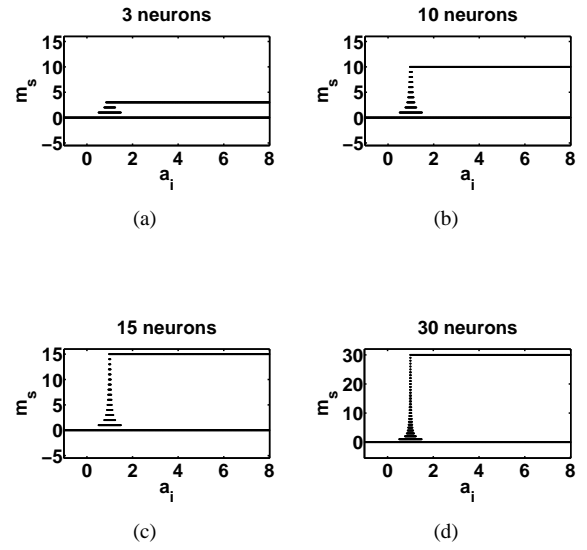


Figure 3. Bifurcation diagrams of the m equation in relation to the a_i parameter for different network sizes n . In black, the stable fixed points. The unstable fixed points are not shown.

state, thus resulting in a continuous attractor. But, as we can see, the weighted sum of sigmoidal functions is just an approximation to a continuous attractor, showing oscillations around the identity line, and only a finite range beyond which the weighted sum diverges from the identity. The points of intersection correspond to the stationary solutions: we therefore get a series of fixed points in line.

3 Stability of network states

We computed the bifurcation diagram of the m equation in relation to the parameters a_i , b_i and h_i for different network sizes n (see a_i bifurcation diagrams Fig. 3). Fig. 3 indicates that if we vary every a_i around 1, even mildly, the system will lose most of its stable fixed points, demonstrating that the system lacks robustness to such *global* perturbations. The perturbation is global because all the neurons will see a change in their corresponding synaptic weights, and in the same direction.

As a consequence, larger networks have smaller intervals in the a_i parameter range (range of stability) for which they show the maximum number of stable points and also an average smaller range of stability for a fixed number of stable points, as verified in Fig. 3. At the same time, larger networks are more robust against *local* perturbations (e.g. knockout of a single neuron; simulations not shown).

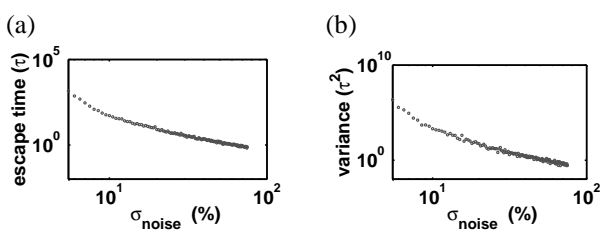


Figure 4. (a) Lifetime of one stable state for different levels of noise. The levels of noise are in units of percentage of the smallest radius of the stable state domain. Mean time of escape from a stable state in τ units. (b) Variance of escape times in τ^2 units.

4 Lifetime of network states

In networks with noisy units, the stable states derived before are no longer stable in the strict term. In fact, due to noise, a system in a stable state has a probability to jump out of it, and go to another one. To address that probability, we looked to the mean lifetime during which the system stays in that state, by performing several numerical simulations.

A network of 10 neurons was simulated with the already presented Seung model [5], and several noise levels were used. All simulations were started with the system in one specific stable state. An escape was counted whenever the system went out of the domain of the fixed point. We waited for 100 escapes for each noise level. The results are shown in Fig. 4.

As we would expect, the smaller the noise level, the bigger the mean lifetime in the state. At very low levels of noise, the mean lifetimes of escape were so big that the simulations were stopped. The variance of the escape time has the same shape as the mean lifetime curve: for low levels of noise, the variance is very big.

These results establish clear predictions related to the lifetime of stable states that could be tested by psychophysics experiments.

5 Two network system

5.1 Persistent activity in mutual inhibitory populations

In neural short-term memory systems, one sometimes finds two functionally distinct and mutual inhibitory populations of neurons that represent the same memory variable [6]. A prime example is the oculomotor system of the goldfish [4]. Each population responds to more than half the memory range: the activity of one population of neurons increases as the second decreases, while the goldfish spans the visual range with eye fixations [4] (see Fig. 5).

Here, we build a model with mutual inhibition and investigate its stability against perturbations.

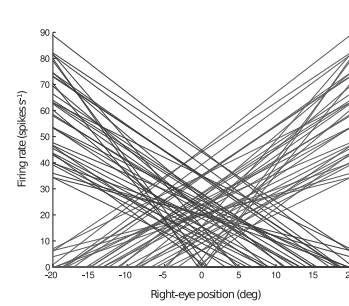


Figure 5. From [4]. Model tuning curves defined by experimentally measured rate vs. position relationships. Positive slopes are respective to right population neurons and negative slopes to left.

5.2 The model

Let's consider two symmetrical and independent networks, built following the procedure presented before

$$\tau \dot{m}_1(t) = -m_1(t) + \sum_{i=1}^n b_i g(a_i m_1(t) + h_i) \quad (6)$$

$$\tau \dot{m}_2(t) = -m_2(t) + \sum_{i=n+1}^{2n} b_i g(a_i m_2(t) + h_i) \quad (7)$$

We can couple these two networks using mutual inhibition so that

$$\tau \dot{m}_1(t) = -m_1(t) + \sum_{i=1}^n b_i g(a_i m_1(t) + c_i m_2(t) + h_i) \quad (8)$$

$$\tau \dot{m}_2(t) = -m_2(t) + \sum_{i=n+1}^{2n} b_i g(a_i m_2(t) + c_i m_1(t) + h_i) \quad (9)$$

The general connectivity matrix ($2n \times 2n$) is assumed to be a rank-2 matrix,

$$T_{ij} = \begin{bmatrix} a_i b_j & c_i b_j \\ c_i b_j & a_i b_j \end{bmatrix}. \quad (10)$$

The stationary solutions are

$$m_{1s} = \sum_{i=1}^n b_i g(a_i m_{1s} + c_i m_{2s} + h_i) \quad (11)$$

$$m_{2s} = \sum_{i=n+1}^{2n} b_i g(a_i m_{2s} + c_i m_{1s} + h_i) \quad (12)$$

Therefore, we get two nullclines, each corresponding to the lines of perfect match between a sum of sigmoidal functions of two variables and a plane. Accordingly, the stationary states correspond to the intersection between the

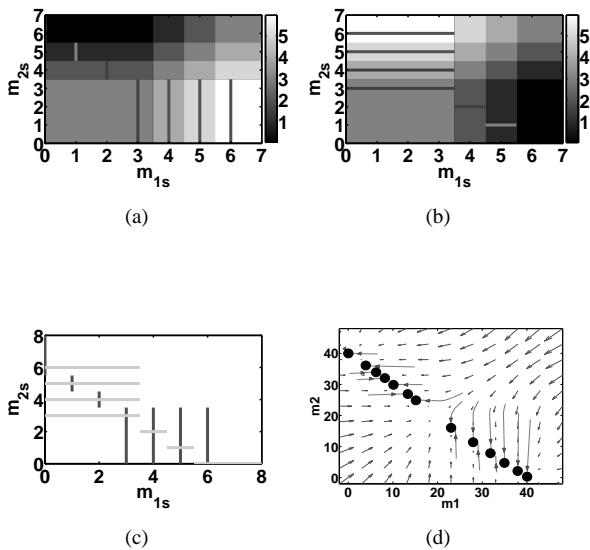


Figure 6. Determination of the stationary solutions. (a) Plot of equation (11). The colored areas are the different values of the input-output functions’s sum versus m_{1s} and m_{2s} . The gray thick lines correspond to the values where m_{1s} is equal to the sum, i.e., one of the two nullclines. (b) Plot of equation (12). (c) Intersection of the nullclines at the fixed points. (d) Vector field of m_1 vs m_2 , for a system of two networks of 40 neurons each.

nullclines. As in the one-dimensional case, we can only approximate a line attractor through a series of fixed points.

Several approximations are possible to get a whole range of fixed points along a line of negative proportionality between m_1 and m_2 as we expect experimentally. For example, one can tune the model similarly to the one dimensional case, with the addition of inhibitory inputs: for each of the two networks, choose $a_i = 0$, $c_i = -1$, $b_i = 1$ for half of it, and $a_i = 1$, $c_i = 0$, $b_i = 1$ for the other half, and span the m_1/m_2 range by choosing different h_i for each neuron. Next, we will look into that possibility.

In Fig. 6a, b, c, we use a system of two networks of 6 neurons each, and show the sum of the input-output functions of the neurons, as well as the nullclines and stationary states. In Fig. 6d, the resultant vector field for a system of two networks of 40 neurons each, showing several stable points (black circles) in line.

5.3 Global perturbations in the 2 network system

We computed the bifurcation diagram of the m_1 and m_2 equations in relation to the parameters a_i , c_i , b_i , and h_i for a system of two networks of 40 neurons each (see a_i bifurcation diagram Fig. 7). Fig. 7 shows the fixed points $m_{1s} - m_{2s}$ as a function of the perturbation added to ev-

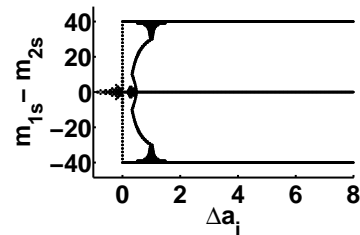


Figure 7. Bifurcation diagram of the the m_1 and m_2 equations in relation to the a_i parameter for a system of two networks of 40 neurons each. Δa_i corresponds to the value of the perturbation added to every a_i . In black, the stable fixed points. The unstable fixed points are not shown.

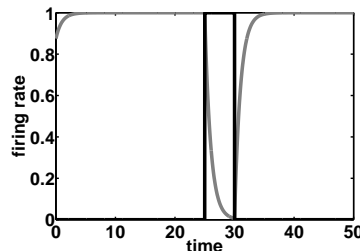


Figure 8. Hyperpolarization of one neuron. In gray, neuron firing rate; in black, hyperpolarization current in adimensional units.

ery a_i . The figure indicates that if we vary every a_i around its original value, even mildly, the system will lose most of its stable fixed points, demonstrating that the system lacks robustness to such *global* perturbations, as in the one dimensional case.

5.4 Local perturbations in the 2 network system

In [7], the authors hyperpolarized one neuron from the oculomotor system of the goldfish for periods of 100ms. The neuron stopped its activity for the period of hyperpolarization but regained its previous activity after it. Our model behaves in accordance with experiments (see Fig. 8).

6 Conclusion

In this study, we performed a first step towards investigating the stability of a neural network that features a set of point attractors, ordered on a line, a line attractor.

We found configurations that capture some of the biological features observed in neural short-term memory systems. In these configurations, the system is robust to particular perturbations.

We furthermore quantified the stability of network states against noise and showed the dependency of the life-

time of short-term memory states on the level of neural noise.

However, one hasn't yet been able to construct realistic line attractor models robust to perturbations, without resorting to additional assumptions about single neurons (the study [8] solves the robustness problem by assuming that neurons are hysteretic, an assumption that has yet to be proven true).

Additionally, in the oculomotor system of the goldfish, mechanisms of retuning of the connectivity matrix to compensate perturbations are known and generally act on longer time scales than the robustness to hyperpolarization perturbations [9] [10], and should be integrated in future work.

This line of research could help us to enlighten architectural principles and parameters's ranges underlying the biological neural networks involved in short-term memory.

The theoretical approach performed here aims at obtaining a deeper view on the functioning of short-term memory systems.

7 Acknowledgements

Pedro Gonçalves is from the PhD Program in Computational Biology of Instituto Gulbenkian de Ciência, supported by Fundação Calouste Gulbenkian, Siemens SA and Fundação para a Ciência e Tecnologia through grant SFRH/BD/33210/2007.

References

- [1] P. S. Goldman-Rakic. Cellular basis of working memory. *Neuron*, 14(3):477–485, Mar 1995.
- [2] Guy Major and David Tank. Persistent neural activity: prevalence and mechanisms. *Curr Opin Neurobiol*, 14(6):675–684, Dec 2004.
- [3] Carlos D Brody, Ranulfo Romo, and Adam Kepecs. Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. *Curr Opin Neurobiol*, 13(2):204–211, Apr 2003.
- [4] Emre Aksay, Itsaso Olasagasti, Brett D Mensh, Robert Baker, Mark S Goldman, and David W Tank. Functional dissection of circuitry in a neural integrator. *Nat Neurosci*, 10(4):494–504, Apr 2007.
- [5] H. S. Seung. How the brain keeps the eyes still. *Proc Natl Acad Sci U S A*, 93(23):13339–13344, Nov 1996.
- [6] Christian K Machens and Carlos D Brody. Design of continuous attractor networks with monotonic tuning using a symmetry principle. *Neural Comput*, 20(2):452–485, Feb 2008.
- [7] E. Aksay, G. Gamkrelidze, H. S. Seung, R. Baker, and D. W. Tank. In vivo intracellular recording and perturbation of persistent activity in a neural integrator. *Nat Neurosci*, 4(2):184–193, Feb 2001.
- [8] Alexei A Koulakov, Sridhar Raghavachari, Adam Kepecs, and John E Lisman. Model for a robust neural integrator. *Nat Neurosci*, 5(8):775–782, Aug 2002.
- [9] Guy Major, Robert Baker, Emre Aksay, Brett Mensh, H. Sebastian Seung, and David W Tank. Plasticity and tuning by visual feedback of the stability of a neural integrator. *Proc Natl Acad Sci U S A*, 101(20):7739–7744, May 2004.
- [10] Guy Major, Robert Baker, Emre Aksay, H. Sebastian Seung, and David W Tank. Plasticity and tuning of the time course of analog persistent firing in a neural integrator. *Proc Natl Acad Sci U S A*, 101(20):7745–7750, May 2004.