A memory frontier for complex synapses

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

Blah blah blah.

1 Introduction

sec:intro

In the field of theoretical neuroscience there are several classic results regarding the storage capacity of synaptic memory, e.g. the Hopfield model's capacity is proportional to the number of neurons, or the Perceptron's capacity is proportional to the number of synapses. However, these rely on one unrealistic feature of these models: the synaptic weights are allowed to take a large number of different values.

Measurements of real synapses have shown that the synaptic efficacy can only take a small number of discrete values, possibly as few as two [1–3] (see [4] for dissenting results). Restricting synaptic weights to a small set of values has a catastrophic effect on the storage capacity of these theoretical models, growing only logarithmically with the size of the system if all else is held fixed [5, 6]. Due to newer memories overwriting older ones, there is a trade-off between learning and remembering: if synapses are too plastic older memories will be wiped out too easily, if they are too rigid it becomes difficult to learn new memories in the first place.

This can be improved when we take into account the fact that there is a complex dynamical system underlying synaptic plasticity (e.g. see [1, 7, 8]). Allowing the synapse to have hidden functional states, with potentiation and depression causing transitions between these states, can dramatically improve their performance.

Much of the previous work on these types of synaptic models has focussed on specific models, to understand their ability to store memories [9–11]. Two examples of these models are shown in Figure 1. we see that they have different memory properties. The serial model has performs very well at one particular timescale, but it performs poorly at other times. The cascade model does not perform quite as well at that time, but it maintains it performance over a wider range of timescales.

We are going to take a different approach. We are going to consider the *entire* space of these models to find upper bounds on the capability of *any* of them. As the performance of a synaptic model is quantified by a memory curve (describing the decay of a memory trace as more and more subsequent memories are stored) rather than a single number, there is no good notion of optimality. What type of memory curve is best will depend on what the synapses are required to do and over what timescales.

We begin in §2 by describing the class of models that we will consider and define the memory curve that we use to quantify their performance. In §3 we will prove upper bounds on two quantities, the initial value of the memory curve and the area under it, as well as finding the models that saturate them. We will use these bounds in §4 to find an envelope curve – one that cannot be crossed by the memory curve of any model at any time.
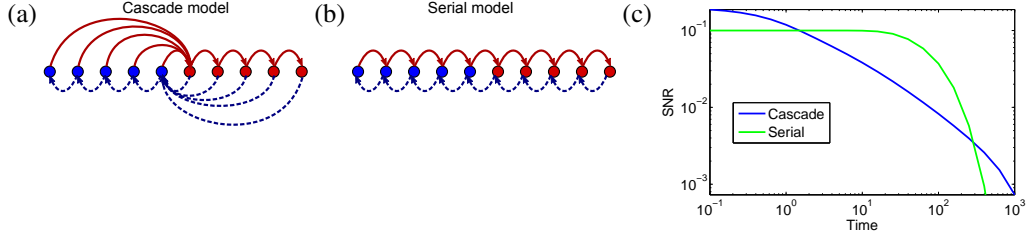
Figure 1: Models of complex synapse. (a) The cascade model of [9], showing transitions between states of high/low synaptic weight (red/blue circles) due to potentiation/depression (solid red/dashed blue arrows). (b) The serial model of [11]. (c) The memory curves of these two models, showing the decay of the signal-to-noise ratio (to be defined in §2) as subsequent memories are stored.

2 Synaptic models and their memory curves

In this section, we will describe the class of models of synaptic plasticity that we are studying and how we quantify their performance as memory storage devices. In the subsequent sections, we will find upper bounds on this performance.

We use a well established formalism for the study of learning and memory with complex synapses (see [9, 10, 12]). In this approach, potentiating and depressing plasticity events occur at random times, with all information about the neural activity and learning rules responsible for them absorbed into their rates. We assume that there are no spatial or temporal correlations in the pattern of potentiating and depressing events, and these events cause Markovian transitions between the internal states of the synapses. As a result of these assumptions, the states of different synapses will be independent, and the system can be fully described by the probability distribution across these states, which we will indicate with the row-vector $\mathbf{p}(t)$.

We also employ an "ideal observer" approach to the memory readout, where the synaptic weights are read directly. This provides an upper bound to the quality of any readout using neural activity. These synaptic weights will be restricted to two values, which we can shift and scale to ± 1.

For any single memory, there will be an ideal pattern of synaptic weights, the N-element vector $\vec{w}_{\text{ideal}}$, that is $+1$ at all synapses that were potentiated and -1 at all synapses that were depressed. The actual pattern of synaptic weights at some later time, t, will be $\vec{w}(t)$. We can use the overlap between these, $\vec{w}_{\text{ideal}} \cdot \vec{w}(t)$, as a measure of the quality of the memory. As $t \to \infty$, the system will return to its steady state distribution which will be uncorrelated with the memory in question. The probability distribution of the quantity $\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty)$ can be used as a "null model" for comparison.

The extent to which the memory has been stored is described by a signal-to-noise ratio (SNR):

$$\text{SNR}(t) = \frac{\langle \vec{w}_{\text{ideal}} \cdot \vec{w}(t) - \vec{w}_{\text{ideal}} \cdot \vec{w}(\infty) \rangle}{\sqrt{\text{Var}(\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty))}}. \tag{1}$$

The noise is essentially $\sqrt{N}$. There is a correction when potentiation and depression are imbalanced, but this will not affect the upper bounds that we will discuss below and will be ignored in the subsequent formulae.

All of the preceding plasticity events will put the system in its steady-state distribution, $\mathbf{p}^\infty$. The memory we are tracking at $t = 0$ will change this to $\mathbf{p}^\infty \mathbf{M}^{\text{pot}}$ in those synapses that are potentiated and $\mathbf{p}^\infty \mathbf{M}^{\text{dep}}$ in those synapses that are depressed, where $\mathbf{M}^{\text{pot/dep}}$ are $M \times M$ matrices of transition probabilities for the Markov processes describing potentiation and depression. As the potentiating/depressing nature of the subsequent memories is independent of $\vec{w}_{\text{ideal}}$, we can average over all sequences, resulting in the evolution of the probability distribution:

$$\frac{d\mathbf{p}(t)}{dt} = r\mathbf{p}(t)\mathbf{W}^{\text{F}}, \qquad \text{where} \quad \mathbf{W}^{\text{F}} = f^{\text{pot}}\mathbf{M}^{\text{pot}} + f^{\text{dep}}\mathbf{M}^{\text{dep}} - \mathbf{I}. \tag{2}$$

Here, r is the rate of the Poisson process describing the timing of plasticity events, $f^{\text{pot/dep}}$ is the fraction of events that are potentiating/depressing and $\mathbf{I}$ is the identity matrix.

2

This results in the following SNR

$$\text{SNR}(t) = \sqrt{N} \left(2 f^{\text{pot}} f^{\text{dep}}\right) \mathbf{p}^\infty \left(\mathbf{M}^{\text{pot}} - \mathbf{M}^{\text{dep}}\right) e^{rt\mathbf{W}^{\text{F}}} \mathbf{w}, \tag{3}$$ `eq:SNRcalc`

where the weight of the synapse when it is in its i'th state is given by the corresponding element of the column vector $\mathbf{w}$. We will frequently refer to this function as the memory curve.

The parameters must satisfy the following constraints:

$$\mathbf{M}_{ij}^{\text{pot/dep}} \in [0,1], \qquad f^{\text{pot/dep}} \in [0,1], \qquad \mathbf{p}^\infty \mathbf{W}^{\text{F}} = 0, \qquad \mathbf{w}_i = \pm 1,$$

$$\sum_j \mathbf{M}_{ij}^{\text{pot/dep}} = 1, \qquad f^{\text{pot}} + f^{\text{dep}} = 1, \qquad \sum_i \mathbf{p}_i^\infty = 1. \tag{4}$$ `eq:constr`

The upper bounds on $\mathbf{M}_{ij}^{\text{pot/dep}}$ and $f^{\text{pot/dep}}$ follow automatically from the other constraints.

The question is: what do these constraints imply for the memory curve above? In practice, to make any statements about finite times, we need to got to the eigenmode description:

$$\mathbf{W}^{\text{F}} = \sum_a -q_a \mathbf{u}^a \mathbf{v}^a, \quad \mathbf{v}^a \mathbf{u}^b = \delta_{ab}, \quad \mathbf{W}^{\text{F}} \mathbf{u}^a = -q_a \mathbf{u}^a, \quad \mathbf{v}^a \mathbf{W}^{\text{F}} = -q_a \mathbf{v}^a. \tag{5}$$ `eq:eigendecomp`

Where q_a are the negative of the eigenvalues, $\mathbf{u}^a$ are the right (column) eigenvectors and $\mathbf{v}^a$ are the left (row) eigenvectors. This allow us to write

$$\text{SNR}(t) = \sqrt{N} \sum_a \mathcal{I}_a e^{-rt/\tau_a}, \qquad \text{where} \qquad \mathcal{I}_a = \left(2 f^{\text{pot}} f^{\text{dep}}\right) \mathbf{p}^\infty (\mathbf{M}^{\text{pot}} - \mathbf{M}^{\text{dep}}) \mathbf{u}^a \mathbf{v}^a \mathbf{w},$$

$$\text{and} \qquad \tau_a = \frac{1}{q_a}. \tag{6}$$ `eq:SNReigen`

We can ask then ask the question: what are the constraints on these quantities implied by the constraints (4)? We will find some of these constraints in the next section.

3 Upper bounds

`sec:bounds`

In this section, we will find some upper bounds on certain properties of the memory curve discussed in the previous section, namely its initial value and the area under it. We will use these bounds to determine an upper bound on the SNR at finite times in the next section.

3.1 Initial SNR

`sec:initial`

Now we will discuss the SNR at $t = 0$:

$$\text{SNR}(0) = \sqrt{N} \left(2 f^{\text{pot}} f^{\text{dep}}\right) \mathbf{p}^\infty \left(\mathbf{M}^{\text{pot}} - \mathbf{M}^{\text{dep}}\right) \mathbf{w}. \tag{7}$$ `eq:init`

We will find an upper bound on this quantity for *all* possible models and also find the model that saturates this bound. As maximizing initial SNR does not place any value on later parts of the memory curve, this will produce models with fast decay that may not be good for memory storage. Nevertheless, the upper bound will be useful in §4.

A useful quantity is the equilibrium probability flux between two disjoint sets of states, $\mathcal{A}$ and $\mathcal{B}$:

$$\boldsymbol{\Phi}_{\mathcal{AB}} = \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{B}} r \mathbf{p}_i^\infty \mathbf{W}_{ij}^{\text{F}}. \tag{8}$$ `eq:flux`

The initial SNR is closely related to the flux between the states with $\mathbf{w}_i = -1$ and those with $\mathbf{w}_j = +1$:

$$\text{SNR}(0) \leq \frac{4\sqrt{N} \boldsymbol{\Phi}_{-+}}{r}. \tag{9}$$ `eq:initflux`

This inequality becomes an equality if potentiation never decreases the synaptic weight and depression never increases it, which should be a property of any sensible model.

To maximize this flux, potentiation from a weak state must be guaranteed to end in a strong state, and depression must do the reverse. An example of such a model is shown in Figure 2(a,b). These
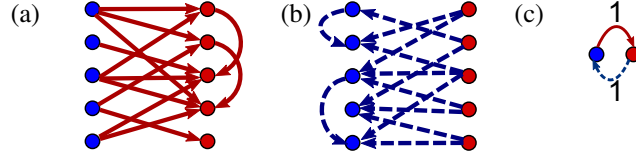
3

Figure 2: Synaptic models that maximize initial SNR. (a) For potentiation, all transitions starting from a weak state lead to a strong state, and the probabilities for all transitions leaving a given weak state sum to 1. (a) Depression is similar to potentiation, but with strong and weak interchanged. (c) The equivalent two state model, with transition probabilities under potentiation and depression equal to one.

models have a property known as "lumpability" (see [13, §6.3] for the discrete time version and [14, 15] for continuous time). They are completely equivalent (i.e. have the same memory curve) as a two state model with transition probabilities equal to 1, as shown in Figure 2(c).

This two state model has the equilibrium distribution $\mathbf{p}^\infty = (f^{\text{dep}}, f^{\text{pot}})$ and its flux is given by $\Phi_{-+} = r f^{\text{pot}} f^{\text{dep}}$. This is maximized when $f^{\text{pot}} = f^{\text{dep}} = \frac{1}{2}$, leading to the upper bound:

$$\text{SNR}(0) \le \sqrt{N}. \tag{10}$$

Whilst this model has high initial SNR, it also has very fast decay – with a timescale $\tau \sim \frac{1}{r}$. As the synapse is very plastic, the initial memory is encoded very easily, but the subsequent memories also overwrite it easily. This is not a good design for a synapse, but the resulting bound on the initial SNR will prove useful in §4.

3.2 Area

Now consider the area under the memory curve:

$$A = \int_0^\infty \mathrm{d}t \ \text{SNR}(t). \tag{11}$$

We will find an upper bound on this quantity as well as the model that saturates it. The area is useful as it also provides an upper bound on the lifetime of a memory, $\tau(\epsilon)$, defined as the time at which $\text{SNR}(\tau) = \epsilon$ (we often take $\epsilon = 1$):

$$A = \int_0^\infty \mathrm{d}t \ \text{SNR}(t) > \int_0^{\tau(\epsilon)} \mathrm{d}t \ \text{SNR}(t) > \int_0^{\tau(\epsilon)} \mathrm{d}t \ \epsilon = \epsilon \tau(\epsilon). \tag{12}$$

However, the area still receives a contribution from the tail of the memory curve, after the SNR has decayed below any useful value. This means that maximizing SNR will favor models with long tails that will not necessarily be good for storing memories.

One important property of the area is that it has a null "scaling" degree of freedom. If all off-diagonal elements of $\mathbf{M}^{\text{pot/dep}}$ are multiplied by a constant, with diagonal elements adjusted to keep row-sums fixed, this will have two effects. First, it will scale up the initial signal creation. Second, it will scale down time. Combining these two effects, it will leave the area invariant.

This feature allows us to ignore the lower bound on the diagonal elements, $\mathbf{M}^{\text{pot/dep}}_{ii} > 0$. If we maximize the area ignoring this constraint, the "scaling" mode can be used to satisfy it without changing the area.

We also develop an important organizing principle using the theory of first passage times. The mean first passage time matrix, $\overline{\mathbf{T}}_{ij}$, is defined as the average time it takes to reach state j for the first time, starting from state j. The diagonal elements are defined to be zero. Then the quantity

$$\eta \equiv \sum_j \overline{\mathbf{T}}_{ij} \mathbf{p}_j^\infty \tag{13}$$

is independent of the starting state i. It is known as Kemeny's constant (see [13, §4.4]). We can define analogous quantities

$$\eta_i^+ = \sum_{j \in +} \overline{\mathbf{T}}_{ij} \mathbf{p}_j^\infty, \qquad \eta_i^- = \sum_{j \in -} \overline{\mathbf{T}}_{ij} \mathbf{p}_j^\infty. \tag{14}$$
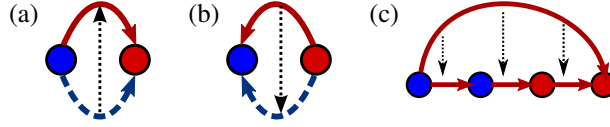
Figure 3: Perturbations that increase the area. (a) Perturbations that increase elements of $\mathbf{M}^{\text{pot}}$ above the diagonal and decrease the corresponding elements of $\mathbf{M}^{\text{dep}}$. It can no longer be used when $\mathbf{M}^{\text{dep}}$ is lower triangular. (b) Perturbations that decrease elements of $\mathbf{M}^{\text{pot}}$ below the diagonal and increase the corresponding elements of $\mathbf{M}^{\text{dep}}$. It can no longer be used when $\mathbf{M}^{\text{pot}}$ is upper triangular. (c) Perturbation that decreases "shortcut" transitions and increases the bypassed "direct" transitions. It can no longer be used when there are only nearest-neighbor "direct" transitions.
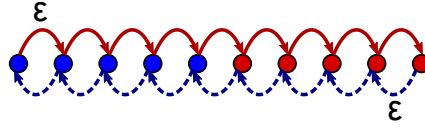

Figure 4: Model that maximizes the area. Unlabeled transitions have probability 1. Labeled transitions have probability $\varepsilon \to 0$.

These can be thought of as the average time it takes to reach the strong/weak states respectively. We can put the states in order of decreasing η_i^+ or increasing η_i^-. Because $\eta_i^+ + \eta_i^- = \eta$ is independent of i, the two orderings are the same.

With the states in this order, we can find perturbations of $\mathbf{M}^{\text{pot/dep}}$ that will always increase the area, whilst leaving the equilibrium distribution, $\mathbf{p}^\infty$, unchanged. Some of these perturbations are shown in Figure 3. The only thing that can prevent these perturbations from increasing the area is when they require the decrease of a matrix element that has already been set to 0. This determines the topology (non-zero transition probabilities) of the model with maximal area. It is of the form shown in Figure 4, with potentiation moving one step to the right and depression moving one step to the left. Any other topology would allow some perturbation to further increase the area.

As these perturbations do not change the equilibrium distribution, this means that the area of *any* model is bounded by that of a linear chain with the same equilibrium distribution. This area is given by

$$A \le \frac{2\sqrt{N}}{r} \sum_k \left[k - \sum_j j\mathbf{p}_j^\infty \right] \mathbf{p}_k^\infty \mathbf{w}_k = \frac{2\sqrt{N}}{r} \sum_k \left| k - \sum_j j\mathbf{p}_j^\infty \right| \mathbf{p}_k^\infty, \tag{15}$$

where we chose $\mathbf{w}_k = \text{sgn}\left[k - \sum_j j\mathbf{p}_j^\infty \right]$. We can then maximize this by pushing all of the equilibrium distribution symmetrically to the two end states. This can be done by reducing the transition probabilities out of these states, as in Figure 4. This makes it very difficult to get out of the end states once they have been entered. The resulting area is

$$A \le \frac{\sqrt{N}(M-1)}{r}. \tag{16}$$

This result is similar to that found numerically by Barrett and van Rossum [12] in a slightly different situation.

The "sticky" end states result in very slow decay, but they also make it difficult to encode the memory in the first place. The memory curve for this model starts off very low, but hovers above zero for a very long time. This is also not a good design for a synapse, but the resulting bound on the area will also prove useful in §4.

4 Memory curve envelope

Now we will look at the implications of the upper bounds found in the previous section for the SNR at finite times. As argued in (6), the memory curve can be written in the form

$$\text{SNR}(t) = \sqrt{N} \sum_a \mathcal{I}_a \text{e}^{-rt/\tau_a}. \tag{17}$$

The upper bounds on the initial SNR, (10), and the area, (16), imply the following constraints on the parameters $\{\mathcal{I}_a, \tau_a\}$:

$$\sum_a \mathcal{I}_a \le 1, \qquad \sum_a \mathcal{I}_a \tau_a \le M - 1. \tag{18}$$

We are not claiming that these are a complete set of constraints: not every set $\{\mathcal{I}_a, \tau_a\}$ that satisfies these inequalities will actually be achievable by a synaptic model. However, any set that violates either inequality will definitely not be achievable.

Now we can pick some fixed time, t_0, and maximize the SNR at that time wrt. the parameters $\{\mathcal{I}_a, \tau_a\}$, subject to the constraints above. This always results in a single nonzero $\mathcal{I}_a$:

$$
\begin{aligned}
t_0 &\le \frac{M-1}{r} &&\implies \text{SNR}(t) = \sqrt{N}\text{e}^{-rt/(M-1)} &&\implies \text{SNR}(t_0) = \sqrt{N}\text{e}^{-rt_0/(M-1)}, \\
t_0 &\ge \frac{M-1}{r} &&\implies \text{SNR}(t) = \frac{\sqrt{N}(M-1)\text{e}^{-t/t_0}}{rt_0} &&\implies \text{SNR}(t_0) = \frac{\sqrt{N}(M-1)}{ert_0}.
\end{aligned}
\tag{19}
$$

Both the initial SNR bound and the area bound are saturated at early times. At late times, only the area bound is saturated. The function $\text{SNR}(t_0)$, the green curve in Figure 5(a), above forms a memory curve envelope. No synaptic model can have an SNR that is greater than this at any time. We can use this to find an upper bound on the memory lifetime, $\tau(\epsilon)$, by finding the point at which the envelope crosses ϵ:

$$\tau(\epsilon) \le \frac{\sqrt{N}(M-1)}{\epsilon er}, \tag{20}$$

where we assume $N > (\epsilon \text{e})^2$.

This leaves the question of whether this bound is achievable. At any time, can we find a model whose memory curve touched the envelope? The red curves in Figure 5(a) show the closest we have come to the envelope with actual models, by repeated numerical optimization with random initialization and by hand designed models.

We see that here is a gap between the upper bound that we can prove and what we can achieve with actual models. There may be other models we haven't found that could beat the ones we have, but it is probably too much to expect them to reach this envelope. That would require saturating the area bound, and we saw that the only model that does that has very low initial SNR. It seems that we need to find additional constraints to add to those we have (18), thereby bringing the envelope down. We do not necessarily need to find *every* constraint. We have already seen that not every constraints is saturated for certain time ranges, so it may be that a small subset of them will be sufficient to find an envelope that is tight.

5 Discussion

References

[1] T. V. P. Bliss and G. L. Collingridge, "A synaptic model of memory: long-term potentiation in the hippocampus," *Nature* **361** (Jan, 1993) 31–39.

[2] C. C. H. Petersen, R. C. Malenka, R. A. Nicoll, and J. J. Hopfield, "All-or-none potentiation at CA3-CA1 synapses," *Proc. Natl. Acad. Sci. U.S.A.* **95** (1998) no. 8, 4732–4737.

[3] D. H. O'Connor, G. M. Wittenberg, and S. S.-H. Wang, "Graded bidirectional synaptic plasticity is composed of switch-like unitary events," *Proc. Natl. Acad. Sci. U.S.A.* **102** (2005) no. 27, 9679–9684.
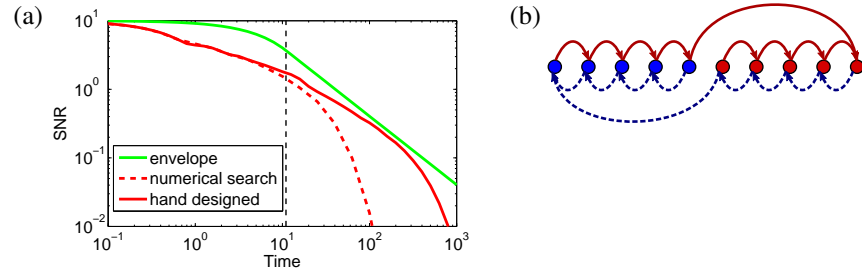
Figure 5: The memory curve envelope for $N = 100$, $M = 12$. (a) An upper bound on the SNR at any time is shown in green. The red dashed curve shows the result of numerical optimization of synaptic models with random initialization. The solid red curve shows the highest SNR we have found with hand designed models. At early times these models are of the form shown in (b) with different numbers of states. At late times they are of the form shown in Figure 4 with different values of ε.

[4] R. Enoki, Y. ling Hu, D. Hamilton, and A. Fine, "Expression of Long-Term Plasticity at Individual Synapses in Hippocampus Is Graded, Bidirectional, and Mainly Presynaptic: Optical Quantal Analysis," *Neuron* **62** (2009) no. 2, 242 – 253.

[5] D. J. Amit and S. Fusi, "Constraints on learning in dynamic synapses," *Network: Computation in Neural Systems* **3** (1992) no. 4, 443–464.

[6] D. J. Amit and S. Fusi, "Learning in neural networks with material synapses," *Neural Computation* **6** (1994) no. 5, 957–982.

[7] D. S. Bredt and R. A. Nicoll, "AMPA Receptor Trafficking at Excitatory Synapses," *Neuron* **40** (2003) no. 2, 361 – 379.

[8] 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** (2009) no. 68, ra19.

[9] S. Fusi, P. J. Drew, and L. F. Abbott, "Cascade models of synaptically stored memories," *Neuron* **45** (Feb, 2005) 599–611.

[10] S. Fusi and L. F. Abbott, "Limits on the memory storage capacity of bounded synapses," *Nat. Neurosci.* **10** (Apr, 2007) 485–493.

[11] C. Leibold and R. Kempter, "Sparseness Constrains the Prolongation of Memory Lifetime via Synaptic Metaplasticity," *Cerebral Cortex* **18** (2008) no. 1, 67–77.

[12] A. B. Barrett and M. C. van Rossum, "Optimal learning rules for discrete synapses," *PLoS Comput. Biol.* **4** (Nov, 2008) e1000230.

[13] J. Kemeny and J. Snell, *Finite markov chains*. Springer, 1960.

[14] C. Burke and M. Rosenblatt, "A Markovian function of a Markov chain," *The Annals of Mathematical Statistics* **29** (1958) no. 4, 1112–1122.

[15] F. Ball and G. F. Yeo, "Lumpability and Marginalisability for Continuous-Time Markov Chains," *Journal of Applied Probability* **30** (1993) no. 3, 518–528.