

# Mean First Passage Memory Lifetimes by Reducing Complex Synapses to Simple Synapses

Terry Elliott<sup>1</sup>

Department of Electronics and Computer Science,  
University of Southampton,  
Highfield,  
Southampton, SO17 1BJ,  
United Kingdom.

**Running Title:** Time-Dependent Simple Synapses.

January 10, 2017.

---

<sup>1</sup>Tel.: +44 (0)23 8059 6000, Fax.: +44 (0)23 8059 2783, E.-mail:  
te@ecs.soton.ac.uk.

## Abstract

Memory models that store new memories by forgetting old ones have memory lifetimes that are rather short and grow only logarithmically in the number of synapses. Attempts to overcome these deficits include “complex” models of synaptic plasticity in which synapses possess internal states governing the expression of synaptic plasticity. Integrate-and-express, filter-based models of synaptic plasticity propose that synapses act as low-pass filters, integrating plasticity induction signals before expressing synaptic plasticity. Such mechanisms enhance memory lifetimes, leading to an initial rise in the memory signal that is in radical contrast to other related, but non-integrative memory models. Because of the complexity of models with internal synaptic states, however, their dynamics can be more difficult to extract compared to “simple” models that lack internal states. Here, we show that by focusing only on processes that lead to changes in synaptic strength, we can integrate out internal synaptic states and effectively reduce complex synapses to simple synapses. For binary-strength synapses, these simplified dynamics then allow us to work directly in the transitions in perceptron activation induced by memory storage rather than in the underlying transitions in synaptic configurations. This permits us to write down master and Fokker-Planck equations that may be simplified under certain, well-defined approximations. These methods allow us to see that memory based on synaptic filters can be viewed as an initial transient that leads to memory signal rise, followed by the emergence of Ornstein-Uhlenbeck-like dynamics that return the system to equilibrium. We may use this approach to compute mean first passage time-defined memory lifetimes for complex models of memory storage.

# 1 Introduction

One approach to circumventing the catastrophic forgetting threshold of the Hopfield model for associative memory (Hopfield, 1982) is to impose bounds on synaptic strength (Nadal *et al.*, 1986; Parisi, 1986). Models of associative memory with bounded strengths learn new memories by forgetting old ones. One particularly appealing approach to bounding synaptic strength is to suppose that synapses exist in only a limited number of discrete states of synaptic strength. Experimental evidence for binary, ternary and even larger numbers of discrete states of synaptic strength exists (Petersen *et al.*, 1998; Montgomery & Madison, 2002, 2004; O’Connor *et al.*, 2005a,b; Bartol *et al.*, 2015) although the interpretation of such evidence can be difficult (Elliott, 2010a), and evidence also supports the possibility that changes in synaptic strength may be discrete, jump-like, all-or-none processes (Yasuda *et al.*, 2003; Bagal *et al.*, 2005; Sobczyk & Svoboda, 2007). Many such memory models and analyses based on discrete synapses in feedforward or recurrent network settings now exist, using a variety of different measures to gauge memory lifetimes (see, for example, Tsodyks, 1990; Amit & Fusi, 1994, Fusi *et al.*, 2005, Leibold & Kempter, 2006, 2008; Rubin & Fusi, 2007; Barrett & van Rossum, 2008; Huang & Amit, 2010, 2011, Lahiri & Ganguli, 2013). Early models are based on “simple” synapses that lack internal states and change strength stochastically with fixed probability in response to memory storage (Tsodyks, 1990). We have termed such a synapse a “stochastic updater”. “Complex” synapses attempt to overcome the problems characteristic of simple synapses by considering internal synaptic states that govern the expression of synaptic plasticity through metaplasticity (see, for example, Fusi *et al.*, 2005; Rubin & Fusi, 2007; Leibold & Kempter, 2008; Lahiri & Ganguli, 2013).

In a developmental setting, we have suggested that synapses may act as low-pass filters (Elliott, 2008) and have shown that by integrating plasticity induction signals and expressing synaptic plasticity only when a filter reaches an upper or lower threshold, fluctuations in large-scale patterns of synaptic connectivity can be powerfully suppressed (Elliott & Lagogiannis, 2009; Elliott, 2011b). When we extend such integrate-and-express, filter-based models of synaptic plasticity to consider memory lifetimes in the context of a feedforward, perceptron-based approach to memory storage with binary-strength synapses, we found that the fidelity of recall of a memory initially *improves* over time, before the memory is slowly forgotten as the system returns to equilibrium (Elliott & Lagogiannis, 2012). This initial enhancement in the memory signal is actually driven by the storage of later memories. These dynamics are in radical contrast to other, non-integrative memory models, in which the memory signal always decays monotonically. These differences result in integrative, filter-based models outperforming non-integrative models in most regions of biologically-relevant parameter space (Elliott, 2016b).

The analysis of complex models of synaptic plasticity can be considerably harder than that of simple models. For example, with a mean first passage time (MFPT) definition of memory lifetimes, it is possible to make considerable progress in deriving exact expressions for memory lifetimes or very good approximations to them for fixed probability stochastic updater synapses (Elliott, 2014). However, an analysis of MFPT-defined memory lifetimes for the full dynamics of complex synapses is likely extremely hard if not intractable. The underlying problem is simply the size of the transition matrices describing changes in the internal states of complex synapses. We have previously shown that considerable analytical progress and much dynamical insight can be achieved by integrating out synapses' internal states and working explicitly

with their changes in strengths (Elliott, 2010b, 2016a). Here, therefore, by focusing only on processes that lead to changes in synaptic strength, we show that the dynamics of complex synapses may be effectively reduced to those of simple synapses for the purposes of studying memory lifetimes. The resulting simple synapses are not, however, stochastic updaters with a fixed or time-independent probability for expressing synaptic plasticity. Instead, they have time- and indeed history-dependent plasticity probabilities. Despite this time- and history-dependence, the reduction of complex synapses to simple synapses without internal states offers considerable analytical benefits, which we explore below.

Our paper is organised as follows. In the next section we summarise our approach to memory lifetimes with a single perceptron using filter-based, binary-strength synapses. In section 3, we perform this reduction of complex, filter-based synapses to simple stochastic updater synapses with time- and history-dependent conditional strength-change probabilities. Then, in section 4 we use these reduced dynamics to derive exact probability and moment generating functions for the activation of the perceptron. Although all these results are exact, much insight can be gained by considering approximations. We develop these in section 5 where we consider the exact master equation governing transitions in perceptron activation, and then consider an approximate form leading to a derivation of a Fokker-Planck equation governing the probability distribution of perceptron activation. In section 6, we use the reduced and approximated dynamics to examine memory lifetimes defined by MFPTs. Finally, in section 7, we discuss our approach and consider future possible work

Parameter or quantity	Description
$N$	Number of synapses.
$S_i(t)$	Strength of synapse $i$ at time $t$ .
$\Theta$	Filter size.
$h(t), h_n$	Tracked memory signal at time $t$ or at memory storage step $n$ .
$\mu(t), \mu_n$	Mean of $h(t)$ or $h_n$ .
$\varphi(t), \varphi_n$	Second moment of $h(t)$ or $h_n$ .
$\sigma(t)$	Standard deviation in $h(t)$ .
$\mathbb{M}^\pm, \mathbb{M}$	$2(2\Theta - 1) \times 2(2\Theta - 1)$ matrices implementing transitions in the joint distribution of a synapse's filter and strength states.
$\mathbf{A}$	$2(2\Theta - 1)$ -dimensional vector describing the equilibrium joint distribution of a synapse's filter and strength states.
$\mathbb{W}^\pm, \mathbb{W}$	$2 \times 2$ matrices implementing transitions in a simple synapse's distribution of strength states.
$p^\pm(t), p_n^\pm$	Conditional strength change probabilities for a complex synapse at time $t$ or at memory storage step $n + 1$ .
$\mathbb{D}_n$	$2 \times 2$ matrix describing $n$ -step transitions in a single synapse's strength.
$X_n, Y_n$	Elements of $\mathbb{D}_n$ , defined in Eq. (3.11).
$\mathcal{G}, \mathcal{M}$	Various symbols representing probability and moment generating functions for perceptron activation.
$\mathbb{H}_n$	$(N + 1) \times (N + 1)$ matrix describing $n$ -step transitions in perceptron activation.
$\mathbb{H}(t)$	Poisson summed form of $\mathbb{H}_n$ relevant to continuous time.

$\mathbb{H}_{n m}$	$(N + 1) \times (N + 1)$ matrix describing transitions in perceptron activation between $h_m$ and $h_n$ , with $n \geq m$ .
$\mathbb{V}_n$	$(N + 1) \times (N + 1)$ matrix describing transitions in perceptron activation between $h_{n-1}$ and $h_n$ .
$\mathbb{V}(t)$	$(N + 1) \times (N + 1)$ matrix generated from $p^\pm(t)$ via Eq. (5.13).
$\mathbf{P}(t), \mathbf{P}_n$	Probability distribution of $h(t)$ or $h_n$ .
$P(h, t)$	Probability distribution of $h$ at time $t$ in Fokker-Planck equation.
$A(h, t), B(h, t)$	Jump moments in Fokker-Planck equation.

**Table 1:** Summary of main parameters and quantities used throughout.

## 2 General Approach to Memory Lifetimes with Filter-Based Synaptic Plasticity

We first provide a general recapitulation of our approach to studying memory lifetimes in the context of filter-based synaptic plasticity. Further details may be found elsewhere (Elliott & Lagogiannis, 2012; Elliott, 2014). In Table 1 we provide a summary of the key mathematical symbols introduced in this and later sections.

### 2.1 Perceptron Formulation in Continuous Time

We consider memory storage for simplicity in the context of a simple, feedforward, perceptron-based approach. Avoiding the use of recurrently-connected, Hopfield-like (Hopfield, 1982) networks of neurons allows many results to be derived exactly. We consider a single perceptron with  $N$  synapses of binary strengths  $S_i(t) \in \{-1, +1\}$ , where  $i = 1, \dots, N$  indexes the synapses and  $t$  denotes continuous time. We will use the notation  $S_i(n)$  with  $n$  a non-negative

integer when we refer to synaptic strengths at discrete time points. We refer to synapses with  $S_i(t) = +1$  as “strong” synapses and those with  $S_i(t) = -1$  as “weak” synapses. This interpretation is legitimate because we can always add an overall constant to these strengths to make them non-negative. To compensate for any induced changes in the perceptron’s firing, we would merely have to change the perceptron’s firing threshold.

The perceptron is required to store a sequence of synaptic memories  $\xi^\alpha$ ,  $\alpha = 0, 1, 2, \dots$ . The components  $\xi_i^\alpha$  take values of  $\pm 1$  with probabilities  $\text{Prob}[\xi_i^\alpha = \pm 1] = g_\pm$ , with  $g_+ + g_- = 1$ . Below we will consider only the balanced case  $g_\pm = \frac{1}{2}$ , but initially we keep  $g_\pm$  general. For simplicity we take the components  $\xi_i^\alpha$  to be uncorrelated across synapses and between memories. Memory  $\xi^0$  is always stored at time  $t = 0^-$  s. We use this formal convenience of  $t = 0^-$  s so that we may refer to the time immediately after the storage of memory  $\xi^0$  as  $t = 0$  s. We refer to memory  $\xi^0$  as the “tracked” memory because we are interested in the fidelity of recall of this particular memory by the perceptron at later times. The recall of this memory will be affected by the subsequent storage of the later memories  $\xi^\alpha$ ,  $\alpha \geq 1$ , as their storage will induce changes in the synaptic strengths  $S_i(t)$  through synaptic plasticity. In a discrete time formalism, these subsequent memories are stored at the discrete time steps  $t = \alpha$  s. Biologically speaking, however, a discrete time approach is somewhat unnatural and memories are more realistically stored as a continuous time process. The simplest continuous time process to consider is a Poisson process of rate  $r$ . We will therefore consider memory storage to occur as a Poisson process, with the later memories  $\xi^\alpha$ ,  $\alpha \geq 1$ , stored at rate  $r$  at times  $t > 0$  s governed by this stochastic process. We may of course convert a discrete time process into a Poisson process by weighting and summing over discrete time events according to the relevant Poisson weighting factors, as employed below.



For the model of synaptic plasticity that we consider here, we may without loss of generality set  $r = 1$  Hz because  $r$  simply acts as an overall factor multiplying time. We retain  $r$  for clarity where necessary, but in simulations we always set  $r = 1$  Hz.

The fidelity of recall of the tracked memory is gauged by measuring the perceptron's activation in response to  $\xi^0$ . If the perceptron has inputs  $x_i \in \{-1, +1\}$  to its  $N$  synapses, then the activation takes the standard form

$$h_{\mathbf{x}}(t) = \frac{1}{N} \sum_{i=1}^N x_i S_i(t), \quad (2.1)$$

and we define  $h(t) = h_{\xi^0}(t)$  as the tracked memory signal or just the memory signal. Depending on how we define memory lifetimes, we are only concerned with whether or not this memory signal is above or below the perceptron's firing threshold or its equilibrium, large time value. We therefore do not need to consider the conversion of this activation to the perceptron's two-level output: if  $h(t)$  is above threshold, then memory  $\xi^0$  is still stored by the perceptron at time  $t$ , otherwise it is not.

We are not interested in the dynamics of  $h(t)$  for any given realisation of the memories  $\xi^\alpha$  but only in the statistics of  $h(t)$  averaged over all possible realisations of these memories. Defining

$$\mu(t) = \mathbb{E}[h(t)], \quad (2.2a)$$

$$\sigma(t)^2 = \text{Var}[h(t)], \quad (2.2b)$$

where  $\mathbb{E}[\cdot]$  and  $\text{Var}[\cdot]$  denote the mean and variance, respectively, the signal-to-noise ratio (SNR) of the perceptron's activation is  $\mathcal{SNR}(t) = [\mu(t) - \mu(\infty)]/\sigma(t)$ , or just  $\mathcal{SNR}(t) = \mu(t)/\sigma(t)$  when  $\mu(t) \rightarrow 0$  as  $t \rightarrow \infty$ . We may then define the

lifetime of the tracked memory as the solution of  $\mathcal{SNR}(\tau_{\text{snr}}) = 1$ , i.e. the time at which  $\mu(t)$  is indistinguishable from its equilibrium value  $\mu(\infty)$  at the level of one standard deviation (Tsodyks, 1990). Alternatively, we may define the memory lifetime by considering the first passage time for the tracked memory signal  $h(t)$  to drop (to or) below some (perhaps perceptron firing) threshold  $\vartheta$ , and average this first passage time over all possible realisations to obtain the MFPT memory lifetime  $\tau_{\text{mfpt}}$  (Elliott, 2014). We shall almost exclusively consider MFPT memory lifetimes here.

The input  $\xi_i^\alpha$  to a particular synapse determines how that synapse’s strength should change. As we consider only an isolated perceptron rather than a recurrently-connected network of neurons, we require only that the perceptron’s activation is above firing threshold, so that the perceptron’s output is  $+1$  rather than  $-1$ . The input  $\xi_i^\alpha$  to synapse  $i$  upon presentation of memory  $\alpha$  is therefore just the plasticity induction signal to this synapse. If  $\xi_i^\alpha = +1$ , then the synapse receives a strengthening or potentiating plasticity induction signal, while if  $\xi_i^\alpha = -1$ , then it receives a weakening or depressing plasticity induction. We now discuss how synapses respond to these plasticity induction signals in our filter-based approach to synaptic plasticity.

## 2.2 Filter-Based Synaptic Plasticity

In order to control fluctuations in synaptic strengths, we have argued that synapses should integrate plasticity induction signals in order to discern any trends in these signals before expressing synaptic plasticity (Elliott, 2008). Such integration can be achieved by a low-pass filter, with synaptic plasticity being expressed only when the filter reaches threshold. A simple implementation of a low-pass filter in a synapse can be achieved by considering an internal

filter state, labelled by letters such as  $I$  and  $J$ , instantiated on a small collection of macromolecules. We have previously discussed the possibility that the filter state could be encoded in the phosphorylation states of kinases and phosphatases at single synapses, such as the CaMKII and PP1 enzymes (see Elliott (2011a) for a fuller discussion). In the context of the storage of memories, we consider a filter with an upper threshold  $+\Theta_+$  and a lower threshold  $-\Theta_-$ , with  $\Theta_{\pm} > 0$  (Elliott & Lagogiannis, 2012). Potentiating induction signals increment the filter state while depressing induction signals decrement the filter state. The allowed filter states are  $I \in \{-(\Theta_- - 1), \dots, +(\Theta_+ - 1)\}$ . If the filter reaches  $+\Theta_+$ , then it is reset to the  $I = 0$  state and potentiation is expressed if the synapse is weak (i.e. the synapse becomes strong); conversely, if the filter reaches  $-\Theta_-$ , then it is reset to the  $I = 0$  state and depression is expressed if the synapse is strong (i.e. the synapse becomes weak). We do not consider  $\pm\Theta_{\pm}$  to be allowed states because they are thresholds at which the synapse is immediately reset to  $I = 0$ .

Because we will restrict to balanced potentiation and depression processes, for which  $g_{\pm} = \frac{1}{2}$ , we consider only symmetric filters for which  $\Theta_{\pm} = \Theta$ . The filter state is then represented as a  $(2\Theta - 1)$ -dimensional vector and transitions in this filter state are implemented by  $(2\Theta - 1) \times (2\Theta - 1)$  matrices. For convenience we index the components and elements of such vectors and matrices by their corresponding filter states, so that for example vector components run from  $-(\Theta - 1)$  to  $+(\Theta - 1)$  rather than from 1 to  $(2\Theta - 1)$ . Let the matrix  $\mathbb{S}^+$  increment the filter state but without implementing the upper threshold process, and let the matrix  $\mathbb{T}^+$  implement only this upper threshold process.

For  $\Theta = 3$ , for example, we have

$$\mathbb{S}^+ = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbb{T}^+ = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The matrix  $\mathbb{S}^+$  just shifts the components of vectors one entry downwards, letting the last component disappear, while  $\mathbb{T}^+$  sends this last  $I = +(\Theta - 1)$  component to  $I = 0$ . Similarly, we denote the corresponding matrices for decrementing filter states and implementing only the lower threshold process by  $\mathbb{S}^-$  and  $\mathbb{T}^-$ . For  $\Theta = 3$ , for example, we have

$$\mathbb{S}^- = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbb{T}^- = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We clearly have that  $\mathbb{S}^- = (\mathbb{S}^+)^T$ , where the superscript T denotes the transpose, and  $\mathbb{T}^-$  sends the  $I = -(\Theta - 1)$  state to  $I = 0$ . The joint distribution of strength and filter states is represented by a  $2(2\Theta - 1)$ -dimensional vector, where the first (second) block of  $(2\Theta - 1)$  of components represents the filter state when the synapse is weak (strong). Transitions in the joint distribution of strength and filter states are then represented by  $2(2\Theta - 1) \times 2(2\Theta - 1)$  matrices. Let the matrices  $\mathbb{M}^\pm$  implement changes in synaptic states in response to potentiating and depressing induction signals, respectively. Schematically representing the block structure of these matrices, we have

$$\mathbb{M}^+ = \left( \begin{array}{c|c} \mathbb{S}^+ & \mathbb{O} \\ \hline \mathbb{T}^+ & \mathbb{S}^+ + \mathbb{T}^+ \end{array} \right) \quad \text{and} \quad \mathbb{M}^- = \left( \begin{array}{c|c} \mathbb{S}^- + \mathbb{T}^- & \mathbb{T}^- \\ \hline \mathbb{O} & \mathbb{S}^- \end{array} \right), \quad (2.3)$$

where  $\mathbb{O}$  is an appropriately sized matrix with entries of zero everywhere. The submatrix  $\mathbb{T}^+$  in the lower left sub-block of  $\mathbb{M}^+$  implements a change in strength from weak to strong via the upper filter threshold process, while its presence in the lower right sub-block implements an upper filter threshold process but without a change in strength because the synapse is already strong. Similarly for  $\mathbb{T}^-$  in  $\mathbb{M}^-$ . Finally, we also define the matrix

$$\mathbb{M} = g_+ \mathbb{M}^+ + g_- \mathbb{M}^-, \quad (2.4)$$

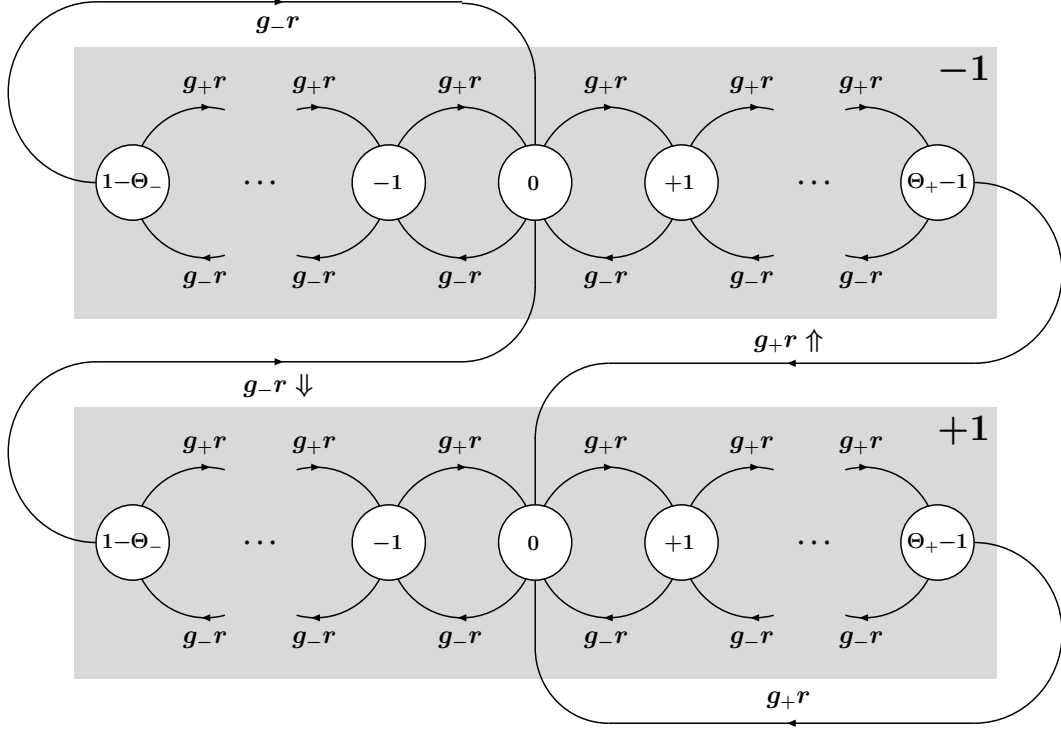
which represents the occurrence of a potentiating induction signal with probability  $\text{Prob}[\xi_i^\alpha = +1] = g_+$  and a depressing induction signal with probability  $\text{Prob}[\xi_i^\alpha = -1] = g_-$ . In terms of its block structure, we have

$$\mathbb{M} = \left( \begin{array}{c|c} g_+ \mathbb{S}^+ + g_- \mathbb{S}^- + g_- \mathbb{T}^- & g_- \mathbb{T}^- \\ \hline g_+ \mathbb{T}^+ & g_+ \mathbb{S}^+ + g_- \mathbb{S}^- + g_+ \mathbb{T}^+ \end{array} \right). \quad (2.5)$$

This matrix encodes the transitions illustrated in Fig. 1.

The transition matrix  $\mathbb{M}$  averages over both potentiating and depressing induction signals. It is therefore the required matrix for averaging over the non-tracked memories  $\xi^\alpha$ ,  $\alpha \geq 1$ . The eigenvector of  $\mathbb{M}$  with unit eigenvalue gives the equilibrium or asymptotic joint distribution of strength and filter states. It is against the background of this equilibrium distribution that the tracked memory  $\xi^0$  is stored. A direct calculation shows that the equilibrium eigenvector, normalised so that it is a probability distribution, takes the form, schematically,

$$\mathbf{A} = \left( \frac{g_-^\ominus}{g_+^\ominus + g_-^\ominus} \mathbf{B}^\top \middle| \frac{g_+^\ominus}{g_+^\ominus + g_-^\ominus} \mathbf{B}^\top \right)^\top, \quad (2.6)$$



**Figure 1:** Transitions between synaptic strength and filter states in response to synaptic plasticity induction signals. The circles represent allowed filter states, with the state indicated by the number enclosed by the circle. Strength states are indicated by the two gray boxes labelled  $-1$  and  $+1$ . Filter states are duplicated between these boxes so that both strength and filter states can be represented, but we stress that each synapse has only a single filter. Transitions between strength and filter states caused by plasticity induction signals are indicated by the lines carrying arrows. Potentiating (depressing) plasticity induction signals occur at a rate  $g_+r$  ( $g_-r$ ) and increment (decrement) the filter state. If the filter is in state  $+(\Theta_+ - 1)$  ( $-(\Theta_- - 1)$ ), then a potentiating (depressing) induction signal will return it to the  $I = 0$  state; if the synapse is weak (strong), potentiation (depression) is expressed (indicated by  $\uparrow$  ( $\downarrow$ )), but if the synapse is strong (weak), it can only remain strong (weak).

where the  $(2\Theta - 1)$ -dimensional vector  $\mathbf{B}$  has components

$$B_I = \begin{cases} \frac{1}{\Theta} \frac{g_+^{\Theta+I} g_-^{-I} - g_-^{\Theta}}{g_+^{\Theta} - g_-^{\Theta}} & \text{for } I < 0 \\ \frac{1}{\Theta} & \text{for } I = 0 \\ \frac{1}{\Theta} \frac{g_+^{\Theta} - g_+^I g_-^{\Theta-I}}{g_+^{\Theta} - g_-^{\Theta}} & \text{for } I > 0 \end{cases} . \quad (2.7)$$

For the particular case  $g_{\pm} = \frac{1}{2}$ , this reduces to

$$B_I = \frac{\Theta - |I|}{\Theta^2} . \quad (2.8)$$

In equilibrium, the probabilities of a synapse being weak or strong are then just

$$\pi_{\pm} = \text{Prob}[S_i(\infty) = \pm 1] = \frac{g_{\pm}^{\Theta}}{g_+^{\Theta} + g_-^{\Theta}} , \quad (2.9)$$

while the conditional filter distribution, conditioned on a particular value of synaptic strength, is just  $\mathbf{B}$ , regardless of the strength. For  $g_{\pm} = \frac{1}{2}$ , we of course have  $\pi_{\pm} = \frac{1}{2}$ .

The tracked memory  $\boldsymbol{\xi}^0$  is stored against the background of this equilibrium distribution. If  $\xi_i^0 = +1$ , then the joint strength and filter probability distribution of synapse  $i$  immediately after the storage of  $\boldsymbol{\xi}^0$  is  $\mathbb{M}^+ \mathbf{A}$ , while if  $\xi_i^0 = -1$ , then it is  $\mathbb{M}^- \mathbf{A}$ . The initial state of any particular synapse at time  $t = 0$  s is therefore in general a mixture of these two distributions, with probabilities  $g_+$  and  $g_-$ . To determine the initial mean perceptron activation, we define  $\mathbf{n} = (1, \dots, 1)^T$  and  $\mathbf{0} = (0, \dots, 0)^T$  to be  $(2\Theta - 1)$ -dimensional vectors all of whose components are unity and zero, respectively, and then we define

the  $2(2\Theta - 1)$ -dimensional vectors

$$\mathbf{\Omega}_+ = (\mathbf{0}^\top \mid \mathbf{n}^\top)^\top, \quad (2.10a)$$

$$\mathbf{\Omega}_- = (\mathbf{n}^\top \mid \mathbf{0}^\top)^\top, \quad (2.10b)$$

which pick out, respectively, the strong and weak strength states from joint strength and filter probability distributions. We also define

$$\mathbf{\Omega} = \mathbf{\Omega}_+ - \mathbf{\Omega}_- = (-\mathbf{n}^\top \mid +\mathbf{n}^\top)^\top, \quad (2.11)$$

which weights weak synapses negatively and strong synapses positively, i.e. according to their strengths. We then have

$$\mu(0) = \mathbf{\Omega}^\top (g_+ \mathbb{M}^+ - g_- \mathbb{M}^-) \mathbf{A}. \quad (2.12)$$

At discrete time step  $n$ , the mean perceptron activation, denoted by  $\mu_n$  in discrete time, is just

$$\mu_n = \mathbf{\Omega}^\top \mathbb{M}^n (g_+ \mathbb{M}^+ - g_- \mathbb{M}^-) \mathbf{A}, \quad (2.13)$$

where the matrix power  $\mathbb{M}^n$  represents  $n$  storage events for the (averaged) memories  $\boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^n$ , and in continuous time, we just perform a Poisson sum over these discrete time values to obtain

$$\begin{aligned} \mu(t) &= \mathbf{\Omega}^\top \left[ \sum_{n=0}^{\infty} \frac{(rt)^n}{n!} e^{-rt} \mathbb{M}^n \right] (g_+ \mathbb{M}^+ - g_- \mathbb{M}^-) \mathbf{A}, \\ &= \mathbf{\Omega}^\top [\exp(rt \mathbb{G})] (g_+ \mathbb{M}^+ - g_- \mathbb{M}^-) \mathbf{A}, \end{aligned} \quad (2.14)$$



where  $\mathbb{G} = \mathbb{M} - \mathbb{I}$  and  $\mathbb{I}$  is the identity matrix.<sup>2</sup> These expressions for  $\mu_n$  and  $\mu(t)$  of course correspond to just the mean strength of any single synapse, as the average perceptron activation in response to the tracked memory just reduces to an average over any single synapse.

We have kept  $g_{\pm}$  general above, but we now restrict exclusively to balanced potentiation and depression processes, for which  $g_{\pm} = \frac{1}{2}$ . This balanced scenario is the case of greatest interest to us and one that we have studied extensively before. Although we may proceed in general with  $g_{\pm} \neq \frac{1}{2}$ , the advantage of setting  $g_{\pm} = \frac{1}{2}$  is that the mixed initial state at  $t = 0$  s effectively collapses down to a single state, at least in terms of the contribution to  $h(t)$ . We may see this explicitly by writing out the two distributions corresponding to  $\mathbb{M}^{\pm} \mathbf{A}$ . For example, for  $\Theta = 3$  we have

$$\mathbb{M}^+ \mathbf{A} = \frac{1}{18} (0, 1, 2, 3, 2 \mid 0, 1, 4, 3, 2)^T, \quad (2.15a)$$

$$\mathbb{M}^- \mathbf{A} = \frac{1}{18} (2, 3, 4, 1, 0 \mid 2, 3, 2, 1, 0)^T. \quad (2.15b)$$

We can see that these two distributions are exact mirror images of each other, i.e.  $\mathbb{M}^+ \mathbf{A}$  read top to bottom is identical to  $\mathbb{M}^- \mathbf{A}$  read bottom to top. For general  $\Theta$ , we obtain

$$\text{Prob}[S_i(0) = \pm 1] = \begin{cases} \frac{1}{2} (1 \pm \mu_0) & \text{for } \xi_i^0 = +1 \\ \frac{1}{2} (1 \mp \mu_0) & \text{for } \xi_i^0 = -1 \end{cases}, \quad (2.16)$$

where  $\mu_0 = \mu(0) = 1/\Theta^2$  is the initial mean memory signal immediately after

---

<sup>2</sup>For the identity and zero matrices  $\mathbb{I}$  and  $\mathbb{O}$  and the unity and zero vectors  $\mathbf{n}$  and  $\mathbf{0}$ , rather than using different symbols for matrices and vectors of different sizes, we use the same symbols. Their sizes are always clear from the context.

the storage of  $\xi^0$ . Defining  $\tilde{S}_i(t) = \xi_i^0 S_i(t)$ , we may write

$$\text{Prob}[\tilde{S}_i(0) = \pm 1] = \frac{1}{2} (1 \pm \mu_0). \quad (2.17)$$

The probability distribution for  $\tilde{S}_i(0)$  is therefore independent of the sign of  $\xi_i^0$ , unlike the probability distribution for  $S_i(0)$ . Furthermore, by examining the action of  $\mathbb{M} = \frac{1}{2} (\mathbb{M}^+ + \mathbb{M}^-)$  on the two distributions  $\mathbb{M}^\pm \mathbf{A}$ , we find that in general  $\mathbb{M}^n \mathbb{M}^\pm \mathbf{A}$  are always mirror images for any  $n \geq 0$ , so that  $\exp(rt \mathbb{G}) \mathbb{M}^\pm \mathbf{A}$  are therefore also mirror images for any  $t \geq 0$  s. Hence, the probability distribution for  $\tilde{S}_i(t)$  is therefore independent of the sign of  $\xi_i^0$  for any  $t \geq 0$  s so that all  $N$  variables  $\tilde{S}_i(t)$  are for all time identically distributed random variables. This behaviour is specific to scenarios in which potentiation and depression processes are treated symmetrically and identically and is discussed in detail and derived fully elsewhere (Elliott, 2016b).

With the tilded strength variables  $\tilde{S}_i(t)$  defined by  $\tilde{S}_i(t) = \xi_i^0 S_i(t)$ , the tracked memory signal  $h(t)$  can be written directly as

$$h(t) = \frac{1}{N} \sum_{i=1}^N \xi_i^0 S_i(t) \equiv \frac{1}{N} \sum_{i=1}^N \tilde{S}_i(t). \quad (2.18)$$

For discrete time steps we will write  $h_n$  instead of  $h(n)$ . For processes that treat potentiation and depression in a symmetrical and balanced manner, the perceptron activation in response to the tracked memory is therefore just an average over the  $N$  identically distributed tilded strength random variables.

The mean and variance then following immediately, as

$$\mu(t) = \mathbb{E}[\tilde{S}(t)] \quad (2.19)$$

$$\begin{aligned} \sigma(t)^2 &= \frac{1}{N} \mathbf{Var}[\tilde{S}(t)] + \left(1 - \frac{1}{N}\right) \mathbf{Cov}[\tilde{S}_i(t), \tilde{S}_j(t)] \\ &= \frac{1}{N} [1 - \mu(t)^2] + \left(1 - \frac{1}{N}\right) \mathbf{Cov}(t), \end{aligned} \quad (2.20)$$

where  $\mathbb{E}[\tilde{S}(t)]$  and  $\mathbf{Var}[\tilde{S}(t)]$  are the mean and variance, respectively, of any one of the synapse's tilded strengths, and  $\mathbf{Cov}(t) = \mathbf{Cov}[\tilde{S}_i(t), \tilde{S}_j(t)]$  is the covariance between any pair of them. Working with the tilded strength variables for balanced processes therefore dramatically simplifies the dynamics of  $h(t)$ , because the mixed initial state with the two distributions  $\mathbb{M}^\pm \mathbf{A}$  for  $\xi_i^0 = \pm 1$  effectively collapses down to a single distribution, in terms of the contribution to  $h(t)$ . For balanced processes, then, we can perform calculations by conditioning only on the initial value  $\tilde{S}_i(0)$  for any given synapse, while for general, unbalanced processes, we must instead condition on the values of  $S_i(0)$  and  $\xi_i^0$  individually rather than through just their product. Although conditioning on  $S_i(0)$  and  $\xi_i^0$  individually is straightforward, working with balanced processes simplifies calculations significantly.

We have previously used renewal methods (Cox, 1962) to evaluate  $\mu(t)$  in this balanced scenario (Elliott & Lagogiannis, 2012). We simply reproduce the result here:

$$\begin{aligned} \mu(t) &= \frac{1}{\Theta^3} \sum_{l=0}^{\Theta-1} \cot^2 \frac{(2l+1)\pi}{4\Theta} \exp \left\{ -rt \left[ 1 - \cos \frac{(2l+1)\pi}{2\Theta} \right] \right\} \\ &\quad - \frac{4}{\Theta^3} \sum_{l=0}^{\lfloor \frac{\Theta-1}{2} \rfloor} \cot^2 \frac{(2l+1)\pi}{2\Theta} \exp \left\{ -rt \left[ 1 - \cos \frac{(2l+1)\pi}{\Theta} \right] \right\}, \end{aligned} \quad (2.21)$$

where  $\lfloor x \rfloor$  is the floor function. Full discussion of the behaviour of  $\mu(t)$  may be

found elsewhere (Elliott & Lagogiannis, 2012). In brief, starting from the initial value of  $\mu(0) = 1/\Theta^2$ ,  $\mu(t)$  initially increases roughly  $\Theta$ -fold to a maximum at a time that grows as  $\Theta^2$ . Following its peak,  $\mu(t)$  then decays back to its equilibrium value, which for  $g_{\pm} = \frac{1}{2}$  is  $\mu(\infty) = 0$ . The initial rise in the mean memory signal in our filter-based, integrative model of synaptic plasticity is in radical contrast to non-integrative models, in which the mean memory signal always decays monotonically from its initial value. This increase occurs because the initial storage of  $\xi^0$  biases synapses that experience  $\xi_i^0 = +1$  to remain or become strong and those that experience  $\xi_i^0 = -1$  to remain or become weak. This biasing occurs because the filter distribution at each synapse is either stepped upwards or stepped downwards by the initial storage event (see Eq. (2.15)). The mean memory signal increases while this bias persists, and only once it has worked out of the system does the mean memory signal then start to return to equilibrium. The behaviour of  $\mu(t)$  will be amply illustrated below.

To compute the variance  $\sigma(t)^2$  in continuous time using the full transition matrix  $\mathbb{M}$  is in general very hard, although its calculation and that of higher-order cumulants is made considerably easier by the approach that we adopt here, as we discuss below. In discrete time, because of the assumed independence of  $\xi_i^\alpha$  across synapses and between memories, synapses evolve independently and thus the discrete-time variance is trivial to compute in terms of the discrete-time mean. However, by driving memory storage as a continuous-time process, correlations in synaptic strength are introduced, leading to a non-zero covariance term  $\text{Cov}(t)$  between synapses' strengths in Eq. (2.20) (Elliott & Lagogiannis, 2012). This occurs in *any* model of synaptic plasticity and not just the filter-based model considered here. We compute the variance, and higher-order cumulants, via a combination of analytical, numerical and simu-

lation methods, described below.

### 3 Transition Matrices for Changes in Synaptic Strength

Memory lifetimes are determined by the dynamics of the tracked memory signal. The MFPT for the perceptron activation in response to the tracked memory to fall (to or) below some threshold  $\vartheta$  can be used to define memory lifetimes (Elliott, 2014). Previously we have extensively studied memory lifetimes defined by MFPTs for a stochastic updater synapse with a fixed probability for expressing a change in synaptic strength. Although a very simple model, such a synapse has the virtue that memory lifetimes defined by MFPTs can be analysed in detail, with exact results available in many instructive limits (Elliott, 2014). However, synapses with internal states are extremely difficult to study analytically, in terms of MFPTs. Although exact analytical results can be derived, they are essentially useless for explicit computations, either analytical or numerical, unless  $N$  is very small (Elliott, 2014). This is because in order to keep track of the entire synaptic configuration over all  $N$  synapses, we would require the tensor product  $\mathbb{M} \otimes \cdots \otimes \mathbb{M}$ , with  $N$  occurrences of the matrix  $\mathbb{M}$ . Equivalently, flattening the product space, we would require a  $[2(2\Theta - 1)]^N \times [2(2\Theta - 1)]^N$  matrix, which in general is intractably large.

Even for a simple synapse, such a matrix would be  $2^N \times 2^N$ . However, for binary-strength synapses, the tracked memory signal  $h(t)$  in continuous time or  $h_n$  in discrete time is a normalised sum over  $N$  tilded strength variables taking values  $\pm 1$ . The perceptron activation is therefore uniquely determined by the number of these variables taking the value  $+1$ , say. For binary-strength,

simple synapses, we can therefore avoid working with the transitions in the entire synaptic configuration and can instead work directly with the transitions in perceptron activation. The matrix governing these transitions for simple, binary-strength synapses is only  $(N + 1) \times (N + 1)$ , which is generally is vastly smaller than  $2^N \times 2^N$  (Elliott, 2014). This shift to working directly with the transitions in perceptron activation is critical to studying MFPT memory lifetimes. This shift is possible because and only because simple synapses do not have internal states.

If it were possible, mathematically-speaking, to reduce a complex synapse with internal states to a simple synapse without internal states, then studying MFPT memory lifetimes for complex synapses could therefore in principle be made more tractable. In order to examine the dynamics of  $h(t)$  that govern memory lifetimes, we are not in fact interested in synapses' internal filter states but only in synapses' strengths, and specifically only in synapses' changes in strength, because only these lead to changes in perceptron activation. Although changes in synapses' internal filter states drive changes in their strengths, only synapses' strengths and not their internal filter states determine perceptron activation and therefore memory lifetimes. Thus, the first step in examining MFPT memory lifetimes is to integrate out synapses' internal filter states and instead work directly with changes in synapses' strengths, ignoring the underlying filter states and the transitions between them. This then permits us to work directly with the dynamics of the perceptron activation.

A simple stochastic updater synapse with fixed probabilities for expressing potentiation and depression can be defined by the two transition matrices

$$\mathbb{W}^+ = \begin{pmatrix} 1 - p^+ & 0 \\ p^+ & 1 \end{pmatrix} \text{ and } \mathbb{W}^- = \begin{pmatrix} 1 & 1 - p^- \\ 0 & p^- \end{pmatrix}, \quad (3.1)$$

where  $\mathbb{W}^+$  is the potentiation matrix that expresses potentiation with probability  $p^+$  and  $\mathbb{W}^-$  is the depression matrix that expresses depression with probability  $p^-$ . The transition matrix

$$\mathbb{W} = \frac{1}{2} (\mathbb{W}^+ + \mathbb{W}^-) = \begin{pmatrix} 1 - \frac{1}{2} p^+ & \frac{1}{2} p^- \\ \frac{1}{2} p^+ & 1 - \frac{1}{2} p^- \end{pmatrix} \quad (3.2)$$

then represents the occurrence of potentiating and depressing induction signals with probabilities  $\text{Prob}[\xi_i^\alpha = \pm 1] = \frac{1}{2}$  for any given (non-tracked) memory storage step. For a given memory storage step with filter-based synapses, we must determine these strength-change probabilities  $p^\pm$  if we are to integrate out the synapses' internal states and work instead with simplified synapses. These probabilities must of course depend on the current full state of the synapse (and therefore on its history of induction signals) and on the current induction signal. They are therefore contingent on a given history of memory storage events and on the initial filter and strength states of the synapse. We stress that in working with such a reduced, simple synapse, we are not arguing that complex synapses are fully equivalent to simple synapses or *vice versa*. We merely use this reduction as a mathematical device to simplify the analysis of the dynamics of the tracked memory signal by essentially throwing away the internal synaptic states.

To determine  $p^\pm$ , let the state vector for a single synapse at any given memory storage step be given schematically by  $(\mathbf{w}^\text{T} | \mathbf{s}^\text{T})^\text{T}$ , where  $\mathbf{w}$  and  $\mathbf{s}$  are  $(2\Theta - 1)$ -dimensional vectors corresponding to the weak and strong synaptic strength states, respectively, and with  $\mathbf{n} \cdot (\mathbf{w} + \mathbf{s}) = 1$ . Under the actions of

$\mathbb{M}^\pm$ , we have

$$\begin{pmatrix} \mathbf{w} \\ \mathbf{s} \end{pmatrix} \xrightarrow{\mathbb{M}^+} \begin{pmatrix} \mathbb{S}^+ \mathbf{w} \\ \mathbb{T}^+ \mathbf{w} + (\mathbb{S}^+ + \mathbb{T}^+) \mathbf{s} \end{pmatrix}, \quad (3.3a)$$

$$\begin{pmatrix} \mathbf{w} \\ \mathbf{s} \end{pmatrix} \xrightarrow{\mathbb{M}^-} \begin{pmatrix} \mathbb{T}^- \mathbf{s} + (\mathbb{S}^- + \mathbb{T}^-) \mathbf{w} \\ \mathbb{S}^- \mathbf{s} \end{pmatrix}. \quad (3.3b)$$

For a general vector  $\mathbf{v}$ , however,  $\mathbf{n} \cdot (\mathbb{S}^\pm + \mathbb{T}^\pm) \mathbf{v} = \mathbf{n} \cdot \mathbf{v}$ , so we see that under the action of  $\mathbb{M}^+$  the strong states gain probability  $\mathbf{n} \cdot \mathbb{T}^+ \mathbf{w}$  while the weak states lose this probability because  $\mathbf{n} \cdot \mathbb{S}^+ \mathbf{w} = \mathbf{n} \cdot \mathbf{w} - \mathbf{n} \cdot \mathbb{T}^+ \mathbf{w}$ . Similarly, under  $\mathbb{M}^-$  the weak states gain probability  $\mathbf{n} \cdot \mathbb{T}^- \mathbf{s}$  while the strong states must lose this probability. The probability of potentiation, conditional on the synapse being weak, is therefore

$$p^+ = \frac{\mathbf{n} \cdot \mathbb{T}^+ \mathbf{w}}{\mathbf{n} \cdot \mathbf{w}} \quad (3.4)$$

and the probability of depression, conditional on the synapse being strong, is

$$p^- = \frac{\mathbf{n} \cdot \mathbb{T}^- \mathbf{s}}{\mathbf{n} \cdot \mathbf{s}}. \quad (3.5)$$

We define  $p^+ = 0$  or  $p^- = 0$  if  $\mathbf{n} \cdot \mathbf{w} = 0$  or  $\mathbf{n} \cdot \mathbf{s} = 0$ , respectively. The probability  $p^+$  is just the probability that a synapses is in filter state  $I = +(\Theta - 1)$  relative to the total probability that the synapse is weak; similarly,  $p^-$  is the probability of being in filter state  $I = -(\Theta - 1)$  relative to its being strong. Defining the 2-dimensional vector of the synapse's strength probabilities with components  $\mathbf{n} \cdot \mathbf{w}$  and  $\mathbf{n} \cdot \mathbf{s}$ , we have that

$$\begin{pmatrix} 1 - \frac{1}{2} p^+ & \frac{1}{2} p^- \\ \frac{1}{2} p^+ & 1 - \frac{1}{2} p^- \end{pmatrix} \begin{pmatrix} \mathbf{n} \cdot \mathbf{w} \\ \mathbf{n} \cdot \mathbf{s} \end{pmatrix} = \begin{pmatrix} \mathbf{n} \cdot \mathbf{w} - \frac{1}{2} \mathbf{n} \cdot \mathbb{T}^+ \mathbf{w} + \frac{1}{2} \mathbf{n} \cdot \mathbb{T}^- \mathbf{s} \\ \mathbf{n} \cdot \mathbf{s} + \frac{1}{2} \mathbf{n} \cdot \mathbb{T}^+ \mathbf{w} - \frac{1}{2} \mathbf{n} \cdot \mathbb{T}^- \mathbf{s} \end{pmatrix}. \quad (3.6)$$

By construction, the strength probabilities in the vector on the right-hand side



(RHS) agree with those from the action of the full  $2(2\Theta - 1) \times 2(2\Theta - 1)$  transition matrix  $\mathbb{M} = \frac{1}{2}(\mathbb{M}^- + \mathbb{M}^+)$  on the full  $2(2\Theta - 1)$ -dimensional state vector  $(\mathbf{w}^T | \mathbf{s}^T)^T$ . That is, the  $2 \times 2$  strength transition matrix  $\mathbb{W}$  on the left-hand side of this equation correctly captures that changes in strength induced by the full filter dynamics. Of course, to compute the conditional strength-change probabilities  $p^\pm$ , we have required the full filter dynamics, but the key feature is that we have a simple synapse with no internal states and with identical strength-change probabilities, at least for one memory storage step.

For multiple memory storage steps, suppose that we start from some initial state vector  $(\mathbf{w}_0^T | \mathbf{s}_0^T)^T$ . Then we write

$$\begin{pmatrix} \mathbf{w}_{n+1} \\ \mathbf{s}_{n+1} \end{pmatrix} = \mathbb{M} \begin{pmatrix} \mathbf{w}_n \\ \mathbf{s}_n \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \mathbf{w}_n \\ \mathbf{s}_n \end{pmatrix} = \mathbb{M}^n \begin{pmatrix} \mathbf{w}_0 \\ \mathbf{s}_0 \end{pmatrix} \quad (3.7)$$

and we define the conditional strength-change probabilities  $p_n^\pm$  by

$$p_n^+ = \frac{\mathbf{n} \cdot \mathbb{T}^+ \mathbf{w}_n}{\mathbf{n} \cdot \mathbf{w}_n} \quad \text{and} \quad p_n^- = \frac{\mathbf{n} \cdot \mathbb{T}^- \mathbf{s}_n}{\mathbf{n} \cdot \mathbf{s}_n}, \quad (3.8)$$

and the corresponding  $2 \times 2$  step-dependent strength-change transition matrices are

$$\mathbb{W}_{n+1} = \begin{pmatrix} 1 - \frac{1}{2} p_n^+ & \frac{1}{2} p_n^- \\ \frac{1}{2} p_n^+ & 1 - \frac{1}{2} p_n^- \end{pmatrix}. \quad (3.9)$$

If  $(\mathbf{w}_0^T | \mathbf{s}_0^T)^T$  is the full state vector immediately after the storage of the tracked memory  $\boldsymbol{\xi}^0$ , then the storage of  $\boldsymbol{\xi}^0$  induces the conditional strength-change probabilities  $p_0^\pm$  that are relevant to the storage of the next memory  $\boldsymbol{\xi}^1$  associated with the application of the  $2 \times 2$  matrix  $\mathbb{W}_1$ . Similarly, the storage of  $\boldsymbol{\xi}^1$  then changes the distribution of strength and filter states and thus induces new conditional strength-change probabilities  $p_1^\pm$  that will govern the subsequent storage of memory  $\boldsymbol{\xi}^2$  via the  $2 \times 2$  matrix  $\mathbb{W}_2$ . And so on. Thus,

the conditional strength-change probabilities  $p_n^\pm$  are calculated from the full  $2(2\Theta - 1)$ -dimensional state vector after the storage of memory  $\xi^n$  and therefore from the vector  $\mathbb{M}^n(\mathbf{w}_0^\text{T}|\mathbf{s}_0^\text{T})^\text{T}$ , and these conditional probabilities govern the storage of the next memory  $\xi^{n+1}$  via the matrix  $\mathbb{W}_{n+1}$ . The product

$$\mathbb{D}_n = \mathbb{W}_n \mathbb{W}_{n-1} \cdots \mathbb{W}_2 \mathbb{W}_1 \quad (3.10)$$

then represents  $n$  applications of  $\mathbb{M}$  to the initial state  $(\mathbf{w}_0^\text{T}|\mathbf{s}_0^\text{T})^\text{T}$ , giving the probabilities of the synapse being weak or strong after the storage of the sequence of memories  $\xi^0, \dots, \xi^n$ . Because we obtain the correct probabilities for a synapse being weak or strong for any single memory storage step from any state, we must therefore also obtain the correct probabilities for any number of memory storage steps from any initial states. Thus,  $\mathbb{D}_n(\mathbf{n} \cdot \mathbf{w}_0, \mathbf{n} \cdot \mathbf{s}_0)^\text{T}$  gives the same probabilities of the synapse being weak or strong as those computed directly from  $\Omega_-^\text{T} \mathbb{M}^n(\mathbf{w}_0^\text{T}|\mathbf{s}_0^\text{T})^\text{T}$  or  $\Omega_+^\text{T} \mathbb{M}^n(\mathbf{w}_0^\text{T}|\mathbf{s}_0^\text{T})^\text{T}$ , respectively. Again, we stress that we require the full internal synaptic dynamics to compute the probabilities  $p_n^\pm$ , and we stress that these probabilities depend on the full initial synaptic state and on the particular sequence of (superposed) induction signals experienced by the synapse. However, they allow us to work with an effective simple synapse and only a  $2 \times 2$  strength transition matrix without internal dynamics. This allows us to focus purely on changes in synapses' strengths, and therefore on the dynamics of the perceptron activation, without having to keep track of all synapses' internal states. The  $2 \times 2$  transition matrices  $\mathbb{W}_n$  depend on the memory storage step, so we refer to this simplified, reduced synapse as a time-dependent stochastic updater (TDSU).

We write the matrix  $\mathbb{D}_n$  in the form

$$\mathbb{D}_n = \begin{pmatrix} \frac{1}{2}(1 + X_n) & \frac{1}{2}(1 - Y_n) \\ \frac{1}{2}(1 - X_n) & \frac{1}{2}(1 + Y_n) \end{pmatrix}, \quad (3.11)$$

which defines the quantities  $X_n$  and  $Y_n$ . As  $n \rightarrow \infty$ , we must asymptote to the correct equilibrium distribution of synaptic strengths governed by the filter dynamics, so

$$\mathbb{D}_n \rightarrow \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad (3.12)$$

since  $\pi_{\pm} = \frac{1}{2}$  for  $g_{\pm} = \frac{1}{2}$ . Thus, we have isolated this equilibrium distribution in the definition in Eq. (3.11) so that  $X_n$  and  $Y_n$  are defined relative to equilibrium. We must then have that  $X_n \rightarrow 0$  and  $Y_n \rightarrow 0$  as  $n \rightarrow \infty$ . Writing  $\mathbb{D}_{n+1} = \mathbb{W}_{n+1}\mathbb{D}_n$ , we obtain simple recurrence relations for  $X_n$  and  $Y_n$ ,

$$X_{n+1} = \left[1 - \frac{1}{2}(p_n^+ + p_n^-)\right] X_n - \frac{1}{2}(p_n^+ - p_n^-), \quad (3.13a)$$

$$Y_{n+1} = \left[1 - \frac{1}{2}(p_n^+ + p_n^-)\right] Y_n + \frac{1}{2}(p_n^+ - p_n^-), \quad (3.13b)$$

with  $X_0 = 1$  and  $Y_0 = 1$  giving the correct form for  $\mathbb{D}_1 \equiv \mathbb{W}_1$ . For  $X_n \rightarrow 0$  and  $Y_n \rightarrow 0$  as  $n \rightarrow \infty$ , we must have that  $p_n^+/p_n^- \rightarrow 1$ , so that over time, the potentiation and depression probabilities must become equal. The recurrence relations in Eq. (3.13) will be used extensively below.

As we have indicated, the probabilities  $p_n^{\pm}$  depend on the initial state vector  $(\mathbf{w}_0^T | \mathbf{s}_0^T)^T$ , and this can be either  $\mathbb{M}^+ \mathbf{A}$  or  $\mathbb{M}^- \mathbf{A}$  depending on the sign of  $\xi_i^0$ . For  $\mathbb{M}^+ \mathbf{A}$ , a direct calculation shows that  $p_0^+ = 2/(\Theta^2 - 1)$  and  $p_0^- = 0$  for  $\Theta > 1$ , while for  $\mathbb{M}^- \mathbf{A}$ ,  $p_0^+ = 0$  and  $p_0^- = 2/(\Theta^2 - 1)$ . These probabilities are merely interchanged, reflecting the equivalence of the  $\tilde{S}_i(t)$  distributions for balanced processes with  $g_{\pm} = \frac{1}{2}$ . We note, furthermore, that in the recurrence relations in Eq. (3.13), under  $p_n^+ \leftrightarrow p_n^-$ , we have  $X_n \leftrightarrow Y_n$ , which again reflects

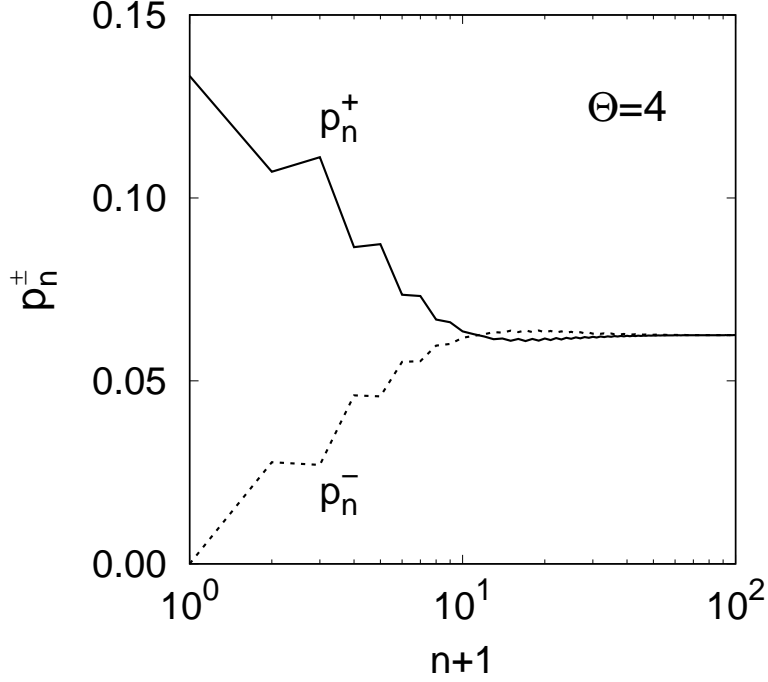
this equivalence. Thus, we take the probabilities  $p_n^\pm$  to be the transition probabilities in the tilded strength variables, as the initial mixed state collapses to a single distribution for balanced processes. To compute them, then, we need only consider the initial distribution associated with  $\mathbb{M}^+ \mathbf{A}$ . With this understanding, the elements of the matrix  $\mathbb{D}_n$  give the transition probabilities for tilded strength states after  $n$  memory storage step. For example, the element  $\frac{1}{2}(1 - X_n)$  is just the transition probability  $\text{Prob}[\tilde{S}_i(n) = +1 \mid \tilde{S}_i(0) = -1]$ .

Although we may use the renewal methods described elsewhere (Elliott & Lagogiannis, 2012) to obtain explicit formulae for arbitrary  $\Theta$  for the conditional strength-change probabilities  $p_n^\pm$ , the resulting expressions are in general very messy. For example, for  $\Theta = 3$ , we obtain

$$p_n^\pm = \frac{6(2^n \pm 2) \mp (\sqrt{3})^n [2 + \sqrt{3} + (-1)^n(2 - \sqrt{3})]}{6(9 \cdot 2^n \pm 4) \mp 2(\sqrt{3})^n [7 + 4\sqrt{3} + (-1)^n(7 - 4\sqrt{3})]},$$

for  $n > 0$ , with  $p_0^+ = \frac{1}{4}$  and  $p_0^- = 0$ . Numerically, it is more efficient simply to construct one-time tables of these conditional probabilities rather than repeatedly use explicit formulae with the associated computational overheads. These tables may be constructed, if necessary, by explicit numerical evaluation of the matrix powers  $\mathbb{M}^n$  when analytical results are not available for any particular form of  $\mathbb{M}$ . To illustrate the behaviour of  $p_n^\pm$ , in Fig. 2 we plot them against  $n$  for  $\Theta = 4$ . We see that initially  $p_n^+ > p_n^-$ , but that they approach each other, overshoot somewhat, and then both asymptote to  $1/\Theta^2$  as  $n \rightarrow \infty$ .

We have discussed the construction of an effective, simple synapse with identical strength-change probabilities only for our filter-based model of synaptic plasticity. It is clear, however, that this construction will in fact go through for any model of synaptic plasticity with internal synaptic states and not just a filter-based model. For any given model, we merely have to identify the



**Figure 2:** Conditional strength-change probabilities  $p_n^\pm$  as functions of  $n$  for  $\Theta = 4$ .

equivalents of the two matrices  $\mathbb{T}^\pm$  that lead directly to strength-change processes, which are the equivalents of threshold processes in a filter-based model. Also, this construction works for general, multi-level discrete synapses and not just for the binary-strength, two-level synapses considered here. Finally, this construction will also work for unbalanced processes with  $g_\pm \neq \frac{1}{2}$ , but then we will have two sets of probabilities  $p_n^\pm$  corresponding to the two different distributions in the mixed initial state.

## 4 Dynamics of Perceptron Activation

Now that we have synaptic dynamics purely in terms of conditional strength-change probabilities, we may determine exactly the probability distribution of the perceptron activation  $h_n$  in discrete time and thus via Poisson summation

the continuous time form  $h(t)$ . In discrete time, with random, uncorrelated memories, synapses are completely independent of each other. Dependence is introduced in continuous time but is entirely due to driving synaptic updates as a continuous-time process: performing Poisson summation fully accounts for the induced dependence. The independence of synapses in discrete time allows us to compute exactly the probability generating function (PGF) or the moment generating function (MGF) for  $h_n$ .

## 4.1 Generating Functions for Perceptron Activation

The scaled perceptron activation  $Nh_n$  is a sum over  $N$  random variables taking values of  $\pm 1$ . The value of  $h_n$  is therefore completely determined by the number of synapses with  $\tilde{S}_i(n) = +1$ , so we may construct generating functions for  $h_n$  or its moments by conditioning on this number. Specifically, we condition on the initial value,  $h_0$ , immediately after the storage of memory  $\boldsymbol{\xi}^0$ . If  $h_0 = 2j/N - 1$ , then precisely  $j$  synapses have  $\tilde{S}_i(0) = +1$  and the remaining  $N - j$  synapses have  $\tilde{S}_i(0) = -1$ . We define the two PGFs

$$\mathcal{G}_n^\pm(x) = \sum_{\sigma \in \{0,1\}} \text{Prob}[\tilde{S}_i(n) = 2\sigma - 1 \mid \tilde{S}_i(0) = \pm 1] x^\sigma, \quad (4.1)$$

where we have used the powers 0 and 1 rather than  $-1$  and  $+1$  because to determine the distribution of  $h_n$  we count only the synapses with  $\tilde{S}_i(n) = +1$ . Just by reading off the elements of the matrix  $\mathbb{D}_n$ , we may immediately write down these PGFs as

$$\mathcal{G}_n^-(x) = \frac{1}{2}(1 + X_n)x^0 + \frac{1}{2}(1 - X_n)x^1, \quad (4.2a)$$

$$\mathcal{G}_n^+(x) = \frac{1}{2}(1 - Y_n)x^0 + \frac{1}{2}(1 + Y_n)x^1. \quad (4.2b)$$

We note that  $\mathcal{G}_0^-(x) = 1$  and  $\mathcal{G}_0^+(x) = x$  since  $X_0 = 1$  and  $Y_0 = 1$ . In discrete time, all  $N$  synapses are independent, so the PGF for the number of synapses with  $\tilde{S}_i(n) = +1$  immediately after the storage of memory  $\xi^n$ , conditional on  $h_0 = 2j/N - 1$ , is given simply by

$$\begin{aligned}\mathcal{G}_n(x \mid h_0 = \tfrac{2j}{N} - 1) &= \sum_{i=0}^N \text{Prob}[h_n = \tfrac{2i}{N} - 1 \mid h_0 = \tfrac{2j}{N} - 1] x^i \\ &= [\mathcal{G}_n^-(x)]^{N-j} [\mathcal{G}_n^+(x)]^j.\end{aligned}\tag{4.3}$$

The coefficients  $\text{Prob}[h_n = \tfrac{2i}{N} - 1 \mid h_0 = \tfrac{2j}{N} - 1]$  in this conditional PGF encode the entire  $n$ -step  $(N+1) \times (N+1)$  transition matrix for the perceptron activation. We denote this matrix by  $\mathbb{H}_n$ . For convenience we index its entries from 0 to  $N$  rather than 1 to  $N+1$  so that we may explicitly write

$$[\mathbb{H}_n]_{ij} = \text{Prob}[h_n = \tfrac{2i}{N} - 1 \mid h_0 = \tfrac{2j}{N} - 1].\tag{4.4}$$

The generating function  $\mathcal{G}_n(x \mid h_0 = \tfrac{2j}{N} - 1)$  therefore generates the  $j$ th column of the matrix  $\mathbb{H}_n$ . We have that  $\mathcal{G}_0(x \mid h_0 = \tfrac{2j}{N} - 1) = x^j$ , so that  $\mathbb{H}_0 \equiv \mathbb{I}$  as required. From Eq. (2.17), we know that  $\text{Prob}[\tilde{S}_i(0) = \pm 1] = \frac{1}{2}(1 \pm \mu_0)$ . Thus, the initial distribution of  $j$  in  $h_0 = 2j/N - 1$  is binomial with parameter  $N$  and probability  $\frac{1}{2}(1 + \mu_0)$ . The unconditional PGF for the number of synapses with  $\tilde{S}_i(n) = +1$  at step  $n$  is then just

$$\begin{aligned}\mathcal{G}_n(x) &= \sum_{i=0}^N \text{Prob}[h_n = \tfrac{2i}{N} - 1] x^i \\ &= \left\{ \frac{1}{2} \left[ 1 + \frac{1}{2}(1 - \mu_0)X_n - \frac{1}{2}(1 + \mu_0)Y_n \right] x^0 \right. \\ &\quad \left. + \frac{1}{2} \left[ 1 - \frac{1}{2}(1 - \mu_0)X_n + \frac{1}{2}(1 + \mu_0)Y_n \right] x^1 \right\}^N.\end{aligned}\tag{4.5}$$

This PGF encodes the entire probability distribution of  $h_n$  via its coefficients  $\text{Prob}[h_n = \frac{2i}{N} - 1]$ . The state vector for  $h_n$  is an  $(N + 1)$ -dimensional vector, again indexed from 0 to  $N$ , with entries  $\text{Prob}[h_n = \frac{2i}{N} - 1]$  for  $i = 0, \dots, N$ . We see that  $h_n$  is itself essentially binomially-distributed (up to a scaling of its range) with parameter  $N$  and probability given by the coefficient of the  $x^1$  term inside the curly brackets in Eq. (4.5). In the limit  $n \rightarrow \infty$ , we obtain

$$\mathcal{G}_n(x) \rightarrow \left(\frac{1}{2}x^0 + \frac{1}{2}x^1\right)^N, \quad (4.6)$$

which of course agrees with the equilibrium distribution of  $\tilde{S}_i(n)$ , since  $\pi_{\pm} = \frac{1}{2}$ .

These conditional and unconditional PGFs for  $h_n$  translate directly into the corresponding MGFs for  $h_n$ . We write

$$\mathcal{M}_n^{\pm}(x) = \sum_{\sigma \in \{-1, +1\}} \text{Prob}[\tilde{S}_i(n) = \sigma \mid \tilde{S}_i(0) = \pm 1] e^{\sigma x/N}, \quad (4.7)$$

where we use  $\sigma x/N$  in the exponent because the divisor automatically accounts for the scaling of  $\sum_{i=1}^N \tilde{S}_i(n)$  by  $N$  in the definition of  $h_n$ . We then obtain

$$\mathcal{M}_n^-(x) = \cosh \frac{x}{N} - X_n \sinh \frac{x}{N}, \quad (4.8a)$$

$$\mathcal{M}_n^+(x) = \cosh \frac{x}{N} + Y_n \sinh \frac{x}{N}, \quad (4.8b)$$

and the conditional MGF for  $h_n$ , given  $h_0$ , is simply

$$\begin{aligned} \mathcal{M}_n(x \mid h_0) = & \left( \cosh \frac{x}{N} - X_n \sinh \frac{x}{N} \right)^{\frac{N}{2}(1-h_0)} \\ & \times \left( \cosh \frac{x}{N} + Y_n \sinh \frac{x}{N} \right)^{\frac{N}{2}(1+h_0)}. \end{aligned} \quad (4.9)$$

From this conditional MGF we then obtain the conditional first- and second-



order moments and thus variance,

$$\mathbb{E}[h_n | h_0] = -\frac{1}{2}(1 - h_0)X_n + \frac{1}{2}(1 + h_0)Y_n, \quad (4.10a)$$

$$\text{Var}[h_n | h_0] = \frac{1}{N} \left[ 1 - \frac{1}{2}(1 - h_0)X_n^2 - \frac{1}{2}(1 + h_0)Y_n^2 \right]. \quad (4.10b)$$

The unconditional MGF for  $h_n$  is

$$\mathcal{M}_n(x) = \mathbb{E}[e^{x h_n}] = \left( \cosh \frac{x}{N} + \mu_n \sinh \frac{x}{N} \right)^N, \quad (4.11)$$

where  $\mu_n = \mathbb{E}[h_n]$  is given by

$$\mu_n = -\frac{1}{2}(1 - \mu_0)X_n + \frac{1}{2}(1 + \mu_0)Y_n. \quad (4.12)$$

The second- (variance), third- (skewness) and fourth-order (kurtosis) cumulants of  $h_n$  follow directly from the cumulant generating function  $\log_e \mathcal{M}_n(x)$  and are

$$\sigma_n^2 = \text{Var}[h_n] = \frac{1}{N}(1 - \mu_n^2), \quad (4.13a)$$

$$\text{Skew}^*[h_n] = -\frac{2\mu_n(1 - \mu_n^2)}{N^2}, \quad (4.13b)$$

$$\text{Kurt}^*[h_n] = \frac{2}{N^3}(1 - \mu_n^2)(3\mu_n^2 - 1), \quad (4.13c)$$

where we use the asterisk to indicate that these cumulants are not normalised by appropriate powers of the variance. Because  $h_n$  is essentially binomially-distributed (up to a scaling of its range), we would expect it to become normally-distributed in the large  $N$  limit. Defining  $\hat{h}_n = (h_n - \mu_n)/\sigma_n$ , its

MGF is  $\widehat{\mathcal{M}}_n(x) = \mathcal{M}_n(x/\sigma_n) e^{-x\mu_n/\sigma_n}$ , and we find that

$$\begin{aligned}\widehat{\mathcal{M}}_n(x) &= \left\{ \frac{1+\mu_n}{2} \exp \left[ +\frac{x(1-\mu_n)}{N\sigma_n} \right] + \frac{1-\mu_n}{2} \exp \left[ -\frac{x(1+\mu_n)}{N\sigma_n} \right] \right\}^N \\ &= \left[ 1 + \frac{x^2}{2N} + \mathcal{O}(N^{-3/2}) \right]^N \xrightarrow{N \rightarrow \infty} \exp \left( \frac{1}{2}x^2 \right),\end{aligned}\tag{4.14}$$

which is the MGF of a normal distribution with zero mean and unit variance. Thus,  $h_n$  is approximately normally-distributed for  $N$  large enough, with mean  $\mu_n$  and variance  $\sigma_n^2 = (1 - \mu_n^2)/N$ .

## 4.2 Continuous-Time Generating Functions

The PGFs and MGFs in discrete time may be converted into PGFs and MGFs in continuous time by performing Poisson summation. For a sequence of quantities  $w_n$  in discrete time, we use the convenient notation,

$$\mathcal{P}\{w_n; t\} = \sum_{n=0}^{\infty} e^{-rt} \frac{(rt)^n}{n!} w_n,\tag{4.15}$$

to define the Poisson sum, so that the continuous time quantity  $w(t)$  is simply

$$w(t) = \mathcal{P}\{w_n; t\}.\tag{4.16}$$

With this notation, the continuous time PGFs are

$$\mathcal{G}(t; x \mid h(0) = \frac{2j}{N} - 1) = \mathcal{P}\left\{ \mathcal{G}_n(x \mid h_0 = \frac{2j}{N} - 1); t \right\},\tag{4.17a}$$

$$\mathcal{G}(t; x) = \mathcal{P}\{ \mathcal{G}_n(x); t \},\tag{4.17b}$$

with the conditional PGF inducing the time-dependent transition matrix

$$\mathbb{H}(t) = \mathcal{P}\{\mathbb{H}_n; t\}. \quad (4.18)$$

The continuous time MGFs are

$$\mathcal{M}(t; x \mid h(0)) = \mathcal{P}\{\mathcal{M}_n(x \mid h_0); t\}, \quad (4.19a)$$

$$\mathcal{M}(t; x) = \mathcal{P}\{\mathcal{M}_n(x); t\}. \quad (4.19b)$$

In the above we have replaced  $h_0$  by  $h(0)$  for notational consistency, but they both refer to the perceptron activation immediately after the storage of memory  $\xi^0$ . In general the Poisson sums cannot be evaluated in closed form, but when the probabilities  $p_n^\pm$  are especially simple, or even constants, the first few moments may be calculated exactly.

We may determine the moments  $\mathbb{E}[h(t)^m]$  of  $h(t)$  from Eq. (4.19b). The (un-normalised) cumulants of  $h(t)$  can then be computed from them or directly from  $\log \mathcal{M}(t; x)$ . With the continuous-time mean  $\mu(t) = \mathbb{E}[h(t)]$ , the variance  $\sigma(t)^2 = \mathbb{E}[h(t)^2] - \mu(t)^2$  is standard, and the skewness and kurtosis are given by

$$\text{Skew}^*(t) = \mathbb{E}[h(t)^3] - 3\mu(t) \mathbb{E}[h(t)^2] + 2\mu(t)^3, \quad (4.20a)$$

$$\begin{aligned} \text{Kurt}^*(t) = & \mathbb{E}[h(t)^4] - 4\mu(t) \mathbb{E}[h(t)^3] - 3\mathbb{E}[h(t)^2]^2 \\ & + 12\mu(t)^2 \mathbb{E}[h(t)^2] - 6\mu(t)^4. \end{aligned} \quad (4.20b)$$

These are general forms independent of the particular, perceptron-based model for  $h(t)$  considered here. For completeness, we write out  $\mathbb{E}[h(t)^m]$  for  $m = 2, 3$

and 4 in terms of the underlying synaptic correlation functions:

$$\mathbb{E}[h(t)^2] = \frac{1}{N} + \frac{(N-1)}{N} \mathbb{E}[\tilde{S}_i(t)\tilde{S}_j(t)], \quad (4.21a)$$

$$\mathbb{E}[h(t)^3] = \frac{(3N-2)}{N^2} \mu(t) + \frac{(N-1)(N-2)}{N^2} \mathbb{E}[\tilde{S}_i(t)\tilde{S}_j(t)\tilde{S}_k(t)], \quad (4.21b)$$

$$\begin{aligned} \mathbb{E}[h(t)^4] &= \frac{(3N-2)}{N^3} + \frac{2(N-1)(3N-4)}{N^3} \mathbb{E}[\tilde{S}_i(t)\tilde{S}_j(t)] \\ &\quad + \frac{(N-1)(N-2)(N-3)}{N^3} \mathbb{E}[\tilde{S}_i(t)\tilde{S}_j(t)\tilde{S}_k(t)\tilde{S}_l(t)], \end{aligned} \quad (4.21c)$$

where the expectation values are over distinct doublets, triplets and quadruplets of synapses.

### 4.3 Conditioning on Step $m > 0$

By conditioning on the number of synapses with  $\tilde{S}_i(0) = +1$ , we have computed transition probabilities for the evolution of  $h_n$  and thus obtained the transition matrix  $\mathbb{H}_n$ . We can use this matrix instead to condition on a different step. Specifically, we write

$$\mathbb{H}_{n|m} = \mathbb{H}_n \mathbb{H}_m^{-1}, \quad (4.22)$$

for  $n \geq m$ . The inverse matrix  $\mathbb{H}_m^{-1}$  evolves the state at step  $m$  backwards to the state at step 0, and the matrix  $\mathbb{H}_n$  then evolves this state at step 0 forwards to the state at step  $n$ . The matrix  $\mathbb{H}_{n|m}$  is therefore the evolution matrix for the perceptron activation conditioned on step  $m$  rather than step 0. We note that  $\mathbb{H}_{n|m} \mathbb{H}_{m|l} = \mathbb{H}_{n|l}$ , so that the Markovian property is automatically satisfied; that  $\mathbb{H}_{n|n} = \mathbb{I}$ ; and that  $\mathbb{H}_{n|0} = \mathbb{H}_n$ . The matrices defined by

$$\mathbb{V}_n = \mathbb{H}_{n|n-1} \quad (4.23)$$

then correspond to the 1-step matrices for the transition from  $h_{n-1}$  to  $h_n$ , with

$$\mathbb{H}_n = \mathbb{V}_n \mathbb{V}_{n-1} \cdots \mathbb{V}_2 \mathbb{V}_1. \quad (4.24)$$

This equation is analogous to Eq. (3.10) for single-synapse transitions but applies to transitions in perceptron activation, so to the entire population of synapses. We then have

$$\mathbb{H}_{n|m} = \begin{cases} \mathbb{V}_n \cdots \mathbb{V}_{m+1} & \text{for } n > m \\ \mathbb{I} & \text{for } n = m \end{cases}, \quad (4.25)$$

so that the transition matrix conditioned on step  $m$  is a product of the  $\mathbb{V}$  matrices starting at  $\mathbb{V}_{m+1}$ .

We may explicitly compute the elements of  $\mathbb{H}_{n|m}$ . We know that the PGF  $[\mathcal{G}_n^-(x)]^{N-j} [\mathcal{G}_n^+(x)]^j$  is the generating function for the  $j$ th column of  $\mathbb{H}_n$ , where the coefficients of the two PGFs  $\mathcal{G}_n^\pm(x)$  are just the elements of the matrix  $\mathbb{D}_n$ . We may confirm by direct calculation that the elements of the matrix  $\mathbb{D}_n \mathbb{D}_m^{-1}$  are the coefficients of the two PGFs, call them  $\mathcal{G}_{n|m}^\pm(x)$ , that are required to construct the generating function for the  $j$ th column of  $\mathbb{H}_{n|m}$ ,

$$\mathcal{G}_{n|m}(x \mid h_m = \frac{2j}{N} - 1) = [\mathcal{G}_{n|m}^-(x)]^{N-j} [\mathcal{G}_{n|m}^+(x)]^j. \quad (4.26)$$

The corresponding MGFs are  $\mathcal{M}_{n|m}^\pm(x)$ , with

$$\mathcal{M}_{n|m}(x \mid h_m) = [\mathcal{M}_{n|m}^-(x)]^{\frac{N}{2}(1-h_m)} [\mathcal{M}_{n|m}^+(x)]^{\frac{N}{2}(1+h_m)}. \quad (4.27)$$

We are especially interested in the particular case  $n = m + 1$ , giving the 1-step transition matrix  $\mathbb{V}_{m+1}$  for transitions in perceptron activation between suc-

cessive memory storage steps. In this case, we may use the recurrence relations in Eq. (3.13) to simplify the expressions for  $\mathcal{G}_{m+1|m}^\pm(x)$  and  $\mathcal{M}_{m+1|m}^\pm(x)$ . We obtain

$$\mathcal{G}_{m+1|m}^-(x) = \left(1 - \frac{1}{2} p_m^+\right) x^0 + \frac{1}{2} p_m^+ x^1, \quad (4.28a)$$

$$\mathcal{G}_{m+1|m}^+(x) = \frac{1}{2} p_m^- x^0 + \left(1 - \frac{1}{2} p_m^-\right) x^1, \quad (4.28b)$$

and

$$\mathcal{M}_{m+1|m}^-(x) = \cosh \frac{x}{N} - (1 - p_m^+) \sinh \frac{x}{N}, \quad (4.29a)$$

$$\mathcal{M}_{m+1|m}^+(x) = \cosh \frac{x}{N} + (1 - p_m^-) \sinh \frac{x}{N}, \quad (4.29b)$$

which lead to

$$\mathbb{E}[h_{m+1} | h_m] = -\frac{1}{2}(1 - h_m)(1 - p_m^+) + \frac{1}{2}(1 + h_m)(1 - p_m^-), \quad (4.30a)$$

$$\text{Var}[h_{m+1} | h_m] = \frac{1}{N} \left[ 1 - \frac{1}{2}(1 - h_m)(1 - p_m^+)^2 - \frac{1}{2}(1 + h_m)(1 - p_m^-)^2 \right]. \quad (4.30b)$$

Notice that the two generating functions  $\mathcal{G}_{m+1|m}^-(x)$  and  $\mathcal{G}_{m+1|m}^+(x)$ , which generate the 1-step transition matrix  $\mathbb{V}_{m+1}$  for transitions in the perceptron activation via  $\mathcal{G}_{m+1|m}(x | h_m = \frac{2j}{N} - 1)$ , also generate the two columns of the 1-step transition matrix  $\mathbb{W}_{m+1}$  in Eq. (3.9) for transitions in a single synapse's strength. In this balanced scenario, therefore, the PGF  $\mathcal{G}_{m+1|m}(x | h_m = \frac{2j}{N} - 1)$  generates both the 1-step transitions for single synapses (with  $N = 1$ ) and the 1-step transitions in perceptron activation. This equivalence between the two processes only occurs for balanced processes.

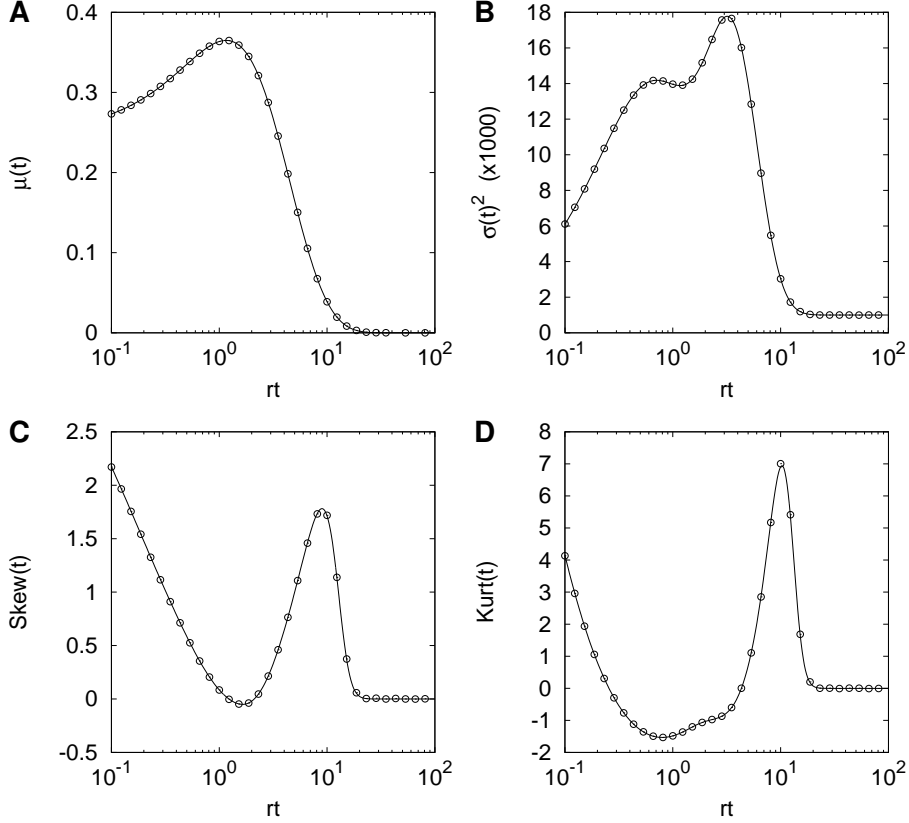
## 4.4 Comparison Between Analytical and Simulation

### Results

With these analytical expressions for the generating functions for  $h(t)$ , we may compare the statistics of  $h(t)$  obtained analytically to those obtained from simulations of full, filter-based synapses. Full details of our simulation protocols may be found elsewhere (Elliott & Lagogiannis, 2012; Elliott, 2014). Simulation results are averaged over  $10^6$  separate simulations in order to obtain good statistics.

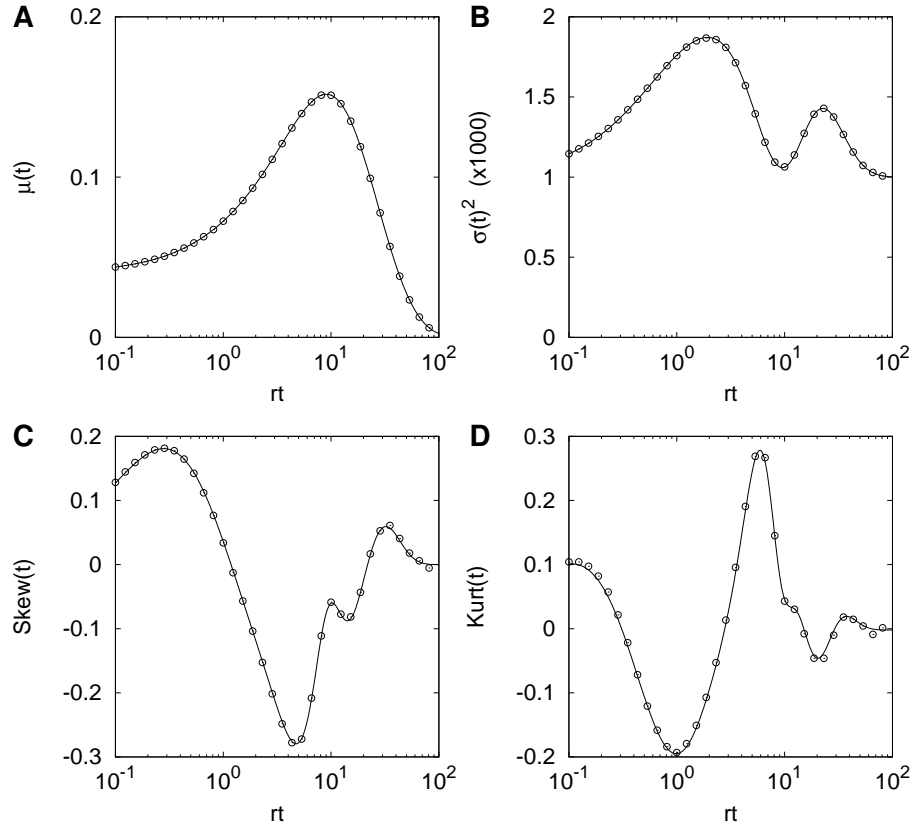
In Fig. 3 (for  $\Theta = 2$ ) and Fig. 4 (for  $\Theta = 5$ ) we plot analytical results for the statistics of  $h(t)$  based on the continuous-time MGF in Eq. (4.19b) and simulation results in which synapses undergo full, filter-based transitions rather than the simplified, purely strength-based transitions associated with conditional strength-change probabilities. We see essentially exact agreement between our analytical results and simulation results averaged over a large enough ensemble. There is a little more noise in the simulation results for  $\Theta = 5$  for the skewness and kurtosis compared to those for  $\Theta = 2$ . This is to be expected as there are more internal states in the former case, so for larger  $\Theta$  larger numbers of individual trials are necessary, especially for higher-order statistics. The essentially exact agreement confirms that simplified synapses with step-dependent strength-change probabilities are completely equivalent to full, filter-based synapses. It also serves to verify our analytical results and validate our simulations protocols.

In computing the moments in Figs. 3 and 4, we have truncated the Poisson sums at a point beyond which further contributions are negligible. Because the Poisson sums cannot in general be evaluated exactly, this procedure constitutes a good numerical approximation scheme. With this numerical approximation



**Figure 3:** Analytical and simulation results for the statistics of  $h(t)$  in continuous time. Analytical results (solid lines) for (A)  $\mu(t)$ , (B)  $\sigma(t)^2$  and the normalised (C) skewness and (D) kurtosis are obtained from the MGF for  $h(t)$ . We have set  $N = 10^3$  and used a filter size of  $\Theta = 2$ . Simulation results (indicated by circles) are obtained by running full simulations of synapses with internal filter-based transitions and averaging over  $10^6$  separate simulations in order to obtain good statistics.





**Figure 4:** Analytical and simulation results for the statistics of  $h(t)$ . The format of this figure is identical to Fig. 3 except that we have used a filter size of  $\Theta = 5$ .

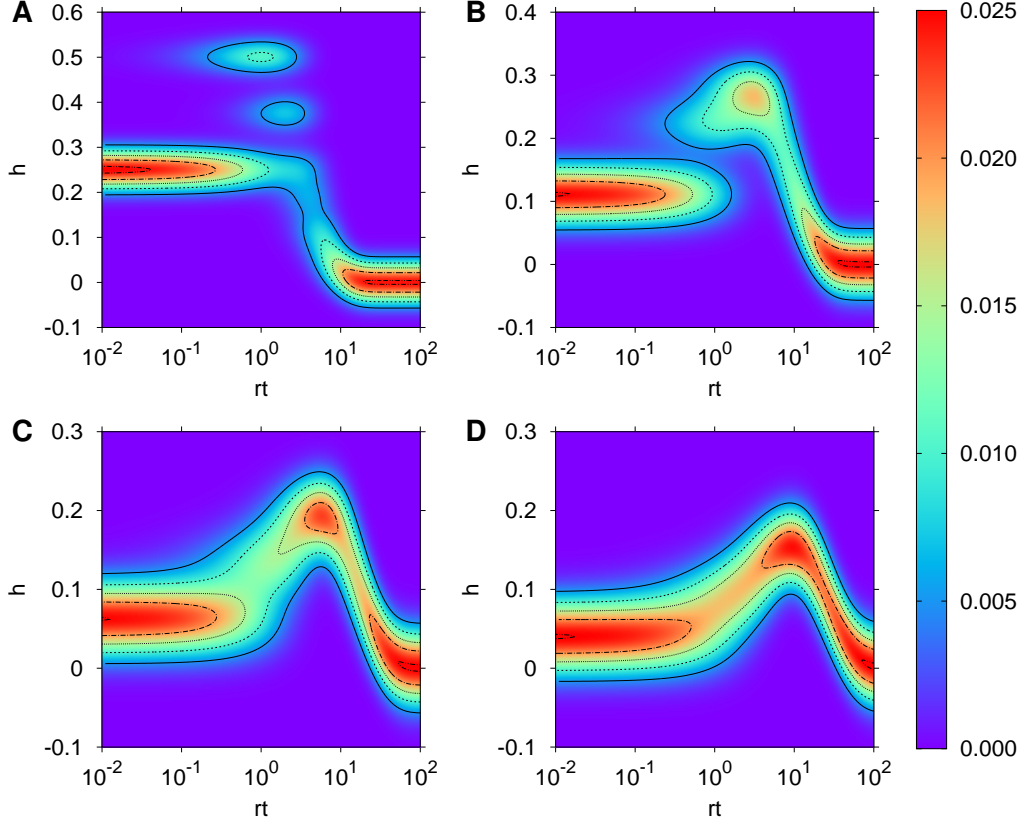
understood, we are able to compute any moment to any desired accuracy using Eq. (4.19b). This computation relies critically on the use of the simplified synaptic dynamics with conditional strength-change probabilities. Although we can derive exact expressions for  $\mu(t)$  using the full, filter-based synaptic dynamics, the calculation of  $\sigma(t)^2$  and specifically the covariance between pairs of synapses' strengths is extremely hard in general (Elliott & Lagogiannis, 2012) and has been possible only for small enough  $\Theta$  (up to around 6, depending on the amount of memory available to computer algebra packages). In general, in order to compute the  $m$ th-order moment, we would need to compute the  $m$ th-order correlation function over  $m$  synapses' (tilded) strengths (see Eq. (4.21)). For the full, filter-based synaptic dynamics, this means considering the tensor product  $\mathbb{M} \otimes \cdots \otimes \mathbb{M}$  with  $m$  occurrences of  $\mathbb{M}$ . Unfortunately, for our filter-based model of synaptic plasticity, the matrix  $\mathbb{M}$  is defective (i.e. it lacks a complete set of eigenvectors), so it is not possible to write down the spectrum of such tensor products directly from the spectrum of  $\mathbb{M}$ . It is therefore necessary to flatten tensor products down to a single matrix, and for  $m$  products, we would obtain a matrix of size  $[2(2\Theta - 1)]^m \times [2(2\Theta - 1)]^m$ . Even for  $\Theta = 2$ , computing the kurtosis, which requires  $m = 4$ , results in a  $1296 \times 1296$  matrix; for  $\Theta = 3$ , this becomes  $10^4 \times 10^4$ . For anything but very small  $\Theta$  and very small  $m$ , such matrices become unmanageably large. Even when a general transition matrix  $\mathbb{M}$  is not defective, the resulting  $m$ -fold sums over eigenvectors and eigenvalues become time consuming. The use of the equivalent TDSU formulation with conditional strength-change probabilities and no internal synaptic states considerably ameliorates these computational problems and permits moments to be calculated to any desired accuracy where otherwise such calculations would be intractable or numerically very intensive.

The PGF for  $h(t)$  in Eq. (4.17b) encodes the entire probability distribution,

call it  $P(h, t)$ , for the perceptron activation or tracked memory signal  $h$  over time  $t$ . Using the same truncation procedure for Poisson sums as above, we may determine  $P(h, t)$  essentially exactly. In Fig. 5 we show superimposed heat and contour maps for  $P(h, t)$  in the  $rt$ - $h$  plane for different choices of  $\Theta$ . Of course,  $h$  takes discrete values in the interval  $[-1, +1]$  with spacing  $2/N$ , but for  $N$  large enough, it appears essentially continuous. We have selected  $N = 10^3$  in this figure. We see clearly for early times the contributions from separate memory storage steps coming in with weightings determined by the relevant factors in the Poisson sums. These contributions are particularly clear for  $\Theta = 2$ . The two isolated “islands” of probability for  $\Theta = 2$  are concentrated around  $rt = 1$  and  $rt = 2$ , which correspond precisely to the locations of the maxima of  $e^{-rt}(rt)^n/n!$  for  $n = 1$  and  $n = 2$ . These early-time contributions serve to increase the mean  $\mu(t)$  by pulling the distribution  $P(h, t)$  towards larger values of  $h$ . They arise because of the imbalance between the probabilities  $p_n^+$  and  $p_n^-$  seen in Fig. 2. This imbalance itself reflects the biasing of filter states by the storage of memory  $\xi^0$ , so that synapses experiencing  $\xi_i^0 = +1$  are initially biased to become or remain strong while those experiencing  $\xi_i^0 = -1$  are initially biased to become or remain weak. When this biasing works out of the system, the distribution of  $h$  relaxes back to being symmetrically distributed around  $h = 0$ , as can be seen for larger times in this figure.

## 5 Fokker-Planck Equation

Although we have determined the evolution of  $h_n$  and thus  $h(t)$  exactly, it is convenient to consider an approximation based on the Fokker-Planck equation. The first- and second-order moments obtained from the Fokker-Planck equation are exact (if the jump moments can be calculated exactly), while higher-



**Figure 5:** Superimposed heat and contour maps for the probability distribution  $P(h, t)$  of the perceptron activation or tracked memory signal,  $h$ , over time,  $t$ . Results are shown for  $N = 10^3$  synapses for filters of size: (A)  $\Theta = 2$ , (B)  $\Theta = 3$ , (C)  $\Theta = 4$ , (D)  $\Theta = 5$ . Contours are shown for probabilities of 0.005, 0.010, 0.015, 0.020 and 0.025, which can be identified from the heat map colour box on the right.

order cumulants in general are not. The higher-order cumulants given by the Fokker-Planck equation are identically zero. This is because either the underlying stochastic process is normally-distributed and the Fokker-Planck equation is in fact a formally-derivable limit; or the Fokker-Planck equation is based on a second-order truncation of the Kramers-Moyal expansion for processes that are not normally-distributed. As  $h_n$  is approximately normally-distributed for  $N$  large enough, it may appear that a Fokker-Planck approach would therefore become exact in the large  $N$  limit. However, the Fokker-Planck equation is also based on a diffusion approximation, while the dynamics of  $h_n$  contains jump processes. Despite the limiting normality of  $h_n$ , a Fokker-Planck approach is thus only ever at best a diffusion-limit approximation to the dynamics of  $h_n$ . It is convenient to consider this diffusion limit because in section 6 we will implement an absorbing boundary when solving the Fokker-Planck equation. This allows us to obtain MFPTs for memory lifetimes in the limit in which jump processes are ignored.

## 5.1 Evolution of the Moments

Before we derive the jump moments for perceptron activation, we first derive exact equations for the evolution of the first- and second-order moments of  $h(t)$  and thus for the evolution of its variance. We can of course compute the moments exactly and directly from the results above, but we obtain equations for their evolution because they will validate our results for the jump moments.

Consider a sequence of values of some quantity  $w$  at each memory storage step in discrete time,  $w_n$  for  $n = 0, 1, 2, \dots$ , and thus its continuous-time form  $w(t) = \mathcal{P}\{w_n; t\}$ . Computing  $\mathcal{P}\{w_{n+1}; t\} = \sum_{n=0}^{\infty} \frac{(rt)^n}{n!} e^{-rt} w_{n+1}$  in which the

sequence is offset by one step, we find

$$\mathcal{P}\{w_{n+1}; t\} - \mathcal{P}\{w_n; t\} \equiv \mathcal{P}\{w_{n+1} - w_n; t\} = \frac{1}{r} \frac{dw(t)}{dt}, \quad (5.1)$$

so that performing a Poisson sum over differences between consecutive steps therefore just returns the time derivative.

To determine  $d\mu(t)/dt$ , we therefore need  $\mu_{n+1} - \mu_n$ . We have the initial condition  $\mu_0 = \mu(0) = 1/\Theta^2$ . From Eq. (4.12) and Eq. (3.13), we obtain

$$\mu_{n+1} - \mu_n = \frac{1}{2}(1 - \mu_n)p_n^+ - \frac{1}{2}(1 + \mu_n)p_n^-. \quad (5.2)$$

The conditional probabilities  $p_n^\pm$  were defined in Eq. (3.8), with the denominators  $\mathbf{n} \cdot \mathbf{w}_n$  and  $\mathbf{n} \cdot \mathbf{s}_n$  being the probabilities of the synapse being weak or strong, respectively. But these latter probabilities are just  $\frac{1}{2}(1 \mp \mu_n)$ , so we can rewrite Eq. (5.2) in the form

$$\mu_{n+1} - \mu_n = f_n^+ - f_n^-, \quad (5.3)$$

where  $f_n^+ = \mathbf{n} \cdot \mathbb{T}^+ \mathbf{w}_n$  and  $f_n^- = \mathbf{n} \cdot \mathbb{T}^- \mathbf{s}_n$ . From the structures of the matrices  $\mathbb{T}^\pm$ , the quantities  $f_n^\pm$  are just the probabilities for a synapse being in filter state  $I = +(\Theta - 1)$  when it is weak or in filter state  $I = -(\Theta - 1)$  when it is strong, respectively. These probabilities are computed in discrete time directly from  $\mathbb{M}^n$  acting on the initial state immediately after the storage of memory  $\boldsymbol{\xi}^0$ , which for balanced processes we may take to be just  $\mathbb{M}^+ \mathbf{A}$ . Thus, they are given directly by the relevant components of the vector  $\mathbb{M}^n \mathbb{M}^+ \mathbf{A}$ . When we perform a Poisson sum over Eq. (5.3) to obtain  $d\mu(t)/dt$ ,  $f_n^\pm$  will turn into  $f^\pm(t) = \mathcal{P}\{f_n^\pm; t\}$ , where equivalently  $f^\pm(t)$  may be determined explicitly from the relevant components of the continuous-time state vector given by

$\exp(rt \mathbb{G}) \mathbb{M}^+ \mathbf{A}$ , since  $\mathcal{P}\{\mathbb{M}^n; t\} \equiv \exp(rt \mathbb{G})$  with  $\mathbb{G} = \mathbb{M} - \mathbb{I}$  (cf. Eq. (2.14)).

Eq. (5.3) therefore becomes, under Poisson summation,

$$\frac{1}{r} \frac{d\mu(t)}{dt} = f^+(t) - f^-(t). \quad (5.4)$$

We now define the continuous-time conditional strength-change probabilities via the equation

$$p^\pm(t) = \frac{f^\pm(t)}{\frac{1}{2}[1 \mp \mu(t)]}, \quad (5.5)$$

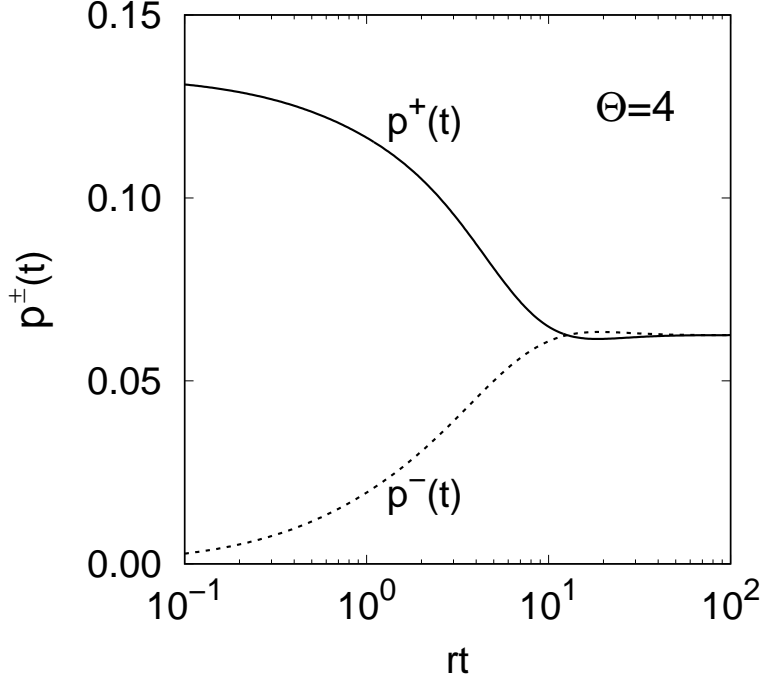
where the two denominators  $\frac{1}{2}[1 \mp \mu(t)]$  are just the continuous-time probabilities for a synapse to be weak or strong, respectively.<sup>3</sup> The continuous-time mean strength  $\mu(t)$  may be computed from the discrete-time mean strengths  $\mu_n$  via  $\mathcal{P}\{\mu_n; t\}$  or, equivalently, directly from  $\boldsymbol{\Omega}^T \exp(rt \mathbb{G}) \mathbb{M}^+ \mathbf{A}$ . With these definitions, Eq. (5.4) becomes

$$\frac{1}{r} \frac{d\mu(t)}{dt} = \frac{1}{2}[1 - \mu(t)]p^+(t) - \frac{1}{2}[1 + \mu(t)]p^-(t), \quad (5.6)$$

with  $\mu(0) = 1/\Theta^2$ , which governs the evolution of  $\mu(t)$  in continuous time. Although the RHSs of Eqs. (5.2) and (5.6) are structurally identical, it is not the case that we have effectively moved from Eq. (5.2) to Eq. (5.6) by replacing the Poisson sum of a product by the product of the Poisson sums: such a replacement would in general be invalid. Rather, the RHSs of Eqs. (5.2) and (5.6) are structurally identical purely because they unpack to the state probabilities  $f_n^\pm$

---

<sup>3</sup> It is critical to note that  $p^\pm(t)$  are given by  $\mathcal{P}\{f_n^\pm; t\}/\mathcal{P}\{\frac{1}{2}[1 \mp \mu_n]\}$  and are *not* given by  $\mathcal{P}\{f_n^\pm/\frac{1}{2}[1 \mp \mu_n]\}$ : the continuous-time conditional strength-change probabilities are the ratios of state probabilities in continuous time to strength probabilities in continuous time.



**Figure 6:** Conditional strength-change probabilities  $p^{\pm}(t)$  as functions of time for  $\Theta = 4$ .

and  $f^{\pm}(t)$ , respectively, which are derived from  $\mathbb{M}^n$  and  $\mathcal{P}\{\mathbb{M}^n; t\} = \exp(rt \mathbb{G})$ , respectively. We note that for dynamics with fixed conditional strength-change probabilities of just  $p_n^{\pm} = p^{\pm}$  (so essentially just a simple synapse with no internal states), Poisson summing Eq. (5.2) does indeed lead directly to Eq. (5.6), with  $p^{\pm}(t) = p^{\pm}$ , as we should expect.

In Fig. 6 we plot  $p^{\pm}(t)$  against  $t$  for the same choices of parameters as for  $p_n^{\pm}$  in Fig. 2. Although we have stressed the fact that the continuous-time conditional strength-change probabilities are not just the Poisson-summed discrete-time conditional strength-change probabilities (see footnote 3), we see nevertheless that  $p^{\pm}(t)$  are very similar in overall profile to  $p_n^{\pm}$ , with the jaggedness in the latter pair being smoothed out in the former pair.

Because all the higher-order moments of  $h_n$  depend only on  $\mu_n$ , the determination of their evolution in terms of  $\mu_n$  is in principle straightforward. The



variance  $\sigma_n^2$  is given in Eq. (4.13a) as  $\sigma_n^2 = (1 - \mu_n^2)/N$ . The second-order moment is just  $\mathbb{E}[h_n^2] = \sigma_n^2 + \mu_n^2$ , which we denote for convenience by  $\varphi_n$ . We have that

$$\varphi_n = \frac{1}{N} + \frac{N-1}{N} \mu_n^2, \quad (5.7)$$

and so by using Eq. (5.2) we obtain

$$\begin{aligned} \varphi_{n+1} - \varphi_n &= \frac{N-1}{N} \left[ \frac{1}{2}(1 - \mu_n)p_n^+ - \frac{1}{2}(1 + \mu_n)p_n^- \right] \\ &\quad \times \left[ 2\mu_n + \frac{1}{2}(1 - \mu_n)p_n^+ - \frac{1}{2}(1 + \mu_n)p_n^- \right]. \end{aligned} \quad (5.8a)$$

We would in general expect the RHS to depend only linearly on both the first- and second-order moments  $\mu_n$  and  $\varphi_n$  (see below), but deriving this equation by exploiting the identity in Eq. (5.7) has effectively replaced the second-order moment  $\varphi_n$  by expressions involving  $\mu_n^2$  on the RHS. We therefore rewrite the RHS by replacing the  $\mu_n^2$  term by  $(N\varphi_n - 1)/(N - 1)$ , to obtain

$$\begin{aligned} \varphi_{n+1} - \varphi_n &= \frac{1}{2}(p_n^+ + p_n^-) \left[ \frac{1}{2}(p_n^+ + p_n^-) - 2 \right] \varphi_n \\ &\quad - \frac{N-1}{N} \left[ \frac{1}{2}(1 - p_n^+)^2 - \frac{1}{2}(1 - p_n^-)^2 \right] \mu_n \\ &\quad + \frac{1}{N} \left[ 1 - \frac{1}{2}(1 - p_n^+)^2 - \frac{1}{2}(1 - p_n^-)^2 \right] + \frac{1}{4}(p_n^+ - p_n^-)^2. \end{aligned} \quad (5.8b)$$

Poisson summing either Eq. (5.8a) or (5.8b) will give us  $d\varphi(t)/dt$  where  $\varphi(t) = \mathcal{P}\{\varphi_n; t\}$ . The two resulting equations for  $d\varphi(t)/dt$  are exact and completely equivalent, but it does not appear to be possible in general to evaluate the Poisson sums on the RHSs and express them purely in terms of  $\mu(t)$  and  $\varphi(t)$ , even when we use the state probabilities  $f_n^\pm = \frac{1}{2}(1 \mp \mu_n)p_n^\pm$  that allowed Eq. (5.2) to be summed. For the particular case that  $p_n^\pm = p^\pm$  are constants, the Poisson sums can be explicitly evaluated.

In Fig. 7 we show for  $N = 10^3$  synapses and various choices of  $\Theta$  that the mean  $\mu(t)$  determined from Eq. (5.6) and the variance  $\sigma(t)^2$  determined via  $\varphi(t)$  from the Poisson-summed Eq. (5.8b) agree exactly with  $\mu(t)$  computed analytically (Elliott & Lagogiannis, 2012) and  $\sigma(t)^2$  determined explicitly from numerical matrix methods. To determine the evolution of  $\varphi(t)$  from Eq. (5.8b), we have as above truncated the Poisson sum at a point beyond which further contributions are negligible. The differential equations are solved by standard numerical methods.

Because we cannot Poisson sum Eq. (5.8b) explicitly in the general case of non-constant  $p_n^\pm$ , in order to make progress we perform an approximation in the following manner. Let the vector  $\mathbf{P}_n$  contain the probability distribution of  $h_n$ , so that the  $i$ th component of  $\mathbf{P}_n$  is  $\text{Prob}[h_n = \frac{2i}{N} - 1]$ . Then

$$\mathbf{P}_{n+1} = \mathbb{V}_{n+1} \mathbf{P}_n, \quad (5.9)$$

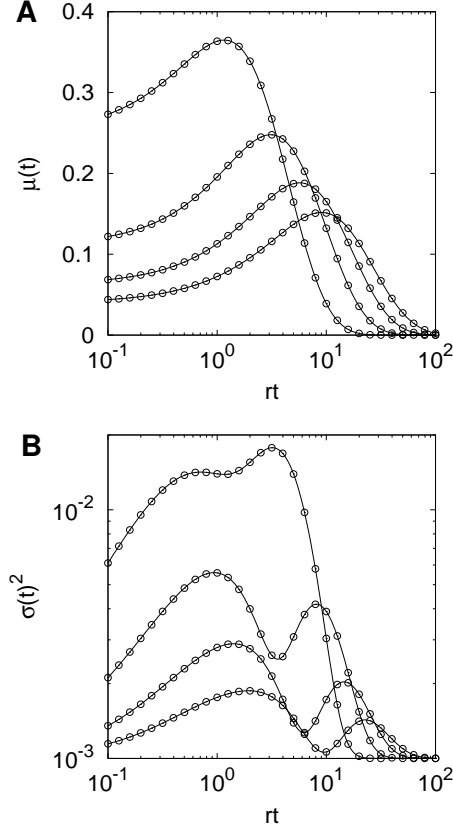
where the elements of  $\mathbb{V}_{n+1}$  depend only on  $p_n^\pm$  from Eq. (4.28). We may use this equation to compute the changes in the moments of  $h_n$  from step to step. For example, for  $N = 2$ , we may write  $\mathbf{P}_n$  in the form

$$\mathbf{P}_n = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & +1 \\ +1 & 0 & +1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \mu_n \\ \phi_n \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(\varphi_n - \mu_n) \\ 1 - \varphi_n \\ \frac{1}{2}(\varphi_n + \mu_n) \end{pmatrix}, \quad (5.10)$$

which gives the required zeroth-, first- and second-order moments of 1,  $\mu_n$  and  $\varphi_n$ , respectively.<sup>4</sup> Eq. (5.9), which is then purely linear in the moments, then

---

<sup>4</sup>For general  $N$  the matrix whose inverse is given on the RHS of Eq. (5.10) has element  $(\frac{2j}{N} - 1)^i$  in row  $i$  and column  $j$ , with the understanding that for  $i = 0$ , these elements are always unity even if  $2j/N - 1 = 0$ . These elements are just



**Figure 7:** Mean and variance in perceptron activation computed either directly or from their associated differential equations. Solid lines show results computed directly while the circles show results computed from differential equations. (A) Means  $\mu(t)$  for  $\Theta = 2, 3, 4$  and  $5$ , moving from top to bottom in the figure; (B) variances  $\sigma(t)^2$  for the same values of  $\Theta$  as in panel A, again moving from top to bottom. Results are shown for  $N = 10^3$  synapses.

explicitly gives Eqs. (5.2) and (5.8b) for  $N = 2$  and in general for arbitrary  $N$  we obtain the correct overall factors in Eqs. (5.8b). From Eq. (5.9), we have  $\mathbf{P}_{n+1} - \mathbf{P}_n = (\mathbb{V}_{n+1} - \mathbb{I})\mathbf{P}_n$ , or

$$\frac{1}{r} \frac{d\mathbf{P}(t)}{dt} = \mathcal{P}\{\mathbb{V}_{n+1}\mathbf{P}_n; t\} - \mathbf{P}(t), \quad (5.11)$$

but we also have from Eq. (5.9) that  $\mathbf{P}_n = \mathbb{H}_n\mathbf{P}_0$  using Eq. (4.24), or

$$\mathbf{P}(t) = \mathcal{P}\{\mathbf{P}_n; t\} = \mathcal{P}\{\mathbb{H}_n; t\}\mathbf{P}(0) \equiv \mathbb{H}(t)\mathbf{P}(0). \quad (5.12)$$

While it appears that this solution of Eq. (5.11) implies that the Poisson sum on the RHS of Eq. (5.11) can be evaluated, explicitly doing so merely results in the essentially tautological and completely useless  $d\mathbf{P}(t)/dt = d\mathbb{H}(t)/dt\mathbf{P}(0)$ .

Except for the case when  $p_n^\pm = p^\pm$ , the Poisson sum on the RHS of Eq. (5.11) cannot therefore be usefully evaluated. However, the structure of this equation suggests that a natural approximation is to replace the Poisson sum  $\mathcal{P}\{\mathbb{V}_{n+1}\mathbf{P}_n; t\}$  with  $\mathbb{V}(t)\mathbf{P}(t)$ , where the time-dependent matrix  $\mathbb{V}(t)$  is generated from the time-dependent probabilities  $p^\pm(t)$ , so that column  $j$  is generated by

$$\begin{aligned} \sum_{i=0}^N x^i [\mathbb{V}(t)]_{ij} &= \left\{ \left[ 1 - \frac{1}{2}p^+(t) \right] x^0 + \frac{1}{2}p^+(t)x^1 \right\}^{N-j} \\ &\quad \times \left\{ \frac{1}{2}p^-(t)x^0 + \left[ 1 - \frac{1}{2}p^-(t) \right] x^1 \right\}^j. \end{aligned} \quad (5.13)$$

We therefore replace the exact distribution  $\mathbf{P}(t) = \mathbb{H}(t)\mathbf{P}(0)$  obtained from

---

all relevant powers of the possible values of the perceptron's activation. This matrix is in fact a transposed Vandermonde matrix. The vector multiplying this inverse matrix in general has components  $\mathbb{E}[h_n^i]$ .

the Poisson sum  $\mathcal{P}\{\mathbf{P}_n; t\}$  by the approximated distribution obtained from the solution of

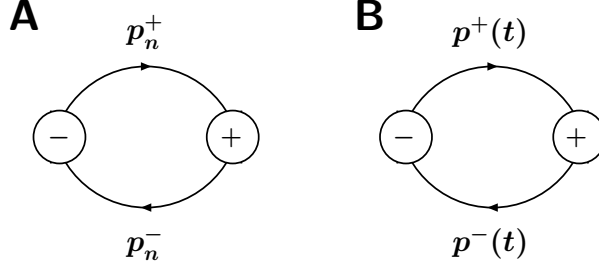
$$\frac{1}{r} \frac{d\mathbf{P}(t)}{dt} = [\mathbb{V}(t) - \mathbb{I}] \mathbf{P}(t), \quad (5.14)$$

with  $\mathbb{V}(t)$  depending on  $p^\pm(t)$ . This approximation is graphically represented in Fig. 8, in which we replace the single-synapse strength-change transitions in Fig. 8A by those shown in Fig. 8B. This approximation in effect elevates the continuous-time probabilities  $p^\pm(t)$  induced by the underlying discrete-time probabilities  $p_n^\pm$  to the status of fully and completely defining the dynamics. Of course, when  $p^\pm(t) = p_n^\pm = p^\pm$  are constants, then this approximation reproduces the exact dynamics, with a Poisson sum over Eq. (5.9) yielding directly Eq. (5.14) with  $\mathbb{V}(t) = \mathbb{V}$ , a constant matrix. In this case, the solution of Eq. (5.14) is  $\mathbf{P}(t) = \mathbb{H}(t) \mathbf{P}(0)$  with  $\mathbb{H}(t) = \exp [rt(\mathbb{V} - \mathbb{I})]$ . In general, however, the solution of Eq. (5.14) gives us the approximated evolution matrix

$$\mathbb{H}(t) = \mathcal{T} \left\{ \exp r \int_0^t d\tau [\mathbb{V}(\tau) - \mathbb{I}] \right\}, \quad (5.15)$$

where  $\mathcal{T}\{\cdot\}$  denotes the time-ordered product (meta-)operator, although in practice we must solve the differential equation in Eq. (5.14) numerically.

With this approximation, we may use Eq. (5.14) to compute the evolution of the moments of  $h(t)$ . Expanding  $\mathbf{P}(t)$  out in terms of its continuous-time moments similarly to Eq. (5.10), we obtain Eq. (5.6) exactly, so that at the level of the dynamics in the mean  $\mu(t)$ , the approximated dynamics reproduce



**Figure 8:** Step- and time-dependent changes in synaptic strength. (A) The full, unapproximated dynamics consist of step-dependent probabilities  $p_n^\pm$  for changes in synaptic strength and thus a master equation for each step for the change in state probabilities per step. (B) We may instead approximate these dynamics by time-dependent probabilities  $p^\pm(t)$  and a single master equation with these time-dependent probabilities. These two processes are only equivalent when  $p_n^\pm = p^\pm(t) = p^\pm$  are constants, independent of the step or time.

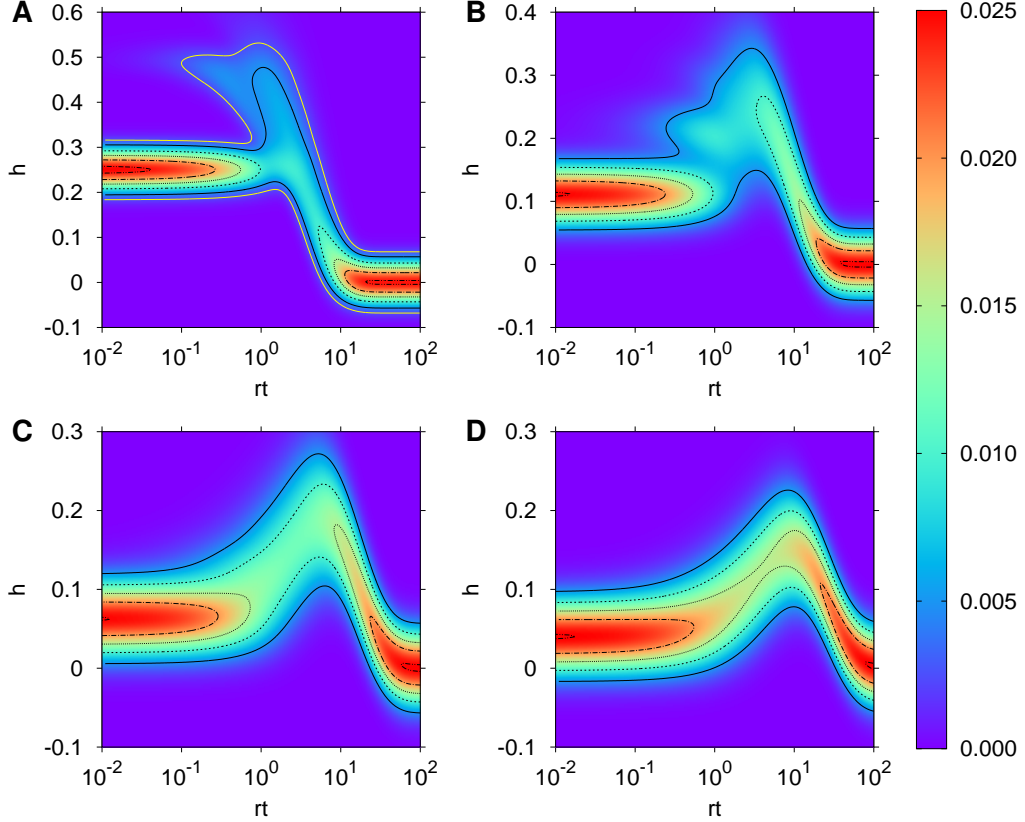
the exact dynamics. For the second-order moment, we obtain

$$\begin{aligned}
\frac{1}{r} \frac{d\varphi(t)}{dt} = & \frac{1}{2} [p^+(t) + p_n^-(t)] \left\{ \frac{1}{2} [p^+(t) + p^-(t)] - 2 \right\} \varphi(t) \\
& - \frac{N-1}{2N} \left\{ [1 - p^+(t)]^2 - [1 - p^-(t)]^2 \right\} \mu(t) \\
& + \frac{1}{N} \left\{ 1 - \frac{1}{2} [1 - p^+(t)]^2 - \frac{1}{2} [1 - p^-(t)]^2 \right\} \\
& + \frac{1}{4} [p^+(t) - p^-(t)]^2,
\end{aligned} \tag{5.16}$$

whose RHS is of course structurally identical to Eq. (5.8b). The result of this approximation is that Poisson sums of products have been replaced by the products of Poisson sums. Further, the approximation has equated  $p^\pm(t)$  with the Poisson sums  $\mathcal{P}\{p_n^\pm; t\}$ , which is incorrect. However, because  $p^\pm(t)$  are defined as ratios of Poisson sums, the approximation has thus essentially replaced a ratio of Poisson sums with a Poisson sum of ratios. The approximation therefore consists in the assumption that Poisson sums factorise both multiplicatively and divisively.

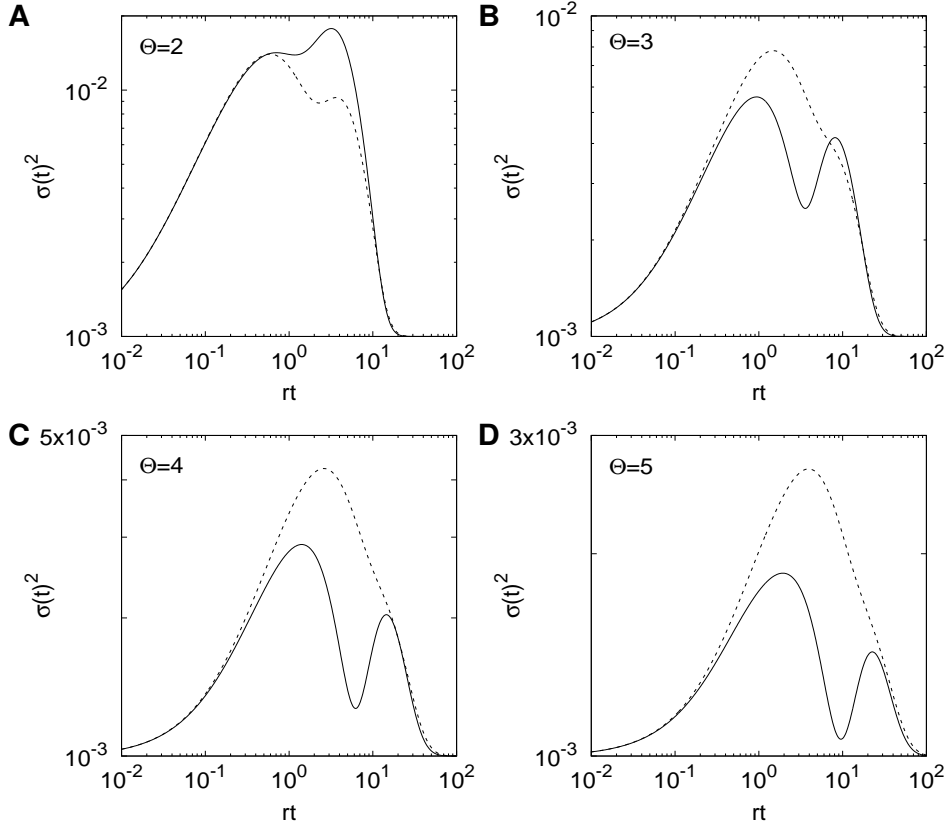
We may numerically solve Eq. (5.14) governing these approximated dynamics to obtain the entire distribution  $P(h, t)$ , and compare it to the distribution for the exact dynamics shown in Fig. 5. From Eq. (5.13), the elements of the matrix  $\mathbb{V}(t)$  are sums of products of binomial coefficients and involve the probabilities  $p^\pm(t)$ . We use  $N = 10^3$ , which is large enough to replace these elements with normal distributions, making numerical solutions more tractable for large  $N$ . Fig. 9 shows the resulting distributions for different choices of  $\Theta$ . Comparing these distributions for the approximated dynamics governed by Eq. (5.14) to those for the exact dynamics in Fig. 5, we see good qualitative agreement. The approximated dynamics tend to smooth out the isolated islands of probability that arise in Fig. 5 as each separate Poisson mode develops, but we clearly see in Fig. 9 for  $\Theta = 2$  and  $\Theta = 3$  these concentrated regions of probability developing at earlier times for larger  $h$  and then connecting with the other regions for smaller  $h$  at later times.

The approximated dynamics reproduce exactly the exact dynamics at the level of the mean but the variances differ. We can see this by comparing Figs. 5 and 9. For example, considering a slice through these maps for  $rt \approx 5$  for  $\Theta = 4$  or  $\Theta = 5$ , it is clear that the approximated dynamics have somewhat broader ranges of non-zero probability than the exact dynamics. By explicitly computing the variance in the approximated dynamics via Eq. (5.16), we directly compare the variances for the exact and approximated dynamics in Fig. 10. For  $\Theta = 2$ , the approximated dynamics underestimate the exact variance while for  $\Theta > 2$  the approximated dynamics (mostly) overestimate the exact variance. The small- and large-time behaviours are in agreement, but the intermediate-time behaviours differ somewhat. Specifically, in the vicinity of the peak in the mean memory signal, the variance in the exact dynamics exhibits a minimum, but for  $\Theta > 2$  the variance in the approximated dynamics



**Figure 9:** Superimposed heat and contour maps for the probability distribution  $P(h, t)$  of the perceptron activation or tracked memory signal,  $h$ , over time,  $t$ , obtained from the approximated dynamics governed by Eq. (5.14). The format of this figure is identical to Fig. 5 except that we have added an additional contour in yellow corresponding to a probability of 0.0025 in panel A, for  $\Theta = 2$ .





**Figure 10:** Comparison between the variance in the exact and approximated dynamics. The variance in the exact dynamics is shown with solid lines while that in the approximated dynamics is shown with dashed lines. Each panel shows results for the indicated value of  $\Theta$ . These variances are determined with  $N = 10^3$  synapses.

is larger and only exhibits an inflexion or undulation at this same location.

## 5.2 Jump Moments

Assuming that we have a standard Fokker-Planck equation of the form

$$\frac{1}{r} \frac{\partial}{\partial t} P(h, t) = -\frac{\partial}{\partial h} [A(h, t) P(h, t)] + \frac{1}{2} \frac{\partial^2}{\partial h^2} [B(h, t) P(h, t)], \quad (5.17)$$

its solution would determine the entire probability distribution  $P(h, t)$  of the perceptron activation or tracked memory signal  $h$  over time  $t$ . The use of a Fokker-Planck equation entails a continuum limit for  $h$ , which for large enough  $N$  is valid. If  $P(h', t' | h, t)$  is the transition probability from  $h$  at time  $t$  to  $h'$  at time  $t' \geq t$ , then the first- and second-order jump moments  $A(h, t)$  and  $B(h, t)$  are determined from

$$M_l(h, t) = \frac{1}{r} \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \int dh' (h' - h)^l P(h', t + \delta t | h, t), \quad (5.18)$$

with  $A(h, t) = M_1(h, t)$  and  $B(h, t) = M_2(h, t)$ .<sup>5</sup> The first- and second-order moments  $\mu(t)$  and  $\varphi(t)$  are then determined from Eq. (5.17) via the equations

$$\frac{1}{r} \frac{d\mu(t)}{dt} = \mathbb{E}[A(h, t)], \quad (5.19a)$$

$$\frac{1}{r} \frac{d\varphi(t)}{dt} = \mathbb{E}[B(h, t)] + 2 \mathbb{E}[hA(h, t)]. \quad (5.19b)$$

We must determine whether we do indeed have a standard Fokker-Planck equation and if so, we must compute the jump moments  $A(h, t)$  and  $B(h, t)$  for perceptron activation and show that they lead to the required evolution equations for  $\mu(t)$  and  $\varphi(t)$  derived in section 5.1.

Eq. (5.11) is a master equation in continuous time for the discrete probability distribution  $\mathbf{P}(t)$  for  $h$  at time  $t$ . The Fokker-Planck equation would be derived from this equation either in certain formal limits or as a truncation of the Kramers-Moyal expansion. However, Eq. (5.11) is itself derived as a

---

<sup>5</sup>Because we prefer to retain an overall factor of  $1/r$  on the left-hand side of the Fokker-Planck equation so that the rate  $r$  manifestly acts as a simple scale factor for time, we have modified this standard definition of the jump moments by dividing through by the rate.

Poisson sum over the discrete-time equation

$$\mathbf{P}_{m+1} - \mathbf{P}_m = (\mathbb{V}_{m+1} - \mathbb{I})\mathbf{P}_m. \quad (5.20)$$

The matrix operator  $\mathbb{V}_{m+1} - \mathbb{I}$  on the RHS induces the jumps in the moments between successive steps, giving the change in the distribution from  $\mathbf{P}_m$  to  $\mathbf{P}_{m+1}$ . Heuristically, in order to move to a continuous probability distribution  $P_m(h)$  at step  $m$  when  $h$  takes continuous rather than discrete values, we merely have to replace the matrix operator  $\mathbb{V}_{m+1} - \mathbb{I}$  with a differential operator that induces the same jump moments. This differential operator is of course precisely that which occurs on the RHS of the Fokker-Planck equation. The matrix  $\mathbb{V}_{m+1}$  induces the conditional first-order moment and variance given in Eq. (4.30). The matrix  $\mathbb{V}_{m+1} - \mathbb{I}$  acting on  $\mathbf{P}_m$  induces the jump moments  $E[(h_{m+1} - h_m)^l | h_m]$ , which can be written down directly from Eq. (4.30). Thus, defining

$$A_m(h) = \frac{1}{2}(1 - h)p_m^+ - \frac{1}{2}(1 + h)p_m^-, \quad (5.21a)$$

$$B_m(h) = \frac{1}{N} \left[ 1 - \frac{1}{2}(1 - h)(1 - p_m^+)^2 - \frac{1}{2}(1 + h)(1 - p_m^-)^2 \right] + \left[ \frac{1}{2}(1 - h)p_m^+ - \frac{1}{2}(1 + h)p_m^- \right]^2, \quad (5.21b)$$

we replace Eq. (5.20) with

$$P_{m+1}(h) - P_m(h) = -\frac{\partial}{\partial h} [A_m(h)P_m(h)] + \frac{1}{2} \frac{\partial^2}{\partial h^2} [B_m(h)P_m(h)], \quad (5.22)$$

in order to move to the continuum limit. We may explicitly confirm by direct

calculation that using Eq. (5.22), we obtain

$$\mu_{m+1} - \mu_m = \mathbb{E}[A_m(h)], \quad (5.23a)$$

$$\varphi_{m+1} - \varphi_m = \mathbb{E}[B_m(h)] + 2\mathbb{E}[hA_m(h)], \quad (5.23b)$$

where the expectation values on the RHSs are evaluated over the probability distribution  $P_m(h)$  with first- and second-order moments  $\mu_m$  and  $\varphi_m$ , respectively, and that the RHSs reproduce the results in Eqs. (5.2) and (5.8b).

The form in Eq. (5.22) is a discrete-time analogue of the Fokker-Planck equation. While it resembles a one-step forward Euler method for obtaining numerical solutions, we stress that it is not. It is an equation that governs the change in distribution of  $h$  between successive memory storage steps, which occur as a Poisson process. In order to move to continuous time, we must of course Poisson sum this equation, obtaining

$$\frac{1}{r} \frac{\partial}{\partial t} P(h, t) = -\frac{\partial}{\partial h} [\mathcal{P}\{A_m(h)P_m(h); t\}] + \frac{1}{2} \frac{\partial^2}{\partial h^2} [\mathcal{P}\{B_m(h)P_m(h); t\}]. \quad (5.24)$$

This equation is not a standard Fokker-Planck equation because the presence of the Poisson sums on the RHS prevents us from pulling out the distribution  $P(h, t)$  from these sums. This failure of factorisation reflects the failure of factorisation in the original master equation in Eq. (5.11). Only when  $A_m(h)$  and  $B_m(h)$  are independent of  $m$  and so constant in time do we obtain a standard Fokker-Planck equation with time-independent jump moments  $A(h)$  and  $B(h)$ . Nevertheless, Eq. (5.24) does necessarily reproduce the correct differential equations for  $\mu(t)$  and  $\varphi(t)$  and so it does generate the correct first- and second-order moments for the tracked memory signal  $h(t)$  in continuous time.

Because of this failure of factorisation, we must move to the approximated dynamics governed by Eq. (5.14). The time-dependent matrix  $\mathbb{V}(t) - \mathbb{I}$  on the RHS of Eq. (5.14) induces jump moments that may be computed directly from the generating function for the columns of  $\mathbb{V}(t)$  in Eq. (5.13). These are of course completely analogous to the jump moments computed from  $\mathbb{V}_{m+1} - \mathbb{I}$ . Thus, we may immediately write down the jump moments for the approximated dynamics as

$$A(h, t) = \frac{1}{2}(1 - h)p^+(t) - \frac{1}{2}(1 + h)p^-(t), \quad (5.25a)$$

$$B(h, t) = \frac{1}{N} \left\{ 1 - \frac{1}{2}(1 - h)[1 - p^+(t)]^2 - \frac{1}{2}(1 + h)[1 - p^-(t)]^2 \right\} \\ + \left[ \frac{1}{2}(1 - h)p^+(t) - \frac{1}{2}(1 + h)p^-(t) \right]^2. \quad (5.25b)$$

These jump moments appear in the standard form of the Fokker-Planck equation stated in Eq. (5.17). Since  $\mathbb{E}[h] = \mu(t)$  and  $\mathbb{E}[h^2] = \varphi(t)$ , we may directly verify that we obtain the correct equations for  $d\mu(t)/dt$  and  $d\varphi(t)/dt$  for the approximated dynamics from Eq. (5.19).

Before considering solutions of this Fokker-Planck equation for the approximated dynamics, we examine the structure of its jump moments by considering the asymptotic behaviour of the continuous-time, conditional strength-change probabilities  $p^\pm(t)$ . In the limit  $t \rightarrow \infty$ , filter and strength states return to equilibrium. The probabilities of a filter being in states  $\pm(\Theta - 1)$  in equilibrium for balanced processes are both just  $1/(2\Theta^2)$ , and the strength probabilities are both  $\frac{1}{2}$ . Thus, we have  $p^\pm(t) \rightarrow p_\infty$  where  $p_\infty = 1/\Theta^2$  is the equilibrium conditional strength-change probability for both weak and strong synapses. We also know that  $p^-(0) = 0$ . Examining Fig. 6, as a qualitatively good approximation

we see that we can write

$$p^-(t) = p_\infty (1 - e^{-\lambda_- rt}), \quad (5.26)$$

where  $\lambda_-$  controls the rate at which  $p^-(t)$  returns to its equilibrium value. This form does not capture the finer details of the slight overshooting of  $p^-(t)$  beyond  $p_\infty$  at intermediate times, but it is sufficient for our purposes here. For  $p^+(0)$ , we have that  $p^+(0) = 2/(\Theta^2 - 1)$  for  $\Theta > 1$ . We write this value as  $p_0$  for simplicity. Again, then, we may write down a qualitatively good approximation for  $p^+(t)$  as

$$p^+(t) = p_\infty + (p_0 - p_\infty)e^{-\lambda_+ rt}, \quad (5.27)$$

where  $\lambda_+$  is the analogue of  $\lambda_-$ . We set  $\lambda_\pm = \lambda$  since from Fig. 6 both  $p^\pm(t)$  appear to return to equilibrium at the same rate, as we might expect from the symmetry of processes with balanced plasticity. We now separate the equilibrium behaviour of the jump moments from the initial transients by writing

$$A(h, t) = A_\infty(h) + A_1(h)e^{-\lambda rt}, \quad (5.28a)$$

$$B(h, t) = B_\infty(h) + B_1(h)e^{-\lambda rt} + B_2(h)e^{-2\lambda rt}. \quad (5.28b)$$

We obtain

$$A_\infty(h) = -p_\infty h, \quad (5.29a)$$

$$B_\infty(h) = p_\infty^2 h^2 + \frac{1}{N} [1 - (1 - p_\infty)^2], \quad (5.29b)$$

and

$$A_1(h) = \frac{1}{2}p_0(1-h) + p_\infty h, \quad (5.30a)$$

$$B_1(h) = \frac{1}{N}(1-p_\infty)[p_0(1-h) - 2p_\infty] - p_\infty h[p_0(1-h) + 2p_\infty h], \quad (5.30b)$$

$$B_2(h) = \frac{1}{N}[p_0 p_\infty(1-h) - \frac{1}{2}(1-h)p_0^2 - p_\infty^2] + [\frac{1}{2}p_0(1-h) + p_\infty h]^2. \quad (5.30c)$$

In the absence of the transients, we have jump moments that define a fixed probability stochastic updater with update probability  $p_\infty = 1/\Theta^2$ . Since  $\mu(0) = 1/\Theta^2 = p_\infty$ , this interpretation is consistent as the mean initial signal for such a synapse is indeed just the update probability. We have previously studied MFPTs for memory lifetimes with such dynamics extensively (Elliott, 2014), and obtained precisely these jump moments (see Eq. (7.5) in Elliott (2014), which is essentially the backward Kolmogorov equation, used for determining passage times). We showed that the dynamics of a fixed probability stochastic updater are essentially just those of the Ornstein-Uhlenbeck (OU) process. The transients modify these dynamics. Although  $A_\infty(h)$  pulls  $h$  towards zero,  $A_1(h)$  increases  $h$ : initially, the full jump moment  $A(h, 0)$  is  $\frac{1}{2}p_0(1-h)$ , which pulls  $h$  towards unity. The balance between  $A_\infty(h)$  and  $A_1(h)e^{-\lambda r t}$  determines whether  $h$  is pulled towards unity or towards zero, with the equilibrium pull towards zero eventually overcoming the transient pull towards unity as the transient decays. We can see these two competing influences explicitly by writing  $A(h, t)$  as

$$A(h, t) = \frac{1}{2}p_0(1-h)e^{-\lambda r t} - p_\infty h(1 - e^{-\lambda r t}).$$

At  $t = 0$ , there is a stable fixed point at  $h = 1$ , at the level of the Liouville

dynamics, while in equilibrium, the stable fixed point is at  $h = 0$ . At intermediate times, there is a stable fixed point at an intermediate value of  $h$ . A filter-based synapse therefore modifies the equilibrium OU-like dynamics by superimposing a transient that pulls  $h$  towards unity.

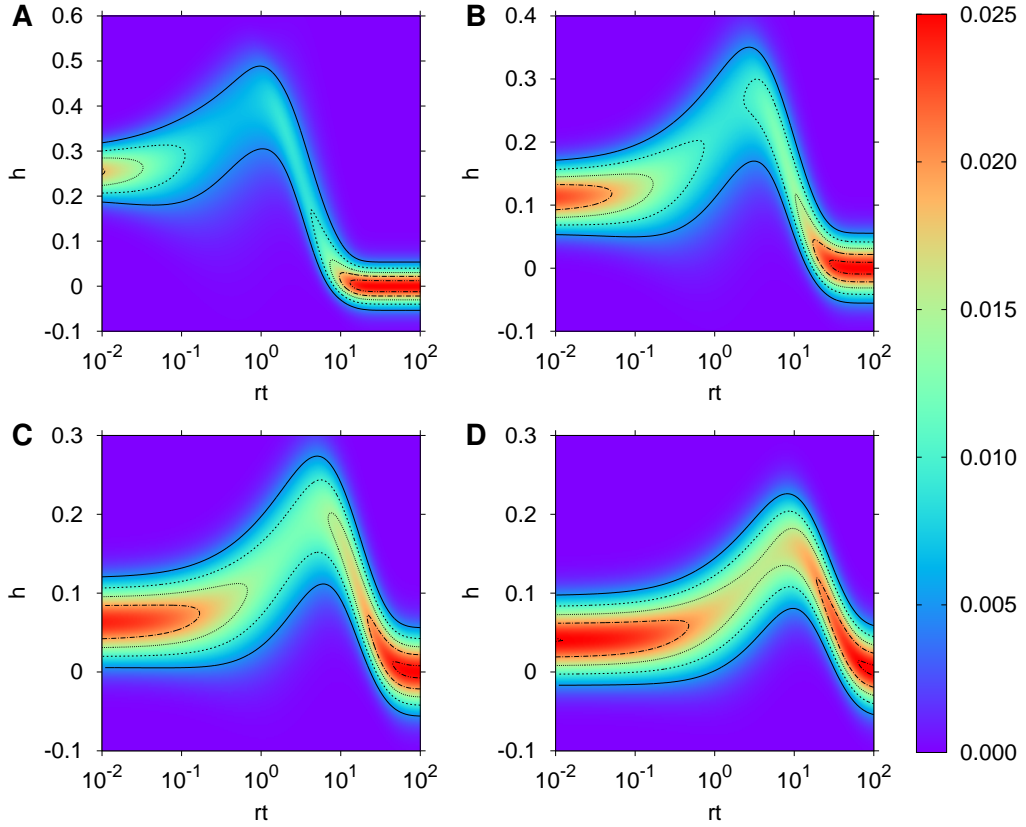
### 5.3 Solutions of Fokker-Planck Equation

We are primarily concerned with using the Fokker-Planck equation for the approximated dynamics in the presence of an absorbing boundary in order to obtain MFPTs for perceptron activation to fall below a defined threshold. Here we therefore only briefly consider unconstrained solutions of the Fokker-Planck, i.e. solutions in the absence of an absorbing boundary.

The Fokker-Planck equation with jump moments in Eq. (5.25) almost certainly cannot be solved analytically. We therefore use standard numerical methods for solving this partial differential equation. To check the integrity of a numerical solution for any given choice of parameters, we verify that it is always non-negative and normalised to an integral of unity. The exact numerical agreement between the mean and variance of the numerically-obtained probability distribution and the mean and variance obtained from the differential equations in Eqs. (5.6) and (5.16) also validates the numerical solution.

In Fig. 11 we show superimposed heat and contour maps for numerical solutions of the Fokker-Planck equation for the approximated dynamics for different choices of filter size. This figure should be compared to Fig. 9, which shows results for the approximated dynamics obtained directly from the master equation in Eq. (5.14), from which the Fokker-Planck equation is obtained as a continuum (and diffusion) limit. We see that for  $\Theta = 4$  and  $\Theta = 5$  (and in general for larger values of  $\Theta$ ), the Fokker-Planck and master equation solutions



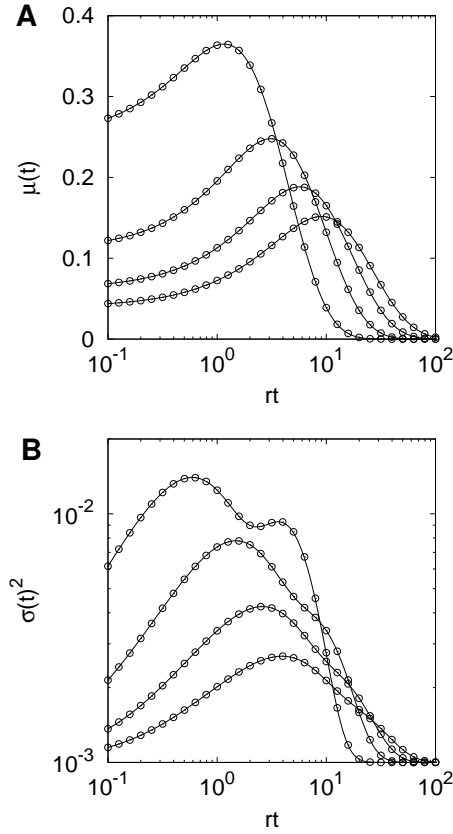


**Figure 11:** Superimposed heat and contour maps for the probability distribution  $P(h, t)$  of the perceptron activation or tracked memory signal,  $h$ , over time,  $t$ , obtained from the Fokker-Planck equation for the approximated dynamics. The format of this figure is identical to Fig. 5.

are very similar, with only relatively small qualitative differences in the overall structures of the solutions. For  $\Theta = 2$  and  $\Theta = 3$ , however, we clearly see the impact of the diffusion approximation. At small times, the solutions from the master equation remain tightly focused around their initial mean. These small-time solutions of the master equation are governed by the initial Poisson mode, which decays as  $e^{-rt}$ , arising from the storage of the tracked memory  $\xi^0$ . As time progresses, subsequent Poisson modes corresponding to the storage of subsequent memories arise in the master equation solution, and these are clearly seen in Fig. 9 (and even more clearly seen in the solutions of the exact dynamics in Fig. 5). However, in the Fokker-Planck equation solutions in Fig. 11, we see that at small times, the solution diffuses outwards from its initial state, and that the entire solution drifts upwards towards higher values of  $h$  rather than developing islands of non-zero probability that then connect with other regions as time progresses. In the regions where the master equation solutions exhibit clearly visible distinct Poisson modes, the Fokker-Planck solutions appear quite diffuse, with probability being spread over a wide range of  $h$  rather than concentrated in distinct, non-overlapping regions. Nevertheless, the mean and variance of the distributions from the Fokker-Planck and master equation solutions must agree, and this is confirmed in Fig. 12.

## 6 Mean First Passage Times

We now consider memory lifetimes defined by the MFPT for the perceptron activation  $h(t)$  to fall below a defined threshold  $\vartheta$ , which could be the perceptron's firing threshold. For balanced potentiation and depression processes,  $\mu(t) \rightarrow 0$  as  $t \rightarrow \infty$ , so we will typically consider a threshold  $\vartheta = 0$  but we will also consider thresholds  $\vartheta > 0$ .



**Figure 12:** Mean and variance in perceptron activation determined either from numerical solutions of the Fokker-Planck equations or from the differential equations governing the evolution of the moments of the approximated dynamics. Solid lines show results from the Fokker-Planck equation while the circles show results from the differential equations. The format of this figure is otherwise identical to Fig. 7.

Previously we gave a general derivation of the lifetime  $\tau_{\text{mfpt}}(\xi^0)$  of memory  $\xi^0$  defined by the MFPT for  $h(t)$  to fall below  $\vartheta$  in a general model with internal synaptic states (see Eq. (3.7) in Elliott (2014)). This result, although exact, is essentially useless for explicit computations, either analytical or numerical, unless  $N$  is very small. However, if we have a transition matrix for changes in the perceptron activation at each memory storage step, as here with the step-dependent matrix  $\mathbb{V}_m$ , then we can instead work directly with these transitions in  $h$  rather than the underlying transitions in the internal states of all  $N$  synapses. The matrix  $\mathbb{V}_m$  in general allows transitions from any value of  $h_{m-1}$  to any value of  $h_m$ . However, we wish to consider transitions between states  $h_{m-1} > \vartheta$  and  $h_m > \vartheta$  in order to impose the threshold on perceptron activation and obtain MFPTs: transitions to disallowed states with  $h_m \leq \vartheta$  should be excluded. Defining the  $(N+1) \times (N+1)$  diagonal matrix

$$\mathbb{P}_{h>\vartheta} = \text{diag}\{ \underbrace{0, \dots, 0}_{h \leq \vartheta \text{ states}} \mid \underbrace{1, \dots, 1}_{h > \vartheta \text{ states}} \}, \quad (6.1)$$

the product  $\mathbb{P}_{h>\vartheta} \mathbb{V}_m$  then imposes the requirement that only transitions to states with  $h > \vartheta$  are allowed, with the states  $h \leq \vartheta$  being projected out. By using the projection property  $\mathbb{P}_{h>\vartheta}^2 = \mathbb{P}_{h>\vartheta}$ , we observe that, for example,

$$(\mathbb{P}_{h>\vartheta} \mathbb{V}_m)(\mathbb{P}_{h>\vartheta} \mathbb{V}_{m-1}) \cdots = (\mathbb{P}_{h>\vartheta} \mathbb{V}_m \mathbb{P}_{h>\vartheta})(\mathbb{P}_{h>\vartheta} \mathbb{V}_{m-1} \mathbb{P}_{h>\vartheta}) \cdots ,$$

so that we may consider the transition matrix between allowed states to be  $\mathbb{P}_{h>\vartheta} \mathbb{V}_m \mathbb{P}_{h>\vartheta}$ , which is essentially just the relevant sub-block of  $\mathbb{V}_m$  for transitions between allowed states.

We define the vector  $\mathbf{P}(h_0)$  to correspond to an initial state in which  $h_0$  takes a definite value, so that  $\mathbf{P}(h_0)$  has an entry of unity at position  $i =$

$\frac{N}{2}(1 + h_0)$  and zeros elsewhere. If  $h_0 > \vartheta$ , then  $\mathbb{P}_{h>\vartheta}\mathbf{P}(h_0) = \mathbf{P}(h_0)$ , but if  $h_0 \leq \vartheta$ , then  $\mathbb{P}_{h>\vartheta}\mathbf{P}(h_0) = \mathbf{0}$ . We then have that

$$\psi_m(h_0) = \mathbf{n}^T(\mathbb{P}_{h>\vartheta}\mathbb{V}_m\mathbb{P}_{h>\vartheta}) \cdots (\mathbb{P}_{h>\vartheta}\mathbb{V}_1\mathbb{P}_{h>\vartheta})[\mathbb{P}_{h>\vartheta}\mathbf{P}(h_0)] \quad (6.2)$$

is the probability that  $h > \vartheta$  during and after the storage of memories  $\boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^m$  from the definite initial state  $h_0$ ; we define  $\psi_0(h_0) = \mathbf{n}^T\mathbb{P}_{h>\vartheta}\mathbf{P}(h_0)$ , with  $\psi_0(h_0) \in \{0, 1\}$  depending on whether  $h_0 \leq \vartheta$  or  $h_0 > \vartheta$ . To include  $\boldsymbol{\xi}^0$ , we merely average  $h_0$  over the initial distribution induced by the storage of  $\boldsymbol{\xi}^0$ . Any states with  $h_0 \leq \vartheta$  are automatically excluded in this averaging. Because  $\mathbb{P}_{h>\vartheta} \neq \mathbb{I}$ ,  $\psi_m(h_0) \rightarrow 0$  as  $m \rightarrow \infty$ , so that it is inevitable that  $h$  will eventually fall (to or) below  $\vartheta$ . The sequence of probabilities  $\psi_0(h_0) \geq \psi_1(h_0) \geq \psi_2(h_0) \geq \dots$  is monotonic decreasing, and the difference  $\psi_m(h_0) - \psi_{m+1}(h_0) \geq 0$  is the probability that  $h$  falls (to or) below  $\vartheta$  at step  $m + 1$ . The MFPT for  $h$  to fall (to or) below  $\vartheta$  from some definite initial state  $h_0$  is then

$$\tau_{\text{mfpt}}(h_0) = \sum_{m=0}^{\infty} (m+1) [\psi_m(h_0) - \psi_{m+1}(h_0)] = \sum_{m=0}^{\infty} \psi_m(h_0). \quad (6.3)$$

In continuous time, we obtain an identical result, up to an overall rate factor appearing on the left-hand side (Elliott, 2014). For  $\mathbb{V}_m = \mathbb{V}$ , a constant matrix independent of  $m$ , the sum in Eq. (6.3) can be evaluated exactly although a matrix inverse must be computed. In general, we must evaluate each probability  $\psi_m(h_0)$  and sum them numerically, terminating the sum when further contributions become negligible. The numerical evaluation of  $\tau_{\text{mfpt}}(h_0)$  and its average  $\langle \tau_{\text{mfpt}}(h_0) \rangle$ , averaged over the initial distribution of  $h_0$  induced by  $\boldsymbol{\xi}^0$ , is feasible for  $N$  up to around  $10^4$ . When we obtain MFPTs from simulations,

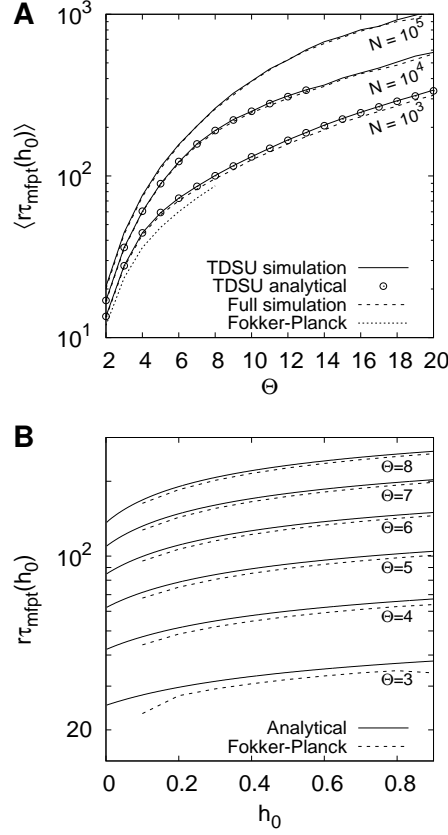
we automatically obtain  $\langle \tau_{\text{mfpt}}(h_0) \rangle$  because we always average over a sufficient number of trials to obtain good statistics.

We may also obtain  $\tau_{\text{mfpt}}(h_0)$  for the approximated dynamics from the Fokker-Planck equation. To do so, we erect an absorbing boundary at  $h = \vartheta$  and solve the Fokker-Planck equation subject to the boundary condition  $P(\vartheta, t) = 0$  and the initial condition  $P(h, 0) = \delta(h - h_0)$  for  $h_0 > \vartheta$ , where  $\delta(x)$  is the Dirac delta function. For  $\langle \tau_{\text{mfpt}}(h_0) \rangle$ , we may change the initial condition to  $P(h, 0) = H(h - \vartheta)P_0(h)$ , where  $H(x)$  is the Heaviside step function, which cuts off the distribution below the boundary at  $h = \vartheta$ , and  $P_0(h)$  is the initial distribution of  $h$  induced by the storage of  $\xi^0$ . We may take  $P_0(h)$  to be a normal distribution with mean  $\mu_0 = 1/\Theta^2$  and variance  $\sigma_0^2 = (1 - \mu_0^2)/N$ . The change in the total probability  $\int_{\vartheta}^{\infty} dh P(h, t)$  gives the MFPT via

$$\tau_{\text{mfpt}}(h_0) = - \int_0^{\infty} dt \, t \frac{d}{dt} \int_{\vartheta}^{\infty} dh P(h, t) = \int_0^{\infty} dt \int_{\vartheta}^{\infty} dh P(h, t), \quad (6.4)$$

which is essentially equivalent to Eq. (6.3). Numerically, when the Fokker-Planck equation is spatially discretised, the probability that moves onto the boundary at  $h = \vartheta$  at each discretised time step, before it is set to zero, gives the change in total probability and may, when weighted by the current time, be explicitly summed to give  $\tau_{\text{mfpt}}(h_0)$ .

In Fig. 13, we show results for  $\langle \tau_{\text{mfpt}}(h_0) \rangle$  and for  $\tau_{\text{mfpt}}(h_0)$  for a threshold  $\vartheta = 0$  and different filter sizes  $\Theta$ . For  $\langle \tau_{\text{mfpt}}(h_0) \rangle$ , in Fig. 13A, we obtain analytical results using Eq. (6.3) for  $N = 10^3$  and  $N = 10^4$ ; results from the Fokker-Planck equation for  $N = 10^3$  and only for smaller values of the filter size  $\Theta$ ; and results by running simulations of either the full, filter-based dynamics without reduction to a simple synapse with step-dependent update probabilities or the reduced, simple dynamics. For the analytical results, we use



**Figure 13:** Comparison of results for memory lifetimes defined by mean first passage times (MFPTs). (A) The MFPT-defined memory lifetime  $\tau_{\text{mfpt}}(h_0)$  averaged over the binomial distribution of the initial memory signal  $h_0$ ,  $\langle \tau_{\text{mfpt}}(h_0) \rangle$ , as a function of filter size  $\Theta$ . Results are shown for  $N = 10^3$ ,  $N = 10^4$  and  $N = 10^5$  synapses as indicated, for simulations of the time-dependent stochastic updater (TDSU) (solid lines); analytical results for the TDSU (circles); full simulations of filter-based dynamics without reduction to the TDSU form (dash lines); numerical results from the Fokker-Planck equation with an absorbing boundary at  $h = 0$  (dotted lines). (B) Memory lifetimes  $\tau_{\text{mfpt}}(h_0)$  as a function of the definite initial state  $h_0$ , so *not* averaged over the distribution of  $h_0$ . Results are shown for  $N = 10^3$  synapses for the different choices of  $\Theta$  as indicated, either for analytical results from the TDSU (solid lines) or for numerical results from the Fokker-Planck equation with an absorbing boundary at  $h = 0$  (dashed lines).

a normal approximation to the binomial probabilities in the transition matrix. We have shown previously that this is an extremely good approximation that allows  $N$  to be pushed to higher values numerically (Elliott, 2014). Even so, for  $N = 10^4$  we have obtained results only up to  $\Theta = 13$  as convergence of the sum in Eq. (6.3) takes too long for larger values of  $\Theta$ . For simulations, we typically average over  $10^8/N$  different trials to obtain good statistics. We see basically exact agreement between the analytical results for the reduced dynamics and their simulation results. Any disagreement is almost entirely due to noise in simulations of a stochastic process and can be removed by averaging over more trials to obtain even better statistics. Any disagreement due to the normal approximation is essentially negligible. The Fokker-Planck results, which are based on the approximated dynamics, qualitatively agree with the other results, but they tend to underestimate  $\langle \tau_{\text{mfpt}}(h_0) \rangle$ . This underestimate is largely due to the increased variance in the approximated dynamics compared to the exact dynamics (see Fig. 10). We have obtained results for the Fokker-Planck equation only for  $N = 10^3$  and smaller  $\Theta$  because it is very hard to obtain stable numerical solutions of the equation for larger parameter choices as these require pushing the numerical solutions out to larger and larger times, for which numerical instabilities tend to develop. We see a small discrepancy between full and reduced simulation results. The discrepancy is larger for smaller  $N$ . This discrepancy is real and not merely due to noisy simulations. We shall explain it momentarily. In Fig. 13B, we plot  $\tau_{\text{mfpt}}(h_0)$  as a function of the definite initial value  $h_0$  for exact results for the reduced, simple dynamics and for the Fokker-Planck equation, for different choices of filter size  $\Theta$  as indicated. Despite the noted slight underestimate of  $\tau_{\text{mfpt}}$  by the Fokker-Planck equation, we see that the quantitative agreement is good and the qualitative agreement, in terms of following the trend in the exact results, is essentially



perfect. The quantitative disagreement is larger for smaller  $h_0$  and smaller  $\Theta$ , for which jump processes, not included in the Fokker-Planck equation because of the diffusion approximation, are more significant.

We observe that the MFPTs in Fig. 13A increase monotonically with increasing  $\Theta$ . This monotonic increase appears to continue indefinitely. If we consider an SNR definition of memory lifetimes, then using the slowest decaying mode in Eq. (2.21), a good estimate of SNR memory lifetime,  $\tau_{\text{snr}}$ , is

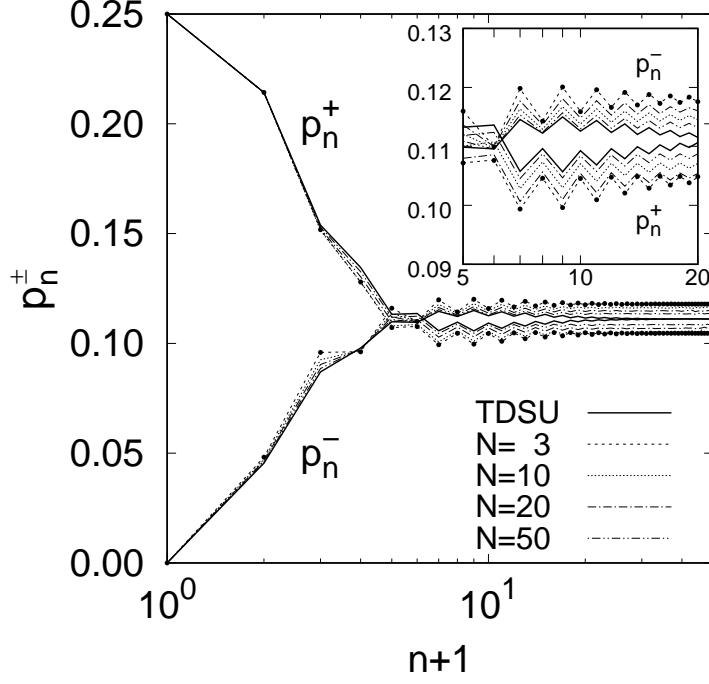
$$r\tau_{\text{snr}} = \frac{4\Theta^2}{\pi^2} \log_e \left( \frac{256}{\pi^4} \frac{N}{\Theta^2} \right). \quad (6.5)$$

Contrary to  $\tau_{\text{mfpt}}$ ,  $\tau_{\text{snr}}$  exhibits a maximum as a function of  $\Theta$ , with the maximum occurring at  $\Theta = \frac{16}{\pi^2} \sqrt{\frac{N}{e}}$ , giving a maximum SNR memory lifetime of  $1024N/(e\pi^6)$ , where  $e$  is the base of natural logarithms. This difference between  $\tau_{\text{snr}}$  and  $\tau_{\text{mfpt}}$  has been examined before (Elliott, 2016a). Although  $\tau_{\text{mfpt}}$  increases indefinitely as a function of  $\Theta$ , the variance in the first passage times grows as  $\Theta$  increases, so that the MFPT eventually becomes indistinguishable from zero at the level of one standard deviation in the first passage time (Elliott, 2016a). Nevertheless, while  $\tau_{\text{snr}}$  appears to exhibit what may be termed a network size effect, this effect may be entirely an artifact of the SNR definition of memory lifetimes. Similar effects in other models of complex synapses are likely similarly artifactual.

The discrepancy between results for the MFPTs for the full and reduced dynamics in Fig. 13A appears to contradict the exact agreement in the statistics of  $h(t)$  explicitly demonstrated in Figs. 3 and 4 and thus perhaps to cast doubt on the claimed complete equivalence between full synaptic dynamics and synaptic dynamics with internal states integrated out, as discussed in section 3. In obtaining these reduced, simple dynamics with the step-dependent

conditional probabilities  $p_n^\pm$  above, it is implicit that the statistical ensemble over which the probabilities  $p_n^\pm$  are defined does not change as a function of the time step  $n$ . The ensemble is time-independent. For time-independent ensembles, the equivalence between the full, complex synaptic dynamics and the reduced, simple, time-dependent synaptic dynamics is indeed exact. Critically, however, in the presence of an absorbing boundary, used to determine MFPTs, the ensemble changes over time. Specifically, when any particular perceptron in the ensemble crosses the  $h = \vartheta$  threshold, it is removed from the ensemble so that it no longer contributes to the determination of memory lifetimes beyond that point. Any particular perceptron that crosses the threshold can do so only by at least one of its synapses changing strength; all of its synapses change their filter states in response to any memory storage step. The perceptron is removed from the ensemble and therefore the subsequent evolution of its filter states no longer contributes to the determination of  $p_n^\pm$  for later time steps. In the presence of an absorbing boundary that changes the statistical ensemble over which  $p_n^\pm$  are defined,  $p_n^\pm$  are therefore dependent on this ensemble. They will therefore also depend on  $N$ , which determines the relative size of the ensemble, since  $p_n^\pm$  are computed as averages over all synapses. In the limit  $N \rightarrow \infty$ , the ensemble remains formally unchanged with absorption at the boundary, so in the large  $N$  limit, we would expect to see the  $p_n^\pm$  defined over a time-dependent, absorbing ensemble to approach the  $p_n^\pm$  defined over a time-independent, non-absorbing ensemble. For MFPTs, we would therefore expect the use of  $p_n^\pm$  defined over a non-absorbing ensemble to constitute a  $1/N$  approximation, becoming exact in the  $N \rightarrow \infty$  limit.

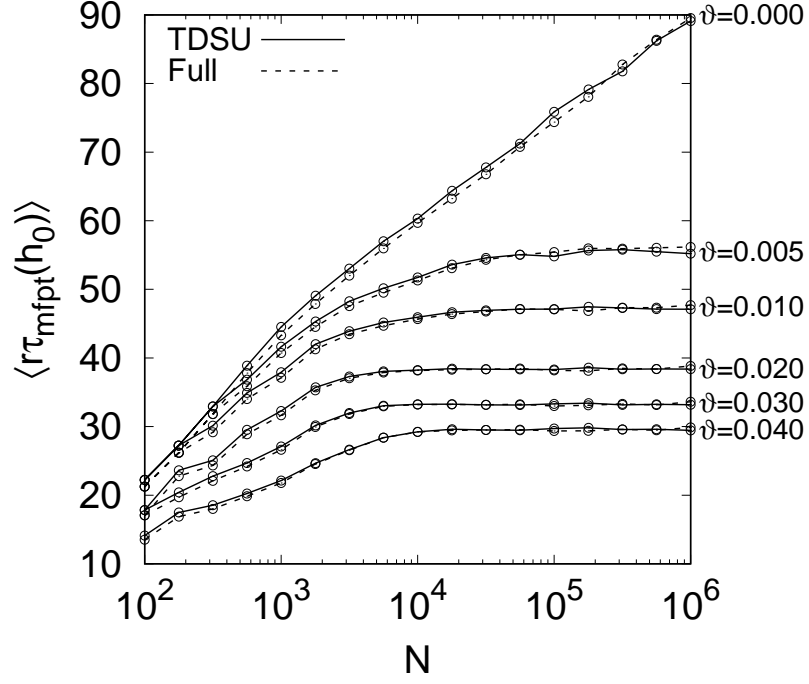
We explore these issues in Fig. 14, where we plot  $p_n^\pm$  for a time-dependent, absorbing ensemble against  $n$  for different values of  $N$ , and we also show for reference  $p_n^\pm$  for the time-independent, non-absorbing ensemble. For  $N =$



**Figure 14:** Conditional strength-change probabilities  $p_n^\pm$ , as a function of time step  $n$ , in the presence of an absorbing boundary. Results for four different values of  $N$  are shown, and for comparison results are also shown for dynamics in the absence of an absorbing boundary so that the ensemble of states over which  $p_n^\pm$  are determined does not change over time. Lines show results obtained from simulations while the solid points show analytical results for  $N = 3$  synapses. As  $N$  increases, the results for  $p_n^\pm$  converge on the solid lines. We have selected a filter size of  $\Theta = 3$ . The inset magnifies the region for  $5 \leq n + 1 \leq 20$ .

3, this value is small enough to permit an explicit analytical or numerical calculation of  $p_n^\pm$  for  $\Theta$  small enough, but in general we must determine  $p_n^\pm$  in the presence of an absorbing boundary (here  $\vartheta = 0$ ) via simulations. The basically exact agreement between the  $N = 3$  analytical or numerical results and those from simulation validates both methods. We see that the absorbing ensemble  $p_n^\pm$ 's do indeed differ from the non-absorbing ensemble  $p_n^\pm$ 's. We also see that the absorbing ensemble  $p_n^\pm$ 's do depend on  $N$ . The absorbing  $p_n^\pm$ 's approach the non-absorbing  $p_n^\pm$ 's as  $n$  increases, but the former asymptote to values that differ from the latter. Specifically, the absorbing  $p_n^-$ 's asymptote to larger values than the non-absorbing  $p_n^-$ 's, and conversely the absorbing  $p_n^+$ 's asymptote to smaller values than the non-absorbing  $p_n^+$ 's. For finite  $N$ , the absorbing  $p_n^\pm$ 's do not converge to each other as  $n$  increases, but they remain separated by an amount that depends on  $N$ . As  $N$  increases, the absorbing  $p_n^\pm$ 's approach the non-absorbing  $p_n^\pm$ 's more and more closely as  $n$  increases before they stabilise on what are essentially their different, asymptotic values. These dynamics explain why MFPTs determined using  $p_n^\pm$ 's from a non-absorbing ensemble differ slightly from those obtained from full, complex synapse simulations in the presence of an absorbing boundary, or equivalently, from an absorbing ensemble of simple synapses. In principle we could employ the absorbing ensemble definitions of  $p_n^\pm$  to obtain exact agreement for MFPTs, but: first, they depend on  $N$  (and on  $\vartheta$ ); second, we cannot obtain them simply except for very small  $N$  and  $\Theta$ . The use of the non-absorbing  $p_n^\pm$ 's, which are independent of  $N$  and are easy to compute, constitutes a  $1/N$  approximation for the determination of MFPTs, which imply the presence of an absorbing boundary. For calculations of the statistics of  $h$  in the absence of an absorbing boundary, the use of the non-absorbing  $p_n^\pm$ 's is, as argued, exact.

Finally, in Fig. 15 we consider how a non-zero threshold  $\vartheta$  affects  $\langle \tau_{\text{mfpt}}(h_0) \rangle$ .



**Figure 15:** Memory lifetimes  $\langle \tau_{\text{mfpt}}(h_0) \rangle$ , defined by mean first passage times for perceptron activation to fall below a firing threshold, as a function of the number of synapses,  $N$ . Results are shown for simulations of the time-dependent stochastic updater (TDSU) (solid lines; data points correspond to the circles) and for full, filter-based simulations without reduction to the TDSU (dashed lines; data points correspond to the circles). Results are shown for six different choices of firing threshold  $\vartheta$  as indicated. We have selected a filter size of  $\Theta = 4$ .

Previously, for a fixed probability simple stochastic updater synapse, we found that for *any* threshold  $\vartheta > 0$  (for balanced processes for which  $\mu(t) \rightarrow 0$  as  $t \rightarrow \infty$ ) the logarithmic growth of  $\langle \tau_{\text{mfpt}}(h_0) \rangle$  as a function of  $N$  is cut off, with  $\langle \tau_{\text{mfpt}}(h_0) \rangle$  asymptoting to a finite,  $\vartheta$ -dependent constant as  $N \rightarrow \infty$  (Elliott, 2014). We observe identical behaviour in Fig. 15 and indeed we find identical behaviour for other models based on synaptic dynamics with internal states. A threshold  $\vartheta > 0$  cuts off the logarithmic growth because the dynamics of  $h$  below  $\vartheta$  become inaccessible, as we have argued before (Elliott, 2014). This behaviour is independent of whether a synapse is simple or complex, or whether it is time-dependent or time-independent: once  $h$  falls below threshold, the exponential relaxation of  $h$  back to an average of zero that would otherwise result in the logarithmic growth of memory lifetimes with  $N$  ceases to be visible, in terms of the output of the perceptron.

Based on an OU limit for a simple stochastic updater synapse with update probability  $p$ , we showed that  $r\tau_{\text{mfpt}}(h_0) \sim p^{-1} \log_e(h_0/\vartheta)$  in the limit  $N \rightarrow \infty$  (Elliott, 2014). By imposing an accessibility criterion on the SNR, requiring that the memory signal is one standard deviation above  $\vartheta > 0$ , we also showed that  $r\tau_{\text{snr}}(h_0) \sim p^{-1} \log_e(h_0/\vartheta) - 1/(\vartheta p \sqrt{N})$ , which agrees with the asymptotic form for  $r\tau_{\text{mfpt}}(h_0)$  but also contains a large  $N$  correction going like  $1/\sqrt{N}$  (Elliott, 2014). For filter-based dynamics, we do not have a fixed  $p$ , but rather the step-dependent conditional probabilities  $p_n^\pm$ . However, we saw in section 5.2 that the filter-based dynamics become OU-like after an initial transient that differs from OU dynamics. This initial transient increases the mean memory signal  $\mu(t)$ . Viewing filter-based dynamics as OU-like beyond the peak in the mean memory signal, we can in fact obtain reasonable quantitative agreement between the asymptotic, large  $N$  values of  $\langle \tau_{\text{mfpt}}(h_0) \rangle$  in Fig. 15 and the OU-limit result  $r\tau_{\text{mfpt}}(h_0) \sim p^{-1} \log_e(h_0/\vartheta)$ . Previously we showed that the peak in

the mean memory signal occurs at  $rt_{\text{peak}} \approx 0.375 \Theta^2$  with amplitude  $\mu(t_{\text{peak}}) \approx 0.766/\Theta$  (Elliott & Lagogiannis, 2012). If we set  $h_0 = 0.766/\Theta$  for simplicity and take  $p = p_\infty = 1/\Theta^2$  beyond the memory signal peak, then we obtain an asymptotic, large  $N$  estimate for MFPTs, including the pre-peak period, of

$$r\tau_{\text{mfpt}}(h_0) \sim \Theta^2 \left( 0.375 + \log_e \frac{0.766}{\Theta \vartheta} \right). \quad (6.6)$$

Ignoring the inclusion of the pre-peak period, the requirement that  $\tau_{\text{mfpt}}(h_0) > 0$  is just the condition that the peak mean memory signal must at least exceed  $\vartheta$ . Plugging  $\Theta = 4$ , used in Fig. 15, into this expression, and using the various non-zero values of the threshold  $\vartheta$  shown in that figure, we obtain values, for increasing  $\vartheta$ , of  $r\tau_{\text{mfpt}} \approx 64, 53, 42, 36$  and  $31$ . These values, especially for larger values of  $\vartheta$ , are in remarkably good agreement with the large  $N$  values of  $\langle \tau_{\text{mfpt}}(h_0) \rangle$  that can be read off from Fig. 15. The agreement is all the more remarkable because of the rather crude estimates involved. These arguments show that we can, to a reasonable approximation, view filter-based dynamics as consisting of an initial, non-OU transient that takes the mean memory signal up to its peak value, following by OU-like dynamics that return the mean memory signal to its equilibrium value. Of course, the transition between these dynamics is not sharp, but viewing it in this way does allow an understanding of the dependence of MFPT memory lifetimes on the perceptron threshold  $\vartheta$ .

## 7 Discussion

Complex models of synaptic plasticity in which synapses possess internal states that control or regulate the expression of synaptic plasticity can in general

be quite difficult to analyse exactly. Conversely, simple stochastic updater synapses that lack internal states are relatively easy to study. By integrating out internal synaptic states and working purely in terms of transitions in synaptic strength, we have shown in earlier work that we can often derive exact results that would otherwise be quite difficult, if not impossible, to obtain by other means (Elliott, 2010b, 2016a). Furthermore, this change of perspective often affords far greater theoretical insight by stripping away the microscopic details and bringing into sharp focus the macroscopic dynamics.

Motivated by the desire to understand MFPT memory lifetimes in models of synaptic plasticity with complex synapses (Elliott, 2014), we have carried through this program of reducing complex, filter-based synapses to simple, stochastic updater synapses, at least for binary-strength synapses. Although we have focused on filter-based synaptic plasticity here, we can perform a similar reduction for any complex model of synaptic plasticity. The price paid for this reduction is that the conditional strength-change probabilities for the resulting stochastic updater synapse are time-dependent rather than time-independent and of course they depend on the history of induction signals and a synapse’s initial state. The conditional probabilities  $p_n^\pm$  must be known in order to use the methods developed here. However, even when analytical results for  $p_n^\pm$  are not available, we need simply construct once and only once an explicit table of these probabilities by extracting  $p_n^\pm$  from the matrix powers  $\mathbb{M}^n$ , which may be determined numerically if necessary. In return for paying the price of time- and history-dependent probabilities, we buy the ability to work directly in the transitions in the perceptron’s activation from memory storage step to memory storage step, ignoring entirely the internal states of the  $N$  synapses that contribute to the perceptron’s activation. This trade brings with it considerable computational, numerical and theoretical benefits.



Computationally, we need run simulations of only stochastic updater synapses using the probabilities  $p_n^\pm$ . When we run full simulations of synapses with internal states, averaging in order to obtain good statistics is achieved by running many (typically millions of) individual trials and then averaging results over all the trials. Each individual trial is a particular realisation of the stochastic dynamics. However, when we run simulations of the reduced, simple dynamics, although we still work with particular realisations of synaptic strengths and memories, some degree of averaging is already present in single trials by the use of the conditional probabilities  $p_n^\pm$  in determining whether synapses should change strength. Single trials are therefore hybrids, combining individual realisations with some element of ensemble averaging. The result is that typically we can run fewer simulations of the reduced dynamics compared to the full dynamics in order to obtain good statistical averaging, and this effect is comparatively more significant for complex synapses with more internal states (e.g. a larger value of  $\Theta$ ) because complex synapses with more internal states require more averaging than complex synapses with fewer internal states in order to obtain good statistics.

Numerically, we have seen that it is much easier to compute, for example, higher-order cumulants with the reduced dynamics compared to the full dynamics. We have exact results for the MGFs of the perceptron activation for the reduced dynamics, and these reproduce the moments (and therefore cumulants) of the full dynamics exactly for non-absorbing ensembles. Although the moments obtained from the MGF for the continuous-time process are Poisson sums that in general cannot be expressed in closed form, we may nevertheless evaluate them numerically to any desired accuracy because the contributions to the sums asymptote to zero. In contrast, for the full dynamics, in order to compute the  $m$ th-order moment we must work with a tensor product of  $m$

matrices, or flatten such products down to single, very large matrices. For non-defective transition matrices, the eigen-structure of the tensor products can be written down immediately, but for defective transitions matrices, this is not possible. In these cases, the reduced dynamics provide a numerically much simpler approach to computing higher-order statistics than working with very large matrices.

The greatest benefits of the reduction of complex synapses to simple synapses are, however, theoretical. First, by allowing us to work directly in the transitions in the perceptron's activation, we are able to derive PGFs and MGFs for the perceptron's activation, with these generating functions being exact. We are able in principle to compute the entire probability distribution for  $h(t)$ ,  $P(h, t)$ . Second, we can write down a master equation for the evolution of the probability distribution of the perceptron activation, either in discrete time or in continuous time. This provides an alternative way of determining the evolution of the moments, from a set of ordinary differential equations. The jump moments derived from the master equation then permit us to move to a Fokker-Planck equation. Because the underlying process is one of memory storage as discrete, punctate events, the continuous-time master and Fokker-Planck equations are expressed as Poisson sums that, unfortunately, do not factorise. However, third, in order to circumvent this non-factorisation, it is necessary to develop an approximation that replaces the discrete-time probabilities  $p_n^\pm$  with the continuous-time probabilities  $p^\pm(t)$ , with the latter not being Poisson sums over the former. While only an approximation, it nevertheless provides considerable theoretical insight into the dynamics of  $h(t)$ . For example, we found that these dynamics can be considered to be composed of an initial transient that leads to an increasing mean memory signal followed by an OU-like process that returns the memory signal to its equilibrium dis-

tribution. This separation of the dynamics into pre- and post-peak phases permitted some degree of quantitative understanding of MFPTs in the presence of a non-zero perceptron firing threshold. Critical to this was viewing the post-peak dynamics as those governed by stochastic updater synapses with a *fixed* strength-change probability  $p_\infty$ , allowing us to use our earlier results for MFPTs for such synapses (Elliott, 2014).

The reduction of complex synapses to simple synapses with time-dependent strength-change probabilities is necessarily exact for non-absorbing ensembles. However, for absorbing ensembles, such as those required for determining MFPTs, the reduction entails what is essentially a  $1/N$  approximation if we use the probabilities  $p_n^\pm$  extracted from a non-absorbing ensemble. For larger and larger  $N$ , we obtain MFPTs for the full and reduced dynamics that agree more and more closely. For absorbing dynamics, the conditional strength-change probabilities  $p_n^\pm$  become ensemble-dependent, but the dependence on the ensemble drops out in the formal limit,  $N \rightarrow \infty$ . Because of the difficulty in determining  $p_n^\pm$  for absorbing dynamics except for extremely small  $N$ , it appears very hard, analytically speaking, to examine precisely how the large  $N$  limit is achieved. Whether any progress can be made in understanding this limit is something that we shall investigate in future work.

An examination of MFPT-defined memory lifetimes with complex synapses reveals that network-size effects or optimality conditions are absent, in contrast to memory lifetimes defined by SNRs (Elliott, 2016a). Although we have only examined this issue with filter-based synapses here, this difference between MFPT and SNR memory lifetimes appears to be present in other models of complex synapses (unpublished observations). Previously, we also showed that for MFPT memory lifetimes for fixed probability stochastic updater synapses, a non-zero firing threshold (for balanced potentiation and depression dynam-

ics) leads to a catastrophic truncation of the dependence of memory lifetimes on the number of synapses,  $N$  (Elliott, 2014). Specifically,  $\tau_{\text{mfp}}t$  asymptotes to a  $\vartheta$ -dependent constant independent of  $N$  for *any* value of  $\vartheta > 0$  or, in general, any value  $\vartheta > \mu(\infty)$ . This truncation happens because the memory signal dynamics below threshold become inaccessible, while the growth of memory lifetimes with  $N$  for a firing threshold  $\vartheta = 0$  or  $\vartheta = \mu(\infty)$  depends critically on the accessibility of this signal (Elliott, 2014). Here, we observe identical behaviour for complex synapses, too. Again, this behaviour must occur quite generally in any complex model of synaptic plasticity for precisely the same reason that it occurs for the fixed probability stochastic updater synapses that we analysed in detail earlier (Elliott, 2014). This truncation of memory lifetimes appears to be disastrous for any such model of real, biological memory because real neurons in real memory systems cannot set their firing thresholds to coincide precisely with their asymptotic, mean membrane potential: the latter cannot be known *a priori* since it depends on the details of the input statistics to which a neuron cannot have prior access. It may be argued that asymptotically, the neuron could achieve such a matching. However, first, the lifetimes of memories stored while the matching of threshold to input statistics is still incomplete would be compromised. Second, the input statistics in real systems are almost certainly not stationary, so the matching can never be complete. Thus, it is unlikely that even an ongoing, dynamic adjustment of threshold to non-stationary input statistics could prevent the truncation of memory lifetimes.

We have performed this reduction of complex synapses to simple synapses for binary-strength, two-state synapses, but as we have indicated, this reduction can be achieved for any number of discrete states of synaptic strength. If there are  $\nu$  states of strength, then the full transition matrices are reduced in

general to  $\nu \times \nu$  matrices rather than  $2 \times 2$  matrices, and the two probabilities  $p_n^\pm$  would generalise to  $2(\nu-1)$  probabilities. The  $\mathbb{W}$  and  $\mathbb{D}$  matrices would generalise immediately, and we would be able to write down generating functions for the transition from the initial synaptic configuration, determined by  $\nu-1$  parameters, to any future synaptic configuration. Such generating functions would be expressed as products over  $\nu$  rather than 2 distinct factors raised to various powers, and each factor could be expressed in terms of a degree  $\nu-1$  polynomial. As a definite synaptic configuration uniquely determines  $h(t)$ , we would still be able to write down the exact distribution of perceptron activation for general, multistate synapses. However, the ability to work directly in the transitions in perceptron activation depends critically on the fact that for binary-strength synapses, any particular value of  $h = 2j/N - 1$  uniquely determines, up to an irrelevant permutation symmetry, the underlying configuration of (tilded) synaptic strengths:  $j$  synapses have (tilded) strength of  $+1$  and  $N-j$  have (tilded) strength of  $-1$ . This isomorphism between perceptron activation and the underlying configuration of (tilded) synaptic strengths is unique to binary synapses. For  $\nu = 3$ , ternary-strength synapses, for example, for strengths  $S_i \in \{-1, 0, +1\}$ , a pair of synapses may have a combined strength of 0 either by each having 0 strength or their having strengths of  $+1$  and  $-1$  in any order. The ordering is an irrelevant permutation symmetry, but for non-binary synapses, the mapping from perceptron activation to the underlying synaptic strength configuration is not unique (factoring out permutations). This non-permutation degeneracy in mapping definite values of  $h$  to synaptic strength configurations means that it is possible that for non-binary synapses, we cannot work directly in the transitions in  $h(t)$ . Nevertheless, if we have the exact distribution of  $h(t)$  available, it is possible that we may still be able to derive jump moments for  $h(t)$  and thus obtain approximations ul-

mately leading to a Fokker-Planck equation. It will be interesting to determine whether we can carry through this program leading to insightful, approximate dynamics for general, multistate synapses in future work.

## References

- Amit, D.J., & Fusi, S. 1994. Learning in neural networks with material synapses. *Neural Comput.*, **6**, 957–982.
- Bagal, A.A., Kao, J.P.Y., Tang, C.-M., & Thompson, S.M. 2005. Long-term potentiation of exogenous glutamate responses at single dendritic spines. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 14434–14439.
- Barrett, A.B., & van Rossum, M.C.W. 2008. Optimal learning rules for discrete synapses. *PLoS Comput. Biol.*, **4**, e1000230.
- Bartol, T.M., Bromer, C., Kinney, J., Chirillo, M.A., Bourne, J.N., Harris, K.M., & Sejnowski, T.J. 2015. Nanoconnectomic upper bound on the variability of synaptic plasticity. *eLife*, **4**, e10778.
- Cox, D.R. 1962. *Renewal Theory*. London: Methuen.
- Elliott, T. 2008. Temporal dynamics of rate-based plasticity rules in a stochastic model of spike-timing-dependent plasticity. *Neural Comput.*, **20**, 2253–2307.
- Elliott, T. 2010a. Discrete states of synaptic strength in a stochastic model of spike-timing-dependent plasticity. *Neural Comput.*, **22**, 244–272.
- Elliott, T. 2010b. A non-Markovian random walk underlies a stochastic model of spike-timing-dependent plasticity. *Neural Comput.*, **22**, 1180–1230.

- Elliott, T. 2011a. The mean time to express synaptic plasticity in stochastic, integrate-and-express models of synaptic plasticity induction. *Neural Comput.*, **23**, 124–159.
- Elliott, T. 2011b. Stability against fluctuations: Scaling, bifurcations and spontaneous symmetry breaking in stochastic models of synaptic plasticity. *Neural Comput.*, **23**, 674–734.
- Elliott, T. 2014. Memory nearly on a spring: A mean first passage time approach to memory lifetimes. *Neural Comput.*, **26**, 1873–1923.
- Elliott, T. 2016a. The enhanced rise and delayed fall of memory in a model of synaptic integration: Extension to discrete state synapses. *Neural Comput.*, **28**, 1927–1984.
- Elliott, T. 2016b. Variations on the theme of synaptic filtering: A comparison of integrate-and-express models of synaptic plasticity for memory lifetimes. *Neural Comput.*, **28**, 2393–2460.
- Elliott, T., & Lagogiannis, K. 2009. Taming fluctuations in a stochastic model of spike-timing-dependent plasticity. *Neural Comput.*, **21**, 3363–3407.
- Elliott, T., & Lagogiannis, K. 2012. The rise and fall of memory in a model of synaptic integration. *Neural Comput.*, **24**, 2604–2654.
- Fusi, S., Drew, P.J., & Abbott, L.F. 2005. Cascade models of synaptically stored memories. *Neuron*, **45**, 599–611.
- Hopfield, J.J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 2554–2558.

- Huang, Y., & Amit, Y. 2010. Precise capacity analysis in binary networks with multiple coding level inputs. *Neural Comput.*, **22**, 660–688.
- Huang, Y., & Amit, Y. 2011. Capacity analysis in multi-state synaptic models: A retrieval probability perspective. *J. Comput. Neurosci.*, **30**, 699–720.
- Lahiri, S., & Ganguli, S. 2013. A memory frontier for complex synapses. *Pages 1034–1042 of: Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., & Weinberger, K.Q. (eds), Advances in Neural Information Processing Systems 26*. Cambridge, MA: MIT Press.
- Leibold, C., & Kempter, R. 2006. Memory capacity for sequences in a recurrent network with biological constraints. *Neural Comput.*, **18**, 904–941.
- Leibold, C., & Kempter, R. 2008. Sparseness constrains the prolongation of memory lifetime via synaptic metaplasticity. *Cerebral Cortex*, **18**, 67–77.
- Montgomery, J.M., & Madison, D.V. 2002. State-dependent heterogeneity in synaptic depression between pyramidal cell pairs. *Neuron*, **33**, 765–777.
- Montgomery, J.M., & Madison, D.V. 2004. Discrete synaptic states define a major mechanism of synapse plasticity. *Trends Neurosci.*, **27**, 744–750.
- Nadal, J.P., Toulouse, G., Changeux, J.P., & Dehaene, S. 1986. Networks of formal neurons and memory palimpsests. *Europhys. Lett.*, **1**, 535–542.
- O’Connor, D.H., Wittenberg, G.M., & Wang, S.S.-H. 2005a. Dissection of bidirectional synaptic plasticity into saturable unidirectional process. *J. Neurophysiol.*, **94**, 1565–1573.
- O’Connor, D.H., Wittenberg, G.M., & Wang, S.S.-H. 2005b. Graded bidirectional synaptic plasticity is composed of switch-like unitary events. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 9679–9684.



- Parisi, G. 1986. A memory which forgets. *J. Phys. A: Math. Gen.*, **19**, L617–L620.
- Petersen, C.C.H., Malenka, R.C., Nicoll, R.A., & Hopfield, J.J. 1998. All-or-none potentiation at CA3-CA1 synapses. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 4732–4737.
- Rubin, D.D.B.D., & Fusi, S. 2007. Long memory lifetimes require complex synapses and limited sparseness. *Front. Comput. Neurosci.*, **1**, 7.
- Sobczyk, A., & Svoboda, K. 2007. Activity-dependent plasticity of the NMDA-receptor fractional  $\text{Ca}^{2+}$  current. *Neuron*, **53**, 17–24.
- Tsodyks, M.V. 1990. Associative memory in neural networks with binary synapses. *Mod. Phys. Lett. B*, **4**, 713–716.
- Yasuda, R., Sabatini, B.L., & Svoboda, K. 2003. Plasticity of calcium channels in dendritic spines. *Nature Neurosci.*, **6**, 948–955.