First Passage Time Memory Lifetimes for Simple, Multistate Synapses: Beyond the Eigenvector Requirement

Terry Elliott[1]

Department of Electronics and Computer Science,

University of Southampton,

Highfield,

Southampton, SO17 1BJ,

United Kingdom.

Running Title: FPT memory lifetimes.

July 4, 2018.

[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-strength synapses are palimpsests, learning new memories by forgetting old ones. Memory lifetimes can be defined by the mean first passage time (MFPT) for a perceptron's activation to fall below firing threshold. By imposing the condition that the vector of possible strengths available to a synapse is a left eigenvector of the stochastic matrix governing transitions in strength, we previously derived results for MFPTs and first passage time (FPT) distributions in models with simple, multistate synapses. This condition permits jump moments to be computed via a 1-dimensional Fokker-Planck approach. Here, we study memory lifetimes in the absence of this condition. To do so we must introduce additional variables, including the perceptron activation, that parametrise synaptic configurations, permitting Markovian dynamics in these variables to be formulated. FPT problems in these variables require solving multi-dimensional partial differential or integral equations. However, the FPT dynamics can be analytically well approximated by focusing on the slowest eigenmode in this higher-dimensional space. We may also obtain a much better approximation by restricting to the two dominant variables in this space, the restriction making numerical methods tractable. Analytical and numerical methods are in excellent agreement with simulation data, validating our methods. These methods prepare the ground for the study of FPT memory lifetimes with complex rather than simple, multistate synapses.

1 Introduction

Associative memories may be stored in a Hopfield network up to a critical loading capacity, beyond which the network catastrophically forgets all memories (Hopfield, 1982; Hertz *et al.*, 1991). However, models that bound synaptic strengths learn new memories by forgetting old ones, creating so-called "palimpsest" memory systems (Nadal *et al.*, 1986; Parisi, 1986). Such bounds could be realised if synapses exist in only a finite number of discrete states of synaptic strength, for which some experimental evidence exists (Petersen *et al.*, 1998; Montgomery & Madison, 2002, 2004; O'Connor *et al.*, 2005a,b; Bartol *et al.*, 2015). Palimpsest models with discrete state synapses in either a feedforward or a recurrent framework have therefore attracted much attention (e.g., Tsodyks, 1990; Amit & Fusi, 1994, Leibold & Kempter, 2006, 2008; Barrett & van Rossum, 2008; Huang & Amit, 2010, 2011; Elliott & Lagogiannis, 2012; Lahiri & Ganguli, 2013). Palimpsest memory lifetimes have been defined using a signal-to-noise criterion or an entirely equivalent ideal observer approach (Tsodkys, 1990; Fusi *et al.*, 2005; Lahiri & Ganguli, 2013; see Elliott, 2016b for their equivalence); signal detection theory (Leibold & Kempter, 2006, 2008); and retrieval probabilities (Huang & Amit, 2010, 2011).

We have introduced a first-passage-time (FPT) approach to defining memory lifetimes in a feedforward setting with a single perceptron for simplicity (Elliott, 2014, 2017a,b). Specifically, we define a memory's lifetime as the mean first passage time (MFPT) for the perceptron's activation to fall below firing threshold, and we may also consider the variability in this lifetime by examining the higher-order statistics of the FPT distribution. Our earlier work was restricted to two-state, binary strength synapses (Elliott, 2014), but recently we considered simple synapses with multiple states of strength (Elliott, 2017b).

To perform this latter analysis, in the context of a Fokker-Planck approach we required that the vector of possible strengths available to a synapse is a left eigenvector of the stochastic matrix defining possible transitions between a synapse's strength states. Using an approximation based on the Ornstein-Uhlenbeck (OU) process (Uhlenbeck & Ornstein, 1930), we derived results for the FPT distribution in extremely good agreement with simulation data.

The eigenvector requirement limits the application of our previous approach. It prevented us from analysing two models discussed in detail below that do not satisfy this condition. Moreover, it prevents us from extending the approach from simple to complex synapses with internal states, where integrating out the internal states leads to time-dependent transition matrices whose eigenvectors will therefore also be time-dependent (Elliott, 2017a). Here, therefore, we develop an alternative approach to the analysis of FPT memory lifetimes that dispenses with the eigenvector requirement. To do so, we must move from dynamics in the 1-dimensional space of the perceptron's activation to dynamics in a $(\nu - 1)$-dimensional space, where ν is the number of states of strength available to a synapse. These extra variables permit a formulation of purely Markovian dynamics in this higher-dimensional space. Working in this larger space is analytically and numerically extremely difficult. However, with good, controlled approximations, half these additional variables decouple, and the dynamics become those of a multi-dimensional OU process. By considering the slowest eigenmode, we obtain a good and only 1-dimensional analytical approximation for computing FPT moments. This approximation breaks down for initial states close to threshold, but to obtain a much better approximation that also works close to threshold, we can restrict to the two dominant variables in the dynamics. Although still analytically problematic, restricting to just these two variables makes the use of numerical methods for

computing FPT moments tractable.

Our paper is organised as follows. In section 2 we review our general approach, discussing our framework, models of plasticity and earlier FPT calculations using the eigenvector condition. Then, in section 3 we extend the dynamics to $\nu-1$ variables, derive the jump moments in the multi-dimensional Fokker-Planck equation, and discuss FPT processes in these variables, including exact results for FPTs. Section 4 approximates these full dynamics by a multi-dimensional OU process, and then develops analytically and numerically tractable approximation methods involving the slowest eigenmode or the two dominant variables. We present a comparison of numerical, analytical and simulation results in section 5. Our approach and future work are discussed in section 6.

2 General Approach

We provide an outline of our general approach to set up the problem, establish notation and conventions, and summarise earlier results. Full details may be found elsewhere (Elliott, 2014, 2016a, 2017b).

2.1 Perceptron Memory

A single perception is required to store a sequence of memories $\boldsymbol{\xi}^\alpha$, $\alpha = 0, 1, 2, \ldots$. The memories are stored sequentially at times governed by a Poisson process of rate r, although we may without loss of generality set $r = 1$ Hz here. We take by convention $\boldsymbol{\xi}^0$ to be stored at time $t = 0^-$ s, so that the time immediately after its storage is $t = 0$ s. Because the storage of later memories modifies the patterns of synaptic strengths embodying earlier memories, memories may degrade over time. To determine the lifetime of a typical memory,

we track the fidelity of recall of memory $\boldsymbol{\xi}^0$ and determine when, on average, the perceptron's activation in response to re-presentation of $\boldsymbol{\xi}^0$ falls below the perceptron's firing threshold, θ. Thus, specifically, we seek to determine the MFPT for the perceptron's activation in response to $\boldsymbol{\xi}^0$ to fall (to or) below θ.

The perceptron has N synapses of strengths $S_i(t)$, $i = 1, \ldots, N$, where these strengths are discrete, with $S_i(t) \in \{\Omega_1, \ldots, \Omega_\nu\}$, so taking $\nu \geq 2$ possible values. These values $\Omega_1 < \Omega_2 < \cdots < \Omega_\nu$ are ordered lowest (weakest) to highest (strongest). We write $\boldsymbol{\Omega} = (\Omega_1, \ldots, \Omega_\nu)^{\mathrm{T}}$, where a superscript T denotes the transpose. With input vector $\boldsymbol{x}$ with components x_i through these N synapses, the perceptron's activation or unthresholded output is the standard form

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

Any non-linearity in generating the perceptron's output from $h_{\boldsymbol{x}}(t)$ is for our purposes here irrelevant, because the memory $\boldsymbol{\xi}^0$ is stored at some future time $t \geq 0$ s provided that $h_{\boldsymbol{\xi}^0}(t) > \theta$. We define $h(t) = h_{\boldsymbol{\xi}^0}(t)$ as the tracked memory signal.

The components ξ_i^α of memory $\boldsymbol{\xi}^\alpha$ are the inputs to the perceptron when the memory is presented either for storage or to gauge the fidelity of its recall. We take these components to be binary-valued, $\xi_i^\alpha \in \{-1, +1\}$, with probabilities $\mathsf{Prob}[\xi_i^\alpha = \pm 1] = g_\pm$, where $g_+ + g_- = 1$. We may consider $g_\pm \neq \frac{1}{2}$, but it is convenient to restrict to the balanced scenario in which $g_\pm = \frac{1}{2}$. If we require a target perceptron activation $h_{\boldsymbol{\xi}^\alpha}(t) > \theta \geq 0$ when $\boldsymbol{\xi}^\alpha$ is stored, then ξ_i^α is the plasticity induction signal to synapse i at memory storage. If $\xi_i^\alpha = +1$ (respectively, $\xi_i^\alpha = -1$), then the synapse should potentiate or strengthen (respectively, depress or weaken). With $g_\pm = \frac{1}{2}$, potentiation and depression are equiprobable and treated symmetrically. For simplicity we take

the induction signals to be uncorrelated (and independent) between synapses and across memories.

Synaptic plasticity, leading to the storage of memories, is a stochastic process driven by the Poisson arrival times of the stochastic plasticity induction signals ξ_i^α. The strength $S_i(t)$ of synapse i is therefore a random variable governed by a probability distribution. This distribution is represented by a ν-dimensional vector with components $\mathsf{Prob}[S_i(t) = \Omega_j]$, $j = 1, \ldots, \nu$. Stochastic transitions in a synapse's strength under potentiation (respectively, depression) are implemented by the $\nu \times \nu$ matrix $\mathbb{M}^+$ (respectively, $\mathbb{M}^-$). To average over all non-tracked memories, with $\alpha > 0$, we consider the superposed transition matrix

$$\mathbb{M} = g_+\mathbb{M}^+ + g_-\mathbb{M}^- = \tfrac{1}{2}\left(\mathbb{M}^+ + \mathbb{M}^-\right) \tag{2.2}$$

for $g_\pm = \frac{1}{2}$. For a Poisson memory storage process, the probability distribution $\boldsymbol{p}(t)$ of any synapse's strength evolves according to the equation

$$\boldsymbol{p}(t) = \exp\left[rt\left(\mathbb{M} - \mathbb{I}\right)\right]\boldsymbol{p}(0), \tag{2.3}$$

where $\mathbb{I}$ is the identity matrix and $\boldsymbol{p}(0)$ is its distribution immediately after the storage of the tracked memory $\boldsymbol{\xi}^0$. In general, the distribution $\boldsymbol{p}(t)$ evolves to the equilibrium distribution governed by the eigenvector of $\mathbb{M}$ with unit eigenvalue. For the two forms of $\mathbb{M}$ considered below, this eigenvector, normalised to a probability distribution, is

$$\boldsymbol{A} = \tfrac{1}{\nu}\mathbf{1} = \tfrac{1}{\nu}(1, \ldots, 1)^{\mathrm{T}}, \tag{2.4}$$

where $\mathbf{1}$ is a ν-dimensional vector with all components unity. In equilibrium, all possible strengths Ω_j of a synapse are therefore equiprobable. It is against

the background of this equilibrium distribution that the tracked memory $\boldsymbol{\xi}^0$ is taken to be stored. For synapses experiencing $\xi_i^0 = +1$ (respectively, $\xi_i^0 = -1$), we have that $\boldsymbol{p}(0) = \mathbb{M}^+\boldsymbol{A}$ (respectively, $\boldsymbol{p}(0) = \mathbb{M}^-\boldsymbol{A}$). Just as we average over all non-tracked memories, we must also average over all possible tracked memories $\boldsymbol{\xi}^0$. Therefore, for $g_\pm = \frac{1}{2}$, all synapses are considered to be in an equiprobable mixture of states governed by the two distributions $\mathbb{M}^\pm\boldsymbol{A}$ at time $t = 0$ s.

When potentiation and depression are treated completely symmetrically, and assuming that the possible synaptic strengths are themselves symmetrically arranged around zero so that $\Omega_i = -\Omega_{\nu+1-i}$ (i.e. $-\Omega_1 = +\Omega_\nu$, $-\Omega_2 = +\Omega_{\nu-1}$, etc.), we have shown previously that while the two conditional distributions $\mathsf{Prob}[S_i(t) = \Omega_j \,|\, \xi_i^0 = \pm 1]$ differ, the two conditional distributions $\mathsf{Prob}[\xi_i^0 S_i(t) = \Omega_j \,|\, \xi_i^0 = \pm 1]$ are identical for all times $t \geq 0$ s. This is true whether a synapse is simple or complex (Elliott, 2016b). Since

$$h(t) = \frac{1}{N}\sum_{i=1}^{N} \xi_i^0 S_i(t), \tag{2.5}$$

the tracked memory signal is therefore a sum over N identically distributed random variables for balanced potentiation and depression processes. It is convenient to define

$$\widetilde{S}_i(t) = \xi_i^0 S_i(t), \tag{2.6}$$

so that $h(t) = \frac{1}{N}\sum_{i=1}^{N} \widetilde{S}_i(t)$, and to focus on these tilded strength variables rather than the untilded strengths. Critically, their initial distribution is not a mixture because they are identical. We may in effect simply set $\xi_i^0 = +1\ \forall i$ below (i.e. set $\boldsymbol{p}(0) = \mathbb{M}^+\boldsymbol{A}\ \forall i$), with the understanding that when we refer to a synapse's strength, we actually refer to its tilded strength.

2.2 Models of Synaptic Plasticity

Previously we considered two different models of synaptic plasticity. These models are defined by the two matrices $\mathbb{M}^{\pm}$.

For the simplest "stochastic updater" synapse, a synapse increases its strength from Ω_j to Ω_{j+1} for $j < \nu$ with probability p when experiencing a potentiating induction signal; if $j = \nu$, the synapse is saturated at its highest strength and cannot increase its strength further. When experiencing a depressing induction signal, a synapse's strength decreases from Ω_j to Ω_{j-1} for $j > 1$ with probability p; if $j = 1$, the synapse is saturated at its lowest strength and cannot decrease its strength further. Writing $\mathbb{M}^{\pm} = \mathbb{I} + p\,\mathbb{C}^{\pm}$, we have that

$$\mathbb{C}^{+} = -\mathrm{diag}\{\underbrace{1,\ldots,1}_{\nu-1},0\} + \mathrm{diag}_{\mathrm{l}}\{\underbrace{1,\ldots,1}_{\nu-1}\}, \tag{2.7a}$$

$$\mathbb{C}^{-} = -\mathrm{diag}\{0,\underbrace{1,\ldots,1}_{\nu-1}\} + \mathrm{diag}_{\mathrm{u}}\{\underbrace{1,\ldots,1}_{\nu-1}\}, \tag{2.7b}$$

where $\mathrm{diag}_{\mathrm{u}}$ and $\mathrm{diag}_{\mathrm{l}}$ refer to the upper and lower diagonals, respectively. The superposed matrix $\mathbb{M} = \frac{1}{2}\left(\mathbb{M}^{+} + \mathbb{M}^{-}\right) = \mathbb{I} + p\,\mathbb{C}$, where the matrix $\mathbb{C}$,

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

implements a symmetric, unbiased random walk on the strengths $\{\Omega_1,\ldots,\Omega_\nu\}$ between two reflecting boundaries. The particular case $\nu = 2$ and $\boldsymbol{\Omega} = (-1,+1)^{\mathrm{T}}$ has been studied before (Tsodyks, 1990), as well as generalisations to $\nu \geq 2$ with strengths Ω_j uniformly spaced between $\Omega_1 = -1$ and $\Omega_\nu = +1$ (Amit & Fusi, 1994; Fusi & Abbott, 2007; Elliott, 2016a). The spectrum of $\mathbb{M}$

is standard. Its orthogonal eigenvectors $\boldsymbol{\varepsilon}^m$, $m = 0, \ldots, \nu-1$, have components

$$\varepsilon_i^m = \frac{\cos \frac{\pi\, m}{2\,\nu}(2\,i-1)}{\cos \frac{\pi\, m}{2\,\nu}(2\,\nu-1)}, \qquad i = 1, \ldots, \nu, \tag{2.9}$$

and corresponding eigenvalues

$$\lambda_m = 1 - p\left(1 - \cos \tfrac{\pi\, m}{\nu}\right). \tag{2.10}$$

We have chosen to normalise the eigenvectors so that $\varepsilon_\nu^m = +1\ \forall m$.

We considered a second plasticity model defined by $\mathbb{M}^{\pm} = \mathbb{I} + p\,\mathbb{Q}^{\pm}$ with

$$\mathbb{Q}^{+} = -\mathrm{diag}\{a_1, \ldots, a_\nu\} + \mathrm{diag}_{\mathrm{l}}\{a_1, \ldots, a_{\nu-1}\}, \tag{2.11a}$$

$$\mathbb{Q}^{-} = -\mathrm{diag}\{a_\nu, \ldots, a_1\} + \mathrm{diag}_{\mathrm{u}}\{a_{\nu-1}, \ldots, a_1\}, \tag{2.11b}$$

where

$$a_i = \frac{i(\nu-i)}{\nu-1}. \tag{2.12}$$

Similarly to $\mathbb{C}$, we define $\mathbb{Q} = \frac{1}{2}\left(\mathbb{Q}^{+} + \mathbb{Q}^{-}\right)$. The elements of $\mathbb{Q}$ are selected to induce a very particular eigenstructure. We can show (Elliott, 2017b) that the matrix $\mathbb{M} = \mathbb{I} + p\,\mathbb{Q}$ has orthogonal eigenvectors $\boldsymbol{\varepsilon}^m$, $m = 0, \ldots, \nu-1$, with components

$$\varepsilon_i^m = \sum_{j=0}^{m} (-1)^j\, b_{\nu-i-j}^m\, {}^m C_j, \qquad i = 1, \ldots, \nu, \tag{2.13}$$

where the symbol b_i^m is shorthand for

$$b_i^m = \frac{{}^{m+i}C_m\ {}^{\nu-1-i}C_m}{{}^{\nu-1}C_m}, \tag{2.14}$$

and ${}^i C_j$ denotes a binomial coefficient with standard conventions for $j < 0$ or

$j > i$ understood. The corresponding eigenvalues are

$$\lambda_m = 1 - p\,\frac{m(m+1)}{2(\nu-1)}. \tag{2.15}$$

We again have the normalisation $\varepsilon_\nu^m = +1\ \forall m$. The elements a_i of $\mathbb{Q}$ are specifically chosen so that $\boldsymbol{\varepsilon}^1$ has components

$$\varepsilon_i^1 = -1 + 2\frac{i-1}{\nu-1}, \tag{2.16}$$

which are uniformly spaced between -1 and $+1$. The overall normalisation of a_i is chosen so that for $\nu = 2$ and $\nu = 3$, $\mathbb{Q} \equiv \mathbb{C}$, so that for these particular values of ν, the two plasticity models are identical. Because $\mathbb{M}$ must be a stochastic matrix, the form of a_i imposes a constraint on ν or p, given by

$$\nu \lessapprox \frac{4}{p} - 1 \quad \text{or} \quad p \lessapprox \frac{4}{\nu+1}. \tag{2.17}$$

The matrix $\mathbb{Q}$ also implements a random walk on the strength states between reflecting boundaries, but unlike $\mathbb{C}$ the transitions probabilities depend on the strength state of the synapse. In particular, a synapse is more likely to potentiate or depress if its strength is close to zero than if its strength is close to saturation.

When discussing general results below, we will refer to $\mathbb{M}$ and $\boldsymbol{\varepsilon}^m$ without specifying the particular form of $\mathbb{M}$. Where necessary, we will write $\mathbb{M}_{\mathrm{C}}$ and $\mathbb{M}_{\mathrm{Q}}$ to distinguish between $\mathbb{M} = \mathbb{I} + p\,\mathbb{C}$ and $\mathbb{M} = \mathbb{I} + p\,\mathbb{Q}$. If we need to distinguish between the two sets of eigenvectors of $\mathbb{M}_{\mathrm{C}}$ and $\mathbb{M}_{\mathrm{Q}}$, we will use $\boldsymbol{s}^m$ and $\boldsymbol{l}^m$, respectively. We also write $\boldsymbol{\mathcal{S}} = \boldsymbol{s}^1$ with $\mathcal{S}_i = \cos\frac{\pi}{2\nu}(2\,i-1)/\cos\frac{\pi}{2\nu}(2\,\nu-1)$ and $\boldsymbol{\mathcal{L}} = \boldsymbol{l}^1$ with $\mathcal{L}_i = -1 + 2(i-1)/(\nu-1)$. We refer to the model with $\mathbb{M} = \mathbb{I} + p\,\mathbb{C}$ (respectively, $\mathbb{M} = \mathbb{I} + p\,\mathbb{Q}$) as the "$\mathbb{C}$ model" (respectively, "$\mathbb{Q}$

model").

2.3 FPT Memory Lifetimes

To determine FPT memory lifetimes, we must examine the dynamics of $h(t)$. Assuming that $h(0) > \theta$, i.e. that the initial storage of $\boldsymbol{\xi}^0$ induces a strong enough initial memory signal, we must determine the probability distribution of the time $t > 0$ s at which $h(t)$ first falls (to or) below θ. The mean of this distribution then gives the MFPT memory lifetime, and its higher order statistics provide information about the variability in this lifetime.

Changes in $h(t)$ are driven by changes in $\widetilde{S}_i(t)$ due to synaptic plasticity induced by ongoing memory storage. In terms of their contribution to $h(t)$, synapses with the same tilded strengths are indistinguishable. For symmetrically arranged strengths with $\Omega_i = -\Omega_{\nu+1-i}$, as assumed here, if $S_i(t) \in \{\Omega_1, \ldots, \Omega_\nu\}$, then so is $\widetilde{S}_i(t)$. Thus, a synaptic configuration is uniquely determined, up to irrelevant permutations, by the number of synapses N_i, $i = 1, \ldots, \nu$, with tilded strengths Ω_i, where $\sum_{i=1}^{\nu} N_i = N$. The number of such configurations is the number of partitions of N into ν parts, or ${}^{N+\nu-1}C_{\nu-1}$. However, for the case for example that $\boldsymbol{\Omega} = \boldsymbol{\mathcal{L}}$, so uniformly or linearly spaced Ω_i with increments of $2/(\nu-1)$, the number of district values of $h(t)$ is just $N(\nu-1)+1$, which is the number of points of uniform spacing $2/[N(\nu-1)]$ between -1 and $+1$, inclusive. The mapping from these ${}^{N+\nu-1}C_{\nu-1}$ distinct synaptic configurations to the $N(\nu-1)+1$ possible values of $h(t)$ is therefore typically massively degenerate, or many-to-one. Only for $\nu = 2$ is the mapping one-to-one, so that a particular value of $h(t)$ uniquely specifies a particular synaptic configuration $\{N_1, N_2 = N - N_1\}$. For $\nu > 2$, a particular value of $h(t)$ does not in general uniquely specify a unique configuration $\{N_1, \ldots, N_\nu\}$.

To study the dynamics of $h(t)$, ideally we require a $[N(\nu-1)+1]\times[N(\nu-1)+1]$ transition matrix (for $\mathbf{\Omega} = \boldsymbol{\mathcal{L}}$) describing the transitions in $h(t)$ at successive non-tracked memory storage steps. The underlying transitions in all N (tilded) synaptic strengths are described by a $\nu^N \times \nu^N$ transition matrix, or a tensor product of N copies of $\mathbb{M}$. Even for $\nu = 2$, these two matrices are typically vastly different in size, with an $(N+1)\times(N+1)$ matrix being manageable even for $N = 10^5$, say, while a $2^N \times 2^N$ matrix is intractably large even for relatively small values of N. But because ${}^{N+\nu-1}C_{\nu-1} = N(\nu-1)+1$ for $\nu = 2$, the transitions in $h(t)$ are Markovian, due to the one-to-one mapping discussed above. Hence, we can compute and use the $(N+1)\times(N+1)$ transition matrix for $h(t)$ to determine MFPTs (Elliott, 2014). However, for $\nu > 2$, the transitions in $h(t)$ alone are not Markovian because of the many-to-one mapping, while of course the underlying transitions in synaptic configurations are Markovian. Essentially, a $[N(\nu-1)+1]\times[N(\nu-1)+1]$ transition matrix for $h(t)$ does not exist in general, in the sense that its powers will not give the correct multistep transitions, because collapsing the dynamics from ${}^{N+\nu-1}C_{\nu-1}$ to just $N(\nu-1)+1$ states loses information.

To overcome this problem for the general case of $\nu \geq 2$, we sought instead to compute the jump moments for $h(t)$, assume a continuum limit, and use these jump moments in a Fokker-Planck equation approach to computing the FPT distribution (Elliott, 2017b). If $P(h,t|h_0,t_0)$ denotes the transition probability from tracked memory signal h_0 at time t_0 (we take $t_0 = 0$ s) to tracked memory signal h at time t, then the Fokker-Planck equation (see, e.g., van Kampen, 1992) is

$$\frac{1}{r}\frac{\partial}{\partial t}P(h,t|h_0,t_0) = -\frac{\partial}{\partial h}\big[A(h)P(h,t|h_0,t_0)\big] + \frac{1}{2}\frac{\partial^2}{\partial h^2}\big[B(h)P(h,t|h_0,t_0)\big], \tag{2.18}$$

where the jump moments are calculated from

$$M_l(x) = \frac{1}{r} \lim_{\delta t \to 0} \frac{1}{\delta t} \int dy\, (y-x)^l P(y, t+\delta t | x, t), \tag{2.19}$$

and $A(x) = M_1(x)$ and $B(x) = M_2(x)$. Using the adjoint or backward equation and with

$$G(h_0, t) = -\frac{\partial}{\partial t} \int_\theta^\infty dh\, P(h, t | h_0, 0) \tag{2.20}$$

being the FPT density for the memory signal to escape from the interval (θ, ∞) for the first time at time t starting from an initial value $h(0) = h_0$, $G(h_0, t)$ satisfies the equation

$$\frac{1}{r}\frac{\partial}{\partial t} G(h_0, t) = A(h_0) \frac{\partial}{\partial h_0} G(h_0, t) + \frac{1}{2} B(h_0) \frac{\partial^2}{\partial h_0^2} G(h_0; t), \tag{2.21}$$

subject to the absorbing boundary condition $G(\theta, t) = \delta(t)$, where $\delta(t)$ is the Dirac delta function. All the moments of the FPT distribution may be computed directly from $G(h_0, t)$. In particular, if $\tau_{\text{mfpt}}(h_0) = \int_0^\infty dt\, t\, G(h_0, t)$ denotes the MFPT conditioned on a specific, initial value of h_0, then $\tau_{\text{mfpt}}(h_0)$ satisfies

$$-\frac{1}{r} = A(h_0) \frac{d}{dh_0} \tau_{\text{mfpt}}(h_0) + \frac{1}{2} B(h_0) \frac{d^2}{dh_0^2} \tau_{\text{mfpt}}(h_0), \tag{2.22}$$

subject to the boundary condition $\tau_{\text{mfpt}}(\theta) = 0$. The unconditional MFPT is then just $\tau_{\text{mfpt}} = \langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0 > \theta}$, where $\langle \cdot \rangle_{h_0 > \theta}$ denotes an average over the initial distribution of h_0 with h_0 above threshold.

To compute $A(x)$ and $B(x)$ explicitly, we used the underlying synaptic dynamics to determine the evolution of $\mu(t) = \mathsf{E}[h(t)]$ and $\varphi(t) = \mathsf{E}[h(t)^2]$, where

$\mathsf{E}[\,\cdot\,]$ denotes the expectation value, and matched them against the equations

$$\frac{1}{r}\frac{d\mu(t)}{dt} = \mathsf{E}\big[A(h)\big], \tag{2.23a}$$

$$\frac{1}{r}\frac{d\varphi(t)}{dt} = \mathsf{E}\big[B(h)\big] + 2\,\mathsf{E}\big[hA(h)\big], \tag{2.23b}$$

which follow directly from the Fokker-Planck equation. To compute $d\mu(t)/dt$ and $d\varphi(t)/dt$ from the underlying synaptic dynamics, it was necessary to take the vector of possible synaptic strengths $\boldsymbol{\Omega}$ to be a left eigenvector of the general transition matrix $\mathbb{M}$, so $\boldsymbol{\Omega}^{\mathrm{T}}\mathbb{M} = \lambda_\Omega \boldsymbol{\Omega}^{\mathrm{T}}$, where λ_Ω is the corresponding eigenvalue. This requirement is necessary to close and solve the system of equations. For $\mathbb{M}_{\mathrm{C}}$, the usual choice of $\boldsymbol{\Omega} = \boldsymbol{\mathcal{L}}$ is not an eigenvector unless $\nu = 2$ or $\nu = 3$. The eigenvector of $\mathbb{M}_{\mathrm{C}}$ "closest" to $\boldsymbol{\mathcal{L}}$ is $\boldsymbol{\mathcal{S}}$, which is a sinusoidal arrangement of strengths exhibiting saturation-like properties at its lower and upper limits. Thus, we considered what we referred to as the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ model. To consider dynamics with $\boldsymbol{\Omega} = \boldsymbol{\mathcal{L}}$, we specifically constructed the transition matrix $\mathbb{M}_{\mathrm{Q}}$ so that it has $\boldsymbol{\mathcal{L}}$ as an eigenvector. We called this model the $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ model. With this eigenvector requirement, we obtained

$$\mathsf{E}\big[A(h)\big] = -\,(1-\lambda_\Omega)\,\mu(t), \tag{2.24a}$$

$$\mathsf{E}\big[B(h)\big] = -\,\big(1-\lambda_\Omega^2\big)\left[\varphi(t) - \frac{1}{N}\langle\boldsymbol{\Omega}^2\rangle\right] - 2\,\mathsf{E}\big[hA(h)\big], \tag{2.24b}$$

where $\langle\boldsymbol{\Omega}^2\rangle$ is a convenient shorthand for $\frac{1}{\nu}\sum_{i=1}^{\nu}\Omega_i^2$. From these two equations we deduced that

$$A(x) = -\,(1-\lambda_\Omega)\,x, \tag{2.25a}$$

$$B(x) = (1-\lambda_\Omega)^2\,x^2 + \frac{1-\lambda_\Omega^2}{N}\,\langle\boldsymbol{\Omega}^2\rangle. \tag{2.25b}$$

When p is small, λ_Ω is close to unity, so we may use the approximation

$$B(x) \approx \frac{2(1-\lambda_\Omega)}{N}\langle \boldsymbol{\Omega}^2 \rangle, \tag{2.26}$$

which then generates a process that is formally identical to the OU process. In this way we were able to derive analytical results from the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ models that agreed well with simulations. Although the eigenvector requirement prevents an analytical study of the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ models using this method, we found that the FPT results from simulations for either the $\mathbb{C}$ model or the $\mathbb{Q}$ model were only very mildly sensitive to the choice of $\boldsymbol{\Omega}$, whether an eigenvector of $\mathbb{M}$ or not. This is because for $\boldsymbol{\Omega}$ not an eigenvector of $\mathbb{M}$, the eigenvector of $\mathbb{M}$ closest to $\boldsymbol{\Omega}$ dominates the dynamics, and $\boldsymbol{\mathcal{L}}$ is closest to $\boldsymbol{\mathcal{S}}$ and *vice versa*.

3 Extension of Dynamics

The deduced forms of the jump moments for $h(t)$ in Eq. (2.25) are strictly speaking illegitimate since in principle there could be additional terms on the right-hand sides whose expectation values are identically zero, although for $\nu = 2$ these specific forms have been obtained by other methods. We now turn to approaching the FPT calculation via a different and more general route, allowing us to drop the eigenvector requirement entirely, so that we can study all four models ($\mathbb{C}\&\boldsymbol{\mathcal{L}}$, $\mathbb{C}\&\boldsymbol{\mathcal{S}}$, $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$) and in principle any other model analytically. This more general method permits a direct calculation of the jump moments without going via their expectation values, but to do so we must formulate a Fokker-Planck equation in $\nu - 1$ variables. We then discuss FPT processes in this higher-dimensional space, including exact methods.

3.1 Extension to $\nu - 1$ Variables

As discussed in section 2, for $\nu > 2$ the dynamics in the variable $h(t)$ alone are not Markovian: we cannot collapse ${}^{N+\nu-1}C_{\nu-1}$ distinct synaptic configurations down to Markovian dynamics in just $N(\nu-1)+1$ states (for $\boldsymbol{\Omega} = \boldsymbol{\mathcal{L}}$). However, by extending the number of variables from just one (namely, $h(t)$) to $\nu - 1$, we can work with a Fokker-Planck equation in these extended variables. A synaptic configuration is (up to irrelevant permutations) uniquely characterised by the numbers N_j, $j = 1, \ldots, \nu$, of synapses with (tilded) strengths Ω_j. The tracked memory signal $h = \frac{1}{N}\sum_{i=1}^{N} \widetilde{S}_i$, where we drop the temporal arguments for convenience, can be written in terms of these synaptic numbers as

$$h = \frac{1}{N}\sum_{j=1}^{\nu} N_j\,\Omega_j = \frac{1}{N}\boldsymbol{N}\cdot\boldsymbol{\Omega}, \tag{3.1}$$

where $\boldsymbol{N} = (N_1, \ldots, N_\nu)^{\mathrm{T}}$ and "$\cdot$" denotes the dot product. The vector $\boldsymbol{N}$ contains $\nu - 1$ independent degrees of freedom because of the constraint $\sum_{i=1}^{\nu} N_i = \boldsymbol{N}\cdot\boldsymbol{1} = N$, but the single variable h encodes only one of them. Thus, we introduce additional, independent degrees of freedom. Let $\boldsymbol{v} = (v_0, v_1, \ldots, v_{\nu-1})^{\mathrm{T}}$ be a ν-dimensional vector of variables, where we take $v_1 \equiv h$. We define these variables via

$$v_m = \frac{1}{N}\boldsymbol{N}\cdot\boldsymbol{e}^m, \tag{3.2}$$

where the $\boldsymbol{e}^m$, $m = 0, \ldots, \nu - 1$, are a set of orthogonal vectors. Since we want $v_1 \equiv h$, we must have $\boldsymbol{e}^1 = \boldsymbol{\Omega}$. Now, if $\boldsymbol{\Omega} = \boldsymbol{\mathcal{L}}$, we already have at hand a convenient set of orthogonal eigenvectors containing $\boldsymbol{\mathcal{L}}$, namely the eigenvectors $\boldsymbol{l}^m$ of the matrix $\mathbb{M}_{\mathrm{Q}}$. Similarly, if $\boldsymbol{\Omega} = \boldsymbol{\mathcal{S}}$, the set of eigenvectors $\boldsymbol{s}^m$ of $\mathbb{M}_{\mathrm{C}}$ contains $\boldsymbol{\mathcal{S}}$. Thus, we set $\boldsymbol{e}^m = \boldsymbol{l}^m$ or $\boldsymbol{e}^m = \boldsymbol{s}^m$ depending on whether

$\boldsymbol{\Omega} = \boldsymbol{\mathcal{L}}$ or $\boldsymbol{\Omega} = \boldsymbol{\mathcal{S}}$. In either case, $\boldsymbol{e}^0 = \mathbf{1} = (1, \ldots, 1)^{\mathrm{T}}$, so the "variable" $v_0 \equiv 1$, representing the constraint $\frac{1}{N}\boldsymbol{N} \cdot \mathbf{1} = 1$, leaving only $\nu - 1$ independent degrees of freedom as required. The choice of a particular set of orthogonal basis vectors depends only on the choice of the vector of possible synaptic strengths $\boldsymbol{\Omega}$. It does *not* depend on the particular choice of transition matrix defining a model of synaptic plasticity: we just happen to choose the eigenvectors of $\mathbb{M}_{\mathrm{C}}$ or $\mathbb{M}_{\mathrm{Q}}$, depending on whether $\boldsymbol{\Omega} = \boldsymbol{\mathcal{S}}$ or $\boldsymbol{\Omega} = \boldsymbol{\mathcal{L}}$, because they are conveniently at hand. For example, if $\boldsymbol{\Omega} = \boldsymbol{\mathcal{S}}$, we would use for convenience $\boldsymbol{e}^m = \boldsymbol{s}^m$ in Eq. (3.2), but the synaptic plasticity transition matrix $\mathbb{M}$ may be either $\mathbb{M}_{\mathrm{C}}$ with eigenvectors $\boldsymbol{\varepsilon}^m = \boldsymbol{s}^m$ or $\mathbb{M}_{\mathrm{Q}}$ with eigenvectors $\boldsymbol{\varepsilon}^m = \boldsymbol{l}^m$, or indeed any other appropriate stochastic matrix.

Because the vectors $\boldsymbol{e}^m$ are orthogonal, it is easy to invert Eq. (3.2), giving the vector $\boldsymbol{N}$,

$$\boldsymbol{N} = N \sum_{m=0}^{\nu-1} \beta_m v_m \boldsymbol{e}^m, \tag{3.3}$$

where $\beta_m^{-1} = \boldsymbol{e}^m \cdot \boldsymbol{e}^m$, or its components

$$N_i = N \sum_{m=0}^{\nu-1} \beta_m v_m e_i^m \equiv N \left[\frac{1}{\nu} + \sum_{m=1}^{\nu-1} \beta_m v_m e_i^m \right]. \tag{3.4}$$

On average $N_i = N/\nu$ in equilibrium, so we see that on average $v_m = 0 \; \forall m \neq 0$ in equilibrium.

The variables in the vector $\boldsymbol{v}$ completely specify the (tilded) synaptic configuration. Consider a single non-tracked memory storage event. Immediately before the event, let $\boldsymbol{v}'$ with components v'_m specify the synaptic configuration; immediately after the event, $\boldsymbol{v}$ specifies the configuration. We can now essentially just write down the conditional moment generating function (MGF) for $\boldsymbol{v}$, conditioned on $\boldsymbol{v}'$. Let the ν columns of the transition matrix $\mathbb{M}$ be

denoted by the ν vectors $\mathbf{M}^j$, $j = 1, \ldots, \nu$. For a single synapse with untilded strength Ω_j, the column $\mathbf{M}^j$ gives the transition probabilities to different untilded strengths during one memory storage event. If $\boldsymbol{y} = (y_1, \ldots, y_\nu)^{\mathrm{T}}$ is a vector of ν dummy variables, the probability generating function (PGF) for this transition is $\mathbf{M}^j \cdot \boldsymbol{y}$ where y_i picks out untilded strength Ω_i immediately after the transition. For a single memory storage step, all N synapses update independently because by assumption they experience independent plasticity induction events. If N'_j, $j = 1, \ldots, \nu$, represent the numbers of synapses with untilded strengths Ω_j before the memory storage event, the conditional PGF

$$\mathcal{G}\left(\boldsymbol{y} \,|\, \boldsymbol{N}'\right) = \prod_{j=1}^{\nu} \left(\mathbf{M}^j \cdot \boldsymbol{y}\right)^{N'_j} \tag{3.5}$$

is the PGF for the number of synapses N_i with untilded strengths Ω_i, $i = 1, \ldots, \nu$, immediately after the storage event. This conditional PGF is just a multinomial in the ν dummy variables y_i, $j = i, \ldots, \nu$. We have been careful to specify untilded strengths in this argument for the purposes of clarity. However, when potentiation and depression processes are treated completely symmetrically, as discussed in section 2, we know that we may in effect simply set $\xi_i^0 = +1 \ \forall i$ in all calculations, consider just untilded strengths, and be guaranteed to obtain the same results as for calculations with tilded strengths. Eq. (3.5) must therefore also be the conditional PGF for a transition from tilded configuration $\boldsymbol{N}'$ to tilded configuration $\boldsymbol{N}$. A careful consideration of the structure of $\mathbb{M}$ for symmetric processes with full conditioning on the values of ξ_i^0 shows this to be the case.

To convert the conditional PGF in Eq. (3.5) to a conditional MGF we replace the vector $(y_1, \ldots, y_\nu)^{\mathrm{T}}$ with $(e^{y_1}, \ldots, e^{y_\nu})^{\mathrm{T}}$. This describes the MGF for a transition between tilded configurations $\boldsymbol{N}'$ and $\boldsymbol{N}$. However, we require

the MGF in terms of $\boldsymbol{v}'$ and $\boldsymbol{v}$ and not $\boldsymbol{N}'$ and $\boldsymbol{N}$. We can simply write $N'_j = N \sum_{m=0}^{\nu-1} \beta_m v'_m e_j^m$ to convert from $\boldsymbol{N}'$ to $\boldsymbol{v}'$. To convert from $\boldsymbol{N}$ to $\boldsymbol{v}$, we note that $\boldsymbol{y} = \sum_{m=1}^{\nu} y_m \, \boldsymbol{\delta}^m$, where the set $\{\boldsymbol{\delta}^1, \ldots, \boldsymbol{\delta}^\nu\}$ is the standard set of Cartesian basis vectors. It is the vector $\boldsymbol{\delta}^i$, say, that picks out synapses with (tilded) strength Ω_i and thus counts the number of synapses N_i with this (tilded) strength. To convert to $\boldsymbol{v}$, instead of projecting out states using $\boldsymbol{\delta}^m$, we use $\boldsymbol{e}^m$ and include the required scaling factor of $1/N$ since $v_m = \frac{1}{N} \boldsymbol{N} \cdot \boldsymbol{e}^m$. To avoid notational awkwardness, we use dummy variables $\boldsymbol{z} = (z_0, \ldots, z_{\nu-1})^{\mathrm{T}}$ rather than $\boldsymbol{y} = (y_1, \ldots, y_\nu)^{\mathrm{T}}$ because $\boldsymbol{v} = (v_0, \ldots, v_{\nu-1})^{\mathrm{T}}$ in contrast to $\boldsymbol{N} = (N_1, \ldots, N_\nu)^{\mathrm{T}}$. The required conditional MGF for transitions from $\boldsymbol{v}'$ to $\boldsymbol{v}$ is then

$$\mathcal{M}\left(\boldsymbol{z} \,|\, \boldsymbol{v}'\right) = \prod_{j=1}^{\nu} \left(\mathbf{M}^j \cdot \boldsymbol{\omega}\right)^{N \sum_{m=0}^{\nu-1} \beta_m v'_m e_i^m} , \tag{3.6}$$

where $\boldsymbol{\omega}$ is a ν-dimensional vector with components

$$\omega_i = \exp\left(\frac{1}{N} \sum_{n=0}^{\nu-1} z_n e_i^n\right) . \tag{3.7}$$

Since $\left(\mathbf{M}^j\right)_i \equiv \mathbb{M}_{ij}$, we may write $\mathcal{M}$ explicitly as

$$\mathcal{M}\left(\boldsymbol{z} \,|\, \boldsymbol{v}'\right) = \prod_{j=1}^{\nu} \left[\sum_{i=1}^{\nu} \exp\left(\frac{1}{N} \sum_{n=0}^{\nu-1} z_n e_i^n\right) \mathbb{M}_{ij}\right]^{N \sum_{m=0}^{\nu-1} \beta_m v'_m e_j^m} . \tag{3.8}$$

This is the required form for the conditional MGF for a transition from configuration $\boldsymbol{v}'$ to configuration $\boldsymbol{v}$ during a single non-tracked memory storage event.

3.2 Fokker-Planck Equation in the v_i Variables

Differentiating $\mathcal{M}(\boldsymbol{z}\,|\,\boldsymbol{v}')$ with respect to the z_i generates the conditional moments for the v_i. After some algebra we obtain the expected changes in the first- and second-order moments induced by a single, non-tracked memory storage event,

$$\mathsf{E}\big[(v_i - v_i')|\,\boldsymbol{v}'\big] = \sum_{m=0}^{\nu-1} \boldsymbol{e}^i \cdot \big(\mathbb{M} - \mathbb{I}\big)\,\boldsymbol{e}^m\,\beta_m\,v_m' \tag{3.9a}$$

$$\begin{aligned}\mathsf{E}\big[(v_i - v_i')(v_j - v_j')|\,\boldsymbol{v}'\big] &= \mathsf{E}\big[(v_i - v_i')|\,\boldsymbol{v}'\big]\,\mathsf{E}\big[(v_j - v_j')|\,\boldsymbol{v}'\big] \\ &\quad + \frac{1}{N}\sum_{m=0}^{\nu-1}\Big\{\big(\boldsymbol{e}^i \circ \boldsymbol{e}^j\big)\cdot\mathbb{M}\,\boldsymbol{e}^m - \big[\big(\mathbb{M}^{\mathrm{T}}\boldsymbol{e}^i\big)\circ\big(\mathbb{M}^{\mathrm{T}}\boldsymbol{e}^j\big)\big]\cdot\boldsymbol{e}^m\Big\}\,\beta_m\,v_m',\end{aligned} \tag{3.9b}$$

where we use the notation $\boldsymbol{a}\circ\boldsymbol{b}$ to indicate a vector with components $(\boldsymbol{a}\circ\boldsymbol{b})_i = a_i b_i$; note that $\mathbf{1}\cdot(\boldsymbol{a}\circ\boldsymbol{b}) \equiv \boldsymbol{a}\cdot\boldsymbol{b}$. These equations are general in the sense that they are valid for a general stochastic matrix $\mathbb{M}$, and as we have not used $\boldsymbol{e}^0 = \mathbf{1}$ and $\mathbf{1}\cdot\boldsymbol{N} = N$ in their derivation, they are also valid for any set of orthogonal vectors $\{\boldsymbol{e}^0, \ldots, \boldsymbol{e}^{\nu-1}\}$. For the particular case that $\boldsymbol{e}^0 = \mathbf{1}$ corresponding to $v_0 = 1 = v_0'$, putting $i = 0$ in these equations returns zero on the right-hand sides as required because $\mathbf{1}$ is a null left eigenvector of the generating matrix $\mathbb{G} = \mathbb{M} - \mathbb{I}$ and also $\mathbf{1}\circ\boldsymbol{e}^j \equiv \boldsymbol{e}^j$.

We now restrict to the particular choice $\boldsymbol{e}^0 = \mathbf{1}$ and use the fact that $v_0 = 1 = v_0'$. We discard v_0 from $\boldsymbol{v}$ and now always write $\boldsymbol{v} = (v_1, \ldots v_{\nu-1})^{\mathrm{T}}$, and similarly for $\boldsymbol{v}'$. Further, the two matrices $\mathbb{M}_{\mathrm{c}}$ and $\mathbb{M}_{\mathrm{q}}$ of interest to us are symmetric, so we also restrict to the case that $\mathbb{M}^{\mathrm{T}} = \mathbb{M}$. Eq. (3.9) then becomes

$$\mathsf{E}\big[(v_i - v_i')|\,\boldsymbol{v}'\big] = \sum_{m=1}^{\nu-1} \boldsymbol{e}^i \cdot \big(\mathbb{M} - \mathbb{I}\big)\,\boldsymbol{e}^m\,\beta_m\,v_m' \tag{3.10a}$$

$$\mathsf{E}\big[(v_i - v_i')(v_j - v_j')\,|\,\boldsymbol{v}'\big] = \mathsf{E}\big[(v_i - v_i')\,|\,\boldsymbol{v}'\big]\,\mathsf{E}\big[(v_j - v_j')\,|\,\boldsymbol{v}'\big] + \frac{1}{N}\sum_{m=1}^{\nu-1}\Big[\mathbb{M}\left(\boldsymbol{e}^i \circ \boldsymbol{e}^j\right) - \left(\mathbb{M}\,\boldsymbol{e}^i\right) \circ \left(\mathbb{M}\,\boldsymbol{e}^j\right)\Big] \cdot \boldsymbol{e}^m\,\beta_m\,v_m' + \frac{1}{N}\frac{1}{\nu}\Big[\beta_i^{-1}\delta_{ij} - \left(\mathbb{M}\,\boldsymbol{e}^i\right) \cdot \left(\mathbb{M}\,\boldsymbol{e}^j\right)\Big], \tag{3.10b}$$

where δ_{ij} is the Kronecker delta symbol, and now $i, j = 1, \ldots, \nu - 1$.

For the jump moments in the Fokker-Planck equation in the variables $v_1, \ldots, v_{\nu-1}$, we require the generalised form of Eq. (2.19), describing the expected changes induced during a small time interval $[t, t+\delta t]$. However, because we have assumed that memory storage occurs as a Poisson process, by definition at most one memory storage event occurs during such an interval. Thus, the expressions in Eq. (3.10) are precisely the jump moments required in the Fokker-Planck equation. We generalise the 1-dimensional jump moments A and B for h in section 2 by defining the $(\nu - 1) \times (\nu - 1)$ matrices $\mathbb{A}$ and $\mathbb{B}$ to have elements $\mathbb{A}_{ij} = \boldsymbol{e}^i \cdot \mathbb{G}\,\boldsymbol{e}^j\,\beta_j$ so that $(\mathbb{A}\boldsymbol{v}')_i \equiv \mathsf{E}\big[(v_i - v_i')\,|\,\boldsymbol{v}'\big]$, and $\mathbb{B}_{ij}(\boldsymbol{v}') = \mathsf{E}\big[(v_i - v_i')(v_j - v_j')\,|\,\boldsymbol{v}'\big]$, where for $\mathbb{B}(\boldsymbol{v}')$ we explicitly indicate its dependence on $\boldsymbol{v}'$. It will be convenient below to write

$$\mathbb{B}(\boldsymbol{v}') = (\mathbb{A}\boldsymbol{v}')(\mathbb{A}\boldsymbol{v}')^{\mathrm{T}} + \mathbb{B}^{(1)}(\boldsymbol{v}') + \mathbb{B}^{(0)}, \tag{3.11}$$

where these three terms have elements that occur, in order, on the right hand side of Eq. (3.10b).

Defining $P(\boldsymbol{v}, t\,|\,\boldsymbol{v}_0, t_0)$ to be the transition probability between configuration $\boldsymbol{v}_0$ at time t_0 (with $t_0 = 0$ s as usual) and configuration $\boldsymbol{v}$ at later time t, the multi-dimensional Fokker-Planck equation (van Kampen, 1992) in the

variables v_i is then

$$\frac{1}{r}\frac{\partial}{\partial t}P(\boldsymbol{v},t\,|\,\boldsymbol{v}_0,t_0) = -\sum_{i=1}^{\nu-1}\frac{\partial}{\partial v_i}\big[(\mathbb{A}\boldsymbol{v})_i\,P(\boldsymbol{v},t\,|\,\boldsymbol{v}_0,t_0)\big] + \frac{1}{2}\sum_{i,j=1}^{\nu-1}\frac{\partial^2}{\partial v_i\partial v_j}\big[\mathbb{B}_{ij}(\boldsymbol{v})P(\boldsymbol{v},t\,|\,\boldsymbol{v}_0,t_0)\big]. \tag{3.12}$$

The backward or adjoint form can also be written down, and we will use it for FPT calculations. The evolution equations for the conditional moments follow directly from the Fokker-Planck equation (van Kampen, 1992). Writing

$$\mathbb{S}_{\boldsymbol{v}_0}(t) = \mathsf{E}\big[\boldsymbol{v}(t)\boldsymbol{v}(t)^{\mathrm{T}}|\boldsymbol{v}_0\big] - \mathsf{E}\big[\boldsymbol{v}(t)|\boldsymbol{v}_0\big]\,\mathsf{E}\big[\boldsymbol{v}(t)^{\mathrm{T}}|\boldsymbol{v}_0\big] \tag{3.13}$$

for the conditional covariance matrix conditioned on $\boldsymbol{v}(t_0)=\boldsymbol{v}_0$, we have

$$\frac{1}{r}\frac{d}{dt}\mathsf{E}[\boldsymbol{v}(t)|\boldsymbol{v}_0] = \mathbb{A}\,\mathsf{E}[\boldsymbol{v}(t)|\boldsymbol{v}_0], \tag{3.14a}$$

$$\frac{1}{r}\frac{d}{dt}\mathbb{S}_{\boldsymbol{v}_0}(t) = \mathbb{A}\,\mathbb{S}_{\boldsymbol{v}_0}(t) + \mathbb{S}_{\boldsymbol{v}_0}(t)\mathbb{A}^{\mathrm{T}} + \mathbb{B}^{(0)} + \mathsf{E}[\mathbb{B}^{(1)}(\boldsymbol{v})|\boldsymbol{v}_0] + \mathbb{A}\,\mathbb{S}_{\boldsymbol{v}_0}(t)\mathbb{A}^{\mathrm{T}} + \mathbb{A}\,\mathsf{E}[\boldsymbol{v}(t)|\boldsymbol{v}_0]\,\mathsf{E}[\boldsymbol{v}(t)|\boldsymbol{v}_0]^{\mathrm{T}}\mathbb{A}^{\mathrm{T}}, \tag{3.14b}$$

subject to the the initial conditions $\mathsf{E}[\boldsymbol{v}(t_0)|\boldsymbol{v}_0]\equiv\boldsymbol{v}_0$ and $\mathbb{S}_{\boldsymbol{v}_0}(t_0)\equiv\mathbb{O}$, a matrix of zeros. Because $\mathbb{B}^{(1)}(\boldsymbol{v})$ is linear in $\boldsymbol{v}$, $\mathsf{E}[\mathbb{B}^{(1)}(\boldsymbol{v})|\boldsymbol{v}_0]$ reduces to a matrix involving the components of $\mathsf{E}[\boldsymbol{v}(t)|\boldsymbol{v}_0]$. Notice that $\mathbb{S}_{\boldsymbol{v}_0}(t)$ depends explicitly on $\boldsymbol{v}_0$ via the explicit dependence of the right hand side of Eq. (3.14b) on $\mathsf{E}[\boldsymbol{v}(t)|\boldsymbol{v}_0]$.

Considering the drift or convective dynamics, the Liouville equation in Eq. (3.14a) has the solution $\mathsf{E}[\boldsymbol{v}(t)|\boldsymbol{v}_0] = e^{rt\mathbb{A}}\,\boldsymbol{v}_0$ (with $t_0=0$ s). Since $\mathbb{A}_{ij} = \boldsymbol{e}^i\cdot\mathbb{G}\,\boldsymbol{e}^j\,\beta_j$, the matrix $\mathbb{A}$ is just the generating matrix $\mathbb{G}$ expressed in the incomplete basis of orthogonal vectors $\{\boldsymbol{e}^1,\ldots,\boldsymbol{e}^{\nu-1}\}$; indeed, we may use the original form in Eq. (3.9a) to extend to the complete basis $\{\boldsymbol{e}^0,\ldots,\boldsymbol{e}^{\nu-1}\}$ for

general $\boldsymbol{e}^0$. It is straightforward to see that $\left(e^{rt\mathbb{A}}\right)_{ij} = \boldsymbol{e}^i \cdot e^{rt\mathbb{G}}\, \boldsymbol{e}^j\, \beta_j$, from which we obtain

$$\mathsf{E}[v_i(t)|\boldsymbol{v}_0] = \boldsymbol{e}^i \cdot e^{rt\mathbb{G}} \sum_j \boldsymbol{e}^j \beta_j\, (\boldsymbol{v}_0)_j = \tfrac{1}{N}\, \boldsymbol{e}^i \cdot e^{rt\mathbb{G}}\, \boldsymbol{N}_0, \tag{3.15}$$

where $(\boldsymbol{v}_0)_j$ is the j^{th} component of $\boldsymbol{v}_0$ and $\boldsymbol{N}_0$ is the definite configuration of tilded synaptic strength numbers corresponding to the definite configuration $\boldsymbol{v}_0$. The evolution of $\mathsf{E}[\boldsymbol{v}(t)|\boldsymbol{v}_0]$ is therefore correctly driven by the underlying generating matrix $\mathbb{G}$, from which the drift matrix $\mathbb{A}$ is obtained by a simple transformation of basis vectors.

We will require explicit forms for the left and right eigenvectors of $\mathbb{A}$. To determine them, for simplicity we complete the basis by including $\boldsymbol{e}^0 = \mathbf{1}$. Let $\mathbb{E}$ be a matrix whose ν rows contain the vectors $\boldsymbol{e}^i$, $i = 0, \ldots, \nu-1$, and define the diagonal matrix $\mathbb{D}^{-1} = \text{diag}\{\beta_0, \ldots, \beta_{\nu-1}\}$ (we choose this convention because $\beta_i^{-1} = |\boldsymbol{e}^i|^2$). Then $\mathbb{A} = \mathbb{E}\,\mathbb{G}\,\mathbb{E}^{\mathrm{T}}\mathbb{D}^{-1}$. The eigenvalues of the symmetric matrix $\mathbb{G}$ are $\lambda_i - 1$ and let its orthonormal eigenvectors be $\widehat{\boldsymbol{\varepsilon}}^{\,i}$, $i = 0, \ldots, \nu-1$, which are just the normalised forms of the eigenvectors $\boldsymbol{\varepsilon}^i$ stated above. Then we can write $\mathbb{G} = \sum_{i=0}^{\nu-1} (\lambda_i - 1)\, \widehat{\boldsymbol{\varepsilon}}^{\,i} \big(\widehat{\boldsymbol{\varepsilon}}^{\,i}\big)^{\mathrm{T}}$ and hence

$$\mathbb{A} = \sum_{i=0}^{\nu-1} (\lambda_i - 1) \big(\mathbb{E}\,\widehat{\boldsymbol{\varepsilon}}^{\,i}\big) \big(\mathbb{D}^{-1}\mathbb{E}\,\widehat{\boldsymbol{\varepsilon}}^{\,i}\big)^{\mathrm{T}} = \sum_{i=0}^{\nu-1} (\lambda_i - 1) \boldsymbol{\rho}^{\,i} \left(\boldsymbol{\lambda}^i\right)^{\mathrm{T}}, \tag{3.16}$$

where the second equality gives the left and right eigenvectors, respectively, as $\boldsymbol{\lambda}^i = \mathbb{D}^{-1}\mathbb{E}\,\widehat{\boldsymbol{\varepsilon}}^{\,i}$ and $\boldsymbol{\rho}^{\,i} = \mathbb{E}\,\widehat{\boldsymbol{\varepsilon}}^{\,i}$. Their components are $\big(\boldsymbol{\lambda}^i\big)_j = \beta_j\, \boldsymbol{e}^j \cdot \widehat{\boldsymbol{\varepsilon}}^{\,i}$ and $\big(\boldsymbol{\rho}^{\,i}\big)_j = \boldsymbol{e}^j \cdot \widehat{\boldsymbol{\varepsilon}}^{\,i}$ and since $\sum_{i=0}^{\nu-1} \beta_i\, \boldsymbol{e}^i \big(\boldsymbol{e}^i\big)^{\mathrm{T}} = \mathbb{E}^{\mathrm{T}}\mathbb{D}^{-1}\mathbb{E} \equiv \mathbb{I}$, we have that $\boldsymbol{\lambda}^i \cdot \boldsymbol{\rho}^{\,j} = \delta_{ij}$, as required. Note that $\mathbb{D}^{-1}$ converts right into left eigenvectors, and therefore $\mathbb{D}$ left into right, and so $\mathbb{A}$ and $\mathbb{A}^{\mathrm{T}}$ are related via the similarity relation $\mathbb{D}^{-1}\mathbb{A}\mathbb{D} = \mathbb{A}^{\mathrm{T}}$. To move to the $(\nu-1)\times(\nu-1)$ form of $\mathbb{A}$ and its $(\nu-1)$-

dimensional eigenvectors, i.e. without the basis vector $\boldsymbol{e}^0 = \mathbf{1}$, we simply strip off any components or elements corresponding to $i = 0$; this works because $\mathbb{A}_{i\,0} = \mathbb{A}_{0\,i} \equiv 0\ \forall i$.

Using these arguments, we may also show that the matrices $\mathbb{A}$ and $\mathbb{B}^{(0)}$ can be simultaneously diagonalised, a result that we will use below to develop approximation methods. We again use the completed basis for convenience. We have that

$$\mathbb{B}^{(0)}_{ij} = \frac{1}{N}\frac{1}{\nu}\Big[\beta_i^{-1}\delta_{ij} - \big(\mathbb{M}\,\boldsymbol{e}^i\big)\cdot\big(\mathbb{M}\,\boldsymbol{e}^j\big)\Big] = -\frac{1}{N}\frac{1}{\nu}\,\boldsymbol{e}^i\cdot\big(2\,\mathbb{G}+\mathbb{G}^2\big)\boldsymbol{e}^j, \qquad (3.17)$$

or just $\mathbb{B}^{(0)} = -\frac{1}{N}\frac{1}{\nu}\,\mathbb{E}\big(2\,\mathbb{G}+\mathbb{G}^2\big)\mathbb{E}^{\mathrm{T}}$. Since $\mathbb{A} = \mathbb{E}\,\mathbb{G}\,\mathbb{E}^{\mathrm{T}}\mathbb{D}^{-1}$ we have $\mathbb{A}^2 = \mathbb{E}\,\mathbb{G}^2\,\mathbb{E}^{\mathrm{T}}\mathbb{D}^{-1}$ and hence we can write

$$\mathbb{B}^{(0)} = -\frac{1}{N}\frac{1}{\nu}\big(2\mathbb{A}+\mathbb{A}^2\big)\mathbb{D}. \qquad (3.18)$$

Define the matrix $\mathbb{T}$ as the matrix whose rows are the normalised left eigenvectors of $\mathbb{A}$, denoted by $\widehat{\boldsymbol{\lambda}}^i = \boldsymbol{\lambda}^i/|\boldsymbol{\lambda}^i|$ for $i = 0,\ldots,\nu-1$. Then the matrix $\mathbb{T}\mathbb{A}\mathbb{T}^{-1}$ is by definition diagonal with the eigenvectors $\lambda_i - 1$ on the diagonal. Considering the matrix $\mathbb{T}\mathbb{B}^{(0)}\mathbb{T}^{\mathrm{T}}$, we have

$$\begin{aligned}\mathbb{T}\mathbb{B}^{(0)}\mathbb{T}^{\mathrm{T}} &= -\frac{1}{N}\frac{1}{\nu}\,\mathbb{T}\big(2\mathbb{A}+\mathbb{A}^2\big)\mathbb{D}\,\mathbb{T}^{\mathrm{T}}\\ &= -\frac{1}{N}\frac{1}{\nu}\left[\mathbb{T}\big(2\mathbb{A}+\mathbb{A}^2\big)\mathbb{T}^{-1}\right]\left[\mathbb{T}\,\mathbb{D}\,\mathbb{T}^{\mathrm{T}}\right]. \qquad (3.19)\end{aligned}$$

The matrix $\mathbb{T}\big(2\mathbb{A}+\mathbb{A}^2\big)\mathbb{T}^{-1}$ is manifestly diagonal with $\lambda_i^2 - 1$ on the diagonal. The matrix $\mathbb{T}\,\mathbb{D}\,\mathbb{T}^{\mathrm{T}}$ is also diagonal because $\mathbb{D}$ converts the left eigenvectors of $\mathbb{A}$ into its right eigenvectors and $\mathbb{T}^{\mathrm{T}}$ has the normalised left eigenvectors as its columns. So $\mathbb{T}\,\mathbb{D}\,\mathbb{T}^{\mathrm{T}}$ is diagonal with $1/\left|\boldsymbol{\lambda}^i\right|^2$ on the diagonal. Hence, the

matrix $\mathbb{T}$ also diagonalises $\mathbb{B}^{(0)}$, leaving the entries $-\frac{1}{N}\frac{1}{\nu}\left(\lambda_i^2-1\right)/\left|\boldsymbol{\lambda}^i\right|^2$ on the diagonal.

Turning to the evolution of the covariance matrix $\mathbb{S}_{\boldsymbol{v}_0}(t)$, we may use Eq. (3.14b) to determine the equilibrium form $\mathbb{S}(\infty)=\lim_{t\to\infty}\mathbb{S}_{\boldsymbol{v}_0}(t)$, where this is expected to be independent of the initial configuration $\boldsymbol{v}_0$. In this limit, Eq. (3.14b) reduces to

$$\mathbb{O}=\mathbb{A}\,\mathbb{S}(\infty)+\mathbb{S}(\infty)\mathbb{A}^{\mathrm{T}}+\mathbb{A}\,\mathbb{S}(\infty)\mathbb{A}^{\mathrm{T}}+\mathbb{B}^{(0)}, \tag{3.20}$$

where indeed $\boldsymbol{v}_0$ drops out, and where $\mathbb{B}^{(0)}$ is given in Eq. (3.18) but using the incomplete basis, so with $\mathbb{D}^{-1}=\mathrm{diag}\{\beta_1,\ldots,\beta_{\nu-1}\}$. Using the similarity relation $\mathbb{D}^{-1}\mathbb{A}\mathbb{D}=\mathbb{A}^{\mathrm{T}}$ we may confirm that

$$\mathbb{S}(\infty)=\frac{1}{N}\frac{1}{\nu}\,\mathbb{D}=\frac{1}{N}\frac{1}{\nu}\,\mathrm{diag}\big\{\left|\boldsymbol{e}^1\right|^2,\ldots,\left|\boldsymbol{e}^{\nu-1}\right|^2\big\} \tag{3.21}$$

satisfies Eq. (3.20), giving the equilibrium covariance matrix. As the MGF of the equilibrium tilded synaptic strength configuration is just $(\boldsymbol{\omega}\cdot\boldsymbol{A})^N$, its covariance matrix follows directly as having elements $\overline{\mathbb{S}}_{ij}(\infty)=\frac{1}{N}\frac{1}{\nu}\,\boldsymbol{e}^i\cdot\boldsymbol{e}^j$, in agreement with Eq. (3.21).

We briefly consider the expressions for the jump moments in the particular case that the eigenvector requirement $\boldsymbol{\Omega}^{\mathrm{T}}\mathbb{M}=\lambda_\Omega\boldsymbol{\Omega}^{\mathrm{T}}$ (or $\mathbb{M}\,\boldsymbol{\Omega}=\lambda_\Omega\boldsymbol{\Omega}$ for symmetric $\mathbb{M}$) studied earlier (Elliott, 2017b) is satisfied. In this case the basis of eigenvectors $\boldsymbol{e}^m$ for defining the extended variables v_m coincides with the set of eigenvectors of $\mathbb{M}$. Hence, the elements of the drift matrix are simply $\mathbb{A}_{ij}=-(1-\lambda_j)\,\delta_{ij}$, so that the drift matrix is diagonal. The diffusion matrix

elements are

$$\mathbb{B}_{ij}(\boldsymbol{v}) = (1-\lambda_i)(1-\lambda_j)v_i v_j + \frac{1}{N}\frac{1}{\nu}\,\beta_j^{-1}(1-\lambda_j^2)\,\delta_{ij} \\ + \frac{1}{N}\sum_{m=1}^{\nu-1}\Big[\mathbb{M}\left(\boldsymbol{e}^i \circ \boldsymbol{e}^j\right) - \lambda_i\lambda_j\left(\boldsymbol{e}^i \circ \boldsymbol{e}^j\right)\Big]\cdot \boldsymbol{e}^m\,\beta_m\,v_m, \tag{3.22}$$

where the sum on the second line corresponds to the elements $\mathbb{B}_{ij}^{(1)}(\boldsymbol{v})$, which cannot be evaluated without using knowledge of the particular form of $\mathbb{M}$ and its eigenvectors. At the level of the mean dynamics, the $\boldsymbol{v}$ variables completely decouple when the eigenvector condition is satisfied, as expected. Focusing, then, on $h \equiv v_1$, writing $\boldsymbol{e}^1 = \boldsymbol{\Omega}$ with $\lambda_1 = \lambda_\Omega$, and using A_h to denote the h-dependent part of the drift term $(\mathbb{A}\boldsymbol{v})_1$ and writing and $B_{hh} = \mathbb{B}_{11}$, we have the explicit forms

$$A_h(h) = -(1-\lambda_\Omega)h \tag{3.23a}$$

$$B_{hh}(h) = (1-\lambda_\Omega)^2 h^2 + \frac{1-\lambda_\Omega^2}{N}\langle \boldsymbol{\Omega}^2\rangle \\ + \frac{1}{N}\sum_{m=1}^{\nu-1}\Big[\mathbb{M}\left(\boldsymbol{\Omega}\circ\boldsymbol{\Omega}\right) - \lambda_\Omega^2\left(\boldsymbol{\Omega}\circ\boldsymbol{\Omega}\right)\Big]\cdot \boldsymbol{e}^m\,\beta_m\,v_m. \tag{3.23b}$$

These jump moments coincide precisely with those given in Eq. (2.25) but with additional terms in $B_{hh}(h)$ arising from $\mathbb{B}_{11}^{(1)}(\boldsymbol{v})$. We will see in section 4.1 that these additional terms can in fact be neglected in a $1/N$ approximation and that, furthermore, their expectation values vanish identically, so that $\mathsf{E}[A_h(h)]$ and $\mathsf{E}[B_{hh}(h)]$ agree precisely with the expected jump moments $\mathsf{E}[A(h)]$ and $\mathsf{E}[B(h)]$ given in Eq. (2.24). Therefore, when the eigenvector requirement is satisfied, we can collapse the $\boldsymbol{v}$ dynamics down to dynamics purely in the perceptron activation h.

3.3 First Passage Processes in the v_i Variables

The backward or adjoint form of Eq. (3.12) is

$$\frac{1}{r}\frac{\partial}{\partial t}P(\boldsymbol{v},t\,|\,\boldsymbol{v}_0,t_0) = +\sum_{i=1}^{\nu-1}(\mathbb{A}\boldsymbol{v}_0)_i\,\frac{\partial}{\partial(\boldsymbol{v}_0)_i}P(\boldsymbol{v},t\,|\,\boldsymbol{v}_0,t_0) + \frac{1}{2}\sum_{i,j=1}^{\nu-1}\mathbb{B}_{ij}(\boldsymbol{v}_0)\,\frac{\partial^2}{\partial(\boldsymbol{v}_0)_i\partial(\boldsymbol{v}_0)_j}P(\boldsymbol{v},t\,|\,\boldsymbol{v}_0,t_0). \tag{3.24}$$

We may use this form to compute the FPT density in the v_i variables. Let $\mathcal{I}$ denote the $(\nu-1)$-dimensional region in (tilded) synaptic strength configuration space in which $v_1 \equiv h > \theta$ and $v_i \in (-\infty,+\infty)$ for $i = 2,\ldots\nu-1$, so that the perceptron's activation in response to the tracked memory exceeds threshold but the values of all the other v_i variables are unconstrained. The density for first escape from $\mathcal{I}$ at some time $t > 0$ s, starting from $\boldsymbol{v}_0 \in \mathcal{I}$ at time $t_0 = 0$ s, is then

$$G(\boldsymbol{v}_0;t) = -\frac{\partial}{\partial t}\int_{\mathcal{I}}\cdots\int d\boldsymbol{v}\,P(\boldsymbol{v},t\,|\,\boldsymbol{v}_0,0), \tag{3.25}$$

where

$$\int_{\mathcal{I}}\cdots\int d\boldsymbol{v} = \int_{\theta}^{+\infty}dv_1\int_{-\infty}^{+\infty}dv_2\cdots\int_{-\infty}^{+\infty}dv_{\nu-1}, \tag{3.26}$$

and satisfies the equation

$$\frac{1}{r}\frac{\partial}{\partial t}G(\boldsymbol{v}_0;t) = \left[\sum_{i=1}^{\nu-1}(\mathbb{A}\boldsymbol{v}_0)_i\,\frac{\partial}{\partial(\boldsymbol{v}_0)_i} + \frac{1}{2}\sum_{i,j=1}^{\nu-1}\mathbb{B}_{ij}(\boldsymbol{v}_0)\,\frac{\partial^2}{\partial(\boldsymbol{v}_0)_i\partial(\boldsymbol{v}_0)_j}\right]G(\boldsymbol{v}_0,t), \tag{3.27}$$

subject to the absorbing boundary condition $G(\boldsymbol{v}_0;t) = \delta(t)$ for $\boldsymbol{v}_0 \in \partial\mathcal{I}$, i.e. when $\boldsymbol{v}_0$ is on the boundary of the region $\mathcal{I}$, which is the $(\nu-2)$-dimensional hyperplane $(\boldsymbol{v}_0)_1 \equiv h_0 = \theta$. The n^{th} order FPT moment for escape from $\mathcal{I}$, starting from $\boldsymbol{v}_0 \in \mathcal{I}$, is just $\tau^{(n)}(\boldsymbol{v}_0) = \int_0^\infty dt\,t^n\,G(\boldsymbol{v}_0;t)$ and for $n \geq 1$ it

satisfies

$$-\frac{1}{r}\,n\,\tau^{(n-1)}(\boldsymbol{v}_0) = \left[\sum_{i=1}^{\nu-1}(\mathbb{A}\boldsymbol{v}_0)_i\,\frac{\partial}{\partial(\boldsymbol{v}_0)_i} + \frac{1}{2}\sum_{i,j=1}^{\nu-1}\mathbb{B}_{ij}(\boldsymbol{v}_0)\,\frac{\partial^2}{\partial(\boldsymbol{v}_0)_i\partial(\boldsymbol{v}_0)_j}\right]\tau^{(n)}(\boldsymbol{v}_0) \tag{3.28}$$

subject to the boundary condition $\tau^{(n)}(\boldsymbol{v}_0) = 0$ for $\boldsymbol{v}_0 \in \partial\mathcal{I}$. We define $\tau^{(0)}(\boldsymbol{v}_0) \equiv 1$. The MFPT $\tau_{\rm mfpt}(\boldsymbol{v}_0)$ is of course just $\tau^{(1)}(\boldsymbol{v}_0)$.

A Fokker-Planck approach to FPT problems entails a diffusion approximation in which processes that lead to escape from $\mathcal{I}$ occur by diffusion across the boundary rather than jumps over it. Let $\mathsf{Prob}[\boldsymbol{v}|\boldsymbol{v}']$ represent the probability of a transition from configuration $\boldsymbol{v}'$ to configuration $\boldsymbol{v}$ during a single non-tracked memory storage step. We have this entire distribution encoded in the conditional MGF in Eq. (3.8). Then for discrete configurations, $\tau_{\rm mfpt}(\boldsymbol{v}_0)$ satisfies the equation

$$\tau_{\rm mfpt}(\boldsymbol{v}_0) = \frac{1}{r} + \sum_{\boldsymbol{v}\in\mathcal{I}}\tau_{\rm mfpt}(\boldsymbol{v})\,\mathsf{Prob}[\boldsymbol{v}|\boldsymbol{v}_0], \tag{3.29a}$$

while for continuous configurations it satisfies the integral equation

$$\tau_{\rm mfpt}(\boldsymbol{v}_0) = \frac{1}{r} + \int_{\mathcal{I}}\cdots\int d\boldsymbol{v}\,\tau_{\rm mfpt}(\boldsymbol{v})\,\mathsf{Prob}[\boldsymbol{v}|\boldsymbol{v}_0]. \tag{3.29b}$$

These equations are generalisations of the standard result in one dimension (van Kampen, 1992). They give *exact* results for $\tau_{\rm mfpt}(\boldsymbol{v}_0)$, in contrast to results from a Fokker-Planck approach with its diffusion approximation. Since we mostly work in the continuum limit we require Eq. (3.29b), but we will also discuss results below when N is small, for which Eq. (3.29a) is required. In general, the $n^{\rm th}$ order FPT moment $\tau^{(n)}(\boldsymbol{v}_0)$ can be shown to satisfy the

integral equation

$$\tau^{(n)}(\boldsymbol{v}_0) = \frac{1}{r}\, n\, \tau^{(n-1)}(\boldsymbol{v}_0) + \int_{\mathcal{I}}\!\cdots\!\int d\boldsymbol{v}\, \tau^{(n)}(\boldsymbol{v})\, \mathsf{Prob}[\boldsymbol{v}|\boldsymbol{v}_0] \tag{3.30}$$

for $n \geq 1$, again with $\tau^{(0)}(\boldsymbol{v}_0) \equiv 1$. We sketch the derivation in appendix A. Defining

$$\widehat{G}(\boldsymbol{v}_0; s) = \int_0^\infty dt\, e^{-st}\, G(\boldsymbol{v}_0; t) = \sum_{n=0}^{\infty} \frac{(-s)^n}{n!} \tau^{(n)}(\boldsymbol{v}_0), \tag{3.31}$$

$\widehat{G}(\boldsymbol{v}_0; s)$ is the Laplace transform of $G(\boldsymbol{v}_0; t)$ with transformed variable s, and is, up to the sign of s, just the MGF of the FPT distribution. By suitably summing Eq. (3.30), it can be massaged into the suggestive form

$$\frac{s}{r}\, \widehat{G}(\boldsymbol{v}_0; s) = \int_{\mathcal{I}}\!\cdots\!\int d\boldsymbol{v}\, \big[\widehat{G}(\boldsymbol{v}; s) - 1\big] \big\{\mathsf{Prob}[\boldsymbol{v}|\boldsymbol{v}_0] - \delta(\boldsymbol{v} - \boldsymbol{v}_0)\big\}, \tag{3.32}$$

where the left hand side is just the Laplace transform of $\frac{1}{r}\frac{\partial}{\partial t} G(\boldsymbol{v}_0; t)$ appearing on the left hand side of Eq. (3.27). In a limit corresponding to the diffusion limit, the right hand side of Eq. (3.32) must therefore correspond to the Laplace transform of the right hand side of Eq. (3.27).

In the continuum limit for simplicity we write $K(\boldsymbol{v}|\boldsymbol{v}') = \mathsf{Prob}[\boldsymbol{v}|\boldsymbol{v}']$ for the kernel under the integral sign in Eq. (3.30). From Eq. (3.10a), the conditional mean of $K(\boldsymbol{v}|\boldsymbol{v}')$ is $\mathsf{E}\big[\boldsymbol{v}|\boldsymbol{v}'\big] = (\mathbb{I} + \mathbb{A})\, \boldsymbol{v}'$, and from Eq. (3.10b) its conditional covariance matrix is

$$\mathsf{E}\big[\boldsymbol{v}\, \boldsymbol{v}^{\mathrm{T}}|\boldsymbol{v}'\big] - \mathsf{E}\big[\boldsymbol{v}|\boldsymbol{v}'\big]\, \mathsf{E}\big[\boldsymbol{v}^{\mathrm{T}}|\boldsymbol{v}'\big] = \mathbb{B}(\boldsymbol{v}') - (\mathbb{A}\boldsymbol{v}')\, (\mathbb{A}\boldsymbol{v}')^{\mathrm{T}} = \mathbb{B}^{(0)} + \mathbb{B}^{(1)}(\boldsymbol{v}'). \tag{3.33}$$

We write $\overline{\mathbb{B}}(\boldsymbol{v}') \equiv \mathbb{B}^{(0)} + \mathbb{B}^{(1)}(\boldsymbol{v}')$. For N large enough, the conditional multinomial distribution $\mathsf{Prob}[\boldsymbol{v}|\boldsymbol{v}']$ can be approximated by the conditional multi-

variate normal distribution,

$$K(\boldsymbol{v}|\boldsymbol{v}') = \frac{1}{\sqrt{\det 2\pi\,\overline{\mathbb{B}}(\boldsymbol{v}')}} \times \exp\left\{-\frac{1}{2}\big[\boldsymbol{v} - (\mathbb{I}+\mathbb{A})\boldsymbol{v}'\big]^{\mathrm{T}}\,\overline{\mathbb{B}}(\boldsymbol{v}')^{-1}\big[\boldsymbol{v} - (\mathbb{I}+\mathbb{A})\boldsymbol{v}'\big]\right\}, \tag{3.34}$$

which describes the transition process during a single non-tracked memory storage event. We will use this kernel below to obtain numerical solutions to Eq. (3.30). If instead of this normal kernel we consider a different, formal kernel $K_0(\boldsymbol{v}|\boldsymbol{v}')$ with the same conditional first- and second-order statistics as $K(\boldsymbol{v}|\boldsymbol{v}')$,

$$K_0(\boldsymbol{v}|\boldsymbol{v}') = \left\{1 - (\mathbb{A}\boldsymbol{v}')\cdot\boldsymbol{\nabla}_{\boldsymbol{v}} + \frac{1}{2}\,\boldsymbol{\nabla}_{\boldsymbol{v}}\cdot\mathbb{B}(\boldsymbol{v}')\boldsymbol{\nabla}_{\boldsymbol{v}}\right\}\delta(\boldsymbol{v}-\boldsymbol{v}'), \tag{3.35}$$

where $\boldsymbol{\nabla}_{\boldsymbol{v}} = (\partial/\partial v_1, \ldots, \partial/\partial v_{\nu-1})^{\mathrm{T}}$ and we remark that $\boldsymbol{\nabla}_{\boldsymbol{v}}$ does not act on functions of $\boldsymbol{v}'$, then Eq. (3.32) reduces identically to the Laplace transform of Eq. (3.27). This formal kernel therefore corresponds to taking the diffusion limit. Because the conditional moments of $K(\boldsymbol{v}|\boldsymbol{v}_0)$ and $K_0(\boldsymbol{v}|\boldsymbol{v}_0)$ match, we expect good agreement between results using $K(\boldsymbol{v}|\boldsymbol{v}_0)$ and $K_0(\boldsymbol{v}|\boldsymbol{v}_0)$, and therefore between the (up to the continuum limit) exact results from Eq. (3.30) and the diffusion-limit results from Eq. (3.28).

To compute FPT moments averaged over the tracked memory $\boldsymbol{\xi}^0$, we require the distribution of the initial tilded synaptic strength configuration $\boldsymbol{v}_0$. The initial distribution of any synapse's tilded strength after the storage of $\boldsymbol{\xi}^0$ is $\boldsymbol{p}(0) = \mathbb{M}^{+}\boldsymbol{A}$, and we write $\boldsymbol{p}(0) = \boldsymbol{p}_0$ for ease. As the N synapses update independently during the storage of $\boldsymbol{\xi}^0$, the MGF of the distribution of $\boldsymbol{v}_0$ is

just $(\boldsymbol{\omega} \cdot \boldsymbol{p}_0)^N$. The moments follow directly, and we obtain

$$\mathsf{E}[(\boldsymbol{v}_0)_i] = \boldsymbol{e}^i \cdot \boldsymbol{p}_0, \tag{3.36a}$$

$$\mathsf{E}[(\boldsymbol{v}_0)_i\,(\boldsymbol{v}_0)_j] - \mathsf{E}[(\boldsymbol{v}_0)_i]\,\mathsf{E}[(\boldsymbol{v}_0)_j] = \frac{1}{N}\big[\,\big(\boldsymbol{e}^i \circ \boldsymbol{e}^j\big) \cdot \boldsymbol{p}_0 - \big(\boldsymbol{e}^i \cdot \boldsymbol{p}_0\big)\,\big(\boldsymbol{e}^j \cdot \boldsymbol{p}_0\big)\,\big], \tag{3.36b}$$

for the means and covariances, respectively, of the components. As the distribution of $\boldsymbol{v}_0$ is again a multinomial distribution, for N large enough we can take it to be well approximated by a multivariate normal distribution with these means and covariances.

4 Approximation as an OU Process

Working with a Fokker-Planck equation in $\nu - 1$ variables is in general difficult unless the equation reduces to that for a Wiener or OU process (see, e.g., Risken, 1984). Fortunately, we can approximate the system an as OU process, as we now show. We then use these OU dynamics to develop tractable analytical and numerical approximation methods for computing FPT moments.

4.1 Decoupling of Odd and Even Variables

The diffusion matrix $\mathbb{B}(\boldsymbol{v})$ is in general $\boldsymbol{v}$-dependent. This dependence arises through the quadratic term $(\mathbb{A}\boldsymbol{v})(\mathbb{A}\boldsymbol{v})^{\mathrm{T}}$ and the linear term $\mathbb{B}^{(1)}(\boldsymbol{v})$. Since $\mathbb{A}$ is just $\mathbb{G}$ in a different basis, the matrix $\mathbb{A}$ is $\mathcal{O}(p)$ because $\mathbb{G}_{\mathrm{C}} = p\,\mathbb{C}$ or $\mathbb{G}_{\mathrm{Q}} = p\,\mathbb{Q}$ for the two matrices $\mathbb{M}_{\mathrm{C}}$ or $\mathbb{M}_{\mathrm{Q}}$, respectively. The parameter p controls the probability of synaptic plasticity and is typically taken to be small compared to unity. Thus, we can drop $\mathcal{O}(p^2)$ terms, and we have shown in our earlier analysis with the eigenvector requirement that this approximation is extremely

good, with negligible impact on results (Elliott, 2017b). We can therefore drop the $(\mathbb{A}\boldsymbol{v})(\mathbb{A}\boldsymbol{v})^{\mathrm{T}}$ term from $\mathbb{B}(\boldsymbol{v})$, which then becomes purely $\mathcal{O}(1/N)$.

The drift terms $(\mathbb{A}\boldsymbol{v})_i$ and diffusion terms $\mathbb{B}_{ij}(\boldsymbol{v})$ are, with this approximation, linear in $\boldsymbol{v}$. We examine the block structures of $\mathbb{A}$ and $\mathbb{B}(\boldsymbol{v})$ by re-ordering the elements of the vector $\boldsymbol{v}$ so that all the "odd" variables of the form v_{2i+1} appear before all the "even" variables of the form v_{2i}. This distinction is warranted by the "parity" difference between the odd basis vectors $\boldsymbol{e}^{2m+1}$ and the even basis vectors $\boldsymbol{e}^{2m}$. By our choice of normalisation we have that $e_\nu^m = +1$ $\forall m$, but $e_1^m = (-1)^m$ $\forall m$ so that the sign of e_1^m depends on the parity of m. This parity difference reflects the more general property that the even basis vectors $\boldsymbol{e}^{2m}$ are symmetric about their centres, meaning that $\boldsymbol{e}_i^{2m} = +\boldsymbol{e}_{\nu+1-i}^{2m}$, while the odd basis vectors $\boldsymbol{e}^{2m+1}$ are antisymmetric, or $\boldsymbol{e}_i^{2m+1} = -\boldsymbol{e}_{\nu+1-i}^{2m+1}$. A critical property of transition matrices $\mathbb{M}$, and therefore of generating matrices $\mathbb{G} = \mathbb{M} - \mathbb{I}$, is that when potentiation and depression processes are treated symmetrically, $\mathbb{M}$ (or $\mathbb{G}$) preserves parity, so that the vectors $\boldsymbol{a}$ and $\mathbb{M}\boldsymbol{a}$ have identical parity (Elliott, 2016b). Finally, it is clear that the dot product of a symmetric and an antisymmetric vector vanishes identically. Using these facts, by examining the forms of the matrices $\mathbb{A}$, $\mathbb{B}^{(0)}$ and $\mathbb{B}^{(1)}(\boldsymbol{v})$, it is easy to see that in the re-ordered basis for $\boldsymbol{v}$ they take the block forms

$$\mathbb{A} \sim \left(\begin{array}{c|c} \mathbb{A}_{\mathrm{odd}} & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{A}_{\mathrm{even}} \end{array}\right) \quad \text{and} \quad \mathbb{B}^{(0)} \sim \left(\begin{array}{c|c} \mathbb{B}^{(0)}_{\mathrm{odd}} & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{B}^{(0)}_{\mathrm{even}} \end{array}\right), \tag{4.1}$$

and

$$\mathbb{B}^{(1)}(\boldsymbol{v}) \sim \left(\begin{array}{c|c} \mathrm{Even} & \mathrm{Odd} \\ \hline \mathrm{Odd} & \mathrm{Even} \end{array}\right), \tag{4.2}$$

by which we mean that only odd and even variables appear in the indicated

blocks. The form for $\mathbb{B}^{(1)}(\boldsymbol{v})$ follows because the circle vector operator "$\circ$" essentially performs an "XNOR" (not exclusive-or) operation on parity.

We are interested in the odd variables specifically because they contain the perceptron activation $v_1 \equiv h$. In terms of the drift or Liouville dynamics controlled by $\mathbb{A}$, the odd and even variables completely decouple. Nor do they mix through the constant part of the diffusion matrix $\mathbb{B}^{(0)}$, since the matrices $\mathbb{A}$ and $\mathbb{B}^{(0)}$ are closely-related. The even variables affect the evolution of the odd variables only through $\mathbb{B}^{(1)}(\boldsymbol{v})$, and this term is $\mathcal{O}(1/N)$, so the even variables' influence on the odd variables' dynamics is relatively weak. In fact, the even variables' influence is extremely weak, as a direct calculation of $\mathsf{E}[v_{2i}(t)]$ shows. We find that $\boldsymbol{p}_0 = \mathbb{M}^{+}\boldsymbol{A}$ can be written as $\boldsymbol{p}_0 = \boldsymbol{A} + \frac{p}{\nu}\,\boldsymbol{\Phi}$ where $\boldsymbol{\Phi} = \boldsymbol{\Delta}$ with $\boldsymbol{\Delta} = (-1, 0, \cdots, 0, +1)^{\mathrm{T}}$ for $\mathbb{M}_{\mathrm{C}}$ or $\boldsymbol{\Phi} = \boldsymbol{\mathcal{L}}$ for $\mathbb{M}_{\mathrm{Q}}$ (Elliott, 2017b). In both cases the vector $\boldsymbol{\Phi}$ is antisymmetric or odd. The distribution $\boldsymbol{p}(t)$ is then given by

$$\boldsymbol{p}(t) = \boldsymbol{A} + \tfrac{p}{\nu} e^{rt\mathbb{G}}\,\boldsymbol{\Phi}, \tag{4.3}$$

and we have that $\mathsf{E}[v_i(t)] = \boldsymbol{e}^i \cdot \boldsymbol{p}(t)$, and $\boldsymbol{e}^i \cdot \boldsymbol{A} \equiv 0$ for $i > 0$. Since $\mathbb{G}$ preserves parity, so does $e^{rt\mathbb{G}}$. Thus, $e^{rt\mathbb{G}}\,\boldsymbol{\Phi}$ is odd while $\boldsymbol{e}^{2i}$ is even. Hence,

$$\mathsf{E}[v_{2i}(t)] = \boldsymbol{e}^{2i} \cdot \boldsymbol{p}(t) \equiv 0, \tag{4.4}$$

so that the mean values of all the even variables vanish for all time. In terms of their typical, mean values, the even variables therefore do not on average influence the dynamics of the odd variables at all, since on average the relevant block of $\mathbb{B}^{(1)}(\boldsymbol{v})$ vanishes. Any contributions that the even variables make to the odd variables' dynamics will arise purely through fluctuations in the even variables' values, and these fluctuations are driven by the diffusion term, which

is already $\mathcal{O}(1/N)$. Hence, the even variables' influence on the evolution of the odd variables arises only at $\mathcal{O}(1/N^{3/2})$, which is smaller than the $\mathcal{O}(1/N)$ diffusion term.

In summary, the even variables influence the odd variables only through terms at most $\mathcal{O}(1/N^{3/2})$, which is beyond the level of approximation of the Fokker-Planck equation at $\mathcal{O}(p)$ with its $\mathcal{O}(1/N)$ diffusion term. Hence, we can simply throw away $\mathbb{B}^{(1)}(\boldsymbol{v})$. The even variables then completely decouple from the odd variables and we can ignore the former. We thus write $\mathbb{B} \approx \mathbb{B}^{(0)}$, which is now a constant matrix independent of $\boldsymbol{v}$. For consistency, we must discard the $\mathcal{O}(p^2)$ terms in $\mathbb{B}^{(0)}$, which is just the $\mathbb{A}^2$ term in Eq. (3.18), so we write

$$\mathbb{B} \approx -\frac{2}{N}\frac{1}{\nu}\,\mathbb{A}\,\mathbb{D}. \tag{4.5}$$

Note that with this approximation, the matrix $\mathbb{T}$ defined above diagonalises both $\mathbb{A}$ and $\mathbb{B}$, i.e. it diagonalises the drift and diffusion matrices that arise in the Fokker-Planck equation.

We saw earlier that when the eigenvector condition is satisfied, $B(h)$ in Eq. (2.25b) and $B_{hh}(h)$ in Eq. (3.23b) differ by terms arising from $\mathbb{B}^{(1)}_{11}(\boldsymbol{v})$. Because of the even parity of $\boldsymbol{\Omega} \circ \boldsymbol{\Omega}$ in Eq. (3.23b), the sum in $B_{hh}(h)$ is over even variables only, so it can be neglected in a $1/N$ approximation, confirming our claim above that when the eigenvector condition is satisfied, the $\boldsymbol{v}$ dynamics can be reduced down to dynamics purely in h.

4.2 Dynamics in the OU Limit

With drift terms of the form $(\mathbb{A}\boldsymbol{v})_i$ and a constant diffusion matrix $\mathbb{B} = -\frac{2}{N}\frac{1}{\nu}\,\mathbb{A}\,\mathbb{D}$, the Fokker-Planck equation in Eq. (3.12) describes an OU process in $\nu - 1$ dimensions, but a process in which the even degrees of freedom are

completely decoupled and can be ignored. We can write down the standard unbounded solution to Eq. (3.12) in this OU limit (van Kampen, 1992). Setting $t_0 = 0$ s and writing $P(\boldsymbol{v} \,|\, \boldsymbol{v}_0; t)$ for $P(\boldsymbol{v}, t \,|\, \boldsymbol{v}_0, t_0)$, we have

$$P(\boldsymbol{v} \,|\, \boldsymbol{v}_0; t) = \frac{1}{\sqrt{\det 2\pi\, \mathbb{S}(t)}} \exp\left[-\frac{1}{2}\left(\boldsymbol{v} - e^{rt\mathbb{A}}\boldsymbol{v}_0\right)^{\mathrm{T}} \mathbb{S}(t)^{-1} \left(\boldsymbol{v} - e^{rt\mathbb{A}}\boldsymbol{v}_0\right)\right], \tag{4.6}$$

where $\mathbb{S}(t)$ is the conditional covariance matrix. In this OU limit, $\mathbb{S}(t)$ satisfies the equation (cf. 3.14b)

$$\frac{1}{r}\frac{d}{dt}\mathbb{S}(t) = \mathbb{A}\,\mathbb{S}(t) + \mathbb{S}(t)\mathbb{A}^{\mathrm{T}} + \mathbb{B}, \tag{4.7}$$

subject to $\mathbb{S}(0) \equiv \mathbb{O}$. Unlike $\mathbb{S}_{\boldsymbol{v}_0}(t)$ for the full dynamics, the conditional covariance matrix $\mathbb{S}(t)$ in the OU limit does not depend on $\boldsymbol{v}_0$ (the right hand side of Eq. (4.7) does not depend on $\boldsymbol{v}_0$), so we have dropped the $\boldsymbol{v}_0$ subscript from $\mathbb{S}_{\boldsymbol{v}_0}(t)$. The standard solution (van Kampen, 1992) for $\mathbb{S}(t)$ is

$$\mathbb{S}(t) = \int_0^t r dt'\, e^{r(t-t')\mathbb{A}}\, \mathbb{B}\, e^{r(t-t')\mathbb{A}^{\mathrm{T}}}. \tag{4.8}$$

Using $\mathbb{B} = -\frac{2}{N}\frac{1}{\nu}\,\mathbb{A}\,\mathbb{D}$ and the similarity relation $\mathbb{D}^{-1}\mathbb{A}\mathbb{D} = \mathbb{A}^{\mathrm{T}}$ we may simplify and evaluate this integral, to obtain the explicit result

$$\mathbb{S}(t) = \frac{1}{N}\frac{1}{\nu}\big[\,\mathbb{I} - \exp\left(2rt\mathbb{A}\right)\big]\mathbb{D}. \tag{4.9}$$

Note the remarkable fact that despite the OU approximation, $\lim_{t\to\infty}\mathbb{S}(t) = \frac{1}{N}\frac{1}{\nu}\,\mathbb{D}$, which agrees with the exact equilibrium form given in Eq. (3.21).

In the OU limit, Eq. (3.28) for the FPT moments can be written as

$$-\frac{1}{r} n\, \tau^{(n-1)}(\boldsymbol{v}_0) = \left[(\mathbb{A}\boldsymbol{v}_0)\cdot\boldsymbol{\nabla}_{\boldsymbol{v}_0} + \frac{1}{2}\boldsymbol{\nabla}_{\boldsymbol{v}_0}\cdot\,\mathbb{B}\,\boldsymbol{\nabla}_{\boldsymbol{v}_0}\right]\tau^{(n)}(\boldsymbol{v}_0), \tag{4.10}$$

because $\mathbb{B} = -\frac{2}{N}\frac{1}{\nu}\,\mathbb{A}\,\mathbb{D}$ is a constant matrix. We may immediately deduce some scaling properties of $\tau^{(n)}(\boldsymbol{v}_0)$ as a function of parameters. First, the drift term $(\mathbb{A}\boldsymbol{v}_0)\cdot\boldsymbol{\nabla}_{\boldsymbol{v}_0}$ is invariant under a change in the scale of $\boldsymbol{v}_0$, $\boldsymbol{v}_0 \to \gamma\,\boldsymbol{v}_0$, while the diffusion term $\boldsymbol{\nabla}_{\boldsymbol{v}_0}\cdot\mathbb{B}\,\boldsymbol{\nabla}_{\boldsymbol{v}_0}$ scales as γ^{-2}. But the diffusion matrix $\mathbb{B}$ scales inversely with N, so N can be absorbed into a re-definition of $\boldsymbol{v}_0$, say $\boldsymbol{v}_0^* = \sqrt{N}\,\boldsymbol{v}_0$. To compensate, the perceptron firing threshold θ must also be rescaled, so $\theta^* = \sqrt{N}\,\theta$ is the rescaled threshold. Second, both $\mathbb{A}$ and $\mathbb{B}$ scale linearly with p in this OU limit and furthermore, for large enough ν, explicit calculation shows that both matrices approximately scale as either ν^{-1} or ν^{-2} for either the $\mathbb{Q}$ model or the $\mathbb{C}$ model, respectively, with the scaling becoming exact in the limit. Thus, we rewrite Eq. (4.10) in the form

$$-n\big(\tfrac{p\,r}{\nu^\kappa}\big)^{n-1}\tau^{(n-1)}(\boldsymbol{v}_0^*) = \Big[\big(\tfrac{\nu^\kappa}{p}\mathbb{A}\boldsymbol{v}_0^*\big)\cdot\boldsymbol{\nabla}_{\boldsymbol{v}_0^*} + \tfrac{1}{2}\boldsymbol{\nabla}_{\boldsymbol{v}_0^*}\cdot\big(\tfrac{N\nu^\kappa}{p}\mathbb{B}\big)\boldsymbol{\nabla}_{\boldsymbol{v}_0^*}\Big]\big(\tfrac{p\,r}{\nu^\kappa}\big)^n\tau^{(n)}(\boldsymbol{v}_0^*), \tag{4.11}$$

subject to $\big(\frac{p\,r}{\nu^\kappa}\big)^n\tau^{(n)}(\boldsymbol{v}_0^*) = 0$ on the absorbing hyperplane $h_0^* = \theta^*$, and with $\kappa = 1$ or $\kappa = 2$ depending on the model. In this equation, $\frac{\nu^\kappa}{p}\mathbb{A}$ and $\frac{N\nu^\kappa}{p}\mathbb{B}$ are independent of p and N, are are approximately independent of ν, becoming increasingly insensitive to ν as ν increases. We write $\mathbb{A}^* = \lim_{\nu\to\infty}\frac{\nu^\kappa}{p}\,\mathbb{A}$ and $\mathbb{B}^* = \lim_{\nu\to\infty}\frac{N\nu^\kappa}{p}\,\mathbb{B}$ so that $\mathbb{A}^*$ and $\mathbb{B}^*$ are just matrices of numbers with no parameter dependence. The differential operator on the RHS of Eq. (4.11) is independent of r, p and N, and is nearly independent of ν. Any dependence on these parameters simply scales the solution or scales the coordinate system. With Eq. (4.11) we have therefore effectively eliminated three parameters entirely and four if $\theta = 0$, in the sense that if we obtain one solution for one set of values of these three or four parameters, then we may simply rescale that solution into solutions for other values of these parameters. In particular, for large ν and $\theta = 0$ we have only two distinct FPT equations based on

the Fokker-Planck approach: we need only solve Eq. (4.11) for one choice of parameters for each of the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ models, and then we have solutions for all choices of parameters. The MFPT in particular therefore scales inversely with p and r (the latter being trivial), linearly or quadratically with ν for large enough ν, and is a function of $\sqrt{N}$. These scaling properties are useful for checking numerical solutions, although in practice grid discretisation makes scaling one numerical solution for one set of parameters into a different solution for a different set of parameters impractical.

For the integral equation approach to computing FPT moments in Eq. (3.30), the full, unapproximated kernel $K(\boldsymbol{v}|\boldsymbol{v}')$ contains the covariance matrix $\overline{\mathbb{B}}(\boldsymbol{v}') = \mathbb{B}^{(0)} + \mathbb{B}^{(1)}(\boldsymbol{v}')$. In the OU approximation, we drop $\mathbb{B}^{(1)}(\boldsymbol{v}')$ to decouple the even variables, so that $\overline{\mathbb{B}} = \mathbb{B}^{(0)}$ becomes a constant matrix, where we have dropped the argument for this reason. For consistency with the $\mathcal{O}(p)$ approximation used to obtain the OU limit, we should also write $\overline{\mathbb{B}} = -\frac{2}{N}\frac{1}{\nu}\,\mathbb{A}\,\mathbb{D} = \mathbb{B}$. However, we note that $K(\boldsymbol{v}|\boldsymbol{v}')$ is, up to the large N normal form, an exact conditional transition probability and thus induces the correct, exact jump moments at all orders. In particular, it induces the $\mathcal{O}(p^2)$ quadratic term $(\mathbb{A}\boldsymbol{v}')(\mathbb{A}\boldsymbol{v}')^{\mathrm{T}}$ in the second-order jump moment, which we discarded in the Fokker-Planck equation to obtain the OU limit. Thus, the integral equation approach to FPTs does not permit a controlled, $\mathcal{O}(p)$ approximation, because it induces jump moments at all orders, unlike the Fokker-Planck approach, which is based on a second-order truncation of the Kramers-Moyal expansion. Below, we will replace $\overline{\mathbb{B}}(\boldsymbol{v}')$ with $\mathbb{B}$ in $K(\boldsymbol{v}|\boldsymbol{v}')$, with the understanding that we may use the $\mathcal{O}(p)$ OU form, $\mathbb{B} = -\frac{2}{N}\frac{1}{\nu}\,\mathbb{A}\,\mathbb{D}$, or the full, $\mathcal{O}(p^2)$ form, $\mathbb{B} = \mathbb{B}^{(0)} = -\frac{1}{N}\frac{1}{\nu}\,(2\mathbb{A} + \mathbb{A}^2)\mathbb{D}$. Provided that p is small enough, the difference between using these two forms will of course in practice be negligible. These considerations also affect the scaling behaviour of $\tau^{(n)}(\boldsymbol{v}_0)$. The transformation $\boldsymbol{v}^* = \sqrt{N}\,\boldsymbol{v}$ under the integral sign

in Eq. (3.30) removes the N dependence as in the Fokker-Planck approach to FPTs, and $\tau^{(n)}(\boldsymbol{v}_0)$ trivially scales with the rate r. However, the scaling with p and ν that we observed for $\tau^{(n)}(\boldsymbol{v}_0)$ in the Fokker-Planck approach does not carry over to the integral equation approach, precisely because the kernel $K(\boldsymbol{v}|\boldsymbol{v}')$ induces jump moments at all orders. Again, though, provided that p is small enough and ν is large enough, we would expect to see approximate scaling behaviour emerge in $\tau^{(n)}(\boldsymbol{v}_0)$ obtained via the integral equation approach.

Performing the transformation of variables $\boldsymbol{x} = \mathbb{T}\boldsymbol{v}_0$ on Eq. (4.10), where $\boldsymbol{x} = (x_1, \ldots, x_{\nu-1})^{\mathrm{T}}$, we obtain

$$-\frac{1}{r} n\, \tau^{(n-1)}(\boldsymbol{x}) = \left[\left(\mathbb{T}\mathbb{A}\mathbb{T}^{-1}\boldsymbol{x} \right) \cdot \boldsymbol{\nabla}_{\boldsymbol{x}} + \frac{1}{2} \boldsymbol{\nabla}_{\boldsymbol{x}} \cdot \left(\mathbb{T}\,\mathbb{B}\,\mathbb{T}^{\mathrm{T}} \right) \boldsymbol{\nabla}_{\boldsymbol{x}} \right] \tau^{(n)}(\boldsymbol{x}). \tag{4.12}$$

We discuss the transformed boundary conditions later. We pick $\mathbb{T}$ as above to be the matrix whose rows are the normalised left eigenvectors of $\mathbb{A}$. Because of the decoupling of odd and even variables, the odd (respectively, even) components of $\boldsymbol{x}$ depend only on the odd (respectively, even) components of $\boldsymbol{v}$. With this choice of $\mathbb{T}$, we know that both $\mathbb{T}\mathbb{A}\mathbb{T}^{-1}$ and $\mathbb{T}\,\mathbb{B}\,\mathbb{T}^{\mathrm{T}}$ are diagonalised by this transformation. Since $\mathbb{B} = -\frac{2}{N}\frac{1}{\nu}\,\mathbb{A}\,\mathbb{D}$, we have $\left(\mathbb{T}\,\mathbb{B}\,\mathbb{T}^{\mathrm{T}} \right)_{ii} = -\frac{2}{N}\frac{1}{\nu}(\lambda_i - 1)\left(\widehat{\boldsymbol{\lambda}}^i \cdot \mathbb{D}\, \widehat{\boldsymbol{\lambda}}^i \right)$ where $\widehat{\boldsymbol{\lambda}}^i \cdot \mathbb{D}\, \widehat{\boldsymbol{\lambda}}^i = \left| \boldsymbol{\lambda}^i \right|^{-2}$, so Eq. (4.12) is just

$$-\frac{1}{r} n\, \tau^{(n-1)}(\boldsymbol{x}) = \sum_{i=1}^{\nu-1} (1 - \lambda_i) \left\{ -x_i \frac{\partial}{\partial x_i} + \frac{1}{N}\frac{1}{\nu} \left(\widehat{\boldsymbol{\lambda}}^i \cdot \mathbb{D}\, \widehat{\boldsymbol{\lambda}}^i \right) \frac{\partial^2}{\partial x_i^2} \right\} \tau^{(n)}(\boldsymbol{x}). \tag{4.13}$$

For the integral equation for FPTs, we have

$$\tau^{(n)}(\boldsymbol{v}_0) = \frac{1}{r}\, n\, \tau^{(n-1)}(\boldsymbol{v}_0) + \int_{\mathcal{I}} \cdots \int d\boldsymbol{v}\, \tau^{(n)}(\boldsymbol{v})\, K(\boldsymbol{v}|\boldsymbol{v}_0), \tag{4.14}$$

where we now write the kernel as

$$K(\boldsymbol{v}|\boldsymbol{v}') = \frac{1}{\sqrt{\det 2\pi\,\mathbb{B}}} \exp\left\{-\frac{1}{2}\big[\boldsymbol{v} - (\mathbb{I}+\mathbb{A})\boldsymbol{v}'\big]^{\mathrm{T}}\,\mathbb{B}^{-1}\big[\boldsymbol{v} - (\mathbb{I}+\mathbb{A})\boldsymbol{v}'\big]\right\}. \tag{4.15}$$

We perform the same $\mathbb{T}$ transformation on the variables in Eq. (4.14), where now we must write $\boldsymbol{x} = \mathbb{T}\boldsymbol{v}_0$ and $\boldsymbol{y} = \mathbb{T}\boldsymbol{v}$. The Jacobian matrix is just $\left|\frac{\partial \boldsymbol{v}}{\partial \boldsymbol{y}}\right| = |\det \mathbb{T}|^{-1}$, and we choose sign conventions for the normalised left eigenvectors $\widehat{\boldsymbol{\lambda}}^i$ so that $\det \mathbb{T} > 0$. Since $\big(\det \mathbb{T}\big)\big(\sqrt{\det \mathbb{B}}\big) = \sqrt{\det\big(\mathbb{T}\,\mathbb{B}\,\mathbb{T}^{\mathrm{T}}\big)}$, Eq. (4.14) can be written as

$$\tau^{(n)}(\boldsymbol{x}) = \frac{1}{r}\,n\,\tau^{(n-1)}(\boldsymbol{x}) + \int_{\mathcal{I}'}\!\!\cdots\!\!\int d\boldsymbol{y}\,\tau^{(n)}(\boldsymbol{y})\,\overline{K}(\boldsymbol{y}|\boldsymbol{x}), \tag{4.16}$$

where $\mathcal{I}'$ denotes the changed integration limits because of the transformation and the modified kernel $\overline{K}(\boldsymbol{x}|\boldsymbol{x}')$ is

$$\begin{aligned}\overline{K}(\boldsymbol{x}|\boldsymbol{x}') = {} & \frac{1}{\sqrt{\det 2\pi\mathbb{T}\,\mathbb{B}\,\mathbb{T}^{\mathrm{T}}}} \\ & \times \exp\left\{-\frac{1}{2}\big[\boldsymbol{x} - \big(\mathbb{I}+\mathbb{T}\mathbb{A}\mathbb{T}^{-1}\big)\boldsymbol{x}'\big]^{\mathrm{T}}\,\big(\mathbb{T}\,\mathbb{B}\,\mathbb{T}^{\mathrm{T}}\big)^{-1}\big[\boldsymbol{x} - \big(\mathbb{I}+\mathbb{T}\mathbb{A}\mathbb{T}^{-1}\big)\boldsymbol{x}'\big]\right\}.\end{aligned} \tag{4.17}$$

Because $\mathbb{T}\mathbb{A}\mathbb{T}^{-1}$ and $\mathbb{T}\,\mathbb{B}\,\mathbb{T}^{\mathrm{T}}$ are diagonal (regardless of whether we use the $\mathcal{O}(p)$ or $\mathcal{O}(p^2)$ form of $\mathbb{B}$), this transformation has therefore decoupled the variables in $\overline{K}(\boldsymbol{x}|\boldsymbol{x}')$, turning it into a simple product over separate, 1-dimensional conditional normal distributions. Unfortunately, the modified integration region $\mathcal{I}'$ prevents us from using this remarkable factorisation to evaluate $\nu - 2$ of the integrals in Eq. (4.16), except for the specific case $\mathbb{T} = \mathbb{I}$, which corresponds to $\mathbb{A}$ itself being diagonal. Nevertheless, the simplicity of the results in Eqs. (4.13) and (4.16) with the factorised kernel in Eq. (4.17) suggests an approximation

method, as we now discuss.

4.3 Approximation Method for $\boldsymbol{\tau^{(n)}(v_0)}$

Eq. (4.13) in the transformed variables $\boldsymbol{x} = \mathbb{T}\boldsymbol{v}_0$ admits of $\nu - 1$ solutions of the form $\tau^{(n)}(\boldsymbol{x}) = \tau^{(n)}(x_i)$, i.e. where the solution depends on only one component, but unless $\mathbb{A}$ is itself diagonal, no such solution nor any linear combination of them satisfies the boundary condition $\tau^{(n)}(\boldsymbol{v}_0) = 0$ on the absorbing hyperplane $h_0 = \theta$. Further, if the integration limits for the y_i, $i \neq 1$, variables in Eq. (4.16) are each between $-\infty$ and $+\infty$, then the integral equation allows a solution of the form $\tau^{(n)}(\boldsymbol{x}) = \tau^{(n)}(x_1)$ because of the factorised kernel $\overline{K}(\boldsymbol{x}|\boldsymbol{x}')$ in Eq. (4.17). In both differential and integral equations for the FPT moments, the dynamics are therefore nearly completely decoupled, but the boundary conditions prevent exact decoupling unless $\mathbb{A}$ is diagonal.

However, if the least negative eigenvalue $\lambda_1 - 1$ of $\mathbb{A}$ is well separated from the other odd eigenvalues, then we expect that FPT processes through the absorbing hyperplane $h_0 = \theta$ will be dominated by the slowest eigenmode corresponding to $x_1 = \widehat{\boldsymbol{\lambda}}^1 \cdot \boldsymbol{v}_0$. With such separation, the other odd eigenmodes will on average decay rapidly to zero compared to x_1. We have that

$$\frac{\lambda_1 - 1}{\lambda_{2m+1} - 1} = \begin{cases} \frac{1-\cos\frac{\pi}{\nu}}{1-\cos\frac{(2m+1)\pi}{\nu}} \sim \frac{1}{(2m+1)^2} & \text{for } \mathbb{C}\&\boldsymbol{\mathcal{L}} \text{ model} \\ \frac{1}{(m+1)(2m+1)} & \text{for } \mathbb{Q}\&\boldsymbol{\mathcal{S}} \text{ model} \end{cases}, \tag{4.18}$$

so that indeed the eigenvalues are well separated. The dynamics can thus be considered as self-restricting to the 1-dimensional subspace governed by x_1. In such a scenario, we may obtain a good approximation to $\tau^{(n)}(\boldsymbol{v}_0)$ by throwing away all but the x_1 eigenmode contributing to Eq. (4.13). Neglecting the other eigenmodes would be expected to have a significant impact only when

$\boldsymbol{v}_0$ is close to the absorbing hyperplane, near which the dynamics of the faster eigenmodes may also be important. Away from the hyperplane, we expect that neglecting the other eigenmodes has little impact, so that FPT processes are dominated by the slowest eigenmode. We can then reduce Eq. (4.13) to

$$-\frac{1}{r}\frac{1}{(1-\lambda_1)}n\tau^{(n-1)}(x_1) = \left\{-x_1\frac{d}{dx_1} + \frac{1}{N}\frac{1}{\nu}\big(\widehat{\boldsymbol{\lambda}}^1\cdot\mathbb{D}\,\widehat{\boldsymbol{\lambda}}^1\big)\frac{d^2}{dx_1^2}\right\}\tau^{(n)}(x_1). \quad (4.19)$$

The location of the absorbing boundary in this 1-dimensional subspace, at say $x_1 = \theta'$, is determined by solving $(\theta', 0, \ldots, 0)^{\mathrm{T}} = \mathbb{T}(\theta, (\boldsymbol{v}_0)_2, \ldots, (\boldsymbol{v}_0)_{\nu-1})^{\mathrm{T}}$ for θ', giving

$$\theta' = \frac{1}{\big|\boldsymbol{\lambda}^1\big|}\frac{\theta}{\big(\boldsymbol{\rho}^1\big)_1} = \frac{1}{\big|\boldsymbol{\lambda}^1\big|}\frac{\theta}{\boldsymbol{e}^1\cdot\widehat{\boldsymbol{\varepsilon}}^{\,1}}, \quad (4.20)$$

and we solve Eq. (4.19) subject to $\tau^{(n)}(\theta') = 0$ for $n \geq 1$. For the particular case of the MFPT, $n = 1$, we therefore have the approximate solution

$$\tau_{\mathrm{mfpt}}(\boldsymbol{v}_0) \approx \begin{cases} \tau_1\big(\widehat{\boldsymbol{\lambda}}^1\cdot\boldsymbol{v}_0\big) - \tau_1(\theta') & \text{for } \widehat{\boldsymbol{\lambda}}^1\cdot\boldsymbol{v}_0 \geq \theta' \\ 0 & \text{for } \widehat{\boldsymbol{\lambda}}^1\cdot\boldsymbol{v}_0 < \theta' \end{cases}, \quad (4.21)$$

where $\tau_1(x_1)$ is the solution of Eq. (4.19) for $n = 1$ with $\tau_1(0) = 0$, given by

$$\begin{aligned}(1-\lambda_1)r\tau_1(x_1) = {}& \frac{\pi}{2}\,\mathrm{erfi}\left(\sqrt{\frac{N\nu}{2\,\widehat{\boldsymbol{\lambda}}^1\cdot\mathbb{D}\,\widehat{\boldsymbol{\lambda}}^1}}\,x_1\right) \\ & - \frac{N\nu x_1^2}{2\,\widehat{\boldsymbol{\lambda}}^1\cdot\mathbb{D}\,\widehat{\boldsymbol{\lambda}}^1}\,{}_2F_2\left(\{1,1\},\{3/2,2\};\frac{N\nu x_1^2}{2\,\widehat{\boldsymbol{\lambda}}^1\cdot\mathbb{D}\,\widehat{\boldsymbol{\lambda}}^1}\right), \quad (4.22)\end{aligned}$$

with erfi and ${}_2F_2$ being the imaginary error function and a hypergeometric function, respectively. Under this approximation, the integral equation for $\tau^{(n)}(\boldsymbol{x})$ in Eq. (4.16) becomes the 1-dimensional integral equation

$$\tau^{(n)}(x_1) = \frac{1}{r}\,n\,\tau^{(n-1)}(x_1) + \int_{\theta'}^{\infty} dy_1\,\tau^{(n)}(y_1)\,\overline{K}_1(y_1|x_1), \quad (4.23)$$

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

$$\overline{K}_1(x|x') = \sqrt{\frac{N\nu}{2\pi\big[(1-\lambda_1^2)\,\widehat{\boldsymbol{\lambda}}^{\,1}\cdot\mathbb{D}\,\widehat{\boldsymbol{\lambda}}^{\,1}\big]}}\exp\left\{-\frac{N\nu\big(x-\lambda_1 x'\big)^2}{2\big[(1-\lambda_1^2)\,\widehat{\boldsymbol{\lambda}}^{\,1}\cdot\mathbb{D}\,\widehat{\boldsymbol{\lambda}}^{\,1}\big]}\right\}, \quad (4.24)$$

which is just the 1-dimensional normal factor in x_1 in the $(\nu-1)$-dimensional kernel $\overline{K}(\boldsymbol{x}|\boldsymbol{x}')$ in Eq. (4.17). Here we have used the $\mathcal{O}(p^2)$ form for $\mathbb{B}$; for the $\mathcal{O}(p)$ form, we replace $(1-\lambda_1^2)$ by $2(1-\lambda_1)$. For the MFPT in particular, we write

$$\tau_{\rm mfpt}(\boldsymbol{v}_0) \approx \tau_2\big(\widehat{\boldsymbol{\lambda}}^{\,1}\cdot\boldsymbol{v}_0\big), \quad (4.25)$$

where $\tau_2(x_1)$ is the solution of

$$\tau_2(x_1) = \frac{1}{r} + \int_{\theta'}^{\infty} dy_1\,\tau_2(y_1)\,\overline{K}_1(y_1|x_1). \quad (4.26)$$

The jump moments $A(x') = \mathsf{E}[(x-x')^1|x']$ and $B(x') = \mathsf{E}[(x-x')^2|x']$ induced by the kernel $\overline{K}_1(x|x')$ are

$$A(x) = -(1-\lambda_1)x, \quad (4.27a)$$

$$B(x) = (1-\lambda_1)^2x^2 + \frac{1-\lambda_1^2}{N}\,\frac{1}{\nu}\,\widehat{\boldsymbol{\lambda}}^{\,1}\cdot\mathbb{D}\,\widehat{\boldsymbol{\lambda}}^{\,1}, \quad (4.27b)$$

which should be compared to Eq. (2.25b). For the particular case that $\mathbb{A}$ is diagonal, when the eigenvector condition is satisfied, $\widehat{\boldsymbol{\lambda}}^1 = (1,0,\ldots,0)^{\rm T}$ so that $x = \widehat{\boldsymbol{\lambda}}^1\cdot\boldsymbol{v} \equiv h$ reduces identically to the perceptron activation, and $\frac{1}{\nu}\,\widehat{\boldsymbol{\lambda}}^{\,1}\cdot\mathbb{D}\,\widehat{\boldsymbol{\lambda}}^{\,1} = \frac{1}{\nu}\,\boldsymbol{e}^1\cdot\boldsymbol{e}^1 \equiv \langle\boldsymbol{\Omega}^2\rangle$. Hence, the jump moment $B(x)$ in Eq. (4.27b) is identical to that in Eq. (2.25b) when the eigenvector condition is satisfied. When the eigenvector condition is not satisfied, Eq. (4.27b) generalises the jump moment by working with the slowest eigenmode instead of the perceptron activation, with $\frac{1}{\nu}\,\widehat{\boldsymbol{\lambda}}^{\,1}\cdot\mathbb{D}\,\widehat{\boldsymbol{\lambda}}^{\,1}$ generalising $\langle\boldsymbol{\Omega}^2\rangle$ found in Eq. (2.25b).

4.4 Restriction of Dynamics to Two Variables

The approximation of discarding all but the slowest eigenmode x_1 greatly simplifies the equations for $\tau^{(n)}(\boldsymbol{v}_0)$, but: first, all odd v_m variables are still present in general in x_1; second, to compute $\langle \tau_{\text{mfpt}}(\boldsymbol{v}_0)\rangle_{\boldsymbol{v}_0|h_0>\theta}$ requires averaging over the entire distribution of the initial configuration $\boldsymbol{v}_0$; third, the approximation may be relatively poor when $\boldsymbol{v}_0$ is close to the absorbing hyperplane $h_0 = \theta$; fourth, obtaining numerical solutions for $\tau^{(n)}(\boldsymbol{v}_0)$ will involve solving a differential or integral equation in $\frac{1}{2}\nu$ (for even ν) or $\frac{1}{2}(\nu-1)$ (for odd ν) odd variables, which in general will be extremely hard. However, by examining $\mathsf{E}[v_m(t)]$ for the odd variables we show that essentially only $v_1(t) \equiv h(t)$ and $v_3(t)$ need be considered for FPT processes, allowing the dynamics to be restricted to just two variables. We need only consider the models for which the eigenvector condition is not satisfied.

First, consider the simpler case of the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model, for which we have that $\mathbb{M}_{\text{Q}}^{+}\boldsymbol{A} = \boldsymbol{A} + \frac{p}{\nu}\boldsymbol{\mathcal{L}}$ where $\boldsymbol{\mathcal{L}} = \boldsymbol{l}^1$, and so

$$\mathsf{E}[v_m(t)] = \frac{p}{v}\,\boldsymbol{s}^m \cdot e^{rt(\mathbb{M}_{\text{Q}}-\mathbb{I})}\boldsymbol{\mathcal{L}} = \frac{p}{\nu}\,e^{-r\left(1-\lambda_1^{\text{Q}}\right)t}\,\boldsymbol{s}^m \cdot \boldsymbol{l}^1, \tag{4.28}$$

where the λ_m^{Q} are the eigenvalues of $\mathbb{M}_{\text{Q}}$ given in Eq. (2.15). Only a single eigenmode contributes to $\mathsf{E}[v_m(t)]$ because $\boldsymbol{\mathcal{L}}$ is an eigenvector of $\mathbb{M}_{\text{Q}}$. All the odd $\mathsf{E}[v_m(t)]$ therefore decay at the same rate governed by $1 - \lambda_1^{\text{Q}} = \frac{p}{\nu-1}$. The overall sizes of the $\mathsf{E}[v_m(t)]$ are controlled by

$$\frac{1}{\nu}\,\boldsymbol{s}^m \cdot \boldsymbol{l}^1 = \frac{[1-(-1)^m]}{2\,\nu(\nu-1)}\,\text{cosec}^2\,\frac{\pi\,m}{2\,\nu} \sim \frac{2\,[1-(-1)^m]}{\pi^2\,m^2}, \tag{4.29}$$

where the last form is for large ν. Relative to $\mathsf{E}[v_1(t)]$, $\mathsf{E}[v_{2m+1}(t)]$ is therefore scaled down by a factor of around $1/(2m+1)^2$, so that $\mathsf{E}[v_5(t)]$ is only 4%

of $\mathsf{E}[v_1(t)]$. Essentially, then, at least in terms of the mean dynamics, we can neglect all but $v_1(t)$ and $v_3(t)$ since the other odd variables are suppressed.

Second, for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model, recalling that $\boldsymbol{\Delta} = (-1, 0, \cdots, 0, +1)^{\mathrm{T}}$, we have that $\mathbb{M}_{\mathrm{C}}^{+}\boldsymbol{A} = \boldsymbol{A} + \frac{p}{\nu}\boldsymbol{\Delta}$, and so

$$\begin{aligned}\mathsf{E}[v_m(t)] &= \frac{p}{v}\boldsymbol{l}^m \cdot e^{rt(\mathbb{M}_{\mathrm{C}}-\mathbb{I})}\boldsymbol{\Delta} \\ &= \frac{p}{\nu}\sum_{n=1}^{\nu-1} e^{-r(1-\lambda_n^{\mathrm{C}})t}\big[1-(-1)^n\big]\frac{\boldsymbol{l}^m\cdot\boldsymbol{s}^n}{\boldsymbol{s}^n\cdot\boldsymbol{s}^n}, \end{aligned} \tag{4.30}$$

where the λ_n^{C} are the eigenvectors of $\mathbb{M}_{\mathrm{C}}$ given in Eq. (2.10). We have used the fact that $\boldsymbol{s}^n \cdot \boldsymbol{\Delta} = 1 - (-1)^n$, so that only odd eigenmodes contribute to the sum. For this case, all odd eigenvectors of $\mathbb{M}_{\mathrm{C}}$ contribute to $\mathsf{E}[v_m(t)]$ because $\boldsymbol{\Delta}$ is not (except for $\nu = 2$ and $\nu = 3$) an eigenvector of $\mathbb{M}_{\mathrm{C}}$. These eigenmodes decay at rates governed by $1 - \lambda_n^{\mathrm{C}} = p\left(1 - \cos\frac{\pi n}{\nu}\right)$, so that modes with larger values of n drop out before those with smaller values. Nevertheless, the slowest-decaying, $n = 1$ mode contributes for any odd m and so for large enough time all $\mathsf{E}[v_m(t)]$ decay at a rate governed by $p\left(1 - \cos\frac{\pi}{\nu}\right)$. Similar to the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ case above, however, the sizes of $\mathsf{E}[v_m(t)]$ are then controlled by $\boldsymbol{l}^m \cdot \boldsymbol{s}^1$ (above it was $\boldsymbol{s}^m \cdot \boldsymbol{l}^1$). We then find that relative to $\mathsf{E}[v_1(t)]$, $\mathsf{E}[v_3(t)]$ is smaller by a factor of $\pi^2/[6(10 - \pi^2)] \approx 12.6$ while $\mathsf{E}[v_5(t)]$ is smaller by a factor of $\pi^4/[15(1008 - 112\pi^2 + \pi^4)] \approx 484.7$, where these forms are valid for large ν. Again, we need only consider the dynamics of $v_1(t)$ and $v_3(t)$ and can neglect the other odd variables.

We illustrate these considerations in Fig. 1 by explicitly plotting $\mathsf{E}[v_i(t)]$ for odd i against t for both these models, for $\nu = 16$. For the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model, the dynamics are especially clear because of the presence of only a single eigenmode. All but $\mathsf{E}[v_1(t)]$ and $\mathsf{E}[v_3(t)]$ are negligible. The dynamics for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model are complicated by the presence of all odd eigenmodes, introducing con-

tributions with different decay rates into each $\mathsf{E}[v_i(t)]$. Nevertheless, at the larger times relevant for FPT processes in $v_1(t) \equiv h(t)$, again only $\mathsf{E}[v_1(t)]$ and $\mathsf{E}[v_3(t)]$ are non-negligible.

FIG. 1 ABOUT HERE.

The variables v_i with larger odd i are weakly coupled to the variables with smaller odd i, because of the former's smaller expectation values. Further coupling also occurs via the diffusion matrix, but as we argued for the even variables, any contributions arising via the diffusion term with be $\mathcal{O}(1/N^{3/2})$ and can be neglected in a $1/N$ approximation. In conclusion, larger odd i variables only very mildly influence the evolution of the smaller odd i variables, especially at larger times. Any strong coupling at small times in the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model will be irrelevant for FPT dynamics occurring on much longer timescales. We can therefore make an approximation by discarding odd variables with large i. While it is better in terms of precision to retain as many odd variables as possible, the requirement of making a tractable approximation demands throwing away as many as possible. Thus, we retain $v_1(t)$ and $v_3(t)$ only, and find that this is good enough.

The variable $v_1(t)$ is just the perceptron activation $h(t)$ of interest to us, and we now write $w(t) = v_3(t)$ for notational ease. We replace the vector $\boldsymbol{v}$ by the 2-dimensional vector $\boldsymbol{u} = (h, w)^{\mathrm{T}}$, and $\boldsymbol{v}_0$ by $\boldsymbol{u}_0 = (h_0, w_0)^{\mathrm{T}}$, etc. We restrict to the 2-dimensional subspace defined by the $\boldsymbol{u}$ variables. We can do this also for the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ models, for which the eigenvector requirement is satisfied, but the h and w dynamics will be completely decoupled for these two cases. We stress that working with the two u_i variables is not equivalent to working with the exact joint marginal distribution of the v_1 and v_3 variables, because by throwing away the other odd variables, we reduce matrices down

in dimension, which in general changes their spectrum. All equations above in the $\boldsymbol{v}$ or $\boldsymbol{v}_0$ variables carry over directly to equations in the $\boldsymbol{u}$ or $\boldsymbol{u}_0$ variables with obvious modifications such as replacing sums and integrals over $\nu - 1$ variables by sums and integrals over only u_1 and u_2. We use these restricted differential or integral equations below to obtain numerical solutions for FPT moments, denoted by $\tau^{(n)}(\boldsymbol{u}_0)$ rather than $\tau^{(n)}(\boldsymbol{v}_0)$, allowing comparison with simulation results or analytical approximations. Because this restriction to $\boldsymbol{u}_0$ does not involve retaining just the slowest eigenmode $x_1 = \widehat{\boldsymbol{\lambda}}^1 \cdot \boldsymbol{v}_0$ as above, we do not expect potential difficulties when $\boldsymbol{u}_0$ is close to the absorbing line $h_0 = \theta$ in this 2-dimensional space.

The matrices $\mathbb{A}$ and $\mathbb{B}$ relevant to the $\boldsymbol{u}$ variables are now 2×2 rather than $(\nu - 1) \times (\nu - 1)$, and for concreteness we write them as

$$\mathbb{A} = \begin{pmatrix} \boldsymbol{e}^h \cdot \mathbb{G}\, \boldsymbol{e}^h & \boldsymbol{e}^h \cdot \mathbb{G}\, \boldsymbol{e}^w \\ \boldsymbol{e}^w \cdot \mathbb{G}\, \boldsymbol{e}^h & \boldsymbol{e}^w \cdot \mathbb{G}\, \boldsymbol{e}^w \end{pmatrix} \begin{pmatrix} \beta_h & 0 \\ 0 & \beta_w \end{pmatrix}, \tag{4.31a}$$

$$\mathbb{B} = -\frac{2}{N}\frac{1}{\nu} \begin{pmatrix} \boldsymbol{e}^h \cdot \mathbb{G}\, \boldsymbol{e}^h & \boldsymbol{e}^h \cdot \mathbb{G}\, \boldsymbol{e}^w \\ \boldsymbol{e}^w \cdot \mathbb{G}\, \boldsymbol{e}^h & \boldsymbol{e}^w \cdot \mathbb{G}\, \boldsymbol{e}^w \end{pmatrix}, \tag{4.31b}$$

and we write

$$\mathbb{D} = \mathrm{diag}\{\beta_h^{-1}, \beta_w^{-1}\}, \tag{4.32}$$

where we define $\boldsymbol{e}^h = \boldsymbol{e}^1$, $\boldsymbol{e}^w = \boldsymbol{e}^3$, $\beta_h = \beta_1$ and $\beta_w = \beta_3$. In appendix B we give explicit results for $\boldsymbol{e}^{u_i} \cdot \mathbb{G}\, \boldsymbol{e}^{u_j}$ and $\beta_{u_i}^{-1}$ for general $\nu \geq 4$ in all four models, as well as the forms for $\mathbb{A}$ and $\mathbb{B}$ for large ν. Of course, $\boldsymbol{e}^w \cdot \mathbb{G}\, \boldsymbol{e}^h = \boldsymbol{e}^h \cdot \mathbb{G}\, \boldsymbol{e}^w$ for symmetric $\mathbb{G}$. In appendix C we list the initial means and covariances of the u_i variables in all four models. For N large enough, we take the distribution of $\boldsymbol{u}_0$ to be a bivariate normal distribution with these means and covariances.

We denote the two normalised left eigenvectors of the 2×2 matrix $\mathbb{A}$ by $\widehat{\boldsymbol{\ell}}^1$

and $\widehat{\boldsymbol{\ell}}^2$. Their corresponding eigenvalues are denoted by $\Lambda_1 - 1$ and $\Lambda_2 - 1$ since the eigenvalues of the full $(\nu-1)\times(\nu-1)$ matrix $\mathbb{A}$ in the $\boldsymbol{v}$ variables are $\lambda_i - 1$. Defining the quantities, $\sigma = \mathbb{A}_{11}+\mathbb{A}_{22}$, $\delta = \mathbb{A}_{11}-\mathbb{A}_{22}$ and $\lambda = +\sqrt{\delta^2 + 4\mathbb{A}_{12}\mathbb{A}_{21}}$, the eigenvalues are $\Lambda_1 - 1 = \frac{1}{2}(\sigma + \lambda)$, $\Lambda_2 - 1 = \frac{1}{2}(\sigma - \lambda)$ with eigenvectors

$$\widehat{\boldsymbol{\ell}}^1 = \frac{1}{\sqrt{\mathbb{A}_{21}^2 + \frac{1}{4}(\lambda-\delta)^2}} \begin{pmatrix} \mp\mathbb{A}_{21} \\ \mp\frac{1}{2}(\lambda-\delta) \end{pmatrix} \approx \begin{pmatrix} 1 \\ +\mathbb{A}_{12}/\delta \end{pmatrix}, \tag{4.33a}$$

$$\widehat{\boldsymbol{\ell}}^2 = \frac{1}{\sqrt{\mathbb{A}_{12}^2 + \frac{1}{4}(\lambda-\delta)^2}} \begin{pmatrix} \pm\frac{1}{2}(\lambda-\delta) \\ \mp\mathbb{A}_{12} \end{pmatrix} \approx \begin{pmatrix} -\mathbb{A}_{21}/\delta \\ 1 \end{pmatrix}, \tag{4.33b}$$

where the upper signs are chosen for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model and the lower for the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model, and the approximations are obtained by writing $\lambda = +\sqrt{\delta^2 + 4\mathbb{A}_{12}\mathbb{A}_{21}} \approx \delta + 2\mathbb{A}_{12}\mathbb{A}_{21}/\delta$ and treating $1/\delta$ as small. Eq. (4.13) for the $\boldsymbol{v}_0$ variables becomes the equation

$$-\frac{1}{r} n\, \tau^{(n-1)}(\boldsymbol{x}) = \sum_{i=1}^{2} (1-\Lambda_i) \left\{ -x_i \frac{\partial}{\partial x_i} + \frac{1}{N}\frac{1}{\nu} \big(\widehat{\boldsymbol{\ell}}^i \cdot \mathbb{D}\, \widehat{\boldsymbol{\ell}}^i\big) \frac{\partial^2}{\partial x_i^2} \right\} \tau^{(n)}(\boldsymbol{x}), \tag{4.34}$$

where now $\boldsymbol{x} = \mathbb{T}\boldsymbol{u}_0$ and $\mathbb{T}$ is the 2×2 matrix whose first and second rows contain the normalised eigenvectors $\widehat{\boldsymbol{\ell}}^1$ and $\widehat{\boldsymbol{\ell}}^2$, respectively. We have chosen conventions for the eigenvectors in Eq. (4.33) so that x_1 is the variable "closest" to h_0, with x_1 increasing when $u_1 \equiv h_0$ increases; and $\Lambda_1 - 1 > \Lambda_2 - 1$, so that $\Lambda_1 - 1$ is the least negative eigenvalue of $\mathbb{A}$. Since the eigenvalues of the 2×2 form for $\mathbb{A}$ are well separated, we can approximate the FPT dynamics governed by Eq. (4.34) by again considering just x_1, so that

$$-\frac{1}{r}\frac{1}{(1-\Lambda_1)} n\, \tau^{(n-1)}(x_1) = \left\{ -x_1 \frac{d}{dx_1} + \frac{1}{N}\frac{1}{\nu} \big(\widehat{\boldsymbol{\ell}}^1 \cdot \mathbb{D}\, \widehat{\boldsymbol{\ell}}^1\big) \frac{d^2}{dx_1^2} \right\} \tau^{(n)}(x_1), \tag{4.35}$$

and the kernel $\overline{K}_1(x|x')$ in the integral equation in Eq. (4.23) becomes

$$\overline{K}_1(x|x') = \sqrt{\frac{N\nu}{2\pi\big[(1-\Lambda_1^2)\,\widehat{\boldsymbol{\ell}}^1\cdot\mathbb{D}\,\widehat{\boldsymbol{\ell}}^1\big]}}\,\exp\left\{-\frac{N\nu\big(x-\Lambda_1 x'\big)^2}{2\big[(1-\Lambda_1^2)\,\widehat{\boldsymbol{\ell}}^1\cdot\mathbb{D}\,\widehat{\boldsymbol{\ell}}^1\big]}\right\}. \quad (4.36)$$

Just as we used the dominant x_1 eigenmode in the full $\boldsymbol{v}_0$ space above to work with a 1-dimensional approximation, we can also use the dominant eigenmode in the restricted $\boldsymbol{u}_0$ space. The MFPT solutions in Eqs. (4.21) and (4.25) for the $\boldsymbol{v}$ variables carry over directly to the $\boldsymbol{u}$ variables with obvious modifications.

If the pair of differential equations in Eqs. (4.19) and (4.35) and the pair of 1-dimensional kernels in Eqs. (4.24) and (4.36) in the integral equation in Eq. (4.23) are to generate similar results for FPT moments, then we require $1-\Lambda_1$ to be numerically very similar to $1-\lambda_1$ and similarly for the pair $\widehat{\boldsymbol{\ell}}^1\cdot\mathbb{D}\,\widehat{\boldsymbol{\ell}}^1$ and $\widehat{\boldsymbol{\lambda}}^1\cdot\mathbb{D}\,\widehat{\boldsymbol{\lambda}}^1$. We have from above that $\widehat{\boldsymbol{\lambda}}^1\cdot\mathbb{D}\,\widehat{\boldsymbol{\lambda}}^1 = 1/\left|\boldsymbol{\lambda}^1\right|^2$, and for the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model we obtain $\left|\boldsymbol{\lambda}^1\right|^2 = 2\big(\nu^2+\nu-3\big)/\big[\nu^2\big(\nu+1\big)\big]$ while for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model we cannot obtain such a simple result. The results for the 2-dimensional version $\widehat{\boldsymbol{\ell}}^1\cdot\mathbb{D}\,\widehat{\boldsymbol{\ell}}^1$ are trivial but messy, but if we employ the approximation given in Eq. (4.33), then we obtain $\widehat{\boldsymbol{\ell}}^1\cdot\mathbb{D}\,\widehat{\boldsymbol{\ell}}^1 \approx \beta_h^{-1}$ for all forms of model. We tabulate some explicit numerical values for these quantities for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ models in Table 1 to permit a direct comparison. For $\nu = 2$ to 5, all corresponding pairs of quantities must coincide exactly because for $\nu = 2$ and $\nu = 3$ the only odd variable is v_1 and we discard v_3 from $\boldsymbol{u}$; and for $\nu = 4$ and $\nu = 5$ the only odd variables are v_1 and v_3, which are precisely u_1 and u_2. Differences only emerge for $\nu \geq 6$, but we see that these amount to a few percent at most for the eigenvalues and are very small for the other pair, and in fact vanishing as ν increases for the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model. Even the approximate forms of the $\widehat{\boldsymbol{\ell}}^i$ eigenvectors in Eq. (4.33), producing the result $\frac{1}{\nu}\beta_h^{-1}$, give very good ($\mathbb{C}\&\boldsymbol{\mathcal{L}}$) or excellent ($\mathbb{Q}\&\boldsymbol{\mathcal{S}}$) agreement. In terms of this approximation

scheme, then, the reduction to the $\boldsymbol{u}$ variables incurs only marginal errors compared to using the whole $\boldsymbol{v}$ space, while the use of the $\boldsymbol{u}$ variables affords a considerable advantage in the calculation of $\langle \tau_{\mathrm{mfpt}}(\boldsymbol{u}_0) \rangle_{\boldsymbol{u}_0|h_0>\theta}$, as we only need to average over a 2-dimensional distribution.

5 Results: Analysis and Simulation

We now turn to a comparison between our analytical or numerical results discussed above and results obtained by simulation. Our simulation protocols have been discussed extensively elsewhere (Elliott & Lagogiannis, 2012; Elliott, 2014). In brief, we run two different forms of simulation, depending on whether we are interested in obtaining $\langle \tau_{\mathrm{mfpt}}(\boldsymbol{v}_0) \rangle_{\boldsymbol{v}_0|h_0>\theta}$, an average of MFPTs over all supra-threshold configurations immediately after the storage of the tracked memory, or $\tau_{\mathrm{mfpt}}(\boldsymbol{v}_0)$, the MFPT for a particular, definite such configuration. For the former, a simulation corresponds to storing a randomly-generated tracked memory $\boldsymbol{\xi}^0$ by updating N synapses whose initial distribution corresponds to the equilibrium distribution. A simulation is then run by storing further, random memories at Poisson-generated times. The simulation is terminated when the perceptron's activation in response to the tracked memory first falls (to or) below threshold. FPT statistics are obtained by averaging over many such simulations, which averages over both the tracked memory and the later memories. Typically we use 10^5 runs for most choices of N but 10^4 runs for $N = 10^5$ for efficiency. For the latter, we must either prepare the simulations in definite, given initial states (particular choices of $\boldsymbol{v}_0$ or $\boldsymbol{u}_0$), or generate initial states as in the former case but record and poll separate statistics for each distinct initial configuration. The second possibility is feasible for small N for which the number of distinct initial configurations is

		$\nu = 2$	$\nu = 3$	$\nu = 4$	$\nu = 5$	$\nu = 6$	$\nu = 10$	$\nu = 20$	$\nu = 40$	$\nu \to \infty$
$\mathbb{C}\&\boldsymbol{\mathcal{L}}$	$\frac{1}{\nu}\widehat{\boldsymbol{\ell}}^1 \cdot \mathbb{D}\widehat{\boldsymbol{\ell}}^1$	1.000000	0.666667	0.558048	0.503043	0.469489	0.408230	0.366418	0.346637	0.327542
	$\frac{1}{\nu}\widehat{\boldsymbol{\lambda}}^1 \cdot \mathbb{D}\widehat{\boldsymbol{\lambda}}^1$	1.000000	0.666667	0.558048	0.503043	0.469505	0.408255	0.366311	0.346367	0.327019*
	$\frac{1}{\nu}\beta_h^{-1}$	1.000000	0.666667	0.555556	0.500000	0.466667	0.407407	0.368421	0.350427	0.333333
	$(1-\Lambda_1)/p$	1.000000	0.500000	0.292893	0.190983	0.133978	0.048951	0.012316	0.003084	$4.937548/\nu^2$
	$(1-\lambda_1)/p$	1.000000	0.500000	0.292893	0.190983	0.133975	0.048944	0.012312	0.003083	$4.934802/\nu^2$
$\mathbb{Q}\&\boldsymbol{\mathcal{S}}$	$\frac{1}{\nu}\widehat{\boldsymbol{\ell}}^1 \cdot \mathbb{D}\widehat{\boldsymbol{\ell}}^1$	1.000000	0.666667	0.588235	0.555556	0.537975	0.513281	0.503276	0.500816	0.500000
	$\frac{1}{\nu}\widehat{\boldsymbol{\lambda}}^1 \cdot \mathbb{D}\widehat{\boldsymbol{\lambda}}^1$	1.000000	0.666667	0.588235	0.555556	0.538462	0.514019	0.503597	0.500916	0.500000
	$\frac{1}{\nu}\beta_h^{-1}$	1.000000	0.666667	0.585786	0.552786	0.535898	0.512543	0.503097	0.500772	0.500000
	$(1-\Lambda_1)/p$	1.000000	0.500000	0.333333	0.250000	0.200954	0.112929	0.053817	0.026261	$1.024725/\nu$
	$(1-\lambda_1)/p$	1.000000	0.500000	0.333333	0.250000	0.200000	0.111111	0.052631	0.025641	$1.000000/\nu$

Table 1. Explicit numerical comparison between different approximations. For the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ models, we show explicit numerical values for the quantities indicated, for different choices of ν, and in the limit $\nu \to \infty$. Even when entries have very simple expressions, we have given numerical values to six decimal places to facilitate easy comparison. We cannot obtain the entry marked with an asterisk as a closed-form analytical expression, but it can be computed via a complicated converging series; we have obtained its numerical value to at least 20 decimal places.

relatively small; the first possibility is essentially mandatory for larger N, for which there are too many configurations (becoming essentially continuous as N increases) for data accumulation over different configurations. Simulations are again then run by storing later random, Poisson memories and terminating when threshold is crossed. In the first possibility, averages are then taken across simulations, which averages over later memories for a definite choice of $\boldsymbol{v}_0$; in the second possibility, averages are taken only over identical initial configurations, so in this case employing multiple simulations both generates (samples over) different initial configurations $\boldsymbol{v}_0$ and averages over later memories for the same initial configurations. For small N and this second possibility, it is necessary to run over 2.5×10^7 simulations to obtain good statistics over a large sample of possible configurations; this is feasible on a desktop computer. For larger N and the first possibility, for each selection of values for the pair h_0 and w_0 corresponding to a choice of $\boldsymbol{u}_0$ we average over 2000 simulations, but must also run over multiple selections of these two initial variables. We find it highly convenient to run these many separate simulations on our central cluster computational facility to accelerate data collection. We also use this facility to solve numerically the integral equations for FPTs, since high memory nodes are indispensable for such computations.

We begin by examining the evolution of the unconditional marginal and joint probability distributions for $h(t)$ and $w(t)$ in the absence of an absorbing boundary. Fig. 2 shows results for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model while Fig. 3 shows the $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ model. We illustrate the marginal distributions by plotting the evolution of the means $\mathsf{E}[h(t)]$ and $\mathsf{E}[w(t)]$ and the one standard deviation regions around them. Results are shown using the exact, full $\boldsymbol{v}(t)$ variables, their OU limit versions, and the restricted $\boldsymbol{u}(t)$ variables. Differences between the exact and their OU forms can appear only in the standard deviations, as their means

are necessarily identical. However, the restricted $\boldsymbol{u}(t)$ variables can differ from the $\boldsymbol{v}(t)$ variables at the level of the means because of the reduction to two dimensions. For the $h(t)$ marginal, all three forms are virtually indistinguishable, with only tiny differences between the exact and OU $\boldsymbol{v}(t)$ forms being discernible at intermediate times for the $\nu = 6$ case. The difference between the full and OU form is somewhat clearer for $w(t)$ for $\nu = 6$ in the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model, but the clearer difference in all the other shown cases is between the $\boldsymbol{v}(t)$ and $\boldsymbol{u}(t)$ forms at intermediate times. This difference is principally due to the differences in the means, as the shift in the one standard deviation regions closely follows the means. Nevertheless, the $\boldsymbol{u}(t)$ variables follow very closely the $\boldsymbol{v}(t)$ variables, and at larger times they become indistinguishable. We also illustrate the evolution of the joint distribution $P(\boldsymbol{u}; t)$ of $h(t)$ and $w(t)$ for the $\boldsymbol{u}(t)$ variables only. We do this by plotting ellipses corresponding to the one standard deviation region (in the sense of Mahalanobis) around their means. For the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model, we see that $w(t)$ moves rapidly towards zero, with $h(t)$ decaying rather more slowly. In the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model, the decays of both $h(t)$ and $w(t)$ are governed by the same, single eigenmode in the $\boldsymbol{v}(t)$ variables, so the sequence of elliptical regions in Fig. 3 takes on the appearance of a sheered cylinder, whereas for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model in Fig. 2 the sequence has a distinct "bend" due the rapid decay of $w(t)$ compared to $h(t)$. The slight distortion in the sheered cylinder for the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model for $\nu = 16$ at earlier times is due to the use of the $\boldsymbol{u}(t)$ variables; with the $\boldsymbol{v}(t)$ variables, the sides of the cylinder would be perfectly straight. The results in these two figures essentially demonstrate the validity of both the OU approximation to the full $\boldsymbol{v}(t)$ variables and their replacement by the restricted $\boldsymbol{u}(t)$ variables. We do not show similar results for the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ models because the $w(t)$ dynamics are irrelevant in these cases, due to the decoupling when the eigenvector condition

is satisfied: only $h(t)$ is relevant, and we have previously shown that the OU approximation in these two cases is very good.

FIG. 2 ABOUT HERE.

FIG. 3 ABOUT HERE.

Turning to MFPTs, we illustrate the differences between all four models in Fig. 4 by plotting MFPTs against "synaptic configuration number" for $N = 25$, for which MFPTs may be computed exactly (albeit numerically) using discrete synaptic configurations and Eq. (3.29a). We do this for $\nu = 4$ and use all three $\boldsymbol{v}$ variables in this case, with $\mathsf{Prob}[\boldsymbol{v}|\boldsymbol{v}_0]$ being determined exactly from the conditional PGF in Eq. (3.5). For all four models, for $N = 25$ and $\nu = 4$ there are ${}^{28}C_3 = 3276$ distinct synaptic configurations, with for $\theta = 0$ exactly half of them having perceptron activation greater than threshold. We must therefore solve the coupled system of linear equations implied by Eq. (3.29a) for these 1638 different MFPTs for each supra-threshold configuration. For $\nu = 4$ we enumerate configurations using the triple (v_1, v_3, v_2), so with the two odd variable ($h \equiv v_1$ and $w \equiv v_3$) first and the single even variable last. In particular, we enumerate these configurations in increasing order of these variables' values, with v_1 taking precedence over v_3 and v_3 over v_2 (i.e. they are ordered first with respect to v_1, then v_3, and then v_2). The "synaptic configuration number" is simply a triple's position in this ordered enumeration of configurations. Fig. 4 plots $\tau_{\mathrm{mfpt}}(\boldsymbol{v}_0)$ against the configuration number of $\boldsymbol{v}_0$. We plot in this way to highlight the primary dependence of $\tau_{\mathrm{mfpt}}(\boldsymbol{v}_0)$ on h_0 and its secondary dependence on w_0. In Fig. 4A for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model, each distinct cluster of data points corresponds to the same value of h_0. As h_0 increases, $\tau_{\mathrm{mfpt}}(\boldsymbol{v}_0)$ increases. Within each cluster, we see a secondary dependence on w_0: as w_0 increases for fixed h_0, $\tau_{\mathrm{mfpt}}(\boldsymbol{v}_0)$ decreases. There is a further, much milder

dependence on the even variable, which is noticeable here because $N = 25$ is well away from N being large enough for the $\mathcal{O}(1/N)$ arguments above to be valid. Nevertheless, we see that the odd variables dominate. Fig. 4B shows results for the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model. The structure visible in Fig. 4A is less clear in Fig. 4B. This is because of the difference between the $\boldsymbol{v}$ variables for $\boldsymbol{e}^m = \boldsymbol{s}^m$ (sigmoidal strengths) rather than $\boldsymbol{e}^m = \boldsymbol{l}^m$ (uniform strengths). For sigmoidal strengths, $\boldsymbol{e}^1 = (-1, +1-\sqrt{2}, -1+\sqrt{2}, +1)^{\mathrm{T}}$, $\boldsymbol{e}^3 = (-1, +1+\sqrt{2}, -1-\sqrt{2}, +1)^{\mathrm{T}}$ and $\boldsymbol{e}^2 = (+1, -1, -1, +1)$. Because of the presence of $\sqrt{2}$ in $\boldsymbol{e}^1$ and $\boldsymbol{e}^3$, the value of h_0 uniquely determines the value of w_0 and *vice versa*, whereas for uniform strengths, this is not the case. Distinct, clear clusters like those in Fig. 4A are therefore much less clear in Fig. 4B. Visually, there are aerofoil-shaped grouping extending horizontally, but the pertinent structures are the near-vertical bands. To make these bands clearer, we bin the values of the variables into bins of width 0.02 and show the resulting data on the same panel but displaced vertically. This vertical banding indicates that for the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model, in configurations with very similar values of h_0, $\tau_{\mathrm{mfpt}}(\boldsymbol{v}_0)$ increases as w_0 increases. This does not contradict the unique correspondence between h_0 and w_0 for this $\nu = 4$ case. For example, a tiny change in h_0 can induce a large change, either upwards or downwards, in the corresponding but unique value of w_0. The MFPT $\tau_{\mathrm{mfpt}}(\boldsymbol{v}_0)$ will then change not because of the tiny change in h_0 but because of the large change in w_0. Specifically, when the values of h_0 are binned so that each bin contains a small range of values of h_0, the range of w_0 within these h_0 bins is large, and $\tau_{\mathrm{mfpt}}(\boldsymbol{v}_0)$ within these bins increases as w_0 increases. Looking more carefully at the unbinned data in Fig. 4B, each very small group of points corresponds to h_0 and w_0 being fixed, and the small variation in $\tau_{\mathrm{mfpt}}(\boldsymbol{v}_0)$ is due to the remaining dependence on the even variable. But as h_0 and w_0 change between these very small groups, $\tau_{\mathrm{mfpt}}(\boldsymbol{v}_0)$ can jump

significantly even when h_0 changes by a very small amount because w_0 can change by a large amount. For the $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ model in Fig. 4C, we again see the distinct, large clusters visible in Fig. 4A, but the variation of $\tau_{\text{mfpt}}(\boldsymbol{v}_0)$ within these clusters is very small. This reflects the decoupling of variables when the eigenvector condition is satisfied, with the remaining variation being visible because N is not large here. To a very good first approximation, each cluster can be regarded as having the same value of $\tau_{\text{mfpt}}(\boldsymbol{v}_0)$; as N increases, the variation within clusters decreases. We observe similar behaviour for the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ model in Fig. 4D, which also satisfies the eigenvector condition. The clusters are again much less distinct because of the $\boldsymbol{e}^m = \boldsymbol{s}^m$ issues discussed above, but essentially the relationship between $\tau_{\text{mfpt}}(\boldsymbol{v}_0)$ and configuration number is unique and monotonic, again reflecting the decoupling of variables. Because the behaviour of the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ models is dominated by this decoupling due to the eigenvector condition being satisfied, we will focus below on the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ models, where the move from h to the $\boldsymbol{v}$ (or the restricted $\boldsymbol{u}$) variables is mandatory for understanding MFPTs.

FIG. 4 ABOUT HERE.

For this $N = 25$ and $\nu = 4$ case, we plot MFPTs as functions of h_0 and w_0 for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ models in Fig. 5, and also on the same plots show results from simulations. In addition, we show results from the opposite extreme, $N = 10^5$ rather than $N = 25$, also with $\theta = 0$. For this large N case we restrict to the $\boldsymbol{u}$ variables, use the continuum limit form in Eq. (4.14) (for $n = 1$), and the normal kernel in Eq. (4.15). Numerical methods are used to solve the 2-dimensional integral equation and obtain $\tau_{\text{mfpt}}(\boldsymbol{u}_0)$. Simulation data for $N = 25$ are obtained by the polling method described above. Very rare (very low probability) configurations therefore do not generate enough samples over

which averages can be taken, so simulation data points are shown only when at least 500 samples are available to generate averages. For $N = 10^5$, simulation data are obtained by preparing configurations with definite h_0 and w_0 values. The initial value of the even variable can either be set to its mean, zero value, or be drawn from its marginal distribution; the results are indistinguishable. This figure essentially serves to validate both our analytical/numerical methods and our simulation methods by showing their essential agreement, up to the inevitable simulation noise.

FIG. 5 ABOUT HERE.

For the $N = 10^5$ case in Fig. 5, the extremely strong dependence of $\tau_{\mathrm{mfpt}}(\boldsymbol{u}_0)$ on w_0 for small h_0 is particularly striking. For the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model, $\tau_{\mathrm{mfpt}}(\boldsymbol{u}_0)$ is large for $w_0 < 0$ but drops rapidly to very close to zero for $w_0 > 0$. This dependence becomes stronger and the transition sharper as ν increases. The behaviour in the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model is opposite that of the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model. Numerically it is very hard to obtain the resolution to take h_0 extremely small, but it is easy to run simulations for very small h_0. For larger h_0, there is still the same dependence on w_0, but the transition weakens and turns into a slight tilting of the landscape. This behaviour is present for $N = 25$, where we can see the dependence of $\tau_{\mathrm{mfpt}}(\boldsymbol{v}_0)$ on w_0 for fixed or nearly fixed h_0 discussed in Fig. 4, but the dependence is far less marked than for large N. We have chosen $N = 10^5$ to illustrate this very strong w_0-dependence for small h_0. In the continuum limit, such dependence is always present by the scaling arguments above with $\boldsymbol{u}_0^* = \sqrt{N}\,\boldsymbol{u}_0$, but as N decreases it is necessary to focus on smaller and smaller values of h_0 to see this strong dependence with its rapid transition around $w_0 = 0$. Of course, for very small N the continuum limit ceases to be valid and the scaling arguments break down.

To explain this strong w_0-dependence for small h_0, we plot flows representing the joint evolution of the conditional expectations $\mathsf{E}[h(t)|\boldsymbol{u}_0]$ and $\mathsf{E}[w(t)|\boldsymbol{u}_0]$ in Fig. 6. The flows represent the Liouville or drift-only dynamics. We have used the formal, $\nu \to \infty$ limit and so plotted flows generated by the matrix $\mathbb{A}^*$. The directions associated with the eigenvectors $\widehat{\boldsymbol{\ell}}^1$ and $\widehat{\boldsymbol{\ell}}^2$ are also shown. At time $t = 0$ s, $\mathsf{E}[\boldsymbol{u}(t)|\boldsymbol{u}_0] \equiv \boldsymbol{u}_0$, so we can refer to the axes at $t = 0$ s as the w_0 and h_0 axes. Flows start on the circle $|\boldsymbol{u}_0| = 1$ at $t = 0$ s and move towards the origin as $t \to \infty$, but the flows are trapped in the quadrants defined by the eigenvectors and so are rotated with respect to the initial axes. In the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model (Fig. 6A) we see that for initial $w_0 > 0$, any initial state between the $\widehat{\boldsymbol{\ell}}^2$ (right-pointing) eigenvector and the $h_0 = 0$ axis gets pushed by the drift across this axis. In contrast, for $w_0 < 0$ the drift pushes the initial state away from the $h_0 = 0$ axis for $h_0 > 0$. In general, for $h_0 > \theta$, with $w_0 > 0$ the drift pushes the system down towards and through threshold, while with $w_0 < 0$ the drift pushes the system away from threshold. As drift dominates diffusion especially for larger choices of N, these dynamics mean that in the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model, for small h_0 and $w_0 > 0$ MFPTs are very small, while for small h_0 and $w_0 < 0$ MFPTs are much larger. Because of the different orientation of the eigenvectors in the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model in Fig. 6B, the opposite dynamics occurs, with $w_0 < 0$ giving small MFPTs for small h_0 and $w_0 > 0$ giving large MFPTs. In the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ models, the two eigenvectors coincide precisely with the two axes, so w_0 does not influence MFPTs at all.

FIG. 6 ABOUT HERE.

We turn now to a comparison between MFPTs obtained from the Fokker-Planck approach in Eq. (4.10) and the integral equation in Eq. (4.14) for $n = 1$, restricted to the $\boldsymbol{u}$ variables. Solutions in both cases are obtained numerically.

Eq. (4.10) is solved by standard numerical methods using an implicit scheme to maximise stability. The absorbing boundary condition at $h = \theta$ (we set $\theta = 0$) is imposed, and additionally we must impose further boundary conditions to obtain numerical solutions. Thus, we solve Eq. (4.10) in a rectangular region in the w_0–h_0 plane with boundary conditions on all sides of the rectangle. However, the solution is largely insensitive to the three artificial boundaries and provided that we are sufficiently far (which is not in fact very far at all) from them, the interior solution is independent of them. We show results using the scaled $\boldsymbol{u}_0^*$ variables to remove the dependence on N. We have checked explicitly that the numerical solutions, up to grid discretisation, do indeed respect this scaling with N by solving for different choices of N and then superimposing the scaled solutions. Rather than showing 3-dimensional plots as in Fig. 5, we now show contour plots of $\tau_{\mathrm{mfpt}}(\boldsymbol{u}_0)$ in the w_0–h_0 plane for greater clarity and less clutter. Fig. 7 shows such a comparison. In fact we show contours of $\frac{pr}{\nu^\kappa}\tau_{\mathrm{mfpt}}(\boldsymbol{u}_0^*)$, which removes the dependence on the choice of p and r, and enables easier comparison between different choices of ν and between the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ models. Results are shown for $\nu = 4$ and for the formal, $\nu \to \infty$ limit. For the $\nu = 4$ case in Figs. 7A and 7B, we see in general good qualitative agreement between the integral equation results (in principle exact, barring issues of numerical solutions, the continuum limit, etc.) and those obtained from the Fokker-Planck approach, with its diffusion limit. For smaller h_0^*, so closer to the threshold $\theta = 0$, we see greater deviations between the two sets of results, reflecting precisely the breaking down of the diffusion approximation in the vicinity of the threshold. In the formal, $\nu \to \infty$ limit in Figs. 7C and 7D, we see excellent agreement between these two sets of results, but again with the differences at small h_0^* due to the diffusion limit approximation in the Fokker-Planck approach. For these results in Figs. 7C and 7D, we note that it is

extremely hard to obtain stable numerical solutions to the integral equation in Eq. (4.14) with large h_0^*, although it is easier to obtain solutions to Eq. (4.10); we have found it impossible, using our numerical methods, to push the integral equation solutions out beyond around $h_0^* \approx 5$–6. Equally, it is very hard to push the Fokker-Planck solution to very small h_0, because doing so requires too fine a mesh, which becomes computationally intractable. Comparing Figs. 7A and 7C for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model with $\nu = 4$ and large ν, respectively, we notice the difference in the sharpness of the transition between $w_0^* < 0$ and $w_0^* > 0$ for smaller h_0^*. For large ν, the tightly packed contours indicate a very sharp transition from large MFPTs to small MFPTs, caused by the drift dynamics discussed above. For the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model in Figs. 7B and 7D, the transition for large ν is sharper than for small ν (note the difference in scales on the w_0^* axes), but not so sharp as for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model.

FIG. 7 ABOUT HERE.

Calculation of MFPTs using 2-dimensional integral equations or partial differential equations is time-consuming and subject to possible numerical instabilities. The approximation developed in Eq. (4.21) (in the $\boldsymbol{u}$ variables) provides a far simpler method for obtaining results in the Fokker-Planck approach. We examine this approximation in Fig. 8 for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model and in Fig. 9 for the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model. We plot cross-sections through the $\tau_{\mathrm{mfpt}}(\boldsymbol{u}_0)$ landscape in the w_0–h_0 plane for various choices of h_0 (multiples of $\mathsf{E}[h_0]$), comparing the approximation in Eq. (4.35) for $n = 1$ with results obtained from the 2-dimensional Fokker-Planck approach, for $\theta = 0$. For $N = 10^4$ and $N = 10^5$, we see in general extremely good agreement in all cases, with differences emerging between the approximation and the numerical, 2-dimensional results only for smaller values of h_0. These smaller values of h_0 are closer

to the threshold $\theta = 0$, and it is precisely in the vicinity of the threshold that we would expect the approximation to be less good. We note, however, that in terms of the absolute rather than relative differences, the approximation is worse for negative w_0 and better for positive w_0, for smaller N. Since $\mathsf{E}[w_0] > 0$, we might therefore expect these differences to be less important for the calculation of $\langle \tau_{\mathrm{mfpt}}(\boldsymbol{u}_0) \rangle_{\boldsymbol{u}_0|h_0>\theta}$. Furthermore, the smaller h_0 lines shown on these graphs correspond to h_0 being at least half $\mathsf{E}[h_0]$ and mostly smaller, so again the errors in the small h_0 regime would not be expected to influence $\langle \tau_{\mathrm{mfpt}}(\boldsymbol{u}_0) \rangle_{\boldsymbol{u}_0|h_0>\theta}$ too much.

FIG. 8 ABOUT HERE.

FIG. 9 ABOUT HERE.

We show explicit results for $\langle \tau_{\mathrm{mfpt}}(\boldsymbol{u}_0) \rangle_{\boldsymbol{u}_0|h_0>\theta}$ with $\theta = 0$ in Fig. 10, plotting $\langle \tau_{\mathrm{mfpt}}(\boldsymbol{u}_0) \rangle_{\boldsymbol{u}_0|h_0>\theta}$ as a function of N for fixed choices of ν (Figs. 10A and 10B) and as a function of ν for fixed choices of N (Figs. 10C and 10D). We show results for the approximation in Eq. (4.21) and its integral equation equivalent in Eq. (4.25) (both in the $\boldsymbol{u}$ variables), and also data obtained from simulations. We see extremely good agreement between simulation data and the integral equation results, and also very good agreement between the integral equation and Fokker-Planck approach, although with the expected deviations in the latter due to the diffusion approximation. The analytical results from the integral equation are mostly indistinguishable from simulation data except in the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model for very small N or larger choices of ν. In contrast, in the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model we see essentially perfect agreement regardless of N or ν. This difference between the two models is due to the parameter dependence of $\mathsf{E}[h_0]$ in both cases. In the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model $\mathsf{E}[h_0] = 2p/\nu$, while in the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model

$\mathsf{E}[h_0] = p\,\mathrm{cosec}^2\frac{\pi}{2\nu}/\left[\nu(\nu-1)\right] \approx 4p/\pi^2 \approx 0.41p$ for large enough ν. For the former, $\mathsf{E}[h_0]$ falls as ν increases, while for the latter, it asymptotes to a constant. The initial distribution in the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model is therefore closer to the threshold θ than that of the $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ model. Further, for smaller N, the initial distribution is broader, while for larger N, it is more tightly focused, so for smaller N more of the distribution will be near or below threshold. Overall, then, results for the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ model will be more subject to the approximation method's sensitivity to the threshold than those for the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model, with smaller N or larger ν revealing this sensitivity, precisely as observed. Nevertheless, given that this approximation method is based on working only with the slowest eigenmode, the agreement between the approximation and simulation is extremely satisfying.

FIG. 10 ABOUT HERE.

Fig. 10 should be compared to Fig. 7 in Elliott (2017b). Analytical results in that figure were obtained only for the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ models, both of which satisfy the eigenvector condition. There, we were able to show only simulation results for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ models as we could not analyse them, although we had argued on the basis of closeness of strength vectors that results for the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ and $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ models should be (and indeed are) very similar, and also for the $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ models. The qualitative features of Fig. 10 are very similar to that earlier figure, whose features were discussed very extensively. We therefore simply note here that while $\tau_{\mathrm{mfpt}}(\boldsymbol{u}_0^*)$ with $\boldsymbol{u}_0^* = \sqrt{N}\,\boldsymbol{u}_0$ is independent of N by the scaling arguments above (for a threshold $\theta = 0$), $\langle\tau_{\mathrm{mfpt}}(\boldsymbol{u}_0)\rangle_{\boldsymbol{u}_0|h_0>\theta}$ does not scale simply with N, for two reasons. First, of course the average over the initial distribution explicitly depends on N, even in the absence of the threshold. Second, the presence of the threshold truncates

the distribution in an N-dependent manner: for large N, there is hardly any effect (if the distribution is almost entirely above threshold), but for small N, a large fraction of the distribution may be below threshold. Furthermore, although $\tau_{\text{mfpt}}(\boldsymbol{u}_0)$ exhibits scaling with ν for large enough ν, $\langle \tau_{\text{mfpt}}(\boldsymbol{u}_0) \rangle_{\boldsymbol{u}_0|h_0>\theta}$ in general does not, because the initial distribution depends explicitly on ν. We see this non-scaling in Fig. 10C for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model, where there are very clear non-linearities in the log-log plot. In contrast, in the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model in Fig. 10D, we see equally clearly the onset of linear-like or nearly linear behaviour in the log-log plot. This behaviour owes its origin to the fact that $\mathsf{E}[h_0]$ loses its ν-dependence for large enough ν, as discussed above. This behaviour however is a very particular property arising from models based on the $\mathbb{M}_{\text{Q}}$ transition matrix.

Finally, in Fig. 11 we show results for $\langle \tau_{\text{mfpt}}(\boldsymbol{u}_0) \rangle_{\boldsymbol{u}_0|h_0>\theta}$ and the one standard deviation region around it, denoted schematically as $\langle \sigma_{\text{fpt}}(\boldsymbol{u}_0) \rangle_{\boldsymbol{u}_0|h_0>\theta}$, where we write $\langle \sigma_{\text{fpt}}(\boldsymbol{u}_0) \rangle^2_{\boldsymbol{u}_0|h_0>\theta} = \langle \tau^{(2)}(\boldsymbol{u}_0) \rangle_{\boldsymbol{u}_0|h_0>\theta} - \langle \tau_{\text{mfpt}}(\boldsymbol{u}_0) \rangle^2_{\boldsymbol{u}_0|h_0>\theta}$, which involves the second moment of the FPT distribution. We again show results using the approximation method in section 4.3, with the second moment obtained from either Eq. (4.19) or Eq. (4.23) for $n = 2$, both restricted to the $\boldsymbol{u}$ variables. Where necessary we also show results obtained from the 2-dimensional integral equation, so not based on the approximation method in section 4.3. We show results as a function of N for the particular but representative choice $\nu = 10$. We also consider varying the threshold θ. For $\theta = 0.00$ in Figs. 11A and 11B, the means are the same as shown in Fig. 10, but we again see extremely good agreement between the approximate analytical and simulation results for the one standard deviation regions around the means; in Fig. 11B for the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model in particular, the agreement between the 1-dimensional integral equation approximation and simulation is essentially perfect. Similarly, for $\theta = 0.01$ in

Figs. 11C and 11D, the agreement is extremely good, especially so again for the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model. This remains true for $\theta = 0.02$ in the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ too, in Fig. 11F, but for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model in Fig. 11E, for larger N we start to see the clear breaking down of the approximation method as $\langle \tau_{\mathrm{mfpt}}(\boldsymbol{u}_0) \rangle_{\boldsymbol{u}_0|h_0>\theta}$ becomes small. In this region we also show the results from the 2-dimensional integral equation, confirming their very close agreement with simulation results. The breakdown in the approximation scheme in this regime can be easily understood. For $\nu = 10$ and $p = 0.1$ in the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model, $\mathsf{E}[h_0] = 0.02$. But in Fig. 11E the threshold is precisely $\theta = 0.02$. So as N increases, almost all the distribution concentrates around the threshold. Hence, as N increases, we move to a regime in which the approximation method, which relies on being sufficiently far away from threshold initially, breaks down.

FIG. 11 ABOUT HERE.

We have discussed extensively the qualitative features of the behaviour seen in Fig. 11 in relation to the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ models, satisfying the eigenvector requirement, elsewhere (Elliott, 2017b). Because of the similarity in results for models with the same underlying transition matrix, this earlier discussion carries over to the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ models studied here, which we were not previously able to analyse. In brief, for $\theta = 0$ with potentiation and depression processes treated symmetrically, $\langle \tau_{\mathrm{mfpt}}(\boldsymbol{u}_0) \rangle_{\boldsymbol{u}_0|h_0>0}$ asymptotes for large N to logarithmic growth with N. For $\theta > 0$, however, this growth is truncated, and $\langle \tau_{\mathrm{mfpt}}(\boldsymbol{u}_0) \rangle_{\boldsymbol{u}_0|h_0>0}$ asymptotes to a non-zero constant as N increases. This behaviour for $\theta > 0$ is clear in the figure, except in Fig. 11E, where we would need to extend to larger N to see the onset of the asymptotic behaviour. For any θ, zero or greater, the variance in the FPT distribution asymptotes to a non-zero constant, at least in the OU approximation. For smaller N and

small θ, the variance swamps the mean in the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ (and $\mathbb{C}\&\boldsymbol{\mathcal{S}}$) model, so that although the mean in robustly positive, the storage of the tracked memory is subject to large variability; as N increases, the variability reduces. For larger θ, however, even for larger N the variance swamps the mean. In contrast, in the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ (and $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$) model, although the smaller N behaviour of the variance is inevitably present, we see in all cases in Fig. 11 that for larger N (in this figure, $N \gtrapprox 2000$ for $\theta = 0.02$), the variability reduces markedly.

6 Discussion

Our previous approach to studying MFPT memory lifetimes for simple synapses with multiple states of strength requires that the vector of possible strengths available to a synapse be a left eigenvector of the strength transition matrix. Even then, in deducing the jump moments from their expectation values, we simply ignored the possibility that the jump moments contain terms whose expectation values identically vanish. Nevertheless, our previous approach gave an extremely good analytical understanding of MFPT lifetimes in the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ models; it allowed an examination of the full FPT distribution, including in particular the variance; and we were able to examine in detail the impact of a non-zero firing threshold θ on memory lifetimes defined by FPT processes. By noting that the linear strength vector $\boldsymbol{\mathcal{L}}$ is close to the sigmoidal strength vector $\boldsymbol{\mathcal{S}}$, we argued that results for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ models, which do not satisfy the eigenvector requirement, should be very similar to those for the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ models, respectively, which do. This is confirmed by simulation results for all four models, but a more general analytical approach, not relying on the eigenvector condition, was highly desirable.

We have developed that more general approach here. In general, purely

Markovian dynamics in only the perceptron activation h do not exist, except for the very special case of binary-strength, $\nu = 2$ synapses (Elliott, 2014). In order to avoid non-Markovian dynamics it is necessary to extend the variables describing a synaptic configuration, so that we work with an orthogonal set of $\nu - 1$ variables, one of which is h. With these full, exact $\boldsymbol{v}$ variables, we are able to write down the one-memory-storage-step conditional MGF. From this we can write down the jump moments, take a continuum limit, and work with a Fokker-Planck equation in $\nu - 1$ variables, erecting an absorbing boundary on the $(\nu - 2)$-dimensional hyperplane $h = \theta$ to study FPT problems through this perceptron firing threshold. In addition, from the conditional PGF, we can obtain the transition probabilities $\mathsf{Prob}[\boldsymbol{v}|\boldsymbol{v}_0]$ and use them in the exact equation governing FPT statistics or, with the continuum limit, in the equivalent multi-dimensional integral equation, avoiding the diffusion limit approximation implicit in a Fokker-Planck approach. Of course, solving FPT problems in $\nu - 1$ variables is in general intractably hard. However, an examination of the structures of the drift and diffusion matrices reveals that we can work in a OU approximation, requiring p to be small enough and N to be large enough, but not so large as to be biologically meaningless. In this approximation, odd and even $\boldsymbol{v}$ variables decouple, and the latter can be thrown away because $h \equiv v_1$ is an odd variable. Further, with this approximation, the drift and diffusion matrices are related via $\mathbb{B} = -\frac{2}{N}\frac{1}{\nu}\,\mathbb{A}\,\mathbb{D}$ (or $\mathbb{B} = -\frac{1}{N}\frac{1}{\nu}(2\mathbb{A} + \mathbb{A}^2)\mathbb{D}$ at $\mathcal{O}(p^2)$). This relationship is extremely powerful, permitting the simultaneous diagonalisation of the $\mathbb{A}$ and $\mathbb{B}$ matrices in the differential equation for FPT statistics and the complete factorisation of the normal kernel in the corresponding integral equation by a transformation of variables. This allows an analytically tractable approximation based on considering only the 1-dimensional subspace defined by the maximal eigenvector

of $\mathbb{A}$, corresponding to the slowest eigenmode. A different, numerically better approximation is based on discarding weakly-coupled variables and restricting the odd variables to just two, $\boldsymbol{u} = (h, w)^{\mathrm{T}}$, whose contributions dominate FPT processes. In this way we obtain a 2-dimensional problem that is numerically approachable but one that is still (likely) analytically intractable in terms of FPT statistics.

We saw that the slowest eigenmode approximation is in fact extremely good, producing results often indistinguishable from simulation data, and breaking down only in the vicinity of the threshold, near which the faster processes associated with the other eigenmodes cannot be ignored. But for configurations initially sufficiently distant from threshold, the excellent agreement of this approximation with simulation or 2-dimensional numerical results indicates, perhaps unsurprisingly, that the FPT process is dominated by the 1-dimensional dynamics in the direction of $\mathbb{A}$'s maximal eigenvector. These dynamics are in the variable $x_1 = \widehat{\boldsymbol{\lambda}}^1 \cdot \boldsymbol{v}_0$ or $x_1 = \widehat{\boldsymbol{\ell}}^1 \cdot \boldsymbol{u}_0$, but the differential or integral equation describing these dynamics is 1-dimensional. The analysis of Eqs. (4.19) and (4.35) is immediate, since it is identical to our earlier analysis with the eigenvector condition: the results carry straight over to this more general case, but now in the x_1 variable rather than the perceptron activation h. The only difference is that we must finally average over either the (supra-threshold) $(\nu - 1)$-dimensional distribution of $\boldsymbol{v}_0$ or the 2-dimensional distribution of $\boldsymbol{u}_0$, rather than the 1-dimensional distribution of h_0. The greatest advantage of this 1-dimensional approximation, however, is that the FPT integral equation, which can be very hard to solve numerically even when restricted to the $\boldsymbol{u}$ variables, reduces to the 1-dimensional equation in Eq. (4.23) with its 1-dimensional kernel $\overline{K}_1(x|x_0)$ in Eq. (4.24). It is relatively easy to solve the 1-dimensional integral equation numerically. However, for the specific

case of $\nu = 2$, binary-strength synapses, we developed approximation methods for solving the integral equation analytically that produced very good results (Elliott, 2014). These methods used the theory of q-series. It will be interesting to see whether they carry over unchanged to the particular problem studied here.

The analysis performed here and in our earlier study (Elliott, 2017b) has looked at simple, multistate synapses. Simple synapses lack internal structure and are therefore easier to study than complex synapses that possess "internal" states that may modify the expression of synaptic plasticity. FPT approaches to memory lifetimes with complex synapses are analytically hard. For the specific case of binary-strength, $\nu = 2$ synapses, we examined MFPTs by developing methods that essentially integrate out the internal states, thereby replacing full transition matrices by simpler, 2×2 matrices (Elliott, 2017a). The price paid for this reduction of complex synapses to simple synapses is that the transition matrices become time- or memory-storage-step-dependent. In such a scenario, the left eigenvectors of the reduced transition matrices inevitably change over time or between storage step. Therefore, there is no possibility that such reduced dynamics will satisfy the eigenvector condition that we used in our earlier study of simple, multistate synapses (Elliott, 2017b). To examine FPT memory lifetimes in complex, multistate synapses using the method of reducing complex synapses to simple ones, an alternative approach, such as the one developed here, is necessary. Our focus here has been developing and validating this approach, but our intent is to apply it to complex, multistate synapses, which we shall do in future work.

Because our focus has been development and validation, we have not stressed biological issues, nor discussed the advantages and disadvantages of FPT approaches to memory lifetimes compared to other approaches. These broader

issues have been discussed extensively elsewhere (e.g., Elliott 2014, 2017b). Furthermore, at least in terms of the averaged forms $\langle \tau_{\mathrm{mfpt}}(\boldsymbol{u}_0) \rangle_{\boldsymbol{u}_0|h_0>\theta}$, our results here for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model are very similar to those for the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ model, and similarly for the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ models, and extensive discussion of the results under the eigenvector requirement, satisfied by the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ and $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ models, can again be found elsewhere (Elliott, 2017b). However, it is worth highlighting the radical difference between models that satisfy the eigenvector condition and those that do not, in terms of the unaveraged forms $\tau_{\mathrm{mfpt}}(\boldsymbol{u}_0)$. Models satisfying the eigenvector condition are insensitive to w_0. However, for small h_0, we saw extreme sensitivity of $\tau_{\mathrm{mfpt}}(\boldsymbol{u}_0)$ to the sign of w_0. This sensitivity is due to the influence of the drift rather than diffusion dynamics, with the drift either pushing the system rapidly through threshold or pushing it away from threshold.

The possibility that memory lifetimes may exhibit such extreme sensitivity to the initial synaptic configuration is intriguing, and may be regarded as a prediction of our approach. Unfortunately, although the perceptron activation $h = \frac{1}{N}\boldsymbol{e}^1 \cdot \boldsymbol{N}$ has a direct neuronal read-out in terms of membrane potential, $w = \frac{1}{N}\boldsymbol{e}^3 \cdot \boldsymbol{N}$ has no such convenient correlate. The variables v_i, $i \geq 2$, are, after all, introduced as auxiliary variables purely for the purposes of analysis. However, they *do* translate directly into numbers of synapses in (tilded) strength configurations. For example, for $\nu = 4$, setting $v_1 = 0$ (since h must be small) and $v_2 = 0$, for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model we have $\frac{1}{N}\boldsymbol{N} = (5-w, 5+3w, 5-3w, 5+w)^{\mathrm{T}}/20$. Saturating w at its minimum or maximum possible values of $\mp 5/3$ gives the two configurations $(1/3, 0, 1/2, 1/6)^{\mathrm{T}}$ and $(1/6, 1/2, 0, 1/3)^{\mathrm{T}}$, respectively. These are tilded configurations, but in an experiment we can directly control whether we potentiate or depress synapses, so with potentiation only or depression only they translate directly into relative probabilities of strength states. These con-

figurations are "skewed", with some states unoccupied. Thus, by driving a neuron's synapses to such skewed initial states, a subsequently presented, initially weakly-encoded memory may then either be very quickly forgotten or else endure for a long time, depending on the direction of skew. Moreover, the endurance of a memory would be driven by the storage of later memories, as they drive the system away from threshold because of the dominating drift dynamics. While it is no doubt formidably hard to perform experiments involving the careful preparation of initial synaptic configurations, this extreme sensitivity of memory lifetimes is a tantalising theoretical possibility. It is also fascinating to speculate on the possibility that memory systems may have evolved mechanisms to avoid such extreme sensitivity; one such mechanism would be to ensure that the left eigenvector condition is satisfied.

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.

A Integral Equation for $\boldsymbol{\tau^{(n)}(v_0)}$

We sketch the derivation of Eq. (3.30). Let $I_m(\boldsymbol{v}|\boldsymbol{v}_0)$ be the density for being at $\boldsymbol{v} \in \mathcal{I}$ immediately after non-tracked memory storage step m, starting from $\boldsymbol{v}_0 \in \mathcal{I}$ and having not escaped from $\mathcal{I}$ at intermediate steps. Clearly $I_0(\boldsymbol{v}|\boldsymbol{v}_0) = \delta(\boldsymbol{v} - \boldsymbol{v}_0)$. Let $f_m(\boldsymbol{v}_0) = \int_{\mathcal{I}}\cdots\int d\boldsymbol{v}\, I_m(\boldsymbol{v}|\boldsymbol{v}_0)$ be the probability of being in $\mathcal{I}$ at step m, again having not escaped at intermediate steps. We have

$$f_m(\boldsymbol{v}_0) = \int_{\mathcal{I}}\cdots\int d\boldsymbol{v} f_{m-1}(\boldsymbol{v})\mathsf{Prob}[\boldsymbol{v}|\boldsymbol{v}_0] \tag{A.1}$$

for $m \geq 1$, with $f_0(\boldsymbol{v}_0) \equiv 1$. Because memories are stored as a Poisson process of rate r, the probability of being in $\mathcal{I}$ at time $t \geq 0$ s, having not escaped at earlier times, is

$$P(\boldsymbol{v}_0; t) = \sum_{m=0}^{\infty} \frac{(rt)^m}{m!} e^{-rt} f_m(\boldsymbol{v}_0). \tag{A.2}$$

As usual, $-\frac{\partial}{\partial t} P(\boldsymbol{v}_0; t)$ is the first escape density, whence we obtain the moments

$$\tau^{(n)}(\boldsymbol{v}_0) = \frac{n}{r^n} \sum_{m=0}^{\infty} \frac{(m+n-1)!}{m!} f_m(\boldsymbol{v}_0), \tag{A.3}$$

for $n > 0$. Using Eq. (A.1) to drop $f_m(\boldsymbol{v}_0)$ down to $f_{m-1}(\boldsymbol{v}_0)$ and writing $\frac{(m+n-1)!}{m!} = \frac{(m-1+n-1)!}{(m-1)!} + (n-1)\frac{(m+n-2)!}{m!}$ for $m > 0$, the right hand side of Eq. (A.3) can be split into two parts that reduce to the two terms on the right hand side of Eq. (3.30), as required.

B Matrix Elements for $\mathbb{A}$ and $\mathbb{B}$

We list here for reference general results for $\boldsymbol{e}^{u_i} \cdot \mathbb{G}\, \boldsymbol{e}^{u_j}$ and $\beta_{u_i}^{-1}$ for $\nu \geq 4$ in all four models. For $\nu = 2$ or $\nu = 3$ we restrict to 1-dimensional dynamics using only $\boldsymbol{e}^h \cdot \mathbb{G}\, \boldsymbol{e}^h$ and β_h^{-1} from the following forms.

$$\left.\begin{aligned} \beta_h^{-1} &= \tfrac{1}{3}\,\nu\,\tfrac{\nu+1}{\nu-1} \\ \beta_w^{-1} &= \tfrac{1}{7}\,\nu\,\tfrac{(\nu+1)(\nu+2)(\nu+3)}{(\nu-1)(\nu-2)(\nu-3)} \end{aligned}\right\} \text{ for } \boldsymbol{\Omega} = \boldsymbol{\mathcal{L}}; \tag{B.1}$$

$$\left.\begin{aligned} \beta_h^{-1} &= \tfrac{1}{2}\,\nu\,\sec^2\tfrac{\pi}{2\nu} \\ \beta_w^{-1} &= \tfrac{1}{2}\,\nu\,\sec^2\tfrac{3\pi}{2\nu} \end{aligned}\right\} \text{ for } \boldsymbol{\Omega} = \boldsymbol{\mathcal{S}}; \tag{B.2}$$

$$\left.\begin{aligned} \boldsymbol{e}^h \cdot \mathbb{G}\, \boldsymbol{e}^h &= -2\,p\,\tfrac{1}{\nu-1} \\ \boldsymbol{e}^h \cdot \mathbb{G}\, \boldsymbol{e}^w &= -2\,p\,\tfrac{1}{\nu-1} \\ \boldsymbol{e}^w \cdot \mathbb{G}\, \boldsymbol{e}^w &= -12\,p\,\tfrac{\nu^2+1}{(\nu-1)(\nu-2)(\nu-3)} \end{aligned}\right\} \text{ for } \mathbb{C}\&\boldsymbol{\mathcal{L}} \text{ model}; \tag{B.3}$$

$$\left.\begin{aligned} \boldsymbol{e}^h \cdot \mathbb{G}\, \boldsymbol{e}^h &= -p\,\nu \tan^2 \tfrac{\pi}{2\nu} \\ \boldsymbol{e}^h \cdot \mathbb{G}\, \boldsymbol{e}^w &= 0 \\ \boldsymbol{e}^w \cdot \mathbb{G}\, \boldsymbol{e}^w &= -p\,\nu \tan^2 \tfrac{3\pi}{2\nu} \end{aligned}\right\} \text{ for } \mathbb{C}\&\boldsymbol{\mathcal{S}} \text{ model;} \tag{B.4}$$

$$\left.\begin{aligned} \boldsymbol{e}^h \cdot \mathbb{G}\, \boldsymbol{e}^h &= -\tfrac{1}{3}\, p\, \tfrac{\nu(\nu+1)}{(\nu-1)^2} \\ \boldsymbol{e}^h \cdot \mathbb{G}\, \boldsymbol{e}^w &= 0 \\ \boldsymbol{e}^w \cdot \mathbb{G}\, \boldsymbol{e}^w &= -\tfrac{6}{7}\, p\, \tfrac{\nu(\nu+1)(\nu+2)(\nu+3)}{(\nu-1)^2(\nu-2)(\nu-3)} \end{aligned}\right\} \text{ for } \mathbb{Q}\&\boldsymbol{\mathcal{L}} \text{ model;} \tag{B.5}$$

$$\left.\begin{aligned} \boldsymbol{e}^h \cdot \mathbb{G}\, \boldsymbol{e}^h &= -\tfrac{1}{48}\, p\, \tfrac{\nu}{\nu-1} \left[5 + \nu^2 + (1-\nu^2)\cos \tfrac{2\pi}{\nu}\right] \sec^4 \tfrac{\pi}{2\nu} \\ \boldsymbol{e}^h \cdot \mathbb{G}\, \boldsymbol{e}^w &= +\tfrac{1}{32}\, p\, \tfrac{\nu}{\nu-1} \left(2 + \sec \tfrac{\pi}{\nu}\right)^2 \sec^4 \tfrac{\pi}{2\nu} \\ \boldsymbol{e}^w \cdot \mathbb{G}\, \boldsymbol{e}^w &= -\tfrac{1}{48}\, p\, \tfrac{\nu}{\nu-1} \left[5 + \nu^2 + (1-\nu^2)\cos \tfrac{6\pi}{\nu}\right] \sec^4 \tfrac{3\pi}{2\nu} \end{aligned}\right\} \text{ for } \mathbb{Q}\&\boldsymbol{\mathcal{S}} \text{ model.} \tag{B.6}$$

Using these forms we may explicitly construct the matrices $\mathbb{A}$ and $\mathbb{B}$. In the limit of large ν, the forms of these matrices are particularly simple.

$$\mathbb{A} = -\frac{p}{\nu^2} \begin{pmatrix} 6 & 14 \\ 6 & 84 \end{pmatrix} \text{ and } \mathbb{B} = +\frac{1}{N}\frac{p}{\nu^2} \begin{pmatrix} 4 & 4 \\ 4 & 24 \end{pmatrix} \text{ for } \mathbb{C}\&\boldsymbol{\mathcal{L}} \text{ model;} \tag{B.7}$$

$$\mathbb{A} = -\frac{p}{\nu^2} \begin{pmatrix} \frac{\pi^2}{2} & 0 \\ 0 & \frac{9\pi^2}{2} \end{pmatrix} \text{ and } \mathbb{B} = +\frac{1}{N}\frac{p}{\nu^2} \begin{pmatrix} \frac{\pi^2}{2} & 0 \\ 0 & \frac{9\pi^2}{2} \end{pmatrix} \text{ for } \mathbb{C}\&\boldsymbol{\mathcal{S}} \text{ model;} \tag{B.8}$$

$$\mathbb{A} = -\frac{p}{\nu} \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix} \text{ and } \mathbb{B} = +\frac{1}{N}\frac{p}{\nu} \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & \frac{12}{7} \end{pmatrix} \text{ for } \mathbb{Q}\&\boldsymbol{\mathcal{L}} \text{ model;} \tag{B.9}$$

$$\mathbb{A} = -\frac{p}{\nu} \begin{pmatrix} \frac{3+\pi^2}{12} & -\frac{9}{16} \\ -\frac{9}{16} & \frac{1+3\pi^2}{4} \end{pmatrix} \text{ and } \mathbb{B} = +\frac{1}{N}\frac{p}{\nu} \begin{pmatrix} \frac{3+\pi^2}{12} & -\frac{9}{16} \\ -\frac{9}{16} & \frac{1+3\pi^2}{4} \end{pmatrix} \text{ for } \mathbb{Q}\&\boldsymbol{\mathcal{S}} \text{ model.} \tag{B.10}$$

C Statistics of Initial Distributions

We list here results for the means and covariances of the initial distributions immediately after the storage of $\boldsymbol{\xi}^0$ in the four models. From Eq. (3.36) restricted to $u_i(0)$, we have

$$\begin{aligned}\mathsf{E}[u_i(0)] &= \boldsymbol{e}^{u_i} \cdot \boldsymbol{p}_0, \\ \mathbb{S}_{ij}(0) &= \frac{1}{N}\big[\,(\boldsymbol{e}^{u_i} \circ \boldsymbol{e}^{u_j}) \cdot \boldsymbol{p}_0 - (\boldsymbol{e}^{u_i} \cdot \boldsymbol{p}_0)\,(\boldsymbol{e}^{u_j} \cdot \boldsymbol{p}_0)\,\big],\end{aligned}$$

where $\boldsymbol{p}_0$ is the initial distribution. Writing $(\boldsymbol{e}^{u_i} \circ \boldsymbol{e}^{u_j}) \cdot \boldsymbol{p}_0 = \mathbb{K}_{ij}(\boldsymbol{u}_0)$ as the elements of the matrix $\mathbb{K}(\boldsymbol{u}_0)$, we have the following results.

$$\mathsf{E}[\boldsymbol{u}_0] = \begin{pmatrix} \frac{2p}{\nu} \\ \frac{2p}{\nu} \end{pmatrix} \text{ and } \mathbb{K}(\boldsymbol{u}_0) = \begin{pmatrix} \frac{1}{3}\frac{\nu+1}{\nu-1} & 0 \\ 0 & \frac{1}{7}\frac{(\nu+1)(\nu+2)(\nu+3)}{(\nu-1)(\nu-2)(\nu-3)} \end{pmatrix} \text{ for } \mathbb{C}\&\boldsymbol{\mathcal{L}} \text{ model;} \tag{C.1}$$

$$\mathsf{E}[\boldsymbol{u}_0] = \begin{pmatrix} \frac{2p}{\nu} \\ \frac{2p}{\nu} \end{pmatrix} \text{ and } \mathbb{K}(\boldsymbol{u}_0) = \begin{pmatrix} \frac{1}{2}\sec^2\frac{\pi}{2\nu} & 0 \\ 0 & \frac{1}{2}\sec^2\frac{3\pi}{2\nu} \end{pmatrix} \text{ for } \mathbb{C}\&\boldsymbol{\mathcal{S}} \text{ model;} \tag{C.2}$$

$$\mathsf{E}[\boldsymbol{u}_0] = \begin{pmatrix} \frac{p}{3}\frac{\nu+1}{\nu-1} \\ 0 \end{pmatrix} \text{ and } \mathbb{K}(\boldsymbol{u}_0) = \begin{pmatrix} \frac{1}{3}\frac{\nu+1}{\nu-1} & 0 \\ 0 & \frac{1}{7}\frac{(\nu+1)(\nu+2)(\nu+3)}{(\nu-1)(\nu-2)(\nu-3)} \end{pmatrix} \text{ for } \mathbb{Q}\&\boldsymbol{\mathcal{L}} \text{ model;} \tag{C.3}$$

$$\mathsf{E}[\boldsymbol{u}_0] = \begin{pmatrix} \frac{p\,\mathrm{cosec}^2\frac{\pi}{2\nu}}{\nu(\nu-1)} \\ \frac{p\,\mathrm{cosec}^2\frac{3\pi}{2\nu}}{\nu(\nu-1)} \end{pmatrix} \text{ and } \mathbb{K}(\boldsymbol{u}_0) = \begin{pmatrix} \frac{1}{2}\sec^2\frac{\pi}{2\nu} & 0 \\ 0 & \frac{1}{2}\sec^2\frac{3\pi}{2\nu} \end{pmatrix} \text{ for } \mathbb{Q}\&\boldsymbol{\mathcal{S}} \text{ model.} \tag{C.4}$$

References

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

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

Bartol, T.M., Bromer, C., Kinney, J., Chirillo, M.A., Bourne, J.N., Harris, K.M., & Sejnowski, T.J. 2015. Nanoconnectomic upper bound on the variability of synaptic plasticity. *eLife*, **4**, e10778.

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

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

Elliott, T., & Lagogiannis, K. 2012. The rise and fall of memory in a model of synaptic integration. *Neural Comput.*, **24**, 2604–2654.

Fusi, S., & Abbott, L.F. 2007. Limits on the memory storage capacity of bounded synapses. *Nature Neurosci.*, **10**, 485–493.

Fusi, S., Drew, P.J., & Abbott, L.F. 2005. Cascade models of synaptically stored memories. *Neuron*, **45**, 599–611.

Hertz, J., Krogh, A., & Palmer, R.G. 1991. *Introduction to the Theory of Neural Comput.*. Redwood City, CA: Addison Wesley.

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. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, Z., & K.Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26* (pp. 1034–1042). 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.

Montgomery, J.M., & Madison, D.V. 2002. State-dependent heterogeneity in synaptic depression between pyramidal cell pairs. *Neuron*, **33**, 765–777.

Montgomery, J.M., & Madison, D.V. 2004. Discrete synaptic states define a major mechanism of synapse plasticity. *Trends Neurosci.*, **27**, 744–750.

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

O'Connor, D.H., Wittenberg, G.M., & Wang, S.S.-H. 2005a. Dissection of bidirectional synaptic plasticity into saturable unidirectional process. *J. Neurophysiol.*, **94**, 1565–1573.

O'Connor, D.H., Wittenberg, G.M., & Wang, S.S.-H. 2005b. Graded bidirectional synaptic plasticity is composed of switch-like unitary events. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 9679–9684.

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

Petersen, C.C.H., Malenka, R.C., Nicoll, R.A., & Hopfield, J.J. 1998. All-or-none potentiation at CA3-CA1 synapses. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 4732–4737.

Risken, H. 1984. *The Fokker-Planck Equation.* Berlin: Springer.

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

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.

Figure Captions

Figure 1: Expected drifts $\mathsf{E}[v_i(t)]$ for the odd variables plotted against time for the (A) $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ and (B) $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ models, for $\nu = 16$. Reading from top to bottom for small times in both panels, the variables correspond to $i = 1, 3, 5, 7, 9, 11, 13$ and 15, in order. The probability parameter p acts as a simple overall scale factor and has been removed, as indicated.

Figure 2: Evolution of marginal and joint distributions for $h(t)$ and $w(t)$ for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model. Results are shown for $\nu = 6$ (panels A, C and E) and $\nu = 16$ (panels B, D and F). The mean and one standard deviation around it of the 1-dimensional marginals for $h(t)$ (panels A and B) and $w(t)$ (panels C and D) are shown, as well as the elliptical one standard deviation region (in the sense of Mahalanobis) around the mean of the 2-dimensional joint distribution for various choices of t (panels E and F). For the 1-dimensional marginals, results are shown for the full, unapproximated forms of the $\boldsymbol{v}$ variables (solid lines), their OU forms (long-dashed lines), and for the restricted $\boldsymbol{u}$ variables (short-dashed lines). For the joint distribution we have used $P(\boldsymbol{u}; t)$ for times from $rt = 0.1$ to either $rt = 10^3$ (E) or $rt = 10^4$ (F) with uniform logarithmic spacing, resulting in a sequence of elliptical regions from top right to bottom left in each panel that indicate how the joint distribution evolves over time. We have set $p = 0.1$ and $N = 10^4$ in all cases.

Figure 3: Evolution of marginal and joint distributions for $h(t)$ and $w(t)$ for the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model. The format and parameter choices in this figure are otherwise identical to Fig. 2.

Figure 4: Exact mean first passage times plotted against synaptic configuration number. Results are shown for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ (panel A), $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ (panel B), $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$(panel C) and $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ (panel D) models. Each datum point corresponds to a particular synaptic configuration, with the enumeration of configurations described in the main text. The insets in panels A and C magnify four-fold the eleventh cluster of data points, corresponding to $h_0 = 0.28$. The data points plotted in red in panel B correspond to those plotted in black but their $\boldsymbol{v}_0$ values have been binned to highlight the vertical structure; the binned data points are displaced vertically to distinguish them clearly from the unbinned data points. We have set $p = 0.1$, $\nu = 4$ and $N = 25$ in all cases and used a threshold $\theta = 0$.

Figure 5: Comparison between exact and simulation results for mean first passage times. MFPTs in the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ (panels A and C) and $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ (panels B and D) models are plotted against h_0 and w_0 for $p = 0.1$ and $\nu = 4$ for a threshold $\theta = 0$ for either $N = 25$ (panels A and B) or $N = 10^5$ (panels B and D). In all panels, red points or lines show analytical results while black points or lines show simulation results. For $N = 25$, analytical results are obtained as discrete data points for each discrete synaptic configuration, while for $N = 10^5$, analytical results are obtained in the continuum limit. In panels A and B, the green crosses on the w_0–h_0 plane indicate the discrete values of w_0 and h_0 corresponding to red MFPTs displayed above. Apparent differences in the thicknesses of the red circles reflect the further degree of freedom in the $\boldsymbol{v}_0$ variables for $\nu = 4$, indicating essentially superimposed data points; similarly for the black data points, except that simulation noise can separate these nearly superimposed points. Simulation data points for $N = 25$ are shown only when the number of samples over which averages are taken exceeds 500;

typically, the number of samples per simulation datum point is in excess of 11,000, so well above 500. In panels C and D contours of $r\tau_{\mathrm{mfpt}}$ are shown in the w_0–h_0 plane with the colours of contours corresponding to analytical or simulation results; contours of $r\tau_{\mathrm{mfpt}} = 5$, 20, 40, 60, 80, 100, 120 are shown.

Figure 6: Evolution of mean flow $\mathsf{E}[\boldsymbol{u}(t)|\boldsymbol{u}_0] = e^{\mathbb{A}^* t}\boldsymbol{u}_0$ for different choices of $\boldsymbol{u}_0 = (h_0, w_0)^{\mathrm{T}}$. Results for the $\mathbb{C}\&\boldsymbol{\mathcal{S}}$ (panel A) and the $\mathbb{Q}\&\boldsymbol{\mathcal{L}}$ (panel B) models are shown. In both panels, the initial states are chosen to lie on the circle $|\boldsymbol{u}_0| = 1$, and flows move towards the origin. The overall scale in these panels is arbitrary for the $\mathbb{A}^*$ matrix. The two arrows in each panel show the directions associated with the eigenvectors of $\mathbb{A}^*$.

Figure 7: Comparison between mean first passage times obtained from the integral equation and Fokker-Planck approach. MFPTs are shown as contours in the w_0^*–h_0^* plane, with dashed contours for results from the integral equation and solid contours from the Fokker-Planck approach. Panels A ($\mathbb{C}\&\boldsymbol{\mathcal{L}}$, $\kappa = 2$) and B ($\mathbb{Q}\&\boldsymbol{\mathcal{S}}$, $\kappa = 1$) show contours of $\frac{p\,r}{\nu^\kappa}\tau_{\mathrm{mfpt}}(\boldsymbol{u}_0^*)$ for the specific choice of $\nu = 4$, while panels C and D show the equivalents in the formal, large ν scaling limit. (A) From bottom right to top left contours correspond in pairs to values of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7; (B) From bottom left to top right, 0.1, 0.22, 0.46, 1, 1.5, 2, 2.5; (C) From bottom right to top left the first ten pairs correspond to 0.0022, 0.0046, 0.01, 0.022, 0.046, 0.1, 0.2, 0.3, 0.4, 0.5 while the three unpaired solid contours correspond to 0.6, 0.7, 0.8; (D) From bottom left to top right the first six pairs correspond to 0.1, 0.22, 0.46, 1, 1.5, 2 while the three unpaired solid contours correspond to 2.5, 3, 3.5. Kinks in a few solid contours correspond to boundary artifacts from numerical solutions. The threshold $\theta = 0$ throughout.

Figure 8: Comparison between mean first passage times obtained numerically and from analytical approximations, in the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model. MFPTs are plotted against w_0 for five different, fixed choices of h_0 in each panel. Results are shown for three different choices of N ($N = 10^3$ in panels A and B; 10^4 in C and D; 10^5 in E and F) and for two different choices of ν ($\nu = 8$ in panels A, C and E; 16 in B, D and F). Solid lines show numerical results obtained from the 2-dimensional Fokker-Planck approach while dashed lines show analytical results based on the approximation in Eq. (4.21). Each pair of different lines styles corresponds to the same choice of h_0, with values moving from top to bottom in each panel of $h_0 = 2^{+2}\mu$, $2^{+1}\mu$, $2^0\mu$, $2^{-1}\mu$ and $2^{-2}\mu$ where $\mu = \mathsf{E}[h_0] = 2\,p/\nu$. We have taken $p = 0.1$ throughout and used a threshold $\theta = 0$. The lower two solid lines in panel F are truncated due to numerical instabilities in the 2-dimensional solution.

Figure 9: Comparison between mean first passage times obtained numerically and from analytical approximations, in the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model. The format of this figure is identical to that for Fig. 8, except that we taken seven choices of h_0, corresponding to $h_0 = 2^{+2}\mu$, $2^{+1}\mu$, $2^0\mu$, $2^{-1}\mu$, $2^{-2}\mu$, $2^{-3}\mu$ and $2^{-4}\mu$ where for the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model $\mu = \mathsf{E}[h_0] = \frac{p\,\mathrm{cosec}^2\frac{\pi}{2\nu}}{\nu(\nu-1)}$.

Figure 10: Mean first passage times averaged over the initial distribution of $\boldsymbol{u}_0$ above threshold as a function of N or ν. Lines show results obtained from the approximation method discussed in section 4.3, with dashed lines showing MFPTs from the Fokker-Planck approach in Eq. (4.21) and solid lines from the integral equation in Eq. (4.25); circles show results from simulations. Panels A and C show results for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model and panels B and D show results for

the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model. From bottom to top in each of panels A and B, each pair of lines corresponds to the choice $\nu = 5$, 10 and 20; in panels C and D, $N = 10^3$, 10^4 and 10^5. We have set $p = 0.1$ and used $\theta = 0$ throughout.

Figure 11: Mean first passage times and one standard deviation in the FPT distribution around them as a function of N, for $\nu = 10$. Black lines show results obtained from the approximation method in section 4.3, with black dashed lines corresponding to the Fokker-Planck approach and black solid lines to the integral equation method; black open circles show results from simulations. The middle pair of black lines in each panel correspond to $r\langle\tau_{\mathrm{mfpt}}(\boldsymbol{u}_0)\rangle_{\boldsymbol{u}_0|h_0>\theta}$ while the upper and lower black pairs correspond to $r\langle\tau_{\mathrm{mfpt}}(\boldsymbol{u}_0)\rangle_{\boldsymbol{u}_0|h_0>\theta} \pm r\langle\sigma_{\mathrm{mfpt}}(\boldsymbol{u}_0)\rangle_{\boldsymbol{u}_0|h_0>\theta}$, respectively; the lower black pair in panel E are missing because their values are negative. Also shown in panel E in red are results obtained from the 2-dimensional integral equation in Eq. (4.14); data points correspond to the small red filled circles, connected by red solid lines. Panels A, C and E show results for the $\mathbb{C}\&\boldsymbol{\mathcal{L}}$ model and panels B, D and F show results for the $\mathbb{Q}\&\boldsymbol{\mathcal{S}}$ model. Each panel indicates the value of θ used. We have set $p = 0.1$ throughout.

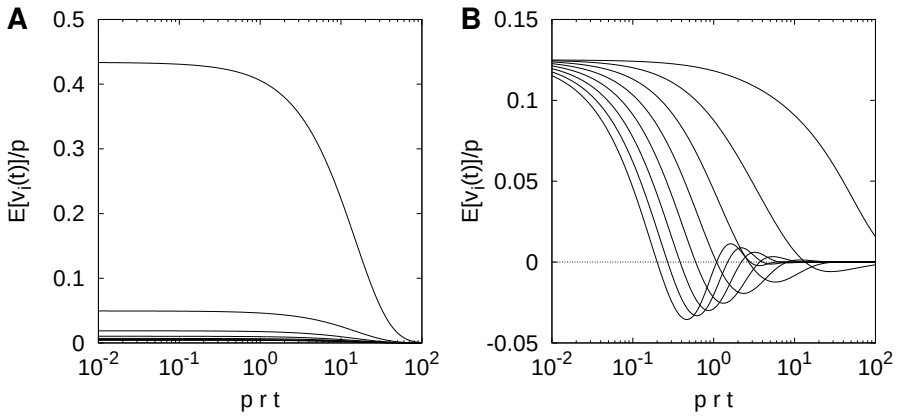

Figure 1

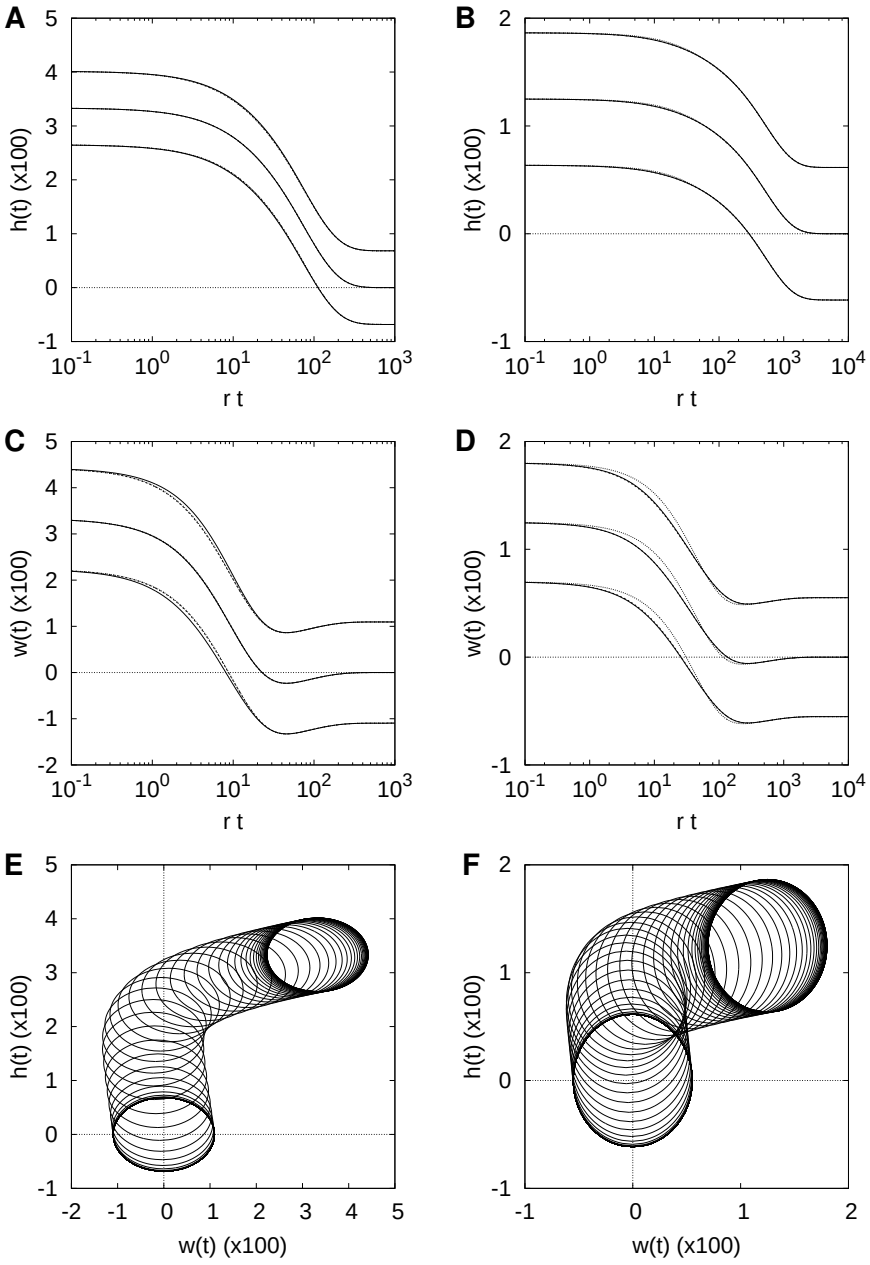

Figure 2

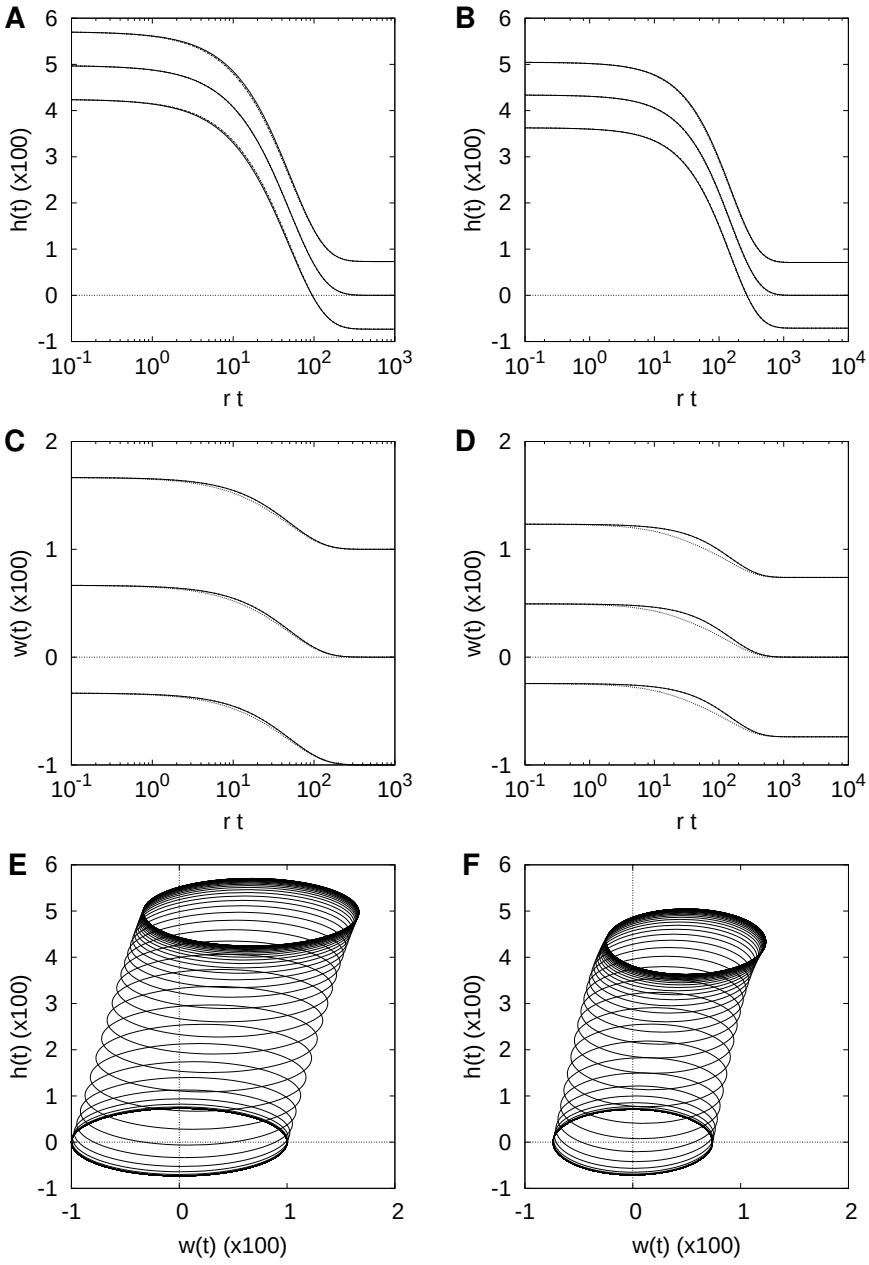

Figure 3

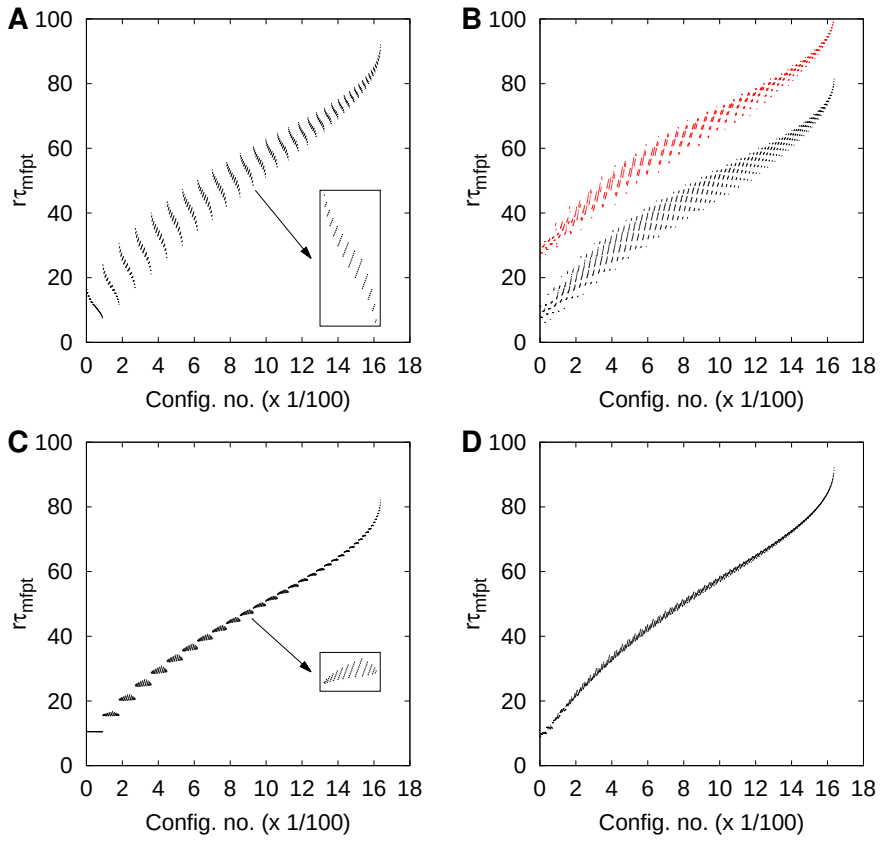
A
B
C
D
100
80
60
40
20
0
0 2 4 6 8 10 12 14 16 18
rτmfpt
Config. no. (x 1/100)

Figure 4

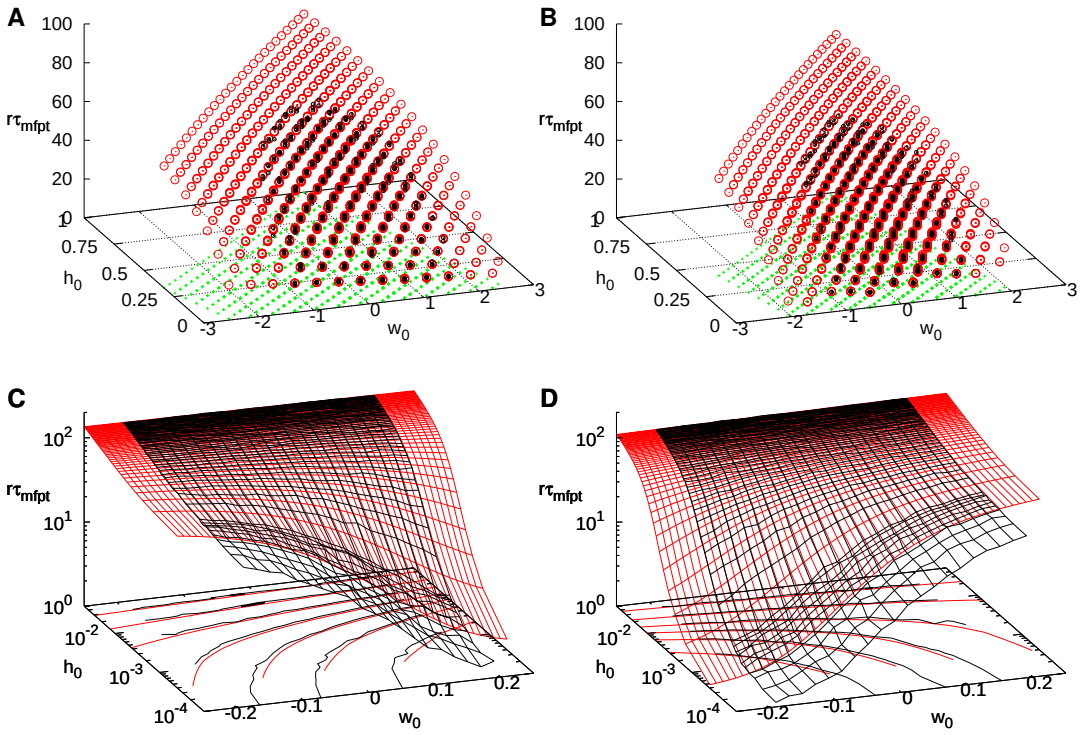

Figure 5

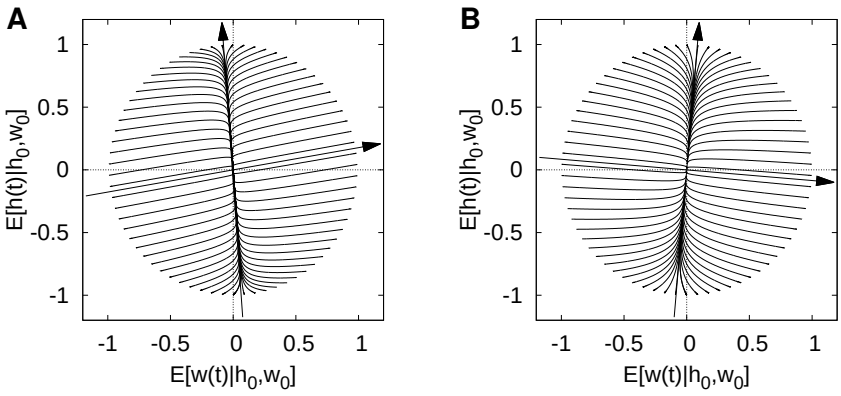

Figure 6

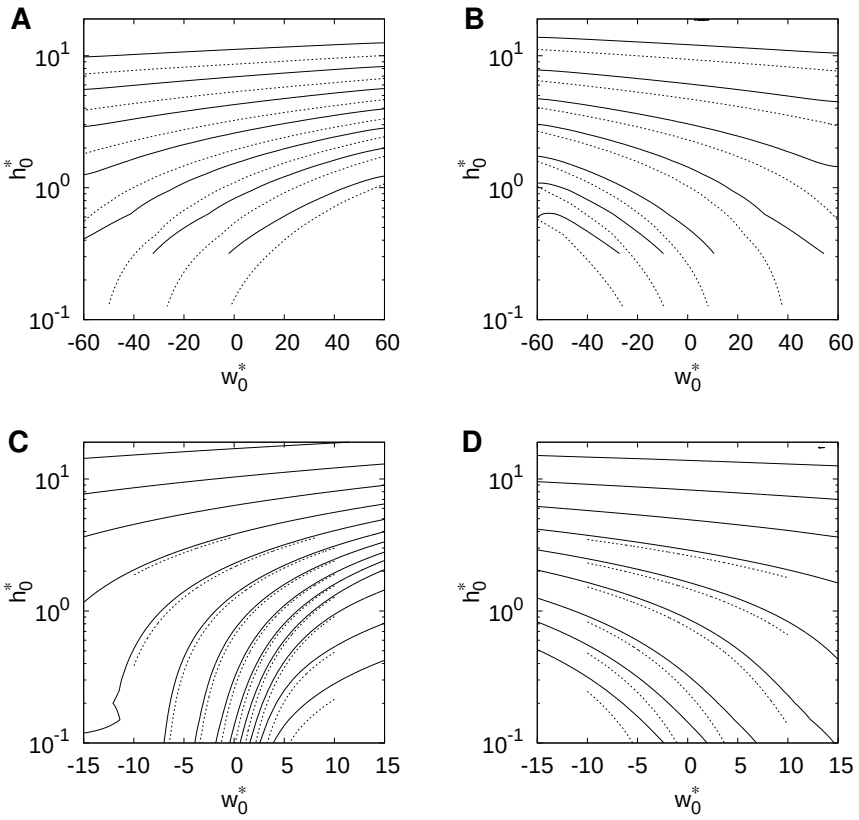

Figure 7

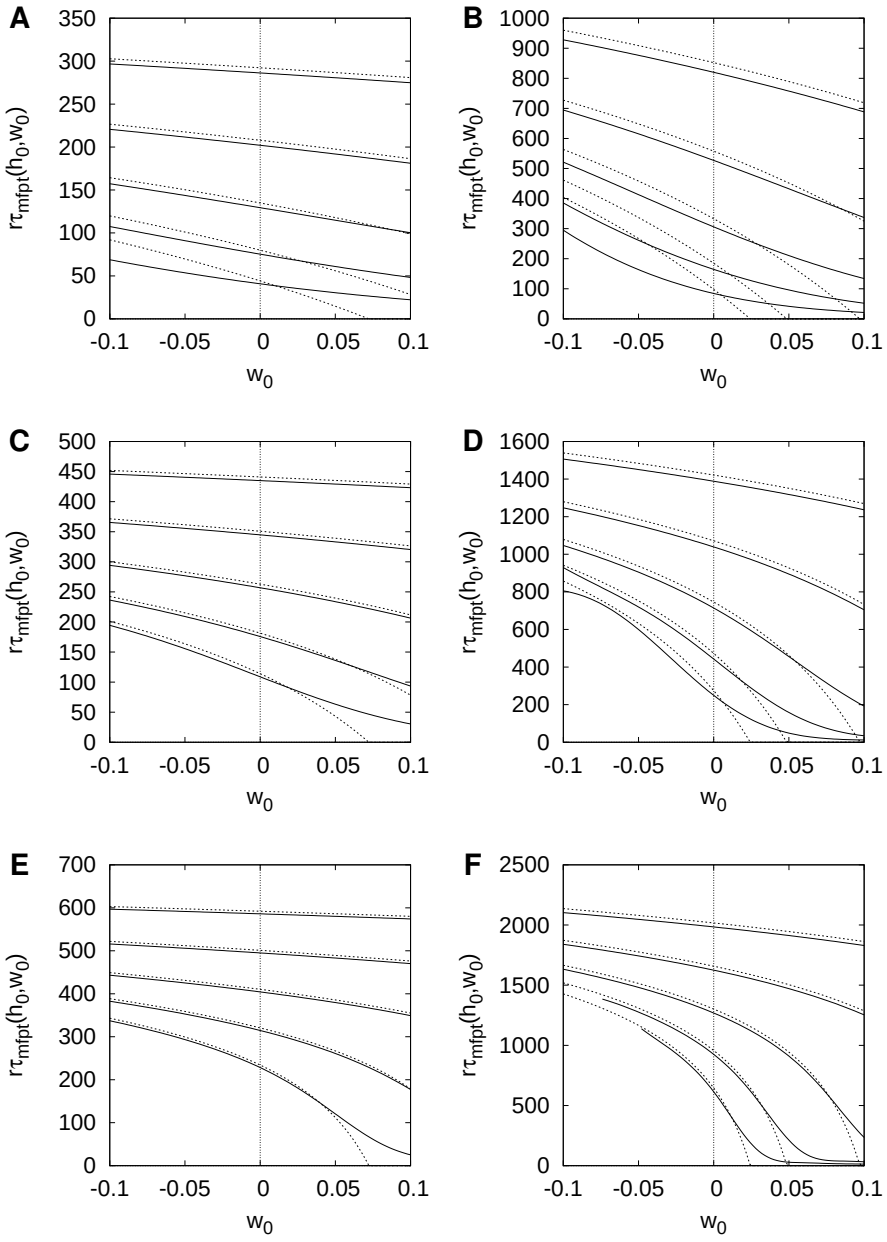

Figure 8

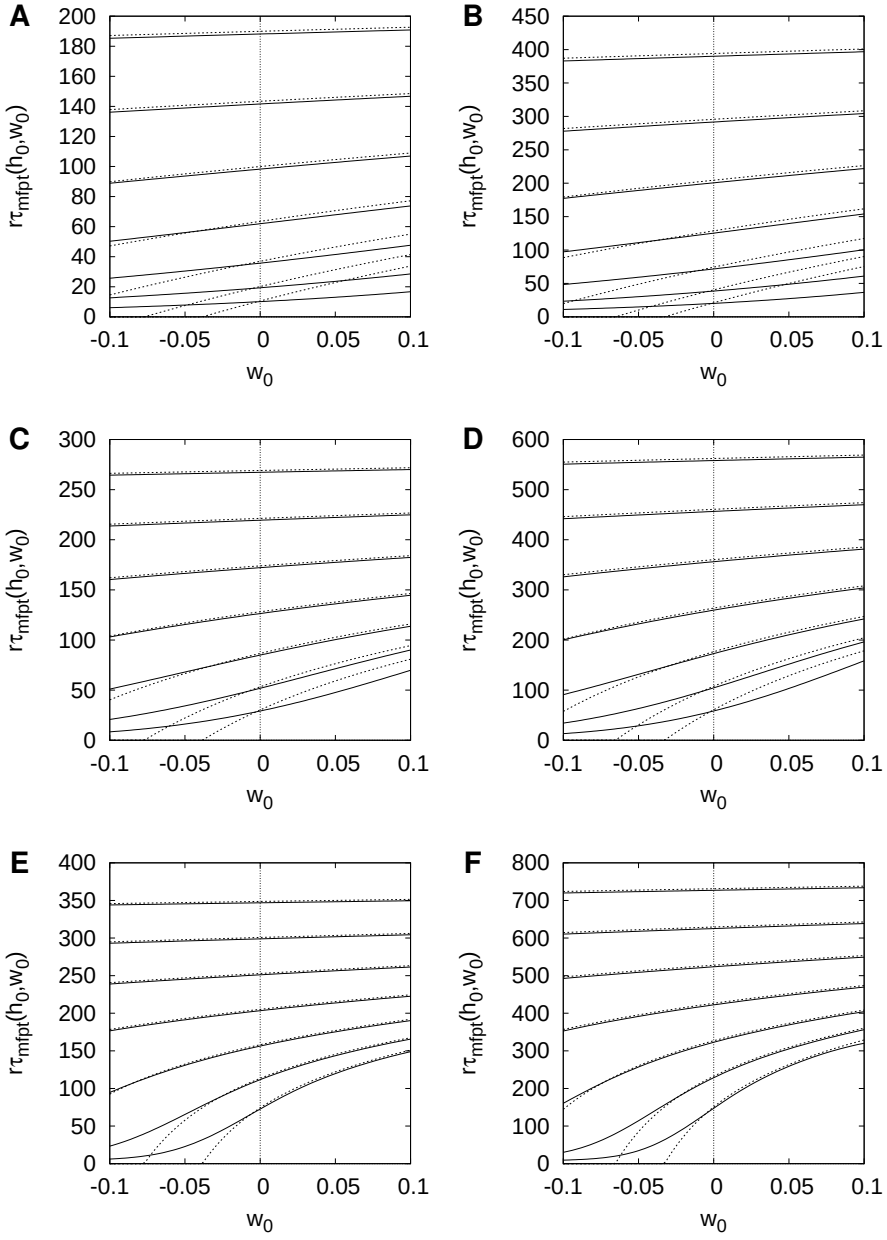

Figure 9

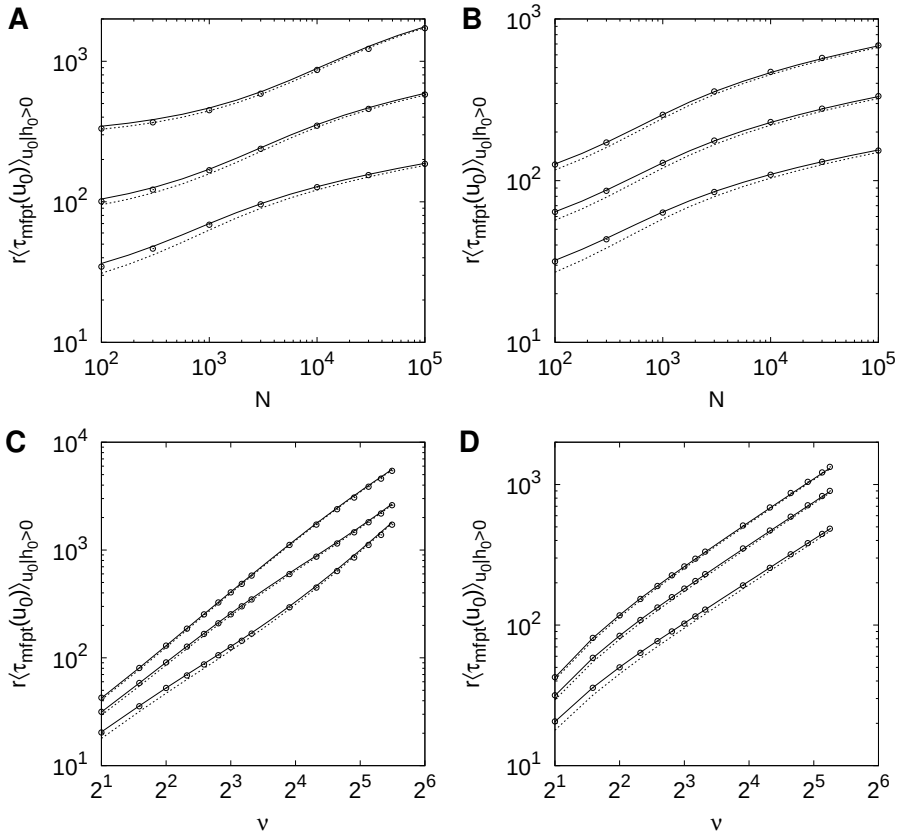

Figure 10

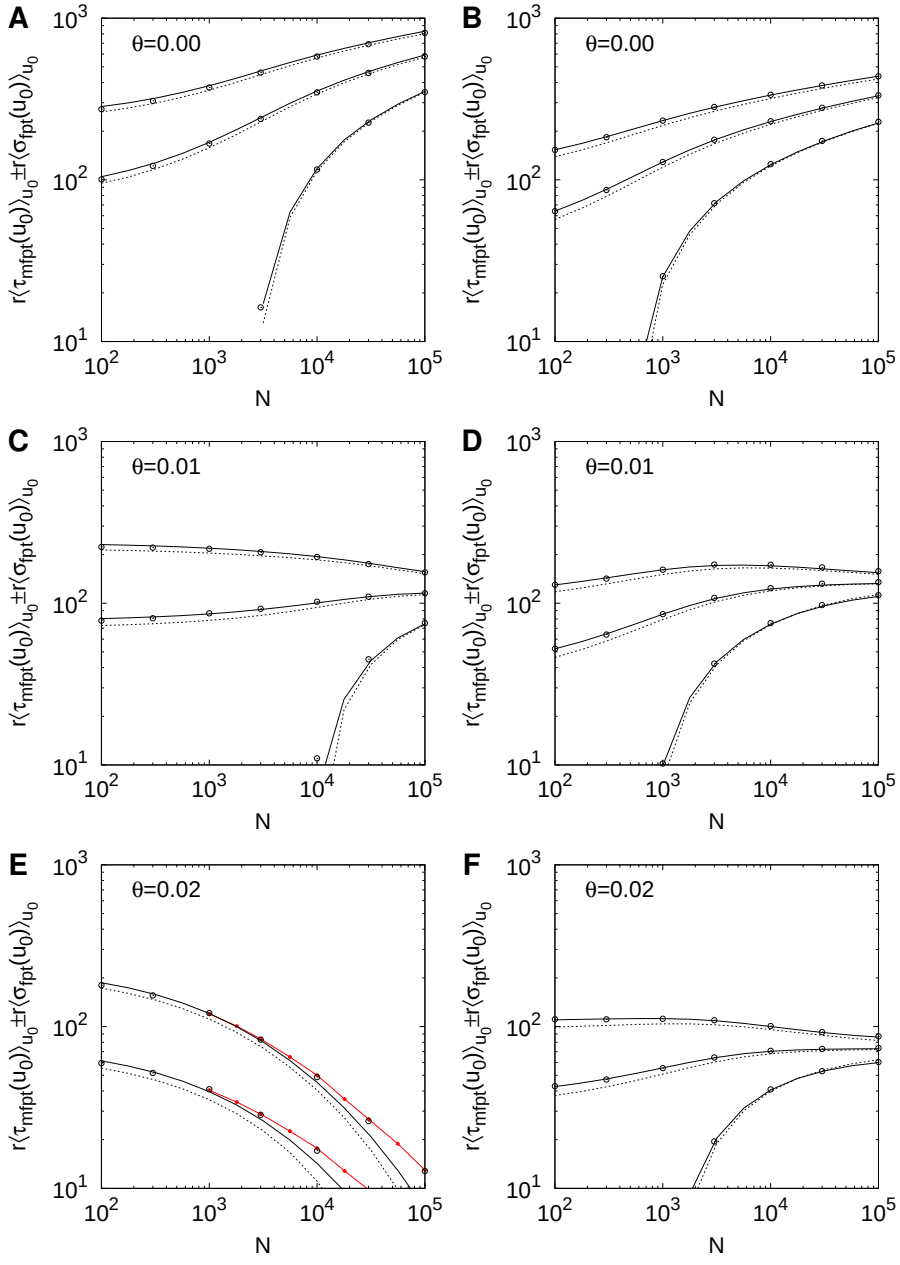
A
θ=0.00
B
θ=0.00
C
θ=0.01
D
θ=0.01
E
θ=0.02
F
θ=0.02
N
r⟨τmfpt(u0)⟩u0±r⟨σfpt(u0)⟩u0

Figure 11