A memory frontier for 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 store new memories rapidly, then memory capacity is $\mathcal{O}(\log N)$).

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

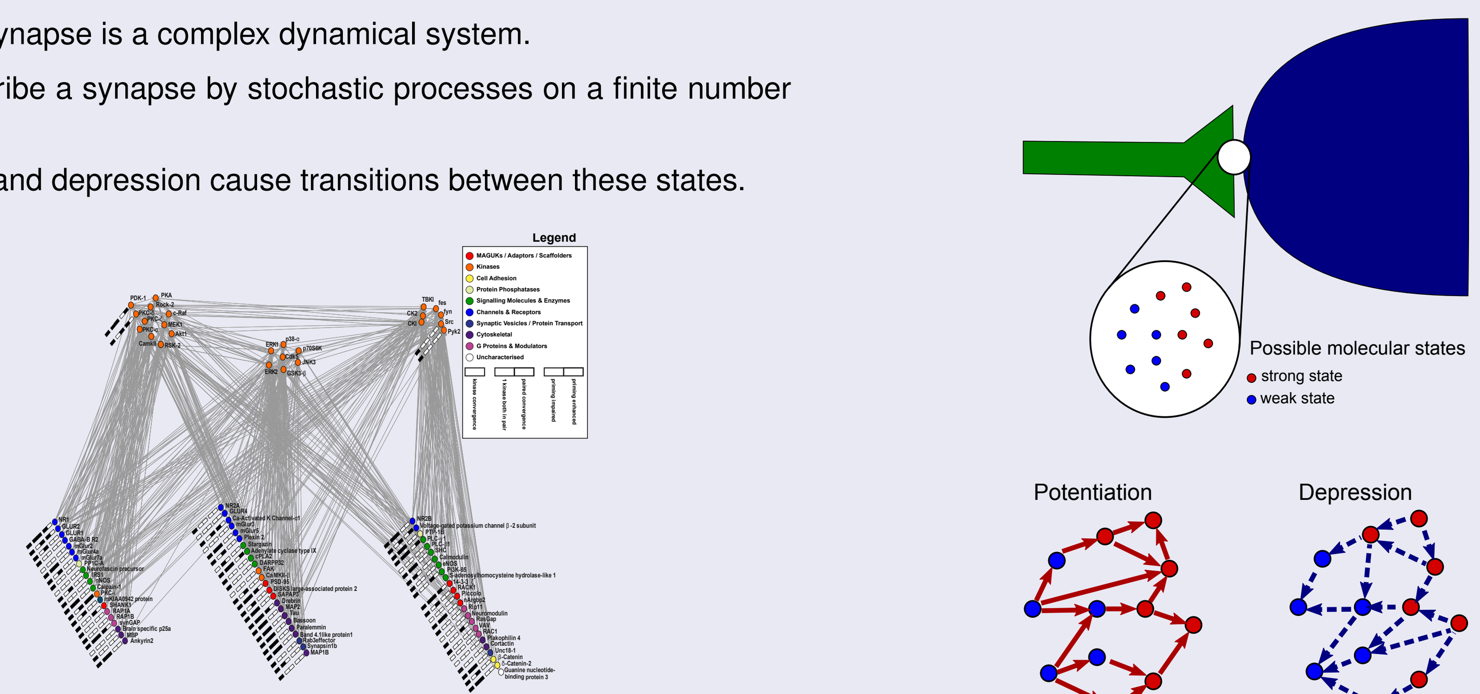
To circumvent this tradeoff, it is essential to enlarge our theoretical conception of a synapse as a single number.

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, M.

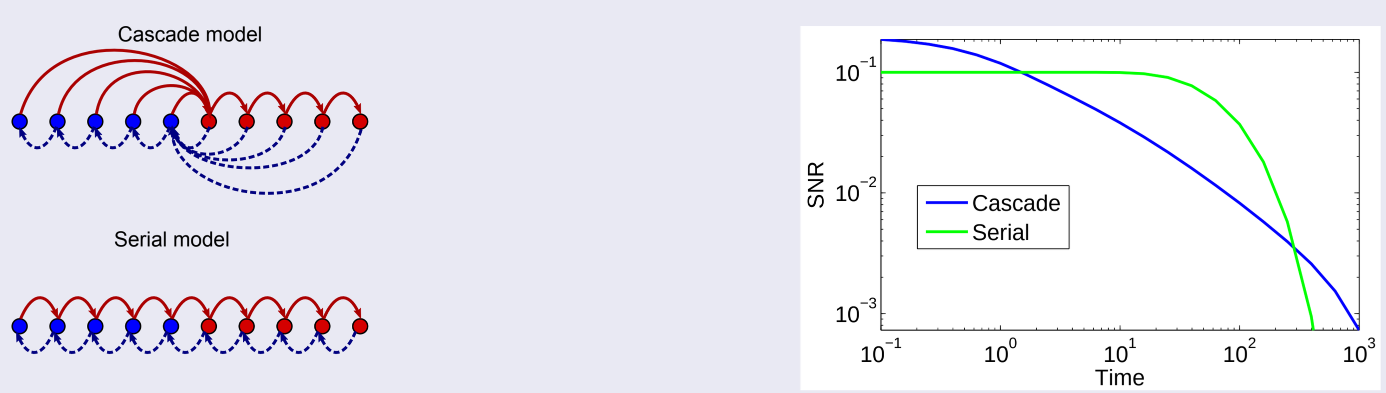
Potentiation and depression cause transitions between these states.



[Coba et al. (2009)]

Cascade and serial models

Two example models of complex synapses with different memory storage properties.



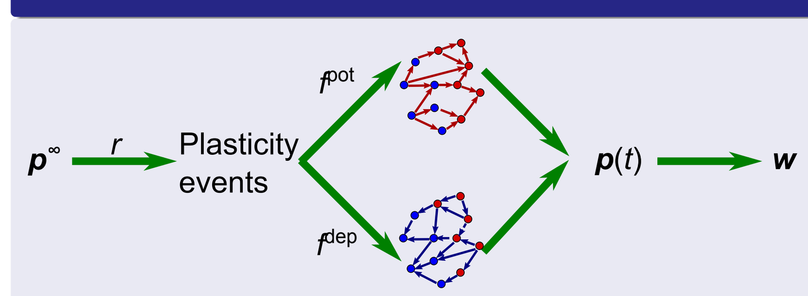
[Fusi et al. (2005), Leibold and Kempter (2008)]

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



Assumptions:

- Candidate plasticity events occur independently at each synapse,
- Each synapse responds with the same state-dependent rules,
- Synaptic weight takes only two values, ± 1.

[Fusi et al. (2005), Fusi and Abbott (2007), Barrett and van Rossum (2008)]

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. To measure memory quality, let $\vec{w}$ be an N-element vector of synaptic strengths,

$$\mathrm{SNR}(t) = \frac{\langle \vec{w}_{\text{ideal}} \cdot \vec{w}(t) \rangle - \langle \vec{w}_{\text{ideal}} \cdot \vec{w}(\infty) \rangle}{\sqrt{\mathrm{Var}(\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty))}}.$$

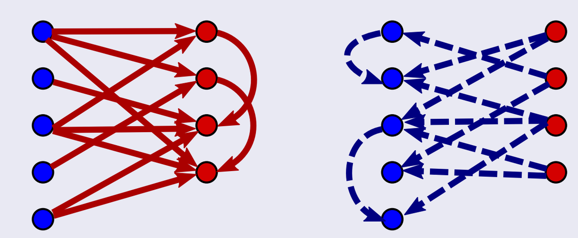
Upper bounds on performance

Initial SNR bound

Initial SNR is closely related to equilibrium flux between strong & weak states

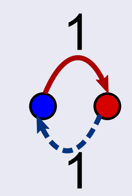
$$\mathrm{SNR}(0) \leq \frac{4\sqrt{N}}{r}\,\Phi_{-+}.$$

Maximized when potentiation guarantees $\vec{w} \to$ strong, depression guarantees $\vec{w} \to$ weak.



$\to$ Equivalent to two-state model

Transitions:



$\Longrightarrow \quad \mathrm{SNR}(t) = \sqrt{N}\,(4 f^{\text{pot}} f^{\text{dep}})\,\mathrm{e}^{-rt}.$

Maximal initial SNR: $\quad \mathrm{SNR}(0) \leq \sqrt{N}.$

Area bound

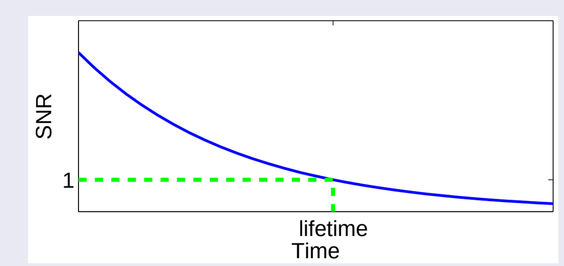
The memory lifetime is bounded by the area under the SNR curve:

$$\mathrm{SNR}(\text{lifetime}) = 1 \quad \Longrightarrow \quad \text{lifetime} < A.$$

We can show that this area has an upper bound:

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

This is saturated by a transition diagram with the serial topology.



Proof: Impose an ordering on the states

Let $\mathbf{T}_{ij}$ be the mean first passage time from state i to state j. The following quantity

$$\eta = \sum_j \mathbf{T}_{ij} \mathbf{p}^\infty_j,$$

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

We define:

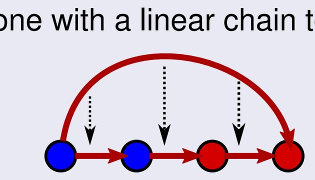
$$\eta^+_i = \sum_{j \in \text{strong}} \mathbf{T}_{ij} \mathbf{p}^\infty_j, \qquad \eta^-_i = \sum_{j \in \text{weak}} \mathbf{T}_{ij} \mathbf{p}^\infty_j.$$

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

Maximal area

Given any synaptic model, we can construct one with a linear chain topology that has

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

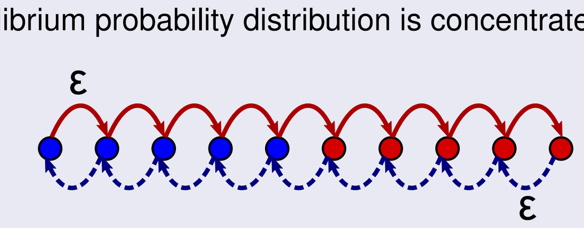


Uses a deformation that reduces "shortcut" transition probabilities and increases the bypassed "direct" ones.

The area of this model is

$$A = \frac{2\sqrt{N}}{r} \sum_k \mathbf{p}^\infty_k \,|k - \langle k \rangle|.$$

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



In the limit $\varepsilon \to 0$.

Eigenmode decomposition

We can split the system along eigenvectors of the stochastic forgetting process:

$$\mathrm{SNR}(t) = \sqrt{N} \sum_a \mathcal{I}_a \mathrm{e}^{-rt/\tau_a}.$$

The upper bounds on initial SNR and area tell us:

$$\sum_a \mathcal{I}_a \leq 1, \qquad \sum_a \mathcal{I}_a \tau_a \leq M-1.$$

- What are the implications for the full memory curve?
- Are there any other important constraints?

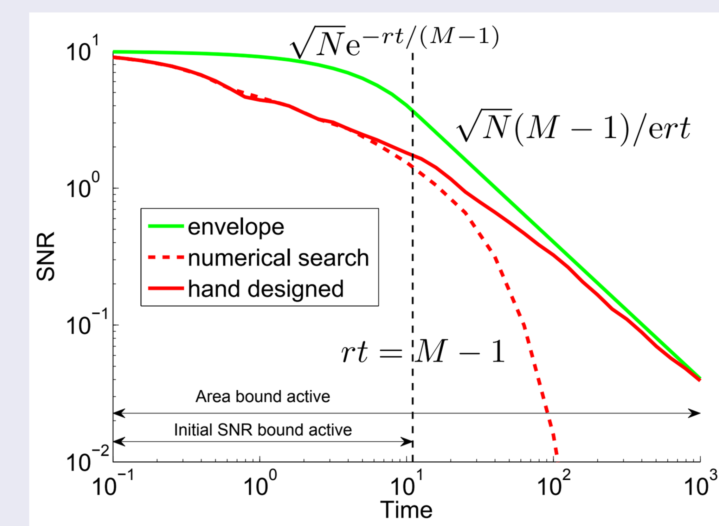
The memory envelope

The frontiers of possibility: a maximal SNR curve

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

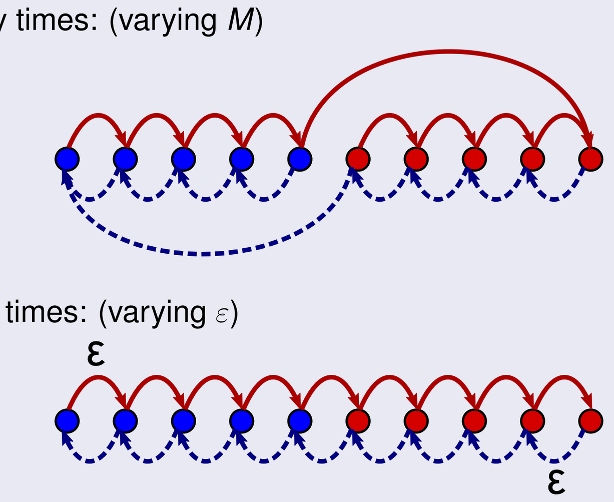
Optimizing the SNR *at one time*, t_0, over the space of such curves, subject to upper bounds on initial SNR and area, yields an upper bound on SNR at t_0 for *any* synaptic model. The resulting optimal memory curve is a single exponential (optimizing at two or more well separated times requires multiple exponentials).

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



Early times: (varying M)

Late times: (varying ε)

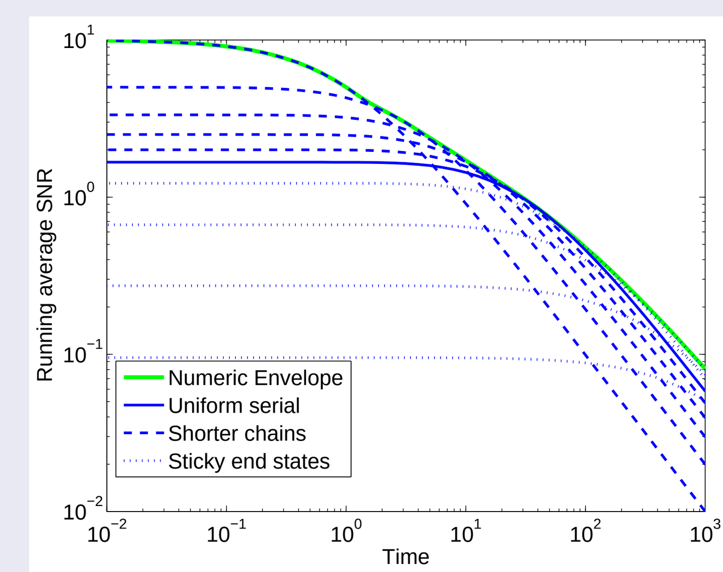


Envelope for running average memory curve

We define the running average SNR:

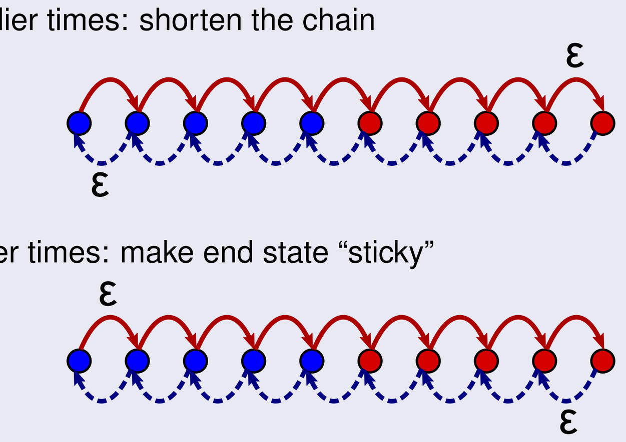
$$\widehat{\mathrm{SNR}}(\tau) = \frac{1}{\tau}\int_0^\infty \mathrm{d}t\, \mathrm{e}^{-t/\tau}\, \mathrm{SNR}(t) \sim \frac{1}{\tau}\int_0^\tau \mathrm{d}t\, \mathrm{SNR}(t)$$

For any τ, this is maximized by a model with the serial topology.



Earlier times: shorten the chain

Later times: make end state "sticky"



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 (M internal states) raises the memory envelope linearly in M for times $> \mathcal{O}(M)$.

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) .
- S. Fusi, P. J. Drew, and L. F. Abbott, "Cascade models of synaptically stored memories", *Neuron*, 45(4):599–611, (Feb, 2005) .
- Christian Leibold and Richard Kempter, "Sparseness Constrains the Prolongation of Memory Lifetime via Synaptic Metaplasticity", *Cerebral Cortex*, 18 (1):67–77, (2008) .
- S. Fusi and L. F. Abbott, "Limits on the memory storage capacity of bounded synapses", *Nat. Neurosci.*, 10(4):485–493, (Apr, 2007) .
- A. B. Barrett and M. C. van Rossum, "Optimal learning rules for discrete synapses", *PLoS Comput. Biol.*, 4(11):e1000230, (Nov, 2008) .

Acknowledgements

SL and SG thank the Swartz Foundation, Burroughs Wellcome Foundation, Stanford Bio-X Neuroventures, Genentech and DARPA for funding, and Larry Abbott, Stefano Fusi, Marcus Benna, David Sussillo and Jascha Sohl-Dickstein for useful conversations.