Accepted Manuscript

A self-organizing short-term dynamical memory network

Callie Federer, Joel Zylberberg

PII: S0893-6080(18)30190-4
DOI: https://doi.org/10.1016/j.neunet.2018.06.008
Reference: NN 3974

To appear in: *Neural Networks*

Received date : 5 January 2018
Revised date : 31 May 2018
Accepted date : 13 June 2018

Please cite this article as: Federer, C., Zylberberg, J., A self-organizing short-term dynamical memory network. *Neural Networks* (2018), https://doi.org/10.1016/j.neunet.2018.06.008

A SELF-ORGANIZING SHORT-TERM DYNAMICAL MEMORY NETWORK

Callie Federer[1] **& Joel Zylberberg**[1,2]
[1]Department of Physiology and Biophysics, University of Colorado Anschutz Medical Campus
[2]Learning in Machines and Brains Program, Canadian Institute For Advanced Research
{callie.federer,joel.zylberberg}@ucdenver.edu

ABSTRACT

Working memory requires information about external stimuli to be represented in the brain even after those stimuli go away. This information is encoded in the activities of neurons, and neural activities change over timescales of tens of milliseconds. Information in working memory, however, is retained for tens of *seconds*, suggesting the question of how time-varying neural activities maintain stable representations. Prior work shows that, if the neural dynamics are in the 'null space' of the representation - so that changes to neural activity do not affect the downstream read-out of stimulus information - then information can be retained for periods much longer than the time-scale of individual-neuronal activities. The prior work, however, requires precisely constructed synaptic connectivity matrices, without explaining how this would arise in a biological neural network. To identify mechanisms through which biological networks can self-organize to learn memory function, we derived biologically plausible synaptic plasticity rules that dynamically modify the connectivity matrix to enable information storing. Networks implementing this plasticity rule can successfully learn to form memory representations even if only 10% of the synapses are plastic, they are robust to synaptic noise, and they can represent information about multiple stimuli.

1 INTRODUCTION

Working memory is a key cognitive function, and it relies on us retaining representations of external stimuli even after they go away. Stimulus-specific elevated firing rates have been observed in the prefrontal cortex during the delay period of working memory tasks, and are the main neural correlates of working memory (Funahashi et al., 1993; Fuster & Alexander, 1971). Perturbations to the delay period neural activities cause changes in the animal's subsequent report of the remembered stimulus representation (Li et al., 2016; Wimmer et al., 2014). These elevated delay-period firing rates are not static but have time-varying dynamics with activities changing over timescales of tens of *milliseconds* (Brody et al., 2003a; Romo et al., 1999), yet information can be retained for tens of *seconds* (Fig. 1A). This suggests the question of how time-varying neural activities keep representing the same information.

Prior work from Druckmann & Chklovskii (2012) shows that, if the neural dynamics are in the "null space" of the representation - so that changes to neural activity do not affect the downstream read-out of stimulus information - then information can be retained for periods much longer than the time-scale of individual neuronal activities (called the FEVER model; Fig. 1B). That model has a severe fine-tuning problem, discussed below. We identified a synaptic plasticity mechanism that overcomes this fine-tuning problem, enabling neural networks to learn to form stable representations.

While the dynamics of neurons in the FEVER model match that which is observed in the monkey prefrontal cortex during a working memory task (Murray et al., 2017), the model itself requires that the network connectivity matrix have one or more eigenvalues very near to unity. According to the Gershgorin Circle Theorem, this will almost surely not happen in randomly-connected networks: fine-tuned connectivity is needed (Zylberberg & Strowbridge, 2017). Druckmann & Chklovskii (2012) suggest a mechanism of Hebbian learning by which this connectivity can be learned. That mechanism requires the read-out weights to form a 'tight frame', which will not necessarily be true in biological circuits. Thus, the prior work leaves it unknown how synaptic plasticity can form and/or maintain functional working memory networks. Here, we identify biologically plausible synaptic plasticity rules that can solve this fine-tuning problem without making strong assumptions like 'tight frame' representations.