Accounting for the dynamics of working memory: Bumps move to decide

Adam Kepecs
Volen Center for Complex Systems
Brandeis University
Waltham MA 02454
kepecs@brandeis.edu

February 24, 2003

Abstract

Recordings from animals during working memory tasks have revealed that single neurons can fire *persistently* without external cues and their firing rate can be persistent *at graded levels* to encode stimuli. Working memory is thought to be mediated by such persistent activity, which is a flexible means of storing multiple values "on-line". Many theoretical studies have focused on the ability of recurrent neural networks to sustain graded levels of activity. Here we account for the dynamics of firing rates during a parametric working memory task reported by Romo and colleagues (1999). We show that an intensity to space conversion can enable a bump attractor network to encode a scalar parameter. Asymmetries in the recurrent connections or in the external inputs cause systematic drift in the bump state, which can account for the behavior of neurons that only encode stimuli during either the beginning or the end of the delay period. We suggest that this network architecture enables information transfer between cortical columns during working memory tasks which could underlie decision processes.

Extended abstract

Working memory refers to the ability of an organism to temporarily hold and manipulate information during cognitive tasks [3]. The biological basis of working memory is beginning to be established with unit recordings from prefrontal cortex during delayed response tasks [9, 8]. For instance, neurons in the prefrontal cortex have been observed to elevate their firing rates during delayed response tasks. Such *persistent activity* is thought to underlie the "on-line" storage of information. Theoretical studies have established that recurrent networks can sustain elevated levels of activity through reverberatory excitation [10]. The standard model that can account for these findings is a "ring" of connected neurons with a *bump attractor* state [18, 2, 4, 5, 7]. However, it appears that multiple levels of activity can be sustained in certain brain regions, such as those responsible for holding eye-position [14, 1] or even by prefrontal neurons during some tasks [15]. Two classes of models have been suggested to account for *graded persistent activity* in integrator networks: line-attractors [6, 17] and bistable, staircase networks [12].

Recent work by Romo and colleagues (1999) has shown, however, that during a parametric working memory task neurons in the prefrontal cortex do not simply exhibit *graded persistent activity* like in integrator networks, but instead dynamically vary their encoding strategy [15]. Monkeys were trained in a flutter discrimination task to discriminate between two flutter stimuli of different frequencies presented sequentially on a fingertip. To perform this task, monkeys had to remember the base frequency of the first flutter for a delay period of 3 seconds until the second flutter was presented. Some neurons in the prefrontal cortex were found to fire persistently at different frequencies corresponding to the base flutter frequency. The dynamics of the neurons fell into three broad categories: neurons that fired early but not late, late but not early or persistently throughout the delay period. Furthermore, each class could be further subdivided into neurons whose firing rate was increasing and neurons whose firing rate for decreasing with increasing flutter frequency [15].

Here we show how a bump attractor network augmented by an appropriate stimulus conversion network can naturally account for all the different types of firing behaviors. Our network architecture leads us to suggest that rich dynamics of graded activity in this task may be used to transfer information between different cortical columns and to enable decision processes.

References

[1] E Aksay, R Baker, HS Seung, and DW Tank. Anatomy and discharge properties of pre-motor neurons in the goldfish medulla that have eye-position signals during fixations. *J Neurophysiol*, 84:1035–49, 2000.

[2] S. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.*, 27:77–87, 1977.

[3] A Baddeley. *Working Memory*. Oxford Unversity Press, London, 1986.

[4] R Ben-Yishai, Lev Bar-Or, R, and H Sompolinsky. Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci. USA*, 92:3844–3848, 1995.

[5] M. Camperi and X.-J. Wang. A model of visuospatial short-term memory in prefrontal cortex: recurrent network and cellular bistability. *J. Comput. Neurosci.*, 5:383–405, 1998.

[6] SC Cannon, DA Robinson, and S Shamma. A proposed neural network for the integrator of the oculomotor system. *Biol Cybern*, 49:127–13, 1983.

[7] A Compte, N Brunel, PS Goldman-Rakic, and Wang X-J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb Cortex*, 10:910–23, 2000.

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

[9] JM Fuster and GE Alexander. Neuron activity related to short-term memory. *Science*, 173:652–4, 1971.

[10] D. O. Hebb. *Organization of behavior*. New York: Wiley, 1949.

[11] JJ Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, 79:2554–8, 1982.

[12] AA Koulakov, S Raghavachari, A Kepecs, and J Lisman. Model for a robust neural integrator. *Nature Neuroscience*, 5:775–782, 2002.

[13] K Kubota and H Niki. Prefrontal cortical unit activity and delayed alternation performance in monkeys. *J Neurophysiol*, 34:337–47, 1971.

[14] DA Robinson. Integrating with neurons. *Annu Rev Neurosci*, 12:33–45, 1989.

[15] R Romo, CD Brody, A Hernandez, and Lemus L. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, 399:470–3, 1999.

[16] HS Seung. How the brain keeps the eyes still. *Proc Natl Acad Sci USA*, 93:13339–13344, 1996.

[17] HS Seung, DD Lee, BY Reis, and DW Tank. Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron*, 26:259–271, 2000.

[18] H. R. Wilson and J. D. Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13:55–80, 1973.

[19] K Zhang. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci*, 16:2112–26, 1996.

[20] D. Zipser, B. Kehoe, G. Littlewort, and J. Fuster. A spiking network model of short-term active memory. *J. Neurosci.*, 13:3406–3420, 1993.