A computational model of working memory using bistability

Marjan Rashidi
Department of Cognitive Sciences
University of California, Irvine
Irvine, California, 92617

***Abstract*— Working memory is significantly involved in several cognitive abilities such as learning and understanding. However, its neural mechanism is still a hot topic and not fully understood. According to some common views, persistent activity is the neural mechanism underlying working memory. This study simulates the working memory activity during the delay period, i.e., persistent activity, in a task that relies on direction selectivity in the visual cortex. Bars with 4 different orientations (0, 90, 180, and 270 degrees) were used as inputs of a Recurrent Neural Network (RNN) at different time steps and the firing rate activity of the prefrontal cortex (PFC) neurons were measured shortly after the stimulation and during the delay period when the stimuli were not presented. The results revealed the bistable activity of the PFC neurons before the stimulation and shortly after the stimulus termination.**

Keywords—working memory, bistability, computational model, recurrent neural network

I. INTRODUCTION

Several studies have focused on memory over the years but still there is no general consensus on a unified categorization of memory [1], [2]. However, short-term, long-term, and working memories are the known categories used to understand and decode the abstract concept of memory. Working memory is the ability that allows us to maintain and manipulate the information for a short time and is used in the execution of goal-directed cognitive tasks [3]. In 1971, two researchers named Fuster and Alexander, worked on neural activity related to short-term memory using a delayed-response task. The results of that study revealed that monkeys showed persistent firing rate activity in their prefrontal cortex during the delay period when no stimulus was presented. This finding suggests that the excitatory reactions of neurons during the delay period occurred because the animal was focusing on the information that had been placed in short-term memory to utilize that information later in the task [4]. This persistent activity during the delay period is most likely related to the memory content because there is no stimulus presented during this period. Other studies revealed that if they prevent this spontaneous activity during the delay period, the performance on the recall tasks will be adversely affected [5]. Many other studies found that cells that show elevated firing rate activity during the delay period when the information is retained are mainly located in the prefrontal cortex of the brain [6], [7], [8].

In working memory tasks, firing rate activities of PFC neurons elevates significantly during the presence of the stimuli [9], [10]. However, during the delay period, after the stimulus termination, and before the response phase, the firing rate activity is still higher than the baseline activity [11], [12]. During this period, the subject needs to hold and retain the relevant information in the working memory and use that information later to produce an appropriate response. Some researchers used the oculomotor delayed-response paradigm to measure the activity of PFC neurons during the delay period in monkeys [13]. In this task, the monkey needs to fixate a spot of light on a TV monitor during the visual stimulation followed by a delay period. At the end of the delay period, the fixation spot is tuned off and the monkey should direct his eye to point to the spot where the previous target was located. Some neurons in prefrontal cortex show the elevated firing rate activity during the delay period when no visual stimuli is presented (Fig.1) [14].

Different theoretical models have been proposed to capture some important neural characteristics of working memory. Many of these models suggest that the recurrent network interactions induce persistent neural activity [15], [16]. These models use lateral inhibition, which can limit the number of concurrent activity patterns that might be maintained. Thus, these models have a limited capacity [17]. Some other working memory models are more flexible and focused on the relationship between working memory representations and short-term synaptic plasticity. However, these models have also a limited capacity and do not consider the neurophysiological features of working memory such as simultaneous persistent and dynamic activities observed in neural data.

Other models used to capture the neural mechanisms underlying working memory are Bump attractor models. These models are biophysically inspired network models that produce a peak of activity that represents a stimulus to be remembered. Bump attractor models simulate the generation of persistent activity and use the neurophysiological data to predict the relationship between neuronal activity, variability, and working memory performance [18]. However, it is still not clear that whether the predictions of the bump attractor models are compatible with real neuronal activity and whether their structure can capture the variability and diversity of the neuronal responses. Modeling the persistent activity and can help us understand the neural mechanisms underlying working memory which could be beneficial to several field of research such as Artificial Intelligence (AI) that could capture the real characteristics of working memory. Solving complicated AI problems could be more possible if we understand how working

memory works and how we can emulate its functionality and structure.

In this study, I used Recurrent Neural Network to capture one of the most important characteristics of the PFC neurons, i.e., bistability, in working memory. I used 4 bars with different orientations as the inputs of the RNN and simulated the one-by-one projections of V1 eye neurons to the PFC neurons during the stimulation and shortly after the stimuli termination. The results revealed the bistable behavior of the PFC neurons.

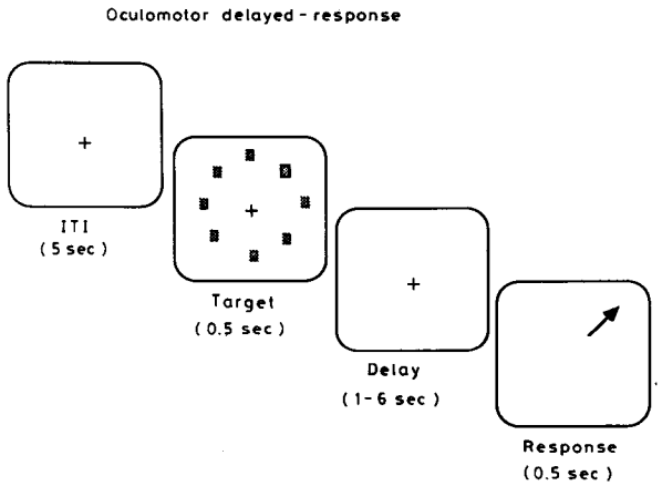

Fig. 1: Different trial phases in oculomotor delayed response paradigm. The task is presented on the TV monitor and the monkey fixates a spot of light during the visual stimulation. The target point disappear during the delay period and the fixation point disappears at the end of the delay period and the monkeys behavior is recorded during the response phase shortly after the delay period [14].

II. Methods

A. Simple connection between the V1 & PFC neurons

First, we defined a simple connection between one V1 and one PFC neuron as we are going to use a specific stimulus with a particular orientation as an input of the V1 neuron. Figure 2 shows the simplified version of this connection. As we present a stimulus with a specific direction, we expect to see the burst of firing activity, shown by "U", in this V1 neuron. The connection between V1 and PFC neuron has a particular weight, shown by "W1" in the figure. The output of the V1 neuron will be its firing rate activity "U" multiplied by its weight "W1" as we can see in the equation (1):

$$V = U.W1 + \text{PFC current rate}.\ W2 \qquad (1)$$

In this model, we also consider the recurrent synaptic feedback for the PFC neurons. This recurrent synaptic feedback shows that in the initial phase, when there is no stimulus presented, we still expect to observe some levels of initial firing rate activity for the PFC neurons. This firing rate activity could be updated using the rate neuron model equation.

For this simplified model, the input to the PFC neuron will be the synaptic feedback from the PFC neuron added to the output from the V1 neuron. Each PFC neuron will get the same input from the V1 neuron, however with different strength because the weights are different for each connection.

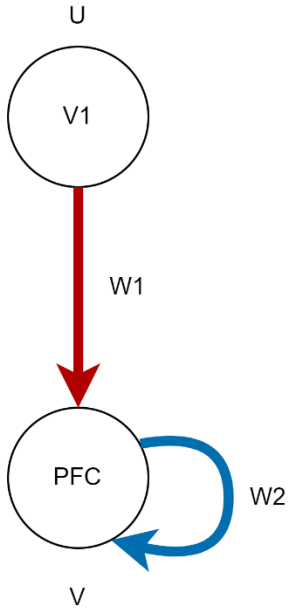

Fig. 2: The simplified connection between one V1 neuron and one PFC neuron. The connection is shown by red arrow. The recurrent synaptic feedback for the PFC neuron is shown by blue arrow. U, V, W1, and W2 are the V1 output, PFC output, the weight for the connection between V1 and PFC neuron, and the weight for the recurrent feedback of the PFC neuron, respectively.

B. General structure of the RNN

As mentioned in the previous section, we are going to use rate neurons to build the structure of the neural network. As mentioned before, the firing rate activity of the PFC neurons could be updated using the rate neuron model equation (2):

$$T_{ri}\ \frac{d_{ri}}{dt} = -ri + f_{ri}\left(\sum_{j} W_{ji}\ \ \frac{r_j}{r_j\,(max)}\right)$$

In which T_{ri} is the time constant for changes in the firing rate of unit i network of neurons. $\frac{d_{ri}}{dt}$ is the rate of change of ri which is the firing rate of unit i. f_{ri} is the function that shows the firing rate curve and the value of this function at any point in time depends on the input to unit i. W_{ji} is the connection strength between unit j and unit i. $\frac{r_j}{r_j\,(max)}$ is the firing rate of unit j divided by its maximum firing rate and is normalized between 0 and 1 and this represent maximum input it gives to other connected units.

Figure 3 shows the general structure of the neural network used in this paper. We used 4 PFC neurons and 4 V1 neurons that project to these PFC neurons through connections with different strength. In this model, all nodes are connected to each other.

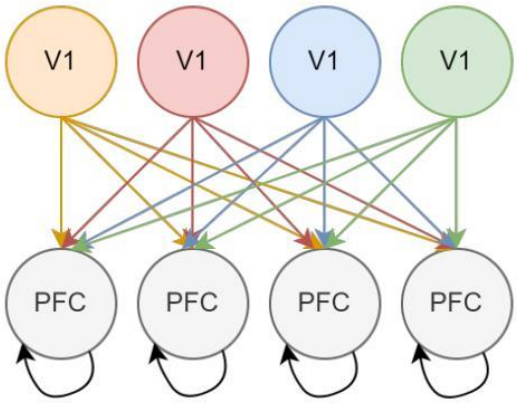

Fig. 3: The structure of the recurrent neural network in this model. Connections of each V1 neurons with PFC neurons are shown in a different color. The recurrent synaptic feedback of the PFC neurons are shown with black arrows.

Different bars with different orientations are inputs to the V1 neurons (Fig. 4). In this model, we did not use any learning algorithm. However, we used 4 bars with the orientations of 0, 90, 180, and 270 degrees to test the efficiency of the model under different conditions when different stimuli with various orientations are presented.

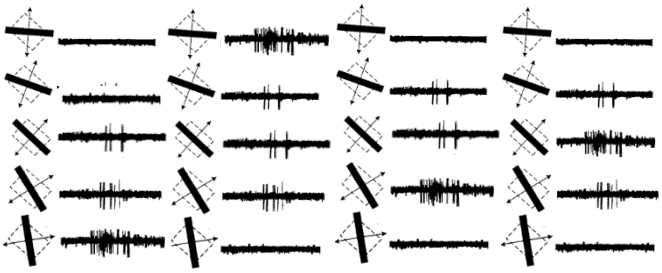

Fig. 4: schematic representations of the bars with different orientations. Each V1 neuron will respond to a different stimulus and show high firing rate activity for a specific orientation (modified from the famous work of Hubel and Weisel)

We used the Gaussian tuning curve equation to generate the tuning curves that capture the V1 firing rates at each time step and for each stimulus orientation. Here is the equation (3):

$$f(s) = r_{max}\,(\exp\,(-\frac{1}{2}(\frac{s-s_{max}}{\sigma_f}))^2 \qquad (3)$$

In this equation, s is the orientation of the stimulus in degrees and $f(s)$ is the firing rate of the neuron for each orientation. s_{max} is the angle at which a specific neuron shows the highest firing rate.

Figure 5 shows the tuning curves for 4 V1 neurons that respond to different stimulus orientations. The highest firing rates are shifted along the X axis as each neuron shows the highest firing rate activity for 0, 90, 180, and 270 degrees of orientations separately.

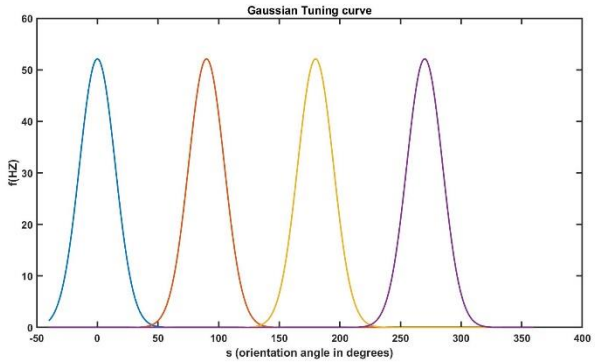

Fig. 5: Gaussian tuning curves for each V1 neuron. Each color shows different firing rate activity of a neuron. These for V1 neurons show the highest rate activity at 0, 90, 180, and 270 degrees of orientations respectively.

III. Results

We simulated the persistent activity of the PFC neurons during the delay period in 4 experiments using the Recurrent Neural Network. We used 4 V1 neurons and 4 PFC neurons that were connected through the connections with different strength and weights. In each experiment, a bar with a different orientation will be the input of the V1 neurons. Each V1 neuron projects to all PFC neurons at each experiment. The results revealed the bistable behavior of a PFC neuron at each experiment. At the initial stage, and before the stimulation, the PFC neurons' firing rate activity gradually increased and stayed stable at a specific value (Fig. 6).

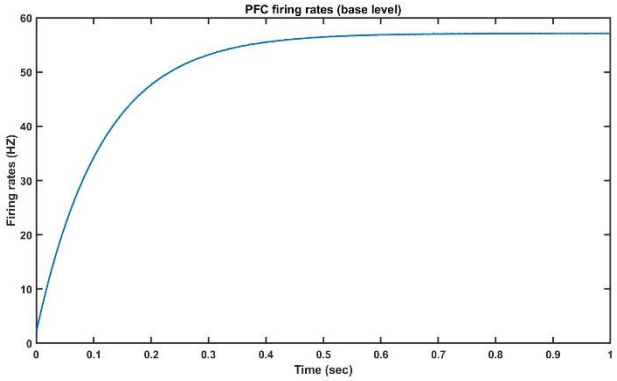

Fig. 6: The baseline firing rate activity of the PFC neurons before the stimulation. The firing rate activity of the PFC neurons start at 2 HZ and stays constant after a short period of time.

A. *Experiment 1*

As mentioned before, in each experiment, a stimulus with a specific orientation is represented to the V1 neurons. In experiment 1, neuron 1 is set to show the highest firing rate activity for the 0 degrees stimulus orientation. All other 3 V1 neurons show different firing rate activity at this orientation. Figure 7 shows the results of the experiment 1. As you can see in the figure, before the stimulation, the PFC neuron 1 shows the low firing rate activity as expected. However, during the stimulation, its firing rate activity peaks significantly, and shortly after the stimulus termination and during the delay period, it still shows elevated firing rate activity which is higher than its baseline levels. This finding emphasizes the fact that the PFC neurons hold the stimulus with the preferred orientation in working memory during the delay period.

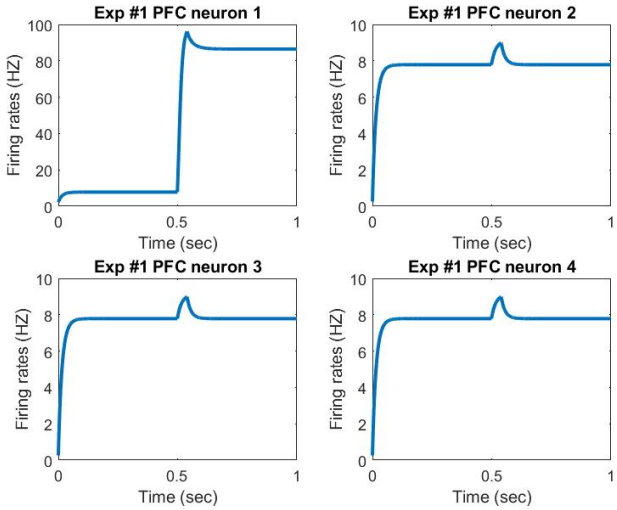

Fig. 7: The results of the simulation for experiment 1. The stimulation period is from 0.5 to 0.54 when the bar with a specific orientation is represented. PFC neuron 1 shows a bistable behavior where it starts with a low firing rate activity before the stimulation, and it peaks during the stimulation. The PFC neuron shows a high firing rate activity during the delay period when no stimulus is presented. PFC neurons 2, 3, and 4 show slightly the same behavior as they show a high firing activity during the stimulation. After the stimulus termination the firing rate activity of these neurons return back to their baseline.

B. *Experiment 2*

In experiment 2, the neuron 2 is set to show the highest firing rate activity for the 90 degrees stimulus orientation. All other 3 V1 neurons show different firing rate activity at this orientation. All conditions are the same for experiment 1 and 2 except that in experiment 2, the stimulus orientation is 90 degrees instead of 0 degrees. The PFC neuron 2 shows the bistable behavior at this experiment. The firing rate activity starts at a very low value and shows a spike during the stimulation which stays constant during the delay period. This bistable behavior would be the potential neural mechanism underlying the working memory.

Figure 8 shows the results of the experiment 2 in which V1 neuron 2 shows the highest firing rate activity for the 90 degrees of stimulus orientation.

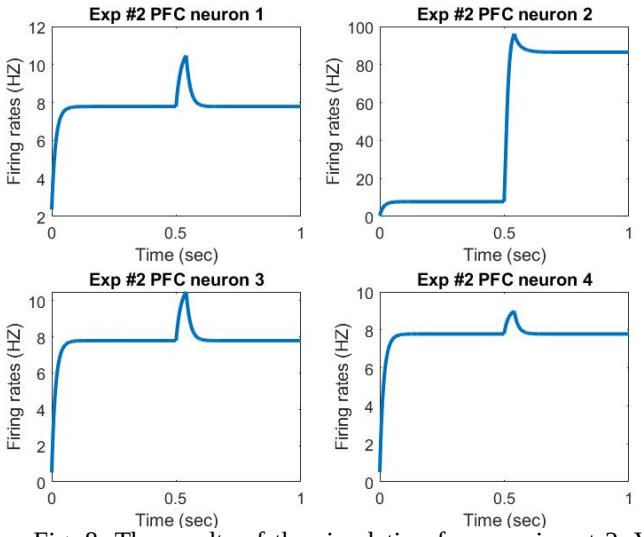

Fig. 8: The results of the simulation for experiment 2. V1 neuron 2 shows the highest firing rate activity for the 90 degrees orientation and PFC neuron 2 shows the bistable behavior accordingly.

C. *Experiment 3*

In experiment 3, a stimulus with 180 degrees of orientation is represented to V1 neurons. V1 neuron number 3 is set to show the highest firing rate activity at this orientation. PFC neuron 3 shows the bistable behavior, with low firing activity in the initial phase and high firing rate activity during the stimulation which stays stable during the delay period.

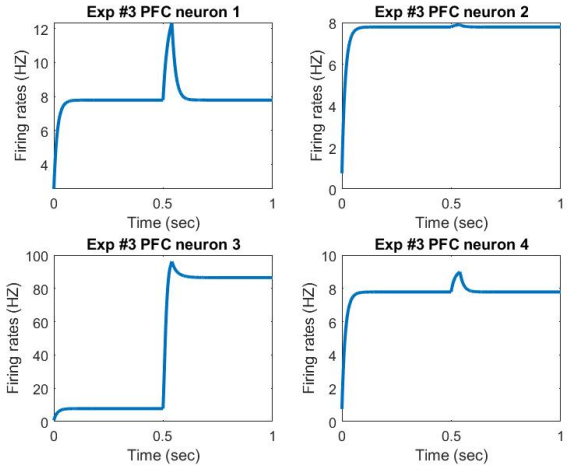

Fig. 9: The results of the simulation for experiment 3. V1 neuron 3 shows the highest firing rate activity for the 180 degrees orientation and PFC neuron 3 shows the bistable behavior accordingly.

D. *Experiment 4*

In Experiment 4, V1 neuron number 4 is set to show the Highest firing rate activity at the 270 degrees of stimulus orientation. The PFC neuron 1, 2, and 3 show the similar firing rate patterns whereas the PFC neuron 4 shows the bistable behavior (Fig. 10).

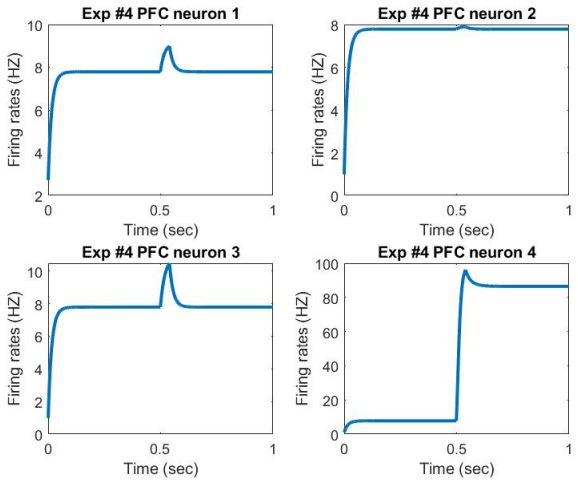

Fig. 10: The results of the simulation for experiment 4. V1 neuron 4 shows the highest firing rate activity for the 180 degrees orientation and PFC neuron 4 shows the bistable behavior accordingly.

IV. Discussion

Artificial neural networks have been widely used to solve computational problems as well as to reveal complicated brain processes. Several studies suggest that artificial neural networks can capture the same fundamental processes performed by the human brain [19], [20]. Some studies used the recurrent neural network to capture the fundamental features of PFC neurons and their ability to retain information in the memory and use that information to execute several cognitive tasks [21]. In this study, we replicated the experimental results revolving around working memory using Recurrent neural network composed of firing rate neurons. The results revealed the bistable behavior of PFC neurons in working memory consisted of low firing rate activity during the initial phase and before the stimulation, and high firing rate activity during the stimulation and the delay period. This behavior is due to the neurons ability to maintain information during the delay period and process this information for future utilization. Understanding the dynamics of working memory helps us build a more comprehensive understanding of the human cognition that could be used in various field of research such as Artificial intelligence.

Bistability rate neurons have been also used in deep neural networks recently, especially those in natural language processing. In particular, switching transformers use a gating mechanism reminiscent of bistable firing rate neurons to create sparse "expert" sub-networks, each of which is thought to become specialized at task-specific processing. This induced sparsity has allowed ever and ever larger transformer models to be trained [22].

Interestingly, some researchers found that inability of neurons to show bistable behavior in the nucleus accumbens is the underlying mechanism of some disorders such as ADHD and anxiety. The lack of bistable neurons reduces the synaptic gating of dopamine input from the hippocampus and PFC on the nucleus accumbens. This effect, decreases individuals' abilities to impede fear responses from amygdala, which in turn leads to anxiety [23].

For the future work, we will expand the current neural network with more V1 and PFC neuron populations and use random projections instead of all to all projections in order to achieve a more realistic result. We will also use BCM as a potential learning algorithms to train the network.

References

[1] N. Cowan, "What are the differences between long-term, short-term, and working memory? Nelson," *NIH Public Access*, vol. 6123, no. 07, pp. 323–338, 2009, doi: 10.1016/S0079-6123(07)00020-9.What.

[2] B. Koopmann-Holm and A. J. O'Connor, "Working memory," *Work. Mem.*, vol. 20, no. 4, pp. 1–82, 2017, doi: 10.4324/9781912282418.

[3] A. Baddeley, "Working memory: Looking back and looking forward," *Nat. Rev. Neurosci.*, vol. 4, no. 10, pp. 829–839, 2003, doi: 10.1038/nrn1201.

[4] T. Eisner, A. F. Kluge, J. E. Carrel, and J. Meinwald, "Neuron Activity Related to Short-Term Memory," no. 11, pp. 652–654.

[5] J. Quintana, J. M. Fuster, and J. Yajeya, "Effects of cooling parietal cortex on prefrontal units in delay tasks," *Brain Res.*, vol. 503, no. 1, pp. 100–110, 1989, doi: 10.1016/0006-8993(89)91709-5.

[6] E. K. Miller, C. A. Erickson, and R. Desimone, "Neural mechanisms of visual working memory in prefrontal cortex of the macaque," *J. Neurosci.*, vol. 16, no. 16, pp. 5154–5167, 1996, doi: 10.1523/jneurosci.16-16-05154.1996.

[7] M. V. Chafee and P. S. Goldman-Rakic, "Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task," *J. Neurophysiol.*, vol. 79, no. 6, pp. 2919–2940, 1998, doi: 10.1152/jn.1998.79.6.2919.

[8] C. Constantinidis and P. S. Goldman-Rakic, "Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex," *J. Neurophysiol.*, vol. 88, no. 6, pp. 3487–3497, 2002, doi: 10.1152/jn.00188.2002.

[9] M. Lundqvist, J. Rose, P. Herman, S. L. L. Brincat, T. J. J. Buschman, and E. K. K. Miller, "Gamma and Beta Bursts Underlie Working Memory," *Neuron*, vol. 90, no. 1, pp. 152–164, 2016, doi: 10.1016/j.neuron.2016.02.028.

[10] K. Wimmer, M. Ramon, T. Pasternak, and A. Compte, "Transitions between multiband oscillatory patterns characterize memory-guided perceptual decisions in prefrontal circuits," *J. Neurosci.*, vol. 36, no. 2, pp. 489–505, 2016, doi: 10.1523/JNEUROSCI.3678-15.2016.

[11] D. Jokisch and O. Jensen, "Modulation of gamma and alpha activity during a working memory task engaging the dorsal or ventral stream," *J. Neurosci.*, vol. 27, no. 12, pp. 3244–3251, 2007, doi: 10.1523/JNEUROSCI.5399-06.2007.

[12] S. Haegens, D. Osipova, R. Oostenveld, and O. Jensen, "Somatosensory working memory performance in humans depends on both engagement and disengagement of regions in a distributed network," *Hum. Brain Mapp.*, vol. 31, no. 1, pp. 26–35, 2010, doi:

10.1002/hbm.20842.

[13] S. Funahashi, C. J. Bruce, and P. S. Goldman-Rakic, “Mnemonic coding of visual space in the monkey’s dorsolateral prefrontal cortex,” *J. Neurophysiol.*, vol. 61, no. 2, pp. 331–349, 1989, doi: 10.1152/jn.1989.61.2.331.

[14] P. S. Goldman-Rakic, “Chapter 16 Cellular and circuit basis of working memory in prefrontal cortex of nonhuman primates,” *Prog. Brain Res.*, vol. 85, no. C, pp. 325–336, 1991, doi: 10.1016/S0079-6123(08)62688-6.

[15] X. J. Wang, “Synaptic reverberation underlying mnemonic persistent activity,” vol. 24, no. 8, pp. 455–463, 2001.

[16] O. Barak and M. Tsodyks, “Working models of working memory,” *Curr. Opin. Neurobiol.*, vol. 25, pp. 20–24, 2014, doi: 10.1016/j.conb.2013.10.008.

[17] G. Swan and B. Wyble, “The binding pool: A model of shared neural resources for distinct items in visual working memory,” *Attention, Perception, Psychophys.*, vol. 76, no. 7, pp. 2136–2157, 2014, doi: 10.3758/s13414-014-0633-3.

[18] J. Barbosa *et al.*, “Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory,” *Nat. Neurosci.*, vol. 23, no. 8, pp. 1016–1024, 2020, doi: 10.1038/s41593-020-0644-4.

[19] S. A. Cadena *et al.*, “Deep convolutional models improve predictions of macaque V1 responses to natural images,” *PLoS Comput. Biol.*, vol. 15, no. 4, pp. 1–27, 2019, doi: 10.1371/journal.pcbi.1006897.

[20] P. Bashivan, K. Kar, and J. J. DiCarlo, “Neural population control via deep image synthesis,” *Science (80-.).*, vol. 364, no. 6439, 2019, doi: 10.1126/science.aav9436.

[21] G. R. Yang, M. R. Joglekar, H. F. Song, W. T. Newsome, and X. J. Wang, “Task representations in neural networks trained to perform many cognitive tasks,” *Nat. Neurosci.*, vol. 22, no. 2, pp. 297–306, 2019, doi: 10.1038/s41593-018-0310-2.

[22] W. Fedus, B. Zoph, and N. Shazeer, “Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity,” *J. Mach. Learn. Res.*, vol. 23, pp. 1–40, 2022.

[23] F. Levy, “Synaptic gating and ADHD: A biological theory of comorbidity of ADHD and anxiety,” *Neuropsychopharmacology*, vol. 29, no. 9, pp. 1589–1596, 2004, doi: 10.1038/sj.npp.1300469.