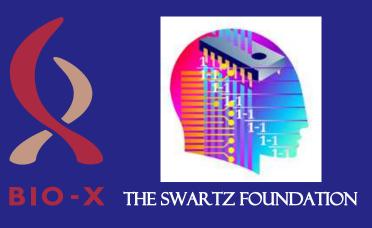


# Learning and memory with complex synapses

Subhaneil Lahiri and Surya Ganguli
Department of Applied Physics, Stanford University, Stanford CA



## Background

## Storage capacity of synaptic memory

A classical perceptron, when used as a recognition memory device, has a memory capacity proportional to the number of synapses *N*.

However, this requires synapses to have a dynamic range also  $\propto N$ . If synaptic efficacies are limited to a fixed dynamic range, this introduces a strong tradeoff between learning and forgetting due to new memories overwriting old. If we wish to rapidly store new memories, then memory capacity is  $\mathcal{O}(logN)$ .

[Amit and Fusi (1992), Amit and Fusi (1994)]

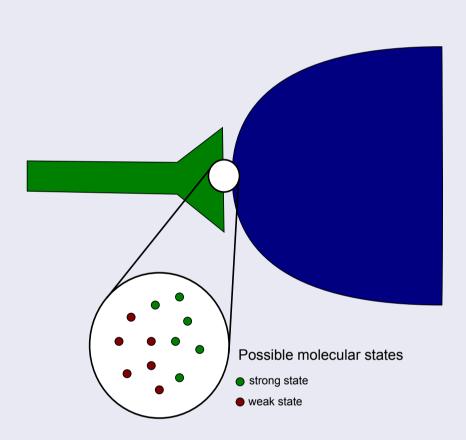
To circumvent this tradeoff, it is essential to enlarge our theoretical conception of a synapse as a single number.

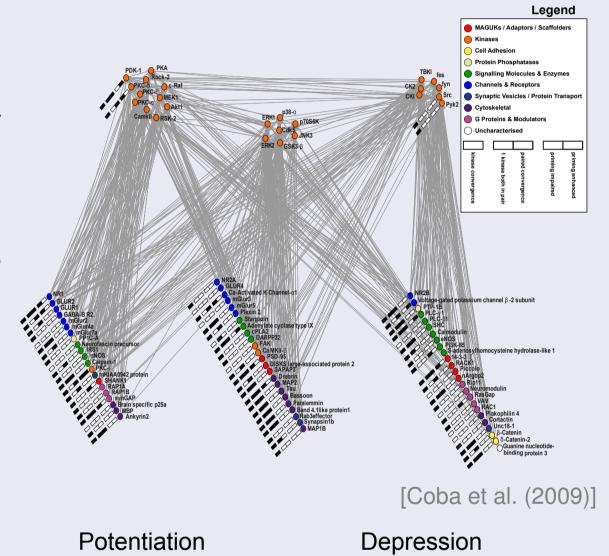
# **Complex synapses**

In reality, a synapse is a complex dynamical system.

We will describe a synapse by stochastic processes on a finite number of states, *n*.

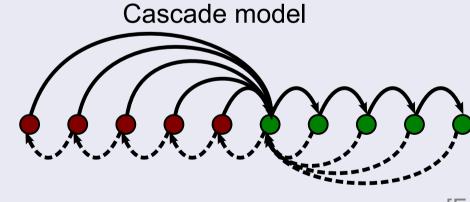
Potentiation and depression cause transitions between these states.

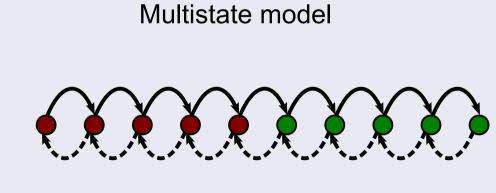




## Cascade and multistate models

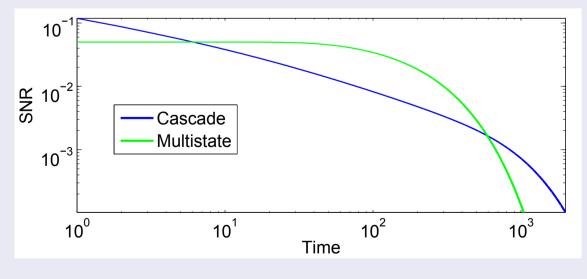
Two example models of complex synapses.





[Fusi et al. (2005), Amit and Fusi (1994), Fusi and Abbott (2007)]

These have different memory storage properties



#### Questions

- Can we understand the space of all possible synaptic models?
- How does the structure (topology) of a synaptic model affect its function (memory curve)?
- How does synaptic complexity (number of states) extend the frontiers of possibility for memory?
- Which synaptic state transition topologies maximize measures of memory?

#### Framework

## Synaptic state transition models

We have two Markov processes describing transition probabilities for potentiation,  $\mathbf{M}^{\text{pot}}$ , and depression,  $\mathbf{M}^{\text{dep}}$ . Plasticity events are potentiating with probability  $f^{\text{pot}}$  and depressing with probability  $f^{\text{dep}}$ .

After the memory we are tracking, subsequent plasticity events occur at rate r, with transition probabilities  $\mathbf{M}^{\text{forget}} = f^{\text{pot}}\mathbf{M}^{\text{pot}} + f^{\text{dep}}\mathbf{M}^{\text{dep}}.$ 

This will eventually return it to the equilibrium distribution,  $\mathbf{p}^{\infty}$ .

## **Memory curve**

We use the ideal observer approach: read synaptic weights directly. This is an upper bound on what could be read from network activity.

Reconstruction probability of a single synapse:

$$s(t) = f^{\text{pot}}P(\text{strong}, t|\text{pot}, 0) + f^{\text{dep}}P(\text{weak}, t|\text{dep}, 0)$$

Alternatively, if  $\vec{w}$  is an N-element vector of synaptic strengths,

Signal = 
$$\langle \vec{w}_{\mathsf{ideal}} \cdot \vec{w}(t) - \vec{w}_{\mathsf{ideal}} \cdot \vec{w}(\infty) \rangle$$
  
Noise =  $\mathsf{Var} (\vec{w}_{\mathsf{ideal}} \cdot \vec{w}(\infty))$ 

If we ignore correlations between different synapses, signal-to-noise ratio:

$$\mathsf{SNR}(t) \sim \sqrt{N}(s(t) - s(\infty)).$$

## **Upper bounds on performance**

#### Area bound

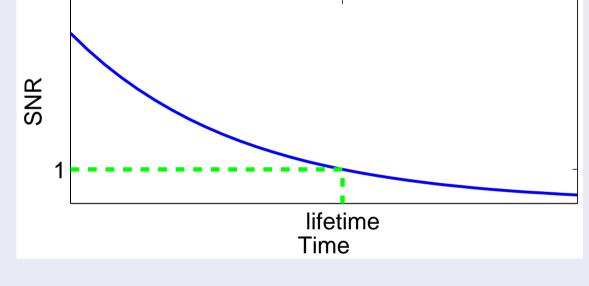
The memory lifetime is bounded by the area under the SNR curve:

$$SNR(lifetime) = 1 \implies lifetime < A.$$

We can show that this area has an upper bound:

$$A \leq \sqrt{N}(n-1)/r$$
.

This is saturated by a transition diagram with a linear chain topology.



# Proof: Impose an ordering on the states

Let  $T_{ii}$  be the mean first passage time from state i to state j. The following quantity

$$\eta = \sum_{j} \mathsf{T}_{ij} \mathsf{p}_{j}^{\infty},$$

is independent of the initial state i. It is known as Kemeney's constant.

[Kemeny and Snell (1960)]

We define:

$$\eta_{\pmb{i}}^+ = \sum_{\pmb{j} \in \mathsf{strong}} \mathsf{T}_{\pmb{i}\pmb{j}} \mathsf{p}_{\pmb{j}}^\infty, \qquad \eta_{\pmb{i}}^- = \sum_{\pmb{j} \in \mathsf{weak}} \mathsf{T}_{\pmb{i}\pmb{j}} \mathsf{p}_{\pmb{j}}^\infty$$

These measure "distance" to the srong/weak states. They can be used to arrange the states in an order (increasing  $\eta^-$  or decreasing  $\eta^+$ ).

## Maximal area

Given any molecular network, we can construct one with the multistate topology that has

- same state order,
- same equilibrium distribution,
- larger area.

Uses a deformation that reduces "shortcut" transition probabilities and increases the bypassed "direct" ones.

The area of this model is

 $A = \frac{2\sqrt{N}}{r} \sum_{k} \mathbf{p}_{k}^{\infty} |k - \langle k \rangle|.$ 

This is maximized when the equilibrium probability distribution is concentrated at both ends.

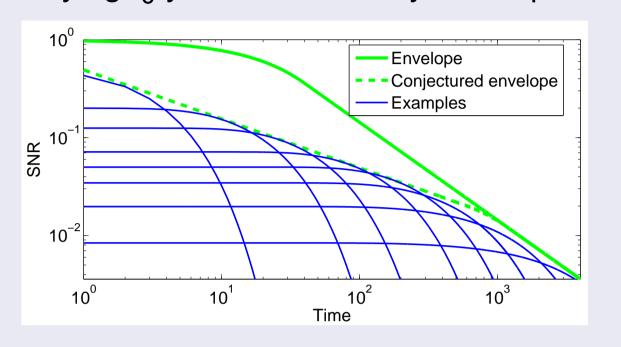
## The memory envelope

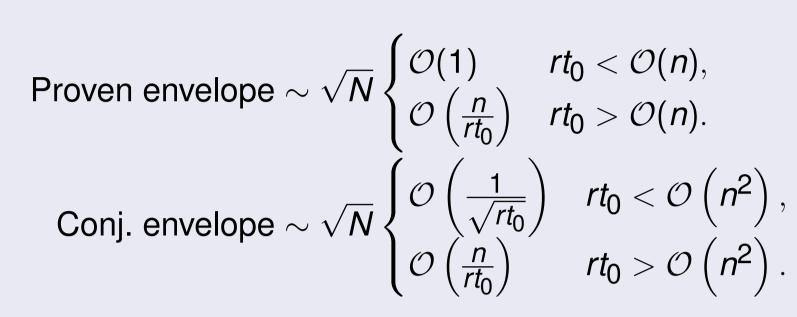
## The frontiers of possibility: a maximal SNR curve

Markovian learning and forgetting  $\implies$  SNR is a sum of exponentials.

Optimizing the SNR at one time,  $t_0$ , over the space of such curves, subject to upper bounds on initial SNR and area, yields an upper bound on SNR at  $t_0$  for any synaptic model. The resulting optimal memory curve is a single exponential.

Varying  $t_0$  yields an memory envelope curve with a power law tail.





## Extra constraint: limits of diffusive learning and forgetting

The envelope above may not be tight. We conjecture an additional constraint, which would yield a tight envelope (dashed line above). Schematically, mode by mode:

$$SNR(0)\sqrt{time-scale} \leq \sqrt{N} \cdot \mathcal{O}(1)$$
.

We have found no model that can exceed this bound. It is saturated by a diffusive chain:

$$SNR(0) \sim \frac{1}{n}$$
, time-scale  $\sim n^2$ .

## Summary

- We have formulated a general theory of learning and memory with complex synapses.
- We can impose an order on the internal states of a synapse through the theory of first passage times.
- The area under the memory curve of any synaptic transition diagram cannot exceed that of a linear chain with the same equilibrium probability distribution.
- We find a memory envelope: a single curve that cannot be exceeded by the memory curve of any synaptic model. Synaptic complexity (n internal states) raises the memory envelope linearly in n for times  $> \mathcal{O}(n^2)$ .
- For times  $< \mathcal{O}(n^2)$  we conjecture the close to optimal synaptic model that reaches the envelope exploits deterministic transitions, resulting in diffusive forgetting.

#### References

- D. J. Amit and S. Fusi, "Constraints on learning in dynamic synapses", Network: Computation in Neural Systems, 3(4):443–464, 1992.
- D. J. Amit and S. Fusi, "Learning in neural networks with material synapses", *Neural Computation*, 6(5):957–982, 1994.
- M. P. Coba, A. J. Pocklington, M. O. Collins, M. V. Kopanitsa, R. T. Uren, S. Swamy, M. D. Croning, J. S. Choudhary, and S. G. Grant, "Neurotransmitters drive combinatorial multistate postsynaptic density networks", *Sci Signal*, 2(68):ra19, 2009, PubMed: 19401593.
- S. Fusi, P. J. Drew, and L. F. Abbott, "Cascade models of synaptically stored memories", Neuron, 45(4):599–611, Feb 2005, PubMed: 15721245.
- S. Fusi and L. F. Abbott, "Limits on the memory storage capacity of bounded synapses", *Nat. Neurosci.*, 10(4): 485–493, Apr 2007, PubMed: 17351638.
- J.G. Kemeny and J.L. Snell, Finite markov chains. Springer, 1960.

# Acknowledgements

SL and SG thank the Swartz Foundation, Burroughs Wellcome Foundation, Stanford Bio-X Neuroventures and DARPA for funding, and Larry Abbott and Stefano Fusi for useful conversations.