

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/333848992>

# Short-term synaptic plasticity expands the operational range of long-term synaptic changes in neural networks

Article in *Neural Networks* · June 2019

DOI: 10.1016/j.neunet.2019.06.002

CITATION

1

READS

272

4 authors:



**Guanxiong Zeng**

Chinese Academy of Sciences

4 PUBLICATIONS 27 CITATIONS

[SEE PROFILE](#)



**Xuhui Huang**

Chinese Academy of Sciences

20 PUBLICATIONS 139 CITATIONS

[SEE PROFILE](#)



**Tianzi Jiang**

Chinese Academy of Sciences

627 PUBLICATIONS 20,409 CITATIONS

[SEE PROFILE](#)



**Shan Yu**

National Institutes of Health

56 PUBLICATIONS 1,393 CITATIONS

[SEE PROFILE](#)

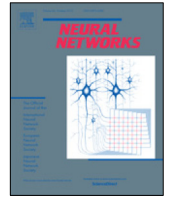
Some of the authors of this publication are also working on these related projects:



Neuro Perfusion [View project](#)



Imaging Genetics [View project](#)



# Short-term synaptic plasticity expands the operational range of long-term synaptic changes in neural networks

Guanxiong Zeng<sup>a,d,1</sup>, Xuhui Huang<sup>a,b,1,\*</sup>, Tianzi Jiang<sup>a,c,d</sup>, Shan Yu<sup>a,c,d,\*\*</sup>

<sup>a</sup> Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, 100190 Beijing, China

<sup>b</sup> Research Center for Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, 100190 Beijing, China

<sup>c</sup> CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, 100190 Beijing, China

<sup>d</sup> University of Chinese Academy of Sciences, 100049 Beijing, China

## ARTICLE INFO

### Article history:

Received 14 October 2018

Received in revised form 23 April 2019

Accepted 3 June 2019

Available online 18 June 2019

### Keywords:

Reservoir computing

Sequence learning and retrieval

Short-term depression

Synaptic heterogeneity

Self-organized criticality

Optimal information processing

## ABSTRACT

The brain is highly plastic, with synaptic weights changing across a wide range of time scales, from hundreds of milliseconds to days. Changes occurring at different temporal scales are believed to serve different purposes, with long-term changes for learning and memory and short-term changes for adaptation and synaptic computation. By studying the performance of reservoir computing (RC) models in a memory task, we revealed that short-term synaptic plasticity is fundamentally important for long-term synaptic changes in neural networks. Specifically, short-term depression (STD) greatly expands the operational range of a neural network in which it can accommodate long-term synaptic changes while maintaining system performance. This is achieved by dynamically adjusting neural networks close to a critical state. The effects of STD can be further strengthened by synaptic weight heterogeneity, resulting in networks that can tolerate very large, long-term changes in synaptic weights. Our results highlight a potential mechanism used by the brain to organize plasticity at different time scales, thereby maintaining optimal information processing while allowing internal structural changes necessary for learning and memory.

© 2019 Elsevier Ltd. All rights reserved.

## 1. Introduction

Plasticity is the ability of the brain to modify its internal structure according to past experiences and/or feedback signals from the environment. Specifically, synaptic plasticity refers to changes in synaptic strength between two neurons, which modulate the efficacy of inter-neuronal communication. There are two major types of synaptic plasticity, i.e., long-term and short-term. Long-term synaptic plasticity (LTP), which lasts from tens of minutes to days or even longer, is believed to serve as the physiological mechanism for memory storage, whereas short-term synaptic plasticity (STP) refers to temporary changes in synaptic strength lasting from hundreds of milliseconds to minutes (Zucker & Regehr, 2002).

There are two forms of STP, which modulate synaptic strength in opposite directions after presynaptic neuron firing: i.e., short-term depression (STD), which reduces synaptic strength due to depletion of neurotransmitters, and short-term facilitation (STF), which enhances synaptic strength due to the increased release probability of neurotransmitters associated with higher intracellular calcium concentration (Abbott & Regehr, 2004; Tsodyks & Wu, 2013; Zucker & Regehr, 2002).

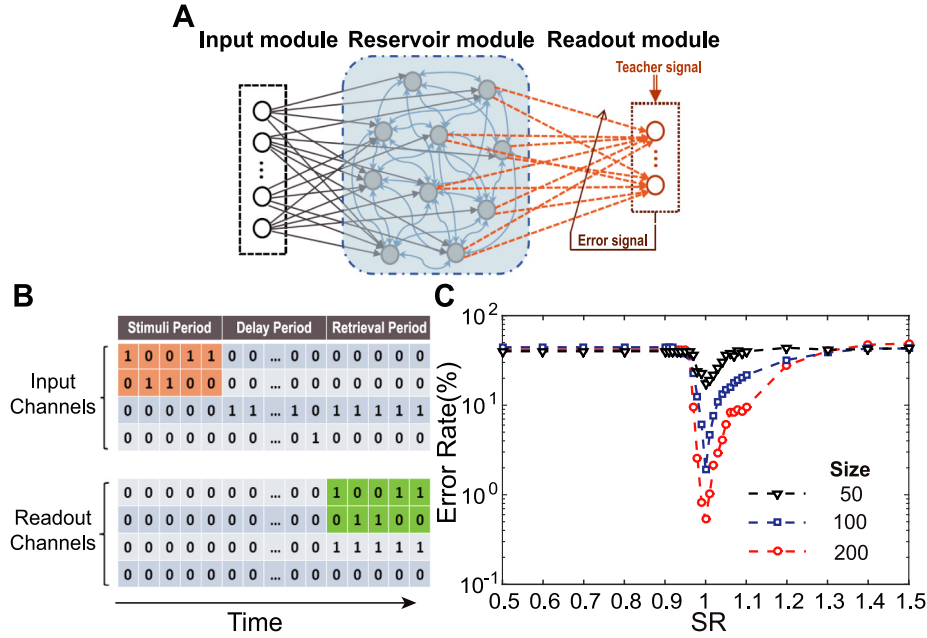
STP plays important roles in synaptic computations and neural functions (Abbott & Regehr, 2004; Klug et al., 2012; Rotman, Deng, & Klyachko, 2011). It can serve as a mechanism for temporal filtering (Abbott & Regehr, 2004; Fortune & Rose, 2001; Rosenbaum, Rubin, & Doiron, 2012; Rotman et al., 2011) and gain control (Abbott, Varela, Sen, & Nelson, 1997), as well as a transient memory buffer (Maass & Markram, 2002), regulator of network memory capacity (Bibitchkov, Herrmann, & Geisel, 2002; Mejias, Hernandezgomez, & Torres, 2012; Mejias, Kappen, Longtin, & Torres, 2013; Mejias & Torres, 2009; Torres, Pantic, & Kappen, 2002), and facilitator of persistent activity or slowing diminishing activity to support working memory (Barak & Tsodyks, 2007; Mi, Li, Wang, & Wu, 2014; Mongillo, Barak, & Tsodyks, 2008; Seeholzer, Deger, & Gerstner, 2018). Furthermore, STP is instrumental for performing probabilistic inference and pattern completion (Leng et al., 2018).

\* Corresponding author at: Research Center for Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, 100190 Beijing, China.

\*\* Corresponding author at: Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, 100190 Beijing, China.

E-mail addresses: [xuhui.huang@ia.ac.cn](mailto:xuhui.huang@ia.ac.cn) (X. Huang), [shan.yu@nlpr.ia.ac.cn](mailto:shan.yu@nlpr.ia.ac.cn) (S. Yu).

<sup>1</sup> These two authors contributed equally to this work.



**Fig. 1.** RC model and its optimal working conditions. **(A)** Diagram of a typical RC model, including input, reservoir, and readout modules. Training is only applied to connections between reservoir and readout. **(B)** Schematic of 5-bit memory task. There are four channels for both input and readout modules. The first two channels are used for loading (orange shadow) and retrieving (green shadow) stimulus patterns in the input and readout modules, respectively. The third channel of input represents a background signal, and the fourth channel is used to cue the start of retrieval. **(C)** Mean error rate percentage (in logarithmic scale) in 5-bit memory task of the RC model with differently sized reservoirs (color coded) plotted as a function of reservoir spectral radius (SR). Mean error rate for each case was calculated by bit errors in the first two readout channels during retrieval in 20 independent trials. Input module was fully connected to the reservoir with weight values randomly sampled from  $-0.5$  to  $0.5$ , with scaling factors for the four input channels set as  $0.02[1 \ 1 \ 0.01 \ 10]$ . Connection probability in the reservoir  $P = 0.03$ ; step duration for update:  $\Delta T = 1$  ms; delay duration for readout:  $T_{\text{delay}} = 150$  ms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

However, the functional roles of LTP and STP appear contradictory. Specifically, LTP changes synaptic strengths to target levels for long-term memory storage, whereas STP constantly changes synaptic strengths away from their target levels according to ongoing activities. Thus far, however, the relationship between these two forms of plasticity and how the brain reconciles them to support various functions remain unclear.

In the present study, we used a reservoir computing (RC) model to address the above question. In the RC model, as shown in Fig. 1A, low dimensional input, usually spanning the temporal domain, is fed into a relatively large recurrent neural network (RNN), called the reservoir, to project the input to high-dimensional space and harvest the capacity of such a high-dimensional, dynamic system in representing information (Jaeger, 2001; Jaeger & Haas, 2004; Lukoševičius & Jaeger, 2009; Lukoševičius, Jaeger, & Schrauwen, 2012; Maass, Natschlager, & Markram, 2002). The reservoir is then projected to a readout module, in which the output of the RC is obtained. In a supervised learning mode, the error between actual output and desired output is used to adjust the connections between the reservoir and readout module using various learning rules. Usually, the reservoir is randomly connected, and its internal weights are fixed (Jaeger, 2001; Jaeger & Haas, 2004; Maass et al., 2002). The RC model is a powerful and versatile tool for tasks such as temporal pattern classification, recognition, prediction, and action sequence control (Lukoševičius & Jaeger, 2009; Lukoševičius et al., 2012), providing an ideal framework to examine how synaptic changes occurring at different time scales can work together to support various network functions.

By using the RC network model to perform information processing tasks, we found that STP can improve the performance of a neural network in terms of robustness to long-term structural changes, e.g., long-term changes in synaptic strength. Specifically, due to the existence of an optimal state for a network to process

information, long-term changes in synaptic strength will likely drive the network away from the optimal state and are, thereby, detrimental to its function. By adaptively modulating the network close to its optimal state, STP stabilizes the system and allows LTP to modify synaptic strength in a much wider range for long-term information storage.

## 2. Models and methods

To examine the factors influencing the performance of the RC model and how they are related to synaptic plasticities at different time scales, two network models were used, including (1) a classic, non-spiking network model, i.e., Echo State Network (ESN) (Jaeger, 2001, 2012; Jaeger & Haas, 2004) and (2) a spiking neural network (SNN) model (Levina, Herrmann, & Geisel, 2007). Both were tested on the same memory task. Below we provide detailed information on the models and the task.

### 2.1. Echo state network model

As outlined above for RC models in general, the ESN contains an input module, randomly connected reservoir with leaky-integrator neurons, and an output/readout module. The neurons in the reservoir follow the equation (Jaeger, 2001, 2012; Jaeger & Haas, 2004):

$$\mathbf{x}(k+1) = (1 - \gamma)\mathbf{x}(k) + \gamma \tanh(\mathbf{W}\mathbf{x}(k) + \mathbf{W}^{\text{in}}\mathbf{u}(k+1)) \quad (1)$$

where  $\mathbf{x}(k)$  is the state vector of the reservoir (with network size  $N$ ) at time  $k$ ,  $\gamma$  is the leaky rate,  $\mathbf{W}$  is the connection weights of the reservoir,  $\mathbf{u}(k)$  is the input vector with dimension  $L$ ,  $\mathbf{W}^{\text{in}}$  is the  $N \times L$  sized weight matrix of connections from the input to the reservoir, and  $\tanh$  is the activation function for each neuron.

## 2.2. Spiking neural network model with STD

To study the functional roles of STP, we used a SNN with STD (Levina et al., 2007) as the reservoir in the RC model. In this model, the dynamics of the membrane potential  $h_i(t)$  of individual neurons can be described by the following equation:

$$\frac{dh_i(t)}{dt} = \frac{1}{N} \sum_{j=1}^N U_{ij} \delta(t - t_j^{\text{sp}} - \tau_D) + \delta_{i,\xi_\tau(t)} I^{\text{bg}} + I_i^{\text{sti}}(t) \quad (2)$$

where the three items on the right side of the equation are the interaction within the network, background noise, and inputs to the network, respectively. In the first item,  $J_{ij}(t)$  is the amount of available neurotransmitters from the synapses connecting neuron  $j$  to neuron  $i$ ,  $U$  is the fraction of neurotransmitter available for release when there is a spike, and  $UJ(t)$  is the amount of released neurotransmitters after a spike.  $\delta(t - t_j^{\text{sp}} - \tau_D)$  describes the spike of neuron  $j$  at  $t_j^{\text{sp}}$  transmitting to a postsynaptic neuron with a delay of  $\tau_D$ . The background noise is loaded to a randomly selected neuron  $\xi_\tau(t)$  with a certain rate  $\tau$ . If  $h_i(t) > \theta$ ,  $h_i(t)$  is reset to  $h_i(t) - \theta$  ( $\theta = 1$ ), indicating a spike generated from neuron  $i$ .

The variable  $J_{ij}(t)$  subjected to STD can be described by the following equation (Levina et al., 2007; Levina, Herrmann, & Geisel, 2009; Mongillo et al., 2008):

$$\frac{dJ_{ij}(t)}{dt} = \frac{1}{\tau_j} \left( \frac{J_{\max}}{U} - J_{ij}(t) \right) - U_{ij}(t) \delta(t - t_j^{\text{sp}}) \quad (3)$$

where  $\tau_j$  is the time constant of the synaptic depression. The value of  $J_{ij}(t)$  is controlled by the weight parameter  $J_{\max}$ , which determines the maximal synaptic weight.

## 2.3. Bit memory task and related learning method

A classic bit memory task was chosen to test the performance of both models (Jaeger, 2012), in which the network needs to remember binary sequences with length equals  $M$  (for details see Fig. 1B). The number of training sequences  $N^{\text{train}}$  for the 5-bit task was 32. Next we explain the task, as well as the training and testing of the network, using the ESN as an example. During the training stage, all 32 5-bit sequences  $\mathbf{u}$  were sequentially fed into the network, the  $M$ -dimensional output of the readout module  $\mathbf{y}$  was generated sequentially by  $\mathbf{y}(k) = \mathbf{W}^{\text{out}} \mathbf{x}(k)$ . Here,  $\mathbf{W}^{\text{out}}$  is the connection matrix between the reservoir and output module, and is adjusted according to the following regression method:

$$\mathbf{W}^{\text{out}} = ((\mathbf{S}^T)^{\dagger} \mathbf{T}^T)^T \quad (4)$$

where  $\mathbf{S}$  is the state collection matrix of all  $\mathbf{x}$  responding to all input sequences  $\mathbf{u}$ , and its size is  $N \times (T^{\text{out}} N^{\text{train}})$ ;  $T^{\text{out}}$  is the length of each training sequence ( $T^{\text{out}} = 2M + T_{\text{delay}}$ ). Similarly,  $\mathbf{T}$  is the target collection matrix of all  $\mathbf{y}$  corresponding to all  $\mathbf{x}$ , and its size is  $M \times (T^{\text{out}} N^{\text{train}})$ .  $^T$  denotes matrix transpose and  $^{\dagger}$  denotes the pseudo-inverse operation. After training, the 32 sequences were presented again, and network performance was quantified by the percentage error based on each bit (whether each bit is incorrect) in corresponding output sequences (for details see Jaeger, 2012). For the spiking neural network model, the duration of each time step for update was set to  $\Delta T = 0.2$  ms, and the duration of each input bit was set to five times the time-step length ( $\tau = 1$  ms). The membrane potentials were fed into the readout module for retrieval.

To examine the robustness of our results, we tested the performance of these models (ESN and SNN with or without STD) on the memory task in a variety of situations, including more difficult tasks with increased sequence length, different scaling factors for inputs, different regression methods for training,

different connection probabilities, and different methods for testing performance, e.g., varied thresholds for binarizing outputs (for ESN) and mean absolute error averaged across all bits (for both ESN and SNN). All results were highly consistent. For simplicity, only typical results are shown in this paper.

## 3. Results

### 3.1. Trade-off between performance and robustness in RC model

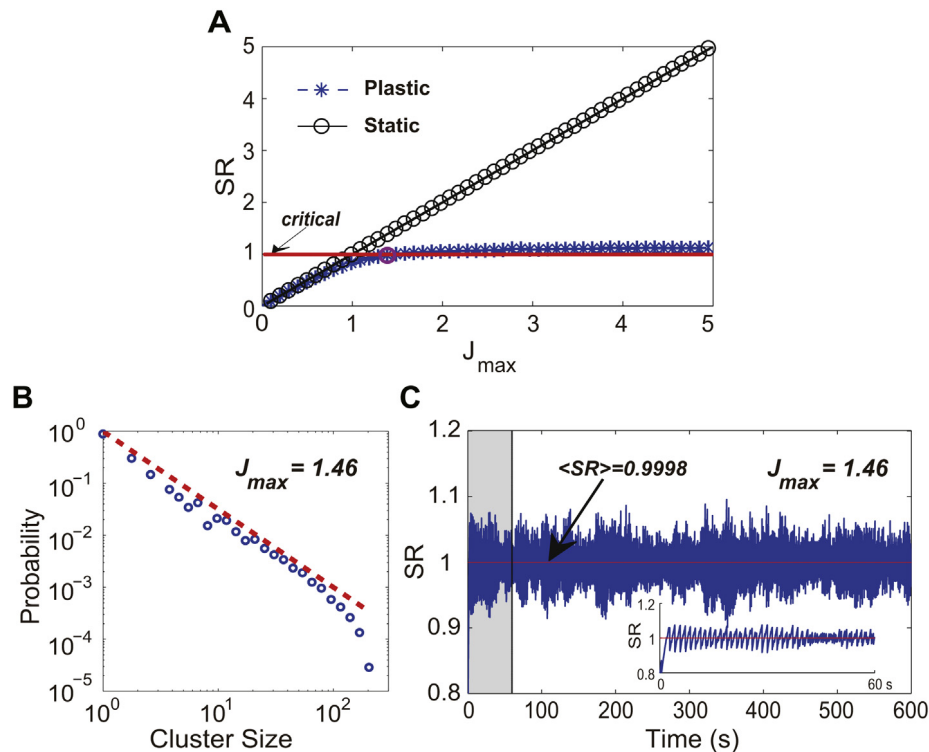
RC model's performance is significantly affected by the structural characteristics of the reservoir network, i.e., spectral radius (SR), which is defined as the largest eigenvalue of the connection matrix  $\mathbf{W}$ . Although the relationship between the SR and network behavior depends on various factors, including the task at hand (Caluwaerts, Wyffels, Dieleman, & Schrauwen, 2013; Jaeger, 2012), and feature inputs (Caluwaerts et al., 2013; Manjunath & Jaeger, 2013; Mayer, 2015; Yildiz, Jaeger, & Kiebel, 2012) (see discussion for details), previous studies have suggested that the value of SR itself can usually insert strong influence by controlling reverberating activity, epidemic threshold, or synchronization within a recurrent network (de Haan et al., 2012; Restrepo, Ott, & Hunt, 2005; Van Mieghem, Omic, & Kooij, 2009). Specifically, in the reservoir network, a small SR can lead to rapid fading of activities and a large SR tends to result in an explosion of activities in the network, both of which can wash out memory trace in the system (Jaeger, 2001, 2012; Jaeger & Haas, 2004).

Here we systematically varied the SR and examined ESN performance in a memory task (see Models and Methods for details). Similar to previously reported results (Jaeger, 2012), we found that the performance of the model for this task peaked at  $\text{SR} = 1$  (Fig. 1C), suggesting that, under such a condition, the memory of the input sequence can be best held and retrieved. Importantly, we found that increasing the size of the reservoir not only improved peak performance, but also increased the sensitivity of the system towards small perturbation in the network structure, as reflected by the abrupt changes in performance in response to small deviations from  $\text{SR} = 1$  (Fig. 1C). This suggests that, in larger systems, a small deviation from the optimal SR can aggregate more easily and give rise to more pronounced effects. On the one hand, these results highlight the vital role of identifying the optimal network structure for enhancing system capability in information processing, as widely appreciated in previous studies (Kinouchi & Copelli, 2006; Larremore, Shew, Ott, & Restrepo, 2011a; Larremore, Shew, & Restrepo, 2011b). On the other hand, our results raised an important yet poorly understood question: with a sufficiently large reservoir, which is necessary for representing complex information, the sensitivity of the system may be too high, such that necessary long-term synaptic changes for memory storage, for example, will cause deviation from the optimal SR and lead to significant impairment in a network's ability to hold information. Thus, the results in Fig. 1C demonstrate the importance of finding a solution for this performance–robustness dilemma, that is, to allow a wide range of long-term synaptic changes in a neural network while stably maintaining the system's peak capability for information processing.

### 3.2. STD Maintains optimal SR under long-term synaptic changes

In the brain, similar dependency between the neural network's state and its information processing capacity is well documented. Both empirical and theoretical studies have suggested that brain networks operate close to their optimal (critical) state to achieve functional benefits (Beggs & Timme, 2012; Shew & Plenz, 2013; Shew, Yang, Petermann, Roy, & Plenz, 2009; Shew, Yang, Yu, Roy,





**Fig. 2.** Self-organized criticality (SOC) in reservoirs induced by STD. **(A)** SR (for static model) or mean SR (for plastic model; calculated by averaging instantaneous SR estimated in 600 seconds) plotted as functions of maximal synaptic weight,  $J_{\max}$ , for reservoirs with (blue) and without (black) STD. Effective SR was calculated as the largest eigenvalue of the instantaneous connection weight matrix ( $UJ(t)/N$ ). Red line:  $SR = 1$ . **(B)** Cluster size distribution (in double-logarithmic scale) is shown for the system with STD when its SR is closest to 1 ( $J_{\max} = 1.46$ , marked with a purple circle in A). Total simulation time of the case is 1800 s. Red dashed line: a power law with exponent of  $-1.5$ . **(C)** For the same system as shown in (B), the dynamic regulation of SR is shown. Red line: mean SR during 600 s. Inset: zoom-in view of the initial 60 s (shaded area in the main figure). Note that SR rapidly approached 1 at the beginning of simulation. Network size  $N = 300$ . Similar results were found in larger networks with  $N$  up to 700. The parameters used for the model were the same as that in [Levina et al. \(2007\)](#). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

& Plenz, 2011; Yang, Shew, Roy, & Plenz, 2012). Moreover, STD has been proposed as an effective way to adaptively maintain the critical state via self-organized criticality (SOC) ([Levina et al., 2007, 2009](#)). Inspired by such a mechanism, we examined if SOC can shed new light on how the brain can reconcile STP and LTP to solve the performance–robustness dilemma demonstrated above.

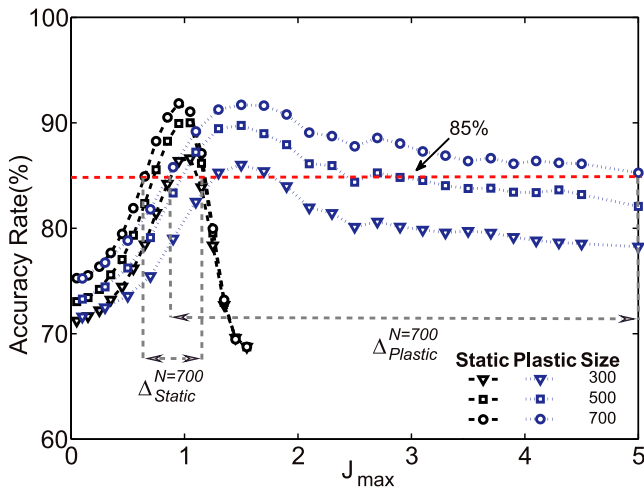
To this end, we first investigated how STD in a SNN-based RC model (see Models and Methods for details) can affect SR under changes in long-term synaptic strength. We found that, with the help of STD, the SR could be maintained close to 1, i.e., sustaining the network in an optimal state, across a relatively wide range of maximal synaptic weights ( $J_{\max} > 1.46$ ; [Fig. 2A](#), blue; “Plastic”). In contrast, in the network with static synaptic weights, i.e., without STD, the SR changed linearly with  $J_{\max}$ , resulting in a very narrow range of synaptic strength to maintain optimal SR ([Fig. 2A](#), black, “Static”). For the RC model with STD, when the SR was closest to 1 (corresponding to  $J_{\max} = 1.46$ ), we found that the dynamics of the reservoir exhibited behavior close to the critical state, including the size distribution of activity cascades (“neuronal avalanches”) following a power-law with the exponent close to  $-1.5$  ([Fig. 2B](#)) ([Beggs & Plenz, 2003, 2004](#); [Levina et al., 2007](#)). Further examination revealed how STD dynamically adjusted the network structure to maintain the SR close to 1 ([Fig. 2C](#)). Together, these results demonstrated that to maintain the SR close to 1 by STD in this model is indeed a SOC process. Thus, the STD-induced SOC scheme may provide an attractive solution to maintain optimal performance of neural networks while accommodating long-term synaptic changes for information storage.

### 3.3. STD-induced SOC can solve performance–robustness dilemma

To examine whether STD-induced SOC can increase the capacity of a neural network to incorporate long-term synaptic changes, we quantified network performance in the same memory task, with synaptic strength changing homogeneously within the network. We found that the sensitivity of the system to parameter perturbation was greatly attenuated without sacrificing performance ([Fig. 3](#), blue), leading to a much higher capacity for the system to accommodate long-term synaptic changes. Importantly, this effect became even more pronounced with the increase in system size. In contrast, the static RC model, i.e., without the STD mechanism, only worked well within a narrow range of synaptic weights ([Fig. 3](#), black).

To confirm the robustness of our results under different conditions and parameter settings, we tested system performance in a variety of situations, including different (1) connection probabilities of reservoirs, (2) difficulty levels of tasks, (3) scaling factors for inputs, (4) regression methods for training, and (5) methods for evaluating performance. All results were highly consistent across these various conditions. For example, the RC models with different connection probabilities ( $P = 0.8, 0.9$  and  $1.0$ ) exhibited very similar behavior of expanding the operational range with STD ([Fig. 4A](#)); similar results were also obtained by the RC models with larger reservoirs (up to 1000 neurons) in more difficult tasks (up to four times more sequences to memorize) ([Fig. 4B](#)).

To explore whether STD-induced SOC could work under various synaptic weight changes, we examined system performance on the same task with synaptic strengths changed heterogeneously in the network. Interestingly, even slight heterogeneity

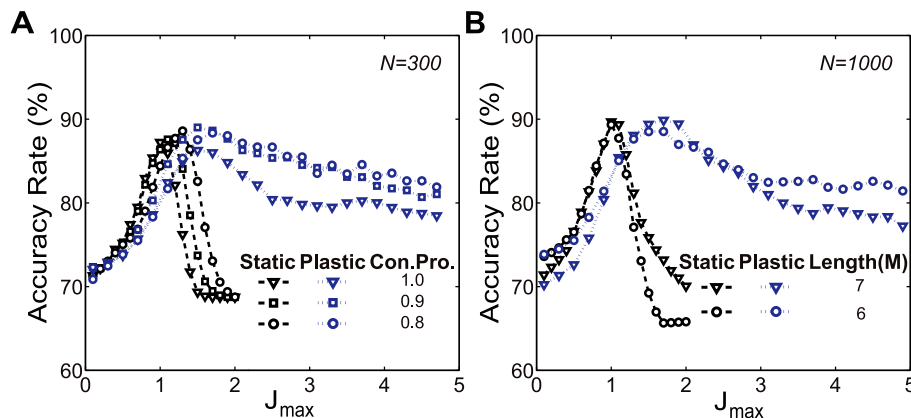


**Fig. 3.** Expansion of operational range for RC model achieved by STD-induced SOC. Mean accuracy rate percentages of RC model with (blue) and without (black) STD-induced SOC plotted as a function of maximal synaptic weight  $J_{\max}$  within the reservoir. For each model type, three reservoir sizes were used. Red dashed line: 85% accuracy. Gray dashed line windows: Operational range within which the model's performance was at least 85% correct for the largest reservoir size ( $N = 700$ ). Mean error rate for each case was calculated by bit errors in the first two readout channels during retrieval in 20 independent trials. Input module was fully connected to reservoir with weight values randomly sampled from 0 to 1, with scaling factors for the four input channels set as 0.14 [1 1 0.45 4]. The reservoir was fully connected. Step duration for update:  $\Delta T = 0.2$  ms; delay duration for readout:  $T_{\text{delay}} = 20$  ms; update interval of bit input  $\tau = 1$  ms; the time constant of STD:  $\tau_j = 2\tau N$ ; the transmission delay:  $\tau_D = \Delta T$ ; the releasable fraction of neurotransmitter:  $U = 0.2$ . We added a leak term to the membrane potential ( $h_i(t)$ ) with a time constant of 30 ms when taking the task. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

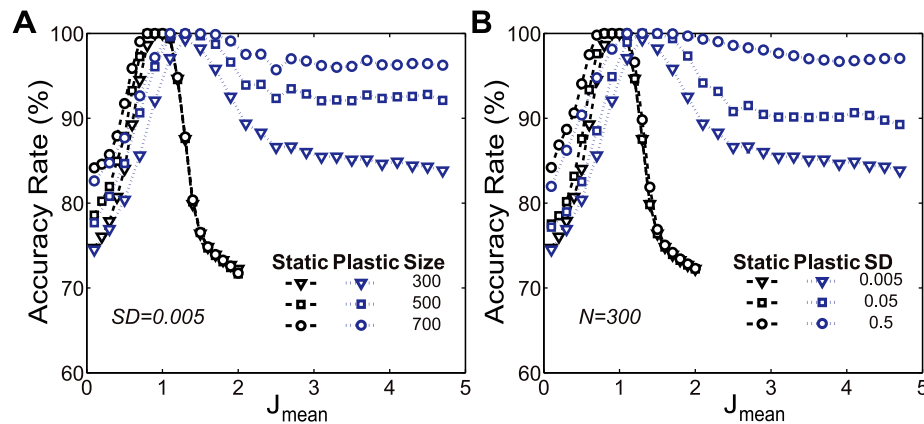
in synaptic weights significantly increased network performance with and without STD (Fig. 5A), likely due to increased representation capacity associated with heterogeneous weights. Similar behaviors have been reported previously (Ju, Xu, Chong, & Van-Dongen, 2013). Importantly, the effects of STD on enlarging the operational range of the network were also observed in networks with heterogeneous synaptic weights (Fig. 5A). In addition, due to enhanced performance, we found that heterogeneity in synaptic weight combined with STD enlarged the operational range even more, as the higher degree of heterogeneity led to a wider operational range with the same STD (Fig. 5B). This provides a potential account for the experimental findings that both STD and a high degree of heterogeneity in synaptic weights co-exist in the brain.

#### 4. Discussion and conclusion

Our results provide important clues on how neural networks can maintain desirable performance in information processing when modifications to their structure are necessary. The optimal state for information processing of RC models is affected by various factors, including network structure (SR), input strength (Caluwaerts et al., 2013; Manjunath & Jaeger, 2013; Mayer, 2015; Yildiz et al., 2012), and the task at hand (Caluwaerts et al., 2013; Jaeger, 2012). Generally, SR is a useful indicator for the optimal network state and, as shown in many previous studies (Jaeger, 2001, 2012; Jaeger & Haas, 2004; Larremore et al., 2011a; Larremore, Shew, Ott, Sorrentino, & Restrepo, 2014; Park, Lee, Kang, Kim, & Kim, 2017; Pei et al., 2012; Rodan & Tino, 2011; Virkar, Shew, Restrepo, & Ott, 2016) the SR slightly less than 1 can be an informative predictor of RC model performance (but it is not a sufficient or necessary condition, see Deng & Zhang, 2007). Furthermore, the SR close to 1 has been linked to the so-called critical state or “edge of chaos” concept (Jaeger & Haas, 2004; Larremore, Carpenter, Ott, & Restrepo, 2012; Larremore et al., 2011a, 2014, 2011b; Lukoševičius & Jaeger, 2009; Lukoševičius et al., 2012), which is long-suggested to be a condition with functional advantages in information processing (Jaeger, 2012; Kinouchi & Copelli, 2006; Langton, 1990; Legenstein & Maass, 2007; Lukoševičius et al., 2012). However, as shown here (Fig. 1C) the sensitivity of network performance to deviation from the optimal (critical) state increased rapidly for larger networks, thus intensifying the difficulty of maintaining the state precisely on this “edge of chaos”. Although in principle the structure of a network can be entirely fixed to avoid structural perturbation, it will deprive the system of important functional benefits associated with synapses that can change their strengths according to either past experiences or feedback signals from the environment. This ability to induce long-term modifications in synaptic strength is the basis for learning and storage of information in neural networks (Caporale & Dan, 2008; Malenka & Bear, 2004). Our results showed that SOC, mediated by a simple form of STP, is very effective at solving the dilemma between maintaining the critical state for optimal information processing and accommodating synaptic changes necessary for long-term information storage. The SOC scheme opens the door to harvesting the functional advantages of long-term synaptic changes while retaining an overall desirable state for processing. It should be noted that various types of synaptic plasticity, such as STD or STD combined with STF, have been demonstrated as possible ways to achieve SOC within certain parameter space (Levina et al., 2007, 2009; Meisel & Gross, 2009; Millman, Mihalas, Kirkwood, & Niebur, 2010), providing



**Fig. 4.** Robustness of expanding operational range by STD in RC models with different connection probabilities of reservoirs (A) and more difficult tasks with longer sequences to memorize (B). Other parameters were the same as in Fig. 3.



**Fig. 5.** STD and synaptic weight heterogeneity together expand the operational range of long-term synaptic changes. **(A)** STD scheme also works for networks with heterogeneous synaptic weights. Performance of network in the same memory task plotted as a function of mean maximal synaptic weights,  $J_{\text{mean}}$ , for static (black) and plastic (blue) systems with various sizes. For both systems, the synaptic weights showed Gaussian distribution with a small degree of heterogeneity ( $J_{\text{mean}} + \xi$ ,  $\xi$ : Gaussian stochastic variable with zero mean and standard deviation SD). **(B)** Effects of STD are more pronounced with larger degrees of synaptic heterogeneity. Performance of network ( $N = 300$ ) plotted for different degrees of synaptic heterogeneity of  $J_{\text{mean}}$  in both static and plastic conditions. Other parameters are as in Fig. 3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

additional options to combine SOC with the benefits of a network that support long-term synaptic changes.

One of the key hypotheses of the current study was that biological neural networks are dynamically organized close to an optimal working state. Accumulating evidence has indicated that neural networks in the brain operate close to criticality (Beggs & Plenz, 2003, 2004; Beggs & Timme, 2012; Shew & Plenz, 2013; Yu, Yang, Shriki, & Plenz, 2013). Numerous studies have also demonstrated that, in such a state, neural networks entertain various functional advantages in terms of information storage, transmission, and processing (Beggs & Timme, 2012; Shew & Plenz, 2013; Shew et al., 2009, 2011; Yang et al., 2012). Moreover, although criticality in simulations is often obtained through fine-tuned control parameters, theoretical studies have suggested that a neural network can be constructed in such a way that criticality becomes an attractor in the system's dynamics (Levina et al., 2007, 2009; Meisel & Gross, 2009; Millman et al., 2010). It has been demonstrated that adaptation to sensory input can dynamically adjust the visual cortex in turtles to the critical state (Shew et al., 2015). Moreover, the critical state can be maintained in the cortex of nonhuman primates, even when network activity levels are increased significantly during motor and cognitive tasks (Yu et al., 2017). Such studies provide strong support for our hypothesis of dynamically modulated criticality in neural networks.

We also predicted a profound relationship between synaptic plasticity manifested at different temporal scales. Although STP seemingly operates in a temporal scale too short to be relevant for long-term information storage, previous studies have revealed that STP can affect network memory capacity, i.e., STD can decrease storage capacity (Bibitchkov et al., 2002; Torres et al., 2002) whereas STF can increase storage capacity (Mejias et al., 2012, 2013; Mejias & Torres, 2009). Our results further indicate that STP may also play a key role in maintaining optimal information processing in the face of long-term synaptic changes. However, this theoretical possibility requires further empirical studies in real brains. Interestingly, in an animal model of Down syndrome – a major intellectual disability – STP but not LTP was reportedly altered, and subjects with abnormal STP exhibit impairment in difficult spatial memory tasks (Witton et al., 2015). As memory is usually linked to LTP rather than STP, such a functional deficit seems paradoxical. Our results provide a potential account for such a phenomenon, suggesting that failure to dynamically maintain the network at the optimal working state,

rather than deficits in LTP itself, may lead to impairments in these animals. It would be informative for future studies to determine whether neuronal dynamics in such animals become suboptimal when synaptic weights are subjected to long-term changes during difficult memory tasks.

In addition to suggesting possible neural mechanisms by which the brain solves the problem of maintaining an optimal state while allowing structural changes, our findings also have potential implications for the design of artificial networks for machine learning tasks. In line with our results suggesting that the critical state is functionally beneficial, recent research has demonstrated that normalizing connection weights in a generative adversarial network (GAN) to achieve  $\text{SR} = 1$  could significantly improve its performance in generating various categories of images (Miyato, Kataoka, Koyama, & Yoshida, 2018). However, further study is needed to examine if STD-induced SOC can provide an effective way to adaptively maintain the SR close to 1 for recurrent neural networks in which connection weights are subject to learning.

In summary, inspired by the criticality in biological neural networks, especially the mechanism of using STP to ensure the critical state as a stable attractor (Levina et al., 2007, 2009), we implemented a STD-induced SOC scheme in the RC model, which automatically regulated the state of the RNN to be close to criticality. Michiels van Kessenich et al. also recently demonstrated that STP and LTP can coexist in a neural network, with STP organizing the network to criticality and LTP enabling the network to learn binary rules such as XOR (van Kessenich, Luković, De Arcangelis, & Herrmann, 2018). In the current study, we showed that STP not only coexists with LTP, but greatly expands the operational range of the system, allowing necessary long-term synaptic changes to be implemented without compromising network performance. These results reveal a basic, yet previously overlooked role of STP in neural networks, and shed new light on how plasticities of different temporal scales can be orchestrated to support brain function.

## Acknowledgments

This work was supported by the National Key Research and Development Program of China (2017YFA0105203), Natural Science Foundation of China (81471368, 11505283, 91732305, 31620103905), the Strategic Priority Research Program of the Chinese Academy of Sciences (CAS) (XDB32040200, XDB32030200), Hundred-Talent Program of CAS (for S.Y.), and Key Research Program of Frontier Sciences, CAS (QYZDJ-SSW-SMC019).



## References

- Abbott, L. F., & Regehr, W. G. (2004). Synaptic computation. *Nature*, 431(7010), 796–803.
- Abbott, L. F., Varela, J. A., Sen, K., & Nelson, S. B. (1997). Synaptic depression and cortical gain control. *Science*, 275(5297), 220–224.
- Barak, O., & Tsodyks, M. (2007). Persistent activity in neural networks with dynamic synapses. *PLoS Computational Biology*, 3(2), 323–332.
- Beggs, J. M., & Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *Journal of Neuroscience*, 23(35), 11167–11177.
- Beggs, J. M., & Plenz, D. (2004). Neuronal avalanches are diverse and precise activity patterns that are stable for many hours in cortical slice cultures. *Journal of Neuroscience*, 24(22), 5216–5229.
- Beggs, J. M., & Timme, N. (2012). Being critical of criticality in the brain. *Frontiers in Physiology*, 3, 14.
- Bibitchkov, D., Herrmann, J., & Geisel, T. (2002). Pattern storage and processing in attractor networks with short-time synaptic dynamics. *Network. Computation in Neural Systems*, 13(1), 115–129.
- Caluwaerts, K., Wyffels, F., Dieleman, S., & Schrauwen, B. (2013). The spectral radius remains a valid indicator of the echo state property for large reservoirs. In *The 2013 international joint conference on neural networks*, IJCNN (pp. 1–6). IEEE.
- Caporale, N., & Dan, Y. (2008). Spike timing-dependent plasticity: a Hebbian learning rule. *Annual Review of Neuroscience*, 31, 25–46.
- Deng, Z., & Zhang, Y. (2007). Collective behavior of a small-world recurrent neural system with scale-free distribution. *IEEE Transactions on Neural Networks*, 18(5), 1364–1375.
- Fortune, E. S., & Rose, G. J. (2001). Short-term synaptic plasticity as a temporal filter. *Trends in Neurosciences (Barking)*, 24(7), 381–385.
- de Haan, W., van der Flier, W. M., Wang, H., Van Mieghem, P. F., Scheltens, P., & Stam, C. J. (2012). Disruption of functional brain networks in Alzheimer's disease: what can we learn from graph spectral analysis of resting-state magnetoencephalography? *Brain Connectivity*, 2(2), 45–55.
- Jaeger, H. (2001). The “echo state” approach to analysing and training recurrent neural networks—with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148(34), 13.
- Jaeger, H. (2012). Long short-term memory in echo state networks: Details of a simulation study. Report, Jacobs University Bremen.
- Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667), 78–80.
- Ju, H., Xu, J. X., Chong, E., & VanDongen, A. M. J. (2013). Effects of synaptic connectivity on liquid state machine performance. *Neural Networks*, 38, 39–51.
- van Kessenich, L. M., Luković, M., De Arcangelis, L., & Herrmann, H. J. (2018). Critical neural networks with short-and long-term plasticity. *Physical Review E*, 97(3), 032312.
- Kinouchi, O., & Copelli, M. (2006). Optimal dynamical range of excitable networks at criticality. *Nature Physics*, 2(5), 348.
- Klug, A., Borst, J. G. G., Carlson, B. A., Kopp-Scheinplig, C., Klyachko, V. A., & Xu-Friedman, M. A. (2012). How do short-term changes at synapses fine-tune information processing? *Journal of Neuroscience*, 32(41), 14058–14063.
- Langton, C. G. (1990). Computation at the edge of chaos - phase-transitions and emergent computation. *Physica D*, 42(1–3), 12–37.
- Larremore, D. B., Carpenter, M. Y., Ott, E., & Restrepo, J. G. (2012). Statistical properties of avalanches in networks. *Physical Review E*, 85(6), 11.
- Larremore, D. B., Shew, W. L., Ott, E., & Restrepo, J. G. (2011a). Effects of network topology, transmission delays, and refractoriness on the response of coupled excitable systems to a stochastic stimulus. *Chaos. An Interdisciplinary Journal of Nonlinear Science*, 21(2), 025117.
- Larremore, D. B., Shew, W. L., Ott, E., Sorrentino, F., & Restrepo, J. G. (2014). Inhibition causes ceaseless dynamics in networks of excitable nodes. *Physical Review Letters*, 112(13), 138103.
- Larremore, D. B., Shew, W. L., & Restrepo, J. G. (2011b). Predicting criticality and dynamic range in complex networks: effects of topology. *Physical Review Letters*, 106(5), 058101.
- Legenstein, R., & Maass, W. (2007). Edge of chaos and prediction of computational performance for neural circuit models. *Neural Networks*, 20(3), 323–334.
- Leng, L., Martel, R., Breitwieser, O., Bytschok, I., Senn, W., Schemmel, J., et al. (2018). Spiking neurons with short-term synaptic plasticity form superior generative networks. *Scientific Reports*, 8(1), 10651.
- Levina, A., Herrmann, J. M., & Geisel, T. (2007). Dynamical synapses causing self-organized criticality in neural networks. *Nature Physics*, 3(12), 857–860.
- Levina, A., Herrmann, J. M., & Geisel, T. (2009). Phase transitions towards criticality in a neural system with adaptive interactions. *Physical Review Letters*, 102(11), 4.
- Lukoševičius, M., & Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3), 127–149.
- Lukoševičius, M., Jaeger, H., & Schrauwen, B. (2012). Reservoir computing trends. *KI-Künstliche Intelligenz*, 26(4), 365–371.
- Maass, W., & Markram, H. (2002). Synapses as dynamic memory buffers. *Neural Networks*, 15(2), 155–161.
- Maass, W., Natschläger, T., & Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(11), 2531–2560.
- Malenka, R. C., & Bear, M. F. (2004). LTP and LTD: An embarrassment of riches. *Neuron*, 44(1), 5–21.
- Manjunath, G., & Jaeger, H. (2013). Echo state property linked to an input: Exploring a fundamental characteristic of recurrent neural networks. *Neural Computation*, 25(3), 671–696.
- Mayer, N. M. (2015). Input-anticipating critical reservoirs show power law forgetting of unexpected input events. *Neural Computation*, 27(5), 1102–1119.
- Meisel, C., & Gross, T. (2009). Adaptive self-organization in a realistic neural network model. *Physical Review E*, 80(6), 6.
- Mejias, J. F., Hernandezgomez, B., & Torres, J. J. (2012). Short-term synaptic facilitation improves information retrieval in noisy neural networks. *EPL*, 97(4), 48008.
- Mejias, J., Kappen, H., Longtin, A., & Torres, J. (2013). Short-term synaptic plasticity and heterogeneity in neural systems. In *AIP conference proceedings: Vol. 1510*, (1), (pp. 185–194). AIP.
- Mejias, J. F., & Torres, J. J. (2009). Maximum memory capacity on neural networks with short-term synaptic depression and facilitation. *Neural Computation*, 21(3), 851–871.
- Mi, Y., Li, L., Wang, D., & Wu, S. (2014). A synaptical story of persistent activity with graded lifetime in a neural system. *Advances in Neural Information Processing Systems*, 352–360.
- Millman, D., Mihalas, S., Kirkwood, A., & Niebur, E. (2010). Self-organized criticality occurs in non-conservative neuronal networks during ‘up’ states. *Nature Physics*, 6(10), 801–805.
- Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. (2018). Spectral normalization for generative adversarial networks, arXiv preprint arXiv:1802.05957.
- Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, 319(5869), 1543–1546.
- Park, J., Lee, B., Kang, S., Kim, P. Y., & Kim, H. J. (2017). Online learning control of hydraulic excavators based on echo-state networks. *IEEE Transactions on Automation Science and Engineering*, 14(1), 249–259.
- Pei, S., Tang, S., Yan, S., Jiang, S., Zhang, X., & Zheng, Z. (2012). How to enhance the dynamic range of excitatory-inhibitory excitable networks. *Physical Review E*, 86(2), 021909.
- Restrepo, J. G., Ott, E., & Hunt, B. R. (2005). Onset of synchronization in large networks of coupled oscillators. *Physical Review E*, 71(3), 036151.
- Rodan, A., & Tino, P. (2011). Minimum complexity echo state network. *IEEE Transactions on Neural Networks*, 22(1), 131–144.
- Rosenbaum, R., Rubin, J., & Doiron, B. (2012). Short term synaptic depression imposes a frequency dependent filter on synaptic information transfer. *PLoS Computational Biology*, 8(6), 18.
- Rotman, Z., Deng, P. Y., & Klyachko, V. A. (2011). Short-term plasticity optimizes synaptic information transmission. *Journal of Neuroscience*, 31(41), 14800–14809.
- Seeholzer, A., Deger, M., & Gerstner, W. (2018). Stability of working memory in continuous attractor networks under the control of short-term plasticity, bioRxiv, 424515.
- Shew, W. L., Clawson, W. P., Pobst, J., Karimippanah, Y., Wright, N. C., & Wessel, R. (2015). Adaptation to sensory input tunes visual cortex to criticality. *Nature Physics*, 11(8), 659–663.
- Shew, W. L., & Plenz, D. (2013). The functional benefits of criticality in the cortex. *Neuroscientist*, 19(1), 88–100.
- Shew, W. L., Yang, H. D., Petermann, T., Roy, R., & Plenz, D. (2009). Neuronal avalanches imply maximum dynamic range in cortical networks at criticality. *Journal of Neuroscience*, 29(49), 15595–15600.
- Shew, W. L., Yang, H. D., Yu, S., Roy, R., & Plenz, D. (2011). Information capacity and transmission are maximized in balanced critical networks with neuronal avalanches. *Journal of Neuroscience*, 31(1), 55–63.
- Torres, J. J., Pantic, L., & Kappen, H. J. (2002). Storage capacity of attractor neural networks with depressing synapses. *Physical Review E*, 66(6), 061910.
- Tsodyks, M., & Wu, S. (2013). Short-term synaptic plasticity. *Scholarpedia*, 8(10), 3153.
- Van Mieghem, P., Omic, J., & Kooij, R. (2009). Virus spread in networks. *IEEE/ACM Transactions on Networking*, 17(1), 1–14.
- Virkar, Y. S., Shew, W. L., Restrepo, J. G., & Ott, E. (2016). Feedback control stabilization of critical dynamics via resource transport on multilayer networks: How glia enable learning dynamics in the brain. *Physical Review E*, 94(4), 042310.
- Witton, J., Padmashri, R., Zinyuk, L. E., Popov, V. I., Kraev, I., Line, S. J., et al. (2015). Hippocampal circuit dysfunction in the Tc1 mouse model of Down syndrome. *Nature Neuroscience*, 18(9), 1291–1298.
- Yang, H. D., Shew, W. L., Roy, R., & Plenz, D. (2012). Maximal variability of phase synchrony in cortical networks with neuronal avalanches. *Journal of Neuroscience*, 32(3), 1061–1072.



- Yildiz, I. B., Jaeger, H., & Kiebel, S. J. (2012). Re-visiting the echo state property. *Neural Networks*, 35, 1–9.
- Yu, S., Ribeiro, T. L., Meisel, C., Chou, S., Mitz, A., Saunders, R., et al. (2017). Maintained avalanche dynamics during task-induced changes of neuronal activity in nonhuman primates. *Elife*, 6, 22.
- Yu, S., Yang, H., Shriki, O., & Plenz, D. (2013). Universal organization of resting brain activity at the thermodynamic critical point. *Frontiers in Systems Neuroscience*, 7, 42.
- Zucker, R. S., & Regehr, W. G. (2002). Short-term synaptic plasticity. *Annual Review of Physiology*, 64, 355–405.