The whole is different from the sum of its parts

Omri Barak[1], Daniel Ben Dayan Rubin[2]

[1] Faculty of Medicine. Technion, Israel

[2] ICRI, Intel Collaborative Research Institute, Intel Labs

Introduction

- Neural networks are able to integrate information over extended periods of time, even though an individual action potential is an event of a few milliseconds.
- A possible explanation is that long timescales arise as an emergent network phenomenon.
- But... individual neurons have a multitude of biophysical processes, with a large range of timescales.
- We investigate the memory capacity of networks composed of neurons with slow timescales.

1 Modeling framework

We use an echo state network, augmented with synaptic depression/facilitation:

$$\dot{x}_i = -x_i + g\textstyle\sum_j J_{ij}\, y_j r_j + aW_i^{in} u(t)$$

$$r_j = 1/(1+e^{-4x_j})$$

$$\dot{y}_j = \begin{cases} \frac{1-y_j}{\tau_y} - U y_j r_j & \text{(Depression)} \\ -\frac{y_j}{\tau_y} + U(1-y_j) r_j & \text{(Facilitation)} \end{cases}$$

$u(t)$ colored noise. Timescale τ_u

The connectivity J is sparse (f=0.1) random, and the synaptic dynamics are a simplified version of the Tsodyks-Markram model. The rate r(x) was chosen with positive values in order to have meaningful interaction with the y variable.

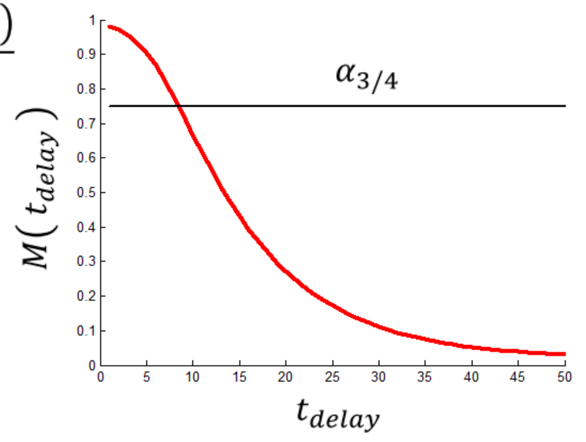
3 Measuring memory capacity

We use the memory function (Jaeger 2001) to estimate the network's memory capacity. It is defined as the correlation between the stimulus delayed by a certain amount and the estimate of the same delayed stimulus using a linear readout of the current network state:

$$M(t_{delay}) = \frac{\boldsymbol{v}^T(t_{delay})\boldsymbol{C}^{-1}\boldsymbol{v}(t_{delay})}{\sigma^2}$$

$$\boldsymbol{v}(t_{delay}) = \langle \boldsymbol{r}(t)\boldsymbol{u}(t-t_{delay}) \rangle_t$$

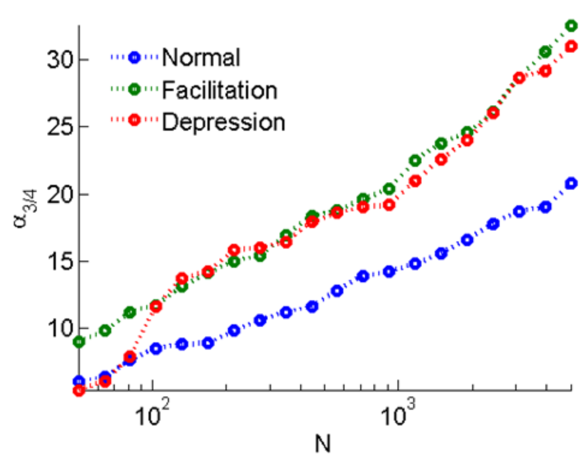
$$\boldsymbol{C} = \langle \boldsymbol{r}(t)\boldsymbol{r}^T(t) \rangle_t$$



4 Slow timescales improve capacity

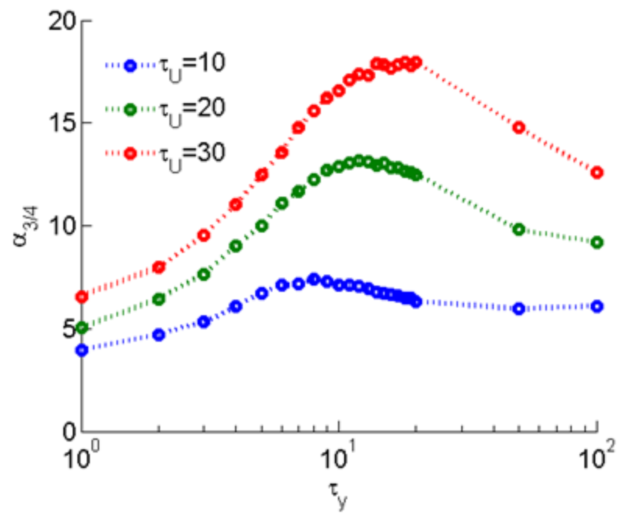
introduction of depression or facilitation improves performance, but does not alter the scaling.

Only a hidden subsystem that is nonlinearly coupled to the firing rates supports the slow process.



Note that the slow processes double the number of dynamic variables (though not the number of readout weights), but the improvement due to them is more than that expected by doubling network size.

5 Input and slow timescales need to match



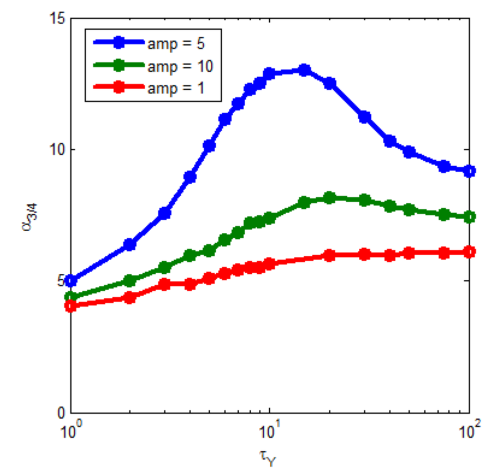
To maximize memory capacity, the slow timescales of the network have to approximately match those of the input.

Note that τ_y is the relaxation time of a nonlinearly coupled hidden variable.

6 Relaxation timescale is not everything

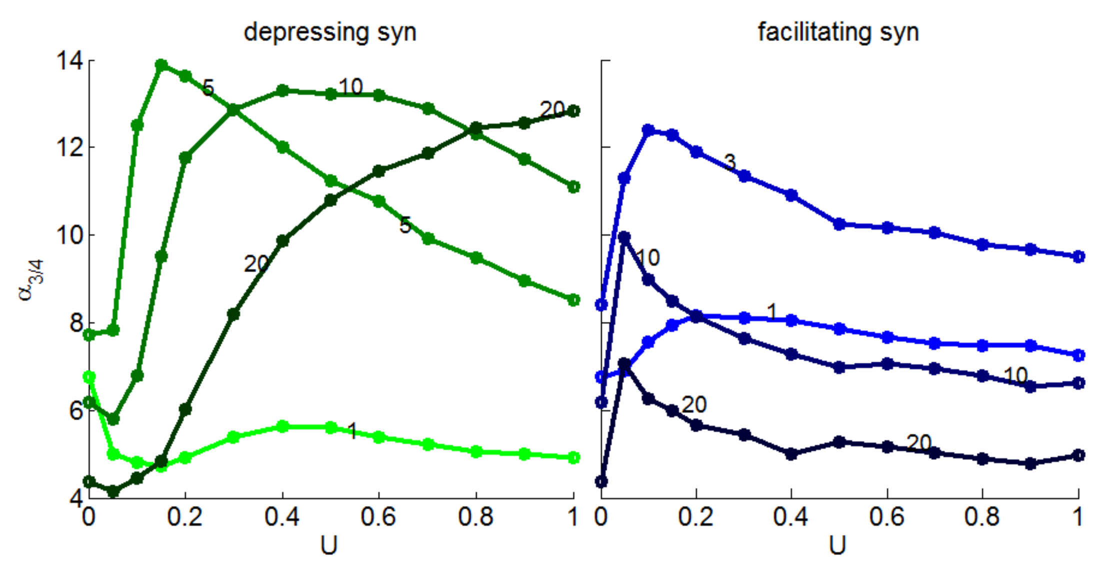
The natural candidate for a source of slow dynamics is the relaxation timescale of depression or facilitation.

Changing either the utilization parameter U, or the input amplitude, the optimal timescale is affected.



While U can be considered as a timescale (transition rate to the depressed/ facilitated state), the amplitude cannot. Rather, the amplitude determines how far away from the origin the system will be, and consequently how nonlinear the dynamics is.

The complex interaction between the degree of nonlinearity and the various slow timescales can also be appreciated by considering performance by varying both U and the synaptic gain g (different color tones):



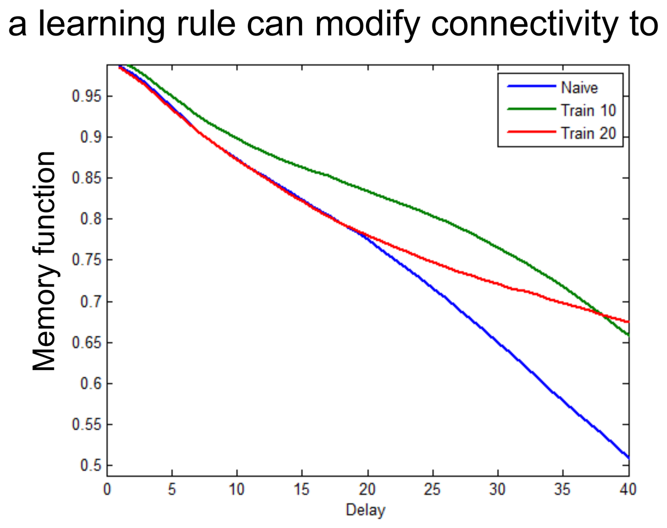
7 Training can emphasize certain timescales

Network topology is known to strongly affect memory capacity [4, 5].

In biological settings, however, it is likely that the same network will be required to perform tasks with different temporal demands.

We thus set out to check whether a learning rule can modify connectivity to obtain relevant improvements in capacity.

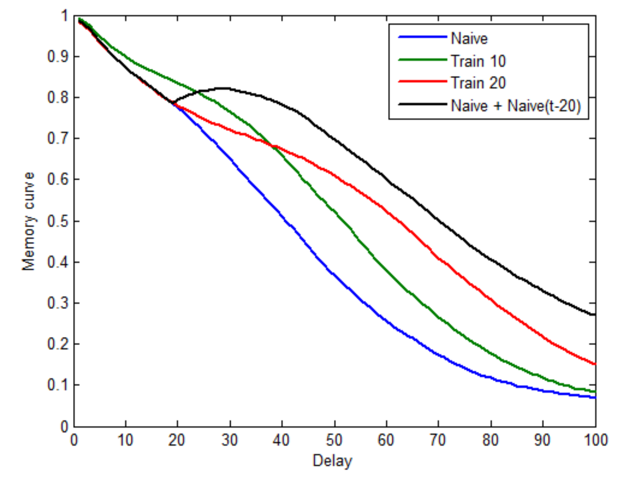
Networks were trained to output a delayed version of the colored noise input using the FORCE algorithm [6]. Note that training for longer delays improves reconstruction of large delays at the expense of early ones.



8 Why training works

How were certain timescales emphasized?

The FORCE algorithm feeds back the readout into the network. If the readout is an estimate of the delayed input, then it is as if there are two inputs - current and delayed - being fed into the network. Thus we might expect the resulting memory curve to be a sum of the original memory curve and a scaled and shifted version of it:



As a rough approximation, we convolve a delayed version with the Naive curve, and then add it back.

9 Conclusions

- Slow processes can contribute to memory capacity, even if they are hidden and non-linearly coupled to the firing rates.
- The timescale of the slow process has to approximately match that of the input.
- The degree of nonlinearity, affected
- Training can emphasize certain timescales by extracting them and feeding back into the network.

References

[1] M. Tsodyks, K. Pawelzik, H. Markram 1998 Neural Comput. 10:821.
[2] Jaeger 2001 GMD Report 152: German NRC for Information Technology.
[3] B Schrauwen, J Defour, D Verstraeten, J Van Campenhout 2007 ICANN 471.
[4] S Ganguli, D Huh, H Sompolinsky 2008 PNAS (USA) 105:18970.
[5] OL White, DD Lee, H Sompolinsky 2004 PRL 92:148102.
[6] D Sussillo, LF Abbot 2009. Neuron. 63(4):423. [7] E Wallace, H Maei, P Latham 2013Neural Comp. 25:1408.

Funding: Intel Collaborative Research Institute

The Lorry I. Lokey Interdisciplinary Center for Life Sciences and Engineering
NETWORK BIOLOGY RESEARCH LABORATORIES

Science and Engineering of Neural Systems