

# A general theory of learning and memory with Complex Synapses

based on work with Surya Ganguli

Subhaneil Lahiri

Stanford University, Applied Physics

October 22, 2015

# Introduction

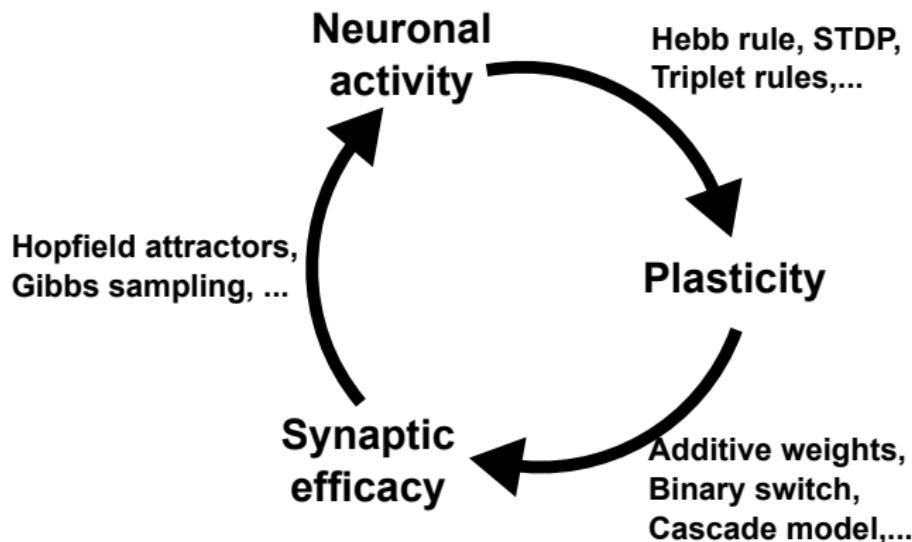
We often model synaptic plasticity as the change of a single number (synaptic weight). In reality, there is a complex dynamical system inside a synapse.

Semi-realistic models of synaptic plasticity have terrible memory without synaptic complexity.

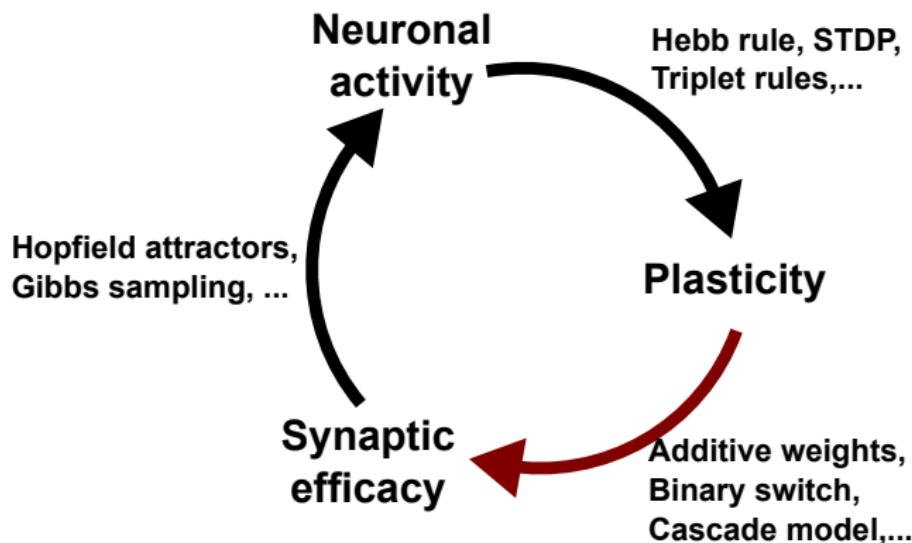
We will study the entire space of a broad class of models of complex synapses to find upper bounds on their performance.

This leads to understanding of what structures are useful for storing memories for different timescales.

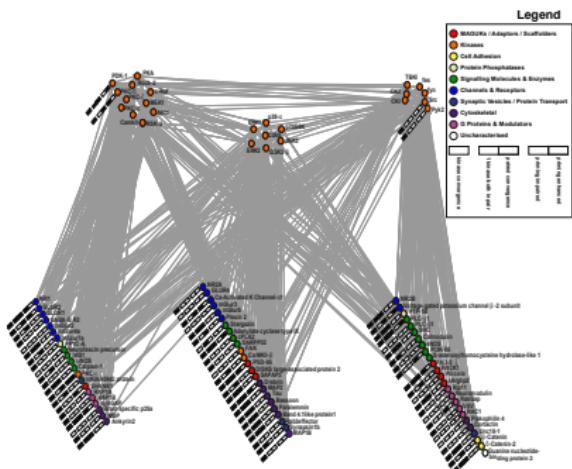
## Synaptic learning and memory



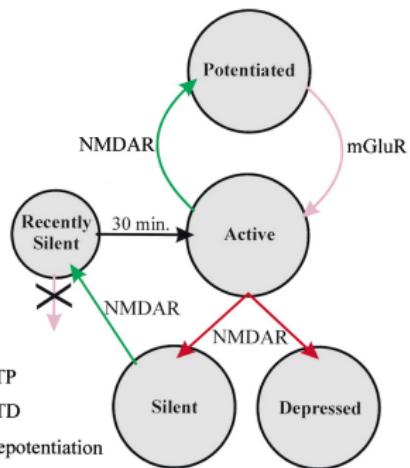
# Synaptic learning and memory



# Synapses are complex



[Coba et al. (2009)]

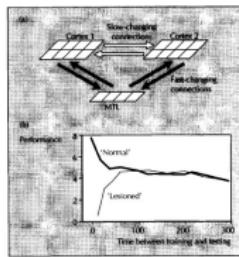


[Montgomery and Madison (2002)]

There is a complex, dynamic system underlying synaptic plasticity.

# Timescales of memory

Memories stored in different places for different timescales

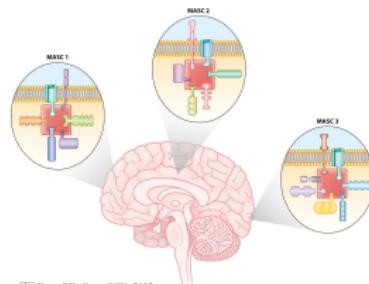


[Squire and Alvarez (1995)]

cf. Cerebellar cortex vs. cerebellar nuclei.

[Krakauer and Shadmehr (2006)]

Different synapses have different molecular structures.



Emes RD, Grant SGN. 2012.  
Annu. Rev. Neurosci. 35:111–31

[Emes and Grant (2012)]

# Outline

- 1 Why complex synapses?
- 2 Modelling synaptic memory
- 3 Upper bounds
- 4 Envelope memory curve
- 5 Experimental tests?

## Section 1

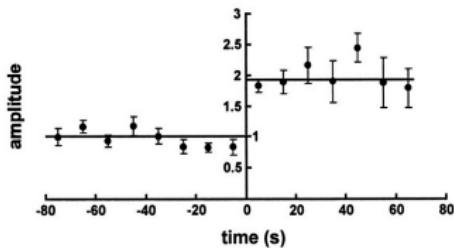
Why complex synapses?

# Storage capacity of synaptic memory

A classical perceptron has a capacity  $\propto N$ , (# synapses).

Requires synapses' dynamic range also  $\propto N$ .

With discrete, finite synapses:  
⇒ new memories overwrite old.



[Petersen et al. (1998), O'Connor et al. (2005)]

When we store new memories rapidly, memory capacity  $\sim \mathcal{O}(\log N)$ .

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

# Trade-off between learning and remembering

Learning   Remembering

Very plastic



# Trade-off between learning and remembering

Learning   Remembering

Very plastic



# Trade-off between learning and remembering

Learning   Remembering

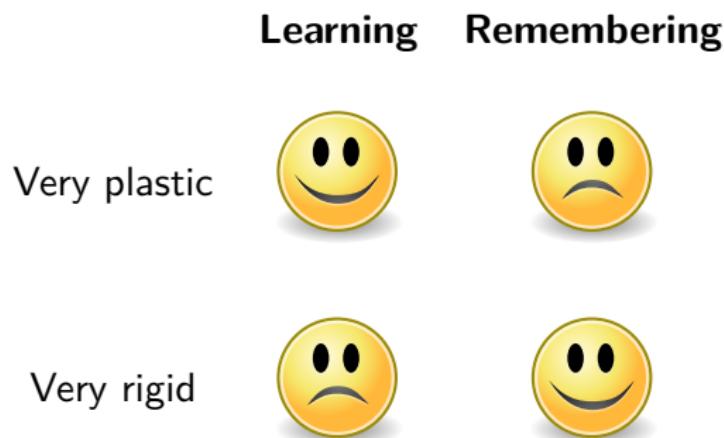
Very plastic



Very rigid



# Trade-off between learning and remembering



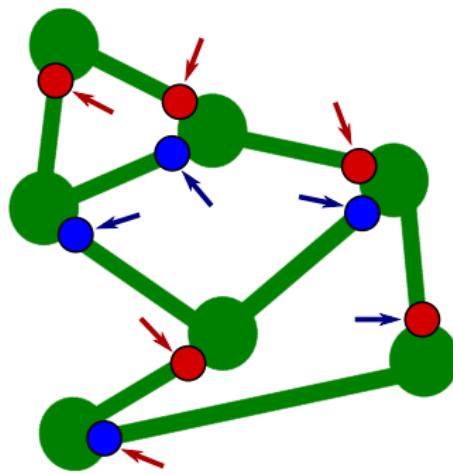
Circumvent tradeoff: go beyond model of synapse as single number.

## Section 2

### Modelling synaptic memory

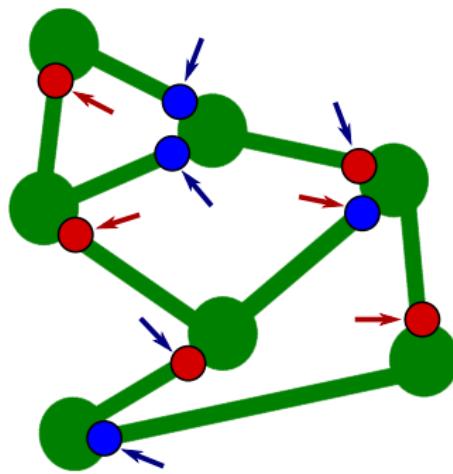
# Recognition memory

Synapses given a sequence of patterns (pot & dep) to store



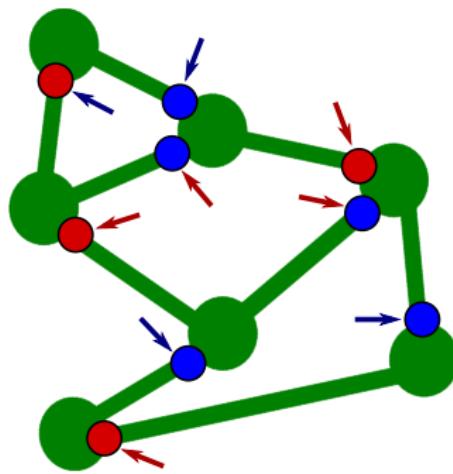
# Recognition memory

Synapses given a sequence of patterns (pot & dep) to store



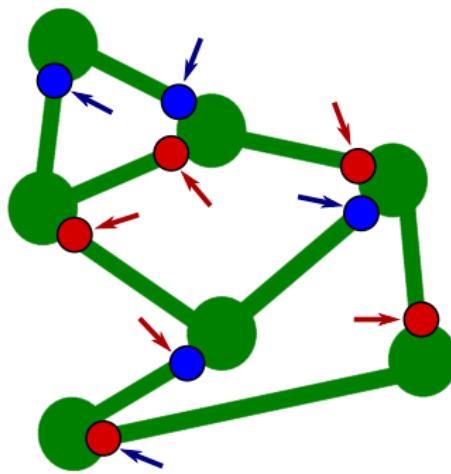
# Recognition memory

Synapses given a sequence of patterns (pot & dep) to store



# Recognition memory

Synapses given a sequence of patterns (pot & dep) to store

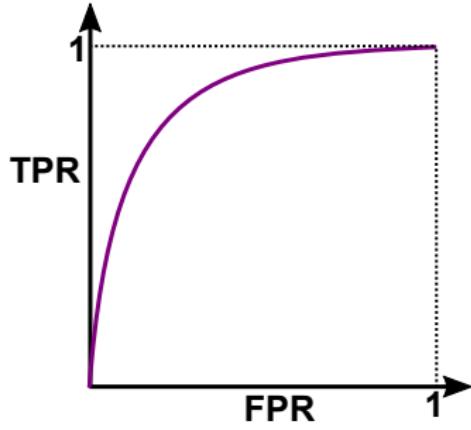


Later: presented with a pattern. Has it been seen before?

# Quantifying memory quality

Have we seen pattern before? Ask if  $\vec{w}_{\text{ideal}} \cdot \vec{w}(t) \geq \theta$ ?

Use  $\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty)$  as a null distribution  $\Rightarrow$  ROC curve:



$$\text{TPR} = \Phi \left( \frac{\text{SNR}(t) + \Phi^{-1}(\text{FPR})}{\text{NNR}(t)} \right),$$

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

$$\text{NNR}(t) = \sqrt{\frac{\text{Var}(\vec{w}_{\text{ideal}} \cdot \vec{w}(t))}{\text{Var}(\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty))}}.$$

Look at:  $\overline{\text{SNR}}(\tau) = \langle \text{SNR}(t) \rangle_{P(t|\tau)}, \quad P(t|\tau) = \frac{e^{-t/\tau}}{\tau}.$

# Models of complex synaptic dynamics

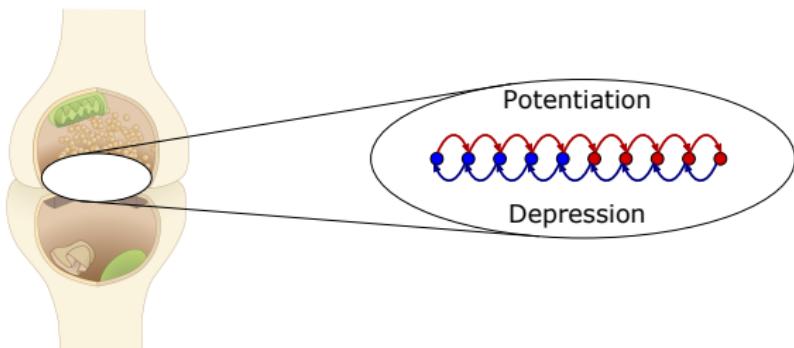


# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
- Candidate plasticity events → transitions between states

weak

strong



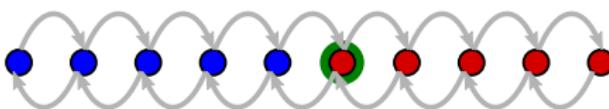
States: #AMPAR, #NMDAR, NMDAR subunit composition,  
CaMK II autophosphorylation, activating PKC, p38 MAPK,...

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

# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
  - Candidate plasticity events → transitions between states
- weak  
● strong

Potentiation event

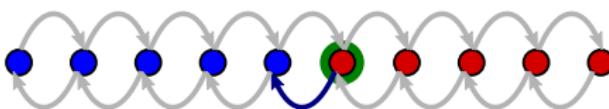


Depression event

# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
  - Candidate plasticity events → transitions between states
- weak  
● strong

Potentiation event

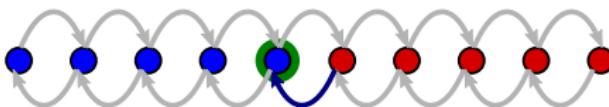


Depression event

# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
  - Candidate plasticity events → transitions between states
- weak  
● strong

Potentiation event

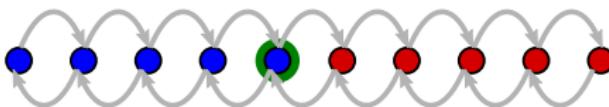


Depression event

# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
  - Candidate plasticity events → transitions between states
- weak  
● strong

Potentiation event

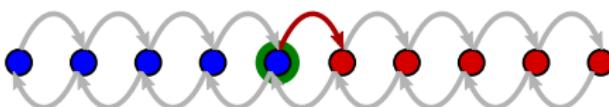


Depression event

# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
  - Candidate plasticity events → transitions between states
- weak  
● strong

## Potentiation event

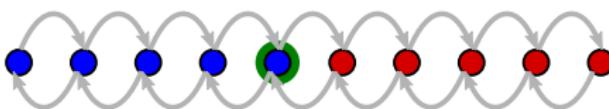


## Depression event

# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
  - Candidate plasticity events → transitions between states
- weak  
● strong

Potentiation event

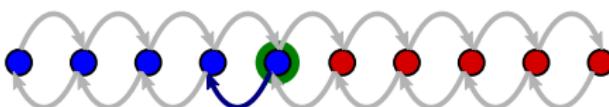


Depression event

# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
  - Candidate plasticity events → transitions between states
- weak  
● strong

Potentiation event



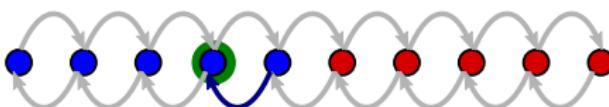
Depression event

Metaplasticity: change propensity for plasticity  
(independent of change in synaptic weight).

# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
  - Candidate plasticity events → transitions between states
- weak  
● strong

Potentiation event



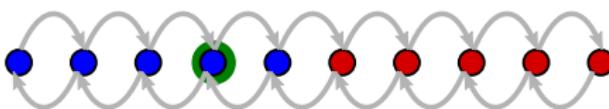
Depression event

Metaplasticity: change propensity for plasticity  
(independent of change in synaptic weight).

# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
  - Candidate plasticity events → transitions between states
- weak  
● strong

Potentiation event



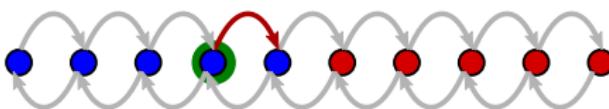
Depression event

Metaplasticity: change propensity for plasticity  
(independent of change in synaptic weight).

# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
  - Candidate plasticity events → transitions between states
- weak  
● strong

## Potentiation event



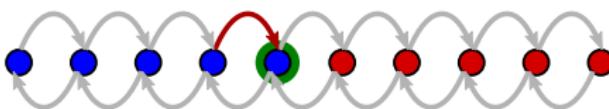
## Depression event

Metaplasticity: change propensity for plasticity  
(independent of change in synaptic weight).

# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
  - Candidate plasticity events → transitions between states
- weak  
● strong

## Potentiation event



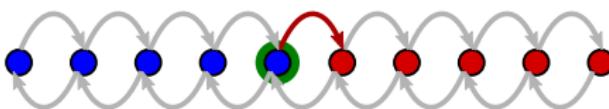
## Depression event

Metaplasticity: change propensity for plasticity  
(independent of change in synaptic weight).

# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
  - Candidate plasticity events → transitions between states
- weak  
● strong

## Potentiation event



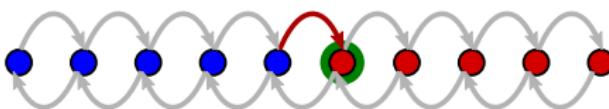
## Depression event

Metaplasticity: change propensity for plasticity  
(independent of change in synaptic weight).

# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
  - Candidate plasticity events → transitions between states
- weak  
● strong

## Potentiation event



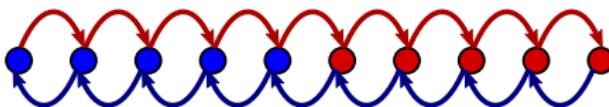
## Depression event

Metaplasticity: change propensity for plasticity  
(independent of change in synaptic weight).

# Models of complex synaptic dynamics

- Internal functional state of synapse → synaptic weight.
  - Candidate plasticity events → transitions between states
- weak  
● strong

## Potentiation

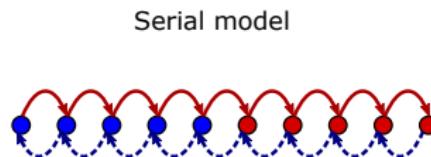
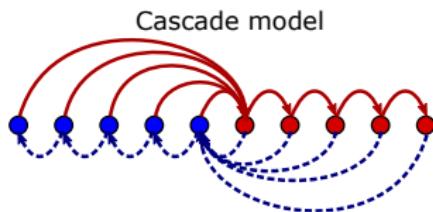


## Depression

Metaplasticity: change propensity for plasticity  
(independent of change in synaptic weight).

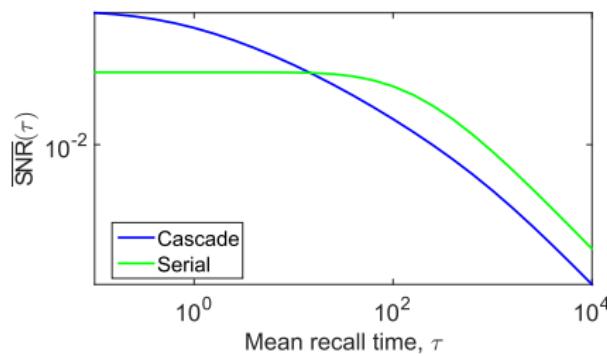
# Example models

Two example models of complex synapses.



[Fusi et al. (2005), Leibold and Kempter (2008), Ben-Dayan Rubin and Fusi (2007)]

These have different memory storage properties

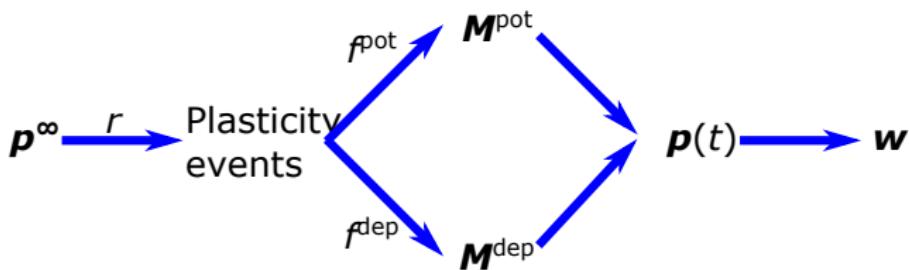


# Questions

- Can we understand the space of *all possible* synaptic models?
- How does structure (topology) of model → function (memory curve)?
- What are the limits on what can be achieved?
- Which transition topologies saturate these limits?
- Can synaptic structure be tuned for different timescales of memory?

# Parameters for synaptic dynamics

There are  $N$  identical synapses with  $M$  internal functional states.



$$\frac{d\mathbf{p}(t)}{dt} = r\mathbf{p}(t)\mathbf{W}^F, \quad \mathbf{W}^F = f^{\text{pot}}\mathbf{M}^{\text{pot}} + f^{\text{dep}}\mathbf{M}^{\text{dep}} - \mathbf{I},$$

$$\mathbf{p}^\infty \mathbf{W}^F = 0.$$

# Constraints

Memory curve given by

$$\text{SNR}(t) = \sqrt{N} (2f^{\text{pot}} f^{\text{dep}}) \mathbf{p}^\infty (\mathbf{M}^{\text{pot}} - \mathbf{M}^{\text{dep}}) \exp(rt\mathbf{W}^F) \mathbf{w},$$

$$\overline{\text{SNR}}(\tau) = \sqrt{N} (2f^{\text{pot}} f^{\text{dep}}) \mathbf{p}^\infty (\mathbf{M}^{\text{pot}} - \mathbf{M}^{\text{dep}}) [\mathbf{I} - r\tau\mathbf{W}^F]^{-1} \mathbf{w}.$$

Constraints:  $\mathbf{M}_{ij}^{\text{pot}/\text{dep}} \in [0, 1]$ ,  $\sum_j \mathbf{M}_{ij}^{\text{pot}/\text{dep}} = 1$ .

Eigenmode decomposition:

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

$$\overline{\text{SNR}}(\tau) = \sqrt{N} \sum_a \frac{\mathcal{I}_a}{1 + r\tau/\tau_a},$$

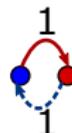
## Section 3

### Upper bounds

# Upper bounds on measures of memory

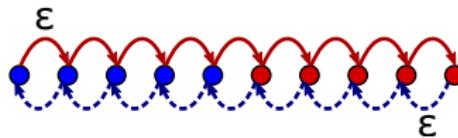
Initial SNR:

$$\text{SNR}(0) = \overline{\text{SNR}}(0) \leq \sqrt{N}.$$



Area under curve:

$$\mathcal{A} = \int_0^\infty dt \text{ SNR}(t) = \lim_{\tau \rightarrow \infty} \tau \overline{\text{SNR}}(\tau) \leq \sqrt{N}(M - 1)/r.$$



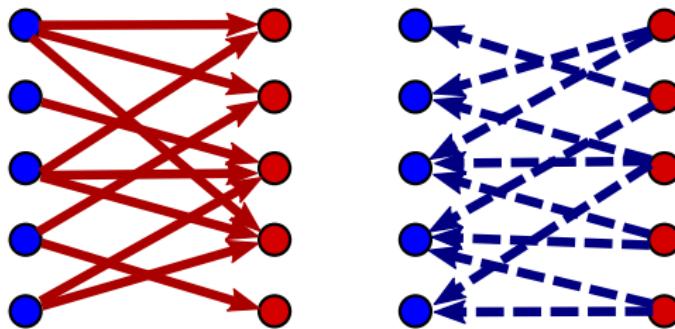
[Lahiri and Ganguli (2013)]

# Initial SNR as flux

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

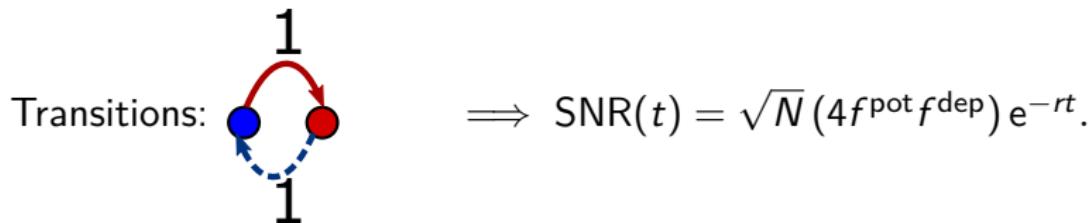
$$\text{SNR}(0) \leq \frac{4\sqrt{N}}{r} \Phi_{-+}.$$

Max when potentiation guarantees  $\mathbf{w} \rightarrow +1$ ,  
depression guarantees  $\mathbf{w} \rightarrow -1$ .



## Two-state model

Two-state model equivalent to previous slide:



Maximal initial SNR:

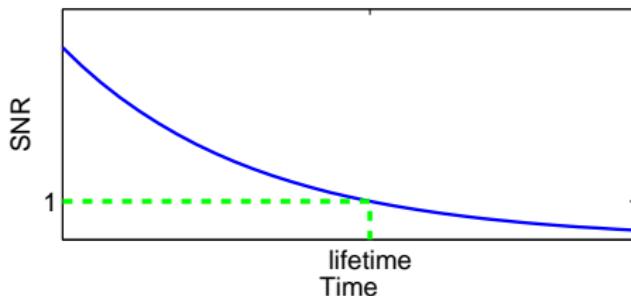
$$\text{SNR}(0) \leq \sqrt{N}.$$

# Area under memory curve

$$\mathcal{A} = \int_0^\infty dt \text{ SNR}(t), \quad \overline{\text{SNR}}(\tau) \rightarrow \frac{\mathcal{A}}{\tau} \quad \text{as} \quad \tau \rightarrow \infty.$$

Area bounds memory lifetime:

$$\begin{aligned}\text{SNR(lifetime)} &= 1 \\ \implies \text{lifetime} &< \mathcal{A}.\end{aligned}$$



This area has an upper bound:

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

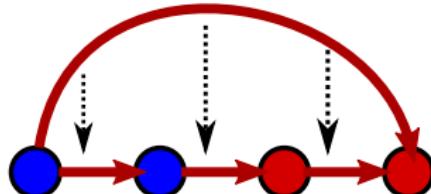
Saturated by a model with linear chain topology.

# Proof of area bound

For any model, we can construct perturbations that

- preserve equilibrium distribution,
- increase area.

details



e.g. decrease “shortcut” transitions, increase bypassed “direct” ones.  
Endpoint: linear chain

The area of this model is

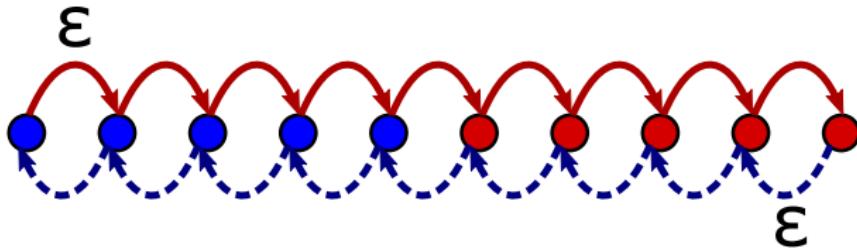
$$A = \frac{2\sqrt{N}}{r} \sum_k \mathbf{p}_k^\infty |k - \langle k \rangle|.$$

Max: equilibrium probability distribution concentrated at both ends.

[Barrett and van Rossum (2008)]

## Saturating model

Make end states “sticky”



Has long decay time, but terrible initial SNR.

$$\lim_{\varepsilon \rightarrow 0} A = \sqrt{N}(M - 1)/r.$$

## Section 4

Envelope memory curve

# Bounding finite time SNR

SNR curve:

$$\overline{\text{SNR}}(\tau) = \sqrt{N} \sum_a \frac{\mathcal{I}_a}{1 + r\tau/\tau_a},$$

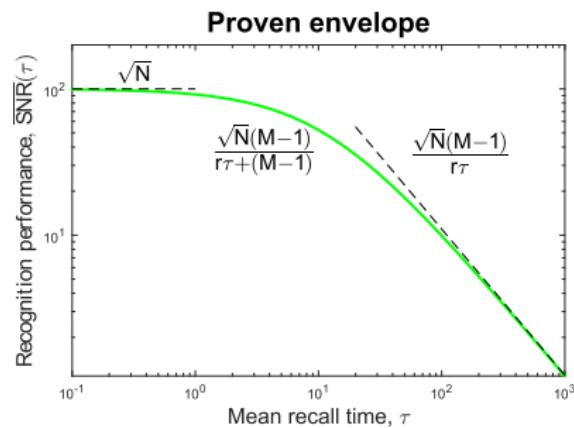
subject to constraints:

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

We can maximise wrt.  $\mathcal{I}_a, \tau_a$ .

# Proven envelope: memory frontier

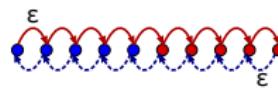
Upper bound on memory curve at *any* time.



Initial SNR:



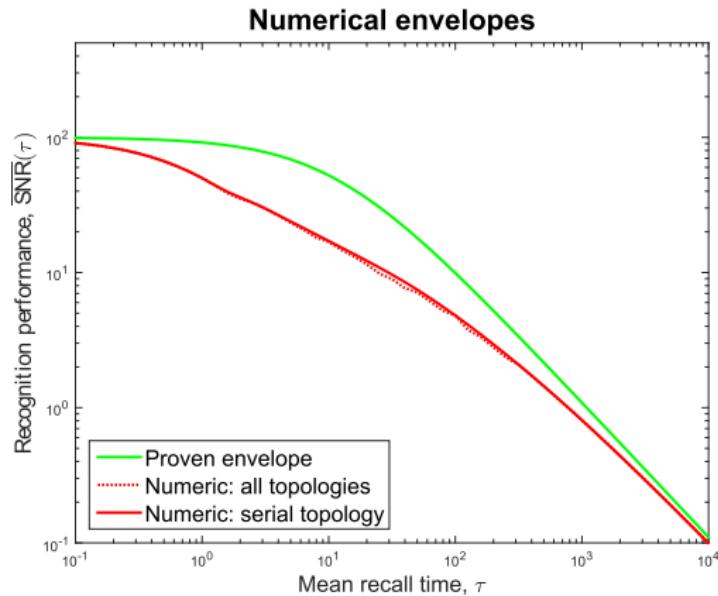
Late times:



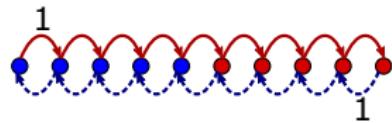
[Lahiri and Ganguli (2013)]

No model can ever go above this envelope. Is it achievable?

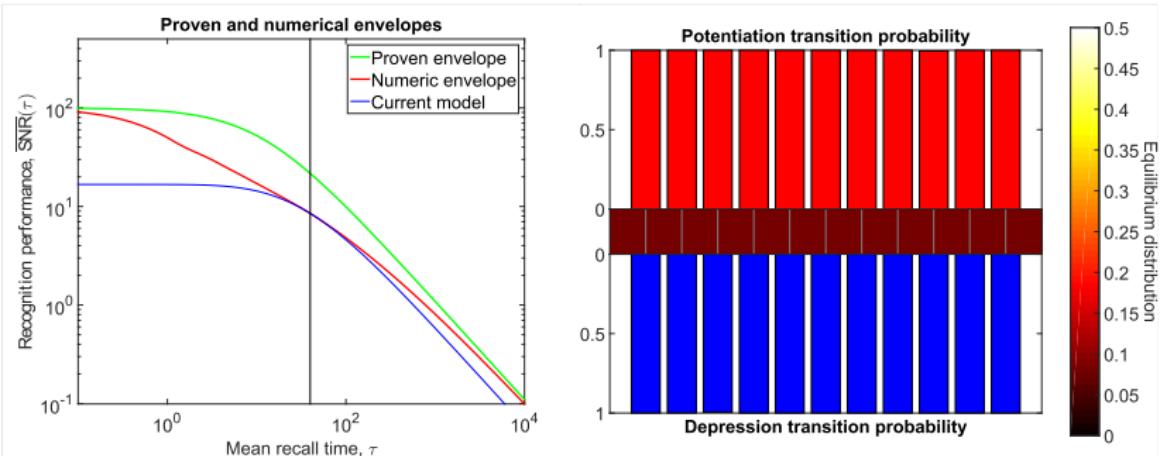
## Achievable envelope



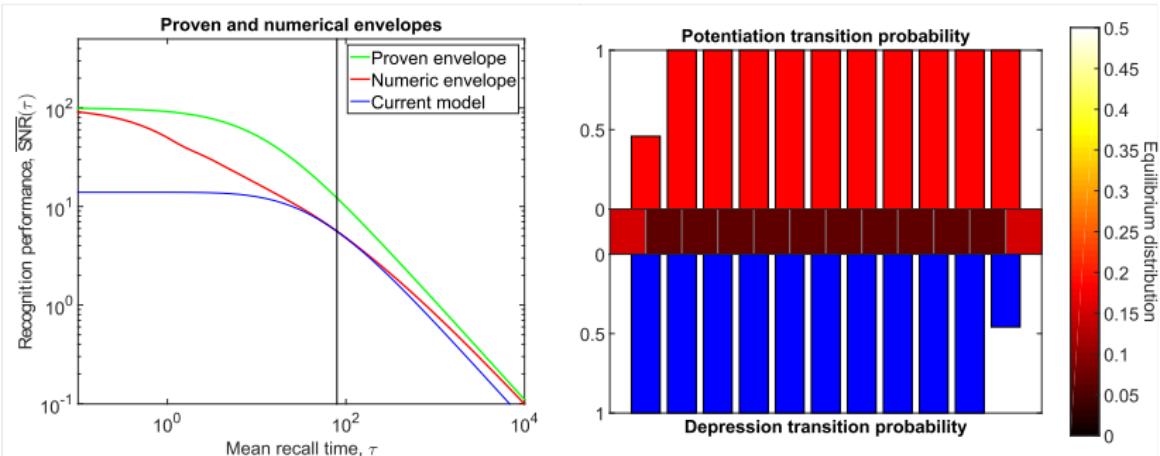
## Serial topology:



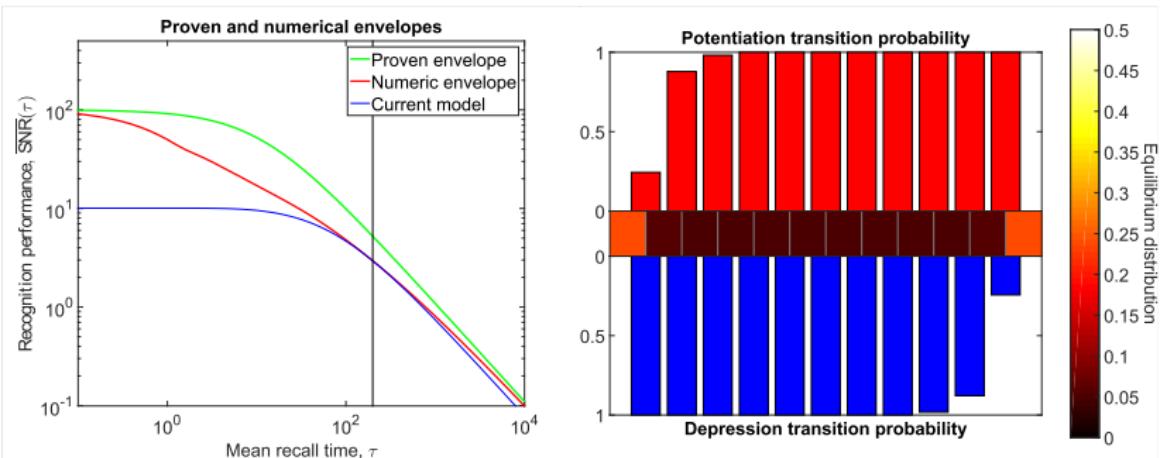
# Models that maximise memory for one timescale



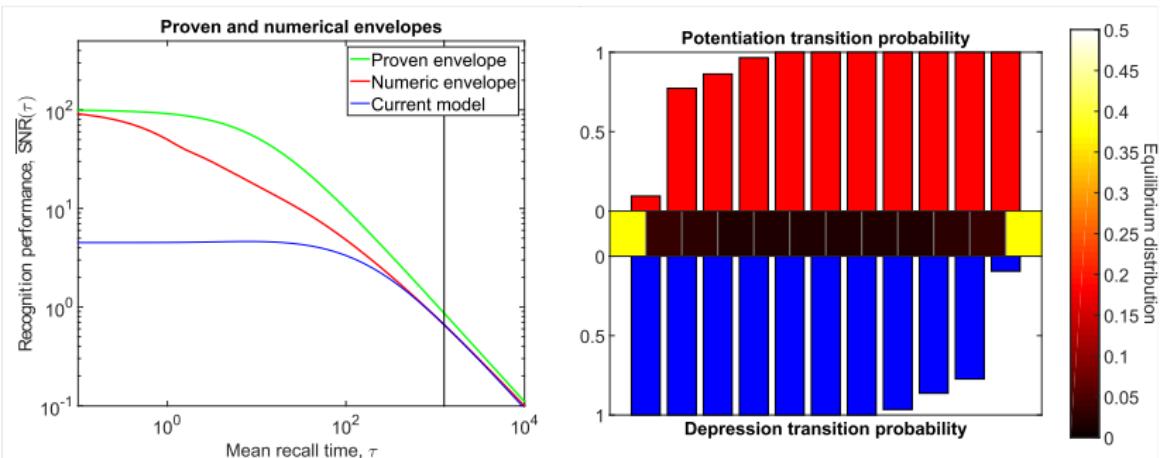
# Models that maximise memory for one timescale



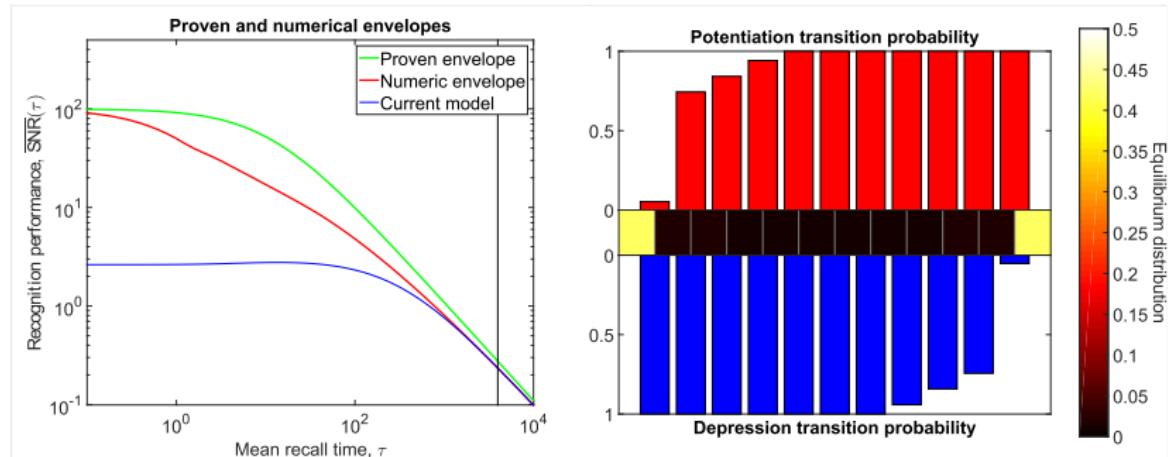
# Models that maximise memory for one timescale



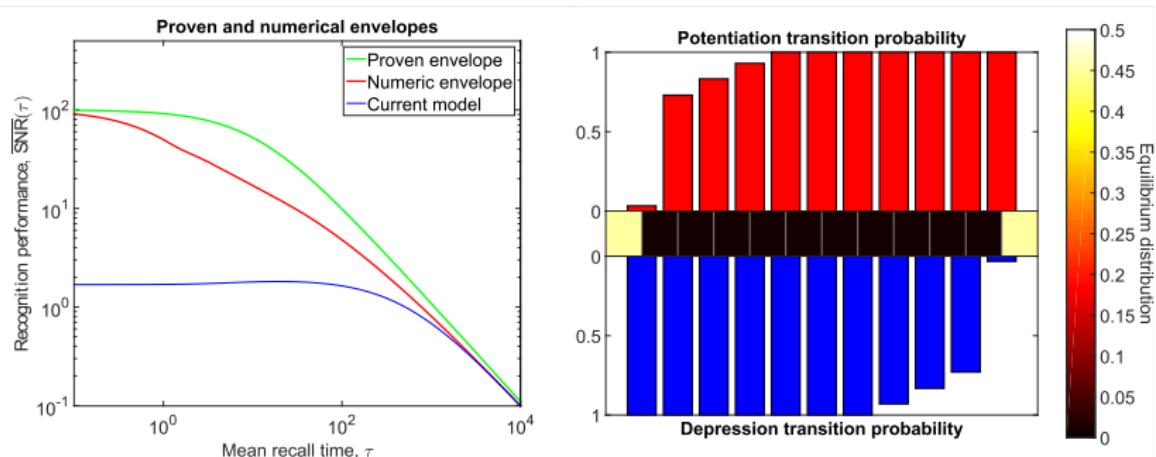
# Models that maximise memory for one timescale



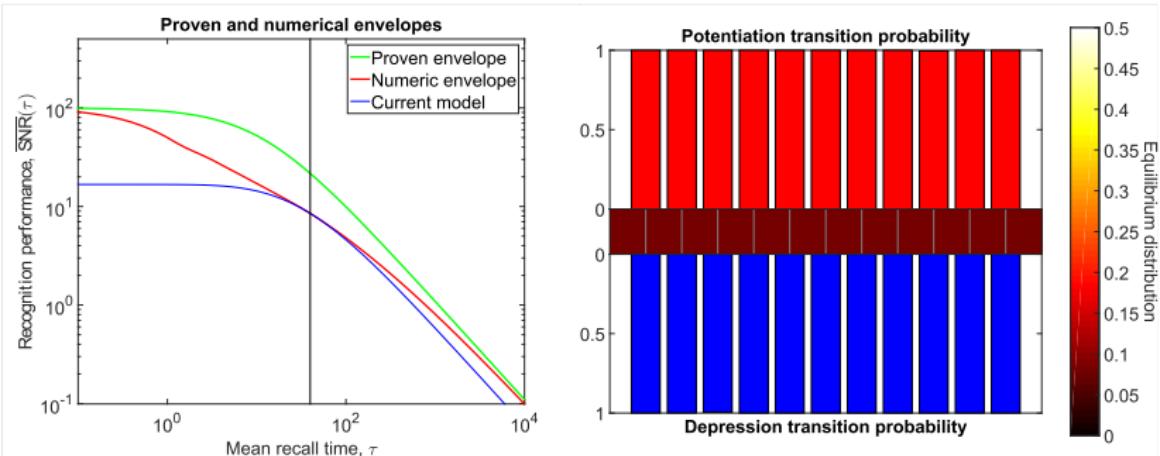
# Models that maximise memory for one timescale



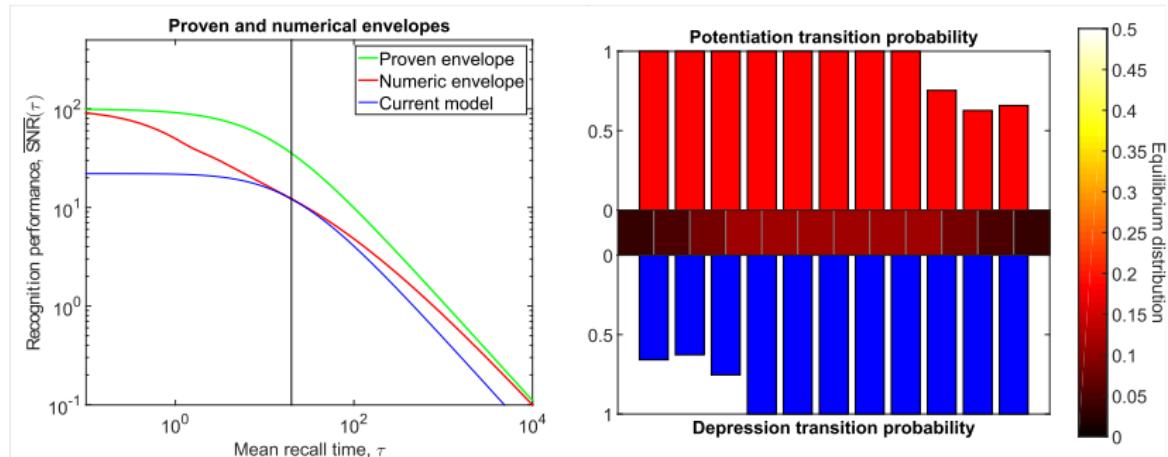
# Models that maximise memory for one timescale



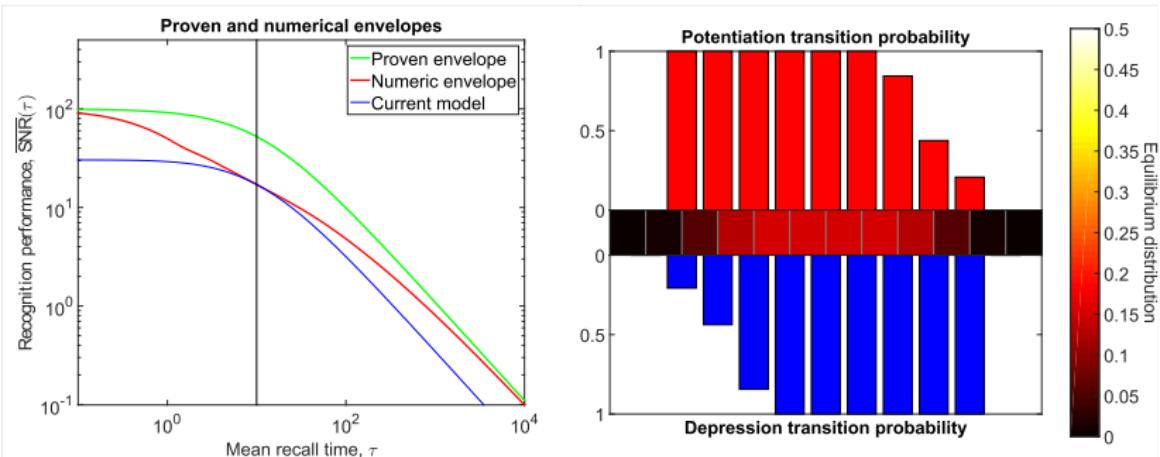
# Models that maximise memory for one timescale



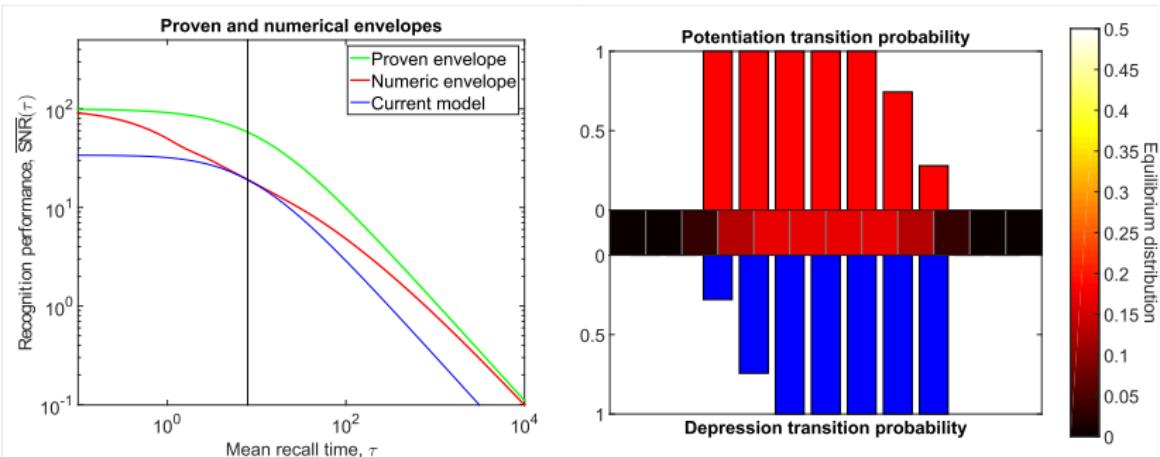
# Models that maximise memory for one timescale



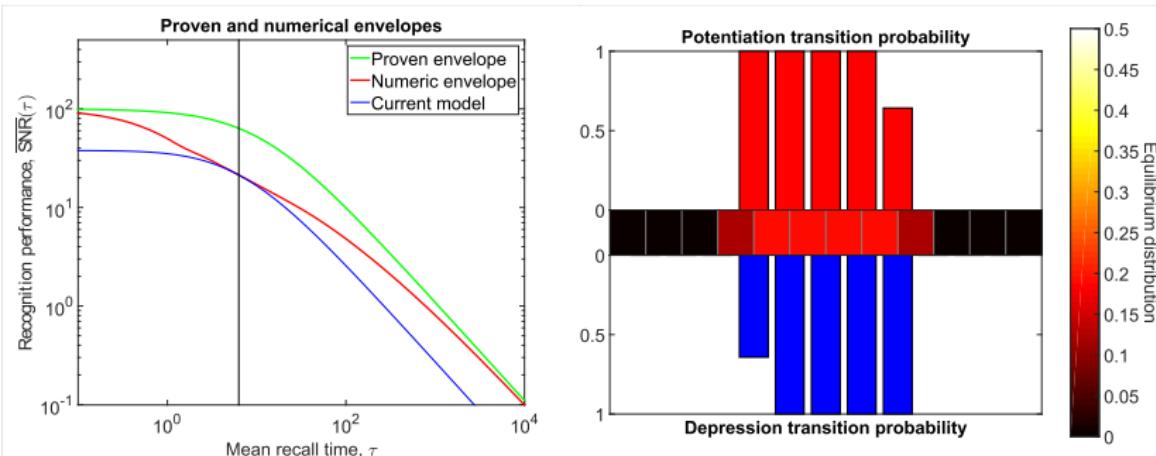
# Models that maximise memory for one timescale



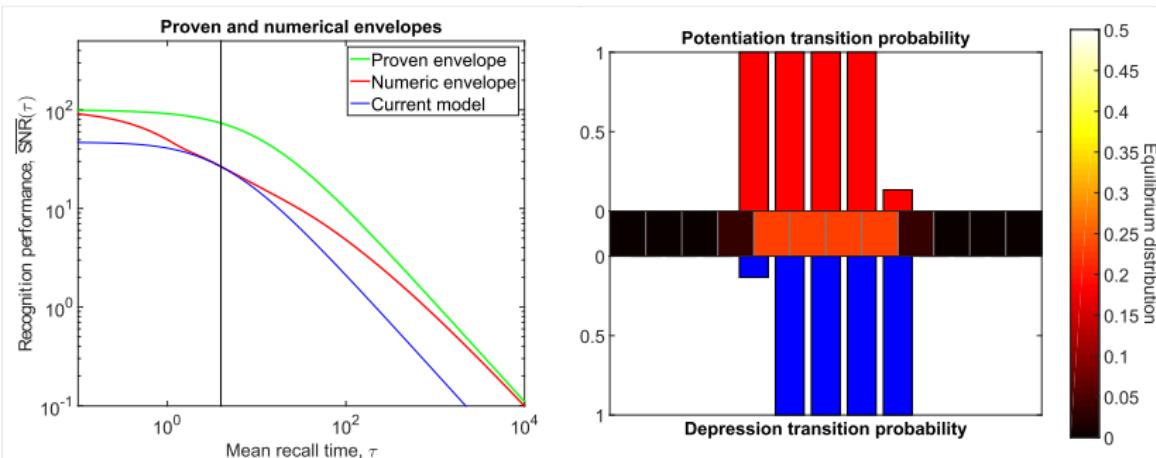
# Models that maximise memory for one timescale



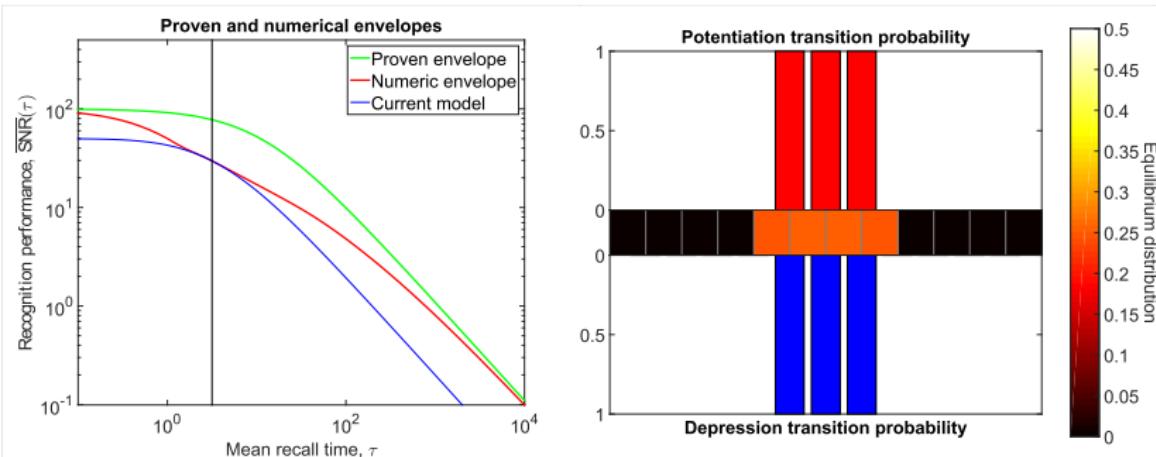
# Models that maximise memory for one timescale



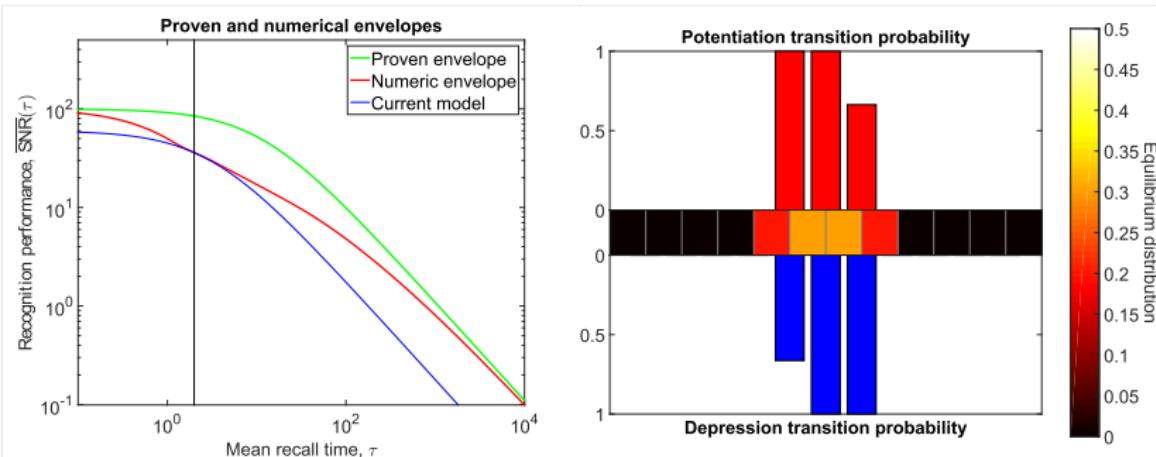
# Models that maximise memory for one timescale



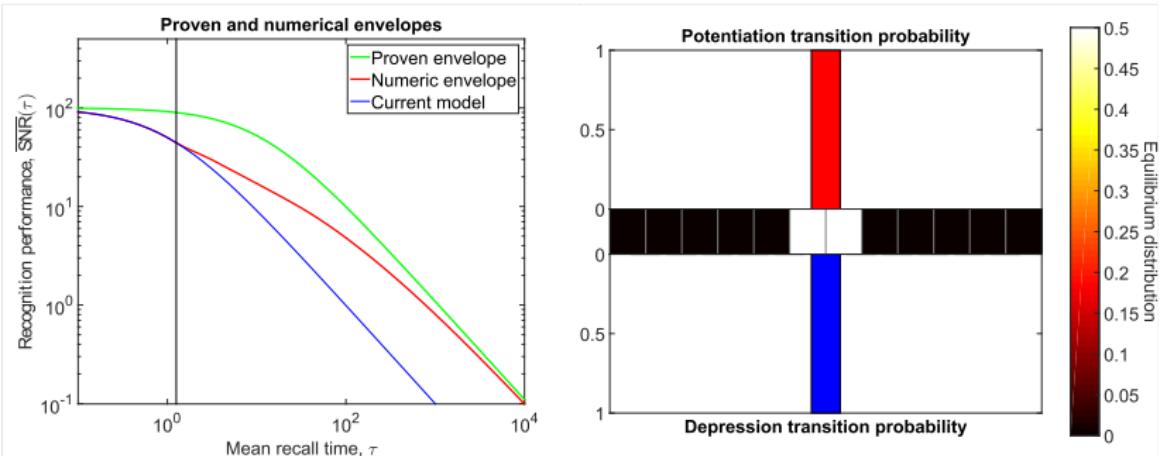
# Models that maximise memory for one timescale



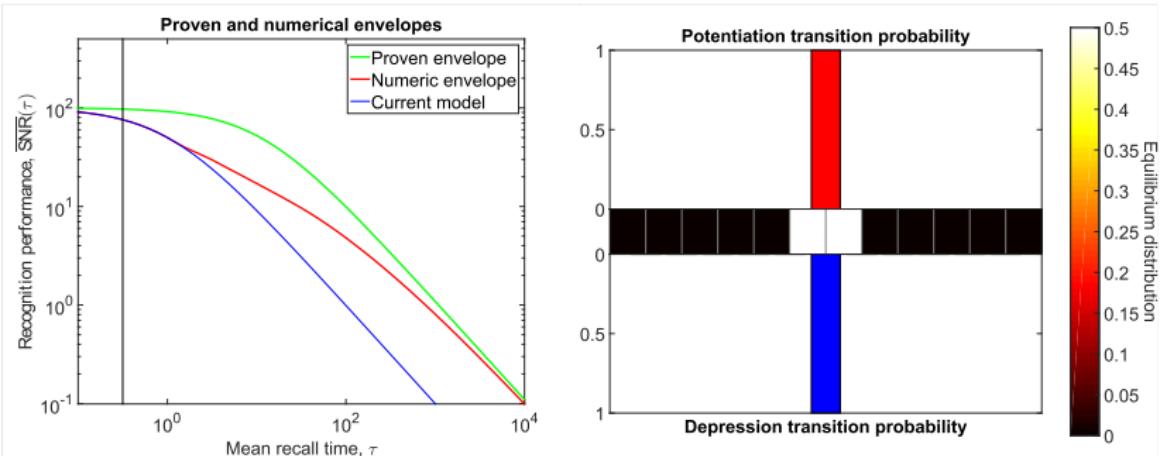
# Models that maximise memory for one timescale



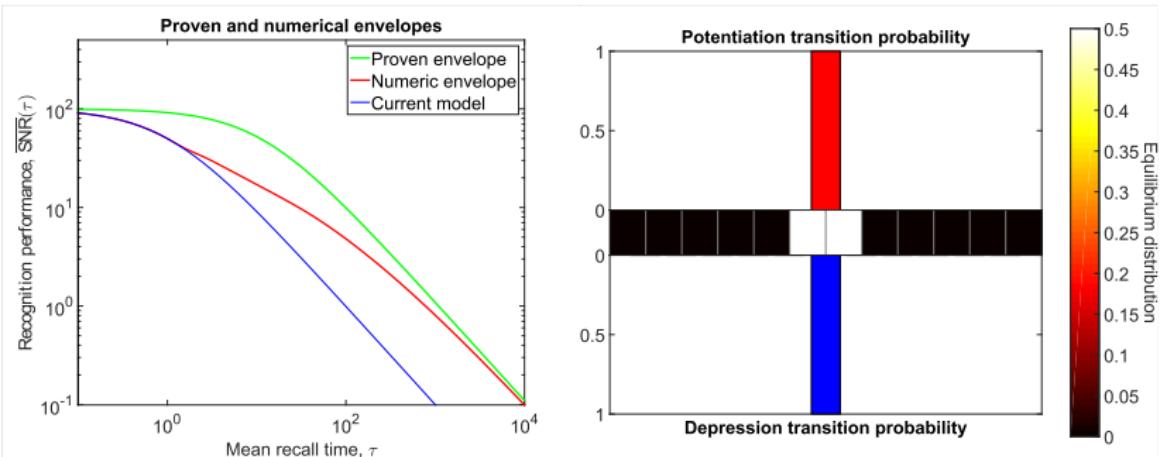
# Models that maximise memory for one timescale



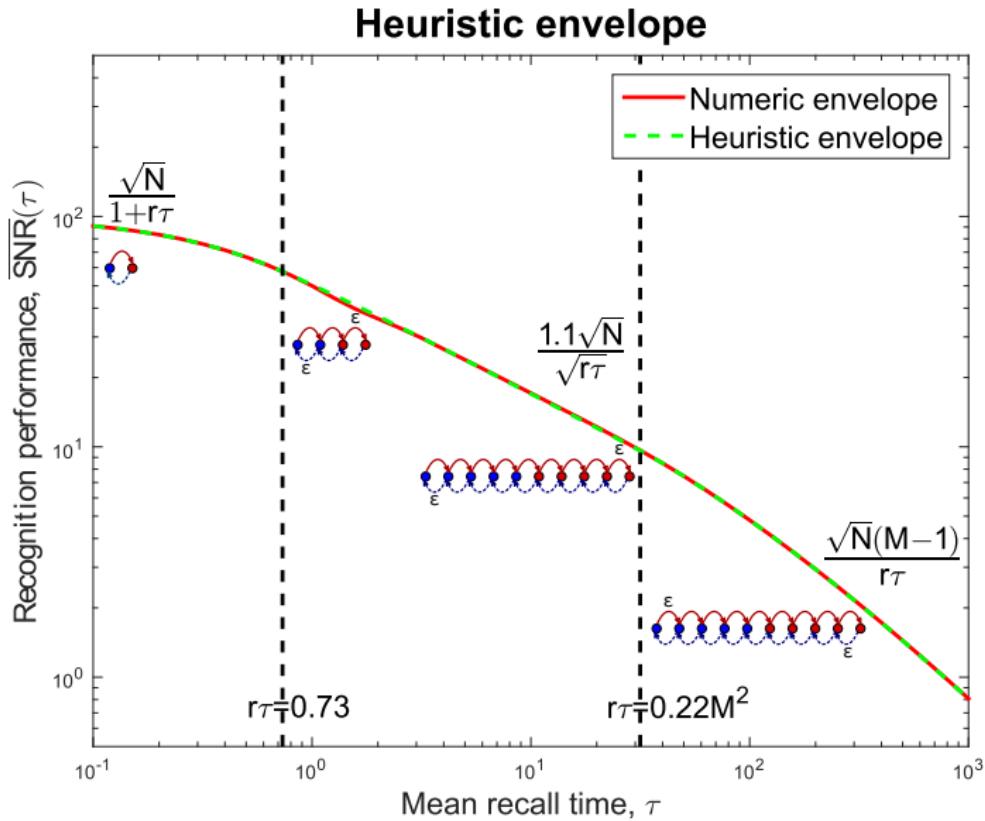
# Models that maximise memory for one timescale



# Models that maximise memory for one timescale



# Heuristic envelope

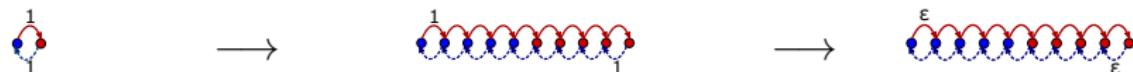


# Synaptic structures for different timescales of memory

Real synapses limited by molecular building blocks.  
Evolution had larger set of priorities.

What can we conclude?

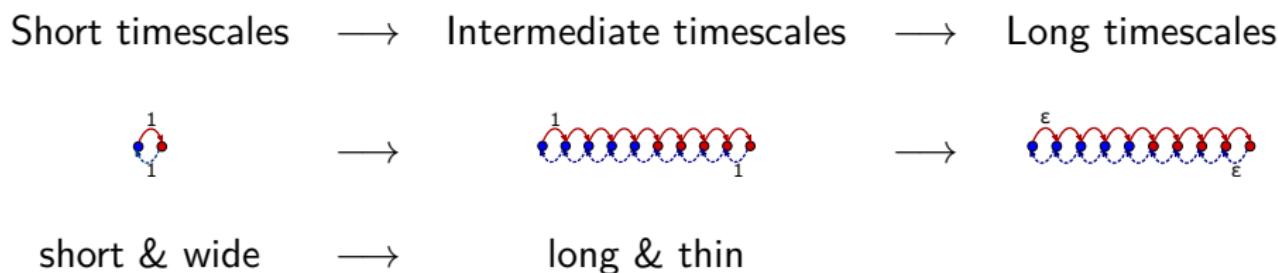
Short timescales → Intermediate timescales → Long timescales



# Synaptic structures for different timescales of memory

Real synapses limited by molecular building blocks.  
Evolution had larger set of priorities.

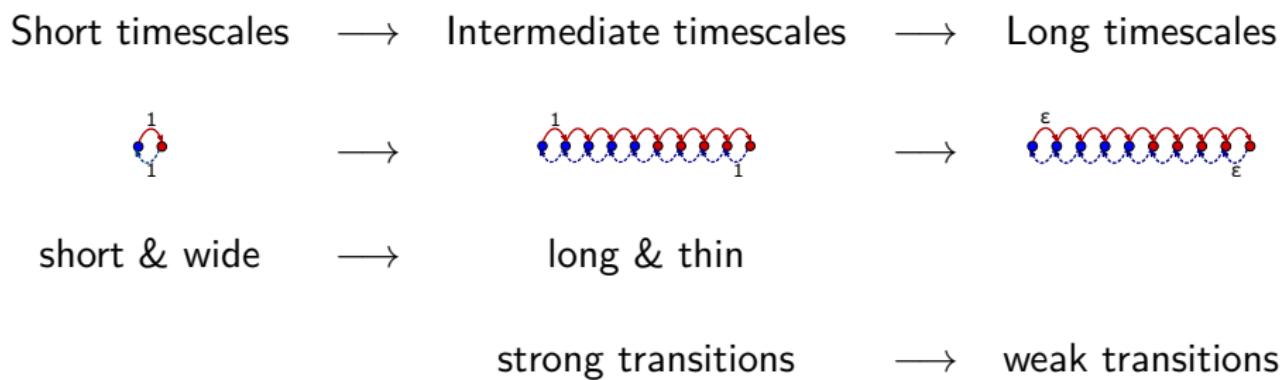
What can we conclude?



# Synaptic structures for different timescales of memory

Real synapses limited by molecular building blocks.  
Evolution had larger set of priorities.

What can we conclude?

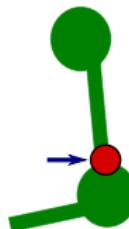


## Section 5

Experimental tests?

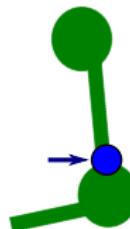
# Proposed Experimental design

Subject a synapse to a sequence of candidate plasticity events.  
Observe the changes in synaptic efficacy.



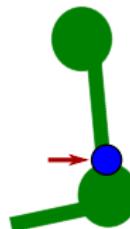
# Proposed Experimental design

Subject a synapse to a sequence of candidate plasticity events.  
Observe the changes in synaptic efficacy.



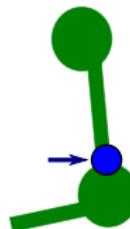
# Proposed Experimental design

Subject a synapse to a sequence of candidate plasticity events.  
Observe the changes in synaptic efficacy.



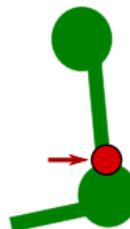
# Proposed Experimental design

Subject a synapse to a sequence of candidate plasticity events.  
Observe the changes in synaptic efficacy.



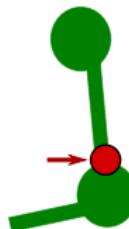
# Proposed Experimental design

Subject a synapse to a sequence of candidate plasticity events.  
Observe the changes in synaptic efficacy.



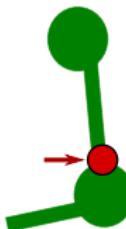
# Proposed Experimental design

Subject a synapse to a sequence of candidate plasticity events.  
Observe the changes in synaptic efficacy.



# Proposed Experimental design

Subject a synapse to a sequence of candidate plasticity events.  
Observe the changes in synaptic efficacy.



## EM algorithms:

Sequence of hidden states → estimate transition probabilities

Transition probabilities → estimate sequence of hidden states

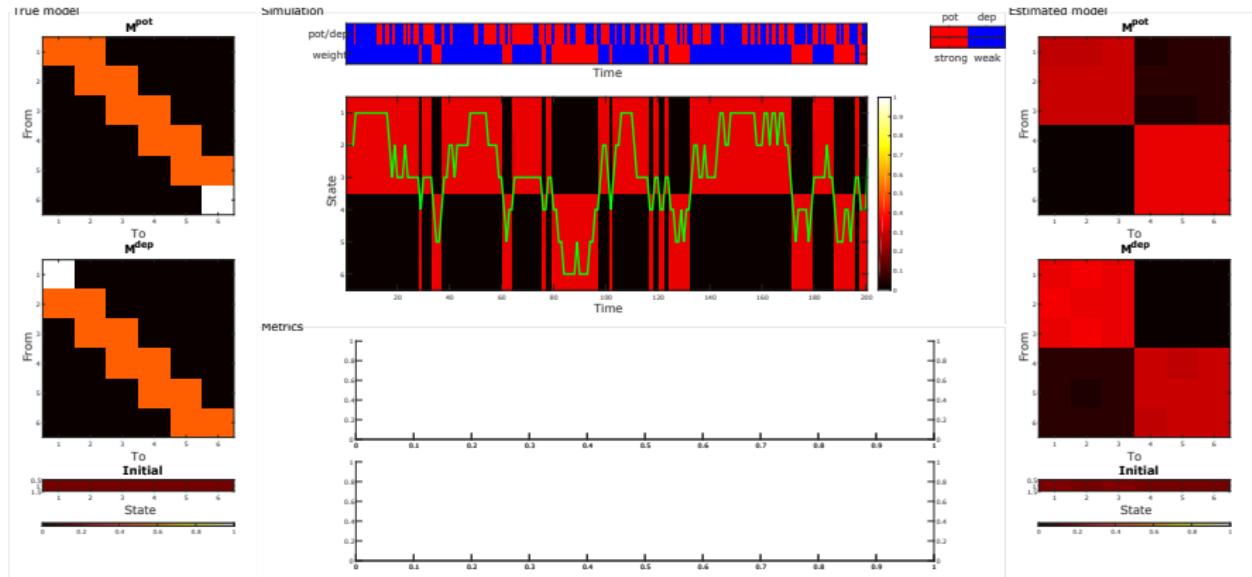
[Baum et al. (1970), Rabiner and Juang (1993), Dempster et al. (2007)]

## Spectral algorithms:

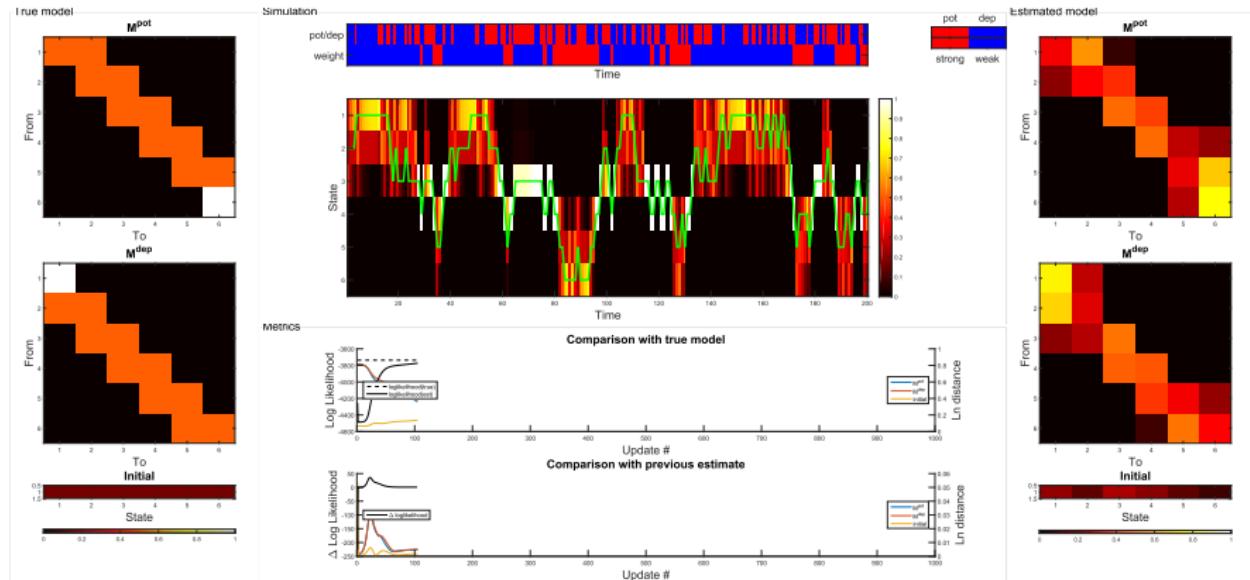
Compute  $P(w_1)$ ,  $P(w_1, w_2)$ ,  $P(w_1, w_2, w_3), \dots$  from data,  
from model.

[Hsu et al. (2008)]

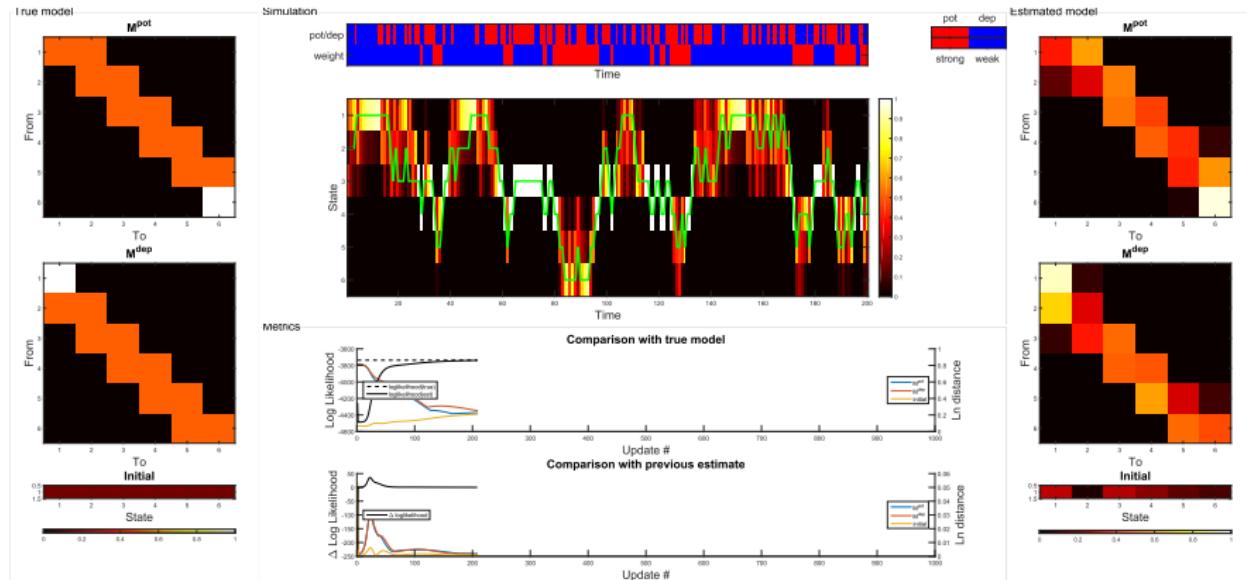
# Fitting algorithm



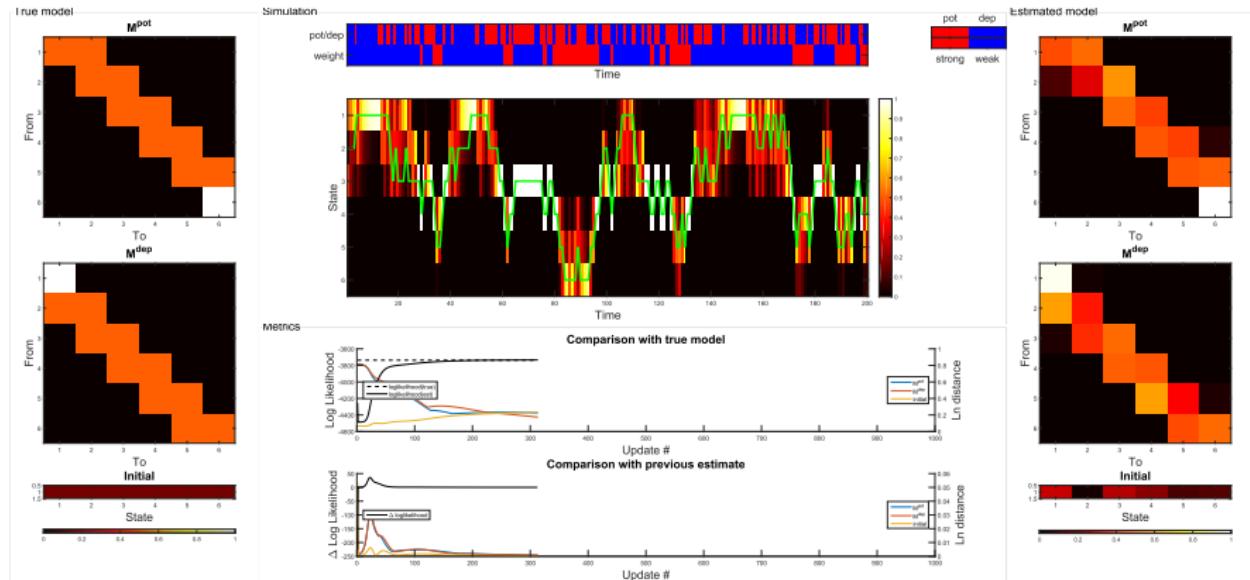
# Fitting algorithm



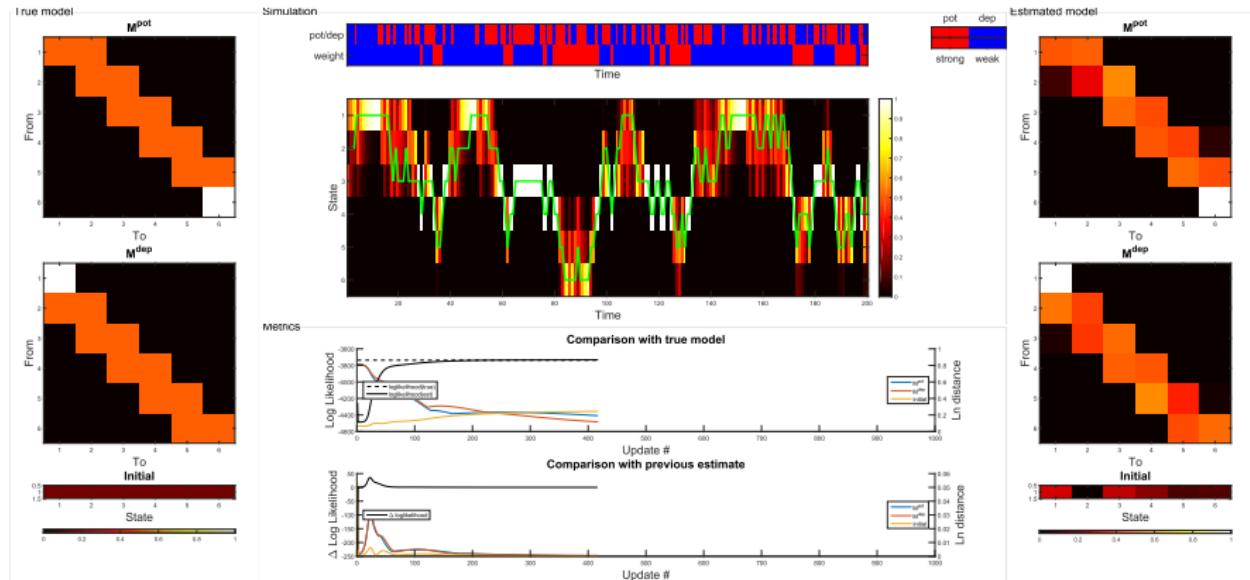
# Fitting algorithm



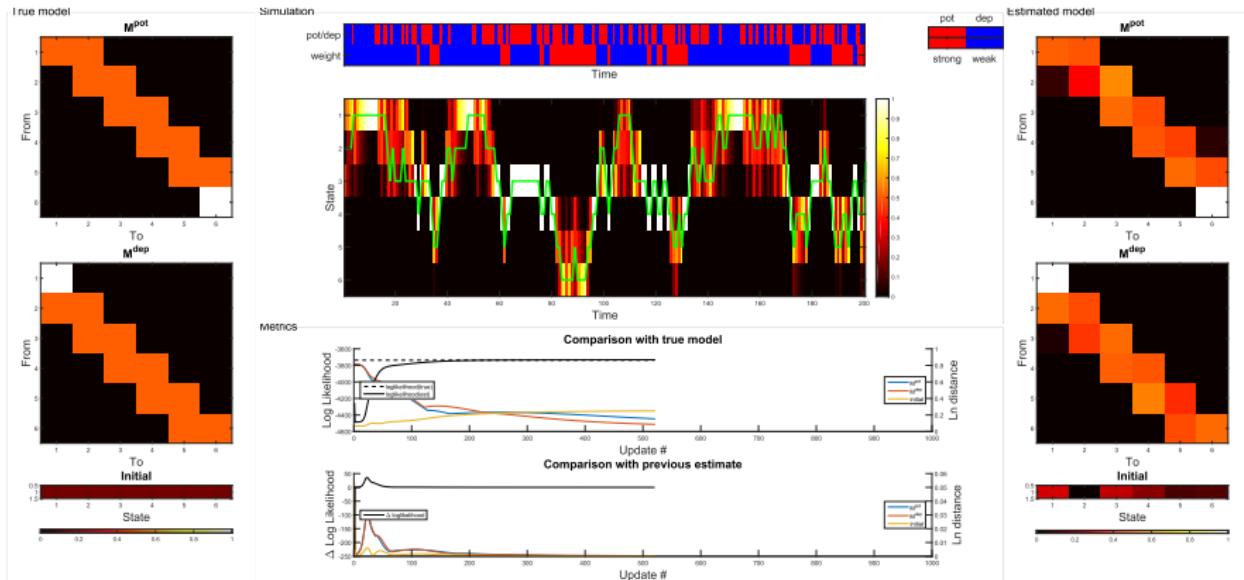
# Fitting algorithm



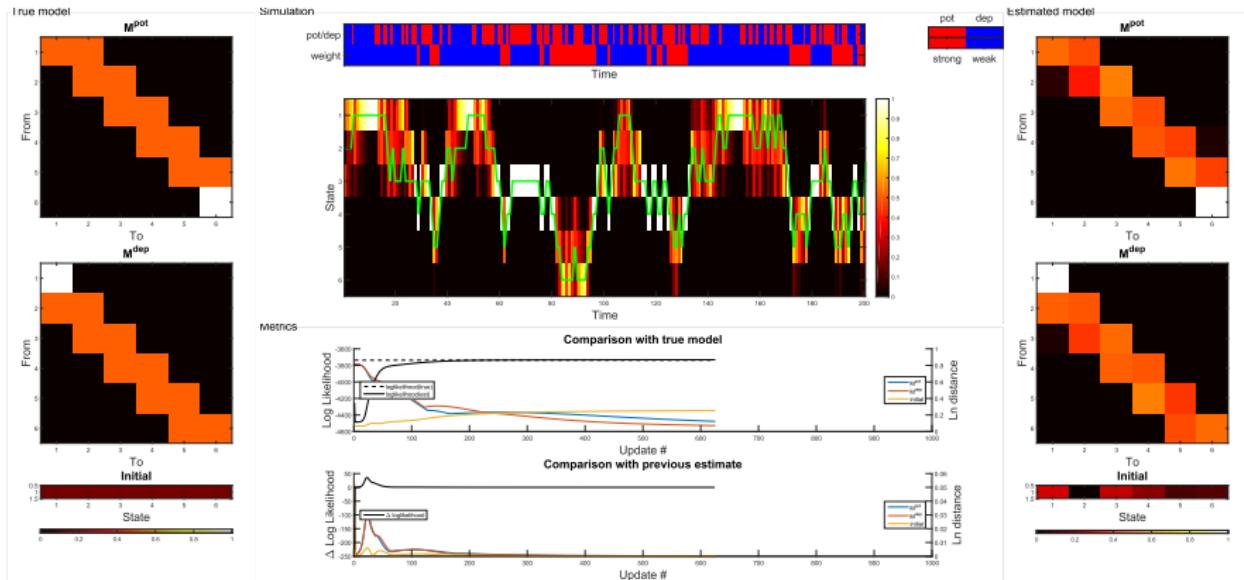
# Fitting algorithm



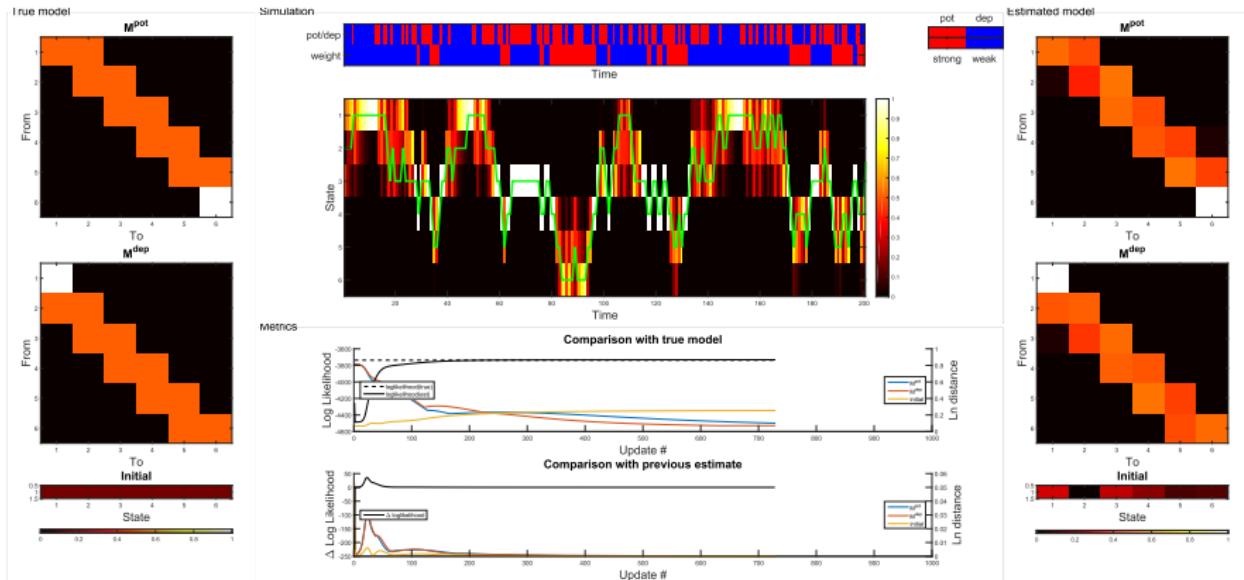
# Fitting algorithm



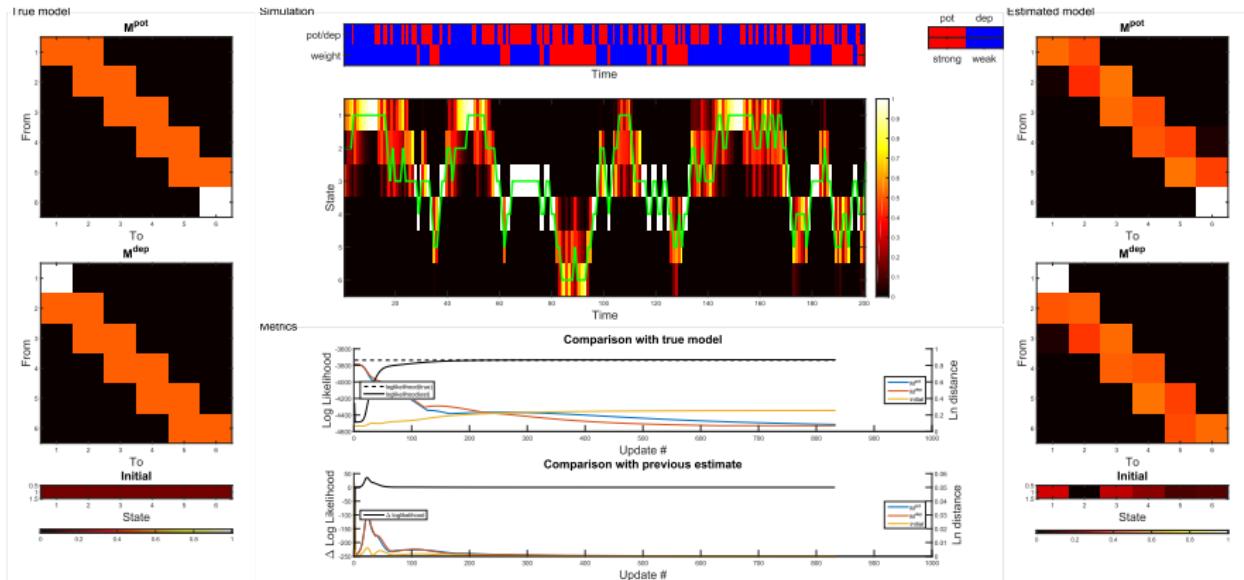
# Fitting algorithm



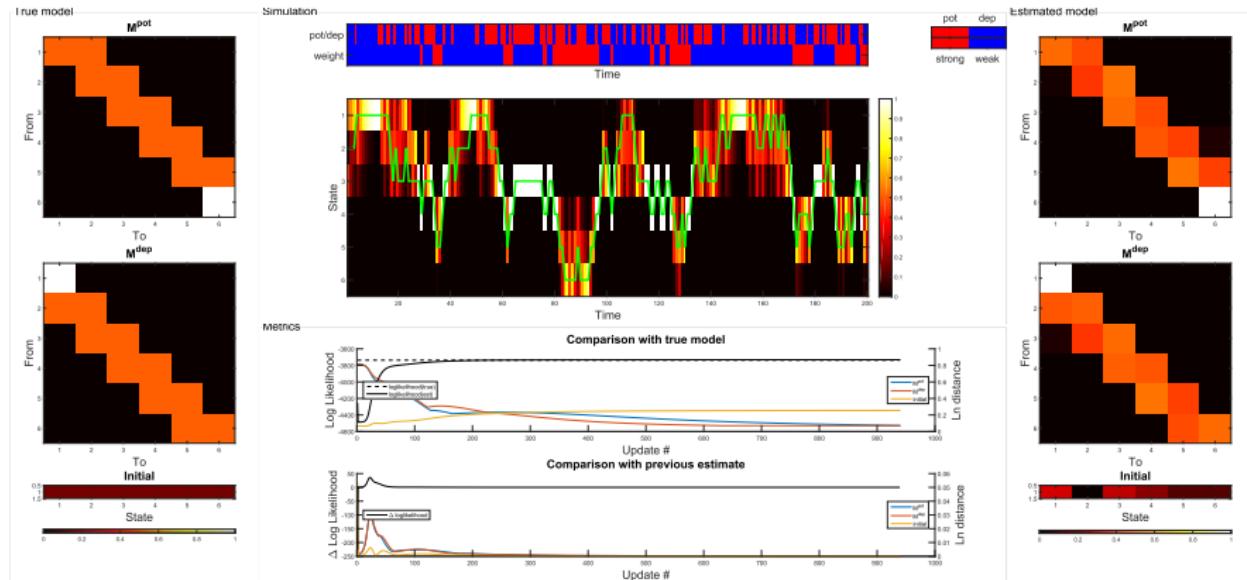
# Fitting algorithm



# Fitting algorithm



# Fitting algorithm



# Experimental problems

- Need single synapses.
- Need long sequences of plasticity events.
- Need to control types of candidate plasticity events.
- Need to measure synaptic efficacies.

When we patch the postsynaptic neuron → Ca washout.

# Summary

- We have formulated a general theory of learning and memory with complex synapses.
- The area under the memory curve of any model < linear chain with same equilibrium 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  $> \mathcal{O}(M)$ .
- We understood which types of synaptic structure are useful for storing memories for different timescales.
- Gap between envelope and what we can achieve at early times?
- Trade-off between SNR at different times?

# Acknowledgements

Thanks to:

- Surya Ganguli
- Stefano Fusi
- Marcus Benna
- David Sussillo
- Jascha Sohl-Dickstein

# References I



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) .

5



Johanna M. Montgomery and Daniel V. Madison.

“State-Dependent Heterogeneity in Synaptic Depression between Pyramidal Cell Pairs”.

*Neuron*, 33(5):765 – 777, (2002) .

5

## References II



Larry R Squire and Pablo Alvarez.

“Retrograde amnesia and memory consolidation: a neurobiological perspective”.

*Current Opinion in Neurobiology*, 5(2):169–177, (April, 1995) .

6



John W Krakauer and Reza Shadmehr.

“Consolidation of motor memory.”.

*Trends in neurosciences*, 29(1):58–64, (January, 2006) .

6



Richard D. Emes and Seth G.N. Grant.

“Evolution of Synapse Complexity and Diversity”.

*Annual Review of Neuroscience*, 35(1):111–131, (2012) .

6

## References III



Carl C. H. Petersen, Robert C. Malenka, Roger A. Nicoll, and John J. Hopfield.

“All-or-none potentiation at CA3-CA1 synapses”.

*Proc. Natl. Acad. Sci. U.S.A.*, 95(8):4732–4737, (1998) .

9



Daniel H. O'Connor, Gayle M. Wittenberg, and Samuel S.-H. Wang.

“Graded bidirectional synaptic plasticity is composed of switch-like unitary events”.

*Proc. Natl. Acad. Sci. U.S.A.*, 102(27):9679–9684, (2005) .

9



D. J. Amit and S. Fusi.

“Constraints on learning in dynamic synapses”.

*Network: Computation in Neural Systems*, 3(4):443–464, (1992) .

9

# References IV



D. J. Amit and S. Fusi.

“Learning in neural networks with material synapses”.

*Neural Computation*, 6(5):957–982, (1994) .

9



S. Fusi, P. J. Drew, and L. F. Abbott.

“Cascade models of synaptically stored memories”.

*Neuron*, 45(4):599–611, (Feb, 2005) .

21

36



S. Fusi and L. F. Abbott.

“Limits on the memory storage capacity of bounded synapses”.

*Nat. Neurosci.*, 10(4):485–493, (Apr, 2007) .

21

## References V

 A. B. Barrett and M. C. van Rossum.  
“Optimal learning rules for discrete synapses”.  
*PLoS Comput. Biol.*, 4(11):e1000230, (Nov, 2008) .

21 45

 Christian Leibold and Richard Kempter.  
“Sparseness Constrains the Prolongation of Memory Lifetime via Synaptic Metaplasticity”.  
*Cerebral Cortex*, 18(1):67–77, (2008) .

36

 Daniel D Ben-Dayan Rubin and Stefano Fusi.  
“Long memory lifetimes require complex synapses and limited sparseness”.  
*Frontiers in computational neuroscience*, 1(November):1–14, (2007) .

36

# References VI



Subhaneil Lahiri and Surya Ganguli.

“A memory frontier for complex synapses”.

In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 1034–1042. NIPS, 2013.

41 49



LE Baum, T Petrie, George Soules, and Norman Weiss.

“A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains”.

*The annals of mathematical statistics*, 41(1):164–171, (1970) .

73 74 75 76 77 78 79

## References VII



Lawrence R Rabiner and Biing-Hwang Juang.

*Fundamentals of speech recognition*, volume 14 of *Signal Processing*.

Prentice Hall, Inc., Upper Saddle River, NJ, USA, 1993.

ISBN 0-13-015157-2.

73 74 75 76 77 78 79



A. P. Dempster, N. M. Laird, and D. B. Rubin.

“Maximum Likelihood from Incomplete Data via the EM Algorithm”.

*Journal of the Royal Statistical Society. Series B (Methodological)*, (October, 2007) .

73 74 75 76 77 78 79



Daniel Hsu, Sham M. Kakade, and Tong Zhang.

“A Spectral Algorithm for Learning Hidden Markov Models”.

pages 1460–1480, (nov, 2008) , arXiv:0811.4413.

73 74 75 76 77 78 79

# References VIII



J.G. Kemeny and J.L. Snell.

*Finite markov chains.*

Springer, 1960.

101

## Techinical detail: ordering states

Let  $\mathbf{T}_{ij}$  = mean first passage time from state  $i$  to state  $j$ . Then:

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

is independent of the initial state  $i$  (Kemeney's constant).

[Kemeny and Snell (1960)]

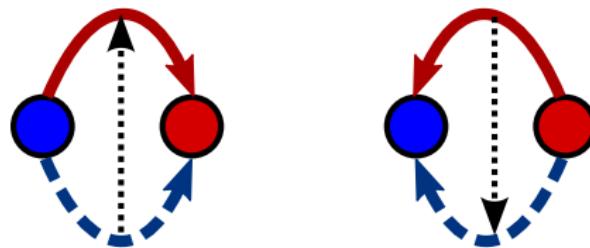
We define:

$$\eta_i^+ = \sum_{j \in \text{strong}} \mathbf{T}_{ij} \mathbf{p}_j^\infty, \quad \eta_i^- = \sum_{j \in \text{weak}} \mathbf{T}_{ij} \mathbf{p}_j^\infty.$$

They can be used to arrange the states in an order (increasing  $\eta^-$  or decreasing  $\eta^+$ ). [back](#)

## Technical detail: upper/lower triangular

With states in order:



Endpoint: potentiation goes right, depression goes left.

[back](#)