



A Computational Model of Working Memory Based on Spike-Timing-Dependent Plasticity

Qiu-Sheng Huang and Hui Wei*

Laboratory of Cognitive Model and Algorithm, Shanghai Key Laboratory of Data Science, Department of Computer Science, Fudan University, Shanghai, China

Working memory is closely involved in various cognitive activities, but its neural mechanism is still under exploration. The mainstream view has long been that persistent activity is the neural basis of working memory, but recent experiments have observed that activity-silent memory can also be correctly recalled. The underlying mechanism of activity-silent memory is considered to be an alternative scheme that rejects the theory of persistent activity. We propose a working memory model based on spike-timing-dependent plasticity (STDP). Different from models based on spike-rate coding, our model adopts temporal patterns of action potentials to represent information, so it can flexibly encode new memory representation. The model can work in both persistent and silent states, i.e., it is compatible with both of these seemingly conflicting neural mechanisms. We conducted a simulation experiment, and the results are similar to the real experimental results, which suggests that our model is plausible in biology.

Keywords: spike-timing-dependent plasticity, working memory, computational model, synaptic plasticity, spiking neural network

OPEN ACCESS

Edited by:

Si Wu,
Peking University, China

Reviewed by:

Dong Song,
University of Southern California,
United States

Antonio Batista,
Universidade Estadual de Ponta
Grossa, Brazil

Kelly Cristiane Iarosz,
University of São Paulo, Brazil

*Correspondence:

Hui Wei
weihui@fudan.edu.cn

Received: 19 November 2020

Accepted: 31 March 2021

Published: 21 April 2021

Citation:

Huang Q-S and Wei H (2021) A
Computational Model of Working
Memory Based on
Spike-Timing-Dependent Plasticity.
Front. Comput. Neurosci. 15:630999.
doi: 10.3389/fncom.2021.630999

INTRODUCTION

Working memory is the ability to preserve information for further processing (Baddeley, 2010). It participates in complex tasks such as reasoning, understanding, and learning, and is an important part of the cognitive function. Reflecting its importance, working memory has been a research object for many years, but its neural mechanism is still not fully understood. As early as 1971, a study on short-term memory by Fuster and Alexander reported that some neurons of the prefrontal cortex maintained a higher firing rate in the delay period than in the intertrial period (Fuster and Alexander, 1971). Similar phenomena were observed in subsequent experiments (Funahashi et al., 1989; Nakamura and Kubota, 1995; Watanabe and Funahashi, 2007; Wimmer et al., 2014). Since there is no external cue stimulus during the delay period, this spontaneous activity is likely related to the memory content. Some studies have found that if researchers try to inhibit the persistent activity of neurons during the delay period, the accuracy of task recall will decrease (Quintana et al., 1989). More importantly, a machine learning model trained by the neuronal spike data collected during persistent activity can predict the behavior of animals in the recall phase (Masse et al., 2017). Therefore, persistent activity is considered the most reasonable candidate mechanism for working memory, and some computational models based on the interconnection of neurons have been proposed. For example, in the ring model (Compte et al., 2000), the connection strength of neurons is determined according to their stimulus selectivity, and the input information is maintained

by mutual excitation of adjacent neurons. Similar models reproduce the phenomena observed in neurobiology experiments, but they are usually only suitable for spatial orientation memory tasks.

Recent studies have found that working memory can be independent of persistent activities, which provides a new way to understand its neural mechanism. For example, under dual tasks, even when an experimental subject makes the correct choice, the spatial selectivity of delayed activities is significantly reduced (Watanabe and Funahashi, 2014), indicating that working memory can be retained in the “activity silence” neural state (Stokes, 2015). Another study found that transient gamma oscillations are accompanied by reactivation of coding and sensory information (Lundqvist et al., 2016), which means gamma oscillations may also be the basis of working memory. Accordingly, models independent of persistent activities have been proposed, in which information is stored by synaptic changes rather than neural activity. This change occurs through specific synaptic mechanisms, such as presynaptic calcium residue (Mongillo et al., 2008) and short-term enhancement (Fiebig and Lansner, 2017; Mi et al., 2017).

Since the observed experimental phenomena are quite different, these two mechanisms are usually considered incompatible. However, because of differences in species types, individual conditions, and experimental methods, the observed phenomena may not address the same problem. Even in research related to persistent activity, only a part of the neurons related to the experimental task show persistent activity, and they may exhibit complex, diverse temporal patterns (Watanabe and Funahashi, 2007). This suggests that these two mechanisms may coexist in the brain or stem from a deeper mechanism. Researchers have recently attempted to integrate the two mechanisms into a unified framework (Kaminski and Rutishauser, 2020). Such work provides a new perspective on the neural mechanism of working memory, and its significance is obvious.

A variety of synaptic plasticity mechanisms have been applied to working memory models. Past models mostly adopted the plasticity mechanism based on facilitation. Since synaptic reinforcement is not selective, these models cannot explain the encoding of novel association (Fiebig and Lansner, 2017). To overcome these shortcomings, some researchers have adopted Hebbian’s rule (Fiebig and Lansner, 2017), whose idea is to use a selective synaptic enhancement mechanism. Following this idea, we propose a working memory model based on STDP. Under the influence of STDP, the neurons can enhance the causal connections and ignore the irrelevant connections (Markram et al., 2012), thus supporting more flexible encoding schemes. Although the STDP rule is usually applied to a computational model on a longer time scale, some experimental phenomena under this rule at different time scales have been observed (Caporale and Dan, 2008). Therefore, it is reasonable to apply STDP to a computational model on a smaller time scale. We conducted three experiments to evaluate our model, following the common delay-match-to-sample (DMS) paradigm. The first two experiments evaluated the model under persistent and silent activity, and the last was compared to a psychology experiment.

The experimental results show that memory items in our model can be well recalled in both neural activity states. The simulation results are similar to real experiment results, suggesting that our model is reasonable and plausible in biology.

MATERIALS AND METHODS

Neural Circuit Model

The neural circuit model represents a functional cluster of neurons in the prefrontal cortex for single-item memory (Figure 1). The circuit consists of $N_E = 24$ excitatory neurons and $N_I = 6$ inhibitory neurons. The number of excitatory neurons is four times that of inhibitory neurons, which meets the proportion of neurobiological findings (Markram et al., 2004). All excitatory neurons are fully connected, and each connection consists of $S_{EE} = 4$ synapses. The strengths of these synapses are initially random values between 0 and 0.14. The connection probability between excitatory neurons and inhibitory neurons is 0.8, and their connection strength is fixed. Inhibitory neurons are used to inhibit excitatory neurons to prevent overexcitation, so there is no connection between them. According to the reports of axon conduction delays in cat and rabbit cerebral cortex (Miller, 1975; Swadlow, 1990), the average conduction delay of connections in our model is set at about 12 ms. Specifically, the conduction delays of connections between excitatory neurons are set to $d_{ij}^k = R_{EE} + k * 3$, where R_{EE} is a random value between 3 and 12, and $k = 0, 1, 2, 3$ is the index corresponding to four synapses of each connection. The delays between inhibitory neurons are random values between 1 and 20 ms. Among the 24 excitatory neurons in the network model, $N_{Input} = 15$ are used to receive the input of a spike sequence, and as output neurons to extract memory items.

Neuron Model

We used the Izhikevich neuron model to simulate the firing behavior of real neurons (Izhikevich, 2003). It combines the advantages of the leaky integrate-and-fire and Hodgkin-Huxley models and can simulate various spiking patterns with relatively simple calculation. The model is represented as follows:

$$v' = 0.04v^2 + 5v + 140 - u + I, \quad (1)$$

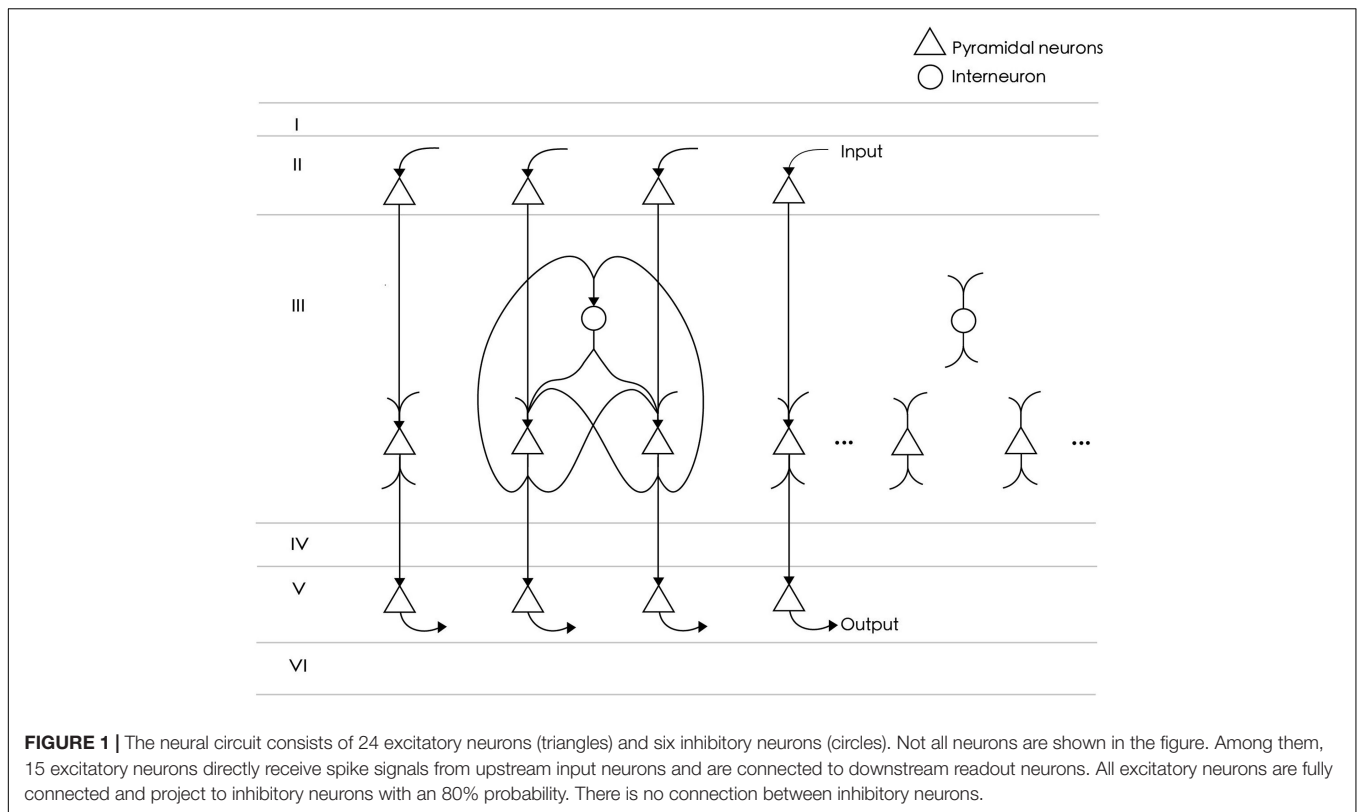
$$u' = a(bv - u), \quad (2)$$

$$\text{if } v > 30\text{mV, then } \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \quad (3)$$

where v is the membrane potential of the neuron, u is the recovery variable of the membrane potential. I is the current input to the neuron, which includes the current from other neurons in the model and noise current, that is:

$$I = I_{\text{input}} + I_{\text{noise}}, \quad (4)$$

where I_{noise} is Gaussian noise with mean ranging from 1.5 to 2 and variance 1.8. It can be reasonably assumed as irrelevant input from nearby cortical tissues. a, b, c, d in Equation 2



and 3 are dimensionless parameters. The model can reproduce different spiking patterns according to the parameter settings. The parameters of excitatory neurons in our model are $a = 0.02$, $b = 0.2$, $c = -65$, $d = 8$, and those of inhibitory neurons are $a = 0.04$, $b = 0.24$, $c = -65$, $d = 2$.

Synapse Model

Synapses are the key parts of neurons, which pass spiking signals to each other. In the main kind of synapse in the nervous system of higher vertebrates, the chemical synapse, an action potential makes the terminal of the presynaptic axon release neurotransmitters, which diffuse across the synaptic cleft and bind to receptor proteins on the postsynaptic membrane, causing the increase or decrease of postsynaptic membrane potential (Kandel et al., 2012). Following the design of Bohte et al. (2002), we describe the time-dependent change of postsynaptic potential caused by an action potential as:

$$V_{PSP}(t) = \frac{t}{\tau} e^{1-t/\tau}, \tag{5}$$

where τ is a time constant, which we set to 4. The equation describes the potential changes caused by a single spike, but in real situations, the high-frequency spikes on a single synapse may cause multiple potential changes to overlap. When synapses fire multiple spikes in a short period, the amplitude of the changes of the postsynaptic membrane potential caused by each spike will steadily decrease due to the depletion of neurotransmitters, which gradually return to normal over time. This effectively reduces the possibility of over-firing of the whole neural circuit.

We refer to the model of Mongillo et al. (2008) to describe this process:

$$x' = \frac{1-x}{\tau_D} - ux\delta(t-t_{sp}), \tag{6}$$

$$u' = \frac{U-u}{\tau_F} + U(1-u)\delta(t-t_{sp}), \tag{7}$$

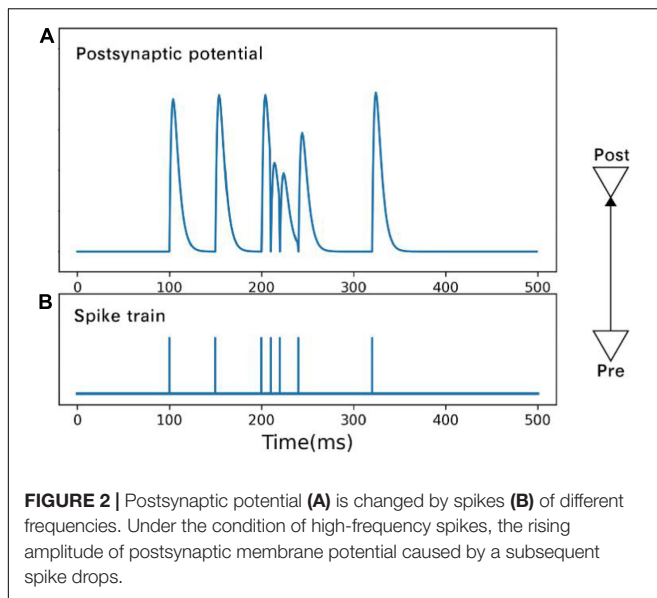
where x is the available resources of the synapse; u is the utilization rate of resources by each spike; $' = d/dt$; δ is the Dirac function; t, t_{sp} are the current time and firing time, respectively, of a spike; U is the benchmark level of utilization; and τ_F, τ_D are recovery time parameters set to 20 and 50, respectively, in our model. The transmission efficiency of a synapse is expressed as xu . Therefore, the final postsynaptic membrane potential at time t can be expressed as:

$$V_{PSP}(t) = \sum_{i=1}^n u_i x_i \frac{t-t_i}{\tau} e^{1-(t-t_i)/\tau}, \tag{8}$$

where u_i, x_i, t_i are the available resources, utilization rate, and arrival time, respectively, of the i th spike. **Figure 2** shows the postsynaptic potential produced by spikes of different frequencies.

Spike-Timing-Dependent Plasticity

Synaptic strength may change under certain induction events (Gustafsson et al., 1989). Research in the late 1990s shows that the difference of firing time between pre- and post-synaptic neurons



has a significant impact on the change of synaptic strength (Gerstner et al., 1996; Markram et al., 1997; Bi and Poo, 1998). The synaptic strength is enhanced when the pre-synaptic neuron fires slightly before the post-synaptic neuron and depressed when they fire in reverse order. Such synaptic plasticity is called spike-timing-dependent plasticity (STDP). The relationship between the difference of firing time and the change of synaptic strength is expressed by STDP function, which can be obtained from spiking activities with appropriate modeling (Song et al., 2015; Robinson et al., 2016). STDP is different at different synapses. It exists not only in excitatory synapses but also in inhibitory synapses (Haas et al., 2006). Hence, there are many forms of STDP function. We apply a traditional form to the model (Song et al., 2000; Bi and Poo, 2001). The STDP function is:

$$\Delta W(\Delta t) = \begin{cases} A_+ \exp(\Delta t / \tau_+) & \text{if } \Delta t < 0 \\ -A_- \exp(\Delta t / \tau_-) & \text{if } \Delta t \geq 0 \end{cases}, \quad (9)$$

where Δt is the time difference between the pre- and post-synaptic neuron; τ_+ , τ_- are parameters that affect the time windows; and A_+ , A_- are the maximum degree of enhancement or weakness, respectively. The spike pairing scheme in this model is the nearest-neighbor scheme (Morrison et al., 2007), i.e., a presynaptic spike is paired with only the last postsynaptic spike, and a postsynaptic spike is paired with only the last presynaptic spike.

Studies have shown that STDP can cause neurons to synchronize or desynchronize (Talathi et al., 2008; Popovych and Tass, 2012; Popovych et al., 2013; Borges et al., 2016, 2017; Lameu et al., 2018), and the synchronization of neural activity is bound up with brain information processing (Baravalle et al., 2019). According to the encoding scheme in this model, the best decoding case is that the neurons are synchronized with specific phases (see section “Encoding and Decoding”). However, strict synchronization, that is, all neurons fire at the same time, is not ideal, because the order information of the spikes is

missed. Previous researches usually set the potentiation and depression windows in STDP to be approximately the same size. After testing, we found that this can easily lead to strict synchronization. Therefore, we increased the depression window. The STDP parameters in our model are: $A_+ = 1$, $A_- = -1$, $\tau_+ = 3$, $\tau_- = 18$.

The synaptic strength is updated according to $W_t = W_{t-1} + \eta \Delta W$, where η is the learning rate. The learning rate should be higher in the working memory model because the sample needs to be memorized quickly. On the other hand, the excessive value will lead to unstable learning. After testing, we set the learning rate to 0.2.

Encoding and Decoding

How the brain encodes information is a key point for the working memory model. We focus on the temporal information carried by the spike, just like our previous work (Wei and Du, 2019). The encoding scheme of memory items in this model comes from rank order coding. This strategy is based on population coding, in which the information is represented by a group of neurons. The idea of order coding is that the firing time order of neurons is critical, and not the exact firing latency time (Gautrais and Thorpe, 1998). This simple idea is plausible in biology. The time from receiving an input to excitation is related to the input's intensity (Heil, 2004; Gollisch and Meister, 2008). Therefore, when there is a stimulus input, neurons encoding different dimensions produce spikes with various excitation latencies to generate a specific spike train.

Applying sequential coding to a computational model has at least two advantages. First, it helps to quickly and accurately identify spike trains. In theory, the complete recognition time of order coding is merely the arrival time difference between the first and last spike, while traditional rate-based coding must count the number of spikes over a time. To reduce the statistical error, the time window is often large; hence, it takes more time to discriminate different spike trains. Second, order coding has high efficiency, and it can represent much information with only a few neurons. If each neuron only fires one spike, then n neurons with order coding can represent $n!$ states, while rate-based coding scheme can represent at most $n + 1$ states (Gautrais and Thorpe, 1998).

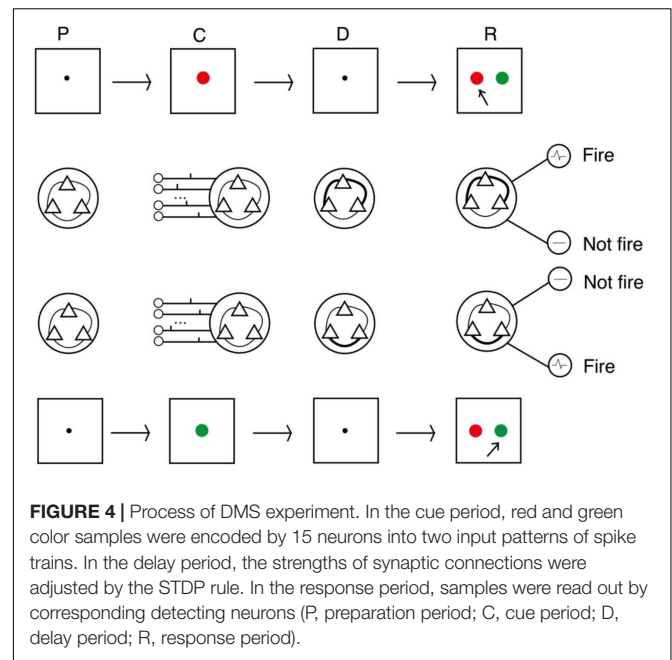
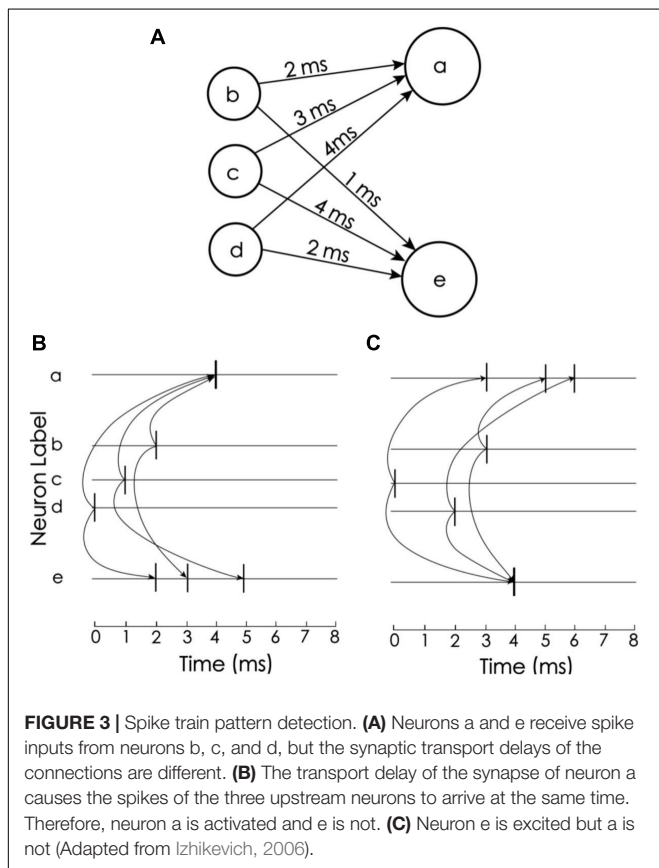
Although the specific firing time of each spike is immaterial to the order coding scheme, the minimum distinguishable time between two spikes must be considered in practice. If the firing times of two spikes are close, then their order may be wrongly resolved under the influence of errors. To simplify the model and improve the recognition accuracy, we add a limitation based on order coding: only one neuron fires a spike during any 4 ms. In this way, the firing time of each neuron can be calculated from a spike train. Considering the case of three neurons, the firing times of neurons 0–2, corresponding to the sequence (1, 3, 2), are 4 ms, 12 ms, and 8 ms.

In this coding scheme, the information can be easily recognized by the corresponding detection neurons. The idea of decoding comes from Izhikevich's polychronization concept (Izhikevich, 2006), i.e., to set the propagation delay so that a train of spikes with a given order arrives at the corresponding

detection neurons at the same time. In the proposed model, there are 15 neurons to encode and decode memory items, but for simplicity, we consider only three neurons to explain the idea (Figure 3). Neurons a and e are detection neurons of different spike trains, and the respective propagation delays between a and e and neurons b, c, and d are 2 ms, 3 ms, 4 ms and 1 ms, 4 ms, 2 ms (Figure 3A). When the firing times of b, c, and d are 2 ms, 1 ms, and 0 ms, respectively (Figure 3B), these spikes arrive at neuron a at the same time, causing it to excite. For neuron e, these spikes arrive at different times. Due to the leakage of the membrane current, these spikes cannot accumulate enough current to excite neuron e. Conversely, if the firing times of b, c, and d are 3 ms, 0 ms, and 2 ms, then e is excited and a is not. Therefore, neurons a and e can be used to detect spike train patterns (2, 1, 0) and (3, 0, 2), respectively.

Experimental Procedure

We conducted three experiments to evaluate the proposed model. The procedures followed the DMS task paradigm, and the sample stimuli were two colors (red and green). The procedure is shown in Figure 4. The first second of an experiment was a preparation period, when the subjects prepared to receive the sample stimulus prompt. At this time, there was no external input, and the model was only affected by random noise current from other neurons in the cerebral cortex. The next second was the cue period. Subjects were prompted by either red or green, and the



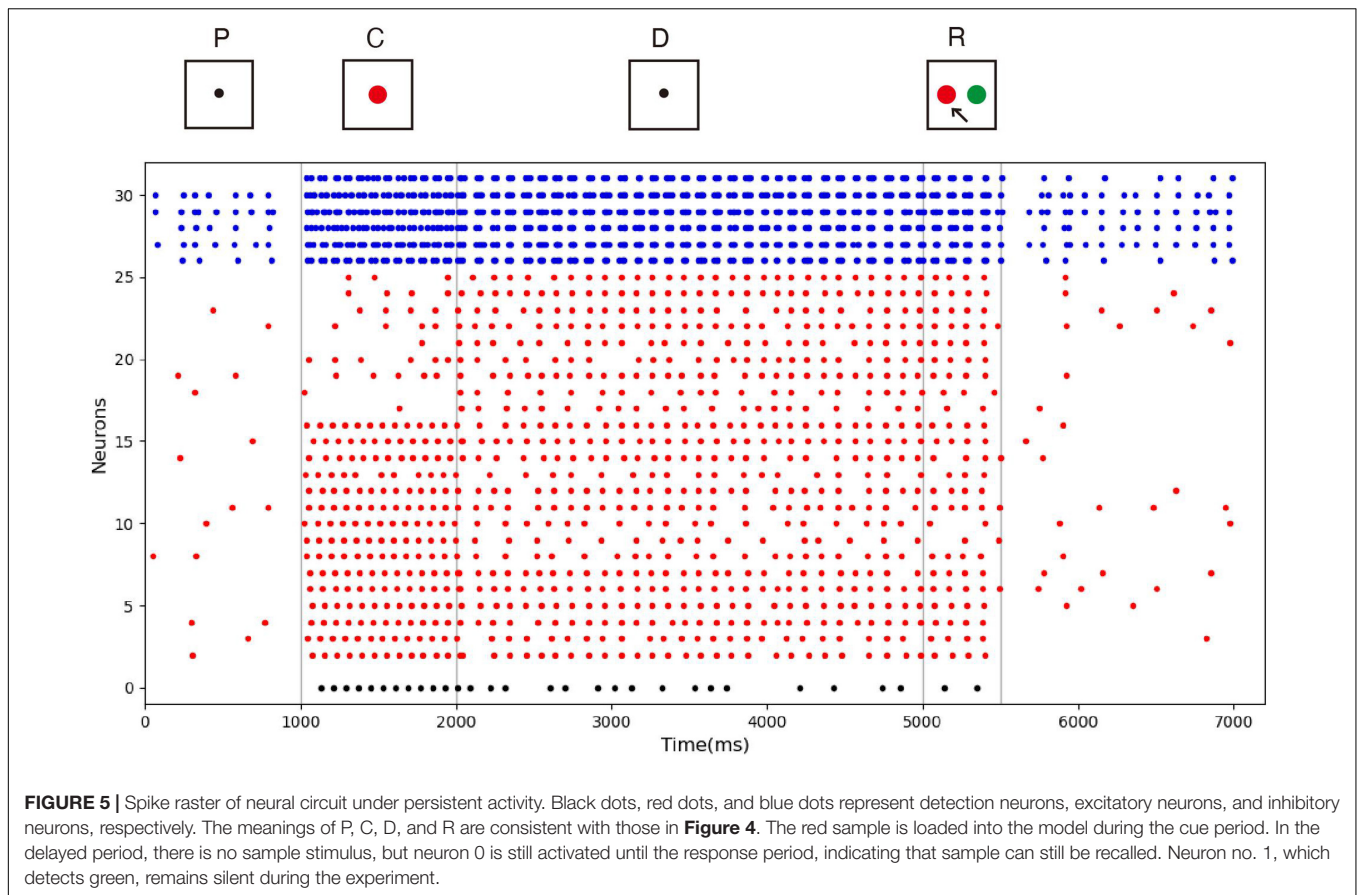
spike train that encoded the color was loaded into the model. The sample stimulus disappeared at the beginning of the delay period. After a 3-s delay, subjects were asked to recall the sample. This was done by imposing a non-specific recall signal to the model, e.g., an extra noise current. The input sample stimuli were encoded by 15 neurons as two different spike train patterns, and decoding was performed by detection neurons corresponding to the two patterns. In the response period, the activities of detection neurons within 0.5 s were counted, and the sample corresponding to the more active neuron was the recalled sample.

RESULTS

Memory Under Persistent Activity

In electrophysiological experiments of working memory, it is usually observed that some neurons relevant to the experiment maintain a firing rate significantly higher than the reference level during the delay period, and this is known as persistent activity. Because of its strong correlation with the memory task, persistent activity theory is dominant in explaining the neural mechanism of working memory. The first experiment tested whether the model can maintain memory and display persistent activity during the delay period. For this purpose, we increased the mean of the noise current by 1.5 for each excitatory neuron. This leads to a slight depolarization, which makes the neuron likely to fire.

The spike raster of the whole neuronal population during the experiment is shown in Figure 5. Neurons 0 and 1 (black) are the detection neurons; neurons 2–25 (red) are excitatory neurons, including 15 neurons (2–16) that receive input; and neurons 26–31 (blue) are inhibitory neurons. The first second is the initial ground state of the model. During this period, the neurons are driven by the noise current from the surrounding cortex, so they irregularly fire at low frequency. During the cue period of the



first second to the second second, a sample stimulus (red) is loaded into the network model, and the network continuously receives the spike train encoded by 15 input neurons. No. 0 neuron is activated at the same time as the detection neuron of red color, while no. 1 neuron, which detects the green color stimulus, is completely silent. The delay period is from the second to fifth second, during which the excitatory neurons show sustained excitation. Although the input of the sample stimulus has disappeared, neuron no. 0 is still activated frequently, which indicates that memory items still exist in the model network. After the response period, the activity of the whole group of neurons decreases and returns to the ground state of low frequency and irregular excitation.

Although there is no unified definition of persistent activity, it is generally believed that persistent activity is related to memory information rather than meaningless random activity. Therefore, some researchers use machine learning methods such as support vector machine (SVM) to analyze the activities of neurons related to working memory. Here, we trained a simple SVM to decode the activity data in the delay period of the model. The firing rates of each neuron were the average spiking activity in 500-ms bins sampled at 200-ms sliding intervals. These data were used as samples to train the SVM.

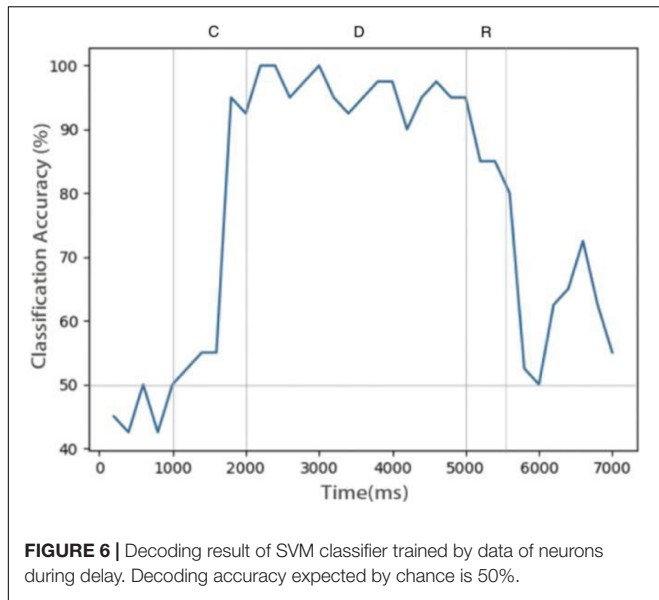
Figure 6 shows the relationship between decoding accuracy and time. Since there are only two types of stimulus, the probability of decoding by chance is 50%. It can be seen that

after the end of the sample cue period, the decoding accuracy rate reaches more than 90% and remains there for the whole delay period, which proves that the persistent activity in the model is indeed related to the stimulus information, rather than purely random activity. During the delay period, the accuracy rate decreases gradually, indicating that the memory is gradually blurred. This is done entirely through STDP, with no other mechanism. We also see that SVM can decode correctly from the release activity after the end of the response period, which indicates that there is still some residual memory.

Memory Under “Silent” State

While the mainstream view is that persistent activity underlines working memory, some believe that memory information can be stored in a “silent” neuron state. Therefore, our second experiment tested whether the model can maintain memory in the silent state. This experiment was much like the first, but the mean of the noise current of each excitatory neuron was reduced by 0.3, so they could not form obvious persistent activities. To show another sample stimulus, green was used as an input. Note that the choice of sample stimulus does not affect the model’s performance.

The spike raster of the whole neuron group in the simulation experiment is shown in **Figure 7**. The first 2 s of the experiment are similar to experiment 1. After the cue period, the neuronal population does not show sustained excitation as in experiment 1,

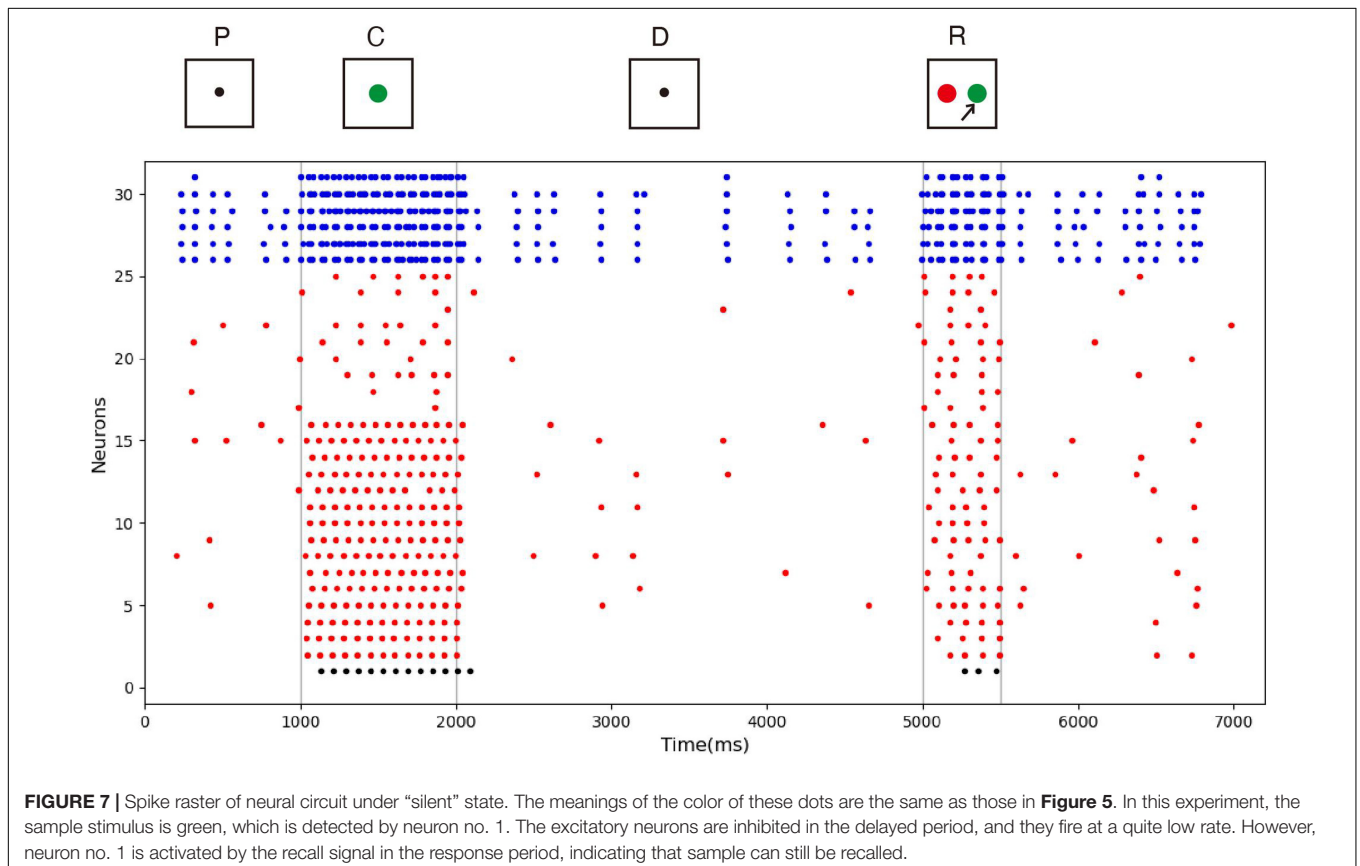


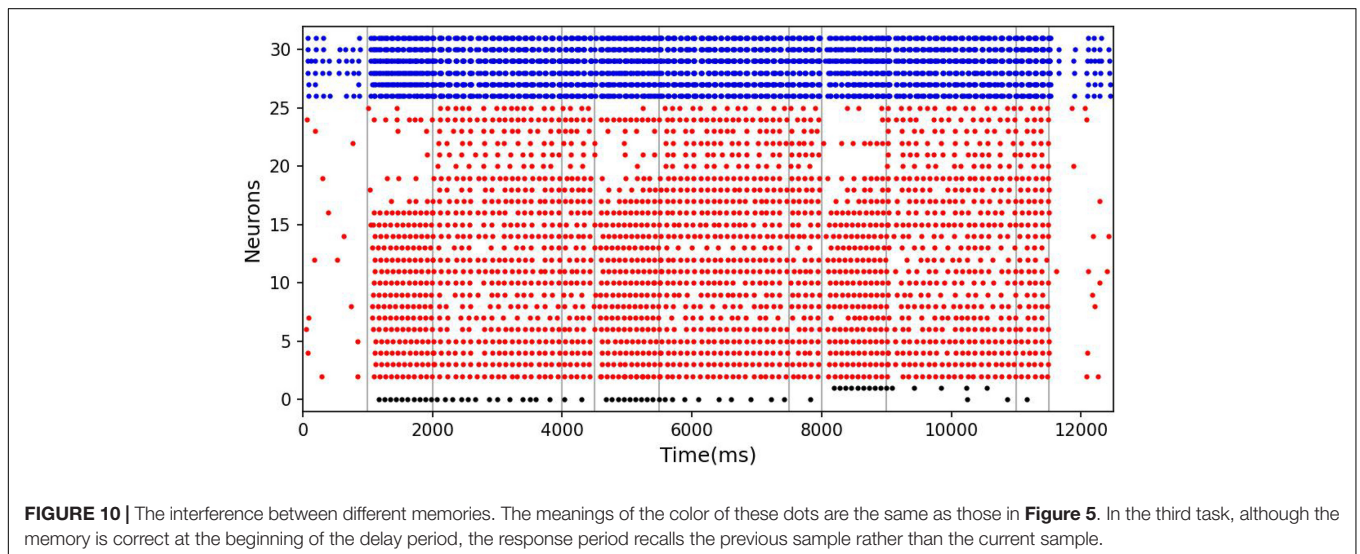
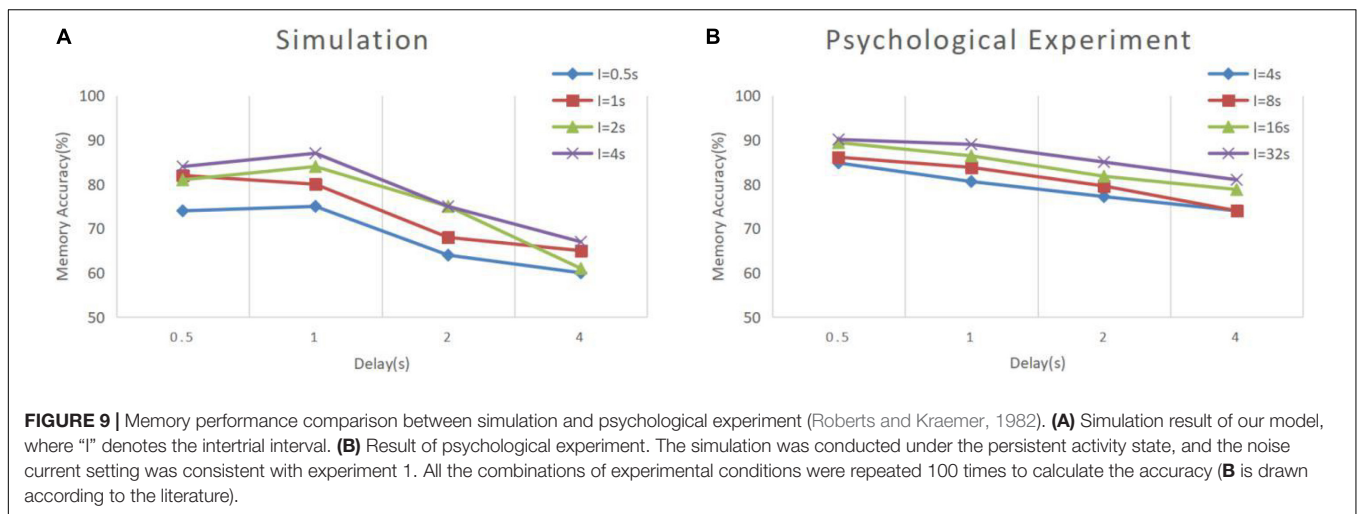
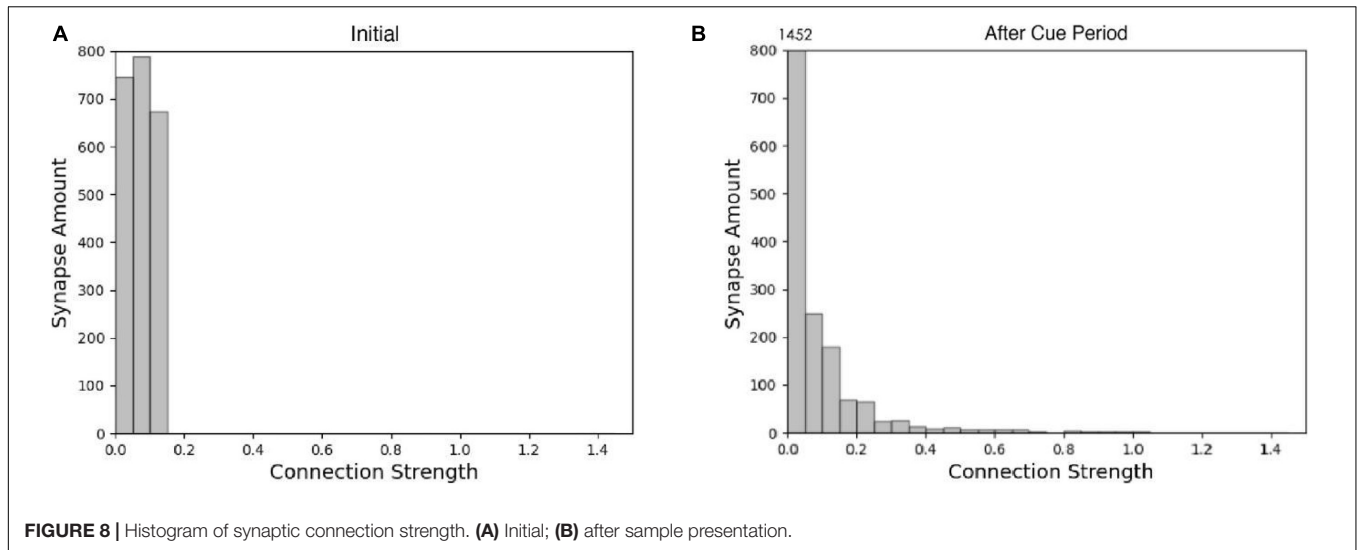
but returns to the low-frequency ground state due to the inhibition. However, despite the lack of sustained activity during the delay period, neuron no. 1, which detects green, is activated under the stimulation of a non-specific recall signal during the response period, implying that the sample can still be accurately

recalled. This indicates that our model preserves information through changes in synaptic connections rather than persistent activity. During the cue period, the information of relative firing time difference between neurons is preserved in synaptic connections by STDP. The recall signal in the response period reproduces the spike pattern of the sample stimulus, which can be discriminated by the detection neurons. **Figure 8** shows the changes in the strengths of synaptic connections after the sample loading. In the initial state, the synaptic connection strength is set at about 0.07 on average. After learning, most of the synaptic strength is close to zero, and only a small part is greater than 0.2, showing a lognormal distribution.

Simulation Experiment

In Experiment 3, we applied the model to continuous DMS tasks and compared the data to that of a pigeon DMS experiment by Roberts and Kraemer (1982). In their experiment, pigeons were kept in cages with three buttons, placed side by side. After the experiment started, the middle button lit and displayed green or red as the sample stimulus. Next were delay and response periods. In the response period, the left and right buttons were illuminated, respectively, in green and red as alternative samples. A pigeon that made the right choice was rewarded, and the experiment was finished; otherwise, the experiment finished directly. The next experiment started after a short interval. Roberts and Kraemer explored the accuracy of pigeon matching





with different combinations of delay intervals and experiment intervals. Our simulation experiment is basically consistent with their experiment, except the sample prompt period and experimental interval were reduced to one-eighth of the original to meet the time scale of our model.

The results of model simulation are shown in **Figure 9A**. For comparison, we draw **Figure 9B** based on the experimental data of Roberts and Kraemer. It can be seen that the simulation results are consistent with the trend reflected by the real experiment. Each line in **Figure 9** shows a downward trend, and the line with the longer experimental interval is placed at the upper position. This indicates that the recall accuracy decreases with the increase of the delay period, and increases with the intertrial interval. It is not surprising that the accuracy drops over time. Under the interference of noise, the synaptic connections formed in the cue period gradually becomes disordered, and finally, the sample cannot be recalled. The increase of accuracy with the intertrial interval may be related to the interference between memories. To test this, we conducted another experiment. We used our model to perform the DMS task three times in succession. The first two samples were the same but different from the third. The experiment was under the persistent activity state, and the delay interval and the intertrial interval were 2000 ms, 0 ms, respectively. We find that there is a possibility that the first two samples interfere with the third sample, as shown in **Figure 10**.

It can be seen that our model has greater volatility and does not strictly follow the trend under some combinations of conditions. We deduce that this is due to the limitations of the small scale of the model. We simulate a memory circuit consisting of only 30 neurons, which is vulnerable to noise currents. The number of neurons in the brain is much greater, and there are many redundant functional circuits, so the psychological experimental performance is more stable. We believe that this difference will be greatly reduced if multiple circuits are combined in a larger memory network.

CONCLUSION AND DISCUSSION

We proposed a working memory neural circuit model based on STDP, an extension of Hebbian's rule, which is considered an important learning rule in the brain's nervous system. It was verified in electrophysiological experiments on multiple time scales. Our experiment explored the feasibility of STDP as a mechanism of working memory and found that it could form a specific synaptic connection state to maintain memory items in a short time scale. The connection state could reproduce the spike pattern of the memory item with non-specific recall signals. In addition, due to the coding scheme based on the spike time information, the model can memorize items flexibly and efficiently without prior training. We tested the model in states of persistent activity and silence, and both could correctly recall the memory item in the response period. We compared the simulation data to real experimental data, and the results were similar, indicating that our model is biologically reasonable.

The internal mechanism of working memory is still under exploration. The traditional view is that persistent activity is the

basis of working memory, but experimental phenomena observed in recent years suggest the possibility that working memory can be preserved in the connection state between neurons without the continuous firing of neurons. Persistent activity theory and connectivity theory are usually regarded as opposing views, but some researchers have integrated them, and think that both persistent and non-persistent mechanisms may exist in the brain. Our model supports this view. However, as pointed out before, memory stored in the silent state is more energy-saving, more stable, and less susceptible to interference than in the persistent activity state (Lundqvist et al., 2018). So, why does the brain not preserve all memory in a silent state? A reasonable explanation is that memory can be read more quickly in a state of persistent activity. In our model, memory reading in the silent state takes time, because the signal must pass to the circuit, and memory information is then passed to the downstream neurons. In a state of persistent activity, the memory content is ready to be read, so there is no delay. In some psychology theories, working memory and attention are considered to be closely related, and memory can be divided into different states according to the attention distribution (Cowan, 1988; Engle et al., 1999; Oberauer, 2002). By distinguishing different states, the brain manages working memory more efficiently. In our model, working memory is preserved through the synaptic connection state, and persistent activity may be an accessory produced by the brain for faster reading during the memory process, suggesting that the phenomena of persistent activity and the silent state may differ according to conditions in the same neural mechanism. Our study provides conjecture for integrating the two mechanisms.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

HW carried out the computational neuroscience model and neural circuit design studies, experiment design, and writing the manuscript. Q-SH designed algorithms, coded simulation program and collected the experimental data, and wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the NSFC Project (Project Nos. 61771146 and 61375122) and the National Thirteen 5-Year Plan for Science and Technology (Project No. 2017YFC1703303).

ACKNOWLEDGMENTS

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of the manuscript.

REFERENCES

- Baddeley, A. (2010). Working memory. *Curr. Biol.* 20, R136–R140. doi: 10.1016/j.cub.2009.12.014
- Baravalle, R., Guisanade, N., Granado, M., Rosso, O. A., and Montani, F. (2019). Characterization of visuomotor/imaginary movements in eeg: an information theory and complex network approach. *Front. Phys.* 7:115. doi: 10.3389/fphys.2019.00115
- Bi, G., and Poo, M. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu. Rev. Neurosci.* 24, 139–166. doi: 10.1146/annurev.neuro.24.1.139
- Bi, G.-Q., and Poo, M.-M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* 18, 10464–10472. doi: 10.1523/jneurosci.18-24-10464.1998
- Bohte, S. M., La Poutre, H., and Kok, J. N. (2002). Unsupervised clustering with spiking neurons by sparse temporal coding and multilayer RBF networks. *IEEE Trans. Neural Netw.* 13, 426–435. doi: 10.1109/72.991428
- Borges, R. R., Borges, F. S., Lameu, E. L., Batista, A. M., Iarosz, K. C., Caldas, I. L., et al. (2016). Effects of the spike timing-dependent plasticity on the synchronisation in a random Hodgkin–Huxley neuronal network. *Commun. Nonlinear Sci. Numer. Simul.* 34, 12–22. doi: 10.1016/j.cnsns.2015.10.005
- Borges, R. R., Borges, F. S., Lameu, E. L., Batista, A. M., Iarosz, K. C., Caldas, I. L., et al. (2017). Spike timing-dependent plasticity induces non-trivial topology in the brain. *Neural Netw.* 88, 58–64. doi: 10.1016/j.neunet.2017.01.010
- Caporale, N., and Dan, Y. (2008). Spike timing-dependent plasticity: a Hebbian Learning rule. *Annu. Rev. Neurosci.* 31, 25–46. doi: 10.1146/annurev.neuro.31.060407.125639
- Compte, A., Brunel, N., Goldman-Rakic, P. S., and Wang, X. J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* 10, 910–923. doi: 10.1093/cercor/10.9.910
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychol. Bull.* 104, 163–191. doi: 10.1037/0033-2909.104.2.163
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., and Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *J. Exp. Psychol. Gen.* 128, 309–331. doi: 10.1037/0096-3445.128.3.309
- Fiebig, F., and Lansner, A. (2017). A spiking working memory model based on Hebbian short-term potentiation. *J. Neurosci.* 37, 83–96. doi: 10.1523/JNEUROSCI.1989-16.2016
- Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* 61, 331–349. doi: 10.1152/jn.1989.61.2.331
- Fuster, J. M., and Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science* 173, 652–654. doi: 10.1126/science.173.3997.652
- Gautrais, J., and Thorpe, S. (1998). Rate coding versus temporal order coding: a theoretical approach. *Biosystems* 48, 57–65. doi: 10.1016/s0303-2647(98)00050-1
- Gerstner, W., Kempter, R., van Hemmen, J. L., and Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature* 383, 76–81. doi: 10.1038/383076a0
- Gollisch, T., and Meister, M. (2008). Rapid neural coding in the retina with relative sparse latencies. *Science* 319, 1108–1111. doi: 10.1126/science.1149639
- Gustafsson, B., Asztely, F., Hanse, E., and Wigstrom, H. (1989). Onset characteristics of long-term potentiation in the guinea-pig hippocampal CA1 region in vitro. *Eur. J. Neurosci.* 1, 382–394. doi: 10.1111/j.1460-9568.1989.tb00803.x
- Haas, J. S., Nowotny, T., and Abarbanel, H. D. (2006). Spike-timing-dependent plasticity of inhibitory synapses in the entorhinal cortex. *J. Neurophysiol.* 96, 3305–3313. doi: 10.1152/jn.00551.2006
- Heil, P. (2004). First-spike latency of auditory neurons revisited. *Curr. Opin. Neurobiol.* 14, 461–467. doi: 10.1016/j.conb.2004.07.002
- Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Trans. Neural Netw.* 14, 1569–1572. doi: 10.1109/TNN.2003.820440
- Izhikevich, E. M. (2006). Polychronization: computation with spikes. *Neural Comput.* 18, 245–282. doi: 10.1162/089976606775093882
- Kaminski, J., and Rutishauser, U. (2020). Between persistently active and activity-silent frameworks: novel vistas on the cellular basis of working memory. *Ann. N. Y. Acad. Sci.* 1464, 64–75. doi: 10.1111/nyas.14213
- Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., and Hudspeth, A. J. (2012). *Principles of Neural Science*. New York, NY: McGraw-hill.
- Lameu, E. L., Macau, E. E. N., Borges, F. S., Iarosz, K. C., Caldas, I. L., Borges, R. R., et al. (2018). Alterations in brain connectivity due to plasticity and synaptic delay. *Eur. Phys. J. Special Topics* 227, 673–682. doi: 10.1140/epjst/e2018-00090-6
- Lundqvist, M., Herman, P., and Miller, E. K. (2018). Working memory: delay activity, yes! persistent activity? maybe not. *J. Neurosci.* 38, 7013–7019. doi: 10.1523/JNEUROSCI.2485-17.2018
- Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., and Miller, E. K. (2016). Gamma and beta bursts underlie working memory. *Neuron* 90, 152–164. doi: 10.1016/j.neuron.2016.02.028
- Markram, H., Gerstner, W., and Sjöström, P. J. (2012). Spike-timing-dependent plasticity: a comprehensive overview. *Front. Synaptic Neurosci.* 4:2. doi: 10.3389/fnsyn.2012.00002
- Markram, H., Lubke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* 275, 213–215. doi: 10.1126/science.275.5297.213
- Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nat. Rev. Neurosci.* 5, 793–807. doi: 10.1038/nrn1519
- Masse, N. Y., Hodnefeld, J. M., and Freedman, D. J. (2017). Mnemonic encoding and cortical organization in parietal and prefrontal cortices. *J. Neurosci.* 37, 6098–6112. doi: 10.1523/JNEUROSCI.3903-16.2017
- Mi, Y., Katkov, M., and Tsodyks, M. (2017). Synaptic correlates of working memory capacity. *Neuron* 93, 323–330. doi: 10.1016/j.neuron.2016.12.004
- Miller, R. (1975). Distribution and properties of commissural and other neurons in cat sensorimotor cortex. *J. Comp. Neurol.* 164, 361–373. doi: 10.1002/cne.901640307
- Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. *Science* 319, 1543–1546. doi: 10.1126/science.1150769
- Morrison, A., Aertsen, A., and Diesmann, M. (2007). Spike-timing-dependent plasticity in balanced random networks. *Neural Comput.* 19, 1437–1467. doi: 10.1162/neco.2007.19.6.1437
- Nakamura, K., and Kubota, K. (1995). Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task. *J. Neurophysiol.* 74, 162–178. doi: 10.1152/jn.1995.74.1.162
- Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *J. Exp. Psychol.* 28, 411–421. doi: 10.1037/0278-7393.28.3.411
- Popovich, O. V., and Tass, P. A. (2012). Desynchronizing electrical and sensory coordinated reset neuromodulation. *Front. Hum. Neurosci.* 6:58. doi: 10.3389/fnhum.2012.00058
- Popovich, O. V., Yanchuk, S., and Tass, P. A. (2013). Self-organized noise resistance of oscillatory neural networks with spike timing-dependent plasticity. *Sci. Rep.* 3:2926. doi: 10.1038/srep02926
- Quintana, J., Fuster, J. M., and Yajeya, J. (1989). Effects of cooling parietal cortex on prefrontal units in delay tasks. *Brain Res.* 503, 100–110. doi: 10.1016/0006-8993(89)91709-5
- Roberts, W. A., and Kraemer, P. J. (1982). Some observations of the effects of intertrial interval and delay on delayed matching to sample in pigeons. *J. Exp. Psychol.* 8, 342–353. doi: 10.1037/0097-7403.8.4.342
- Robinson, B. S., Berger, T. W., and Song, D. (2016). Identification of stable spike-timing-dependent plasticity from spiking activity with generalized multilinear modeling. *Neural Comput.* 28, 2320–2351. doi: 10.1162/NECO_a_00883
- Song, D., Chan, R. H., Robinson, B. S., Marmorelis, V. Z., Opris, I., Hampson, R. E., et al. (2015). Identification of functional synaptic plasticity from spiking activities using nonlinear dynamical modeling. *J. Neurosci. Methods* 244, 123–135. doi: 10.1016/j.jneumeth.2014.09.023
- Song, S., Miller, K. D., and Abbott, L. F. (2000). Competitive Hebbian Learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.* 3, 919–926. doi: 10.1038/78829
- Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* 19, 394–405. doi: 10.1016/j.tics.2015.05.004
- Swadlow, H. A. (1990). Efferent neurons and suspected interneurons in S-1 forelimb representation of the awake rabbit: receptive fields and axonal properties. *J. Neurophysiol.* 63, 1477–1498. doi: 10.1152/jn.1990.63.6.1477

- Talathi, S. S., Hwang, D. U., and Ditto, W. L. (2008). Spike timing dependent plasticity promotes synchrony of inhibitory networks in the presence of heterogeneity. *J. Comput. Neurosci.* 25, 262–281. doi: 10.1007/s10827-008-0077-7
- Watanabe, K., and Funahashi, S. (2007). Prefrontal delay-period activity reflects the decision process of a saccade direction during a free-choice ODR task. *Cereb. Cortex* 17(Suppl. 1), i88–i100. doi: 10.1093/cercor/bhm102
- Watanabe, K., and Funahashi, S. (2014). Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nat. Neurosci.* 17, 601–611. doi: 10.1038/nn.3667
- Wei, H., and Du, Y. F. (2019). A temporal signal-processing circuit based on spiking neuron and synaptic learning. *Front. Comput. Neurosci.* 13:41. doi: 10.3389/fncom.2019.00041
- Wimmer, K., Nykamp, D. Q., Constantinidis, C., and Compte, A. (2014). Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.* 17, 431–439. doi: 10.1038/nn.3645

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Huang and Wei. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.