First Passage Time Memory Lifetimes for Multistate, Filter-Based Synapses

Terry Elliott[1]

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

Running Title: FPT memory lifetimes.

January 25, 2020.

[1] Tel: +44 (0)23 8059 6000, Fax: +44 (0)23 8059 2783, Email: te@ecs.soton.ac.uk.

Abstract

Models of associative memory with discrete state synapses learn new memories by forgetting old ones. In contrast to non-integrative models of synaptic plasticity, models with integrative, filter-based synapses exhibit an initial rise in the fidelity of recall of stored memories. This rise to a peak is driven by a transient process, and is then followed by a return to equilibrium. In a series of papers, we have employed a first passage time (FPT) approach to define and study memory lifetimes, incrementally developing our methods, from both simple and complex binary-strength synapses, to simple multistate synapses. Here, we complete this work by analysing FPT memory lifetimes in multistate, filter-based synapses. To achieve this, we integrate out the internal filter states so that we can work with transitions only in synaptic strength. We then generalise results on polysynaptic generating functions from binary-strength to multistate synapses, allowing us to examine the dynamics of synaptic strength changes in an ensemble of synapses rather than just a single synapse. To derive analytical results for FPT memory lifetimes, we partition the synaptic dynamics into two distinct phases, the first, pre-peak phase studied with a drift-only approximation, the second, post-peak phase studied with approximations to the full strength transition probabilities. These approximations capture the underlying dynamics very well, as demonstrated by the extremely good agreement between results obtained by simulating our model and results obtained from the Fokker-Planck or integral equation approaches to FPT processes.

1 Introduction

Associative memory systems with bounded synaptic strengths store new memories by forgetting old ones (Nadal *et al.*, 1986; Parisi, 1986), overcoming the catastrophic forgetting of the Hopfield model (Hopfield, 1982). Studying memory systems with discrete strength synapses has therefore become popular (see, e.g., Tsodyks, 1990; Amit & Fusi, 1994; Leibold & Kempter, 2006; Barrett & van Rossum, 2008; Elliott & Lagogiannis, 2012). Such systems must be able to store new memories while ensuring that already-stored memories have long lifetimes. In most discrete synapse models, the fidelity of recall of a memory falls monotonically over time. The emphasis in these models has therefore been in prolonging this decay process, typically via appealing to sparse coding (Tsodyks & Feigel'man, 1988; Rubin & Fusi, 2007; Leibold & Kempter, 2008) or synaptic metaplasticity mechanisms (Fusi *et al.*, 2005). But the memory trace continues to fall monotonically, and these approaches essentially serve merely to dilate time rather than address the underlying problem.

The underlying problem is that the storage of new memories induces activity-driven fluctuations in the synaptic strengths that encode old memories. We have proposed "integrate-and-express" models of synaptic plasticity to control fluctuations in synaptic strength in a developmental context, to stabilise patterns of synaptic connectivity such as ocular dominance columns in primary visual cortex (Elliott, 2008; Elliott & Lagogiannis, 2009). In these models, synaptic plasticity induction signals are integrated by a synaptic low pass filter, attenuating high-frequency noise and passing low-frequency signals. When the filter reaches threshold, the synapse expresses synaptic plasticity, changing its strength. A synaptic filtering mechanism could be implemented in the macromolecular machinery present at single synapses (Elliott, 2011). For example, the phosphorylation state of the collective CaMKII-PP1 switch (Ferrell, 1996; Lisman & Zhabotinsky, 2001; Pi & Lisman, 2008) may encode a discrete

filter state, as we have previously proposed (Elliott & Lagogiannis, 2009; Elliott, 2011). The calcium transients induced by individual plasticity induction events would lead to the phosphorylation of the CaMKII holoenzyme and embody the sub-threshold integration phase, while its autocatalytic threshold would implement the filter threshold process and lead to the expression of a change in synaptic strength. When such integrative mechanisms are applied to associative memory tasks, we found that the memory trace actually rises, driven by the storage of later memories, before it finally reaches a peak and then decays away (Elliott & Lagogiannis, 2012). Such integrative, filter-based models of synaptic plasticity outperform their non-integrative competitors, in terms of memory longevity, in almost all biologically relevant regions of parameter space (Elliott, 2016b).

To study memory lifetimes, we must have a definition of memory lifetimes. Memory lifetimes can be defined by a variety of different methods, including the single-to-noise ratio (SNR) (Tsodyks, 1990), the equivalent "ideal observer" approach (Fusi *et al.*, 2005; Lahiri & Ganguli, 2013; Elliott, 2016b), via signal detection theory (Leibold & Kempter, 2006, 2008), and using retrieval probabilities (Huang & Amit, 2010, 2011). Recently, in a feed-forward setting, we defined memory lifetimes by the mean first passage time (MFPT) for a perceptron's activation to fall below firing threshold (Elliott, 2014). This first passage time (FPT) approach is theoretically sounder and biologically more relevant than the SNR approach, and reveals that the SNR method can often give a very poor and misleading indication of memory lifetimes (Elliott, 2014, 2017a,b, 2019).

Understanding FPT memory lifetimes in systems with filter-based synapses is analytically very difficult. In a series of papers, we have gradually developed the machinery to do this (Elliott, 2014, 2017a,b, 2019). In the present paper, we complete this program by studying FPT memory lifetimes in the presence

of synapses with an arbitrary number of discrete states of strength. First, we provide in section 2 a recapitulation of our general approach to memory storage with filter-based synapses. In section 3 we integrate out the internal filter states of synapses, so that we can work with the probabilities that synapses change strength without the added complexity of their internal dynamics. Then, in section 4, we construct the polysynaptic generating functions using these probabilities that describe how a perceptron's entire set of synapses evolves over time. We then apply these results in section 5 to the calculation of FPT statistics, developing approximation methods to simplify and render tractable the calculations. In section 6 we compare these analytical results to simulations. Finally, in section 7, we briefly discussion our methods.

2 General Formulation and Setup

We first outline our approach to using a perceptron in continuous time for memory storage, and then we discuss our filter-based model of synaptic plasticity that allows the perceptron's synapses to change strength during memory storage.

2.1 Continuous-Time Perceptron Approach

We consider memory storage in a feed-forward setting with a single perceptron operating in continuous time. The perceptron has N synapses, indexed by letters such as i and j, where $i, j = 1, \ldots, N$, with strengths $S_i(t)$, where $t \geq 0$ s denotes time. Each synapse's strength is discrete, at any given time taking one of n distinct values from the ordered set $\{\Omega_1, \ldots, \Omega_n\}$, with

$$\Omega_a = -1 + 2\,\frac{a-1}{n-1}, \tag{2.1}$$

where we use letters such as a and b, where $a, b = 1, \ldots, n$, to index these discrete strength states, so that the possible strengths are uniformly spaced in the interval $[-1, +1]$. With inputs x_i through these N synapses, the perceptron has activation given as standard by

$$h_{\boldsymbol{x}}(t) = \frac{1}{N} \sum_{i=1}^{N} x_i \, S_i(t), \tag{2.2}$$

where $\boldsymbol{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$, T denoting the transpose. Here we are concerned only with whether or not the perceptron's activation exceeds its firing threshold, ϑ: any non-linearity in converting the activation into a firing rate is not relevant in a purely feed-forward analysis. Just as $h_{\boldsymbol{x}}(t)$ is scaled by $1/N$, we could have allowed the range of the strengths Ω_a to depend on n, rather than being fixed in the interval $[-1, +1]$, and then scaled $h_{\boldsymbol{x}}(t)$ accordingly. The choice of the range and scale of the strengths Ω_a and of the perceptron's activation $h_{\boldsymbol{x}}(t)$ is purely a matter of convention, because we merely have to change the definition of the threshold ϑ to compensate. In particular, although the range of the strengths Ω_a extends from negative to positive values, we should not think of these values as corresponding to a transition from inhibitory to excitatory synapses, but merely an arrangement from weaker to stronger synapses.

The perceptron stores a sequence of "memories" $\boldsymbol{\xi}^{\alpha}$, with letters such as α and β, where $\alpha, \beta = 0, 1, 2, \ldots$, indexing these memories. By suitable choices of conventions, it is convenient to restrict the components ξ_i^{α} to the values ± 1. We set the probabilities of these values to $\mathsf{Prob}[\xi_i^{\alpha} = \pm 1] = g_{\pm}$, with $g_+ + g_- = 1$, restricting to the symmetric, unbiased case of $g_{\pm} = \frac{1}{2}$. The components ξ_i^{α} are taken for simplicity to be independent both between synapses and across memories. Memory $\boldsymbol{\xi}^0$ is always stored at time $t = 0$ s. The later memories $\boldsymbol{\xi}^{\alpha}$ ($\alpha > 0$) are stored at times $t > 0$ s governed by a Poisson process of rate r. Here

we may without loss of generality set $r = 1$ Hz, or just replace the dimensionful t with the dimensionless rt. Strictly we should distinguish between the time of the storage of a memory and the time immediately thereafter (e.g. $t = 0$ s and $t = 0^+$ s for $\boldsymbol{\xi}^0$), but the distinction is not required provided that we are consistent.

All memories are stored by (possible) changes in the strengths of the perceptron's N synapses. The storage of the later memories $\boldsymbol{\xi}^\alpha$ ($\alpha > 0$) will in general affect the fidelity of recall of memory $\boldsymbol{\xi}^0$. We refer to memory $\boldsymbol{\xi}^0$ as the tracked memory and to the later memories $\boldsymbol{\xi}^\alpha$ ($\alpha > 0$) as the non-tracked memories. The perceptron's activation in response to the re-presentation (but not the re-storage) of $\boldsymbol{\xi}^0$ is

$$h(t) \equiv h_{\boldsymbol{\xi}^0}(t) = \frac{1}{N}\sum_{i=1}^{N} \xi_i^0 \, S_i(t) \equiv \frac{1}{N}\sum_{i=1}^{N} \widetilde{S}_i(t), \tag{2.3}$$

where we define the tilded synaptic strengths by $\widetilde{S}_i(t) = \xi_i^0 \, S_i(t)$, and we frequently write $h(t)$ for $h_{\boldsymbol{\xi}^0}(t)$ for simplicity. Consider a particular realisation of the tracked memory $\boldsymbol{\xi}^0$, of the non-tracked memories $\boldsymbol{\xi}^\alpha$ ($\alpha > 0$) and their Poisson storage times, and of the underlying synaptic states and dynamics leading to the storage of these memories. If $h_{\boldsymbol{\xi}^0}(0) \leq \vartheta$, then $\boldsymbol{\xi}^0$ is deemed not to be successfully stored because its immediate re-presentation (but not re-storage) does not induce the perceptron to fire. If $h_{\boldsymbol{\xi}^0}(0) > \vartheta$, then $\boldsymbol{\xi}^0$ is successfully stored. With the subsequent storage of non-tracked memories, there will then be some later time $\tau_{\rm fpt}(\boldsymbol{\xi}^0)$ for which upon the re-presentation of $\boldsymbol{\xi}^0$, the perceptron's activation first falls to or below firing threshold, so that $h_{\boldsymbol{\xi}^0}(\tau_{\rm fpt}(\boldsymbol{\xi}^0)) \leq \vartheta$ but $h_{\boldsymbol{\xi}^0}(t) > \vartheta$ for $t < \tau_{\rm fpt}(\boldsymbol{\xi}^0)$. This FPT provides one definition of the lifetime of $\boldsymbol{\xi}^0$ for any particular realisation; if $h_{\boldsymbol{\xi}^0}(0) \leq \vartheta$, then we define $\tau_{\rm fpt}(\boldsymbol{\xi}^0) \equiv 0$ s. Averaging over all realisations of the non-tracked memories defines the MFPT $\tau_{\rm mfpt}(\boldsymbol{\xi}^0)$ conditioned on the particular tracked

memory $\boldsymbol{\xi}^0$ and any associated initial synaptic configuration. Then averaging further over $\boldsymbol{\xi}^0$ and any associated initial synaptic configuration defines the unconditional MFPT, $\tau_{\text{mfpt}} = \langle \tau_{\text{mfpt}}(\boldsymbol{\xi}^0) \rangle_{\boldsymbol{\xi}^0}$. Writing $h^{(0)} = h_{\boldsymbol{\xi}^0}(0)$, we may express the MFPT in the alternative form $\tau_{\text{mfpt}} = \langle \tau_{\text{mfpt}}(h^{(0)}) \rangle_{h^{(0)} > \vartheta}$, which has the advantage that the distribution of $h^{(0)}$ depends explicitly on the joint distribution of the initial synaptic configuration and the tracked memory $\boldsymbol{\xi}^0$.

2.2 Filter-Based Synaptic Plasticity

In a purely feed-forward setting in which a perceptron should fire in response to the re-presentation of a memory, the component ξ_i^α is the plasticity induction signal to synapse i upon the storage of memory α. For $\xi_i^\alpha = +1$ (respectively, -1), the synapse should potentiate (respectively, depress) in order to participate in the perceptron's storage of memory α. In integrate-and-express, filter-based models of synaptic plasticity, these induction signals do not necessarily lead to the immediate expression of synaptic plasticity, but rather they are integrated by single synapses before plasticity is expressed. Specifically, we propose that a single synapse increments (respectively, decrements) an internal filter state in response to a potentiating (respectively, depressing) induction signal, with potentiation (respectively, depression) being expressed by the synapse (if possible) when the synapse's filter state reaches an upper (respectively, lower) threshold. We label these filter states by letters such as I and J, and because we will consider only equiprobable potentiating and depressing induction signals ($g_\pm = \frac{1}{2}$), we need only consider symmetric filters with upper and lower thresholds of $+\Theta$ and $-\Theta$, respectively. Although this assumption of balanced potentiation and depression may appear implausible, we have previously argued (Elliott, 2016a) that in order to prevent saturation and to maintain a full dynamic range, synaptic plasticity likely operates in regimes governed by fixed-point dynamics (see, e.g., Bienenstock *et al.*, 1982; Burkitt *et al.*, 2004; Appleby

& Elliott, 2006): in such regimes, potentiation and depression would be automatically adjusted to be balanced, on average. With a symmetric filter, we have the possible values $I, J \in \{-(\Theta-1), \ldots, +(\Theta-1)\}$ for filter states, with the filter state being immediately reset to zero upon threshold being reached. Such resetting dynamics are the most natural for general multistate, $n \geq 2$ synapses, while other dynamics are possible for specifically binary-strength (or "bistate"), $n = 2$ synapses (Elliott, 2016b). These transitions in filter state in response to plasticity induction signals can be summarised as

$$\begin{aligned} \xi_i^\alpha = +1 \Rightarrow & \begin{cases} I \mapsto I+1 & \text{for } I < +(\Theta-1) \\ I \mapsto 0 \;\&\; \Uparrow & \text{for } I = +(\Theta-1) \end{cases} \\ & \qquad , \\ \xi_i^\alpha = -1 \Rightarrow & \begin{cases} I \mapsto I-1 & \text{for } I > -(\Theta-1) \\ I \mapsto 0 \;\&\; \Downarrow & \text{for } I = -(\Theta-1) \end{cases} \end{aligned} \tag{2.4}$$

where the symbols "$\Uparrow$" and "$\Downarrow$" indicate the expression of potentiation and depression, respectively, i.e. $S_i \to \max\{+1, S_i + \frac{2}{n-1}\}$ and $S_i \to \min\{-1, S_i - \frac{2}{n-1}\}$, respectively, where the maximum and minimum functions ensure that a synapse that is saturated at its upper or lower strength cannot potentiate or depress further. Fig. 1 graphically illustrates these transitions.

We represent the filter state of a synapse by a $(2\,\Theta-1)$-dimensional vector, where we use the filter state itself to index the vector's components, so that they run from $-(\Theta-1)$ to $+(\Theta-1)$ rather than the more conventional 1 to $(2\,\Theta-1)$. Let the $(2\,\Theta-1) \times (2\,\Theta-1)$ matrices $\mathbb{F}^\pm$ implement a change in a synapse's filter state in response to a potentiating or depressing induction signal without representing the resetting to zero transitions at threshold, and let the two matrices $\mathbb{T}^\pm$ represent only these resetting to zero transitions at

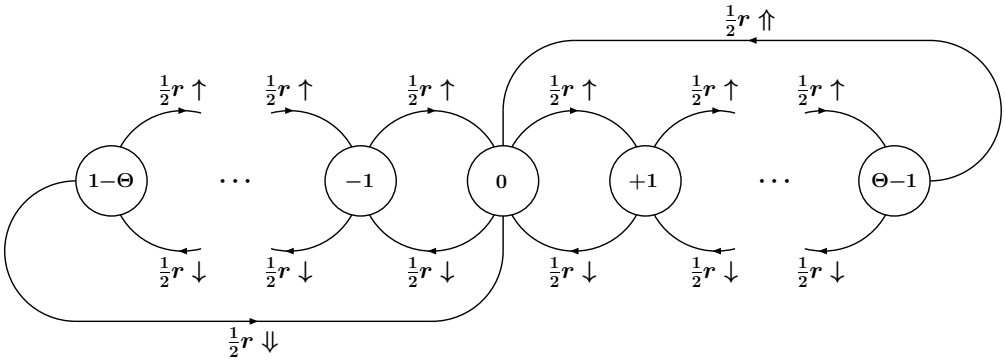


Figure 1: Transitions between filter states leading to the expression of synaptic plasticity at filter thresholds. Filter states are represented by the circled numbers, while the transitions between them are indicated by the directed curves labelled by their rates and associated potentiation (arrows $\uparrow$ and $\Uparrow$) and depression (arrows $\downarrow$ and $\Downarrow$) signals. A potentiation (respectively, depression) signal in state $+(\Theta-1)$ (respectively, $-(\Theta-1)$) leads to an increment (respectively, decrement) in synaptic strength where possible, indicated by the arrow $\Uparrow$ (respectively, $\Downarrow$).

upper or lower threshold. The matrices $\mathbb{F}^{\pm}$ are simply the step operators,

$$\mathbb{F}^{+} = \mathrm{diag}_{\mathrm{l}}\{\underbrace{1,\ldots,1}_{2\Theta-2}\} \text{ and } \mathbb{F}^{-} = \mathrm{diag}_{\mathrm{u}}\{\underbrace{1,\ldots,1}_{2\Theta-2}\}, \tag{2.5}$$

where $\mathrm{diag}_{\mathrm{l}}$ and $\mathrm{diag}_{\mathrm{u}}$ mean lower and upper diagonal matrices, respectively, and the non-zero components of the threshold matrices $\mathbb{T}^{\pm}$ are

$$\left[\mathbb{T}^{-}\right]_{0,-(\Theta-1)} = 1 \text{ and } \left[\mathbb{T}^{+}\right]_{0,+(\Theta-1)} = 1, \tag{2.6}$$

where again we index elements of such matrices using the filter states, so the above defines an element on the middle rows of these two matrices and in their first and last columns, respectively. With n possible states of synaptic strength, the joint strength and filter state of a synapse is represented by a $(2\,\Theta-1)n$-dimensional vector, where we order the blocks of components so

that the a^{th} block of $(2\,\Theta - 1)$ components represents the filter state when the synapse has strength Ω_a. Changes in joint strength and filter state are then represented by the two $(2\,\Theta - 1)n \times (2\,\Theta - 1)n$ matrices $\mathbb{M}^{\pm}$ for potentiating and depressing induction signals, respectively. For the particular case of $n = 4$, for example, these matrices take the block forms

$$\mathbb{M}^{+} = \left(\begin{array}{c|c|c|c} \mathbb{F}^{+} & & & \\ \hline \mathbb{T}^{+} & \mathbb{F}^{+} & & \\ \hline & \mathbb{T}^{+} & \mathbb{F}^{+} & \\ \hline & & \mathbb{T}^{+} & \mathbb{F}^{+}+\mathbb{T}^{+} \end{array}\right), \tag{2.7a}$$

and

$$\mathbb{M}^{-} = \left(\begin{array}{c|c|c|c} \mathbb{F}^{-}+\mathbb{T}^{-} & \mathbb{T}^{-} & & \\ \hline & \mathbb{F}^{-} & \mathbb{T}^{-} & \\ \hline & & \mathbb{F}^{-} & \mathbb{T}^{-} \\ \hline & & & \mathbb{F}^{-} \end{array}\right), \tag{2.7b}$$

with all other blocks being zero. The submatrices $\mathbb{F}^{\pm}$ on the diagonals of $\mathbb{M}^{\pm}$ represent changes in the filter state of a synapse without a change in strength state. The submatrices $\mathbb{T}^{\pm}$ on the lower and upper diagonals represent threshold events in which the synapse changes strength with its filter state being reset to zero, while their appearance on the diagonals represent threshold events without a corresponding change in strength because the synapse is saturated at its upper or lower strength.

The two matrices $\mathbb{M}^{\pm}$ describe transitions in the joint strength and filter state of a single synapse in response to a particular induction signal. As we must average over the non-tracked memories to determine MFPTs, we must consider the superposed transition matrix

$$\mathbb{M} = g_{+}\mathbb{M}^{+} + g_{-}\mathbb{M}^{-} = \tfrac{1}{2}\left(\mathbb{M}^{+} + \mathbb{M}^{-}\right), \tag{2.8}$$

describing the change in synaptic state in response to the probabilistic oc-

currence of a potentiating or a depressing induction signal, with probabilities $g_\pm = \frac{1}{2}$. Even with a definite initial synaptic state, the probabilistic treatment of the later memories, including their Poisson times, means that the joint strength and filter state of a synapse is a stochastic quantity, describing the joint probability distribution of a synapse's strength and filter state. With ongoing non-tracked memory storage, a synapse's joint strength and filter state will in general approach the (suitably normalised) equilibrium distribution governed by the maximal eigenvector of the stochastic matrix $\mathbb{M}$. We have computed this unit eigenvector elsewhere (Elliott, 2016a). Defining the filter distribution $\boldsymbol{B}$ with components

$$B_I = \frac{1}{\Theta^2}\left(\Theta - |I|\right), \tag{2.9}$$

we may confirm that the $(2\,\Theta - 1)n$-dimensional vector $\boldsymbol{A}$ taking the block form

$$\boldsymbol{A}^{\mathrm{T}} = \frac{1}{n}\big(\underbrace{\boldsymbol{B}^{\mathrm{T}}\big|\cdots\big|\boldsymbol{B}^{\mathrm{T}}}_{n}\big), \tag{2.10}$$

in which the distribution $\boldsymbol{B}$ occurs once for each strength state, corresponds to the equilibrium state. Therefore, in equilibrium, filter states are distributed according to $\boldsymbol{B}$ regardless of strength state, and strength states are uniformly distributed with probability $1/n$ regardless of filter state.

The tracked memory $\boldsymbol{\xi}^0$ is taken to be stored against the background of this equilibrium distribution. For $\xi_i^0 = \pm 1$, the state of synapse i immediately after the storage of the tracked memory is $\mathbb{M}^{\pm}\boldsymbol{A}$. To average over all possible tracked memories to determine τ_{mfpt}, we must treat each synapse as being in an equiprobable mixture (and not superposition) of the states $\mathbb{M}^{\pm}\boldsymbol{A}$. However, when potentiation and depression processes are treated in a completely symmetrical, unbiased manner, we have previously shown that the two distributions $\mathbb{M}^{\pm}\boldsymbol{A}$ are, as vectors, mirror images of each other: $\mathbb{M}^{-}\boldsymbol{A}$ read

from bottom to top is identical to $\mathbb{M}^{+}\boldsymbol{A}$ read from top to bottom (Elliott & Lagogiannis, 2012; Elliott, 2016a). Because the superposed matrix $\mathbb{M}$ also treats potentiation and depression symmetrically when $g_{\pm} = \frac{1}{2}$, the two distributions $\mathbb{M}^{\alpha}\,\mathbb{M}^{\pm}\boldsymbol{A}$, corresponding to the two possible states of a synapse after $\alpha \geq 0$ non-tracked memories have been stored, are also mirror images. So too therefore are the two distributions $e^{rt(\mathbb{M}-\mathbb{I})}\mathbb{M}^{\pm}\boldsymbol{A}$, corresponding to the two possible states of a synapse after a time $t \geq 0$ s has elapsed ($\mathbb{I}$ is the identity matrix). While the two conditional probabilities $\mathsf{Prob}[S_i(t) = \Omega_a \,|\, \xi_i^0 = \pm 1]$ differ, the two conditional probabilities $\mathsf{Prob}[\widetilde{S}_i(t) = \Omega_a \,|\, \xi_i^0 = \pm 1]$ are therefore identical when the Ω_a are symmetically arranged around their mean, as here, because the inclusion of the initial induction signal in the tilded strength exactly reverses this mirror image structure. Essentially, for $\xi_i^0 = -1$, positive and negative filter indices and positive and negative strength values are swapped around, compared to $\xi_i^0 = +1$, in terms of their contribution to the perceptron's activation. Since $h(t) = \frac{1}{N}\sum_{i=1}^{N}\widetilde{S}_i(t)$, $h(t)$ is therefore a sum over N identically distributed random variables, and in terms of dynamics of $h(t)$, there is no need to distinguish between the two states $\mathbb{M}^{\pm}\boldsymbol{A}$ in the mixture of any synapse's initial distribution after the storage of the tracked memory. We may in effect simply set $\xi_i^0 = +1$ for all synapses, without any loss of generality, which significantly simplifies calculations.

3 Integrating Out Filter Dynamics

In principle the computation of MFPT memory lifetimes is completely routine, from a purely algorithmic point of view. However, in practice their computation is intractable for all but the simplest of problems with very small N. Part of the difficulty is that while the internal filter dynamics determine when changes in synaptic strength occur, only these changes in strength are relevant to changes in the perceptron's activation; the processes by which the strength

changes actually occur are not relevant. To simplify the computation of MFPTs, we therefore now integrate out the internal filter dynamics, effectively reducing a complex synapse to a simple synapse. Previously we did this for specifically bistate synapses (Elliott, 2017b), but we now extend to the general multistate case.

3.1 Time-Dependent Stochastic Updater Formulation

The simplest model of synaptic plasticity in the context of examining memory lifetimes is a simple synapse without internal states that changes strength in response to plasticity induction signals with some fixed probability (Tsodyks, 1990), a model that we have called a "stochastic updater" (SU) (Elliott & Lagogiannis, 2012). When we reduce a complex synapse with internal states to a simple synapse without them, the probability that a synapse changes strength becomes memory storage step- or time-dependent, so we refer to such a synapse as a time-dependent stochastic updater (TDSU) (Elliott, 2017b).

We wish to integrate out the internal filter states of a filter-based synapse so that instead of working with $(2\,\Theta - 1)n \times (2\,\Theta - 1)n$ transition matrices describing transitions in the joint strength and filter state of a synapse, we work with only $n \times n$ transition matrices describing changes in only the synapse's strength state. Let the vector $\boldsymbol{\Sigma}^\alpha$ represent the probability distribution of a synapse's joint strength and filter state upon the storage of memory α. Clearly $\boldsymbol{\Sigma}^\alpha = \mathbb{M}^\alpha\,\boldsymbol{\Sigma}^0$, where $\boldsymbol{\Sigma}^0 = \mathbb{M}^\pm \boldsymbol{A}$ depending on the sign of $\xi_i^0 = \pm 1$. We write $\boldsymbol{\Sigma}^\alpha$ in the block form

$$\boldsymbol{\Sigma}^{\alpha\,\mathrm{T}} = \left(\boldsymbol{\Sigma}_1^{\alpha\,\mathrm{T}} \,|\, \cdots \,|\, \boldsymbol{\Sigma}_n^{\alpha\,\mathrm{T}}\right), \tag{3.1}$$

where the $(2\,\Theta - 1)$-dimensional vectors $\boldsymbol{\Sigma}_a^\alpha$ correspond to the various distributions of a synapse's filter states when in the various strength states a. Defining

the vector $\mathbf{1}$ to be a vector all of whose components are unity,[2] the probability that the synapse is in strength state a is just $\mathbf{1} \cdot \mathbf{\Sigma}_a^\alpha$, where a "$\cdot$" denotes the dot product. We write $\mathbf{\Pi}^\alpha$ as the vector describing the probability distribution of a synapse's strength states, where its components are just $\Pi_a^\alpha = \mathbf{1} \cdot \mathbf{\Sigma}_a^\alpha$. In a filter-based framework, a synapse can only change strength in response to an induction signal if its filter is in the state $+(\Theta - 1)$ for a potentiating induction signal, or $-(\Theta - 1)$ for a depressing induction signal. Let $f_a^{(\alpha)\pm}$ denote the probabilities that a synapse in strength state a at memory storage step α is in the filter states $\pm(\Theta - 1)$.[3] Then

$$f_a^{(\alpha)\pm} = \mathbf{1} \cdot \mathbb{T}^\pm \mathbf{\Sigma}_a^\alpha. \tag{3.2}$$

We may then write down the system of equations

$$\Pi_1^{\alpha+1} - \Pi_1^\alpha = g_- f_2^{(\alpha)-} - g_+ f_1^{(\alpha)+}, \tag{3.3a}$$

$$\Pi_a^{\alpha+1} - \Pi_a^\alpha = g_- f_{a+1}^{(\alpha)-} + g_+ f_{a-1}^{(\alpha)+} - g_- f_a^{(\alpha)-} - g_+ f_a^{(\alpha)+}, \tag{3.3b}$$

$$\Pi_n^{\alpha+1} - \Pi_n^\alpha = g_+ f_{n-1}^{(\alpha)+} - g_- f_n^{(\alpha)-}, \tag{3.3c}$$

for $a = 2, \ldots, n-1$ to describe the change in the probability distribution of a synapse's strength states from step α to step $\alpha + 1$. We have not specified $g_\pm = \frac{1}{2}$ in these equations for generality and clarity. To express Eq. (3.3) in the form of a discrete time Master equation involving the strength vectors $\mathbf{\Pi}^\alpha$

---

[2] We do not in general need to specify the size of the vector $\mathbf{1}$ as it will always be clear from the context. This is also true of the identity matrix $\mathbb{I}$.

[3] We surround α with parentheses in a symbol such as $f_a^{(\alpha)\pm}$ since α is merely an index or label and does not indicate a power. We adopt this convention whenever there is risk of such ambiguity.

and step-dependent strength transition matrices, we define the quantities

$$p_a^{(\alpha)\pm} = \frac{f_a^{(\alpha)\pm}}{\Pi_a^\alpha}, \tag{3.4}$$

which are nothing but the *conditional* strength-change probabilities, conditioned on the synapse being in a particular strength state. If $\Pi_a^\alpha = 0$, then we define $p_a^{(\alpha)\pm} = 0$. Eq. (3.4) also defines the two probabilities $p_n^{(\alpha)+}$ and $p_1^{(\alpha)-}$, but they are not relevant because of saturation. We may then write $f_a^{(\alpha)\pm} = p_a^{(\alpha)\pm}\,\Pi_a^\alpha$ in Eq. (3.3), and defining the two $n \times n$ strength transition matrices $\mathbb{W}_\alpha^\pm$ by

$$\mathbb{W}_{\alpha+1}^+ = \mathrm{diag}\{1 - p_1^{(\alpha)+}, \ldots, 1 - p_{n-1}^{(\alpha)+}, 1\} + \mathrm{diag}_\mathrm{l}\{p_1^{(\alpha)+}, \ldots, p_{n-1}^{(\alpha)+}\}, \tag{3.5a}$$

$$\mathbb{W}_{\alpha+1}^- = \mathrm{diag}\{1, 1 - p_2^{(\alpha)-}, \ldots, 1 - p_n^{(\alpha)-}\} + \mathrm{diag}_\mathrm{u}\{p_2^{(\alpha)-}, \ldots, p_n^{(\alpha)-}\}, \tag{3.5b}$$

for $\alpha \geq 0$, setting $\mathbb{W}_0^\pm \equiv \mathbb{I}$, and writing

$$\mathbb{W}_\alpha = g_+ \mathbb{W}_\alpha^+ + g_- \mathbb{W}_\alpha^- = \tfrac{1}{2}\left(\mathbb{W}_\alpha^+ + \mathbb{W}_\alpha^-\right), \tag{3.6}$$

Eq. (3.3) can then be written as

$$\mathbf{\Pi}^{\alpha+1} = \mathbb{W}_{\alpha+1}\mathbf{\Pi}^\alpha. \tag{3.7}$$

This expresses the strength distribution $\mathbf{\Pi}^{\alpha+1}$ at memory storage step $\alpha + 1$ in terms of the action of a strength transition matrix on the distribution $\mathbf{\Pi}^\alpha$ at memory storage step α. It is therefore analogous to $\mathbf{\Sigma}^{\alpha+1} = \mathbb{M}\,\mathbf{\Sigma}^\alpha$. Since $\mathbf{\Sigma}^\alpha = \mathbb{M}^\alpha \mathbf{\Sigma}^0$, we also write $\mathbf{\Pi}^\alpha = \mathbb{D}_\alpha\,\mathbf{\Pi}^0$, where we define

$$\mathbb{D}_\alpha = \mathbb{W}_\alpha \mathbb{W}_{\alpha-1} \cdots \mathbb{W}_1 \mathbb{W}_0, \tag{3.8}$$

so that $\mathbb{D}_\alpha$ is the purely strength transition matrix describing the successive

storage of the first α non-tracked memories (with $\mathbb{D}_0 \equiv \mathbb{I}$ via $\mathbb{W}_0 \equiv \mathbb{I}$). It is therefore the analogue of $\mathbb{M}^\alpha$ describing the full joint filter and strength transitions.

We defined the $\mathbb{W}$ matrices in Eq. (3.5) using $\alpha + 1$ on the left-hand side (LHS) because the storage of memory $\boldsymbol{\xi}^\alpha$ changes the joint probability distribution of a synapse's strength and filter states, and this distribution then determines the probabilities $p_a^{(\alpha)\pm}$, which are the conditional probabilities that upon the storage of the next memory $\boldsymbol{\xi}^{\alpha+1}$, a synapse will potentiate or depress. Hence, the matrices $\mathbb{W}_{\alpha+1}^{\pm}$ are determined by the storage of the earlier memories $\boldsymbol{\xi}^0, \ldots, \boldsymbol{\xi}^\alpha$, but they determine how memory $\boldsymbol{\xi}^{\alpha+1}$ will be stored.

Since the storage of the memories $\boldsymbol{\xi}^\alpha$ represents a set of discrete temporal events, the discrete memory storage-step formulation in terms of the conditional strength-change probabilities $p_a^{(\alpha)\pm}$ and corresponding matrices $\mathbb{W}_{\alpha+1}^{\pm}$ at each storage step is natural. However, these memory storage steps occur as a Poisson process of rate r. So let $\boldsymbol{\Sigma}(t)$ represent the probability distribution of a synapse's joint strength and filter state at time t, where $\boldsymbol{\Sigma}(t) = e^{rt(\mathbb{M}-\mathbb{I})}\boldsymbol{\Sigma}(0)$ with $\boldsymbol{\Sigma}(0) \equiv \boldsymbol{\Sigma}^0 = \mathbb{M}^{\pm}\boldsymbol{A}$, and write $\boldsymbol{\Sigma}(t)$ in a block form analogous to that in Eq. (3.1) for $\boldsymbol{\Sigma}^\alpha$, so that $\boldsymbol{\Sigma}_a(t)$ is the sub-vector of $\boldsymbol{\Sigma}(t)$ corresponding to strength state a, and define the probability distribution of strength states $\boldsymbol{\Pi}(t)$ with components $\Pi_a(t) = \mathbf{1} \cdot \boldsymbol{\Sigma}_a(t)$. By the same reasoning as above, we may then immediately write down the equivalent of Eq. (3.3),

$$\frac{1}{r}\frac{d\Pi_1(t)}{dt} = g_- f_2^-(t) - g_+ f_1^+(t), \tag{3.9a}$$

$$\frac{1}{r}\frac{d\Pi_a(t)}{dt} = g_- f_{a+1}^-(t) + g_+ f_{a-1}^+(t) - g_- f_a^-(t) - g_+ f_a^+(t), \tag{3.9b}$$

$$\frac{1}{r}\frac{d\Pi_n(t)}{dt} = g_+ f_{n-1}^+(t) - g_- f_n^-(t), \tag{3.9c}$$

where

$$f_a^{\pm}(t) = \mathbf{1} \cdot \mathbb{T}^{\pm}\boldsymbol{\Sigma}_a(t). \tag{3.10}$$

Of course, Eq. (3.9) also follows directly from Eq. (3.3) by performing a Poisson sum over Eq. (3.3): the LHSs of Eq. (3.3) turn into the derivatives on the LHSs of Eq. (3.9); and $\sum_{\alpha=0}^{\infty} e^{-rt}\frac{(rt)^\alpha}{\alpha!} f_a^{(\alpha)\pm} \equiv f_a^{\pm}(t)$ precisely because $\mathbf{\Sigma}(t) = e^{rt(\mathbb{M}-\mathbb{I})}\mathbf{\Sigma}^0 \equiv \sum_{\alpha=0}^{\infty} e^{-rt}\frac{(rt)^\alpha}{\alpha!}\mathbf{\Sigma}^\alpha$. Defining the continuous-time conditional strength-change probabilities in analogy with Eq. (3.4),

$$p_a^{\pm}(t) = \frac{f_a^{\pm}(t)}{\Pi_a(t)}, \tag{3.11}$$

and the corresponding transition matrices

$$\mathbb{W}^+(t) = \mathrm{diag}\{1 - p_1^+(t), \ldots, 1 - p_{n-1}^+(t), 1\} + \mathrm{diag}_{\mathrm{l}}\{p_1^+(t), \ldots, p_{n-1}^+(t)\}, \tag{3.12a}$$

$$\mathbb{W}^-(t) = \mathrm{diag}\{1, 1 - p_2^-(t), \ldots, 1 - p_n^-(t)\} + \mathrm{diag}_{\mathrm{u}}\{p_2^-(t), \ldots, p_n^-(t)\}, \tag{3.12b}$$

$$\mathbb{W}(t) = g_+\mathbb{W}^+(t) + g_-\mathbb{W}^-(t) = \tfrac{1}{2}\left[\mathbb{W}^+(t) + \mathbb{W}^-(t)\right], \tag{3.12c}$$

Eq. (3.9) then becomes the time-dependent Master equation

$$\frac{1}{r}\frac{d}{dt}\mathbf{\Pi}(t) = [\,\mathbb{W}(t) - \mathbb{I}\,]\,\mathbf{\Pi}(t). \tag{3.13}$$

The formal solution of this equation can be written as $\mathbf{\Pi}(t) = \overline{\mathbb{U}}(t)\mathbf{\Pi}(0)$ where

$$\overline{\mathbb{U}}(t) = \boldsymbol{\mathcal{T}}\left\{\exp r\int_0^t dt_1\,[\mathbb{W}(t) - \mathbb{I}]\right\}, \tag{3.14}$$

where $\boldsymbol{\mathcal{T}}\{\cdot\}$ denotes the time-ordered product (meta-)operator, although numerical solutions are required in practice. Using the matrices $\mathbb{D}_\alpha$ in Eq. (3.8),

we can also define the Poisson sum

$$\begin{aligned}\mathbb{U}(t) &= e^{-rt}\sum_{\alpha=0}^{\infty}\frac{(rt)^{\alpha}}{\alpha!}\mathbb{D}_{\alpha} \\ &= e^{-rt}\left[\mathbb{I} + rt\,\mathbb{W}_1 + \frac{(rt)^2}{2!}\mathbb{W}_2\mathbb{W}_1 + \frac{(rt)^3}{3!}\mathbb{W}_3\mathbb{W}_2\mathbb{W}_1 + \ldots\right], \qquad (3.15)\end{aligned}$$

leading to the alternative representation of the evolution of $\mathbf{\Pi}(t)$ as $\mathbf{\Pi}(t) = \mathbb{U}(t)\mathbf{\Pi}(0)$.

We note that although the discrete- and continuous-time conditional strength-change probabilities in Eqs. (3.4) and (3.11) are structurally identical, in terms of ratios of probabilities, mathematically they are very different objects. While Poisson summing $f_a^{(\alpha)\pm}$ leads immediately to $f_a^{\pm}(t)$ because these functions are purely linear in $\mathbf{\Sigma}$, it is *not* the case that Poisson summing $p_a^{(\alpha)\pm}$ gives $p_a^{\pm}(t)$. In particular, while say $p_a^{+}(t)$ is the ratio of two Poisson sums (the ratio of $f_a^{+}(t)$ and $\Pi_a(t)$), the Poisson sum of $p_a^{(\alpha)+}$ would be a Poisson sum of the ratio of $f_a^{(\alpha)+}$ and Π_a^{α}. But the Poisson sum of a ratio is not the ratio of two Poisson sums. Hence, despite their interpretations as conditional probabilities and despite the likelihood that their values are therefore very similar, $p_a^{\pm}(t)$ and $p_a^{(\alpha)\pm}$ are mathematically very different, and in particular the continuous-time version is much more complicated than the discrete-time version.

This important mathematical and conceptual difference between $p_a^{\pm}(t)$ and $p_a^{(\alpha)\pm}$ informs the question of the existence of the two representations of the evolution of $\mathbf{\Pi}(t)$, via the Master equation in Eq. (3.13) with solution $\mathbf{\Pi}(t) = \overline{\mathbb{U}}(t)\mathbf{\Pi}(0)$, and via the Poisson-summed evolution matrix $\mathbb{U}(t)$ in Eq. (3.15) automatically giving $\mathbf{\Pi}(t) = \mathbb{U}(t)\mathbf{\Pi}(0)$. By construction of the $\mathbb{W}_{\alpha}$ matrices and the fact that the underlying discrete memory storage process is fundamental, the Poisson-summed matrix $\mathbb{U}(t)$ is the evolution matrix in continuous time describing the evolution of the probability distribution of a synapse's strength states. It then follows trivially, by definition, that $\mathbf{\Pi}(t) = \mathbb{U}(t)\mathbf{\Pi}(0)$.

Given the difference between $p_a^{\pm}(t)$ and $p_a^{(\alpha)\pm}$, it is clear that $\mathbb{W}(t)$ cannot be the Poisson sum of the $\mathbb{W}_\alpha$ matrices. Yet we also have the explicit solution $\mathbf{\Pi}(t) = \overline{\mathbb{U}}(t)\mathbf{\Pi}(0)$, where $\overline{\mathbb{U}}(t)$ is a time-ordered product matrix exponential involving $\mathbb{W}(t)$. Clearly, there is only one correct matrix evolution operator for this problem, and so the question arises as to whether or not $\mathbb{U}(t)$ and $\overline{\mathbb{U}}(t)$ are in fact the very same matrix? By direct evaluation using numerical solutions, we find that while $\mathbb{U}(t)$ and $\overline{\mathbb{U}}(t)$ are indeed very similar (see later), they are not identical. However, by construction, $\mathbb{U}(t)\mathbf{\Pi}(0)$ and $\overline{\mathbb{U}}(t)\mathbf{\Pi}(0)$ are always identical, when $\mathbf{\Pi}(0)$ is the correct initial state. While two matrices $\mathbb{A}$ and $\mathbb{B}$ are identical if, and only if, $\mathbb{A}\boldsymbol{x} = \mathbb{B}\boldsymbol{x}$ for all non-zero vectors $\boldsymbol{x}$, we require the equality $\mathbb{U}(t)\boldsymbol{x} = \overline{\mathbb{U}}(t)\boldsymbol{x}$ only for precisely one vector, $\boldsymbol{x} = \mathbf{\Pi}(0)$, so $\mathbf{\Pi}(0)$ must be a null eigenvector of the in general non-zero matrix $\mathbb{U}(t) - \overline{\mathbb{U}}(t)$. While the matrix $\mathbb{U}(t)$ is the correct time evolution matrix for single synaptic strength states, the matrix $\overline{\mathbb{U}}(t)$ only captures the dynamics at the level of the mean by describing correctly the evolution of $\mathbf{\Pi}(t)$, given $\mathbf{\Pi}(0)$. However, given that $\mathbb{U}(t)$ and $\overline{\mathbb{U}}(t)$ are numerically very similar, and that $\overline{\mathbb{U}}(t)$ is easy to obtain from numerical solutions while $\mathbb{U}(t)$ is a cumbersome Poisson sum that cannot be evaluated exactly but must be truncated at large times, it is typically far more efficient to use $\overline{\mathbb{U}}(t)$ as an approximation to $\mathbb{U}(t)$. This approximation incurs errors that are negligible compared to those from other approximations that we will be forced to make later.

In summary, using the synaptic strength states $\mathbf{\Pi}^\alpha$ or $\mathbf{\Pi}(t)$ and the strength transition matrices $\mathbb{W}_\alpha^{\pm}$ or $\mathbb{W}^{\pm}(t)$, we have integrated out the full filter dynamics contained in the joint states $\mathbf{\Sigma}^\alpha$ or $\mathbf{\Sigma}(t)$ and joint transition matrices $\mathbb{M}^{\pm}$. However, while $\mathbb{M}^{\pm}$ are constant matrices independent of α or t, the matrices $\mathbb{W}_\alpha^{\pm}$ or $\mathbb{W}^{\pm}(t)$ depend explicitly on the memory storage step or time. This step- or time-dependence is the price to be paid for reducing a complex synapse with internal states to a simple synapse without them. While we require the

full, complex synapse formulation in order to determine $\boldsymbol{\Sigma}$, the TDSU formulation eliminates the underlying filter states and allows us to work directly with changes in synaptic strength. It is these changes, and only these changes, that are relevant to changes in the perceptron's activation and thus are relevant to the calculation of passages times for this activation to cross firing threshold.

3.2 Calculation of $p_a^{(\alpha)\pm}$ and $p_a^{\pm}(t)$

We now calculate the strength transition probabilities, $p_a^{(\alpha)\pm}$ and $p_a^{\pm}(t)$. To obtain these, we require either $\boldsymbol{\Sigma}_a^{\alpha}$ or $\boldsymbol{\Sigma}_a(t)$ for use in Eqs. (3.4) or (3.11). Calculating $\boldsymbol{\Sigma}(t)$ is simpler, and then the $\boldsymbol{\Sigma}^{\alpha}$ follow just by undoing the Poisson sum. The initial full synaptic state $\boldsymbol{\Sigma}(0)$ after the storage of $\boldsymbol{\xi}^0$ is either $\mathbb{M}^+\boldsymbol{A}$ or $\mathbb{M}^-\boldsymbol{A}$, but as argued above, it suffices to consider only the case $\boldsymbol{\Sigma}(0) = \mathbb{M}^+\boldsymbol{A}$.

To determine $\boldsymbol{\Sigma}(t)$, we employ the renewal methods found in Elliott (2016a), which we follow closely but using a slightly modified notation. Let $F_{I|J}(t)$ denote the probability of a transition from initial filter state J to final filter state I in time t without either filter threshold being reached. This probability is just the I, J element of the matrix $\exp\left\{rt\left[\frac{1}{2}\left(\mathbb{F}^+ + \mathbb{F}^-\right) - \mathbb{I}\right]\right\}$, and so $F_{I|J}(t) = F_{J|I}(t)$ for symmetric thresholds. Let $G_J^{\pm}(t)$ denote the densities for first reaching the two filter thresholds $\pm\Theta$ at time t. Since we reach upper or lower threshold by taking a single step from the states $+(\Theta - 1)$ or $-(\Theta - 1)$ with rates rg_+ or rg_-, respectively, these densities are just

$$G_J^{\pm}(t) = rg_{\pm}F_{\pm(\Theta-1)|J}(t), \tag{3.16}$$

where here $g_{\pm} = \frac{1}{2}$. Let $H_J(t)$ denote the probability of not having reached either threshold in time t from initial filter state J. We can write $H_J(t)$ in

either of the two forms,

$$H_J(t) = \sum_{I=-(\Theta-1)}^{+(\Theta-1)} F_{I|J}(t), \tag{3.17a}$$

$$H_J(t) = 1 - \int_0^t dt_1 \left[G_J^+(t_1) + G_J^-(t_1) \right], \tag{3.17b}$$

where in the first form, $H_J(t)$ is expressed as the probability of being in any filter state at time t without threshold having been reached, and in the second form, it is expressed as the complement of the probability of having first reached threshold at any time in the interval $[0, t]$. By writing down and solving the Master equation for a symmetric random walk between two absorbing boundaries at $-\Theta$ and $+\Theta$, we can explicitly obtain expressions for the Laplace transforms of $G_J^{\pm}(t)$ (Elliott, 2011),

$$\widehat{G}_J^{\pm}(s) = \frac{\left[\Phi_+(s)\right]^{\Theta \pm J} - \left[\Phi_-(s)\right]^{\Theta \pm J}}{\left[\Phi_+(s)\right]^{2\Theta} - \left[\Phi_-(s)\right]^{2\Theta}}, \tag{3.18}$$

where a hat over a quantity denotes the Laplace transform, s is the transformed variable, and $\Phi_{\pm}(s)$ are the two solutions of $\Phi^2 - 2(1 + s/r)\Phi + 1 = 0$; for notational convenience we will frequently drop the argument of a Laplace transform. We see that $\widehat{G}_{+J}^{+}(s) = \widehat{G}_{-J}^{-}(s)$, a property due to symmetric processes, and hence $\widehat{G}_0^+(s) = \widehat{G}_0^-(s)$, so below we will just write $\widehat{G}_0(s) = \widehat{G}_0^{\pm}(s)$.

To obtain $\mathbf{\Sigma}(t)$, we set up and solve a system of renewal equations for the transition probabilities $P_{I,a|J,b}(t)$ from initial filter state J and strength state b to final filter state I and strength state a in time t. These are given in Eq. (3.3) of Elliott (2016a), so we do not reproduce them here. Defining the set of $n \times n$

matrices

$$\mathbb{G}_J(t) = G_J^-(t)\,\mathrm{diag}_{\mathrm{u}}\{\overbrace{1,\ldots,1}^{n-1}\} + G_J^+(t)\,\mathrm{diag}_{\mathrm{l}}\{\overbrace{1,\ldots,1}^{n-1}\}$$
$$+ \mathrm{diag}\{G_J^-(t), \underbrace{0,\ldots,0}_{n-2}, G_J^+(t)\}, \tag{3.19}$$

and also the $n \times n$ matrices $\mathbb{P}_{I|J}(t)$ with elements $P_{I,a|J,b}(t)$ (i.e. $\left[\mathbb{P}_{I|J}(t)\right]_{a,b} = P_{I,a|J,b}(t)$), the system of renewal equations can be written compactly in the Laplace transformed form

$$\widehat{\mathbb{P}}_{I|J}(s) = \widehat{F}_{I|J}(s)\,\mathbb{I} + \widehat{\mathbb{P}}_{I|0}(s)\,\widehat{\mathbb{G}}_J(s) \tag{3.20a}$$
$$= \widehat{F}_{I|J}(s)\,\mathbb{I} + \widehat{F}_{I|0}(s)\,\widehat{\mathbb{Y}}(s)^{-1}\,\widehat{\mathbb{G}}_J(s), \tag{3.20b}$$

where $\widehat{\mathbb{Y}}(s) = \mathbb{I} - \widehat{\mathbb{G}}_0(s)$ or $\widehat{\mathbb{Y}}(s)^{-1} = \sum_{\alpha=0}^{\infty}\left[\widehat{\mathbb{G}}_0(s)\right]^{\alpha}$, with Eq. (3.20b) following from Eq. (3.20a) by setting $J = 0$ and solving for $\widehat{\mathbb{P}}_{I|0}(s)$. The combinatorial interpretation of Eq. (3.20b) is discussed extensively in Elliott (2016a).

Using the initial state $\boldsymbol{\Sigma}(0) = \mathbb{M}^+\boldsymbol{A}$, we construct the n-dimensional vectors $\boldsymbol{\sigma}_J$ by picking out from $\mathbb{M}^+\boldsymbol{A}$ the probability distribution of strength states for a specific filter state. These are given in Eq. (3.10) of Elliott (2016a) as

$$\boldsymbol{\sigma}_J = \begin{cases} \frac{1}{n}\frac{1}{\Theta^2}\left(\Theta - |J-1|\right)\mathbf{1} & \text{for } J \neq 0 \\ \frac{1}{n}\frac{1}{\Theta^2}\left(\Theta\,\mathbf{1} + \boldsymbol{w}\right) & \text{for } J = 0 \end{cases}, \tag{3.21}$$

where $\boldsymbol{w} = (-1, \overbrace{0,\ldots,0}^{n-2}, +1)^{\mathrm{T}}$. The joint strength and filter distribution $\boldsymbol{\Sigma}(t)$ is then obtained by considering the set of n-dimensional vectors

$$\boldsymbol{P}_I(t) \equiv \sum_{J=-(\Theta-1)}^{+(\Theta-1)} \mathbb{P}_{I|J}(t)\,\boldsymbol{\sigma}_J, \tag{3.22}$$

and we write $P_{I,a}(t) = \left[\boldsymbol{P}_I(t)\right]_a$. These are just the components of $\boldsymbol{\Sigma}(t)$ since $\left[\boldsymbol{\Sigma}_a(t)\right]_I = \left[\boldsymbol{P}_I(t)\right]_a$. Further, we have that $\boldsymbol{\Pi}(t) \equiv \sum_{I=-(\Theta-1)}^{+(\Theta-1)}\boldsymbol{P}_I(t)$ with

components $\Pi_a(t) \equiv \sum_{I=-(\Theta-1)}^{+(\Theta-1)} P_{I,a}(t)$, and that $f_a^{\pm}(t) \equiv P_{\pm(\Theta-1),a}(t)$. With the expression for $\widehat{\mathbb{G}}_J\,\boldsymbol{\sigma}_J$ given in Eq. (3.11) of Elliott (2016a) we may use Eq. (3.20b) to write down $\widehat{P}_{I,a}(s)$. For this we need $\widehat{\mathbb{Y}}(s)^{-1}\sum_{J=-(\Theta-1)}^{+(\Theta-1)}\widehat{\mathbb{G}}_J(s)\,\boldsymbol{\sigma}_J$, which we obtain as

$$\widehat{\mathbb{Y}}^{-1}\sum_{J=-(\Theta-1)}^{+(\Theta-1)}\widehat{\mathbb{G}}_J\,\boldsymbol{\sigma}_J = \frac{1}{n\,\Theta^2}\Bigg\{\frac{2\sum_{J>0}(\Theta-J)(\widehat{G}_J^{+}+\widehat{G}_J^{-})}{1-2\,\widehat{G}_0}\,\mathbf{1}+\frac{2\,\Theta\,\widehat{G}_0}{1-2\,\widehat{G}_0}\,\mathbf{1} + \Big[1+2\sum_{J>0}(\widehat{G}_J^{+}-\widehat{G}_J^{-})\Big]\widehat{\mathbb{Y}}^{-1}\boldsymbol{w}-\boldsymbol{w}\Bigg\}. \tag{3.23}$$

Using the generating function method given in section 3.2.2 of Elliott (2016a), we may evaluate $\widehat{\mathbb{Y}}^{-1}\boldsymbol{w}$ as having the components

$$\big[\widehat{\mathbb{Y}}^{-1}\boldsymbol{w}\big]_a = \frac{1}{\widehat{G}_0}\frac{\Phi_+^{\Theta\,a}-\Phi_+^{\Theta(n+1-a)}}{(\Phi_+^{\Theta}-1)(\Phi_+^{\Theta\,n}+1)}. \tag{3.24}$$

We note for later use the antisymmetry $\big[\widehat{\mathbb{Y}}^{-1}\boldsymbol{w}\big]_a = -\big[\widehat{\mathbb{Y}}^{-1}\boldsymbol{w}\big]_{n+1-a}$. We also require $\sum_{J=-(\Theta-1)}^{+(\Theta-1)} F_{I|J}(t)\,\boldsymbol{\sigma}_J$ to compute $\boldsymbol{P}_I(t)$. However, since

$$p_a^{\pm}(t) = \frac{f_a^{\pm}(t)}{\Pi_a(t)} = \frac{P_{\pm(\Theta-1),a}(t)}{\sum_{I=-(\Theta-1)}^{+(\Theta-1)} P_{I,a}(t)}, \tag{3.25}$$

we need only evaluate $\sum_{J=-(\Theta-1)}^{+(\Theta-1)} F_{I|J}(t)\,\boldsymbol{\sigma}_J$ at $I=\pm(\Theta-1)$ or with a sum over I, so we can use Eqs. (3.16) and (3.17) to rewrite $F_{I|J}(t)$ in terms of $G_J^{\pm}(t)$

and $H_J(t)$. The relevant sums are then found to be

$$\sum_{J=-(\Theta-1)}^{+(\Theta-1)} \widehat{F}_{\pm(\Theta-1)|J}\,\boldsymbol{\sigma}_J = \frac{2}{r}\frac{1}{n\,\Theta^2}\Big[\sum_{J>0}(\Theta-J)(\widehat{G}_J^{+}+\widehat{G}_J^{-})\,\mathbf{1} + \sum_{J>0}(\widehat{G}_J^{+}-\widehat{G}_J^{-})\,\mathbf{1}+\widehat{G}_0\,(\Theta\,\mathbf{1}+\boldsymbol{w})\Big], \quad (3.26a)$$

$$\sum_{J=-(\Theta-1)}^{+(\Theta-1)} \widehat{H}_J\,\boldsymbol{\sigma}_J = \frac{1}{s}\frac{1}{n\,\Theta^2}\Big\{-2\sum_{J>0}(\Theta-J)(\widehat{G}_J^{+}+\widehat{G}_J^{-})\,\mathbf{1}+(1-2\,\widehat{G}_0)\,\boldsymbol{w} + \Theta\big[(1-2\,\widehat{G}_0)+\Theta-1\big]\,\mathbf{1}\Big\}, \quad (3.26b)$$

where in first equation, the result does not depend on the choice of sign of $\pm(\Theta-1)$ on the LHS. Pulling everything together, using Eq. (3.18) to compute the various sums involving $\widehat{G}_J^{\pm}$ in terms of Φ_+, after a great deal of tedious algebra we eventually obtain the final results

$$\widehat{f}_a^{+} = \frac{2}{r}\frac{1}{n\,\Theta^2}\Big\{\frac{\Phi_+ + 1}{\Phi_+ - 1}\,\frac{1}{\Phi_+^{\Theta}+1}\,\frac{\Phi_+^{\Theta\,a}-\Phi_+^{\Theta(n+1-a)}}{\Phi_+^{\Theta\,n}+1} + \frac{2\,\Phi_+(\Phi_+^{\Theta}+1)-\Phi_+^2-\Phi_+^{\Theta}}{(\Phi_+-1)^2\,(\Phi_+^{\Theta}+1)}\Big\}, \quad (3.27a)$$

$$\widehat{f}_a^{-} = \frac{2}{r}\frac{1}{n\,\Theta^2}\Big\{\frac{\Phi_+ + 1}{\Phi_+ - 1}\,\frac{1}{\Phi_+^{\Theta}+1}\,\frac{\Phi_+^{\Theta\,a}-\Phi_+^{\Theta(n+1-a)}}{\Phi_+^{\Theta\,n}+1} + \frac{\Phi_+^2+\Phi_+^{\Theta}}{(\Phi_+-1)^2\,(\Phi_+^{\Theta}+1)}\Big\}, \quad (3.27b)$$

$$\widehat{\Pi}_a = \frac{1}{s}\,\frac{1}{n} + \frac{1}{s}\,\frac{1}{n\,\Theta^2}\,\frac{\Phi_+ + 1}{\Phi_+ - 1}\,\frac{(\Phi_+^{\Theta}-1)^2}{\Phi_+^{\Theta}(\Phi_+^{\Theta}+1)}\,\frac{\Phi_+^{\Theta\,a}-\Phi_+^{\Theta(n+1-a)}}{\Phi_+^{\Theta\,n}+1}. \quad (3.27c)$$

The identities

$$\frac{1}{2}\Big[\Pi_a(t)+\Pi_{n+1-a}(t)\Big] = \frac{1}{n}, \quad (3.28a)$$

$$\frac{1}{2}\Big[f_a^{+}(t)+f_{n+1-a}^{-}(t)\Big] = \frac{1}{n}\,\frac{1}{\Theta^2}, \quad (3.28b)$$

follow automatically from the antisymmetry of the components of $\widehat{\mathbb{Y}}^{-1}\boldsymbol{w}$ and the fact that $2\,\Phi_+/(\Phi_+-1)^2 \equiv r/s$.

It remains to determine the inverse Laplace transforms of the expressions in Eq. (3.27). These expressions are in terms of Φ_+, but Φ_+ is a solution of the equation $\Phi^2 - 2(1 + s/r)\Phi + 1 = 0$, and we need to invert with respect to s and not Φ_+. We note, however, that all the poles in Φ_+ of the expressions in Eq. (3.27) are of unit modulus, and that when we factorise the denominators, a pole $e^{+i\varphi}$ and its complex conjugate $e^{-i\varphi}$ combine via

$$\frac{2\,\Phi_+}{(\Phi_+ - e^{+i\varphi})(\Phi_+ - e^{-i\varphi})} = \frac{2\,\Phi_+}{\Phi_+^2 + 1 - 2\,\Phi_+ \cos\varphi} = \frac{r}{s - r(\cos\varphi - 1)} \tag{3.29}$$

to give a simple pole in s at $r(\cos\varphi - 1) \leq 0$. Using this fact, the evaluation of the inverse Laplace transforms is straightforward if tiresome. We eventually obtain

$$\begin{aligned} n\,\Theta^2\, f_a^{\pm}(t) = \\ & 1 + \frac{4}{\Theta}\Big[\pm 1 - (-1)^a\,\mathrm{C}_n(a)\Big] \sum_{m=0}^{\left\lfloor\frac{\Theta-1}{2}\right\rfloor} \cos^2\tfrac{(2m+1)\pi}{2\Theta}\, e^{-rt\left[1-\cos\frac{(2m+1)\pi}{\Theta}\right]} \\ & - \frac{4}{\Theta n} \sum_{\substack{m=0 \\ n\nmid 2m+1}}^{\left\lfloor\frac{n\Theta-1}{2}\right\rfloor} \cos\tfrac{(2m+1)(2a-1)\pi}{2n}\, \frac{\cos^2\frac{(2m+1)\pi}{2n\Theta}}{\cos\frac{(2m+1)\pi}{2n}}\, e^{-rt\left[1-\cos\frac{(2m+1)\pi}{n\Theta}\right]}, \end{aligned} \tag{3.30a}$$

$$\begin{aligned} n\,\Pi_a(t) = \\ & 1 - \frac{4}{\Theta^3}\Bigg\{(-1)^a\,\mathrm{C}_n(a) \sum_{m=0}^{\left\lfloor\frac{\Theta-1}{2}\right\rfloor} \cot^2\tfrac{(2m+1)\pi}{2\Theta}\, e^{-rt\left[1-\cos\frac{(2m+1)\pi}{\Theta}\right]} \\ & + \frac{1}{n} \sum_{\substack{m=0 \\ n\nmid 2m+1}}^{\left\lfloor\frac{n\Theta-1}{2}\right\rfloor} \cos\tfrac{(2m+1)(2a-1)\pi}{2n}\, \cot^2\tfrac{(2m+1)\pi}{2n\Theta}\, \frac{\sin^2\frac{(2m+1)\pi}{2n}}{\cos\frac{(2m+1)\pi}{2n}}\, e^{-rt\left[1-\cos\frac{(2m+1)\pi}{n\Theta}\right]}\Bigg\}, \end{aligned} \tag{3.30b}$$

with $\lfloor\cdot\rfloor$ being the floor function, and where the function

$$\mathrm{C}_n(a) = \begin{cases} 1 & \text{for } n \text{ even} \\ \frac{n-2\,a+1}{n} & \text{for } n \text{ odd} \end{cases}, \tag{3.31}$$

depends on the parity of n, and where the condition $n \nmid 2m+1$ means that values of m for which n divides $2m+1$ are excluded from the sum, with this condition biting only when n is odd.

For the results at the discrete memory storage step α rather than at continuous time t, we need $f_a^{(\alpha)\pm}$ and Π_a^α instead of $f_a^\pm(t)$ and $\Pi_a(t)$. To obtain these from Eq. (3.30), we simply undo the Poisson sums by replacing the exponentials in time of the form $e^{-rt(1-\cos\varphi)}$ with powers in α of the form $\cos^\alpha\varphi$, where for $\alpha = 0$ this is always unity regardless of φ. Of course, the equivalent identities in Eq. (3.28) continue to hold at each discrete memory storage step. When it is not necessary to specify continuous- or discrete-time variables, such as in Eq. (3.28), we just drop the time argument or step superscript for convenience and/or generality.

Using the result for $\Pi_a(t)$ in Eq. (3.30b) we may evaluate $\mu_1(t) \equiv \mathsf{E}[S(t)] = \sum_{a=1}^n \Omega_a\,\Pi_\alpha(t)$, the mean strength of a single synapse, where $\mathsf{E}[\,\cdot\,]$ denotes the expectation value. Because the perceptron's activation in response to the tracked memory, $h(t) = \frac{1}{N}\sum_{i=1}^N \widetilde{S}_i(t)$, is an average over N synapses' identically-distributed tilded strengths, we also then have that $\mathsf{E}[h(t)] \equiv \mu_1(t)$ as we need only consider $\boldsymbol{\Sigma}(0) = \mathbb{M}^+\boldsymbol{A}$ in the initial mixture. We refer to $\mu_1(t)$ as the mean tracked memory signal, or just the mean memory signal. In Elliott (2016a) we obtained $\widehat{\mu}_1(s)$ and then inverse Laplace transformed it,

with the result

$$\mu_1(t) = \frac{4}{n(n-1)\Theta^3}\left(\frac{1}{n}\sum_{m=0}^{\left\lfloor\frac{n\Theta-1}{2}\right\rfloor}\cot^2\tfrac{(2m+1)\pi}{2n\Theta}\,e^{-rt\left[1-\cos\frac{(2m+1)\pi}{n\Theta}\right]}\right.$$
$$\left. - n\sum_{m=0}^{\left\lfloor\frac{\Theta-1}{2}\right\rfloor}\cot^2\tfrac{(2m+1)\pi}{2\Theta}\,e^{-rt\left[1-\cos\frac{(2m+1)\pi}{\Theta}\right]}\right). \tag{3.32}$$

This also follows directly from Eq. (3.30b), the restriction $n \nmid 2m+1$ on the second sum falling away. The mean μ_1^α at memory storage step α is automatic.

It is instructive to derive the distribution of synaptic strengths $\mathbf{\Pi}(t)$ for a simple, multistate SU synapse that changes strength with fixed probability p in response to a plasticity induction signal, and we will require this distribution below. The $n \times n$ transition matrices in this case are just

$$\mathbb{W}_{\rm su}^{+} = \mathrm{diag}\big\{\underbrace{1-p,\ldots,1-p}_{n-1},1\big\} + p\,\mathrm{diag}_{\rm l}\big\{\underbrace{1,\ldots,1}_{n-1}\big\}, \tag{3.33a}$$

$$\mathbb{W}_{\rm su}^{-} = \mathrm{diag}\big\{1,\underbrace{1-p,\ldots,1-p}_{n-1}\big\} + p\,\mathrm{diag}_{\rm u}\big\{\underbrace{1,\ldots,1}_{n-1}\big\}, \tag{3.33b}$$

which are independent of α, and we write $\mathbb{W}_{\rm su} = \frac{1}{2}\left(\mathbb{W}_{\rm su}^{+} + \mathbb{W}_{\rm su}^{-}\right)$. The eigenvalues λ_m and corresponding orthonormal eigenvectors $\boldsymbol{\epsilon}^m$ with components ϵ_a^m, $m = 0,\ldots,n-1$, of the matrix $\mathbb{W}_{\rm su}$ are given by (Elliott, 2019)

$$\lambda_m = 1 - p\left(1-\cos\frac{m\pi}{n}\right), \tag{3.34a}$$

and

$$\epsilon_a^m = \begin{cases} \sqrt{\frac{1}{n}} & \text{for } m = 0 \\ \sqrt{\frac{2}{n}}\cos\frac{m(2a-1)\pi}{2n} & \text{for } m = 1,\ldots,n-1 \end{cases}. \tag{3.34b}$$

The equilibrium distribution is just $\boldsymbol{A} = \frac{1}{n}\,\mathbf{1}$, so $\mathbf{\Pi}(0) = \mathbb{W}_{\rm su}^{+}\boldsymbol{A} = \frac{1}{n}(\mathbf{1} + p\,\boldsymbol{w})$.

Writing $e^{rt(\mathbb{W}_{\rm su}-\mathbb{I})}$ via the eigen-decomposition of $\mathbb{W}_{\rm su}$, we obtain

$$\mathbf{\Pi}(t) = \frac{1}{n}\left\{\mathbf{1} + p\sqrt{\tfrac{2}{n}}\sum_{m=1}^{n-1}[(-1)^m - 1]\cos\tfrac{m\pi}{2n}\, e^{-prt(1-\cos\frac{m\pi}{n})}\,\boldsymbol{\epsilon}^m\right\}. \tag{3.35}$$

Only odd values of m contribute to the sum, so we arrive at

$$n\,\Pi_a(t) = 1 - \frac{4\,p}{n}\sum_{m=0}^{\lfloor\frac{n-1}{2}\rfloor}\cos\tfrac{(2m+1)(2a-1)\pi}{2n}\,\cos\tfrac{(2m+1)\pi}{2n}\,e^{-prt\left[1-\cos\frac{(2m+1)\pi}{n}\right]} \tag{3.36}$$

for the components. These components also satisfy the identity in Eq. (3.28a). If we set $\Theta = 1$ in Eq. (3.30b), then the first sum vanishes and the second sum simplifies precisely to the form in Eq. (3.36) with $p = 1$. We also obtain

$$\mu_1(t) = \frac{4\,p}{n^2(n-1)}\sum_{m=0}^{\lfloor\frac{n-1}{2}\rfloor}\cot^2\tfrac{(2m+1)\pi}{2n}\,e^{-prt\left[1-\cos\frac{(2m+1)\pi}{n}\right]}, \tag{3.37}$$

for $\mathsf{E}[h(t)]$ for an SU, which is Eq. (3.61) in Elliott (2016a).

3.3 Illustrative Behaviours

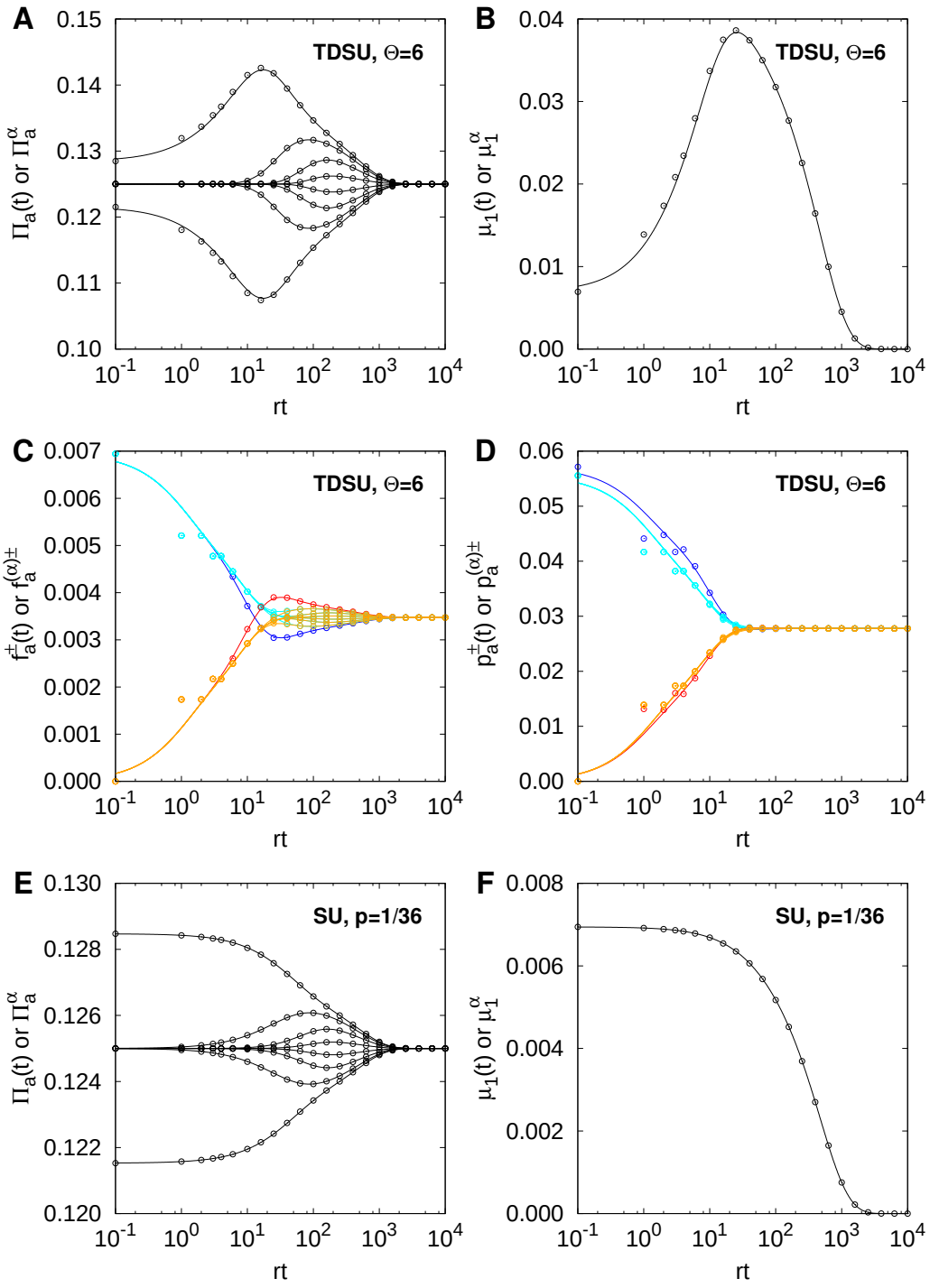
With these explicit results for Π_a, μ_1, $f_a^{\pm}$ and $p_a^{\pm}$ in hand, we may now illustrate the typical behaviour of these functions, and also examine the difference between $\mathbb{U}(t)$ and $\overline{\mathbb{U}}(t)$.

In Fig. 2 we plot these functions against rt for a TSDU with $\Theta = 6$ and $n = 8$ using both continuous- and discrete-time results. We also show results for an SU with $p = 1/\Theta^2$, where this value of p corresponds to the equilibrium value of all forms of $p_a^{\pm}$ for a TDSU. To enable comparison between the continuous and discrete forms, we imagine the discrete steps occurring at unit increments of rt, or we just set $r = 1$ Hz. Fig. 2A shows Π_a for the TDSU. Because $\Pi_a(t)$ is a Poisson sum over the Π_a^{α}, the former should rapidly converge to the latter, with noticeable differences only at smaller times. Initially, $\mathbf{\Pi}^0 = \frac{1}{n}\left(\mathbf{1} + \frac{1}{\Theta^2}\boldsymbol{w}\right)$

so that only the terminal probabilities Π_1^0 and Π_n^0 differ from $1/n$. In discrete time, the non-terminal Π_a^α's remain at $1/n$ for smaller values of α, but pairwise (Π_a and Π_{n+1-a}) peel away from $1/n$ as α increases.[4] The continuous-time non-terminal forms therefore slowly move away from $1/n$ as t increases. Most strikingly, however, Π_n initially increases and Π_1 therefore initially decreases: a synapse becomes increasingly likely to occupy the highest strength state and so increasingly unlikely to occupy the lowest strength state. Therefore, the mean memory signal $\mathsf{E}[h] = \mu_1$ initially increases, as we see in Fig. 2B. We have extensively discussed the behaviour and causes of this initial increase in the mean memory signal elsewhere (Elliott & Lagogiannis, 2012; Elliott, 2016a). In brief, the initial storage of $\boldsymbol{\xi}^0$ shifts the filter state of a synapse, say i, either upwards or downwards depending on the sign of ξ_i^0. If the synapse does not immediately change its strength, this change in filter state biases the synapse on average either to potentiate ($\xi_i^0 = +1$) or to depress ($\xi_i^0 = -1$) in response to future memory storage steps, leading to the initial rise in the memory signal. The signal reaches a peak at some future time τ_{peak} and starts falling only once this bias has worked out of the system.

In Figs. 2C and 2D we show the behaviours of $f_a^\pm$ and $p_a^\pm$, respectively. We do not show f_n^+ or f_1^- in Fig. 2C because they are not relevant to the conditional strength change probabilities. At small time steps, all the $f_a^{(\alpha)+}$ are identical. Then $f_1^{(\alpha)+}$ and $f_n^{(\alpha)+}$ peel away, each taking different values, with the non-terminal $f_a^{(\alpha)+}$'s remaining identical. This process continues pairwise, with $f_2^{(\alpha)+}$ and $f_{n-1}^{(\alpha)+}$ peeling away later, then $f_3^{(\alpha)+}$ and $f_{n-2}^{(\alpha)+}$ later still, etc., in a process similar to that for Π_a^α in Fig. 2A. An identical process occurs for

[4] Immediately above Eq. (2.14) in Elliott (2016a) we incorrectly stated that the non-terminal Π_a^α's remain at $1/n$ for all time. However, the result in Eq. (2.14) is valid because it only depends on $\boldsymbol{\Pi} - \boldsymbol{A}$ being antisymmetric. No other results in Elliott (2016a) depend on this incorrect statement.


A
TDSU, Θ=6
$\Pi_a(t)$ or Π_a^α
rt
B
TDSU, Θ=6
$\mu_1(t)$ or μ_1^α
rt
C
TDSU, Θ=6
$f_a^\pm(t)$ or $f_a^{(\alpha)\pm}$
rt
D
TDSU, Θ=6
$p_a^\pm(t)$ or $p_a^{(\alpha)\pm}$
rt
E
SU, p=1/36
$\Pi_a(t)$ or Π_a^α
rt
F
SU, p=1/36
$\mu_1(t)$ or μ_1^α
rt

the $f_a^{(\alpha)-}$'s, except that they rise from zero initially. We note the opposite behaviours of the pairs f_a^+ and f_{n+1-a}^-, consistent with the identity in Eq. (3.28b). The Poisson-summed, continuous time forms $f_a^{\pm}(t)$ of course exhibit very similar behaviours to the discrete forms. The two terminal cases f_1^+ and f_n^- differ from the non-terminal forms at earlier times, and overshoot the equilibrium value $1/(n\Theta^2)$ to greater extents than the non-terminal forms. This reflects the different dynamics of saturated and non-saturated strength states, with the former having only one immediately neighbouring strength state and the latter two such states. For $p_a^{\pm}$ in Fig. 2D, we see somewhat similar behaviour to the $f_a^{\pm}$. For the discrete time values, we again have successive pairwise peeling away, except that the initial terminal values also differ, because of the initial terminal values of Π_a. Despite this peeling away, the numerical values

---

Figure 2 (previous page): Behaviours of Π_a, μ_1 and $f_a^{\pm}$ for a TDSU with $\Theta = 6$ and $n = 8$, and for an SU with $p = 1/\Theta^2$ and $n = 8$. In all panels, continuous-time results are shown as solid lines and discrete-time results are shown as circles; the discrete-time point $\alpha = 0$ has been moved to $rt = 0.1$ on this logarithmic scale so that it can be displayed without excessive distortion. (A) Π_a for the TDSU for $a = 1, \ldots, 8$. Moving from bottom to top of the graph, the results correspond to $\Pi_1, \ldots, \Pi_8$. (B) μ_1 for the TDSU. (C) $f_a^{\pm}$ for the TDSU. Results in blue correspond to f_1^+, while those in cyan correspond to $f_2^+, \ldots, f_7^+$, with these latter running in order from bottom to top at $rt = 10^2$. Results in red correspond to f_n^-, while those in orange correspond to $f_{n-1}^-, \ldots, f_2^-$, with these latter running in that order from top to bottom at $rt = 10^2$. (D) $f_a^{\pm}$ for the TDSU. Line styles are identical to those in panel C; all cyan results are indistiguishable on this graph, as are all orangle results. (E) Π_a for the SU, in an identical format to and with the same ordering as panel A. (F) μ_1 for the SU.

of the non-terminal $p_a^{(\alpha)+}$ remain extremely similar, and are visually indistinguishable in Fig. 2D; similarly for $p_a^{(\alpha)-}$. Thus, the changes in the non-terminal Π_a's almost exactly cancel the changes in the non-terminal $f_a^{\pm}$'s. As we have noted, the $p_a^{\pm}(t)$ are *not* merely Poisson sums of the $p_a^{(\alpha)\pm}$. Nevertheless, the continuous-time forms reflect the behaviour of the discrete-time forms, because the numerical values of the Π_a and $f_a^{\pm}$ are similar at equivalent discrete- and continuous-time points. Comparing Figs. 2B and 2D, we see that all the probabilities $p_a^{\pm}$ have essentially converged and are indistinguishable from their equilibrium value of $1/\Theta^2$ by the time τ_{peak} that μ_1 reaches its maximum value. The conditional strength change probabilities $p_a^{\pm}$ can therefore be well approximated by their equilibrium value long before other dynamical quantities, such as μ_1 and Π_a, have reached (or come close to) their equilibrium values. We shall exploit this behaviour later to simplify FPT calculations.

For comparison with the TDSU, we show results for an asymptotic probability-matched SU in Figs. 2E and 2F. Comparing Figs. 2A and 2E, the terminal Π_a's for an SU relax monotonically to their equilibrium value, in contrast to their behaviour in the TDSU. Although the non-terminal Π_a's appear similar between the TDSU and SU, we note the difference in scale of the ordinates on these two graphs: the TDSU non-terminal probabilities rise or fall to greater extents than their SU equivalents, and indeed for the TDSU, Π_2 and Π_{n-1} reach maximum and minimum values that are above and below, respectively, the initial values of Π_1 and Π_n for an SU. For the mean memory signal in Fig. 2F, we see that it falls monotonically to its equilibrium value of zero, in contrast to the initial rise in the TDSU's signal in Fig. 2B.

Above we discussed in general terms the two matrices $\overline{\mathbb{U}}(t)$ and $\mathbb{U}(t)$ in Eqs. (3.14) and (3.15). Having determined the probabilities $p_a^{\pm}(t)$ and $p_a^{(\alpha)\pm}$ from which the matrices $\mathbb{W}(t)$ and $\mathbb{W}_\alpha$ and thence $\overline{\mathbb{U}}(t)$ and $\mathbb{U}(t)$ are constructed, we can now be more concrete. Specifically, in Fig. 3 we plot the

relative normed difference $\|\mathbb{U}(t) - \overline{\mathbb{U}}(t)\|_2/\|\mathbb{U}(t)\|_2$ as a function of time for different choices of Θ and n. We obtain $\mathbb{U}(t)$ directly from its Poisson sum in Eq. (3.15), truncating the sum when further contributions are negligible. We obtain $\overline{\mathbb{U}}(t)$ by numerically solving the matrix differential equation $d\overline{\mathbb{U}}(t)/dt = [\,\mathbb{W}(t) - \mathbb{I}\,]\,\overline{\mathbb{U}}(t)$ subject to the initial condition $\overline{\mathbb{U}}(0) = \mathbb{I}$. For $\Theta = 2$ in Fig. 3A, the relative normed difference is less than 0.03 for n up to 8, although it increases as n increases but with the rate of increase falling. As Θ increases, the difference decreases, so that for $\Theta = 8$, the relative difference is well under 0.001 for $n = 8$. These numbers correspond to the largest relative difference between $\mathbb{U}(t)$ and $\overline{\mathbb{U}}(t)$, which occurs in the vicinity of τ_{peak}, the time at which $\mu_1(t)$ reaches its maximum. The fact that this relative difference is typically so small indicates that the two matrices $\mathbb{U}(t)$ and $\overline{\mathbb{U}}(t)$ are almost always very similar, and sufficiently so that $\overline{\mathbb{U}}(t)$ may be used as an approximation to $\mathbb{U}(t)$.

Explicit, closed-form results for $\mathbb{U}(t)$ and in particular $\overline{\mathbb{U}}(t)$ are likely impossible to obtain, but we can compute the first few terms in their Taylor expansions for small, definite choices of Θ and n, and it is instructive to do so. For $n = 3$ and $\Theta = 4$, for example, we find

$$\mathbb{U}(t) = \mathbb{I} + rt \begin{pmatrix} -\frac{1}{15} & 0 & 0 \\ +\frac{1}{15} & -\frac{1}{16} & 0 \\ 0 & +\frac{1}{16} & 0 \end{pmatrix} + \frac{(rt)^2}{2!} \begin{pmatrix} +\frac{17}{960} & +\frac{15}{1024} & 0 \\ -\frac{1}{48} & +\frac{11}{2304} & +\frac{1}{72} \\ +\frac{1}{320} & -\frac{179}{9216} & -\frac{1}{72} \end{pmatrix} + \cdots , \tag{3.38a}$$

$$\overline{\mathbb{U}}(t) = \mathbb{I} + rt \begin{pmatrix} -\frac{1}{15} & 0 & 0 \\ +\frac{1}{15} & -\frac{1}{16} & 0 \\ 0 & +\frac{1}{16} & 0 \end{pmatrix} + \frac{(rt)^2}{2!} \begin{pmatrix} +\frac{1}{60} & +\frac{1}{64} & 0 \\ -\frac{1}{48} & +\frac{1}{256} & +\frac{1}{68} \\ +\frac{1}{240} & -\frac{5}{256} & -\frac{1}{68} \end{pmatrix} + \cdots . \tag{3.38b}$$

Pushing out the calculation to higher powers in rt is possible, but we merely obtain increasingly bewildering matrix coefficients. The matrices $\mathbb{U}(t)$ and $\overline{\mathbb{U}}(t)$

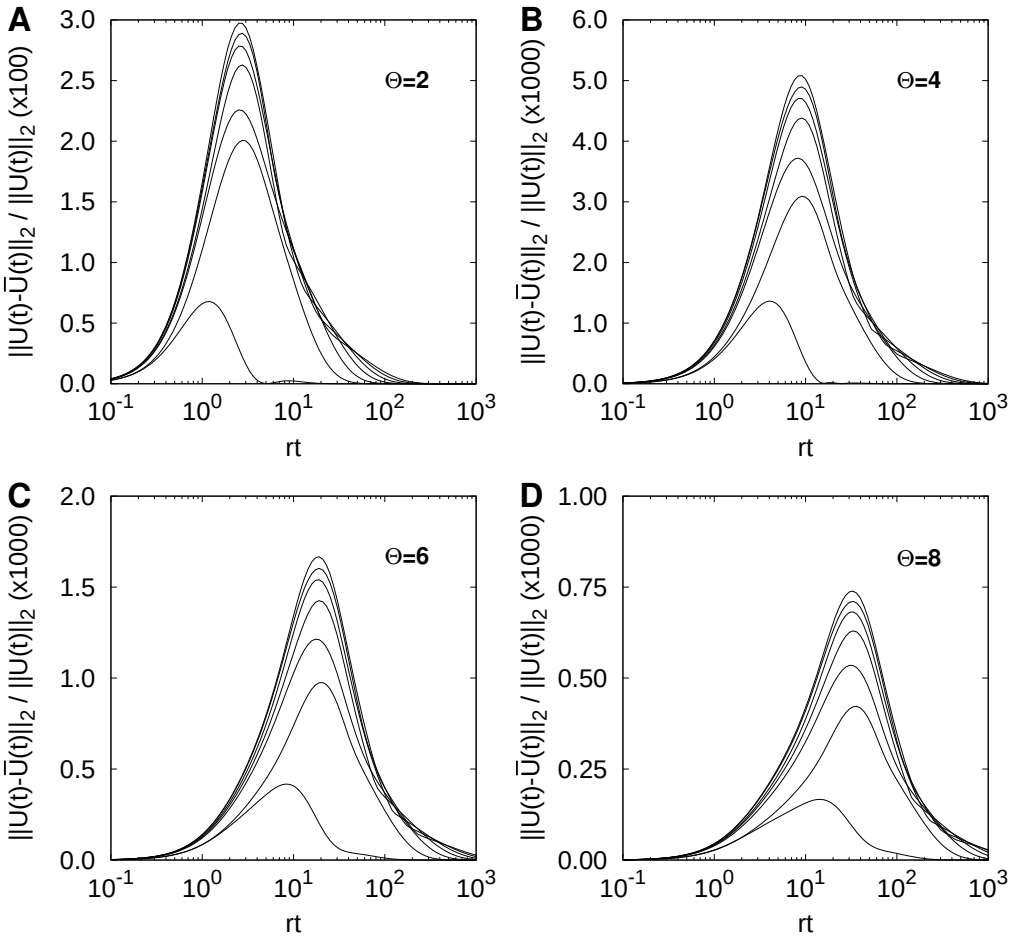


Figure 3: Relative normed difference $\|\mathbb{U}(t) - \overline{\mathbb{U}}(t)\|_2 / \|\mathbb{U}(t)\|_2$ as a function of rt, for various choices of n and Θ. In each panel we take a fixed choice of Θ as indicated, and show results for $n = 2, \ldots, 8$, corresponding to the lines running bottom to top in each panel. Note the difference in scaling in panel A compared to the other panels.

start to differ at $\mathcal{O}\big((rt)^2\big)$. We know, however, that $\mathbb{U}(t)\,\mathbf{\Pi}(0) = \overline{\mathbb{U}}(t)\,\mathbf{\Pi}(0)$ so that $\mathbf{\Pi}(0) = \frac{1}{n}\left(\mathbf{1} + \frac{1}{\Theta^2}\boldsymbol{w}\right)$ must be a null eigenvector of $\mathbb{U}(t) - \overline{\mathbb{U}}(t)$. This must be true order-by-order in rt. Successively determining the eigenvectors and eigenvalues of the matrix differences order-by-order in the expansions in Eq. (3.38), we do indeed find precisely that $\mathbf{\Pi}(0) = \frac{1}{3}\left(\mathbf{1} + \frac{1}{16}\boldsymbol{w}\right)$ is always a null eigenvector.

4 Dynamics of Polysynaptic Configurations

In the previous section we integrated out a synapse's internal filter states so that we could consider transitions only in a synapse's strength states and not in its joint strength and filter states. We now use this TDSU formulation of the dynamics of single synapses to examine the dynamics of multiple synapses.

4.1 Generating Functions for Polysynaptic Configurations

A single filter-based synapse is defined by its strength state, taking one of n distinct values, and by its filter state, independently taking one of $2\,\Theta - 1$ distinct values. The size of the space of all possible, distinguishable joint strength and filter states for N synapses is therefore $[(2\,\Theta - 1)n]^N$. Since we have integrated out the filter states, we need only distinguish the strength states, so the size of the space is reduced to n^N. The same is true for the tilded strength states. However, the perceptron's activation is a linear sum over all (identically distributed) tilded synaptic strengths, $h(t) = \frac{1}{N}\sum_{i=1}^{N}\widetilde{S}_i(t)$. This sum is determined by the number $N_a(t)$ of synapses that have (tilded) strength Ω_a, $a = 1, \ldots, n$. Which particular synapses have which particular strengths, provided that $N_a(t)$ of them have (tilded) strength Ω_a, is irrelevant because the perceptron cannot distinguish between synapses just so long as they have the same (tilded) strength. Thus, a synaptic configuration is actually determined for our purposes by the numbers $N_a(t)$, where $\sum_{a=1}^{N} N_a(t) = N$. The number

of such indistinguishable configurations is the number of partitions of N into the n parts $N_1(t), \ldots, N_n(t)$, allowing for assignments of zero to some but not all of the $N_a(t)$. This number is the binomial coefficient ${}^{N+n-1}C_{n-1}$, which grows polynomially in N, and is typically vastly smaller that n^N, which grows exponentially in N.

We are not so much interested in the dynamics of these ${}^{N+n-1}C_{n-1}$ polysynaptic configurations as in the dynamics of the perceptron's activation, and specifically in the MFPT for it to cross firing threshold. Ideally, we would like to derive a transition matrix describing the transitions in perceptron activation. The number of distinct, distinguishable values that $h(t)$ can take is determined by the smallest change in its value, which is $2/[N(n-1)]$, where $2/(n-1)$ is the smallest possible change in a single synapse's strength. Therefore, $h(t)$ can take $N(n-1)+1$ distinct values, which is the number of points between -1 and $+1$, inclusive, with spacing $2/[N(n-1)]$. The required transition matrix would then be $[N(n-1)+1] \times [N(n-1)+1]$ in size.

The number $N(n-1)+1$ of distinct values of $h(t)$ is in general very much smaller than the number ${}^{N+n-1}C_{n-1}$ of indistinguishable polysynaptic configurations. The mapping from synaptic configurations to values of perceptron activation is therefore in general highly degenerate: distinct synaptic configurations characterised by different sets of numbers $N_a(t)$ can produce the same value of $h(t)$. For example, a pair of synapses with strengths $+1$ and -1 contribute to $h(t)$ identically to a pair of synapses both of whose strengths are zero, when zero is an allowed value of strength. For the particular case of $n = 2$, we have that ${}^{N+n-1}C_{n-1} \equiv N(n-1)+1$, so the mapping from synaptic configuration to perceptron activation is one-to-one and unique, and we can work exclusively with the latter and dispense with the former. However, for $n > 2$, the state of the perceptron's activation does not uniquely determine the underlying polysynaptic configuration. As transitions in the latter determine

transitions in the former and not *vice versa*, for $n > 2$ we cannot formulate the dynamics purely in terms of the transitions in $h(t)$.

Instead, we must extend the dynamics to include further polysynaptic functions that collectively fully specify the underlying polysynaptic state, as we did in Elliott (2019) for simple, multistate synapses without internal states. We want the set of these functions to include the perceptron's activation, so that in deriving the dynamics of the entire set, we also include the activation. We defined $h(t)$ above in continuous time, but we now also define $h^{(\alpha)}$ in the obvious way as the perceptron's activation upon the storage of memory α, with $h(t)$ being a Poisson sum over the $h^{(\alpha)}$. Similarly, we have $N_a(t)$, which define the vector $\boldsymbol{N}(t)$ with these components, and then also $\boldsymbol{N}^\alpha$ with components N_a^α. We write $\boldsymbol{\Omega}$ as the vector of possible synaptic strengths, with components Ω_a. Then we may write

$$h = \frac{1}{N}\,\boldsymbol{\Omega}\cdot\boldsymbol{N}, \tag{4.1a}$$

$$1 = \frac{1}{N}\,\mathbf{1}\cdot\boldsymbol{N}, \tag{4.1b}$$

where we have that $\mathbf{1}\cdot\boldsymbol{\Omega} \equiv 0$. The vector $\boldsymbol{N}$ fully specifies the (tilded) synaptic configuration, subject to the constraint in Eq. (4.1b). We pick a set of n orthogonal vectors $\boldsymbol{e}^a$, $a = 1, \ldots, n$, containing the two vectors $\mathbf{1}$ and $\boldsymbol{\Omega}$ and define the variables

$$v_a = \frac{1}{N}\,\boldsymbol{e}^a\cdot\boldsymbol{N}, \tag{4.2}$$

and write $\boldsymbol{v}$ as the vector with components v_a. Because the vectors $\boldsymbol{e}^a$ are orthogonal and complete, and because $\boldsymbol{N}$ fully specifies the (tilded) synaptic configuration, the vector $\boldsymbol{v}$, which includes h as one of its components, therefore also completely specifies the configuration. Previously (Elliott, 2019) we used

the orthogonal vectors $\boldsymbol{\varepsilon}^m$, $m = 0, \ldots, n-1$, with components

$$\varepsilon_a^m = \sum_{j=0}^{m} (-1)^j \, b_{n-a-j}^m \, {}^m C_j, \text{ where } b_i^m = \frac{{}^{m+i}C_m \, {}^{n-1-i}C_m}{{}^{n-1}C_m}, \tag{4.3}$$

to define v_m, where $\boldsymbol{\varepsilon}^0 = \mathbf{1}$ and $\boldsymbol{\varepsilon}^1 = \boldsymbol{\Omega}$. These vectors are normalised so that $\varepsilon_n^m = +1$, and we may see that $\varepsilon_a^m = (-1)^m \, \varepsilon_{n+1-a}^m$, so that "even" ("odd") vectors $\boldsymbol{\varepsilon}^{2i}$ ($\boldsymbol{\varepsilon}^{2i+1}$) are symmetric (antisymmetric) when reflected about their centres. Allowing the vector $\boldsymbol{\varepsilon}^0$ would lead to $v_0 = \frac{1}{N} \, \boldsymbol{\varepsilon}^0 \cdot \boldsymbol{N}$, but as we wish to maintain the convention that n-dimensional vectors have components indexed by letters such as $a, b = 1, \ldots, n$, we re-order this previous set of vectors by defining the vectors $\boldsymbol{e}^a$ via

$$\boldsymbol{e}^a = \begin{cases} \boldsymbol{\varepsilon}^a & \text{for } a = 1, \ldots, n-1 \\ \boldsymbol{\varepsilon}^0 & \text{for } a = n \end{cases} . \tag{4.4}$$

Hence $\boldsymbol{e}^n \equiv \mathbf{1}$ and $v_n \equiv 1$ instead of $v_0 \equiv 1$. This change in position of $\mathbf{1}$ in the set also maintains the fully dynamical variables $v_1, \ldots, v_{n-1}$ in their original ordering, and so we may continue to refer to these variables as "even" or "odd" variables depending on the parity of a; this was a critical distinction in our earlier analysis because of the symmetry properties of $\boldsymbol{e}^a$ (Elliott, 2019). Defining the $n \times n$ matrix $\mathbb{E}$ whose a^{th} row contains the vector $\boldsymbol{e}^a$, we can write

$$\boldsymbol{v} = \frac{1}{N} \, \mathbb{E} \, \boldsymbol{N}, \tag{4.5}$$

and so also $\boldsymbol{N} = N \, \mathbb{E}^{-1} \, \boldsymbol{v}$. Because $\mathbb{E}$ is a matrix whose rows are a set of orthogonal vectors, we have $\mathbb{E} \, \mathbb{E}^{\mathrm{T}} = \mathrm{diag}\{|\boldsymbol{e}^1|^2, \ldots, |\boldsymbol{e}^n|^2\} \equiv \mathbb{Q}$, defining the diagonal matrix $\mathbb{Q}$. Thus, $\mathbb{E}^{-1} = \mathbb{E}^{\mathrm{T}} \, \mathbb{Q}^{-1}$, so that $\mathbb{E}^{-1}$ is just a matrix whose columns are proportional to the vectors $\boldsymbol{e}^a$. Because it is more convenient for the set of vectors $\boldsymbol{e}^a$ to be orthogonal rather than orthonormal, so that $\boldsymbol{e}^1 \equiv \Omega$ and hence $v_1 \equiv h$, the non-identity scaling matrix $\mathbb{Q}$ appears in $\mathbb{E}^{-1}$.

The vectors $\boldsymbol{N}$ and $\boldsymbol{v}$ are of course random variables subject to probability distributions. To write down the unconditional probability generating functions (PGFs) for the distributions of $\boldsymbol{N}^\alpha$ and $\boldsymbol{v}^\alpha$ at each memory storage step, we observe that the PGF for a single synapse's (tilded) strength state is just $\boldsymbol{\Pi}^\alpha \cdot \boldsymbol{x}$, where $\boldsymbol{x}$ is a vector of dummy variables x_a and $\boldsymbol{\Pi}^\alpha$ was derived above. The coefficient of x_a gives the probability Π_a^α that the synapse has (tilded) strength Ω_a. Since each of the N synapses updates at each storage step independently of the others, the PGF for the entire polysynaptic configuration is just

$$\mathcal{G}_{\boldsymbol{N}}^{(\alpha)}(\boldsymbol{x}) = \left(\boldsymbol{\Pi}^\alpha \cdot \boldsymbol{x}\right)^N , \tag{4.6}$$

where now the coefficient of x_a gives the probability that N_a^α synapses have (tilded) strength Ω_a. Differentiation with respect to x_a produces expectation values for N_a^α, etc., so that, for example, $\mathsf{E}[N_a^\alpha] = N\,\Pi_a^\alpha$. To convert N_a^α into v_b^α we write $v_b^\alpha = \frac{1}{N}\sum_a e_a^b\, N_a^\alpha$, so we must introduce dummy variables y_b for the PGF of $\boldsymbol{v}^\alpha$ such that differentiation with respect to y_b automatically generates the required multiplicative factor $\frac{1}{N}\, e_a^b$ and the sum over a. The required PGF for $\boldsymbol{v}^\alpha$ is therefore

$$\mathcal{G}^{(\alpha)}(\boldsymbol{y}) = \left[\sum_{a=1}^{n} \Pi_a^\alpha \left(\prod_{c=1}^{n} y_c^{\frac{1}{N}e_a^c}\right)\right]^N , \tag{4.7}$$

and the corresponding MGF is

$$\mathcal{M}^{(\alpha)}(\boldsymbol{z}) = \left[\sum_{a=1}^{n} \Pi_a^\alpha \, \exp\left(\frac{1}{N}\sum_{c=1}^{n} e_a^c\, z_c\right)\right]^N . \tag{4.8}$$

The components $\mu_a^\alpha \equiv \mathsf{E}[v_a^\alpha]$ of the mean vector $\boldsymbol{\mu}^\alpha$ and elements $[\,\mathbb{S}_\alpha\,]_{a,b} \equiv$

$\mathsf{Cov}[v_a^\alpha, v_b^\alpha]$ of the covariance matrix $\mathbb{S}_\alpha$ then follow as

$$\mu_a^\alpha = \boldsymbol{e}^a \cdot \boldsymbol{\Pi}^\alpha, \tag{4.9a}$$

$$[\,\mathbb{S}_\alpha\,]_{a,b} = \frac{1}{N}\left[\left(\boldsymbol{e}^a \circ \boldsymbol{e}^b\right) \cdot \boldsymbol{\Pi}^\alpha - \left(\boldsymbol{e}^a \cdot \boldsymbol{\Pi}^\alpha\right)\left(\boldsymbol{e}^b \cdot \boldsymbol{\Pi}^\alpha\right)\right], \tag{4.9b}$$

where the notation $\boldsymbol{x} \circ \boldsymbol{y}$ denotes a vector with components $[\boldsymbol{x} \circ \boldsymbol{y}]_a = x_a\, y_a$. Because the N_a^α are multinomially distributed with probabilities Π_a and parameter N from Eq. (4.6) and the v_a^α are just linear transforms of the N_a^α, for large enough N we can as a good approximation where necessary just take the v_a^α to be distributed as a multivariate Gaussian with the statistics in Eq. (4.9).

The components Π_a satisfy the identity in Eq. (3.28a), and the vectors $\boldsymbol{e}^a$ ($a \neq n$) are either symmetric or antisymmetric, depending on the parity of a. Hence, we may immediately deduce that since $\mu_a = \boldsymbol{e}^a \cdot \boldsymbol{\Pi}$, the means of the even dynamical variables v_{2i} ($2i \neq n$) automatically vanish at all times or memory storage steps. This result depends on the identity in Eq. (3.28a), which essentially follows directly from the assumption of symmetric, unbiased processes: any model built on such an assumption must respect the identity. The result also depends on the symmetry properties of the $\boldsymbol{e}^a$ vectors. These properies, however, are purely a consequence of our definition of the vector components in Eq. (4.3). We require only that the set of $\boldsymbol{e}^a$ vectors includes $\boldsymbol{e}^1 \equiv \boldsymbol{\Omega}$ and $\boldsymbol{e}^n \equiv \mathbf{1}$, and that the total set is orthogonal and for convenience normalised so that $e_n^a = +1$. The construction of the vectors $\boldsymbol{e}^2, \ldots, \boldsymbol{e}^{n-1}$ with these requirements leaves $\frac{1}{2}(n-2)(n-3)$ (for $n > 1$) degrees of freedom in their specification. For $n = 3$, we uniquely have $\boldsymbol{e}^2 = (+1, -2, +1)^{\mathrm{T}}$ (or up to an overall multiplicative constant without the normalisation), but for $n \geq 4$, the $\boldsymbol{e}^a$ need not in general respect the symmetry properties. The vanishing of the even dynamical means is thus essentially purely a matter of our choice in how we extend the polysynaptic variables beyond just the perceptron's activation.

The choice in Eq. (4.3) is, however, extremely convenient precisely because of the symmetry properties and the resulting vanishing of the even dynamical means.

Since Eq. (4.9a) gives the mean vector as $\boldsymbol{\mu}^{\alpha} = \mathbb{E}\,\boldsymbol{\Pi}^{\alpha}$, we can re-express the generating functions purely in terms of $\boldsymbol{\mu}^{\alpha}$. For example, the PGF is

$$\mathcal{G}^{(\alpha)}(\boldsymbol{y}) = \left\{ \sum_{a=1}^{n} \left[\mathbb{E}^{-1}\boldsymbol{\mu}^{\alpha}\right]_{a} \left(\prod_{c=1}^{n} y_{c}^{\frac{1}{N}e_{a}^{c}} \right) \right\}^{N} . \tag{4.10}$$

Because the equivalent MGF depends only on $\boldsymbol{\mu}^{\alpha}$, all higher-order moments of $\boldsymbol{v}^{\alpha}$ are therefore only functions of the first-order moment $\boldsymbol{\mu}^{\alpha}$. For $n = 2$, we have that $\boldsymbol{e}^{1} = (-1, +1)^{\mathrm{T}}$ and $\boldsymbol{e}^{2} = (+1, +1)^{\mathrm{T}}$, and since $\mu_{2}^{\alpha} = \mathsf{E}[v_{2}^{\alpha}] \equiv 1$ because $v_{2}^{\alpha} \equiv 1$, with $z_{2} = 1$ the MGF for $h^{(\alpha)} \equiv v_{1}^{\alpha}$ becomes the extremely simple

$$\mathcal{M}^{(\alpha)}(z_{1}) = \left(\cosh \frac{z_{1}}{N} + \mu_{1}^{\alpha} \sinh \frac{z_{1}}{N} \right)^{N} , \tag{4.11}$$

which is Eq. (4.11) in Elliott (2017b). We have generalised that result to those here for $n \geq 2$, although the $n = 2$ is especially elegant.

Using $\boldsymbol{\mu}^{\alpha} = \mathbb{E}\,\boldsymbol{\Pi}^{\alpha}$ we can compute the change in the mean vector between memory storage steps by writing

$$\boldsymbol{\mu}^{\alpha+1} - \boldsymbol{\mu}^{\alpha} = \mathbb{E}\,\boldsymbol{\Pi}^{\alpha+1} - \boldsymbol{\mu}^{\alpha} = \mathbb{E}\,\mathbb{W}_{\alpha+1}\boldsymbol{\Pi}^{\alpha} - \boldsymbol{\mu}^{\alpha}, \tag{4.12}$$

or

$$\boldsymbol{\mu}^{\alpha+1} - \boldsymbol{\mu}^{\alpha} = \mathbb{E}\,(\,\mathbb{W}_{\alpha+1} - \mathbb{I}\,)\,\mathbb{E}^{-1}\boldsymbol{\mu}^{\alpha}. \tag{4.13}$$

Because of the difference between $p_{a}^{(\alpha)\pm}$ and $p_{a}^{\pm}(t)$ discussed above, we cannot immediately Poisson sum the right-hand side (RHS) of Eq. (4.13). We must instead Poisson sum $\mathbb{W}_{\alpha+1}\boldsymbol{\Pi}^{\alpha}$ on the RHS of Eq. (4.12) by the same arguments leading up to Eq. (3.13), using $f_{a}^{(\alpha)\pm}$ instead of $p_{a}^{(\alpha)\pm}$. We obtain $\mathbb{W}(t)\,\boldsymbol{\Pi}(t) =$

$\mathbb{W}(t)\,\mathbb{E}^{-1}\boldsymbol{\mu}(t)$, and hence

$$\frac{1}{r}\frac{d}{dt}\,\boldsymbol{\mu}(t) = \mathbb{E}\big[\,\mathbb{W}(t) - \mathbb{I}\,\big]\mathbb{E}^{-1}\boldsymbol{\mu}(t). \tag{4.14}$$

Eqs. (4.13) and (4.14) are the generalisations to $n \geq 2$ of the equivalent Eqs. (5.2) and (5.6) given in Elliott (2017b) for the specifically $n = 2$ case. In both Eq. (4.13) and Eq. (4.14), the underlying monosynaptic transition matrix $\mathbb{W}$ is transformed via the action of the matrices $\mathbb{E}$ and $\mathbb{E}^{-1}$ on its left and right, respectively. This action corresponds to the same transformation of variables from the monosynaptic state probability vector $\boldsymbol{\Pi}$ to the polysynaptic mean vector $\boldsymbol{\mu}$, since $\boldsymbol{\mu} = \mathbb{E}\,\boldsymbol{\Pi}$. The evolution equation for $\boldsymbol{\mu}(t)$ in Eq. (4.14) therefore also follows automatically from the Master equation for $\boldsymbol{\Pi}(t)$ in Eq. (3.13) under this transformation. Since the formal solution of Eq. (3.13) is $\boldsymbol{\Pi}(t) = \overline{\mathbb{U}}(t)\boldsymbol{\Pi}(0)$, we can then write the formal solution of Eq. (4.14) as $\boldsymbol{\mu}(t) = \mathbb{E}\,\overline{\mathbb{U}}(t)\mathbb{E}^{-1}\boldsymbol{\mu}(0)$, with a transformed $\overline{\mathbb{U}}(t)$ matrix.

Since $[\,\mathbb{E}\,]_{n,a} = \boldsymbol{e}^n_a = \mathbf{1}_a \equiv 1$, the change in μ_n per storage step or in time is just $\mathbf{1}\cdot[\,\mathbb{W} - \mathbb{I}\,]\,\mathbb{E}^{-1}\boldsymbol{\mu}$. As $\mathbb{W}$ is a stochastic matrix, this change vanishes, as it should: $v_n \equiv 1$, hence $\mu_n \equiv 1$ always. Taking out this fixed $\mu_n(t) = 1$ term on the RHS of Eq. (4.14), we may write Eq. (4.14) component-wise as

$$\frac{1}{r}\frac{d}{dt}\,\mu_a(t) = \frac{1}{n}\left\{\boldsymbol{e}^a\cdot\big[\,\mathbb{W}(t) - \mathbb{I}\,\big]\,\mathbf{1}\right\} + \sum_{b=1}^{n-1}\left\{\boldsymbol{e}^a\cdot\big[\,\mathbb{W}(t) - \mathbb{I}\,\big]\,\boldsymbol{e}^b\right\}\beta_b\,\mu_b(t), \tag{4.15}$$

where $\beta_b = 1/|\boldsymbol{e}^b|^2$. Provided that $\mathbb{W}(t)$ is not symmetric (in which case $\mathbf{1}$ would be a null right eigenvector of $\mathbb{W}(t) - \mathbb{I}$), the RHS of this equation contains an inhomogeneous term that drives the evolution of the $\mu_a(t)$ for $a \neq n$ in addition to the homogeneous terms. In particular, the inhomogeneous term is responsible for driving the initial rise in the mean perceptron activation, as we discuss later.

We defer the discussion of the evolution of the covariance matrix, whose

elements are given in Eq. (4.9b), until section 4.3. We note however that if we set either $a = n$ or $b = n$, then the RHS of Eq. (4.9b) correctly vanishes, because $\mathbf{1} \cdot \mathbf{\Pi} \equiv 1$ and $\mathbf{1} \circ \boldsymbol{e} \equiv \boldsymbol{e}$ for any vector $\boldsymbol{e}$.

The "variable" $v_n \equiv 1$ is not dynamical but just encodes the constraint $\mathbf{1} \cdot \boldsymbol{N} \equiv N$. Its corresponding dummy variables in Eqs. (4.7) and (4.8) can therefore be set to $y_n = 1$ or $z_n = 0$, respectively, without any loss of generality. Further, setting say any $y_a = 1$ corresponds to marginalising or averaging over the associated dynamical variable v_a. Thus, we can reduce these functions to the marginal generating functions for any single v_a variable by setting all other dummy variables to unity or zero. In particular, the marginal PGF for the perceptron's activation is then just

$$\mathcal{G}^{(\alpha)}(y_1) = \left(\sum_{a=1}^{n} \Pi_a^\alpha \, y_1^{\frac{1}{N}\Omega_a} \right)^N = y_1^{-1} \left(\sum_{a=1}^{n} \Pi_a^\alpha \, y_1^{\frac{2}{N}\frac{a-1}{n-1}} \right)^N . \tag{4.16}$$

This PGF is a function of y_1 with powers ranging between -1 and $+1$ with spacing $2/[N(n-1)]$, representing the resolution of the perceptron's activation discussed above. Here, these different powers of y_1 merely index different possible percepton activations, so we can instead use

$$y^{N(n-1)/2} \, \mathcal{G}^{(\alpha)}\big(y^{N(n-1)/2}\big) \equiv \left(\sum_{a=1}^{n} \Pi_a^\alpha \, y^{a-1} \right)^N \tag{4.17}$$

as the PGF for the scaled perceptron's activation, where the powers now run from 0 to $N(n-1)$ in unit steps. A synapse of strength Ω_a contributes a term y^{a-1} weighted by the probability Π_a^α of that strength, with N independently-updating synapses contributing N such monosynaptic PGFs to the overall product. It is remarkable that such a simple PGF fully captures the full complexity of the perceptron's activation, given that it represents a polysynaptic function. This simplicity is due to the use of the symmetric processes considered here. For the marginal MGF, we can use the antisymmetry of $\mathbf{\Omega}$ and the

identity in Eq. (3.28a) to write

$$\mathcal{M}^{(\alpha)}(z_1) = \left\{ \sum_{a=1}^{\lfloor \frac{n+1}{2} \rfloor} \left[\frac{2}{n} \cosh \frac{\Omega_a z_1}{N} + \left(2\,\Pi_a^\alpha - \frac{2}{n} \right) \sinh \frac{\Omega_a z_1}{N} \right] \right\}^N , \qquad (4.18)$$

where the sum could also run from $a = 1 + \lfloor n/2 \rfloor$ to $a = n$. For $n = 2$, this reduces to Eq. (4.11) since $\Omega_1 = -1$ and $1 - 2\,\Pi_1^\alpha \equiv \mu_1^\alpha$. Eq. (4.18) is therefore perhaps the most natural generalisation of Eq. (4.11) to the $n \geq 2$ marginal MGF for $h^{(\alpha)}$.

To obtain results in continuous time, we must Poisson sum Eq. (4.7) or Eq. (4.8) and then take derivatives to obtain the moments. This does not affect the means, but the Poisson sum introduces terms in the continuous-time covariance matrix that are not present in the discrete-time form (Elliott & Lagogiannis, 2012). In particular, although synapses update independently, driving their updates synchronously via a Poisson process correlates their strengths, while in discrete time, these correlations are absent. Dealing specifically with the marginal distribution for $v_1 \equiv h$, using Eqs. (4.8) and (4.9) we obtain

$$\begin{aligned} \mathsf{Var}\big[h(t)\big] = \frac{1}{N} \left[\frac{1}{n} |\, \mathbf{\Omega} \,|^2 - \mu_1(t)^2 \right] \\ + \left(1 - \frac{1}{N} \right) \left\{ \sum_{\alpha=0}^{\infty} e^{-rt} \frac{(rt)^\alpha}{\alpha!} \left(\mu_1^\alpha \right)^2 - \mu_1(t)^2 \right\} , \end{aligned} \qquad (4.19)$$

for the variance in the perceptron's continuous-time activation, where we have used $(\mathbf{\Omega} \circ \mathbf{\Omega}) \cdot \mathbf{\Pi}(t) = \frac{1}{n} |\, \mathbf{\Omega} \,|^2$, which follows from the vector $\mathbf{\Omega} \circ \mathbf{\Omega}$ being symmetric and the components of $\mathbf{\Pi}(t)$ satisfying the identity in Eq. (3.28a). The second term on the RHS is absent in the discrete-time form. The term in curly brackets represents the covariance in strengths between any pair of synapses, and is an irreducible contribution to $\mathsf{Var}[h(t)]$ in the limit $N \to \infty$. Using the result for μ_1^α implied by the expression for $\mu_1(t)$ in Eq. (3.32), we can directly evaluate the remaining Poisson sum in Eq. (4.19), although the

final form is extremely messy. Previously, we have been unable to obtain exact, closed-form results for $\mathsf{Var}[h(t)]$ except for small values of Θ for bistate synapses (Elliott & Lagogiannis, 2012). By directly integrating out the internal filter states, the TDSU formulation of the dynamics has permitted general calculations that would otherwise be extremely difficult or impossible.

The entire continuous-time marginal distribution for $h(t)$ is given by the PGF

$$\mathcal{G}(y_1;t) = \sum_{\alpha=0}^{\infty} e^{-rt}\frac{(rt)^\alpha}{\alpha!}\left(\sum_{a=1}^{n}\Pi_a^\alpha\, y_1^{\frac{1}{N}\Omega_a}\right)^N. \tag{4.20}$$

In principle we can perform a multinomial expansion and then use the results for Π_a^α to evaluate the Poisson sum. In practice we obtain only unusable formal results. Nevertheless, we can use Eq. (4.20) to obtain the explicit marginal probability distribution for $h(t)$ at any time t. In Fig. 4 we give examples of the evolution of this distribution for different choices of n and Θ, for $N = 10^3$ synapses. We show superimposed heat and contour maps, representing the probability for any allowed value of h at time t. Since N is reasonably large, the allowed values of $h(t)$ appear continuous rather than discrete in this figure. For $n = 2$ (top row), we have shown Fig. 4A ($\Theta = 2$) and a version of Fig. 4B ($\Theta = 4$) before for the $n = 2$ TDSU (Elliott, 2016a), but our present work allows us to show results for any $n \geq 2$. The shift in the distribution of $h(t)$ towards higher values for increasing $t < \tau_{\text{peak}}$ is clear. For $\Theta = 2$ (left column), the first Poisson mode corresponding to the storage of $\boldsymbol{\xi}^1$ can be clearly seen as a separate branch of higher probability, and for the particular case of $n = 2$ (Fig. 4A) even the second Poisson mode for $\boldsymbol{\xi}^2$ can be seen. Other early Poisson modes cannot be distinguished as their contributions merge together. These early Poisson modes lead to the rising mean memory signal. Later Poisson modes, corresponding to the storage of later memories, lead to the return of the system to equilibrium for $t > \tau_{\text{peak}}$.

As commented earlier, the distribution of the discrete v_a^α variables at each

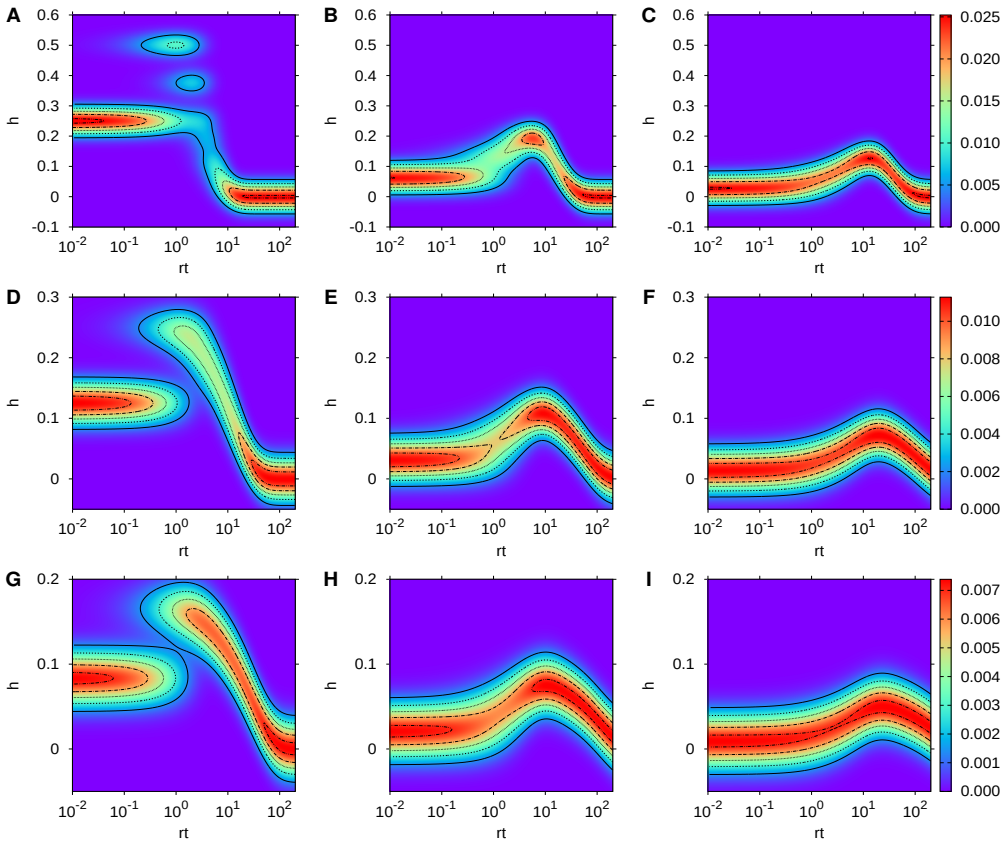


Figure 4: Superimposed heat and contour maps of the probability distribution $P(h;t)$ of the tracked memory signal h at time t for $N = 10^3$ synapses, for various choices of Θ and n. In the top row, $n = 2$; in the middle row, $n = 4$; in the bottom row, $n = 6$. In the left column, $\Theta = 2$; in the middle column, $\Theta = 4$; in the right column, $\Theta = 6$. Each row has its own, common colour scale shown to the right of each row. In the first row, 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. In the second row, the contours correspond to probabilities of 0.002, 0.004, 0.006, 0.008 and 0.010. In the third row, they correspond to 0.0014, 0.0028, 0.0042, 0.0056 and 0.0070.

memory storage step can to a good approximation be taken as a multivariate Gaussian with the statistics given in Eq. (4.9). Clearly, however, at least for $\Theta = 2$ (left column of Fig. 4), the distribution of $h(t)$ for intermediate times is not Gaussian, but rather is bimodal or even trimodal. For larger Θ, the distributions are unimodal and well approximated by Gaussians with time-dependent means and variances, but the manifestly non-Gaussian distributions for $\Theta = 2$ throw into sharp relief the radical difference between the discrete-time and continuous-time formulations of the dynamics of memory storage.

Since $\mathbf{\Pi}(0) = \frac{1}{n}(\mathbf{1} + \frac{1}{\Theta^2}\boldsymbol{w})$, we have that $\mu_1(0) = \frac{2}{n}\frac{1}{\Theta^2}$, so for fixed Θ, these is an overall shifting down of the distribution towards smaller $h(t)$ as n increases. This is reflected in the change in the range of the ordinates in the three rows of Fig. 4. We note also the change in the range of the colour scales for these rows, but this change reflects the fact that as n increases, the number of distinct values $N(n-1)+1$ that $h(t)$ can take increases. Thus, although the distribution for $h(t)$ shifts towards smaller values of $h(t)$ and so is more focused as n increases, the distribution is spread out over more discrete points in this narrower range, so the probability of any particular value of $h(t)$ decreases as n increases.

4.2 Generating Functions for Polysynaptic Transitions

The generating functions above give the probabilities or moments of entire polysynaptic configurations at each memory storage step or for all time. Because they are constructed using the probabilities Π_a^α, which were calculated in section 3.2, it has not been necessary to explicitly examine the transitions in polysynaptic configurations at each step. For computing FPTs, however, we must constrain these transitions by excluding transitions to configurations in which $h^{(\alpha)} \leq \vartheta$. We therefore now consider generating functions for the transitions in polysynaptic configurations that occur during a single memory

storage step. Constraints or boundary conditions can then be applied to these transitions to exclude those that are forbidden.

Using the results in Elliott (2019) adapted to the step-dependent matrices $\mathbb{W}_\alpha$, we can write down the conditional PGF for the probability distribution of $\boldsymbol{N}^{\alpha+1}$, conditioned on a definite configuration $\boldsymbol{N}^\alpha$. This distribution completely specifies how the entire (tilded) polysynaptic configuration changes during the storage of memory $\boldsymbol{\xi}^\alpha$, given a definite configuration $\boldsymbol{N}^\alpha$. From Eq. (3.5) of Elliott (2019), we have

$$\mathcal{G}_{\boldsymbol{N}}\big(\boldsymbol{x} \,|\, \boldsymbol{N}^\alpha\big) = \prod_{b=1}^{n} \left(\sum_{a=1}^{n} x_a \left[\mathbb{W}_{\alpha+1}\right]_{a,b} \right)^{N_b^\alpha} . \tag{4.21}$$

If we remove the conditioning by summing over $\mathcal{G}_{\boldsymbol{N}}\big(\boldsymbol{x} \,|\, \boldsymbol{N}^\alpha\big)$ weighted by the multinomial probability distribution of $\boldsymbol{N}^\alpha$, then we obtain the unconditional PGF $\mathcal{G}_{\boldsymbol{N}}^{(\alpha+1)}(\boldsymbol{x})$ by using the fact that $\boldsymbol{\Pi}^{\alpha+1} = \mathbb{W}_{\alpha+1}\boldsymbol{\Pi}^\alpha$. The conditional PGF in Eq. (4.21) expresses the fact that a synapse in strength state b at step α will undergo a transition governed by the matrix $\mathbb{W}_{\alpha+1}$ during the storage of $\boldsymbol{\xi}^{\alpha+1}$, leading to strength state a with probability $\left[\mathbb{W}_{\alpha+1}\right]_{a,b}$, with PGF $\sum_{a=1}^{n} x_a \left[\mathbb{W}_{\alpha+1}\right]_{a,b}$ for this monosynaptic transition (see Elliott (2019) for a more careful discussion of tilded versus untilded strengths in this argument). The polysynaptic transition PGF as a product of N of these individual, monosynaptic PGFs follows by independence. In terms of the $\boldsymbol{v}$ variables, we then have the PGF

$$\mathcal{G}\big(\boldsymbol{y} \,|\, \boldsymbol{v}^\alpha\big) = \prod_{b=1}^{n} \left[\sum_{a=1}^{n} \left(\prod_{c=1}^{n} y_c^{\frac{1}{N} e_a^c} \right) \left[\mathbb{W}_{\alpha+1}\right]_{a,b} \right]^{N\left[\mathbb{E}^{-1}\boldsymbol{v}^\alpha\right]_b} , \tag{4.22}$$

where the powers in this product are just $N\left[\mathbb{E}^{-1}\boldsymbol{v}^\alpha\right]_b \equiv N_b^\alpha$, corresponding to the same powers in $\mathcal{G}_{\boldsymbol{N}}\big(\boldsymbol{x} \,|\, \boldsymbol{N}^\alpha\big)$, but written in the terms of $\boldsymbol{v}^\alpha$. This PGF encodes the entire set of polysynaptic transition probabilities, $\mathsf{Prob}[\, \boldsymbol{v}^{\alpha+1} \,|\, \boldsymbol{v}^\alpha \,]$.

Since $v_n \equiv 1$, for $n = 2$ the transition probabilities $\mathsf{Prob}[\,\boldsymbol{v}^{\alpha+1} \,|\, \boldsymbol{v}^{\alpha}\,]$ are completely equivalent to $\mathsf{Prob}[\,h^{(\alpha+1)} \,|\, h^{(\alpha)}\,]$, and hence they define an $(N+1)\times(N+1)$ transition matrix. For $n > 2$, the transition probabilities $\mathsf{Prob}[\,\boldsymbol{v}^{\alpha+1} \,|\, \boldsymbol{v}^{\alpha}\,]$ most naturally define a "transition tensor", but the transitions can be encoded into an ${}^{N+n-1}C_{n-1} \times {}^{N+n-1}C_{n-1} \sim \mathcal{O}\,(N^{n-1}/(n-1)!) \times \mathcal{O}\,(N^{n-1}/(n-1)!)$ matrix. For FPTs in the percepton's activation, however, we are not concerned with the full set of transitions $\mathsf{Prob}[\,\boldsymbol{v}^{\alpha+1} \,|\, \boldsymbol{v}^{\alpha}\,]$ but only $\mathsf{Prob}[\,v_1^{\alpha+1} \,|\, v_1^{\alpha}\,]$. We can obtain $\mathsf{Prob}[\,v_1^{\alpha+1} \,|\, \boldsymbol{v}^{\alpha}\,]$ directly from $\mathsf{Prob}[\,\boldsymbol{v}^{\alpha+1} \,|\, \boldsymbol{v}^{\alpha}\,]$ just by marginalising over the other $v_a^{\alpha+1}$ variables, i.e. by setting $y_a = 1$ for $a \neq 1$ in Eq. (4.22). To obtain $\mathsf{Prob}[\,v_1^{\alpha+1} \,|\, v_1^{\alpha}\,]$, we write

$$\begin{aligned}\mathsf{Prob}[\,v_1^{\alpha+1} \,|\, v_1^{\alpha}\,] &= \frac{\mathsf{Prob}[\,v_1^{\alpha+1} \,\&\, v_1^{\alpha}\,]}{\mathsf{Prob}[\,v_1^{\alpha}\,]} = \sum_{\boldsymbol{v}^{\alpha}\{v_1^{\alpha}\}} \frac{\mathsf{Prob}[\,v_1^{\alpha+1} \,\&\, \boldsymbol{v}^{\alpha}\{v_1^{\alpha}\}\,]}{\mathsf{Prob}[\,v_1^{\alpha}\,]} \\ &= \sum_{\boldsymbol{v}^{\boldsymbol{\alpha}}\{v_1^{\alpha}\}} \mathsf{Prob}[\,v_1^{\alpha+1} \,|\, \boldsymbol{v}^{\alpha}\{v_1^{\alpha}\}\,] \, \frac{\mathsf{Prob}[\,\boldsymbol{v}^{\alpha}\{v_1^{\alpha}\}\,]}{\mathsf{Prob}[\,v_1^{\alpha}\,]}, \qquad (4.23)\end{aligned}$$

where we use the notation $\boldsymbol{v}^{\alpha}\{v_1^{\alpha}\}$ to denote a configuration with a definite value of v_1^{α}, and so the sum is over all such configurations with this definite value. The ratio $\mathsf{Prob}[\,\boldsymbol{v}^{\alpha}\{v_1^{\alpha}\}\,]/\mathsf{Prob}[\,v_1^{\alpha}\,]$ is just the probability of any particular configuration with definite activation v_1^{α} relative to the total probability of all such configurations. All the quantities on the RHS of Eq. (4.23) can be extracted, if somewhat laboriously, from $\mathcal{G}^{(\alpha)}\big(\boldsymbol{y}\big)$ in Eq. (4.7) and $\mathcal{G}\big(\boldsymbol{y} \,|\, \boldsymbol{v}^{\alpha}\big)$ in Eq. (4.22), and so we can in principle explicitly compute these elements of the $[N(n-1)+1] \times N[(n-1)+1]$ transition matrix in the perceptron's activation for $n \geq 2$.

For $n = 2$, the ratio $\mathsf{Prob}[\,\boldsymbol{v}^{\alpha}\{v_1^{\alpha}\}\,]/\mathsf{Prob}[\,v_1^{\alpha}\,]$ in Eq. (4.23) is identically equal to unity because there is only one configuration $\boldsymbol{v}^{\alpha}\{v_1^{\alpha}\}$ with perceptron activation v_1^{α}; indeed, in this case Eq. (4.23) is a trivial identity. For $n = 2$, we obtain the transition matrix $\mathsf{Prob}[\,h^{(\alpha+1)} \,|\, h^{(\alpha)}\,]$ directly from the generating function $\mathcal{G}\big(\boldsymbol{y} \,|\, \boldsymbol{v}^{\alpha}\big)$. To obtain FPTs for $n = 2$, we used a projection

matrix in conjunction with this h transition matrix to remove transitions to states with $h^{(\alpha+1)} \leq \vartheta$ (Elliott, 2017b). Given that we have just obtained, at least in principle, the h transition matrix for $n > 2$, it would appear that we could also project out forbidden h transitions and obtain FPTs more generally. However, Eq. (4.23) involves the configuration probabilities at step α via the ratio $\mathsf{Prob}[\,\boldsymbol{v}^{\alpha}\{v_1^{\alpha}\}\,]/\mathsf{Prob}[\,v_1^{\alpha}\,]$. Although we appear to have these via $\mathcal{G}^{(\alpha)}\big(\boldsymbol{y}\big)$, critically these probabilities were computed by allowing for all possible transitions. But for Eq. (4.23) to apply to a calculation in which we project out forbidden transitions, then $\mathsf{Prob}[\,\boldsymbol{v}^{\alpha}\{v_1^{\alpha}\}\,]/\mathsf{Prob}[\,v_1^{\alpha}\,]$ must also be calculated by disallowing forbidden transitions. It is therefore ironic that in obtaining an h transition matrix to use with a projection matrix to remove forbidden transitions, for $n > 2$ this h transition matrix must itself be obtained by computing state probabilities in which forbidden transitions have already been excluded. The method of projecting out forbidden transitions from an h transition matrix is therefore only useful when the h transition matrix is "stateless", and this is true only for $n = 2$. Thus, we must use other methods to obtain FPTs for $n > 2$.

4.3 Fokker-Planck Equations

The transition probabilities $\mathsf{Prob}[\,\boldsymbol{v}^{\alpha+1}\,|\,\boldsymbol{v}^{\alpha}\,]$ and the equivalent ${}^{N+n-1}C_{n-1} \times {}^{N+n-1}C_{n-1}$ transition matrix *are* stateless and could be used to obtain FPTs for $n > 2$. However, the polynomial growth of ${}^{N+n-1}C_{n-1}$ with N severely restricts the largest values of N for which we could reasonably obtain FPTs. Even for $n = 2$, we were restricted to an upper value of N of around 10^4, mainly because of long convergence times (Elliott, 2017b).

One way to avoid time-consuming and numerically difficult calculations with large matrices is to move to the continuum limit and formulate the FPT problem via a Fokker-Planck equation with an absorbing boundary (van Kam-

pen, 1992). We therefore need the jump moments for the memory storage process. The conditional MGF for transitions in the $\boldsymbol{v}$ variables corresponding to the conditional PGF in Eq. (4.22) is

$$\mathcal{M}\big(\boldsymbol{z}\,|\,\boldsymbol{v}^{\alpha}\big) = \prod_{b=1}^{n}\left\{\sum_{a=1}^{n}\exp\left(\frac{1}{N}\sum_{c=1}^{n} z_c\, e_a^c\right)\big[\mathbb{W}_{\alpha+1}\big]_{a,b}\right\}^{N\left[\mathbb{E}^{-1}\boldsymbol{v}^{\alpha}\right]_b}, \tag{4.24}$$

which is Eq. (3.8) in Elliott (2019) generalised to $\mathbb{W}_{\alpha+1}$. We then obtain

$$\mathsf{E}\big[\big(\boldsymbol{v}^{\alpha+1}-\boldsymbol{v}^{\alpha}\big)|\,\boldsymbol{v}^{\alpha}\big] = \mathbb{E}\left(\mathbb{W}_{\alpha+1}-\mathbb{I}\right)\mathbb{E}^{-1}\,\boldsymbol{v}^{\alpha}, \tag{4.25a}$$

$$\begin{aligned}\mathsf{E}\big[\big(v_a^{\alpha+1}-v_a^{\alpha}\big)\big(v_b^{\alpha+1}-v_b^{\alpha}\big)|\,\boldsymbol{v}^{\alpha}\big] &= \mathsf{E}\big[\big(v_a^{\alpha+1}-v_a^{\alpha}\big)|\,\boldsymbol{v}^{\alpha}\big]\,\mathsf{E}\big[\big(v_b^{\alpha+1}-v_b^{\alpha}\big)|\,\boldsymbol{v}^{\alpha}\big]\\ &+\frac{1}{N}\Big\{\big(\boldsymbol{e}^a\circ\boldsymbol{e}^b\big)^{\mathrm{T}}\,\mathbb{W}_{\alpha+1}-\big[\big(\mathbb{W}_{\alpha+1}^{\mathrm{T}}\boldsymbol{e}^a\big)\circ\big(\mathbb{W}_{\alpha+1}^{\mathrm{T}}\boldsymbol{e}^b\big)\big]^{\mathrm{T}}\Big\}\,\mathbb{E}^{-1}\boldsymbol{v}^{\alpha},\end{aligned} \tag{4.25b}$$

describing the expected changes in the first- and second-order moments during the storage of $\boldsymbol{\xi}^{\alpha+1}$, conditioned on the definite state $\boldsymbol{v}^{\alpha}$. Again, these are just Eq. (3.9) of Elliott (2019) suitably generalised. In obtaining these equations, we have kept the vectors $\boldsymbol{e}^a$ general, in the sense that we have not used the fact that $\boldsymbol{e}^n \equiv \mathbf{1}$, so they are valid for any (non-singular) transformation matrix $\mathbb{E}$. In a standard Fokker-Planck equation, if the matrices $\mathbb{W}_{\alpha}$ were in fact constants independent of α, then Eq. (4.25) would give the jump moments (van Kampen, 1992). Eq. (4.25a) would give the drift vector and Eq. (4.25b) the elements of the diffusion matrix. We define the drift matrix

$$\mathbb{A}_{\alpha} = \mathbb{E}\left(\mathbb{W}_{\alpha}-\mathbb{I}\right)\mathbb{E}^{-1}, \tag{4.26}$$

so that the drift vector in Eq. (4.25a) is $\mathbb{A}_{\alpha+1}\boldsymbol{v}^{\alpha}$. This drift matrix appears in Eq. (4.13), operating on $\boldsymbol{\mu}^{\alpha}$ and giving the unconditional change $\boldsymbol{\mu}^{\alpha+1}-\boldsymbol{\mu}^{\alpha}$. We write the $\boldsymbol{v}$-dependent matrix $\mathbb{B}_{\alpha}(\boldsymbol{v})$ as

$$\mathbb{B}_{\alpha}(\boldsymbol{v}) = \left(\mathbb{A}_{\alpha}\boldsymbol{v}\right)\left(\mathbb{A}_{\alpha}\boldsymbol{v}\right)^{\mathrm{T}} + \overline{\mathbb{B}}_{\alpha}(\boldsymbol{v}), \tag{4.27}$$

where the associated matrix $\overline{\mathbb{B}}_\alpha(\boldsymbol{v})$ has elements

$$\big[\,\overline{\mathbb{B}}_\alpha(\boldsymbol{v})\,\big]_{a,b} = \frac{1}{N}\Big\{\big(\boldsymbol{e}^a\circ\boldsymbol{e}^b\big)^{\mathrm{T}}\,\mathbb{W}_\alpha - \big[\big(\mathbb{W}_\alpha^{\mathrm{T}}\boldsymbol{e}^a\big)\circ\big(\mathbb{W}_\alpha^{\mathrm{T}}\boldsymbol{e}^b\big)\big]^{\mathrm{T}}\Big\}\,\mathbb{E}^{-1}\boldsymbol{v}, \tag{4.28}$$

so that the diffusion matrix with elements given in Eq. (4.25b) is $\mathbb{B}_{\alpha+1}(\boldsymbol{v}^\alpha)$. We note that with $\boldsymbol{e}^n \equiv \mathbf{1}$, $\big[\,\overline{\mathbb{B}}_\alpha(\boldsymbol{v})\,\big]_{n,b} = \big[\,\overline{\mathbb{B}}_\alpha(\boldsymbol{v})\,\big]_{a,n} \equiv 0$ because $\mathbf{1}\circ\boldsymbol{e}\equiv\boldsymbol{e}$ and $\mathbf{1}$ is a left eigenvector of the stochastic matrix $\mathbb{W}_\alpha$ with unit eigenvalue.

The standard Fokker-Planck equation is written in continuous time. For our memory storage problem, the underlying process occurs at discrete time points corresponding to the storage of each memory, where these time points are distributed according to a Poisson process. We must therefore write down a discrete memory-storage-step Fokker Planck equation and then move to continuous time by Poisson summation (Elliott, 2017b). We first write down a Fokker-Planck equation for the conditional probability distribution $P_{\alpha+1}(\boldsymbol{v}\,|\,\boldsymbol{v}^\alpha)$, describing the distribution of $\boldsymbol{v}$ at storage step $\alpha+1$, conditioned on the definite state $\boldsymbol{v}^\alpha$ at step α. This function is defined so that, for example,

$$\int d\boldsymbol{v}\,\boldsymbol{v}\,P_{\alpha+1}(\boldsymbol{v}\,|\,\boldsymbol{v}^\alpha) \equiv \mathsf{E}[\boldsymbol{v}^{\alpha+1}\,|\,\boldsymbol{v}^\alpha], \tag{4.29}$$

the integral being over the whole of the n-dimensional space. Because $\boldsymbol{v}^\alpha$ is a definite state, its probability distribution is $\delta(\boldsymbol{v}-\boldsymbol{v}^\alpha)$, where $\delta(\cdot)$ is in general a multi-dimensional Dirac delta function. The required conditional Fokker-Planck equation for this discrete process is then

$$\begin{aligned} &P_{\alpha+1}(\boldsymbol{v}\,|\,\boldsymbol{v}^\alpha) - \delta(\boldsymbol{v}-\boldsymbol{v}^\alpha) \\ &\quad= -\sum_{a=1}^{n}\frac{\partial}{\partial v_a}\Big\{\big[\mathbb{A}_{\alpha+1}\boldsymbol{v}\big]_a\,\delta(\boldsymbol{v}-\boldsymbol{v}^\alpha)\Big\} \\ &\qquad+\frac{1}{2}\sum_{a,b=1}^{n}\frac{\partial^2}{\partial v_a\partial v_b}\Big\{\big[\mathbb{B}_{\alpha+1}(\boldsymbol{v})\big]_{a,b}\,\delta(\boldsymbol{v}-\boldsymbol{v}^\alpha)\Big\}. \end{aligned} \tag{4.30}$$

For the moment we keep $\mathbb{E}$ and therefore $\boldsymbol{e}^n$ general, so that v_n is not necessarily constant and derivatives with respect to it make sense. We may see that this is the correct equation by confirming that the jump moments induced by it agree with those in Eq. (4.25). For example, multiplying by $\boldsymbol{v}$ and integrating with respect to it produces precisely Eq. (4.25a). To remove the conditioning on the definite state $\boldsymbol{v}^\alpha$ in Eq. (4.30), we require the unconditional distribution at step α, which we write as $P_\alpha(\boldsymbol{v})$. Thus, $\boldsymbol{v}^\alpha$ in Eq. (4.30) is distributed as $P_\alpha(\boldsymbol{v}^\alpha)$, and so multiplying Eq. (4.30) by $P_\alpha(\boldsymbol{v}^\alpha)$ and integrating with respect to $\boldsymbol{v}^\alpha$, we obtain the unconditional equation

$$P_{\alpha+1}(\boldsymbol{v}) - P_\alpha(\boldsymbol{v}) = -\sum_{a=1}^{n} \frac{\partial}{\partial v_a}\Big\{ \big[\mathbb{A}_{\alpha+1}\boldsymbol{v}\big]_a P_\alpha(\boldsymbol{v})\Big\} + \frac{1}{2}\sum_{a,b=1}^{n} \frac{\partial^2}{\partial v_a \partial v_b}\Big\{ \big[\mathbb{B}_{\alpha+1}(\boldsymbol{v})\big]_{a,b} P_\alpha(\boldsymbol{v})\Big\}, \tag{4.31}$$

where $\int d\boldsymbol{v}^\alpha P_{\alpha+1}(\boldsymbol{v}\,|\,\boldsymbol{v}^\alpha)\, P_\alpha(\boldsymbol{v}^\alpha) \equiv P_{\alpha+1}(\boldsymbol{v})$ by definition of the conditional probability.

Eq. (4.13) follows by using Eq. (4.31) to obtain the change in the first-order moment. Writing $\mathbb{J}_\alpha$ as the matrix of second-order moments with elements $[\,\mathbb{J}_\alpha\,]_{a,b} = \mathsf{E}[v_a^\alpha\, v_b^\alpha]$, Eq. (4.31) gives its change as

$$\mathbb{J}_{\alpha+1} - \mathbb{J}_\alpha = \mathbb{A}_{\alpha+1}\,\mathbb{J}_\alpha + \mathbb{J}_\alpha\,\mathbb{A}_{\alpha+1}^{\mathrm{T}} + \mathbb{A}_{\alpha+1}\,\mathbb{J}_\alpha\,\mathbb{A}_{\alpha+1}^{\mathrm{T}} + \overline{\mathbb{B}}_{\alpha+1}(\boldsymbol{\mu}^\alpha). \tag{4.32}$$

This equation is the generalisation to $n \geq 2$ of the equivalent Eq. (5.8b) given in Elliott (2017b) for the specifically $n = 2$ case. Since $\mathbb{J}_\alpha = \mathbb{S}_\alpha + \boldsymbol{\mu}^\alpha \boldsymbol{\mu}^{\alpha\,\mathrm{T}}$, we could have used Eq. (4.9b) instead to directly derive Eq. (4.32) without using Eq. (4.31). However, that derivation would have been considerably more messy and less transparent, and would have required re-writing terms $\mu_a^\alpha \mu_b^\alpha$ quadratic in the mean as terms involving $[\,\mathbb{J}_\alpha]_{a,b}$ linear in the second-order moments with the help of Eq. (4.9b). Eq. (4.31) therefore induces the correct changes in the

first- and second-order moments, consistent with those implied by Eq. (4.9).

With $\boldsymbol{e}^n \equiv \mathbf{1}$, the n^{th} column or row of $\mathbb{J}_\alpha$ contains the vector $\boldsymbol{\mu}^\alpha$. Eq. (4.32) must therefore reduce to Eq. (4.13) when restricted to that column or row. We already know that the corresponding entries in $\overline{\mathbb{B}}_{\alpha+1}(\boldsymbol{\mu}^\alpha)$ are zero. Clearly, the n^{th} column of $\mathbb{A}_{\alpha+1}\mathbb{J}_\alpha$ is just the vector $\mathbb{A}_{\alpha+1}\boldsymbol{\mu}^\alpha$, which is precisely the RHS of Eq. (4.13), so the remaining two terms on the RHS of Eq. (4.32), $\mathbb{J}_\alpha\,\mathbb{A}_{\alpha+1}^{\mathrm{T}}$ and $\mathbb{A}_{\alpha+1}\,\mathbb{J}_\alpha\,\mathbb{A}_{\alpha+1}^{\mathrm{T}}$, must have zeros in this column. Since $\mathbb{A}_{\alpha+1} = \mathbb{E}\,(\,\mathbb{W}_{\alpha+1} - \mathbb{I}\,)\,\mathbb{E}^{-1}$, with $\boldsymbol{e}^n \equiv \mathbf{1}$ the n^{th} row of $\mathbb{E}\,(\,\mathbb{W}_{\alpha+1} - \mathbb{I}\,)$ contains only zeros because $\mathbf{1}$ is a null left eigenvector of $\mathbb{W}_{\alpha+1} - \mathbb{I}$, so the same is true of $\mathbb{A}_{\alpha+1}$. Therefore the n^{th} column of $\mathbb{A}_{\alpha+1}^{\mathrm{T}}$ contains only zeros, and so the two matrices $\mathbb{J}_\alpha\,\mathbb{A}_{\alpha+1}^{\mathrm{T}}$ and $\mathbb{A}_{\alpha+1}\,\mathbb{J}_\alpha\,\mathbb{A}_{\alpha+1}^{\mathrm{T}}$ do indeed have zeros in that column. Hence, the n^{th} column of Eq. (4.32) does reduce to Eq. (4.13), and since $\mathbb{J}_\alpha$ is symmetric, so does its n^{th} row.

In principle we can iteratively solve Eq. (4.31) using the known initial distribution $P_0(\boldsymbol{v})$ (using the cumulants in Eq. (4.9) for $\alpha = 0$ to define a multivariate Gaussian), imposing any boundary conditions on the solution at each step, and taking enough terms in the Poisson sum $P(\boldsymbol{v};t) \equiv \sum_{\alpha=0}^{\infty} e^{-rt}\frac{(rt)^\alpha}{\alpha!}P_\alpha(\boldsymbol{v})$ to obtain the continuous-time probability distribution $P(\boldsymbol{v};t)$ to any desired accuracy. In practice, this is intractably hard. Directly Poisson summing Eq. (4.31), we have

$$\begin{aligned}\frac{1}{r}\frac{\partial}{\partial t}P(\boldsymbol{v};t) = &-\sum_{a=1}^{n}\frac{\partial}{\partial v_a}\sum_{\alpha=0}^{\infty}e^{-rt}\frac{(rt)^\alpha}{\alpha!}\Big\{\big[\mathbb{A}_{\alpha+1}\boldsymbol{v}\big]_a\,P_\alpha(\boldsymbol{v})\Big\}\\ &+\frac{1}{2}\sum_{a,b=1}^{n}\frac{\partial^2}{\partial v_a\partial v_b}\sum_{\alpha=0}^{\infty}e^{-rt}\frac{(rt)^\alpha}{\alpha!}\Big\{\big[\mathbb{B}_{\alpha+1}(\boldsymbol{v})\big]_{a,b}\,P_\alpha(\boldsymbol{v})\Big\}. \qquad (4.33)\end{aligned}$$

We know that the time-dependent matrix $\mathbb{W}(t) - \mathbb{I}$ correctly describes the dynamics of $\mathbf{\Pi}(t)$ in Eq. (3.13), and $\mathbb{A}(t) \equiv \mathbb{E}\,[\,\mathbb{W}(t) - \mathbb{I}\,]\,\mathbb{E}^{-1}$ correctly describes the dynamics of $\boldsymbol{\mu}(t)$ in Eq. (4.14). Hence, the first Poisson sum on the RHS of Eq. (4.33) can be replaced by $[\,\mathbb{A}(t)\boldsymbol{v}\,]_a\,P(\boldsymbol{v};t)$, which induces the correct

RHS in Eq. (4.14). The second Poisson sum on the RHS of Eq. (4.33) cannot, however, be evaluated in closed-form. Any such expression would be required to induce the correct contribution to the change in the continuous-time second-order moment matrix $\mathbb{J}(t)$. The question of whether we can evaluate this second Poisson sum is therefore equivalent to whether we can evaluate the Poisson sum of the RHS of Eq. (4.32), the Poisson sum of the LHS giving just $\frac{1}{r}\frac{d}{dt}\mathbb{J}(t)$. The critical step in evaluating the Poisson sum of the RHS of Eq. (4.12) for the mean dynamics was to express the RHS in terms of $f_a^{(\alpha)\pm}$. In doing so, we obtained a purely linear function in the $f^{(\alpha)}$'s, the Poisson sum of which as $f(t)$ was immediate. Since $\mathbb{J}_\alpha$ is a matrix of second moments (except for the n^{th} column and row), it is a quadratic function of $\boldsymbol{v}^\alpha$, so it is impossible to express the RHS of Eq. (4.32) as a purely linear function of $f_a^{(\alpha)\pm}$. Hence, the Poisson sum of the RHS of Eq. (4.32) cannot be evaluated in closed form, and so the second Poisson sum on the RHS of Eq. (4.33) cannot be replaced by a closed-form expression that would induce the correct contribution to the non-existent closed-form expression for the Poisson sum of the RHS of Eq. (4.32). This issue of the non-factorisation of Poisson sums of products and ratios was discussed extensively in Elliott (2017b) for bistate synapses. There, it was necessary to make the approximation that Poisson sums of products and ratios factorise, generating products and ratios of Poisson sums.

With such an approximation, Eq. (4.33) becomes a standard Fokker-Planck equation (van Kampen, 1992), taking the form

$$\begin{aligned}\frac{1}{r}\frac{\partial}{\partial t}P(\boldsymbol{v};t) = &-\sum_{a=1}^{n}\frac{\partial}{\partial v_a}\Big\{\big[\mathbb{A}(t)\boldsymbol{v}\big]_a P(\boldsymbol{v};t)\Big\} \\ &+\frac{1}{2}\sum_{a,b=1}^{n}\frac{\partial^2}{\partial v_a \partial v_b}\Big\{\big[\mathbb{B}(\boldsymbol{v};t)\big]_{a,b} P(\boldsymbol{v};t)\Big\}, \qquad (4.34)\end{aligned}$$

where $\mathbb{B}(\boldsymbol{v};t) = [\mathbb{A}(t)\boldsymbol{v}]\,[\mathbb{A}(t)\boldsymbol{v}]^{\mathrm{T}} + \overline{\mathbb{B}}(\boldsymbol{v};t)$ with

$$\big[\,\overline{\mathbb{B}}(\boldsymbol{v};t)\,\big]_{a,b} = \frac{1}{N}\Big(\big(\boldsymbol{e}^a\circ\boldsymbol{e}^b\big)^{\mathrm{T}}\,\mathbb{W}(t) - \big\{\big[\mathbb{W}(t)^{\mathrm{T}}\boldsymbol{e}^a\big]\circ\big[\mathbb{W}(t)^{\mathrm{T}}\boldsymbol{e}^b\big]\big\}^{\mathrm{T}}\Big)\,\mathbb{E}^{-1}\boldsymbol{v}. \quad (4.35)$$

The induced changes in the mean vector and covariance matrix are then

$$\frac{1}{r}\frac{d}{dt}\,\boldsymbol{\mu}(t) = \mathbb{A}(t)\boldsymbol{\mu}(t), \quad (4.14)$$

$$\frac{1}{r}\frac{d}{dt}\,\mathbb{S}(t) = \mathbb{A}(t)\mathbb{S}(t) + \mathbb{S}(t)\mathbb{A}(t)^{\mathrm{T}} + \mathbb{A}(t)\mathbb{S}(t)\mathbb{A}(t)^{\mathrm{T}} + [\mathbb{A}(t)\boldsymbol{\mu}(t)]\,[\mathbb{A}(t)\boldsymbol{\mu}(t)]^{\mathrm{T}} + \overline{\mathbb{B}}(\boldsymbol{\mu}(t);t), \quad (4.36)$$

where we have reproduced the exact Eq. (4.14) for completeness. Eq. (4.34) with the associated cumulant Eqs. (4.14) and (4.36) generalise the results in both Elliott (2017b) and Elliott (2019) from only the $n = 2$ TDSU and only the $n \geq 2$ SU, respectively, to the general $n \geq 2$ TDSU. With $p_a^{\pm}(t)$ constants, the matrix $\mathbb{W}(t)$ is constant, so with $p_a^{\pm}(t) \to p$, Eq. (4.34) is exact (modulo the usual approximations in the Fokker-Planck approach), not requiring the approximation that Poisson sums of products and ratios factorise.

With $\boldsymbol{e}^n \equiv \mathbf{1}$, derivatives with respect to $v_n \equiv 1$ are meaningless. Restricting to the specific choice $\boldsymbol{e}^n \equiv \mathbf{1}$, we work with the dynamical variables $v_1, \ldots, v_{n-1}$ and so write $\boldsymbol{v}^* = (v_1, \ldots, v_{n-1})^{\mathrm{T}}$ and $\boldsymbol{v} = \big(\text{—}\,\boldsymbol{v}^{*\mathrm{T}}\,\text{—}\,|\,1\big)^{\mathrm{T}}$. We also use an asterisk to denote the upper $(n-1)\times(n-1)$ submatrix of an $n \times n$ matrix, and for $\mathbb{A}$ we also write its n^{th} column as $\big(\text{—}\,\boldsymbol{\mathcal{A}}^{*\mathrm{T}}\,\text{—}\,|\,0\big)^{\mathrm{T}}$ so that, schematically,

$$\mathbb{A} = \left(\begin{array}{c|c} \mathbb{A}^* & \boldsymbol{\mathcal{A}}^* \\ \hline \mathbf{0}^{\mathrm{T}} & 0 \end{array}\right), \quad (4.37)$$

where $\mathbf{0}$ is a vector of zeros and

$$\mathcal{A}_a^* = \frac{1}{n}\,\boldsymbol{e}^a \cdot (\,\mathbb{W} - \mathbb{I}\,)\,\mathbf{1}. \quad (4.38)$$

We then write the Fokker-Planck equation in Eq. (4.34) as

$$\frac{1}{r}\frac{\partial}{\partial t}P(\boldsymbol{v}^*;t) = -\sum_{a=1}^{n-1}\frac{\partial}{\partial v_a^*}\Big\{\big[\boldsymbol{\mathcal{A}}^*(t) + \mathbb{A}^*(t)\boldsymbol{v}^*\big]_a P(\boldsymbol{v}^*;t)\Big\} + \frac{1}{2}\sum_{a,b=1}^{n-1}\frac{\partial^2}{\partial v_a^*\partial v_b^*}\Big\{\big[\mathbb{B}^*(\boldsymbol{v}^*;t)\big]_{a,b} P(\boldsymbol{v}^*;t)\Big\}, \tag{4.39}$$

where $\mathbb{B}^*(\boldsymbol{v}^*) = (\mathbb{A}^*\boldsymbol{v}^*)\,(\mathbb{A}^*\boldsymbol{v}^*)^{\mathrm{T}} + \mathbb{B}_1^*(\boldsymbol{v}^*) + \mathbb{B}_0^*$ and

$$[\,\mathbb{B}_1^*(\boldsymbol{v}^*)\,]_{a,b} = \frac{1}{N}\sum_{c=1}^{n-1}\Big\{\big(\boldsymbol{e}^a\circ\boldsymbol{e}^b\big)\cdot\mathbb{W}\,\boldsymbol{e}^c - \big[\big(\mathbb{W}^{\mathrm{T}}\boldsymbol{e}^a\big)\circ\big(\mathbb{W}^{\mathrm{T}}\boldsymbol{e}^b\big)\big]\cdot\boldsymbol{e}^c\Big\}\,\beta_c\,v_c^* + \mathcal{A}_a^*\big[\mathbb{A}^*\boldsymbol{v}^*\big]_b + \big[\mathbb{A}^*\boldsymbol{v}^*\big]_a\mathcal{A}_b^*, \tag{4.40a}$$

$$[\,\mathbb{B}_0^*\,]_{a,b} = \mathcal{A}_a^*\mathcal{A}_b^* + \frac{1}{Nn}\Big\{\big(\boldsymbol{e}^a\circ\boldsymbol{e}^b\big)\cdot\mathbb{W}\,\mathbf{1} - \big(\mathbb{W}^{\mathrm{T}}\boldsymbol{e}^a\big)\cdot\big(\mathbb{W}^{\mathrm{T}}\boldsymbol{e}^b\big)\Big\}. \tag{4.40b}$$

We have dropped the temporal arguments in Eqs. (4.38) and (4.40) because these definitions also carry straight over to the discrete Fokker-Planck equation in Eq. (4.31) when similarly restricted to $\boldsymbol{v}^*$. The vector $\boldsymbol{\mathcal{A}}^*(t)$ in the first term on the RHS of Eq. (4.39) induces the inhomogeneous term that we saw on the RHS of Eq. (4.15); such an inhomogeneous term is also present in the equivalent discrete dynamics. Restricting to $\boldsymbol{v}^*$ has also induced a $\boldsymbol{v}^*$-independent contribution to the diffusion matrix, in the form of $\mathbb{B}_0^*$. Compared to the equivalent matrix $\mathbb{B}^{(0)}$ defined via Eq. (3.11) in Elliott (2019), $\mathbb{B}_0^*$ is modified by the inhomogeneous contribution $\boldsymbol{\mathcal{A}}^*\boldsymbol{\mathcal{A}}^{*\mathrm{T}}$. There are also modifications to $\mathbb{B}_1^*(\boldsymbol{v}^*)$ compared to its equivalent $\mathbb{B}^{(1)}(\boldsymbol{v})$ in Elliott (2019). In addition to the explicit time dependence of $\mathbb{W}(t)$, compared to an SU, a TDSU also contains inhomogeneities that contribute to both the drift or convective dynamics and to the diffusive dynamics. Only if $\mathbb{W}$ is symmetric are these inhomogeneities absent. In moving from general $\boldsymbol{v}$ to $\boldsymbol{v}^*$ with $\boldsymbol{e}^n \equiv 1$, we explicitly see these inhomogeneities for asymmetric $\mathbb{W}$, but it is often more convenient to keep $\boldsymbol{e}^n$ and thus v_n general so that results can be expressed as compactly as possible,

with the inhomogeneities hidden as factors of the non-dynamical v_n variable.

In principle the Fokker-Planck equation in Eq. (4.39) now allows us to compute MFPTs for memory lifetimes. Given an initial state $P(\boldsymbol{v}^*;0)$ defined by the storage of $\boldsymbol{\xi}^0$, we solve the Fokker-Planck equation for $P(\boldsymbol{v}^*;t)$ subject to the boundary condition $P(\boldsymbol{v}^*;t)=0$ on the hyperplane defined by $\boldsymbol{\delta}_1^*\cdot\boldsymbol{v}^*=\vartheta$, where $\boldsymbol{\delta}_1^*=(1,0,\ldots,0)^{\mathrm{T}}$, which hyperplane just defines the location of the perceptron's firing threshold $v_1\equiv h=\vartheta$ in this $(n-1)$-dimensional space. The MFPT is then just

$$\tau_{\mathrm{mfpt}}=\int_0^\infty dt\int_{\boldsymbol{\delta}_1\cdot\boldsymbol{v}^*>\vartheta}d\boldsymbol{v}^*P(\boldsymbol{v}^*;t). \tag{4.41}$$

In practice we cannot obtain analytical solutions to Eq. (4.39), and even for the bistate case, numerical solution are extremely hard to obtain, it being very difficult to push solutions out to sizeable rt without numerical instabilities developing (Elliott, 2017b).

For an SU, we showed (Elliott, 2019) that we can in fact approximate the dynamics of memory storage as an Ornstein-Uhlenbeck (OU) (Uhlenbeck & Ornstein, 1930) process. This approximation depends on: first, neglecting quadratic terms in the diffusion matrix; second, observing that odd (v_{2i+1}) and even (v_{2i}) variables completely decouple in the drift matrix and the constant part of the diffusion matrix; third, observing that odd and even variables couple in the linear part of the diffusion matrix only via terms that are of order $1/N^{3/2}$, because the even variables always have zero means. With an $\mathcal{O}(1/N)$ approximation, we can then approximate the diffusion matrix by retaining only its $\boldsymbol{v}$-independent, constant part. Even and odd variables then completely decouple, and since $h\equiv v_1$, we need only consider the odd variables. Furthermore, the drift and (approximated) diffusion matrices can then be simultaneously diagonalised. This diagonalisation radically simplifies the differential or integral equations for FPT moments, but the absorbing boundary condition continues

to couple the odd variables together. However, it usually suffices to work with the slowest eigenmode defined by the maximal eigenvector of the drift matrix $\mathbb{A}^*_{\text{su}} = \mathbb{E}\,(\mathbb{W}^*_{\text{su}} - \mathbb{I})\,\mathbb{E}^{-1}$. For the SU, then, we can reduce the problem to a 1-dimensional OU process in the dominant eigenmode, writing down explicit analytical solutions for the MFPT (or indeed the entire FPT density) from the Fokker-Planck equation, or solving 1-dimensional integral equations from the (exact) integral equation approach to solving FPT problems (van Kampen, 1992). We can move beyond the dominant eigenmode and extend to the two dominant modes where strictly necessary, but the single mode provides a remarkably good approximation in most cases.

4.4 Dynamical Comparison of TDSU and Matched SU

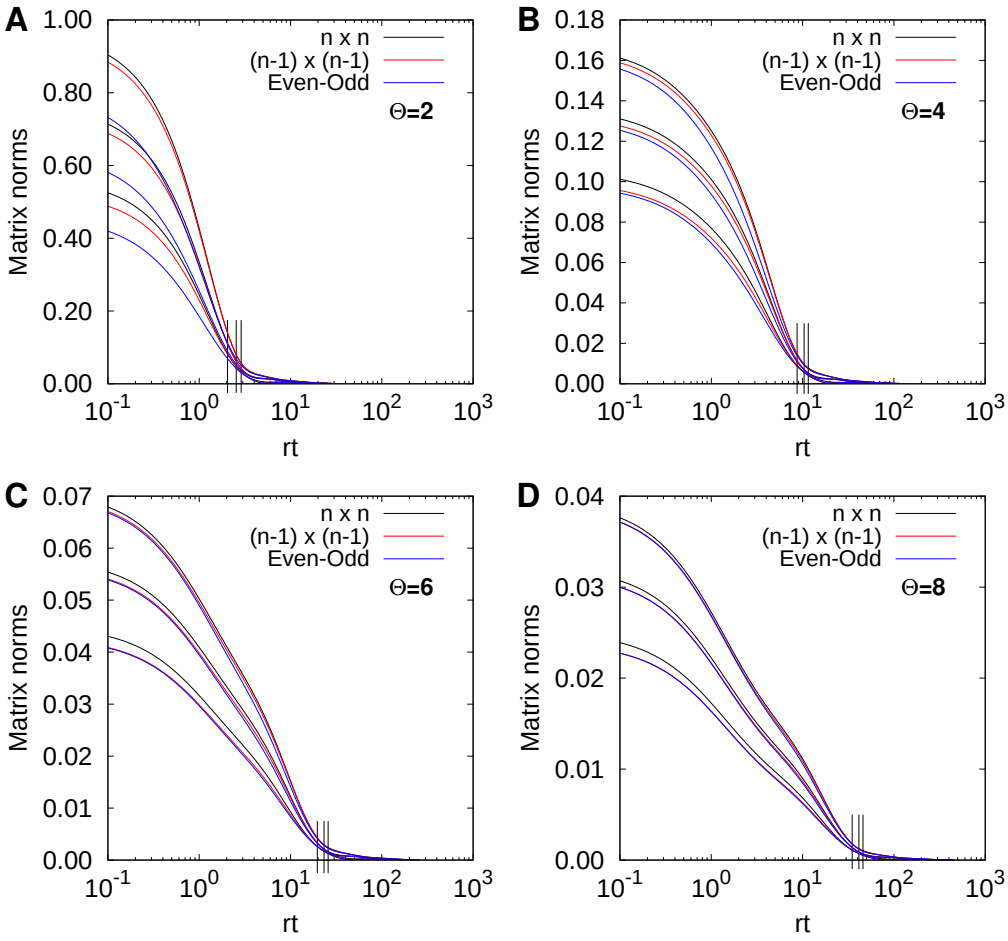
For the same approximation methods that work for FPT calculations with an SU to carry over unchanged to the TDSU, we require that the TDSU's even dynamical variables v_{2i} ($2i \neq n$) have zero means. This is already assured, as discussed above, because of the particular choice of the vectors $\boldsymbol{e}^a$, with the even vectors being symmetric.

We also require that the even and odd variables decouple in the TDSU's drift matrix $\mathbb{A}^*(t)$. To examine this, in Fig. 5 we plot, for various choices of Θ and n, the 2-norm of the matrix formed from the elements of $\mathbb{A}^*(t)$ that just couple the even and odd variables together (blue lines). We explicitly see that this norm is non-zero, and therefore the even and odd variables are coupled together in $\mathbb{A}^*(t)$, breaking the decoupling that is critical to tractable FPT calculations in an SU. However, the coupling rapidly drops to very close to zero, and is already small at time $t = \tau_{\text{peak}}$. Also shown in this figure (red lines) is the normed difference $\|\mathbb{A}^*(t) - \mathbb{A}^*_{\text{su}}\|_2$ between the dynamical TDSU drift matrix and that for an SU with matched $p = 1/\Theta^2$. We have used $\|\mathbb{A}^*(t) - \mathbb{A}^*_{\text{su}}\|_2$ and not the relative normed difference $\|\mathbb{A}^*(t) - \mathbb{A}^*_{\text{su}}\|_2 / \|\mathbb{A}^*(t)\|_2$

so that $\|\mathbb{A}^*(t) - \mathbb{A}^*_{\rm su}\|_2$ can be directly compared with the norm of the even-odd coupling matrix. For $\Theta = 2$ in Fig. 5A, these two norms are clearly different, although $\|\mathbb{A}^*(t) - \mathbb{A}^*_{\rm su}\|_2$ also drops rapidly to zero and is nearly so even at $t = \tau_{\rm peak}$. For the other choices of Θ in the other panels, these two norms are extremely similar and are almost indistinguishable for larger values of Θ. We conclude from this, at least for $\Theta > 2$, that the even-even and odd-odd coupling elements in $\mathbb{A}^*(t)$ are typically very similar to those in $\mathbb{A}^*_{\rm su}$, and that the principal difference between these two matrices is the presence of non-zero even-odd coupling elements in $\mathbb{A}^*(t)$.

Of course, the TDSU is not just a probability-matched SU: the TDSU's mean memory signal initially rises, while an SU's signal monotonically falls. For completeness, we therefore also plot in the same Fig. 5 the normed difference $\|\mathbb{A}(t) - \mathbb{A}_{\rm su}\|_2$, which includes the inhomogeneous contributions to the mean dynamics (black lines). We see that $\|\mathbb{A}(t) - \mathbb{A}_{\rm su}\|_2$ and $\|\mathbb{A}^*(t) - \mathbb{A}^*_{\rm su}\|_2$ follow each other closely, but are always distinguishable, with the former being larger than the latter. We further see that $\|\mathbb{A}(t) - \mathbb{A}_{\rm su}\|_2$ has also dropped to close to zero at $t = \tau_{\rm peak}$, showing that the inhomogeneous terms in $\mathbb{A}(t)$ drop out at around the time that the mean memory signal reaches its peak. These inhomogeneous terms are on the RHS of Eq. (4.15), or are defined explicitly in Eq. (4.38), and appear in the Fokker-Planck equation in Eq. (4.39).

To pursue the role of the inhomogeneous terms in the mean dynamics, in Fig. 6 we plot $\mathcal{A}^*_a(t)$ against time, for the particular choice $n = 8$ and $\Theta = 6$. We also plot $\mu_a(t)$ against time by solving Eq. (4.15) both with and without the inhomogeneous terms present. With the homogeneous terms, we see the characteristic initial rise in $\mu_1(t)$, but also in the three other odd variables' means. While $\mu_1(t)$ rises, reaches a peak and falls back to zero, the other odd means oscillate around zero before asymptoting to zero; they relax to zero on timescales faster than $\mu_1(t)$. The number of finite solutions


A
B
C
D
n x n
(n-1) x (n-1)
Even-Odd
Θ=2
Θ=4
Θ=6
Θ=8
Matrix norms
rt

of $\mu_{2i+1}(t) = 0$ is i, with this being determined by the number of changes in sign of the components of $\boldsymbol{e}^{2i+1}$. The three even variables' means are all zero, always. Turning off the homogeneous terms, instead of initially rising, the odd variables' means fall, with $\mu_1(t)$ falling monotonically to zero, reminiscent of an SU. Indeed, were we to plot the means for a matched SU in this figure, the SU's odd means would be visually indistinguishable from the TDSU's odd means, although there are small numerical differences. For smaller parameter choices (not shown), the difference would be noticeable but still small. For the odd means, then, switching off the inhomogeneous terms in the TDSU's dynamics renders them extremely similar to those for a matched SU. For the even means, which all vanish for the full TDSU and in an SU, the absence of the inhomogeneous terms allows them to become non-zero. Thus, removing the homogeneous terms from the mean dynamics does not turn a TDSU into

---

Figure 5 (previous page): Various matrix 2-norms plotted against time, showing the convergence of the TDSU's strength transition matrix to that of a matched SU. Each panel shows results for the choice of Θ as indicated, and within each panel three different choices of n are used, corresponding to $n = 4$ (lowest triplet of lines), $n = 6$ (middle triplet of lines) and $n = 8$ (highest triplet of lines). Blue lines correspond to the 2-norm of the matrix formed from only the even-odd coupling elements in the $(n-1) \times (n-1)$ matrix $\mathbb{A}^*(t)$. Red lines show the normed difference $\|\mathbb{A}^*(t) - \mathbb{A}^*_{\text{su}}\|_2$; these matrices do not include inhomogeneous coupling terms. Black lines show the full normed difference $\|\mathbb{A}(t) - \mathbb{A}_{\text{su}}\|_2$, where the $n \times n$ matrix $\mathbb{A}(t)$ includes inhomogeneous coupling terms. Equivalent results cannot be shown for $n = 2$ because there are no even variables in this case, and results for $n = 3$ are very similar to those for $n = 4$. The three vertical bars in each panel indicate the locations of $r\tau_{\text{peak}}$ for each value of n, with $r\tau_{\text{peak}}$ increasing as n increases.

a matched SU because of the even means' anomalous behaviour, but the odd means do become extremely SU-like.

Examining the evolution of the inhomogeneous terms explicitly in Fig. 6B, we see once again that they are all very small around $t = \tau_{\text{peak}}$, or small compared to their values of $t = 0$ s. The odd inhomogeneous terms are initially positive, which drives the initial rise in the odd means. They fall and can oscillate around zero, contributing to the oscillations in the corresponding means. The even means are always exactly zero, but we have seen that there are even-odd coupling terms in the drift matrix that would without other contributions drive the even means to non-zero values. The even inhomogeneous terms, which are initially negative, exactly oppose the even-odd coupling terms, keeping the even means at precisely zero. When these even inhomogeneous terms are removed, it is the even-odd coupling terms that cause the even means to become non-zero via the non-zero odd means.

In summary, for $t < \tau_{\text{peak}}$ the mean dynamics of the TDSU are mostly dominated by inhomogeneous terms. Without them, the TDSU's odd means evolve very similarly to a matched SU's odd means, while the TDSU's even means exhibit mildly anomalous behaviour. Pre-peak, then, a TDSU is essentially a matched SU with additional inhomogeneities. For $t > \tau_{\text{peak}}$, we have that $\mathbb{A}(t) \approx \mathbb{A}_{\text{su}}$, so that the post-peak mean dynamics of the TDSU can to a good approximation be replaced by those of a matched SU. In this regime, the inhomogeneous terms and the even-odd coupling terms in $\mathbb{A}(t)$ can be neglected. We further confirm that the post-peak TDSU essentially behaves like a matched SU by directly comparing $\mu_1(t)$ for a TDSU with $t \geq \tau_{\text{peak}}$ to the suitably scaled and shifted $\mu_1(t)$ for an SU with $t \geq 0$ s. In particular, we write

$$[\mu_1(t)]_{\text{tdsu}} \approx \frac{[\mu_1(\tau_{\text{peak}})]_{\text{tdsu}}}{[\mu_1(0)]_{\text{su}}} \, [\mu_1(t - \tau_{\text{peak}})]_{\text{su}}, \text{ for } t \geq \tau_{\text{peak}}. \tag{4.42}$$

The overall scaling ensures that we have the correct value of $[\mu_1(t)]_{\text{tdsu}}$ at

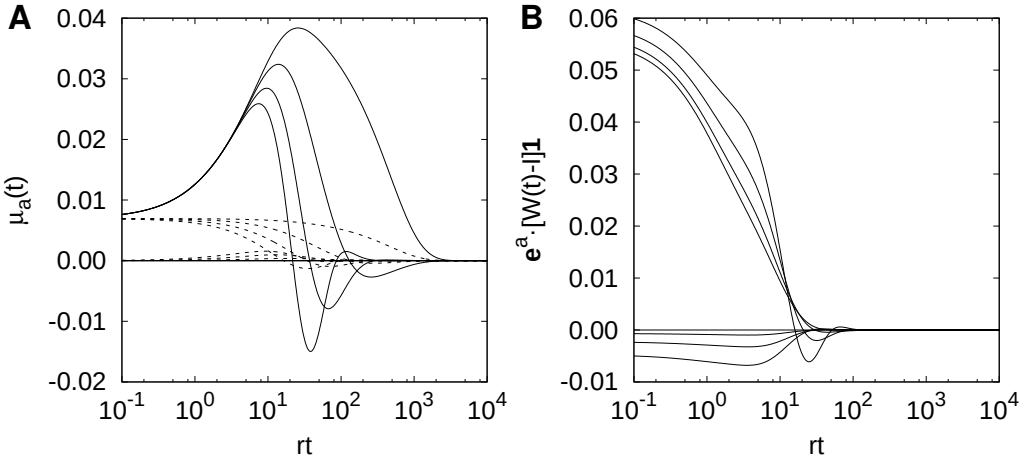


Figure 6: Evolution of the means $\mu_a(t)$ with and without inhomogeneous terms, for $\Theta = 6$ and $n = 8$. (A) Solid lines show the full dynamical evolution of $\mu_a(t)$ in the presence of the inhomogeneous terms $\mathcal{A}^*_a(t)$, while the dashed lines show their evolution in the absence of $\mathcal{A}^*_a(t)$. The four solid lines correspond, from top to bottom, reading the maxima, to $a = 1$, 3, 5 and 7, and so are the odd variables; the full dynamical even variables have zero means. The four highest dashed lines correspond to the same odd variables, again in order from top to bottom. Without the inhomogeneous terms, the three even variables means' now become slightly non-zero. (B) The seven scaled inhomogeneous terms $n\,\mathcal{A}^*_a(t) = \boldsymbol{e}^a \cdot [\,\mathbb{W}(t) - \mathbb{I}\,]\,\mathbf{1}$, with the odd ones being positive at $rt = 10^{-1}$ and ordered bottom to top on the graph as $a = 1$, 3, 5 and 7, and with the even ones being negative at $rt = 10^{-1}$ and ordered from top to bottom as $a = 2$, 4 and 6.

$t = \tau_{\text{peak}}$, and the temporal offset in $[\mu_1(t - \tau_{\text{peak}})]_{\text{su}}$ ensures that the dynamics of the SU start when the TDSU reaches peak. We compare this approximation to the exact form of $[\mu_1(t)]_{\text{tdsu}}$ in Fig. 7 for various choices of Θ and n. We see that indeed the post-peak TDSU's mean memory signal is well approximated by a suitably transformed SU's signal, especially for n away from its smaller possible values.

5 TDSU Drift-Only Pre-Peak and SU Post-Peak Approximation

We have seen that in terms of the mean dynamics, a TDSU can to a very good approximation be replaced by a matched SU at times $t \geq \tau_{\text{peak}}$. This behaviour reflects the fact that the conditional strength change probabilities $p_a^{\pm}$ are very close to their equilibrium value, $1/\Theta^2$, at time $t = \tau_{\text{peak}}$, even though $\mu_1(t)$ is at this time at its maximum value. While we focused above on the dynamics of the TDSU's drift matrix $\mathbb{A}(t)$ in relation to $\mathbb{A}_{\text{su}}$, the early equilibration of the probabilities $p_a^{\pm}$ will also ensure that the TDSU's diffusion matrix will be very similar to that of an SU for $t \geq \tau_{\text{peak}}$. Indeed, because all jump moments at any order are determined from Eq. (4.24), the convergence of $\mathbb{W}_\alpha$ to $\mathbb{W}_{\text{su}}$ will ensure that the TDSU's dynamics at all orders will converge to those of an SU with $p = 1/\Theta^2$. Although this is a relaxation process and not a sharp, threshold-like process, to a good approximation a post-peak TDSU behaves at all orders just like a matched SU. Therefore, to make FPT calculations tractable, as an approximation we replace the post-peak TDSU by an SU with $p = 1/\Theta^2$. The TDSU dynamics are therefore separated into a pre-peak phase and a post-peak phase. The dynamics between these two phases must be matched at $t = \tau_{\text{peak}}$ not only by using an SU with $p = 1/\Theta^2$ for $t \geq \tau_{\text{peak}}$, but also by "initialising" the SU at $t = \tau_{\text{peak}}$ with the TDSU's distribution $P(\boldsymbol{v}; \tau_{\text{peak}})$ of the $\boldsymbol{v}$ variables at signal peak. We first discuss how this distribution $P(\boldsymbol{v}; \tau_{\text{peak}})$ is obtained, then revisit FPT results for an SU, and finally examine FPT statistics with

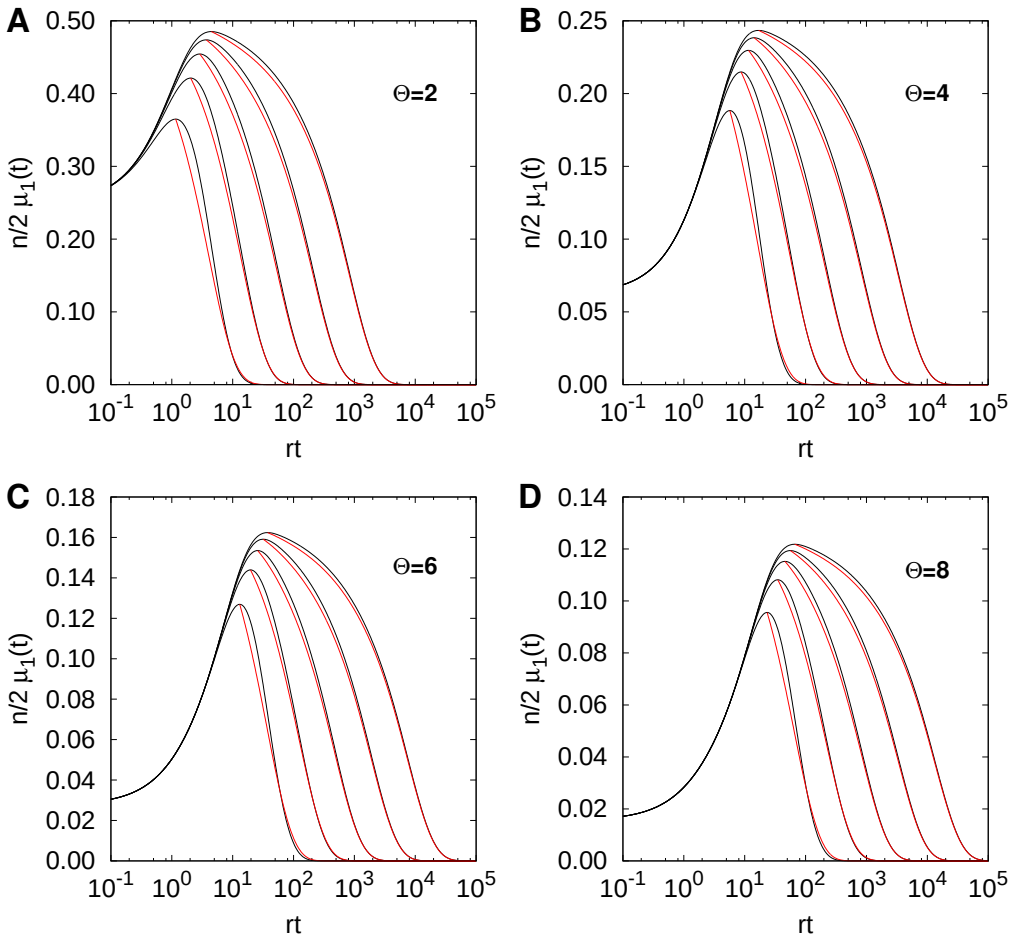


Figure 7: Comparison between the evolution of $\mu_1(t)$ for the full, filter-based dynamics (black lines) and the post-peak, matched SU approximation in Eq. (4.42) (red lines). Each panel indicates the value of Θ used, and within each panel, we take $n = 2^1$, 2^2, 2^3, 2^4 and 2^5, reading from bottom to top in each graph. The mean $\mu_1(t)$ is scaled by $n/2$ within each panel for easier comparison.

this approximation.

5.1 TDSU Drift-Only Pre-peak Dynamics

To determine $P(\boldsymbol{v};\tau_{\text{peak}})$ exactly, we still have to solve the Fokker-Planck equation in Eq. (4.34) in the presence of the absorbing hyperplane condition $\boldsymbol{\delta}_1 \cdot \boldsymbol{v} = \vartheta$, where now $\boldsymbol{\delta}_1$ is an n- rather than $(n-1)$-dimensional vector with components $[\boldsymbol{\delta}_1]_a = \delta_{1,a}$, with $\delta_{a,b}$ being the Kronecker delta function. Even though τ_{peak} is typically small enough that we could perhaps obtain stable numerical solutions, such solutions would be entirely uninformative, offering no analytical insight. We must therefore also find ways of approximating the TDSU's pre-peak dynamics. For $t \leq \tau_{\text{peak}}$, we know that $\mu_1(t)$ increases. In fact, it increases approximately Θ-fold (Elliott & Lagogiannis, 2012; Elliott, 2016a). The variable v_1 is therefore on average rapidly pulled or pushed away from any firing threshold ϑ, if $v_1(0) > \vartheta$. We saw earlier that these dynamics are dominated by inhomogeneous terms in the drift equation. Further, it is usually the case that $\tau_{\text{peak}} \ll \tau_{\text{mfpt}}$, so any approximations that we make for $t \leq \tau_{\text{peak}}$ are unlikely to have a significant impact on the determination of FPT processes. So, since the pre-peak dynamics are dominated by drift or convective processes that move the distribution $P(\boldsymbol{v};t)$ away from the absorbing hyperplane $\boldsymbol{\delta}_1 \cdot \boldsymbol{v} = \vartheta$, we make the pre-peak approximation that the TDSU's dynamics are dominated by drift, and that diffusion can be neglected.

The Fokker-Planck equation in Eq. (4.34) therefore reduces to the Liouville equation

$$\frac{1}{r}\frac{\partial}{\partial t}P(\boldsymbol{v};t) = -\boldsymbol{\nabla}_{\boldsymbol{v}} \cdot \Big\{ \big[\mathbb{A}(t)\boldsymbol{v}\big]\, P(\boldsymbol{v};t) \Big\}. \tag{5.1}$$

With a definite initial state $\boldsymbol{v}^0$, so that $P(\boldsymbol{v};0) = \delta\big(\boldsymbol{v} - \boldsymbol{v}^0\big)$, the solution of the Liouville equation is

$$P(\boldsymbol{v};t) = \delta\big(\boldsymbol{v} - \overline{\mathbb{V}}(t)\boldsymbol{v}^0\big), \tag{5.2}$$

where we write $\overline{\mathbb{V}}(t) = \mathbb{E}\,\overline{\mathbb{U}}(t)\mathbb{E}^{-1}$. This solution corresponds to the deterministic flow $\boldsymbol{v}(t) = \overline{\mathbb{V}}(t)\boldsymbol{v}(0)$ satisfying

$$\frac{1}{r}\frac{d}{dt}\,\boldsymbol{v}(t) = \mathbb{E}\,[\,\mathbb{W}(t) - \mathbb{I}\,]\,\mathbb{E}^{-1}\boldsymbol{v}(t) \equiv \mathbb{A}(t)\boldsymbol{v}(t), \tag{5.3}$$

where $\boldsymbol{v}(0) \equiv \boldsymbol{v}^0$. With a general initial distribution $P(\boldsymbol{v};0)$, the solution of the Liouville equation is

$$P(\boldsymbol{v};t) = \int d\boldsymbol{v}^0 P(\boldsymbol{v}^0;0)\,\delta\big(\boldsymbol{v} - \overline{\mathbb{V}}(t)\boldsymbol{v}^0\big) = \frac{1}{\big|\det\overline{\mathbb{V}}(t)\big|}P\big(\overline{\mathbb{V}}(t)^{-1}\boldsymbol{v};0\big). \tag{5.4}$$

Taking the initial distribution to be a Gaussian multivariate with mean vector $\boldsymbol{\mu}(0)$ and covariance matrix $\mathbb{S}(0)$, Eq. (5.4) gives

$$\begin{aligned} P(\boldsymbol{v};t) = {} & \frac{1}{\sqrt{\det 2\pi\left[\,\overline{\mathbb{V}}(t)\,\mathbb{S}(0)\,\overline{\mathbb{V}}(t)^{\mathrm{T}}\right]}} \\ & \times \exp\left\{-\frac{1}{2}\left[\boldsymbol{v} - \overline{\mathbb{V}}(t)\boldsymbol{\mu}(0)\right]^{\mathrm{T}}\left[\,\overline{\mathbb{V}}(t)\,\mathbb{S}(0)\,\overline{\mathbb{V}}(t)^{\mathrm{T}}\right]^{-1}\left[\boldsymbol{v} - \overline{\mathbb{V}}(t)\boldsymbol{\mu}(0)\right]\right\}, \end{aligned} \tag{5.5}$$

which is also a Gaussian multivariate, with mean vector $\boldsymbol{\mu}(t) = \overline{\mathbb{V}}(t)\boldsymbol{\mu}(0)$ and covariance matrix $\mathbb{S}(t) = \overline{\mathbb{V}}(t)\,\mathbb{S}(0)\,\overline{\mathbb{V}}(t)^{\mathrm{T}}$. The mean $\boldsymbol{\mu}(t)$ of course satisfies Eq. (4.14), while this covariance matrix $\mathbb{S}(t)$ satisfies

$$\frac{1}{r}\frac{d}{dt}\,\mathbb{S}(t) = \mathbb{A}(t)\mathbb{S}(t) + \mathbb{S}(t)\mathbb{A}(t)^{\mathrm{T}}, \tag{5.6}$$

which is just Eq. (4.36) with the contributions due to the diffusion matrix $\mathbb{B}(\boldsymbol{v};t)$ stripped out, and also follows directly from Eq. (5.1).

The Gaussian multivariate in Eq. (5.5) is valid for a general vector $\boldsymbol{e}^n$. However, for the specific case $\boldsymbol{e}^n \equiv \mathbf{1}$, the initial covariance matrix with elements given by Eq. (4.9b) satisfies $[\,\mathbb{S}_0\,]_{n,b} = [\,\mathbb{S}_0\,]_{a,n} \equiv 0$ and is therefore singular. In this case Eq. (5.5) factorises, giving $P(\boldsymbol{v};t) = P(\boldsymbol{v}^*;t) \times \delta(v_n - 1)$, because v_n

is not dynamical. Writing $\overline{\mathbb{V}}(t)$ in the schematic form (cf. Eq. (4.37))

$$\overline{\mathbb{V}}(t) = \left(\begin{array}{c|c} \overline{\mathbb{V}}^*(t) & \overline{\boldsymbol{\mathcal{V}}}^*(t) \\ \hline \mathbf{0}^{\mathrm{T}} & 1 \end{array} \right), \tag{5.7}$$

the distribution $P(\boldsymbol{v}^*;t)$ in the dynamical $\boldsymbol{v}^*$ variables is a Gaussian multivariate with mean vector $\boldsymbol{\mu}^*(t) = \overline{\boldsymbol{\mathcal{V}}}^*(t) + \overline{\mathbb{V}}^*(t)\boldsymbol{\mu}^*(0)$ and covariance matrix $\mathbb{S}^*(t) = \overline{\mathbb{V}}^*(t)\,\mathbb{S}^*(0)\,\overline{\mathbb{V}}^*(t)^{\mathrm{T}}$, where the initial statistics are given in Eq. (4.9), and where again we see an inhomogeneous term in the (exact) evolution of the mean vector. For simplicity of notation we keep $\boldsymbol{e}^n$ general below, but with this factorisation of $P(\boldsymbol{v};t)$ understood for the specific case of $\boldsymbol{e}^n \equiv \mathbf{1}$.

We note the appearance of Eq. (5.3) with solution $\boldsymbol{v}(t) = \overline{\mathbb{V}}(t)\boldsymbol{v}(0)$, which appears to generalise the merely mean dynamics in Eq. (4.14) with solution $\boldsymbol{\mu}(t) = \overline{\mathbb{V}}(t)\boldsymbol{\mu}(0)$. The Fokker-Planck equation, however, is usually correct only at the level of the mean and covariance, ignoring other issues such as any diffusion limit (van Kampen, 1992). The Liouville equation is therefore usually correct only at the level of the mean. Hence, the appearance of $\overline{\mathbb{V}}(t)$ to describe the evolution of $\boldsymbol{v}(t)$ rather than merely $\boldsymbol{\mu}(t)$ is entirely consistent, at the level of approximation of the Liouville dynamics. However, we could have formulated the drift-only dynamics without using the Fokker-Planck equation. The matrix $\mathbb{U}(t)$ in Eq. (3.15) is the *exact* evolution matrix describing the evolution of a single synapse's probability distribution of strength states, via the equation $\boldsymbol{\Pi}(t) = \mathbb{U}(t)\boldsymbol{\Pi}(0)$. This equation captures correctly not only the mean dynamics but also all higher-order statistics, for a single synapse. For the polysynaptic variables, we have $\boldsymbol{v}(t) = \frac{1}{N}\,\mathbb{E}\,\boldsymbol{N}(t)$ and $\frac{1}{N}\,\mathsf{E}[\boldsymbol{N}(t)] \equiv \boldsymbol{\Pi}(t)$, and so $\mathsf{E}[\boldsymbol{v}(t)] = \mathbb{E}\,\mathbb{U}(t)\,\mathbb{E}^{-1}\mathsf{E}[\boldsymbol{v}(0)] \equiv \mathbb{V}(t)\mathsf{E}[\boldsymbol{v}(0)]$, which defines $\mathbb{V}(t)$. The equivalent approximation to the Liouville equation is thus to elevate this mean equation $\mathsf{E}[\boldsymbol{v}(t)] = \mathbb{V}(t)\mathsf{E}[\boldsymbol{v}(0)]$ to $\boldsymbol{v}(t) = \mathbb{V}(t)\boldsymbol{v}(0)$, which obviously fails to capture correctly the fluctuations in $\boldsymbol{v}$. Via this alternative route, we have the

deterministic flow $\boldsymbol{v}(t) = \mathbb{V}(t)\boldsymbol{v}(0)$ rather than $\boldsymbol{v}(t) = \overline{\mathbb{V}}(t)\boldsymbol{v}(0)$, and so $\mathbb{V}(t)$ rather than $\overline{\mathbb{V}}(t)$ would appear in Eq. (5.5). Because $\mathbb{U}(t)\boldsymbol{\Pi}(0) = \overline{\mathbb{U}}(t)\boldsymbol{\Pi}(0)$, we have that $\mathbb{V}(t)\boldsymbol{\mu}(0) = \overline{\mathbb{V}}(t)\boldsymbol{\mu}(0)$, and so the appearance of $\overline{\mathbb{V}}(t)\boldsymbol{\mu}(0)$ in Eq. (5.5) is correct: using Eq. (5.5), we of course obtain the correct mean dynamics. However, with the Liouville approximation, we are not strictly-speaking entitled to calculate the covariance matrix using Eq. (5.5). If we do, then clearly we obtain a result that is incorrect because of the missing diffusion matrix, but it is also incorrect because of the more subtle issue that, in the Liouville approximation, the dynamics of $\boldsymbol{v}(t)$ are identical to those of $\boldsymbol{\mu}(t)$. This is reflected in the covariance matrix in Eq. (5.5) being $\overline{\mathbb{V}}(t)\,\mathbb{S}(0)\,\overline{\mathbb{V}}(t)^{\mathrm{T}}$ and not $\mathbb{V}(t)\,\mathbb{S}(0)\,\mathbb{V}(t)^{\mathrm{T}}$. However, we have seen that the relative normed difference between $\mathbb{U}(t)$ and $\overline{\mathbb{U}}(t)$ is typically very small, and this is also true of $\mathbb{V}(t)$ and $\overline{\mathbb{V}}(t)$. The error incurred in using $\overline{\mathbb{V}}(t)$ rather than $\mathbb{V}(t)$ in Eq. (5.5) will therefore be swamped by the error incurred in using a drift-only approximation to obtain $P(\boldsymbol{v}; \tau_{\mathrm{peak}})$.

5.2 SU Post-peak FPT Processes

For the post-peak period, we approximate the dynamics by a matched SU with $p = 1/\Theta^2$, where the "initial" distribution of this SU is the TDSU's distribution $P(\boldsymbol{v}; \tau_{\mathrm{peak}})$ at signal peak. To avoid awkward temporal offsetting such as that in Eq. (4.42), we just reset time at τ_{peak}, taking this point as the start of time, $t = 0$ s, for the SU.

For an SU with an OU approximation, the odd and even dynamical v_a variables decouple and, moreover, the drift and diffusion matrices can be simultaneously diagonalised (Elliott, 2019). We showed that it suffices to consider the dynamics of the two slowest variables v_1 and v_3 for $n \geq 4$ and that we may in fact usually just work with the slowest eigenmode defined by these two variables, reducing the FPT problem to one dimension (Elliott, 2019). For $n = 2$ or

$n = 3$, the problem is already in just the single variable v_1, with v_2 having decoupled for $n = 3$. We use the 2-dimensional vector $\boldsymbol{u} = (v_1, v_3)^{\mathrm{T}}$ to restrict to the two relevant variables, with the drift and diffusion matrices also being 2×2 matrices restricted to these two variables. We also write $\mathbb{Q} = \mathrm{diag}\{|\boldsymbol{e}^1|^2, |\boldsymbol{e}^3|^2\}$, which is the relevant restriction of the full form, $\mathbb{Q} = \mathbb{E}\,\mathbb{E}^{\mathrm{T}}$. From Eq. (4.31) and appendix B in Elliott (2019), these three matrices are given explicitly by

$$\mathbb{A} = -\frac{2\,p}{n(n+1)} \begin{pmatrix} 3 & 7\,\frac{(n-2)(n-3)}{(n+2)(n+3)} \\ 3 & 42\,\frac{n^2+1}{(n+2)(n+3)} \end{pmatrix}, \tag{5.8a}$$

$$\mathbb{B} = +\frac{1}{N}\,\frac{4\,p}{n(n-1)} \begin{pmatrix} 1 & 1 \\ 1 & 6\,\frac{n^2+1}{(n+2)(n+3)} \end{pmatrix}, \tag{5.8b}$$

$$\mathbb{Q} = +\,n\,\frac{n+1}{n-1} \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{7}\,\frac{(n+2)(n+3)}{(n-2)(n-3)} \end{pmatrix}. \tag{5.8c}$$

Let the two normalised left eigenvectors of $\mathbb{A}$ be $\widehat{\boldsymbol{\ell}}^1$ and $\widehat{\boldsymbol{\ell}}^2$ with eigenvalues $\Lambda_1 - 1$ and $\Lambda_2 - 1$, respectively, with conventions chosen so that $\Lambda_1 \geq \Lambda_2$ and $\widehat{\ell}^1_1 > 0$, and let the matrix $\mathbb{L}$ have its first and second rows as $\widehat{\boldsymbol{\ell}}^{1\,\mathrm{T}}$ and $\widehat{\boldsymbol{\ell}}^{2\,\mathrm{T}}$. Then clearly $\mathbb{L}\mathbb{A}\mathbb{L}^{-1}$ is diagonal, but so too is $\mathbb{L}\mathbb{B}\mathbb{L}^{\mathrm{T}}$, with entries of $-\frac{2}{N}\frac{1}{n}(\Lambda_i - 1)\big[\,\widehat{\boldsymbol{\ell}}^i \cdot \mathbb{Q}\,\widehat{\boldsymbol{\ell}}^i\,\big]$ on the diagonal (Elliott, 2019). If $\boldsymbol{u}^0$ represents a definite initial choice of configuration variables, then performing the transformation of variables $\boldsymbol{x} = \mathbb{L}\boldsymbol{u}^0$ diagonalises the drift and diffusion matrices in the differential equations governing FPT moments. Considering just the slowest mode x_1 corresponding to eigenvalue $\Lambda_1 - 1$, the m^{th} FPT moment $\tau^{(m)}(x_1)$ satisfies the equation

$$-\frac{1}{r}\frac{1}{(1-\Lambda_1)}\,m\,\tau^{(m-1)}(x_1) = \left\{-x_1\frac{d}{dx_1} + \frac{\widehat{\boldsymbol{\ell}}^1 \cdot \mathbb{Q}\,\widehat{\boldsymbol{\ell}}^1}{Nn}\frac{d^2}{dx_1^2}\right\}\tau^{(m)}(x_1), \tag{5.9}$$

where $\tau^{(0)}(x_1) \equiv 1$. This is Eq. (4.35) in Elliott (2019). The boundary condi-

tion is $\tau^{(m)}(\vartheta') = 0$ for $m \geq 1$, where

$$\vartheta' = \vartheta \begin{pmatrix} 1 \\ \widehat{\ell}_1^2/\widehat{\ell}_2^2 \end{pmatrix} \cdot \widehat{\boldsymbol{\ell}}^1, \tag{5.10}$$

which corresponds to the location of the firing threshold ϑ in the x_1 direction; the first vector in this expression is a right eigenvector of $\mathbb{A}$ with eigenvalue $\Lambda_1 - 1$. For $x_1 < \vartheta'$, we define $\tau^{(m)}(x_1) \equiv 0$ for $m \geq 1$.

The tower of coupled equations in Eq. (5.9) can be solved iteratively, but we can also derive the Laplace transform of the entire FPT density. Up to the sign of the argument, this is just the MGF for the FPT process. Writing

$$\widehat{G}_{\text{fpt}}(x, s) = \sum_{m=0}^{\infty} \frac{(-s)^m}{m!} \tau^{(m)}(x), \tag{5.11}$$

by using Eq. (5.9) we obtain

$$\frac{s}{r} \frac{1}{(1 - \Lambda_1)} \widehat{G}_{\text{fpt}}(x_1, s) = \left\{ -x_1 \frac{d}{dx_1} + \frac{\widehat{\boldsymbol{\ell}}^1 \cdot \mathbb{Q} \widehat{\boldsymbol{\ell}}^1}{Nn} \frac{d^2}{dx_1^2} \right\} \widehat{G}_{\text{fpt}}(x_1, s), \tag{5.12}$$

subject to the boundary condition $\widehat{G}_{\text{fpt}}(\vartheta', s) = 1$. By imposing a second boundary condition that we take to infinity, the solution to this equation is (see Elliott (2017a) for the solution of a very similar equation)

$$\widehat{G}_{\text{fpt}}(x_1, s) = \frac{H_{-s/[r(1-\Lambda_1)]}\left(x_1 \sqrt{\frac{Nn}{2\widehat{\boldsymbol{\ell}}^1 \cdot \mathbb{Q} \widehat{\boldsymbol{\ell}}^1}} \right)}{H_{-s/[r(1-\Lambda_1)]}\left(\vartheta' \sqrt{\frac{Nn}{2\widehat{\boldsymbol{\ell}}^1 \cdot \mathbb{Q} \widehat{\boldsymbol{\ell}}^1}} \right)}, \tag{5.13}$$

where $H_{-a}(y)$ is a Hermite polynomial of non-integer order. Expanding as a power series in $-s$ returns the FPT moments for $x_1 \geq \vartheta'$. For the particular case of $\vartheta = 0$, we can use an image construction method using the unbounded solution of the Fokker-Planck equation for the OU process to directly obtain

$G(x_1,t)$, without using the Laplace transform (Elliott, 2017a). We obtain

$$G(x_1,t) = r(1-\Lambda_1)\sqrt{\frac{2}{\pi}\frac{Nn}{\widehat{\boldsymbol{\ell}}^1\cdot\mathbb{Q}\,\widehat{\boldsymbol{\ell}}^1}}\,\frac{x_1\,e^{-(1-\Lambda_1)rt}}{\left[1-e^{-2(1-\Lambda_1)rt}\right]^{3/2}} \times\exp\left\{-\frac{Nn}{2\,\widehat{\boldsymbol{\ell}}^1\cdot\mathbb{Q}\,\widehat{\boldsymbol{\ell}}^1}\frac{x_1^2\,e^{-2(1-\Lambda_1)rt}}{\left[1-e^{-2(1-\Lambda_1)rt}\right]}\right\}, \tag{5.14}$$

which is just Eq. (3.22) of Elliott (2017a) adapted to our problem here.

We may also obtain FPTs for the SU using the integral equation approach (van Kampen, 1992). In the absence of any approximations such as the continuum limit, this method provides exact results for FPTs. In contrast, the approach based on the Fokker-Planck equation requires a diffusion approximation, which eliminates the jump processes that occur during memory storage. However, given that we have had to make several approximations in order to obtain a tractable formulation of the FPT problem, the integral equation approach will not be exact here. Nevertheless, we compare results from both methods. In the slowest x_1 variable, the FPT moments are given by the integral equation

$$\tau^{(m)}(x_1) = \frac{1}{r}\,m\,\tau^{(m-1)}(x_1) + \int_{\vartheta'}^{\infty} dy_1\,\tau^{(m)}(y_1)\,\overline{K}_1(y_1\,|\,x_1), \tag{5.15}$$

where $\overline{K}_1(x\,|\,x')$ is the 1-dimensional kernel

$$\overline{K}_1(x\,|\,x') = \sqrt{\frac{Nn}{2\pi\big[2(1-\Lambda_1)\,\widehat{\boldsymbol{\ell}}^1\cdot\mathbb{Q}\,\widehat{\boldsymbol{\ell}}^1\big]}}\,\exp\left\{-\frac{Nn\big(x-\Lambda_1 x'\big)^2}{2\big[2(1-\Lambda_1)\,\widehat{\boldsymbol{\ell}}^1\cdot\mathbb{Q}\,\widehat{\boldsymbol{\ell}}^1\big]}\right\}. \tag{5.16}$$

These are Eqs. (4.23) and (4.36) in Elliott (2019). This kernel is a 1-dimensional Gaussian that arises from the factorisation of an $(n-1)$-dimensional Gaussian in the $\boldsymbol{v}^*$ variables when the SU drift and diffusion matrices $\mathbb{A}^*_{\text{su}}$ and $\mathbb{B}^*_{\text{su}}$ are diagonalised, with this $(n-1)$-dimensional kernel itself being a Gaussian approximation to the full one-memory-storage-step transition probabilities in the

SU. We can write down the integral equation for the entire FPT density (Elliott, 2019), but we directly solve Eq. (5.15) by numerical methods for the first two FPT moments.

5.3 First Passage Times for Filter-Based Synapses

Working for the moment in the full $\boldsymbol{v}$ variables, with the above approximations we write the conditional MFPT $\tau_{\text{mfpt}}(\boldsymbol{v}^0)$ for a definite initial polysynaptic configuration $\boldsymbol{v}^0$ as $\tau_{\text{mfpt}}(\boldsymbol{v}^0) = \tau_{\text{peak}} + \tau^{(1)}\big(\widehat{\boldsymbol{\lambda}}^1 \cdot \overline{\mathbb{V}}(\tau_{\text{peak}})\boldsymbol{v}^0\big)$, where $\widehat{\boldsymbol{\lambda}}^1$ is the full eigenvector corresponding to the restricted eigenvector $\widehat{\boldsymbol{\ell}}^1$. The state $\overline{\mathbb{V}}(\tau_{\text{peak}})\boldsymbol{v}^0$ is $\boldsymbol{v}^0$ after its drift-only evolution, and the transformed variable $x_1 = \widehat{\boldsymbol{\lambda}}^1 \cdot \overline{\mathbb{V}}(\tau_{\text{peak}})\boldsymbol{v}^0$ is the slowest mode of the matched SU's full dynamics. To obtain the unconditional MFPT τ_{mfpt}, we use either of the two forms

$$\tau_{\text{mfpt}} = \int\limits_{\mathfrak{R}[\boldsymbol{v}^0]} d\boldsymbol{v}^0 \Big[\tau_{\text{peak}} + \tau^{(1)}\big(\widehat{\boldsymbol{\lambda}}^1 \cdot \overline{\mathbb{V}}(\tau_{\text{peak}})\boldsymbol{v}^0\big)\Big] P(\boldsymbol{v}^0; 0), \tag{5.17a}$$

$$\tau_{\text{mfpt}} = \int\limits_{\mathfrak{R}\left[\overline{\mathbb{V}}(\tau_{\text{peak}})^{-1}\boldsymbol{v}^0\right]} d\boldsymbol{v}^0 \Big[\tau_{\text{peak}} + \tau^{(1)}\big(\widehat{\boldsymbol{\lambda}}^1 \cdot \boldsymbol{v}^0\big)\Big] P(\boldsymbol{v}^0; \tau_{\text{peak}}), \tag{5.17b}$$

which are equivalent in virtue of Eq. (5.4), and where $\mathfrak{R}[\boldsymbol{v}^0]$ denotes the integration region in the n-dimensional space. With the absorbing hyperplane $\boldsymbol{\delta}_1 \cdot \boldsymbol{v}^0 = \vartheta$, $\mathfrak{R}[\boldsymbol{v}^0]$ in Eq. (5.17a) corresponds to the region $\boldsymbol{\delta}_1 \cdot \boldsymbol{v}^0 > \vartheta$, so that we integrate over only that part of the initial distribution $P(\boldsymbol{v}^0; 0)$ that is above firing threshold, where $\boldsymbol{v}^0$ corresponds to an initial TDSU configuration at $t = 0$ s. The condition $\mathfrak{R}\big[\,\overline{\mathbb{V}}(\tau_{\text{peak}})^{-1}\boldsymbol{v}^0\big]$ in Eq. (5.17b) corresponds to the region $\boldsymbol{\delta}_1 \cdot \big[\,\overline{\mathbb{V}}(\tau_{\text{peak}})^{-1}\boldsymbol{v}^0\,\big] > \vartheta$, so that we integrate over only that part of the evolved distribution $P(\boldsymbol{v}^0; \tau_{\text{peak}})$ that was initially above firing threshold at $t = 0$ s (hence, the same part as in Eq. (5.17a)), where $\boldsymbol{v}^0$ now corresponds to an initial matched SU configuration at $t = \tau_{\text{peak}}$. This is because the time-dependent hyperplane equation $\boldsymbol{\delta}_1 \cdot \big[\,\overline{\mathbb{V}}(t)^{-1}\boldsymbol{v}\,\big] = \vartheta$ corresponds to the

evolution of the initial hyperplane equation $\boldsymbol{\delta}_1 \cdot \boldsymbol{v} = \vartheta$ under the deterministic flow $v(t) = \overline{\mathbb{V}}(t)v(0)$. Thus, as the initial distribution $P(\boldsymbol{v}; 0)$ evolves into $P(\boldsymbol{v}; t)$, the flow of the initial hyperplane $\boldsymbol{\delta}_1 \cdot \boldsymbol{v} = \vartheta$ into $\boldsymbol{\delta}_1 \cdot \left[\overline{\mathbb{V}}(t)^{-1} \boldsymbol{v} \right] = \vartheta$ keeps track of where the initial hyperplane cleaved the initial distribution.

Imposing the condition $\boldsymbol{\delta}_1 \cdot \boldsymbol{v}^0 > \vartheta$ on initial TSDU configurations $\boldsymbol{v}^0$ at $t = 0$ s is consistent with FPT calculations in which the hyperplane $\boldsymbol{\delta}_1 \cdot \boldsymbol{v} = \vartheta$ is an absorbing boundary that removes all subsequent configurations $\boldsymbol{v}$ whose activations falls below firing threshold. However, although consistent, this naïve approach systematically underestimates unconditional memory lifetimes for filter-based synapses. As we have seen, the mean tracked memory signal actually increases, reaching its peak at τ_{peak}. While there are some initial TDSU configurations $\boldsymbol{v}^0$ that exceed firing threshold, there are also others that initially fall below threshold, but that at around τ_{peak} are above it. In some parameter regimes, the mean memory signal itself may initially fall below threshold, then at some later time rise above threshold, and then finally fall back below it. By imposing the condition $\boldsymbol{\delta}_1 \cdot \boldsymbol{v}^0 > \vartheta$ on initial TDSU configurations, we lose this complexity and nuance in filter-based dynamics, losing contributions from configurations who activations are initially below threshold but that will (on average) rise above threshold at around τ_{peak}. Were we to probe such systems experimentally, we would see an initial period during which the tracked memory is not recalled, followed by a later period during which it is recalled. In such a case, we would define the lifetime of the memory as the last time at which the memory is recalled: we would not define the lifetime to be zero because it was not initially recalled.

With the pre-peak, drift-only approximation above, the distribution $P(\boldsymbol{v}; t)$ in Eq. (5.5) at $t = \tau_{\text{peak}}$ is obtained from purely deterministic dynamics and, furthermore, without imposing the absorbing boundary condition at $\boldsymbol{\delta}_1 \cdot \boldsymbol{v} = \vartheta$ during the evolution. The pre- and post-peak approximations therefore allow

us very easily to modify the passage time calculations by instead imposing the condition $\boldsymbol{\delta}_1 \cdot \boldsymbol{v}^0 > \vartheta$ on the initial, matched SU configurations at $t = \tau_{\text{peak}}$ rather than on the initial TDSU configurations at $t = 0$ s. Hence, with this approach, we write

$$\tau_{\text{mfpt}} = \int\limits_{\mathfrak{R}[\boldsymbol{v}^0]} d\boldsymbol{v}^0 \Big[\tau_{\text{peak}} + \tau^{(1)}\big(\widehat{\boldsymbol{\lambda}}^1 \cdot \boldsymbol{v}^0\big)\Big] P(\boldsymbol{v}^0; \tau_{\text{peak}}), \tag{5.17c}$$

where we integrate the initial matched SU configurations $\boldsymbol{v}^0$ at $t = \tau_{\text{peak}}$ over the region $\boldsymbol{\delta}_1 \cdot \boldsymbol{v}^0 > \vartheta$. We will consider and compare both Eq. (5.17b) and Eq. (5.17c) below. For convenience, we define the quantity τ_ϑ as the time at which the firing threshold condition is first applied to the distribution. In Eq. (5.17a), we have $\tau_\vartheta = 0$ s, corresponding to the "naïve" approach, and in Eq. (5.17c), we have $\tau_\vartheta = \tau_{\text{peak}}$, corresponding to the "nuanced" approach.

We work not with the full $\boldsymbol{v}$ variables, but instead with the restricted $\boldsymbol{u}$ variables. To apply Eqs. (5.17), we require $P(\boldsymbol{u}; t)$ instead of $P(\boldsymbol{v}; t)$, where the former is the Gaussian bivariate marginal distribution of the latter in the v_1 and v_3 variables. We must also restrict the integration region. If the region is defined by $\boldsymbol{\delta}_1 \cdot \big[\,\overline{\mathbb{V}}(t)^{-1}\boldsymbol{v}\,\big] > \vartheta$ for some t, then we marginalise over this hyperplane condition so that its average projection onto the v_1–v_3 plane is

$$\big[\,\overline{\mathbb{V}}(t)^{-1}\big]_{1,1} v_1 + \big[\,\overline{\mathbb{V}}(t)^{-1}\big]_{1,3} v_3 > \vartheta - \sum_{a \neq 1,3} \big[\,\overline{\mathbb{V}}(t)^{-1}\big]_{1,a} \mu_a(t). \tag{5.18}$$

In fact, the sum on the RHS makes a negligible difference to our results, and we could drop it. For $\tau_\vartheta = 0$ s, we take $t = \tau_{\text{peak}}$, and for $\tau_\vartheta = \tau_{\text{peak}}$, we take $t = 0$ s, since $\overline{\mathbb{V}}(0) \equiv \mathbb{I}$. Defining

$$\boldsymbol{\Delta}(t) = \begin{pmatrix} \big[\,\overline{\mathbb{V}}(\tau_{\text{peak}} - t)^{-1}\big]_{1,1} \\ \big[\,\overline{\mathbb{V}}(\tau_{\text{peak}} - t)^{-1}\big]_{1,3} \end{pmatrix}, \tag{5.19}$$

and

$$\varphi(t) = \vartheta - \sum_{a\neq 1,3} \left[\overline{\mathbb{V}}(\tau_{\text{peak}} - t)^{-1} \right]_{1,a} \mu_a(\tau_{\text{peak}} - t), \tag{5.20}$$

we can then write the unconditional MFPT for filter-based synapses as

$$\tau_{\text{mfpt}} = \int\limits_{\boldsymbol{\Delta}(\tau_\vartheta)\cdot\boldsymbol{u}^0 > \varphi(\tau_\vartheta)} d\boldsymbol{u}^0 \Big[\tau_{\text{peak}} + \tau^{(1)}\big(\widehat{\boldsymbol{\ell}}^1 \cdot \boldsymbol{u}^0\big) \Big] P(\boldsymbol{u}^0; \tau_{\text{peak}}), \tag{5.21}$$

where we set τ_ϑ as required. We will also require the unconditional variance in the FPT, which we denote by σ^2_{fpt}, and is obtained from the second moment of the FPT distribution,

$$\sigma^2_{\text{fpt}} + \tau^2_{\text{mfpt}} = \int\limits_{\boldsymbol{\Delta}(\tau_\vartheta)\cdot\boldsymbol{u}^0 > \varphi(\tau_\vartheta)} d\boldsymbol{u}^0 \Big[\tau^2_{\text{peak}} + 2\,\tau_{\text{peak}}\,\tau^{(1)}\big(\widehat{\boldsymbol{\ell}}^1 \cdot \boldsymbol{u}^0\big) + \tau^{(2)}\big(\widehat{\boldsymbol{\ell}}^1 \cdot \boldsymbol{u}^0\big) \Big] P(\boldsymbol{u}^0; \tau_{\text{peak}}). \tag{5.22}$$

The moments $\tau^{(m)}(x_1)$ are obtained either via Eq. (5.13) or from Eq. (5.15), corresponding to the FPT moments for the post-peak SU with matched $p = 1/\Theta^2$.

In order to understand the general trends in some of our results below, we will need some analytical understanding of the conditional FPT cumulants, $\tau_{\text{mfpt}}(\boldsymbol{u}^0)$ and $\sigma^2_{\text{fpt}}(\boldsymbol{u}^0)$, for a definite configuration $\boldsymbol{u}^0$. In a post-peak SU approximation, we essentially need only look at the FPT cumulants for the matched SU, and we have extensively examined the behaviour of the FPT process for an SU before (Elliott, 2017a, 2019). Writing $\tau_{\text{su}}(\boldsymbol{u}^0)$ and $\sigma^2_{\text{su}}(\boldsymbol{u}^0)$ for the first two FPT cumulants for an SU, by combining the results of our earlier analyses based on the Fokker-Planck approach in the OU limit (Elliott, 2017a, 2019) we have

$$r(1-\Lambda_1)\,\tau_{\text{su}}(\boldsymbol{u}^0) \sim \begin{cases} \frac{1}{2}\left[\gamma + \log_e \frac{2Nn\left(\widehat{\boldsymbol{\ell}}^1\cdot\boldsymbol{u}^0\right)^2}{\widehat{\boldsymbol{\ell}}^1\cdot\mathbb{Q}\,\widehat{\boldsymbol{\ell}}^1}\right] & \text{for } \vartheta = 0 \\ \log_e \frac{\widehat{\boldsymbol{\ell}}^1\cdot\boldsymbol{u}^0}{\vartheta'}, & \text{for } \vartheta > 0 \end{cases}, \tag{5.23}$$

and

$$r^2(1-\Lambda_1)^2\,\sigma_{\rm su}^2(\boldsymbol{u}^0) \sim \begin{cases} \frac{\pi^2}{8} & \text{for } \vartheta = 0 \\ 0 & \text{for } \vartheta > 0 \end{cases}, \tag{5.24}$$

where γ is Euler's constant, and where these results hold for N sufficiently large. However, for N sufficiently large, the distribution $P(\boldsymbol{u}^0;\tau_{\rm peak})$ is also tightly concentrated around its mean $\mathsf{E}\left[\boldsymbol{u}(\tau_{\rm peak})\right]$, i.e. $P(\boldsymbol{u}^0;\tau_{\rm peak}) \approx \delta(\boldsymbol{u}^0 - \mathsf{E}\left[\boldsymbol{u}(\tau_{\rm peak})\right])$. Thus, provided that $\mu_1(\tau_\vartheta) > \vartheta$, the integrals in Eqs. (5.21) and (5.22) collapse and we have

$$\tau_{\rm mfpt} \sim \tau_{\rm peak} + \tau_{\rm su}\left(\mathsf{E}\left[\boldsymbol{u}(\tau_{\rm peak})\right]\right), \tag{5.25}$$

$$\sigma_{\rm fpt}^2 \sim \sigma_{\rm su}^2\left(\mathsf{E}\left[\boldsymbol{u}(\tau_{\rm peak})\right]\right), \tag{5.26}$$

for the asymptotic behaviours of the unconditional TDSU FPT cumulants, where $\tau_{\rm su}$ and $\sigma_{\rm su}^2$ are obtained using the asymptotic forms in Eqs. (5.23) and (5.24). The resulting asymptotic expression for $\tau_{\rm mfpt}$ is rather messy, but if we make the somewhat coarse approximation that $\widehat{\boldsymbol{\ell}}^1$ points principally in the v_1 direction, so that $\widehat{\boldsymbol{\ell}}^1 \approx (1,0)^{\rm T}$, which ignores the v_3 dynamics, then we obtain

$$\tau_{\rm mfpt} \approx \tau_{\rm peak} + \frac{1}{r(1-\Lambda_1)} \begin{cases} \frac{1}{2}\left[\gamma + \log_e \frac{2Nn\,\mu_1(\tau_{\rm peak})^2}{|\boldsymbol{\Omega}|^2}\right] & \text{for } \vartheta = 0 \\ \log_e \frac{\mu_1(\tau_{\rm peak})}{\vartheta} & \text{for } \vartheta > 0 \end{cases}. \tag{5.27}$$

We have previously (Elliott, 2016a) obtained an approximation for $\tau_{\rm peak}$,

$$r\tau_{\rm peak} \approx \frac{2}{\cos\frac{\pi}{n\Theta} - \cos\frac{\pi}{\Theta}} \log_e\left(n\frac{\cos\frac{\pi}{2\Theta}}{\cos\frac{\pi}{2n\Theta}}\right) \sim \frac{4\Theta^2}{\pi^2}\log_e n, \tag{5.28}$$

where the second form is for large enough n and Θ. Obtaining a decent approximation for $\mu_1(\tau_{\rm peak})$ is more difficult. The best that we could do is

$$\mu_1(\tau_{\rm peak}) \approx \frac{2}{n}\frac{8}{\Theta\pi^2}\left(1+\frac{1}{n}\right)n^{-2/(n^2-1)} \sim \frac{2}{n}\frac{8}{\Theta\pi^2}, \tag{5.29}$$

where the second form is for large enough n. Replacing the messy eigenvalue $\Lambda_1 - 1$ of the 2×2 matrix $\mathbb{A}$ in Eq. (5.8a) by the corresponding eigenvalue $\lambda_1 - 1 = -\frac{1}{\Theta^2}\left(1 - \cos\frac{\pi}{n}\right) \sim -\frac{\pi^2}{2n^2\Theta^2}$ of the full matrix $\mathbb{A}$, where the λ_a are given in Eq. (3.34a), we obtain for $\vartheta = 0$,

$$r\tau_{\text{mfpt}} \approx \frac{4\Theta^2}{\pi^2}\log_e n + \frac{n^2\Theta^2}{\pi^2}\left(\gamma + \log_e \frac{1536N}{n^2\Theta^2\pi^4}\right), \tag{5.30}$$

for the large N behaviour of τ_{mfpt}, for large enough n and Θ.

To determine how large N must be for the onset of asymptotic behaviour, we previously compared $\left\langle G(h^{(0)}, t)\right\rangle_{h^{(0)}>\vartheta}$ to its mean field form $G(\langle h^{(0)}\rangle, t)$ for a simple, multistate synapse satisfying an eigenvalue requirement, for $\vartheta = 0$ (Elliott, 2017a). We obtained the simple condition $\mathsf{E}[h^{(0)}]\sqrt{Nn/|\boldsymbol{\Omega}|^2} \gtrapprox 2$. To obtain a similar condition for the TDSU, we should in principle use Eq. (5.14), taking the average over the relevant part of the distribution $P(\boldsymbol{u}^0; \tau_\vartheta)$, but any resulting condition(s) would be very messy. Instead, we just use the same coarse approximation for $\widehat{\boldsymbol{\ell}}^1$ as above, and for $\tau_\vartheta = 0$ s we take $\mathsf{E}[h^{(0)}] = \mu_1(0) = \frac{2}{n}\frac{1}{\Theta^2}$, while for $\tau_\vartheta = \tau_{\text{peak}}$ we take $\mathsf{E}[h^{(0)}] = \mu_1(\tau_{\text{peak}}) \approx \frac{2}{n}\frac{8}{\Theta\pi^2}$ from Eq. (5.29). We then obtain the conditions

$$N \gtrapprox \begin{cases} \frac{1}{3}\frac{n+1}{n-1}\, n^2\Theta^4 \approx \frac{1}{3}\, n^2\Theta^4 & \text{for } \tau_\vartheta = 0 \text{ s} \\ \frac{\pi^4}{192}\frac{n+1}{n-1}\, n^2\Theta^2 \approx \frac{1}{2}\, n^2\Theta^2 & \text{for } \tau_\vartheta = \tau_{\text{peak}} \end{cases}, \tag{5.31}$$

and we find that they provide good estimates of how large N must be for the onset of asymptotic behaviour in both cases, for $\vartheta = 0$.

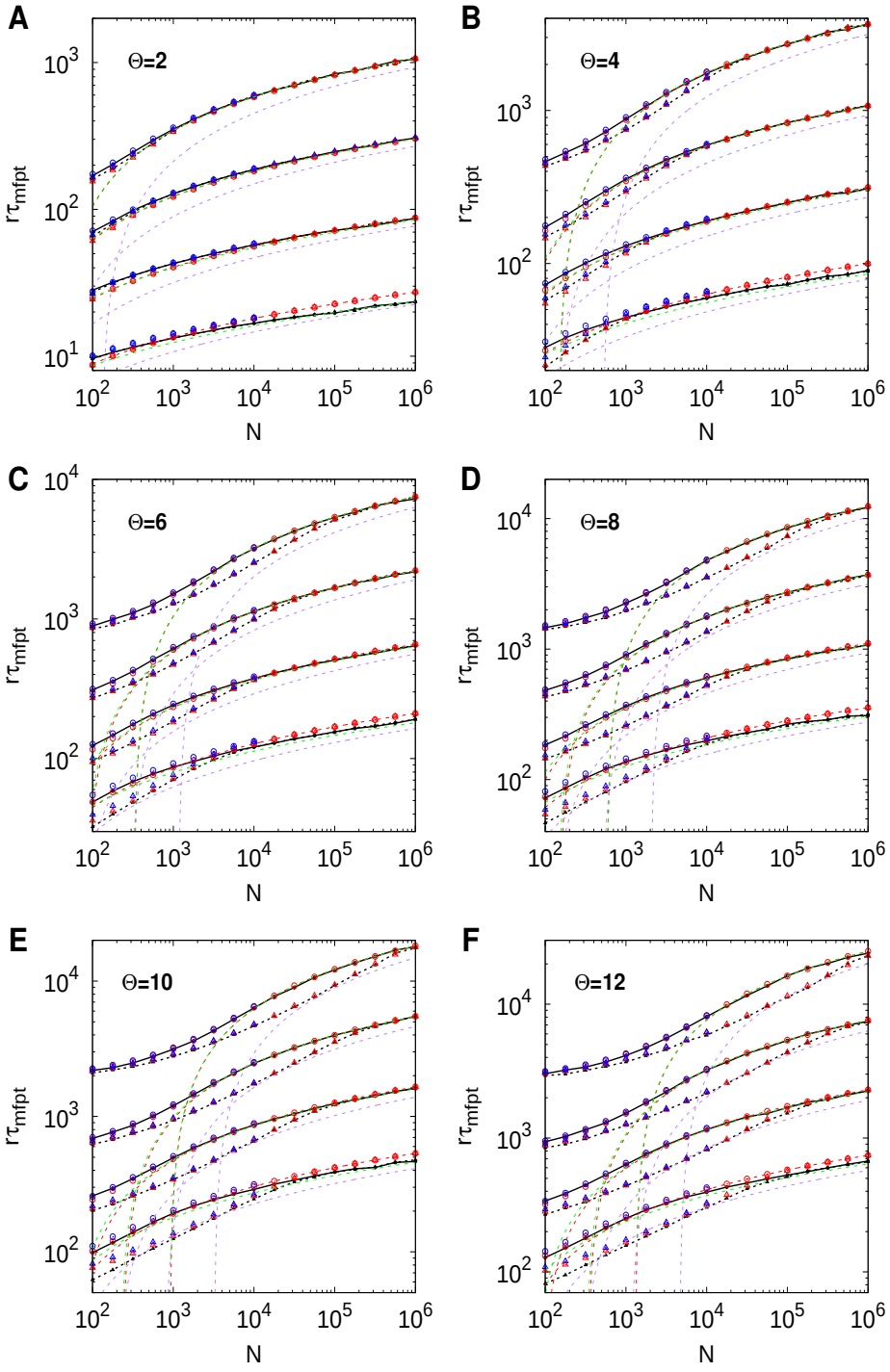
6 Comparison Between Analytical and Simulation Results

We may now compare the results for FPT statistics obtained from the above approximation methods to results from simulations of a perceptron storing memories using filter-based synapses. Our simulation methods are discussed

extensively elsewere (Elliott & Lagogiannis, 2012; Elliott, 2014, 2016a). Each simulation constitutes a single realisation of a particular sequence of memory storage events and initial synaptic configuration, so we must average over multiple simulations in order to obtain the unconditional MFPT τ_{mfpt} and variance σ^2_{fpt}. The number of runs over which we average depends principally on N, with fewer runs for larger N because there is intrinsically more self-averaging for larger N. However, larger N requires more simulation time because MFPTs are larger for larger N, so we must balance obtaining good statistics against simulation time. Furthermore, more averaging is required for higher-order statistics compared to lower-order statistics. Typically, we average over between 10^2 and 10^3 simulations for $N = 10^6$, depending on n and Θ, and over 10^5 simulations for $N = 10^2$.

In Fig. 8 we plot both forms of τ_{mfpt} (for $\tau_\vartheta = 0$ s and $\tau_\vartheta = \tau_{\text{peak}}$) against N for various choices of Θ and n, for the particular choice of firing threshold $\vartheta = 0$. Results are shown for simulations, for the Fokker-Planck approach to FPTs using both the results in Eq. (5.21) and their asymptotic form in Eqs. (5.25) and (5.30), and for the integral equation approach to FPTs. For this last, we restrict to solving the integal equations for smaller values of N, up to 10^4. For the other methods, we take N up to 10^6, but this value is already somewhat biologically unrealistic: Purkinje cells, which have the largest numbers of synapses, have up to around 2.5×10^5 (Napper & Harvey, 1988), although most neurons have considerably fewer.

We focus first on comparing simulation, Fokker-Planck and integral equation results in Fig. 8. For bistate synapses, we see small differences between analytical and simulation results, although the analytical results nevertheless capture the trends in the simulation results perfectly, in every detail. We also see differences between the Fokker-Planck and integral equation results for $n = 2$, which are in general less noticeable for larger n. Given that we


A
Θ=2
B
Θ=4
C
Θ=6
D
Θ=8
E
Θ=10
F
Θ=12
rτmfpt
N

saw earlier that our approximation methods are less good for very small n (see Figs. 3, 5 and 7), this difference between analytical and simulation results is not surprising. Further, we already have exact analytical results for bistate synapses (Elliott, 2017b), so we have other methods available in this case. For larger values of n, we see extremely good agreement between analytical and simulation results, especially for larger values of N. This agreement is both surprising and remarkable, given the approximation methods used to obtain the analytical results. It demonstrates that indeed for n away from two, the filter-based TDSU is very well approximated by drift-only pre-peak dynamics and matched SU post-peak dynamics, with a sharp transition between them.

Turning to the differences between the two forms of τ_{mfpt} for $\tau_\vartheta = 0$ s and $\tau_\vartheta = \tau_{\text{peak}}$ in Fig. 8, we see that, as expected, results for $\tau_\vartheta = 0$ s systemat-

---

Figure 8 (previous page): Mean first passage memory lifetimes for filter-based synapses plotted as a function of N, for various choices of Θ and n, and with a perceptron firing threshold, $\vartheta = 0$. Each panel shows results for the indicated choice of Θ, and within each panel we take (reading from bottom to top in each graph) $n = 2^1$, 2^2, 2^3 and 2^4. Results are shown for both forms of MFPT, $\tau_\vartheta = \tau_{\text{peak}}$ and $\tau_\vartheta = 0$ s. Black solid ($\tau_\vartheta = \tau_{\text{peak}}$) and dashed ($\tau_\vartheta = 0$ s) lines show the results of simulations, with simulation data points shown by small black circles and triangles, respectively. Red circles and triangles indicate results from the Fokker-Planck approach to computing FPTs, while the red and green dashed lines show the corresponding asymptotic forms of the Fokker-Planck results in Eq. (5.25) and Eq. (5.30), respectively. Blue circles and triangle show results from the integral equation approach. Results for these latter are obtained only up to $N = 10^4$ synapses. Also shown for comparison, as dashed purple lines, are memory lifetimes determined from an SNR criterion.

ically underestimate those for $\tau_\vartheta = \tau_{\rm peak}$. For larger N, they converge to the same asymptotic behaviour, with this convergence being guaranteed for $\vartheta = 0$ because $\mu_1(\tau_\vartheta)$ always exceeds zero. For smaller Θ, the differences between the results tend to be smaller, while for larger Θ, the differences are larger, with larger values of N being required for convergence. This reflects the difference between $P(\boldsymbol{u}^0; 0)$ in Eq. (5.17a) and $P(\boldsymbol{u}^0; \tau_{\rm peak})$ in Eq. (5.17c), and specifically the relative sizes of the initial signal $\mu_1(0) = \frac{2}{n}\frac{1}{\Theta^2}$ and the peak signal $\mu_1(\tau_{\rm peak}) \approx \frac{2}{n}\frac{8}{\Theta\pi^2}$ in relation to the threshold ϑ. The latter is nearly Θ-fold larger than the former, and so for any finite N, more of the distribution $P(\boldsymbol{u}^0; \tau_{\rm peak})$ is above threshold than $P(\boldsymbol{u}^0; 0)$, with this difference being larger for large N. We also see a trend for small N that increasing n reduces the difference between the two forms of $\tau_{\rm mfpt}$. For small N, the two distributions $P(\boldsymbol{u}^0; 0)$ and $P(\boldsymbol{u}^0; \tau_{\rm peak})$ are quite diffuse, and increasing n reduces both $\mu_1(0)$ and $\mu_1(\tau_{\rm peak})$. Overall, the two forms of $\tau_{\rm mfpt}$ can differ by up to around two-fold, for the parameters used in this figure, but their general trends are similar.

Comparing the behaviour of $\tau_{\rm mfpt}$ for smaller and larger values of N in Fig. 8, we see clear differences. For larger N, we see the transition to the asymptotic, logarithmic growth of $\tau_{\rm mfpt}$ with N in Eq. (5.25), and we note also that the approximate asymptotic form in Eq. (5.30) is an extremely good approximation to the exact asymptotic form in Eq. (5.25) except for $n = 2$. We also see that $\tau_{\rm mfpt}$ for $\tau_\vartheta = \tau_{\rm peak}$ transitions to its asymptotic behaviour much more quickly that does the $\tau_\vartheta = 0$ s form, confirming the different Θ-dependences in Eq. (5.31). This difference is again a question of the relative distances of $\mu_1(0)$ and $\mu_1(\tau_{\rm peak})$ from $\vartheta = 0$. Using Eq. (5.31) to obtain an estimate of how large N must be for the Fokker-Planck form to transition to asymptotic behaviour, we find that (taking typical examples and rounding to the most significant figure) for $\Theta = 6$ and $n = 4$ (panel C) we need $N \gtrapprox 7 \times 10^3$

for $\tau_\theta = 0$ s and $N \gtrapprox 3 \times 10^2$ for $\tau_\theta = \tau_{\rm peak}$, while for $\Theta = 10$ and $n = 16$ (panel E) we need $N \gtrapprox 9 \times 10^5$ for $\tau_\theta = 0$ s and $N \gtrapprox 1 \times 10^4$ for $\tau_\theta = \tau_{\rm peak}$, in agreement with the numbers in these panels. For smaller N, there are non-uniformities in $\tau_{\rm mfpt}$, reflecting the idiosyncratic dynamics of the approach to asymptotic behaviour, governed by the sizes of $\mu_1(\tau_{\rm peak})$ and $\mu_1(0)$.

For comparison, in Fig. 8 we also show memory lifetimes defined according to an SNR criterion. This defines the memory lifetime $\tau_{\rm snr}$ as the largest finite solution of $\mathsf{E}\left[h(\tau_{\rm snr})\right]^2 = \mathsf{Var}\left[h(\tau_{\rm snr})\right]$; when a solution does not exist, $\tau_{\rm snr}$ is defined to be zero. We have derived and extensively examined $\tau_{\rm snr}$ for filter-based synapses before. For $\vartheta = 0$, to a good approximation we have

$$r\tau_{\rm snr} \approx \frac{1}{2\left(1-\cos\frac{\pi}{n\Theta}\right)} \log_e \left[\frac{48N}{n^4(n^2-1)\Theta^6} \cot^4 \frac{\pi}{2n\Theta}\right] \sim \frac{n^2\Theta^2}{\pi^2} \log_e \frac{768N}{n^2\Theta^2\pi^4}, \tag{6.1}$$

when this is positive (Elliott, 2016a); the last form is for n large enough. The memory lifetimes $\tau_{\rm mfpt}$ and $\tau_{\rm snr}$ can be somewhat similar, but $\tau_{\rm mfpt}$ is always non-zero while $\tau_{\rm snr}$ can be zero for smaller values of N, indicating that $\tau_{\rm snr}$ is only asymptotically valid and in general is a very poor measure of memory lifetimes for small N. For larger N in this figure, $\tau_{\rm mfpt}$ and $\tau_{\rm snr}$ can differ by a constant, or nearly so, for some parameter choices, indicating that the coefficient of $\log_e N$ in their expressions in Eqs. (5.25) and (6.1) are identical, or nearly so. Indeed, comparing the approximation to the asymptotic form of $\tau_{\rm mfpt}$ in Eq. (5.30) to the large n form of the expression for $\tau_{\rm snr}$ in Eq. (6.1), these expressions are identical as functions of N, differing only by additive but n- and Θ-dependent constants.

We note that both Eq. (5.30) and Eq. (6.1) exhibit maxima as a function of n or Θ, with other parameters held fixed, suggesting optimal choices of n or Θ to maximise $\tau_{\rm mfpt}$ and $\tau_{\rm snr}$ for fixed N. However, these optima are illusory. Specifically, the full rather than asymptotic forms of $\tau_{\rm mfpt}$ never exhibit such

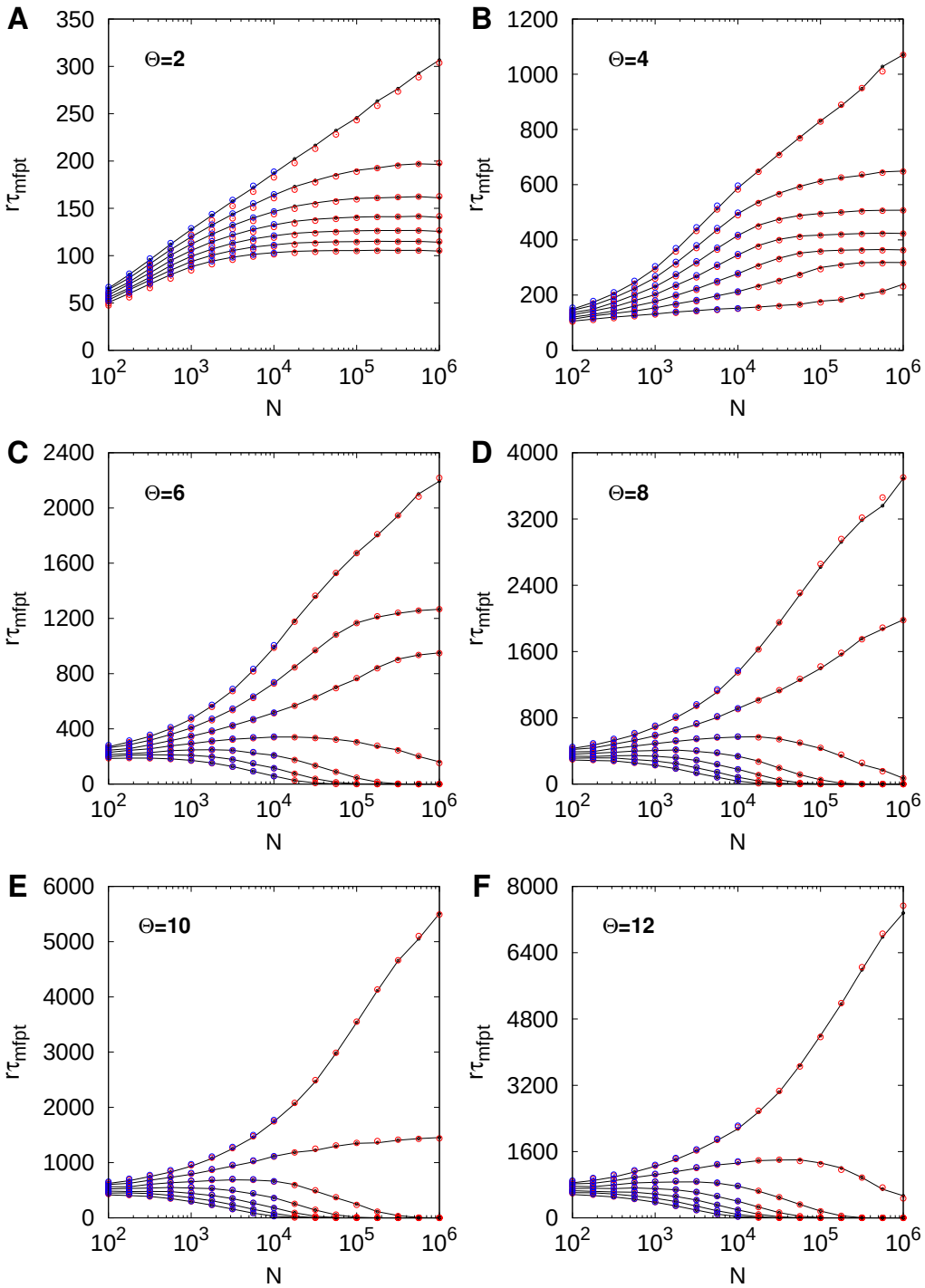
maxima, because the apparently optimal choice of n or Θ that maximises the asymptotic form of τ_{mfpt} for a fixed value of N ensures that that value of N does not satisfy either condition in Eq. (5.31). In particular, using only the second, dominant term in Eq. (5.30), τ_{mfpt} is maximised when $n^2\Theta^2 = 1536e^{\gamma-1}N/\pi^4 \approx 10.33N$. Since $1 \not> 3.44\,\Theta^2$ (for $\tau_\vartheta = 0$ s) and $1 \not> 5.17$ (for $\tau_\vartheta = \tau_{\text{peak}}$), both conditions in Eq. (5.31) are violated and so the asymptotic form used to derive this optimal condition is never attained; indeed, it does not even come close to being attained. Thus, both forms of τ_{mfpt} in fact never exhibit maxima as functions of n or Θ, but increase monotonically with them. Because τ_{snr} is also only asymptotically valid, its maximum as a function of n or Θ is therefore also spurious. We do not labour these issues here because we have discussed them at length in relation to simulation results or SU synapses in earlier work, and in particular we have discussed when SNR memory lifetimes are an acceptable substitute for MFPT memory lifetimes (Elliott, 2016a, 2017a).

We examine the impact of non-zero ϑ on the two forms of τ_{mfpt} for the particular choice of $n = 8$ in Figs. 9 and 10, for $\tau_\vartheta = 0$ s and $\tau_\vartheta = \tau_{\text{peak}}$, respectively. Results are shown for the seven choices, $\vartheta = 0.0000$, 0.0025, 0.0050, 0.0075, 0.0100, 0.0125 and 0.0150. We note that in these figures we see more noise in the simulation results for large N and for larger MFPTs. In Fig. 9 for $\tau_\vartheta = 0$ s, the location of the initial mean $\mu_1(\tau_\vartheta) = \frac{2}{n}\frac{1}{\Theta^2}$ relative to the threshold ϑ determines the overall behaviour. For $\Theta = 2$ or 4, $\mu_1(\tau_\vartheta) = \frac{1}{16}$ or $\frac{1}{64}$, both being in excess of the largest value of ϑ used in this figure. Hence, in all cases in panels A and B, for $\vartheta > 0$, τ_{mfpt} always eventually assumes the asymptotic behaviour in Eq. (5.25) for large enough N. For $\Theta = 6$ in panel C, $\mu_1(\tau_\vartheta) = \frac{1}{144} \approx 0.0069$, which sits in the middle of the range of values of ϑ considered here. For $\mu_1(\tau_\vartheta) < \vartheta$, as N increases, the initial distribution $P(\boldsymbol{u}^0; 0)$ mostly falls below firing threshold with an increasingly small tail above it, so in these cases τ_{mfpt} drops to zero as N increases. As Θ increase

further, $\mu_1(\tau_\vartheta)$ drops further, until for all non-zero values of ϑ for $\Theta = 12$ used in panel F, τ_{mfpt} always goes to zero as N increases.

In contrast, for $\tau_\vartheta = \tau_{\text{peak}}$ in Fig. 10, we have that the peak mean $\mu_1(\tau_\vartheta) \approx \frac{2}{n}\frac{8}{\Theta\pi^2}$. Even for $\Theta = 12$ in panel F, $\mu_1(\tau_\vartheta) \approx 0.0169$, so in excess of the largest value of ϑ used in this figure. Hence, for all non-zero choices of ϑ in Fig. 10, τ_{mfpt} assumes its asymptotic behaviour in Eq. (5.25). We would need to take n or Θ larger to start to see τ_{mfpt} dropping to zero as N increases. We have not shown the constant, asymptotic, non-zero values of τ_{mfpt} in Eq. (5.25) for $\vartheta > 0$ in these figures to avoid unnecessary clutter, but we confirm that they are indistinguishable from the exact results shown. We have already observed that for $\vartheta = 0$, τ_{mfpt} for $\tau_\vartheta = \tau_{\text{peak}}$ asymptotes to logarithmic behaviour faster than for $\tau_\vartheta = 0$ s. Comparing the $\vartheta = 0.0150$ case for $\Theta = 4$ in panel B of Figs. 9 and 10, we also see faster attainment of the constant, non-zero asymptotic behaviour for $\tau_\vartheta = \tau_{\text{peak}}$ compared to $\tau_\vartheta = 0$ s. Overall, then, MFPTs for $\tau_\vartheta = \tau_{\text{peak}}$ are more stable and better behaved than for $\tau_\vartheta = 0$ s, unless n or Θ are taken to be large.

In Figs. 11 and 12 for $\tau_\vartheta = 0$ s and $\tau_\vartheta = \tau_{\text{peak}}$, respectively, we plot σ_{fpt} against N for different choices of Θ and ϑ for the particular case of $n = 8$. These figures are identical to Figs. 9 and 10, except that they show σ_{fpt} rather than τ_{mfpt}. We note that, as expected, there is more noise in the simulation results for σ_{fpt} compared to τ_{mfpt}. Overall, we again see extremely good or excellent agreement between analytical and simulation results. We see discrepancies between simulation and Fokker-Planck results for $\Theta = 2$ at larger values of N. To determine whether these are due to our approximations or are intrinsic to the diffusion approximation of the Fokker-Planck equation itself, we have taken the integral equation results to larger values of N for $\Theta = 2$. We see that the integral equation results agree extremely well with the simulation results at larger N for $\vartheta > 0$. The deviations observed in the Fokker-Planck results


A
Θ=2
B
Θ=4
C
Θ=6
D
Θ=8
E
Θ=10
F
Θ=12
rτmfpt
N

are therefore due to the diffusion approximation, which will be more significant as N increases and for smaller values of Θ, for which jump processes will be more significant. For $\vartheta = 0$, σ_{fpt} always asymptotes to a non-zero constant for large enough N. For $\Theta = 10$ and $\Theta = 12$ in Fig. 11, we would need to take N implausibly large to see this, although in Fig. 12 these cases attain their asymptotic behaviours more rapidly. For $\vartheta > 0$, the variance in the FPT always goes to zero for N large enough, at least according to the Fokker-Planck approach, but the manner in which it goes to zero is strongly dependent on Θ and n, at least in Fig. 11. Indeed, again, we see that for $\tau_\vartheta = \tau_{\text{peak}}$ in Fig. 12, σ_{fpt} is overall more stable and better behaved than for $\tau_\vartheta = 0$ s in Fig. 11. For example, for $\tau_\vartheta = 0$ s, σ_{fpt} can rise significantly before finally asymptoting (for both $\vartheta = 0$ and $\vartheta > 0$), while for $\tau_\vartheta = \tau_{\text{peak}}$, any increases tend to be much smaller. Furthermore, for $\tau_\vartheta = 0$ s, σ_{fpt} can be maintained at much higher values for larger values of N compared to the $\tau_\vartheta = \tau_{\text{peak}}$ case. Although for larger ϑ the results for $\tau_\vartheta = 0$ s do in general asymptote to zero more quickly than for $\tau_\vartheta = \tau_{\text{peak}}$, this is only because τ_{mfpt} also goes to zero for $\tau_\vartheta = 0$ s in these cases, while for $\tau_\vartheta = \tau_{\text{peak}}$, τ_{mfpt} asymptotes to a non-zero constant.

Finally, in Figs. 13 and 14 for $\tau_\vartheta = 0$ s and $\tau_\vartheta = \tau_{\text{peak}}$, respectively, we plot the one stardard deviation region defined by σ_{fpt} around the MFPT τ_{mfpt}, as a function of N, for $n = 8$. To avoid clutter, we restrict to $\vartheta = 0.0000$, 0.0025,

---

Figure 9 (previous page): Mean first passage memory lifetimes (with $\tau_\vartheta = 0$ s) for filter-based synapses plotted as a function of N, for various choices of Θ and ϑ, and with $n = 8$ in all cases. Reading from top to bottom in each panel, the data sets correspond to $\vartheta = 0.0000$, 0.0025, 0.0050, 0.0075, 0.0100, 0.0125 and 0.0150. Black lines and circles show simulation results, while red and blue circles show analytical results from the Fokker-Planck and integral equation approaches, respectively.

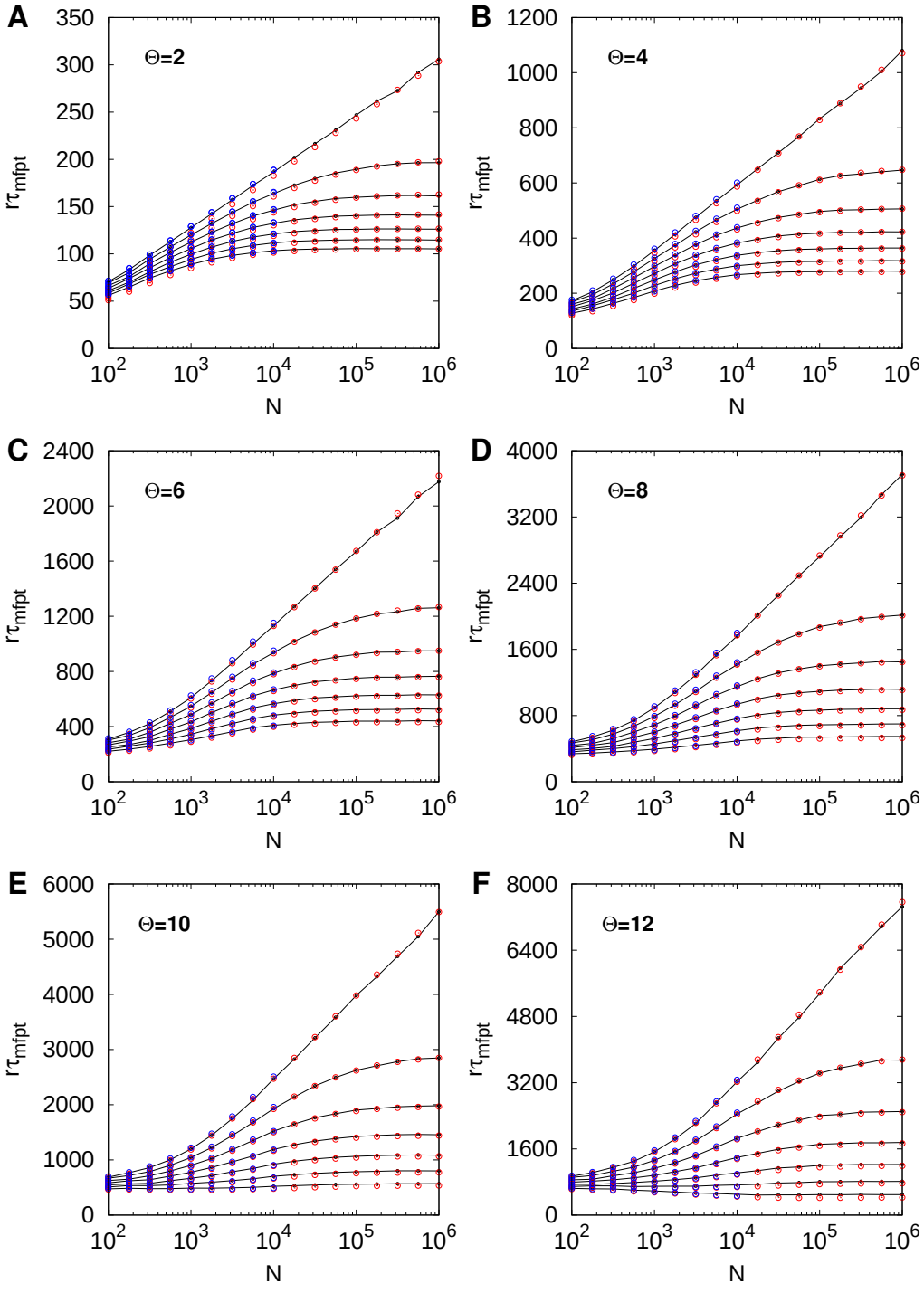


Figure 10: Mean first passage memory lifetimes (with $\tau_\vartheta = \tau_{\text{peak}}$) for filter-based synapses plotted as a function of N. The format of this figure is otherwise identical to Fig. 9

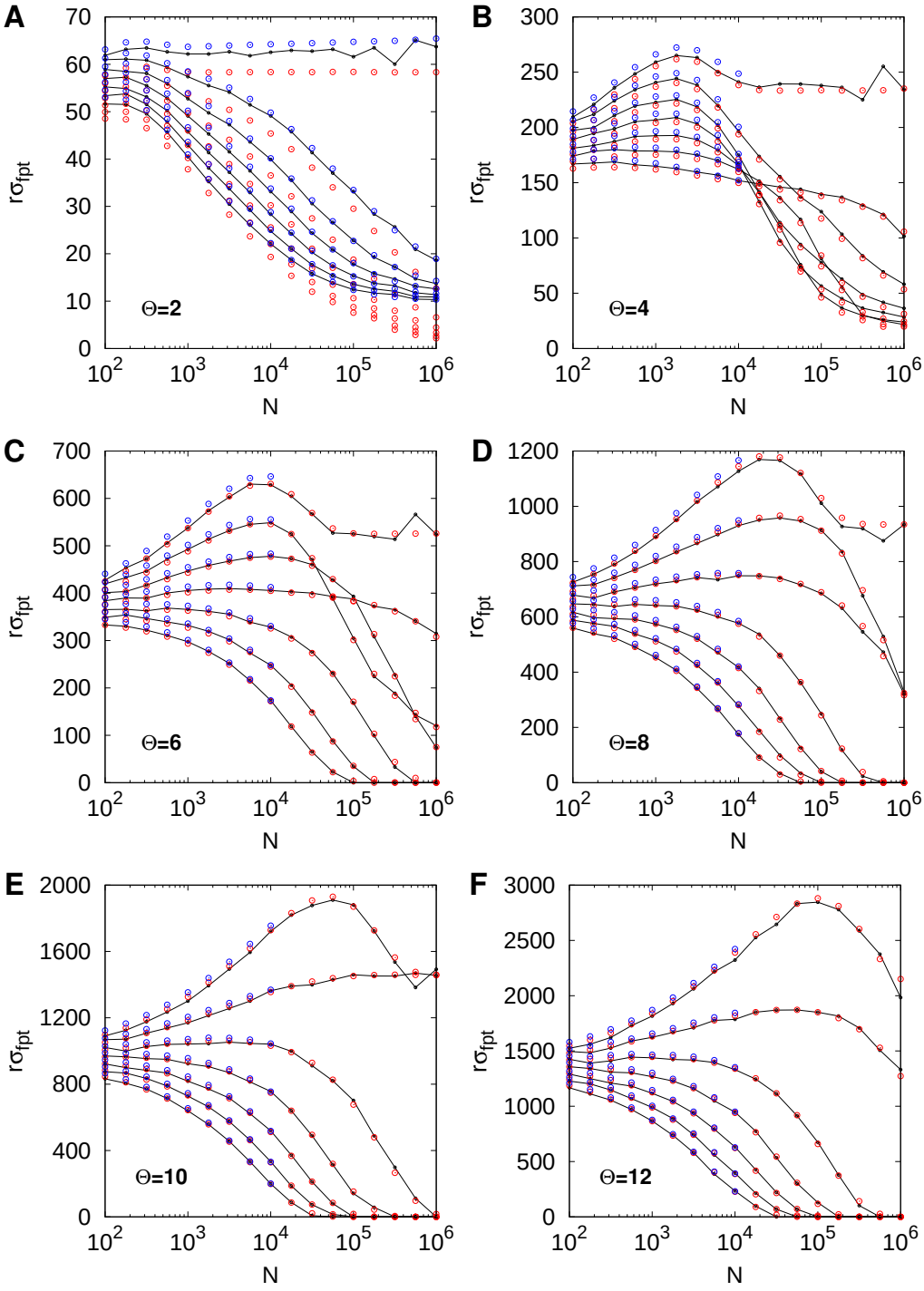


Figure 11: Standard deviation in the first passage time memory lifetimes (with $\tau_\vartheta = 0$ s) for filter-based synapses plotted as a function of N, for various choices of Θ and ϑ, and with $n = 8$ in all cases. The format of this figure is otherwise identical to Fig. 9, except that we also show results for the integral equation approach for $\Theta = 2$ up to $N = 10^6$ rather than the standard $N = 10^4$.

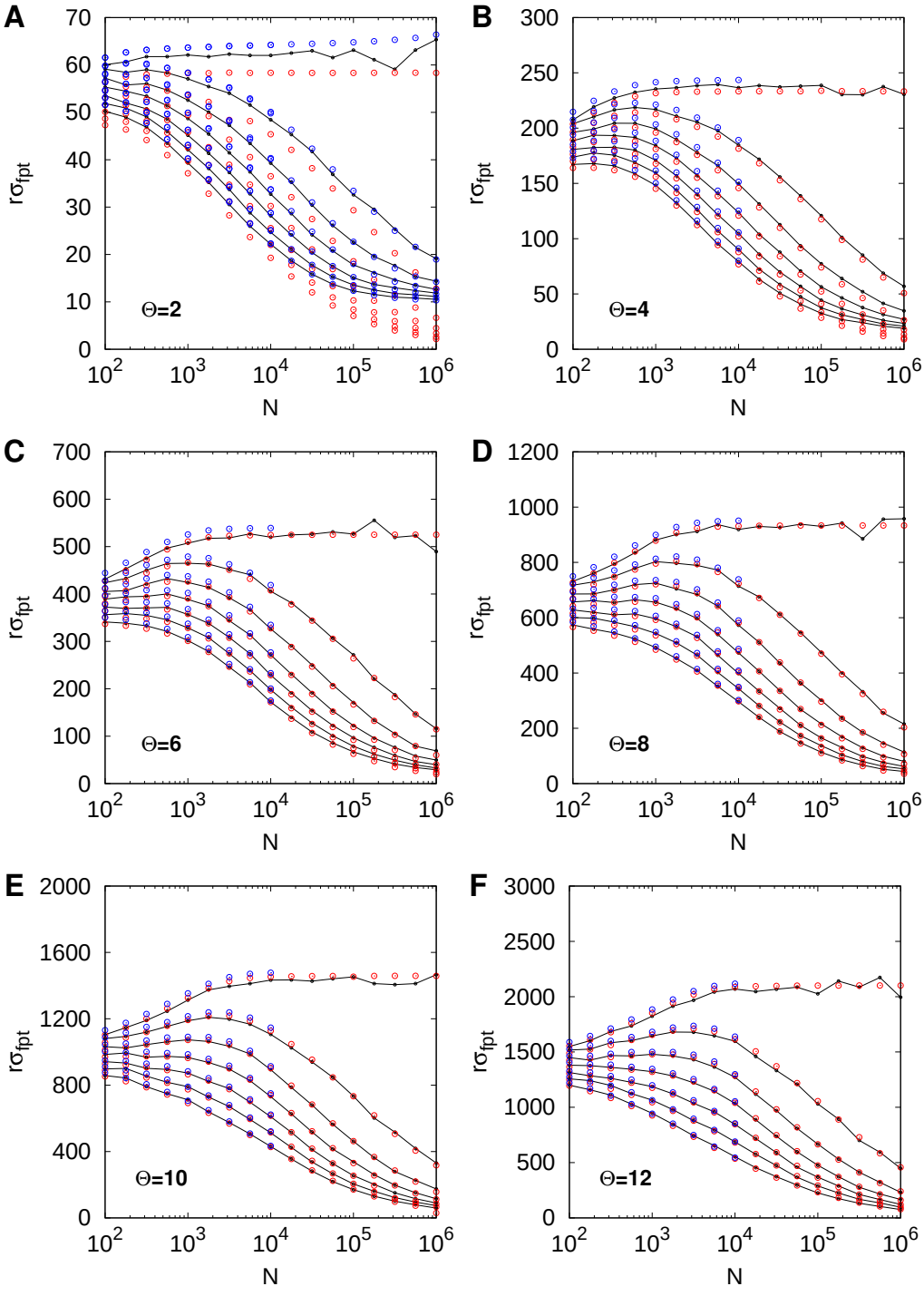


Figure 12: Standard deviation in the first passage time memory lifetimes (with $\tau_\vartheta = \tau_{\text{peak}}$) for filter-based synapses plotted as a function of N. The format of this figure is otherwise identical to Fig. 11.

0.0050 and 0.0075, and show results only from the Fokker-Planck approach. For $\tau_\vartheta = 0$ s in Fig. 13, when $\mu_1(\tau_\vartheta) > \vartheta$, $\tau_{\rm mfpt}$ is robustly positive, and indeed eventually grows logarithmically with N for $\vartheta = 0$. Thus, on average in these regimes, memories can have substantial lifetimes. However, the variability in memory lifetimes, indicated by the 1σ regions, can be significant in some parameter regimes. In general, provided that $\mu_1(\tau_\vartheta) > \vartheta$, increasing N always ensures that $\tau_{\rm mfpt}$ is not swamped by $\sigma_{\rm fpt}$, so that not only are memory lifetimes rebustly positive on average, but the variability in them is, relatively speaking, low. But, achieving this suppression of the variability can require taking N implausibly large, for this particular case of $\tau_\vartheta = 0$ s. However, for $\tau_\vartheta = \tau_{\rm peak}$ in Fig. 14, we see a striking difference. Because of the dynamics of $\tau_{\rm mfpt}$ and $\sigma_{\rm fpt}$ discussed in relation to Figs. 10 and 12, memory lifetimes do not fall to zero for $\vartheta > 0$ and larger values of Θ, and the variance in them is better behaved than for $\tau_\vartheta = 0$ s, not being subject to large increases and not being sustained at high levels for large values of N. What is particularly noteworthy about the results in Fig. 14 is that in each panel for fixed Θ, the value of N for which the entire 1σ region around the mean lifts above zero, is roughly independent of ϑ. In Fig. 13 this is emphatically not the case, with larger ϑ requiring larger N, but only when $\tau_{\rm mfpt}$ remains non-zero. Although MFPT memory lifetimes are robustly positive below this threshold value of N in Fig. 14, above it, the relative variability in them reduces significantly, compared to Fig. 13. Morever, even for $\Theta = 12$, the threshold value of N is only around a few thousand synapses (for $n = 8$), whereas in Fig. 13, even for $\vartheta = 0$, we require N in excess of 10^4 synapses.

7 Discussion

In a series of papers, we have incrementally developed an FPT approach to memory lifetimes for synapses with discrete strengths: first, for simple bistate

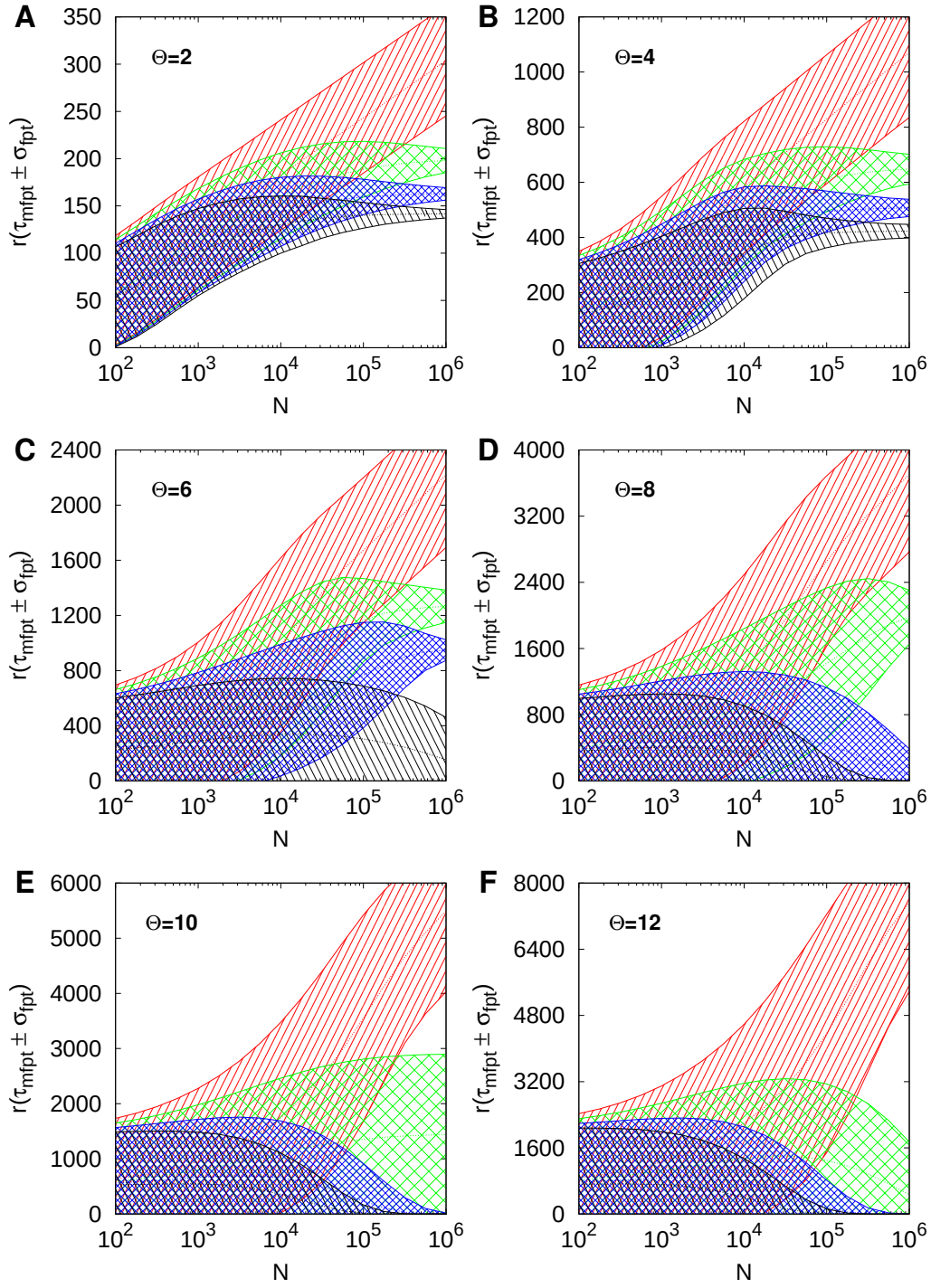


Figure 13: One standard deviation region around the mean first passage memory lifetime (with $\tau_\vartheta = 0$ s) for filter-based synapses plotted as a function of N, for various choices of Θ and ϑ, and with $n = 8$ in all cases. The format of this figure is essentially identical to Fig. 9, except that we consider only $\vartheta = 0.0000$ (red regions), 0.0025 (green regions), 0.0050 (blue regions) and 0.0075 (black regions).

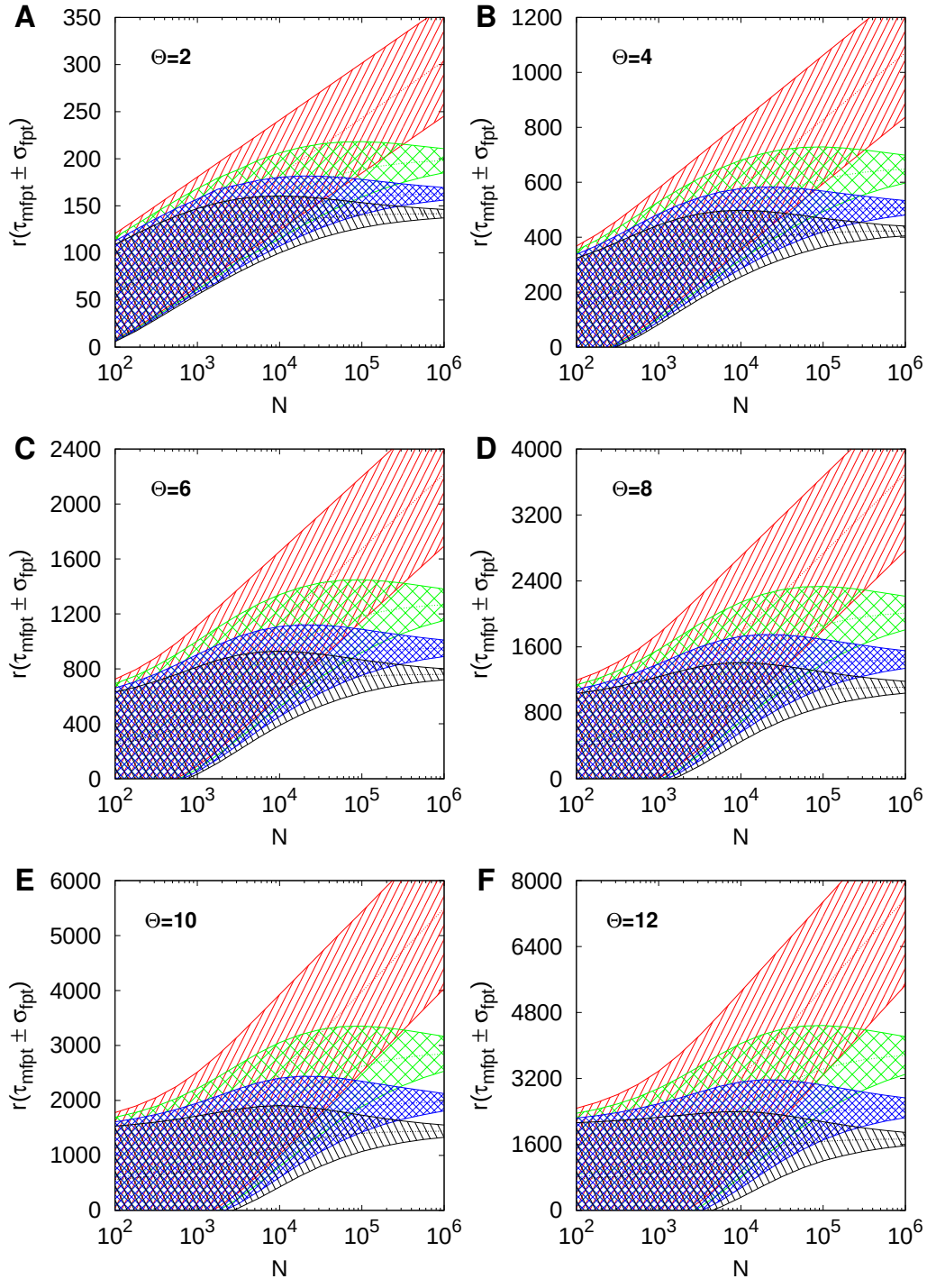


Figure 14: One standard deviation region around the mean first passage memory lifetime (with $\tau_\vartheta = \tau_{\text{peak}}$) for filter-based synapses plotted as a function of N. The format of this figure is otherwise identical to Fig. 13.

synapses (Elliott, 2014); second, for complex (filter-based) bistate synapses (Elliott, 2017b); third, for simple multistate synapses (Elliott, 2017a, 2019). The present paper completes this program of work by considering complex (filter-based) multistate synapses. While the more traditional, SNR approach to memory lifetimes (Tsodyks, 1990) is certainly analytically much easier in comparison, we have shown in the past (Elliott, 2014, 2016a, 2017a,b, 2019), and once again confirmed here, that the analytical effort of applying the FPT approach pays dividends. For example: memory lifetimes do not, in fact, grow logarithmically with the number of synapses, unless the perceptron's firing threshold is precisely (and implausibly) tuned to match its equilibrium mean memory signal. Or: optimality conditions are never actually realised, because these conditions assume asymptotic behaviour, which is never attained when those conditions are satisfied.

In applying the FPT approach to filter-based, multistate synapses, we first integrated out the internal filter states to construct TDSU synapses, which are simple synapses whose probabilities of undergoing changes in strength during subsequent memory storage events are time- or memory-storage-step dependent. Without this simplification of the underlying synaptic dynamics, it would have been impossible to develop the approximation methods required to apply the FPT approach. In particular, a filter-based TDSU synapse can be seen to become an effective SU synapse, with $p = 1/\Theta^2$, on a timescale that is much faster than FPT processes. This "collapse" of synaptic processes essentially ensures that the underlying dynamics can be cleanly separated into pre-peak and post-peak phases, with the post-peak dynamics approximated as a simple SU. Given the initial domination of synaptic dynamics by the rise in the mean memory signal, the pre-peak dynamics are also safely approximated by drift-only processes, meaning that it is not necessary to solve the full Fokker-Planck equation but just the Liouville equation in the pre-peak regime.

The remarkable agreement of this approximation scheme with simulation results demonstrates that filter-based synapses can to an extremely good approximation be regarded as simple, stochastic updaters beyond memory peak (i.e., that the internal filter states have essentially equilibrated by this time). What really distinguishes a filter-based synapse from an SU synapse is the initial, pre-peak transient phase driven by the presence of relatively short-lived inhomogeneous terms that rapidly drive the system away from, rather than towards, equilibrium. The approximation is therefore not, as it were, merely mathematical in character, but in fact captures the fundamental features of the underlying dynamics. It essentially suffices to replace a filter-based synapse, at memory signal peak, by its drift-only distribution, which arises from the initial transient, and by subsequent SU dynamics, which returns the system to equilibrium.

Elsewhere we have discussed the fact that many approaches to defining and analysing memory lifetimes appear to be predicated on the implicit assumption that the mean tracked memory signal decays monotonically over time (Elliott, 2016b). Given that all other, non-integrative models do exhibit only monotonic decay, perhaps this is not too surprising. In applying an SNR criterion to define memory lifetimes in filter-based synapses, we have explicitly allowed for the possibility that the tracked memory's initial SNR may be below unity, but that it may subsequently rise above unity as the mean signal rises (Elliott & Lagogiannis, 2012; Elliott, 2016a,b), and above we defined the SNR memory lifetime as precisely the largest finite solution of the SNR equation. In moving to a FPT definition of memory lifetimes, however, matters are somewhat more subtle. If we average over only that part of the initial distribution at $t = 0$ s that is above firing threshold in order to obtain unconditional FPT statistics, then we throw away that part of the distribution that will, on average, rise above threshold at some later time and contribute to memory recall. In

our earlier simulations of filter-based synapses, we took precisely this "naïve" approach to FPT memory lifetimes (Elliott, 2016a, 2017b), although in our analytical work on bistate TDSU synapses we did consider the pre- and post-peak periods in obtaining very rough estimates of FPT memory lifetimes (Elliott, 2017b). The result is that in our earlier simulation-based work, we both underestimated mean FPT memory lifetimes and overestimated the variance in them.

The pre-peak drift-only and post-peak matched SU approximations permit us very easily to overcome this shortcoming. Instead of applying the threshold condition to the initial distribution at $t = 0$ s, it is applied to the distribution at $t = \tau_{\text{peak}}$, both analytically and in simulation. This is, however, a methodology that is based on an approximation scheme, and in principle we would like to liberate the passage time calculation of memory lifetimes from this method. To achieve this, intuitively it appears that we should consider three classes of process: 1) those realisations that are sub-threshold at $t = 0$ s and remain sub-threshold; 2) those that are sub-threshold at $t = 0$ s and become supra-threshold later, before becoming sub-threshold again; 3) those that are supra-threshold at $t = 0$ s. The third class requires a standard FPT calculation, with a downward transition through threshold. The second requires an initial, upward passage through threshold followed by a second, downward passage. Unfortunately, such a classification blurs the distinction between drift and diffusion processes. For example, any initially sub-threshold realisation is evitably subject to fluctuations of any size if we wait long enough, so it is impossible to know whether an upward transition through threshold is due to drift or diffusion. Thus, while this classification is intuitive, it is intuitive precisely because it is based on our understanding of how filter-based dynamics operate; and, after all, the approximation methods used here deploy exactly this understanding to obtain FPT memory lifetimes. This discussion perhaps reflects

the more general difficulty of trying to define memory lifetimes in models in which the memory signal is non-monotonic, and indeed could vary systematically up and down over days, months and years. Perhaps even trying to define a generic approach to memory lifetimes in a model-independent manner represents a fundamental failure to accept the full complexity of memory, and memory systems, in biological systems, and is just a modeller's fantasy?

Acknowledgements

I acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.

References

Amit, D.J., & Fusi, S. 1994. Learning in neural networks with material synapses. *Neural Comput.*, **6**, 957–982.

Appleby, P.A., & Elliott, T. (2006). Stable competitive dynamics emerge from multispike interactions in a stochastic model of spike-timing-dependent plasticity. *Neural Comput.*, **18**, 2414–2464.

Barrett, A.B., & van Rossum, M.C.W. 2008. Optimal learning rules for discrete synapses. *PLoS Comput. Biol.*, **4**, e1000230.

Bienenstock, E.L., Cooper, L.N., & Munro, P.W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *J. Neurosci.*, **2**, 32–48.

Burkitt, A.N., Meffin, H., & Grayden, D.B. (2004). Spike-timing-dependent plasticity: The relationship to rate-based learning for models with weight dynamics determined by a stable fixed point. *Neural Comput.*, **16**, 885–940.

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. 2011. The mean time to express synaptic plasticity in stochastic, integrate-and-express models of synaptic plasticity induction. *Neural Comput.*, **23**, 124–159.

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. 2017a. First passage time memory lifetimes for simple, multistate synapses. *Neural Comput.*, **29**, 32193259.

Elliott, T. 2017b. Mean first passage memory lifetimes by reducing complex synapses to simple synapses. *Neural Comput.*, **29**, 14681527.

Elliott, T. 2019. First passage time memory lifetimes for simple, multistate synapses: Beyond the eigenvector requirement. *Neural Comput.*, **31**, 8–67.

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.

Ferrell, J.E. 1996. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends Biochem. Sci.*, **21**, 460–466.

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. *Cereb. Cortex*, **18**, 67–77.

Lisman, J., & Zhabotinsky, A.M. 2001. A model of synaptic memory: A CaMKII/PP1 switch that potentiates transmission by organizing an AMPA receptor anchoring assembly. *Neuron*, **31**, 191–201.

Nadal, J.P., Toulouse, G., Changeux, J.P., & Dehaene, S. 1986. Networks of formal neurons and memory palimpsests. *Europhys. Lett.*, **1**, 535–542.

Napper, R.M., & Harvey, R.J. 1988. Number of parallel fiber synapses on an individual Purkinje cell in the cerebellum of the rat. *J. Comp. Neurol.*, **274**, 168–177.

Parisi, G. 1986. A memory which forgets. *J. Phys. A: Math. and Gen.*, **19**, L617–L620.

Pi, H.J., & Lisman, J.E. 2008. Coupled phosphatase and kinase switches produce the tristability required for long-term potentiation and long-term depression. *J. Neurosci.*, **28**, 13132–13138.

Rubin, D.D.B.D., & Fusi, S. 2007. Long memory lifetimes require complex synapses and limited sparseness. *Front. Comput. Neurosci.*, **1**, 7.

Tsodyks, M.V. 1990. Associative memory in neural networks with binary synapses. *Mod. Phys. Lett. B*, **4**, 713–716.

Tsodyks, M.V., & Feigel'man, M.V. 1988. The enhanced storage capacity in neural networks with low activity levels. *Europhys. Lett.*, **6**, 101–105.

Uhlenbeck, G.E., & Ornstein, L.S. 1930. On the theory of Brownian motion. *Phys. Rev.*, **36**, 823–841.

van Kampen, N.G. 1992. *Stochastic Processes in Physics and Chemistry.* Amsterdam: Elsevier.