LETTER

Why is our capacity of working memory so large?

Thomas P. Trappenberg
Faculty of Computer Science, Dalhousie University
6050 University Avenue, Halifax, Nova Scotia B3H 1W5, Canada
E-mail: tt@cs.dal.ca

(Submitted on September 12, 2003)

Abstract— A neuropysiological correlate of short-term memory has been found in the ability of some prefrontal neurons to hold task relevant information through ongoing neuronal activity. Such experimental findings, and the corresponding hypothesis of the mechanisms that enables the ongoing neuronal activity, have been captured successfully with continuous attractor neural network models. However, the basic model does only support the short-term storage of one item at a time. In this paper it is shown that more items can be stored in the model if the basic model is augmented with physiological plausible stabilization mechanisms. This modified model supports the storage of a few items consistent with the behavioral capacity limits of working memory.

Keywords—Working Memory, Capacity, Continuous Attractor Neural Networks

1. Introduction

The concept of working memory has emerged in cognitive psychology to describe the specific memory requirements that are used in everyday mental tasks such as reading, planning actions, comprehending complex relations, and facilitating learning [1,2]. Thus, working memory is a meta-concept of a memory system that cuts across other traditional memory concepts such as short-term and long-term memory, or declarative and implicit memory. The limited storage capacity of working memory was first popularized by Miller [3] who suggested a capacity limit of around 7±2 items. However, as there is no precise agreement on the definition of working memory, and indeed there is the possibility that different working memory systems exist [1], it is not surprising that different numbers have been implicated with the capacity limit of working memory, such as the number four [4].

Working memory requires the ability to hold information, or references to information in long-term memory, over a short period of time. Hence, a capacity limit of short term memory would support the limited capacity of working memory. The neurophysiological correlates and mechanisms of short-term memory are therefore important factors in understanding central cognitive tasks. It is technically easy to store large amounts of digital information in small memory devices, and it has puzzled researchers for a long time why the ability of humans to hold items in memory over a short period of time is limited to an astonishing small number. The capacity limit is typically subject and task dependent, and it is the small magnitude of this number, rather than its absolute value, that is of continuous surprise to researchers.

Several suggestions for the reason behind the capacity limit have been made over the years, including limited channel capacities [5], limited attentional resources [6], spurious synchronization [7], or the use of subcycles in brain oscillations [8]. However, in light of the recent understanding of the physiological mechanisms underlying short-term memory [9] it is more puzzling how more than one item can be stored in ongoing neural activity at a time. Here I show how continuous attractor neural network (CANN) models, which capture our understanding of the physiological mechanisms underlying short-term memory, can be augmented with biologically plausible mechanisms closely related to properties of NMDA-mediated ion channels to yield capacity limits consistent with psychophysical findings. Thus, the possible explanation outlined here is based on converging evidence of the physiological realization of short-term memory in the brain, and several suggestions are made how such a hypothesis can be further verified experimentally.

2. Continuous Attractor Neural Network Models of Short-Term Memory

Funahashi, Bruce, and Goldman-Rakic [10] were among the first to demonstrate continuous activity of prefrontal neurons in the delay period of a memory guided saccade task, which they suggested is a physiological reflection of short-term memory. The mechanism that enables such neural activity was recently captured by a model based on continuous attractor neural networks (CANNs) [9]. Such networks, illustrated in Figure 1, are abstractions of neuronal networks in which an ongoing node activity is enabled through feedback connections from the neural layer to itself. Such recurrent networks are commonly used as models for associative memories, where the memory states correspond to point attractors of the equivalent dynamical system imprinted through Hebbian learning. CANN models are a specific subclass of such models with continuous manifolds of point attractors [11]. They are marked by a specific interaction structure in the network, in which the effective strengths of the recurrent connections are dependent solely on the distance between nodes, and are effectively positive (excitatory) for nearby nodes, and negative (inhibitory) for distant nodes. Such effective network structures have long been proposed to be crucial for information processing in the brain [12,13].

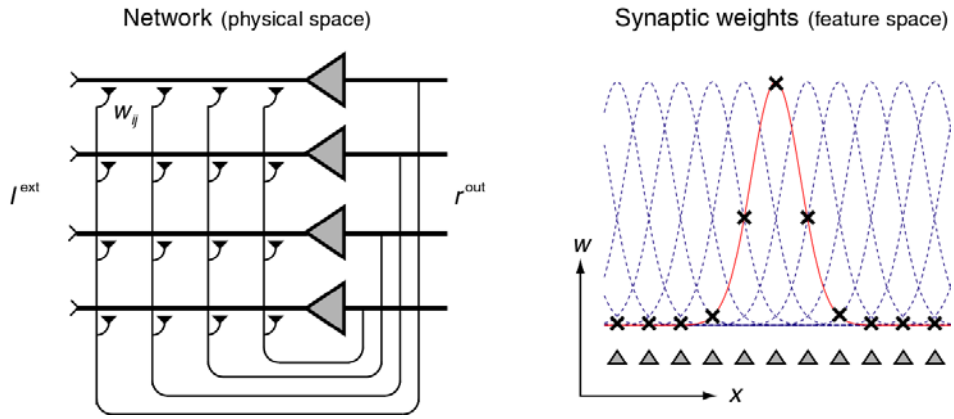

Figure 1. Recurrent network model used in this study. The weights w_{ij} depend only on the distance between nodes, with positive values (excitatory) for short distances and negative values (inhibitory) for long distances.

The nodes in the networks used to demonstrate the hypothesis discussed here represent population averages of spike counts of neurons with similar response properties. Such models are an appropriate representation of average integrate-and-fire neuron activities in the adiabatic limit [13, 14], and such a limit is appropriate for the exploration of attractor states. The activity u of the nodes is thereby governed by the dynamic equation

$$\tau du_i/dt = -u_i + \sum_j w_{ij}\, g(u_i) + I^{ext}_i, \qquad (1)$$

where u_i is the activity of the node, I^{ext} is the initial external input given to the network, and τ represents the time scale of the dynamics. The matrix element w_{ij} represents the connectivity weight between node i and j, which is given by training the network on all possible Gaussian patterns $I^{ext}_i = \exp(-(i-i_0)^2/2\,\sigma^2)$ with the Hebbian rule

$$w_{ij} = 1/(\sqrt{\pi}\,\sigma) \sum_{i0}\; I^{ext}_i\, I^{ext}_j. \qquad (2)$$

The width σ was taken to be $2\pi/80$ in the experiments illustrated in this paper.

A major feature of CANN models is illustrated in the leftmost panel of Figure 2 which shows the activity of nodes in the network. The network was stimulated by an external stimulus centred around the 5th node in this network of 100 nodes, and network activity in a form of an activity packet, which is related to the original input stimulus, was maintained after the external stimulus was removed at time $t=10\tau$. The important question asked here is how many such activity packets can be maintained at the same time. The answer is that, in the basic model with global inhibition, only one activity packet can be active asymptotically. Indeed, such networks implement a winner-take-all process by suppressing asymptotically all but one activity packet. There are many indications that such conflict resolution mechanisms are used in the brain, for example in the integration of endogenous and exogenous information to direct saccadic eye movements [15], or the decoding of population

codes [16]. However, the utilization of CANN network mechanisms for the storage of short-term memories in this basic form does only support one memory item to be stored at any given point in time.

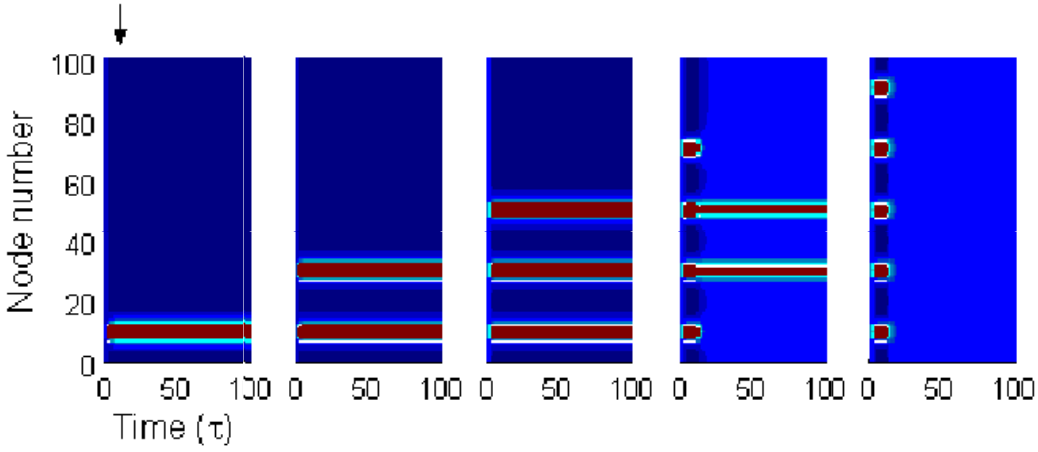

Figure 2: Experiment of the network response with different numbers of initially excited sites in the network. The external stimulus is withdrawn after the time t=10τ, which is indicated with an arrow at the leftmost sub-graph.

3. The enhanced model with stabilization

The previously proposed model of physiological short-term memory [9] is consistent with a too severe limit of short-term memory, and the question becomes how can this number be increased in such systems to reflect the behavioural findings. A possible solution to this problem is based on a mechanism that was proposed to solve another problem in the biologically realistic CANN models, that of the drift of activity packets due to noise in the system [17]. The solution proposed by Stringer et al. [18] is based on the increased sensitivity of a node to become active following a previous active period, which can be modelled in CANNs by an decrease of the firing threshold of the nodes that were active in a proceeding time window. Such a mechanism can be enabled by NMDA-mediated ion channels in neurons [19], in which the Mg^{2+} blockage of the ion channel is removed only after a neuron become active. While Lisman et al. [19] proposed the NMDA-style stabilization to be the sole mechanism of short-term memory, the hypothesis here is that a balance between competitive long-range inhibition and the cooperative short-range excitation, which is enhanced by NMDA mechanisms, is realized in the brain, and that this system is the root of the psychophysically observed working memory capacity.

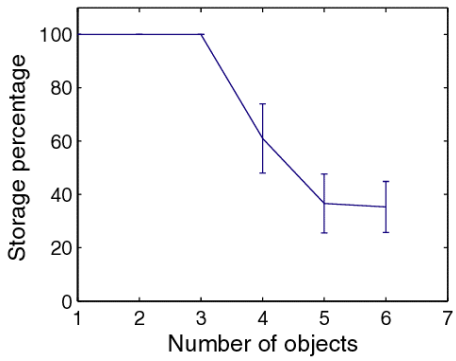

Figure 3: Average percentage of activity package and standard deviation for different numbers of non-overlapping objects at random locations. This network has a load capacity of 3 objects.

Stringer et al. [20] have recently demonstrated that the combination of the NMDA-style stabilization mechanism with CANN models enables the concurrent stabilization of two activity packets. In Figure 2 it is demonstrated that even three activity packets can be stable in such networks with the given choice of parameters (see methods). However, with the chosen parameters it is not possible to stabilize more than 3 activity packets in the experiment shown in Figure 2, where the initial external stimuli were placed in a systematic equidistant manner. The results of an additional experiment, in which a varying number of objects are placed at random

locations, excluding possible overlaps, is shown in Figure 3. This graph shows the average percentage and standard deviation of asymptotically stable states in 100 trials for different numbers of randomly placed objects. The sharp decline in the number of concurrent activity packets is consistent with psychophysical evidence of the limited capacity of working memory [21].

These experiments suggest that the capacity limit of short-term memory is based on a limited excitation level that can be realized in competitive neural systems. Indeed, a short global excitation of the model layer was used by Comte et al. [9] to extinguish the firing of prefrontal neurons after the initiation of a saccade. It is demonstrated here that this feature is present in rate models, and that this is not a characteristics of the implementation with spiking neurons.

4. Discussion

It is possible to achieve different numbers of concurrently stabilizable activity packets by varying physiological parameters such as the strength of the NMDA effect and the width of the interaction structure. However, realistic physiological parameters lead typically to a small number of concurrent activity packets consistent with the capacity limit of working memory in the literature. A crucial physiological experiment to verify the hypothesis outlined here is to verify the existence of concurrent activity packets in neurons associated with short-term memory, such as prefrontal or parietal neurons. This should be achievable with cell recordings using an experimental paradigm similar to that used by Funahashi et al. [10], in which the memory-guided saccade has to be initiated to one of several memorised locations, where the specific final target location in each trial is determined by the go cue. Hence, the subject has to hold location specific information for several locations in short-term memory until the go cue is presented, for which corresponding ongoing neural activity in different neurons should be detectable. Additional experiments should further attempt to study the effects of blocking NMDA receptors to verify more directly the involvement of such channels in short-term memory.

Many further predictions can be derived from the hypothesis presented here such as the variation of the number of possible concurrent activity packets with the width of the receptive fields of the neurons. The hypothesis expressed here in the form of a concrete model can guide further studies of the factors that influence working memory limits and the consequences such limits bear. This study not only sheds light on the possible information processing in the brain, but the more precise understanding of the working-memory capacity has also several important technical implications. The concept of a capacity limit is already regarded as an important guiding principle for the design of human-machine interaction systems such as computer interfaces, and a more specific knowledge of the factors that drive capacity limits to their lower bounds or might increase capacity are necessary to systemize and optimise such efforts and designs.

References

[1] Cowan, N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* **24**, 87-185 (2001).
[2] Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* **63**, 81-97 (1956).
[3] Broadbent, D.E. *Perception and Communication* (Pergamon Press, London, 1958). [4] Cowan, N. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin* **104**, 163-191 (1988).
[5] Luck, S. J. & Vogel, E. K. Response from Luck and Vogel (Response to Commentary by Nelson Cowan). Trends in Cognitive Sciences **2**, 78-80 (1998).
[6] Lisman, J.E. & Idiart, M.A.P. Storage of 7±2 short-term memories in oscillatory subcycles. *Science* **267**, 1512-1515 (1995).
[9] Compte, A., Brunel, N., Goldman-Rakic, P.S. & Wang, XJ. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* **10**, 910-23 (2000).
[8] Baddeley, A.D. *Working Memory* (Clarendon Press, Oxford 1986).
[9] Miyake A. & Shah, P., (eds.) *Models of Working Memory* (Cambridge University Press, Cambridge 1999).
[10] Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol* **61**, 331-349 (1989).
[11] Seung, H. S. Learning continuous attractors in recurrent networks. *Advances in Neural Information*

Processing Systems **10**, 654-660 (1998).
[12] Grossberg, S. Contour enhancement, short-term memory, and constancies in reverberating neural networks, *Studies in Applied Mathematics* **52**, 217-257 (1973).
[13] Wilson, H. R. & Cowan, J. D. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue, *Kybernetik* **13**, 55-80 (1973).
[14] Gerstner, W. Population Dynamics of Spiking Neurons: Fast Transients, Asynchronous States, and Locking. *Neural Computation* **12**, 43-89 (2000).
[15] Trappenberg, T.P, Dorris, M., Klein, R.M. & Munoz, D.P. A Model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus, *Journal of Cognitive Neuroscience* **13**, 256-271 (2001).
[16] Deneve, S., Latham, P.E. & Pouget, A. Reading population codes: a neural implementation of ideal observers, *Nature Neuroscience* **2**, 740-745 (1999)
[17] Zhang, K. Representation of spatial orientation by the intrinsic dynamics of head-direction cell ensembles: a theory. *Journal of Neuroscience* **16**, 2112—2126 (1996)
[18] Stringer, S.M., Trappenberg, T.P., Rolls, E.T. & Araujo, I.E.T. Self-organising continuous attractor networks and path integration: One-dimensional models of head direction cells. *Network: Computation in Neural Systems* **13**, 217-242 (2002).
[19] Lisman, J.E., Fellous, J.M. & Wang, X.J. A role for NMDA-receptor channels in working memory. *Nature Neuroscience.* **1**, 273-5 (1998).
[20] Stringer, S.M., Rolls, E.T., Trappenberg, T.P. & de Araujo, I.E.T. Self-organizing continuous attractor networks and motor function. *Neural Networks* **16**, 161-182 (2003).
[21] Luck, S.J. & Vogel, E.K. The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279-281 (1997).

Thomas P. Trappenberg is an associate professor of computer science at Dalhousie University where he also lectures on computational neuroscience. He is the author of the textbook Fundamentals in Computational Neuroscience published by Oxford University Press, 2002. His research interest includes information processing principles of the brain, neural dynamics, memory, and applications of machine learning to data classification and feature selection. (Home page: http://www.cs.dal.ca/~tt)