

# First Passage Time Memory Lifetimes for Simple, Multistate Synapses

Terry Elliott<sup>1</sup>

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

**Running Title:** FPT memory lifetimes.

July 2, 2017.

---

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

## Abstract

Memory models based on synapses with discrete and bounded strengths store new memories by forgetting old ones. Memory lifetimes in such memory systems may be defined in a variety of ways. A mean first passage time (MFPT) definition overcomes much of the arbitrariness and many of the problems associated with the more usual signal-to-noise ratio (SNR) definition. We have previously computed MFPT lifetimes for simple, binary-strength synapses that lack internal, plasticity-related states. In simulation we have also seen that for multistate synapses, optimality conditions based on SNR lifetimes are absent with MFPT lifetimes, suggesting that such conditions may be artifactual. Here we extend our earlier work by computing the entire first passage time (FPT) distribution for simple, multistate synapses, from which all statistics including the MFPT lifetime may be extracted. For this, we develop a Fokker-Planck equation using the jump moments for perceptron activation. Two models are considered that satisfy a particular eigenvector condition that this approach requires. In these models, MFPT lifetimes do not exhibit optimality conditions, while in one but not the other, SNR lifetimes do exhibit optimality. Thus, not only are such optimality conditions artifacts of the SNR approach, but they are also strongly model-dependent. By examining the variance in the FPT distribution, we may identify regions in which memory storage is subject to high variability, although MFPT lifetimes are nevertheless robustly positive. In such regions, SNR lifetimes are typically (defined to be) zero. FPT-defined memory lifetimes therefore provide an analytically superior approach and also have the virtue of being directly related to a neuron's firing properties.

# 1 Introduction

Imposing limits on synaptic strengths turns an otherwise catastrophically forgetting Hopfield (1982) network into a “palimpsest” memory that learns new memories by forgetting old ones (Nadal *et al.*, 1986; Parisi, 1986). Models of palimpsest memory with discrete, multistate synapses using feedforward or recurrent networks have become the subject of intensive study in recent years (Tsodyks, 1990; Amit & Fusi, 1994, Fusi *et al.*, 2005, Leibold & Kempster, 2006, 2008; Rubin & Fusi, 2007; Barrett & van Rossum, 2008; Huang & Amit, 2010, 2011; Elliott & Lagogiannis, 2012; Lahiri & Ganguli, 2013; Elliott, 2016a,b). Such models may be based on “simple” synapses that lack internal, plasticity-related states, or “complex” synapses that possess internal states that may affect the expression of synaptic plasticity.

To be viable models of biological memory, memories in palimpsest models must be sufficiently long-lived. Several approaches to defining palimpsest memory lifetimes exist, including the signal-to-noise ratio (SNR) (Tsodyks, 1990) and equivalent so-called “ideal observer” variants (Fusi *et al.*, 2005; Lahiri & Ganguli, 2013; see Elliott (2016b) for a discussion of their complete equivalence); signal detection theory (Leibold & Kempster, 2006, 2008); and retrieval probabilities (Huang & Amit, 2010, 2011). In a feedforward setting with a single perceptron for simplicity, we have also considered the mean first passage time (MFPT) for the perceptron’s activation to fall below firing threshold (Elliott, 2014). An MFPT approach to memory lifetimes overcomes many of the difficulties of an SNR approach and shows that the latter is only asymptotically valid in the limit of a large number of synapses (Elliott, 2014). We have also observed in simulation that conditions on the number of states of synaptic strength that appear to optimise SNR memory lifetimes are not respected by

MFPT lifetimes, suggesting that such optimality conditions are artifacts of the SNR approach (Elliott, 2016a).

We may obtain exact analytical results for MFPT lifetimes for any synaptic model but the results are essentially useless for explicit computations. For the specific case of simple, binary-strength synapses, we may reduce the difficulty of the calculations by considering transitions in the perceptron's activation at successive memory storage steps (Elliott, 2014). This allows us to derive approximation methods and reduce the dynamics of memory decay to an Ornstein-Uhlenbeck (OU) process (Uhlenbeck & Ornstein, 1930). It is also possible to make some progress in understanding MFPT memory lifetimes for complex synapses with binary strengths by integrating out the internal states and working directly in the transitions in synapses' strengths (Elliott, 2017). For general, multistate synapses however, whether simple or complex, we cannot work directly in the transitions in the perceptron's activation, as discussed below. Here, we show that for simple synapses, we can obtain the entire first passage time (FPT) distribution from a Fokker-Planck equation when the vector of strengths available to a synapse is an eigenvector of the stochastic matrix governing changes in synapses' strengths. Provided that the actual vector of possible synaptic strengths is sufficiently close to an eigenvector, our results give good approximations, so this eigenvector requirement is not too restrictive.

Our paper is organised as follows. In section 2 we define our general formalism and review the derivation of analytical results for MFPTs for simple, binary-strength synapses. In section 3 for simple, multistate synapses we set up a Fokker-Planck equation, derive the required jump moments, and then obtain the FPT distribution. In section 4 we consider two different synaptic models respecting the eigenvector requirement. In section 5, we derive SNR memory lifetimes for the purposes of comparison with MFPT memory lifetimes. We

examine our results in section 6, comparing analytical and simulation results, and considering the differences between SNR and MFPT memory lifetimes, but also considering the variance in FPT-defined lifetimes. Finally, in section 7, we briefly discuss our approach.

## 2 General Formalism and Previous Results

We first summarise our general approach to studying memory lifetimes in a feedforward, perceptron-based formulation. We then discuss the simplest possible model of synaptic plasticity for palimpsest memory. We finally briefly review our previous analysis of MFPT memory lifetimes for simple, binary-strength synapses. Full details may be found elsewhere (Elliott, 2014).

### 2.1 Perceptron Memory

A single perceptron with  $N$  synapses of strengths  $S_i(t)$ ,  $i = 1, \dots, N$ , at time  $t \geq 0$  s and input vector  $\mathbf{x}$  with components  $x_i$  has normalised total input or activation or unthresholded output defined by

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

We are concerned only with whether or not  $h_{\mathbf{x}}(t)$  is above the perceptron's firing threshold, defined as  $\theta$ . The synaptic strengths  $S_i(t)$  take values from a discrete set. For binary-strength synapses, these values are taken to be  $S_i(t) \in \{-1, +1\}$ . For multistate synapses with  $\nu$  discrete levels of strength, so for  $\nu > 2$ , we will consider different possible choices of this set of values below.

The perceptron sequentially stores memories  $\boldsymbol{\xi}^\alpha$ , indexed by  $\alpha = 0, 1, 2, \dots$ , with components  $\xi_i^\alpha$ . These memories may be presented as a discrete time

process, or more realistically for biological memory storage as a continuous time process, which we take to be a Poisson process of rate  $r$ . The first memory  $\xi^0$  is always presented at time  $t = 0^-$  s, where we use this formal device of  $t = 0^-$  s rather than  $t = 0$  s so that we may refer to the time immediately after the storage of  $\xi^0$  as time  $t = 0$  s. The components  $\xi_i^\alpha$  take binary values  $\xi_i^\alpha \in \{-1, +1\}$  with probabilities  $\text{Prob}[\xi_i^\alpha = \pm 1] = g_\pm$ , with  $g_+ + g_- = 1$ . Any particular memory  $\xi^\alpha$  is deemed to be stored at time  $t$  provided that the perceptron's activation upon re-presentation of the memory exceeds threshold,  $h_{\xi^\alpha}(t) > \theta$ . As we will assume that  $\theta \geq 0$ , the perceptron's output is required to be positive for memory storage. The component  $\xi_i^\alpha$  is therefore the plasticity induction signal to synapse  $i$  upon storage of memory  $\alpha$ . Consistent with our previous work, we set  $g_\pm = \frac{1}{2}$ , so that potentiation ( $\xi_i^\alpha = +1$ ) and depression ( $\xi_i^\alpha = -1$ ) processes are balanced.

To assess memory lifetimes, we track the fidelity of recall of the first memory  $\xi^0$  as the later memories  $\xi^\alpha$ ,  $\alpha > 0$ , are stored. The storage of these later memories leads to changes in synaptic strengths that may affect the recall of  $\xi^0$ . We refer to memory  $\xi^0$  as the tracked memory and we define

$$h(t) = h_{\xi^0}(t) = \frac{1}{N} \sum_{i=1}^N \xi_i^0 S_i(t), \quad (2.2)$$

and refer to  $h(t)$  as the tracked memory signal. As the memories  $\xi^\alpha$  are stochastic in nature and the Poisson times at which they are stored are random variables, the memory signal  $h(t)$  is a random variable governed by a probability distribution. Its mean and variance,

$$\mu(t) = \text{E}[h(t)], \quad (2.3a)$$

$$\sigma(t)^2 = \text{Var}[h(t)], \quad (2.3b)$$

are used to define the SNR  $\mathcal{SNR}(t) = [\mu(t) - \mu(\infty)]/\sigma(t)$ , and the SNR memory lifetime of any particular model is typically defined as the solution  $\tau_{\text{snr}}$  of  $\mathcal{SNR}(\tau_{\text{snr}}) = 1$ . Some variants of the SNR approach use  $\sigma(\infty)$  rather than  $\sigma(t)$  in the denominator of  $\mathcal{SNR}(t)$ , but this approach is less well justified from a statistical point of view (Elliott, 2016b).

The SNR definition of memory lifetime suffers from a number of difficulties that we have previously described (Elliott, 2014). First, there is some arbitrariness in defining  $\tau_{\text{snr}}$  via  $\mathcal{SNR}(\tau_{\text{snr}}) = 1$ : we could use any other positive number on the right-hand side instead. Second, the SNR considers only the variance as a possible source of fluctuations that may render the memory signal indistinguishable from its equilibrium value. Third, SNR memory lifetimes differ depending on whether memories are stored as a discrete time process or as a continuous time process. Fourth, because the SNR mixes different signal statistics, it is not a quantity that can be read out directly from a neuron’s membrane potential, and so it is not a quantity of immediate relevance to the system whose memory dynamics are being studied.

## 2.2 Stochastic Updater Synapses

The simplest possible model of synaptic plasticity for memory storage is based on a simple, binary-strength synapse that expresses with probability  $p$  a change in synaptic strength (if possible) when the synapse experiences a plasticity induction signal (Tsodyks, 1990). We refer to such a synapse as a “stochastic updater”. The strength  $S_i(t)$  of synapse  $i$  is a random variable. For a binary-strength synapse, the probability distribution of a synapse’s strength is represented by a 2-dimensional vector, where the first (respectively, second) entry of the vector is the probability that  $S_i(t) = -1$  (respectively,  $S_i(t) = +1$ ). The

stochastic transitions in a synapse's strength in response to plasticity induction signals are represented by  $2 \times 2$  stochastic or transition matrices given by

$$\mathbb{M}^+ = \begin{pmatrix} 1-p & 0 \\ p & 1 \end{pmatrix} \text{ and } \mathbb{M}^- = \begin{pmatrix} 1 & p \\ 0 & 1-p \end{pmatrix}, \quad (2.4)$$

for  $\xi_i^\alpha = \pm 1$ , respectively. Because we average over the sequence of memories rather than consider any particular realisation, the relevant transition matrix for the storage of the non-tracked,  $\alpha > 0$  memories is

$$\mathbb{M} = \frac{1}{2} (\mathbb{M}^+ + \mathbb{M}^-). \quad (2.5)$$

As  $t \rightarrow \infty$ , any synapse's strength state asymptotes to the equilibrium distribution defined by the eigenvector associated with the unit eigenvalue of  $\mathbb{M}$ ,  $\mathbf{A} = \frac{1}{2} (1, 1)^T$ , where the superscript T denotes the transpose. The tracked memory  $\boldsymbol{\xi}^0$  is stored against the background of this equilibrium distribution at  $t = 0^-$  s. For synapses experiencing  $\xi_i^0 = +1$  (respectively,  $\xi_i^0 = -1$ ), their states at  $t = 0$  s are governed by the probability distribution  $\mathbb{M}^+ \mathbf{A}$  (respectively,  $\mathbb{M}^- \mathbf{A}$ ). Because we average over the initial memory  $\boldsymbol{\xi}^0$ , any synapse is initially in a state that is an equiprobable mixture of the two distributions  $\mathbb{M}^\pm \mathbf{A}$ .

At some future time  $t$ , the distribution of strengths of synapse  $i$  is given by  $e^{rt(\mathbb{M}-\mathbb{I})} \mathbb{M}^\pm \mathbf{A}$ , depending on the sign of  $\xi_i^0 = \pm 1$ , where  $\mathbb{I}$  is the identity matrix. Computing these two distributions explicitly and defining  $\tilde{S}_i(t) = \xi_i^0 S_i(t)$ , we obtain

$$\text{Prob} \left[ \tilde{S}_i(t) = \pm 1 \right] = \frac{1}{2} (1 \pm p e^{-prt}), \quad (2.6)$$

regardless of the sign of  $\xi_i^0$ , so that all  $N$  of the  $\tilde{S}_i(t)$  variables are identically



distributed. Because the tracked memory signal  $h(t)$  in Eq. (2.2) is just a (normalised) sum over these  $N$  tilded strength variables, it is therefore just a (normalised) sum over  $N$  identically distributed random variables. For balanced potentiation and depression processes,  $g_{\pm} = \frac{1}{2}$ , the mixture of states governed by the two distributions  $\mathbb{M}^{\pm}\mathbf{A}$  therefore collapses in terms of their contribution to the evolution of  $h(t)$ . This result is in fact quite general for synaptic plasticity processes that treat potentiation and depression completely symmetrically (Elliott, 2016b) and holds not only for  $\nu = 2$ , binary-strength stochastic updater synapses but also for their generalisation below to multi-state,  $\nu > 2$  synapses. We therefore need not consider the initial synaptic state immediately after the storage of  $\xi^0$  to be a mixture of the two distributions  $\mathbb{M}^{\pm}\mathbf{A}$  but can instead consider, say, only  $\mathbb{M}^+\mathbf{A}$  and work directly with the  $S_i(t)$  variables rather than their tilded forms  $\tilde{S}_i(t)$ , in effect simply setting  $\xi_i^0 \equiv +1$  for all synapses. This dramatic simplification is possible only for balanced and symmetric processes.

### 2.3 MFPTs for Binary Stochastic Updater Synapses

To overcome the shortcomings in the SNR approach discussed above, we consider the FPT for the perceptron’s activation to fall below threshold (Elliott, 2014). For any particular realisation of the sequence of memories  $\xi^{\alpha}$ ,  $h(t)$  will first fall (to or) below threshold at some time  $\tau$ . We average over all possible realisations of the memories to obtain the MFPT, and this defines the MFPT memory lifetime  $\tau_{\text{mfpt}}$ . The MFPT memory lifetime overcomes all the shortcomings of the SNR memory lifetime (Elliott, 2014).

To calculate the MFPT memory lifetime for stochastic updater synapses, we observe that the tracked memory signal  $h(t)$  is a (normalised) sum over  $N$

variables taking values of either  $+1$  or  $-1$  for binary-strength synapses. Its value is therefore uniquely determined by the number of these variables taking value  $+1$ : if  $j$  of them take value  $+1$ , then  $h(t) = [(+1)j + (-1)(N - j)]/N = 2j/N - 1$ . We may use this observation to compute the transition probability for the perceptron activation between successive memory storage steps (Elliott, 2014). Let  $h_\alpha$  denote the perceptron activation immediately after the storage of memory  $\alpha$ . The initial distribution  $h_0$  immediately after the storage of  $\xi^0$  is

$$\text{Prob} \left[ h_0 = \frac{2j}{N} - 1 \right] = {}^N C_j \left( \frac{1+p}{2} \right)^j \left( \frac{1-p}{2} \right)^{N-j}, \quad (2.7)$$

where  ${}^N C_j$  denotes a binomial coefficient. The transition probability between successive values of the activation is

$$\begin{aligned} \text{Prob} \left[ h_{\alpha+1} = \frac{2i}{N} - 1 \mid h_\alpha = \frac{2j}{N} - 1 \right] \\ = \sum_{k=0}^j {}^j C_k \left( \frac{p}{2} \right)^k \left( 1 - \frac{p}{2} \right)^{j-k} {}^{N-j} C_{i-j+k} \left( \frac{p}{2} \right)^{i-j+k} \left( 1 - \frac{p}{2} \right)^{N-i-k}, \end{aligned} \quad (2.8)$$

where the usual conventions regarding binomial coefficients apply. Using these transitions in perceptron activation, we derived an expression for the MFPT for the activation to fall (to or) below  $\theta$  from an initial activation  $h_0 > \theta$ . Letting  $\tau_{\text{mfpt}}(h_0)$  denote this MFPT, we have

$$\tau_{\text{mfpt}}(h_0) = \frac{1}{r} + \sum_{h>\theta} \tau_{\text{mfpt}}(h) \text{Prob}[h|h_0]. \quad (2.9)$$

We may move to a continuum limit for  $h$  when  $N$  is large enough, in excess of around 100. In this limit, the two distributions in Eqs. (2.7) and (2.8) may be replaced with Gaussian distributions with matched (conditional) means and

variances. In this limit Eq. (2.9) becomes

$$\tau_{\text{mfpt}}(h_0) = \frac{1}{r} + \int_{\theta}^{\infty} dh \tau_{\text{mfpt}}(h) K_0(h|h_0), \quad (2.10)$$

where  $K_0$  is the continuum Gaussian kernel corresponding to Eq. (2.8). Assuming that we can solve the equation for  $\tau_{\text{mfpt}}(h_0)$ , then  $\tau_{\text{mfpt}} = \langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$ , where  $\langle \cdot \rangle_{h_0}$  denotes an average over the initial distribution for  $h_0$ .

MFPT equations of the form in Eq. (2.10) are rarely soluble except for a handful of particular kernels. Previously we replaced the Gaussian kernel  $K_0$  with a formal expansion using the Dirac delta function  $\delta(x)$ ,

$$K_1(h|h_0) = \delta(h - h_0) + p h_0 \delta'(h - h_0) + \frac{1}{2} \left( p^2 h_0^2 + \frac{1 - q^2}{N} \right) \delta''(h - h_0), \quad (2.11)$$

where the primes denote differentiation with respect to the argument and we write  $q = 1 - p$ . This formal kernel has the same conditional mean and variance as Eq. (2.8). Eq. (2.10) then becomes the differential equation

$$-\frac{1}{r} = -p h_0 \frac{d}{dh_0} \tau_{\text{mfpt}}(h_0) + \frac{1}{2} \left( p^2 h_0^2 + \frac{1 - q^2}{N} \right) \frac{d^2}{dh_0^2} \tau_{\text{mfpt}}(h_0) \quad (2.12)$$

for  $h_0 > \theta$  and the solution  $\tau_{\text{mfpt}}(h_0) = 1/r$  for  $h_0 < \theta$ . For  $p$  small enough, Eq. (2.12) becomes

$$-\frac{1}{r} = -p h_0 \frac{d}{dh_0} \tau_{\text{mfpt}}(h_0) + \frac{p}{N} \frac{d^2}{dh_0^2} \tau_{\text{mfpt}}(h_0), \quad (2.13)$$

which is the equation governing the MFPT for the OU process. We defer discussion of the solutions of these equations to the next section.

### 3 Fokker-Planck Approach to FPT Distribution

The ability to work directly in the transitions in the perceptron's activation with each memory storage event and essentially ignore the details of the underlying transitions in all  $N$  synapses' strengths is critical to our derivation of MFPT results for binary-strength synapses. In this way, we need consider only transition matrices that are  $(N + 1) \times (N + 1)$  rather than  $2^N \times 2^N$  in size. For binary-strength synapses, this is possible because the number of synapses with (tilded) strength  $+1$  uniquely determines the perceptron's activation and, conversely, the perceptron's activation uniquely determines the number of such synapses. For  $\nu > 2$ , however, although the configuration of synaptic strengths uniquely determines the perceptron's activation, the perceptron's activation does not in general uniquely (even up to trivial permutation symmetries) determine the configuration of synaptic strengths. For example, for  $\nu = 3$  and with  $S_i(t) \in \{-1, 0, +1\}$ , any pair of synapses may have (tilded) strengths of  $+1$  and  $-1$  (in any order) or both may have strengths of  $0$ : both these strength configurations contribute identically to the perceptron's activation. This degeneracy only increases as  $\nu$  increases. To determine the statistics of the FPT process for the perceptron's activation for general  $\nu$ , we therefore cannot directly use the transitions in the perceptron's activation and we must find a different method.

#### 3.1 Fokker-Planck Formulation

Let  $P(h, t|h_0, t_0)$  denote the transition probability from an initial activation  $h_0$  at time  $t_0$  (here  $t_0 = 0$ ) to a final activation  $h$  at time  $t \geq t_0$ . The Fokker-Planck

or forward Chapman-Kolmogorov equation is then

$$\frac{1}{r} \frac{\partial}{\partial t} P(h, t|h_0, t_0) = -\frac{\partial}{\partial h} [A(h)P(h, t|h_0, t_0)] + \frac{1}{2} \frac{\partial^2}{\partial h^2} [B(h)P(h, t|h_0, t_0)], \quad (3.1)$$

while the adjoint or backward Chapman-Kolmogorov equation is

$$\frac{1}{r} \frac{\partial}{\partial t} P(h, t|h_0, t_0) = +A(h_0) \frac{\partial}{\partial h_0} P(h, t|h_0, t_0) + \frac{1}{2} B(h_0) \frac{\partial^2}{\partial h_0^2} P(h, t|h_0, t_0). \quad (3.2)$$

The functions  $A(x)$  and  $B(x)$  are the infinitesimal jump moments,

$$M_l(x) = \frac{1}{r} \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \int dy (y - x)^l P(y, t + \delta t|x, t), \quad (3.3)$$

with  $A(x) = M_1(x)$  and  $B(x) = M_2(x)$ . Because the transitions in the perceptor's activation are subject to potentially large jump processes in which the activation can jump across the firing threshold  $\theta$ , the use of the Fokker-Planck equation constitutes a diffusion limit approximation.

If we impose the absorbing boundary condition  $P(\theta, t|h_0, 0) = 0$  on the solution of the Fokker-Planck equation, the density  $G(h_0, t)$  for the system to escape from the interval  $(\theta, \infty)$  for the first time at time  $t$  from an initial state  $h_0 > \theta$  (at time  $t_0 = 0$ ) in this interval is given by

$$G(h_0, t) = -\frac{\partial}{\partial t} \int_{\theta}^{\infty} dh P(h, t|h_0, 0). \quad (3.4)$$

Using the backward equation we obtain

$$\frac{1}{r} \frac{\partial}{\partial t} G(h_0, t) = A(h_0) \frac{\partial}{\partial h_0} G(h_0, t) + \frac{1}{2} B(h_0) \frac{\partial^2}{\partial h_0^2} G(h_0, t). \quad (3.5)$$

Since  $\tau_{\text{mfpt}}(h_0) = \int_0^\infty dt t G(h_0, t)$ , we have

$$-\frac{1}{r} = A(h_0) \frac{d}{dh_0} \tau_{\text{mfpt}}(h_0) + \frac{1}{2} B(h_0) \frac{d^2}{dh_0^2} \tau_{\text{mfpt}}(h_0) \quad (3.6)$$

for  $h_0 > \theta$ . Comparing Eq. (3.6) to Eq. (2.12) for  $\nu = 2$ , we see that they are structurally identical, indicating that the use of the kernel  $K_1(h|h_0)$  constitutes a diffusion approximation in which jump processes have been neglected. In addition to the MFPT, we can also obtain all the FPT statistics from Eq. (3.5). Laplace transforming this equation, with  $\widehat{G}(h_0, s)$  denoting the Laplace transform of  $G(h_0, t)$  with transformed variable  $s$ , and using the fact that  $G(h_0, 0) \equiv 0$  for  $h_0 > \theta$ , we have

$$\frac{s}{r} \widehat{G}(h_0, s) = A(h_0) \frac{d}{dh_0} \widehat{G}(h_0, s) + \frac{1}{2} B(h_0) \frac{d^2}{dh_0^2} \widehat{G}(h_0, s). \quad (3.7)$$

This equation is solved subject to the two boundary conditions  $\widehat{G}(\theta, s) = 1$  and  $\widehat{G}(\theta^*, s) = 1$  and then we take the limit  $\theta^* \rightarrow \infty$  in order to remove the influence of the second boundary at  $\theta^*$ . As the moment generating function (MGF) of a density is just its Laplace transform up to the sign of  $s$ ,  $\widehat{G}(h_0, -s)$  is just the MGF of the FPT distribution. To be able to determine the FPT distribution and all its moments, we need the jump moments  $A(x)$  and  $B(x)$ .

## 3.2 Determination of Jump Moments

With  $\mu(t) = \mathbf{E}[h(t)]$  as before and defining  $\varphi(t) = \mathbf{E}[h(t)^2]$ , we can obtain the evolution of these moments from the Fokker-Planck equation using

$$\frac{1}{r} \frac{d\mu(t)}{dt} = \mathbf{E}[A(h)], \quad (3.8a)$$

$$\frac{1}{r} \frac{d\varphi(t)}{dt} = \mathbf{E}[B(h)] + 2\mathbf{E}[hA(h)]. \quad (3.8b)$$

If we can derive these two evolution equations via another method, then we can deduce the form of the jump moments  $A(x)$  and  $B(x)$ .

Since  $h(t) = \frac{1}{N} \sum_{i=1}^N \tilde{S}_i(t)$ , we have that

$$\mu(t) = \mathbf{E}[\tilde{S}_1(t)], \quad (3.9a)$$

$$\varphi(t) = \frac{1}{N} \mathbf{E}[\tilde{S}_1(t)^2] + \left(1 - \frac{1}{N}\right) \mathbf{E}[\tilde{S}_1(t)\tilde{S}_2(t)], \quad (3.9b)$$

where we have used the fact that all the  $\tilde{S}_i(t)$  are identically distributed to single out any particular pair of synapses, here just  $i = 1$  and  $i = 2$ . We can also simply set  $\xi_i^0 \equiv 1 \forall i$  and so compute the expectation values using only strength rather than tilded strength variables. It is a property of the synaptic plasticity models considered below that for any choice of  $\nu$ ,  $\mathbf{E}[\tilde{S}_1(t)^2] \equiv \mathbf{E}[S_1(t)^2]$  in Eq. (3.9b) is independent of time. We denote by  $\boldsymbol{\Omega}$  the  $\nu$ -dimensional vector of possible synaptic strengths available to a multistate synapse, so that  $S_i(t) \in \{\Omega_1, \dots, \Omega_\nu\}$ . These components are ordered weakest to strongest. Then we will show below that

$$\mathbf{E}[S_1(t)^2] = \frac{1}{\nu} \sum_{i=1}^{\nu} \Omega_i^2 \equiv \langle \boldsymbol{\Omega}^2 \rangle, \quad (3.10)$$

where we use the final form as convenient shorthand notation.

The evolution of the quantities on the right-hand sides of Eq. (3.9) involve only the dynamics of a single synapse via the mean  $\mu(t) = \mathbf{E}[S_1(t)]$  and the joint dynamics of a pair of synapses via the correlation function  $\psi(t) = \mathbf{E}[S_1(t)S_2(t)]$ . For general  $\nu \geq 2$ , let the general  $\nu \times \nu$  transition matrix for a synapse's strength be  $\mathbb{M}$ , which for  $\nu > 2$  will generalise the particular form of  $\mathbb{M}$  in Eq. (2.5). Let  $\mathbf{P}_1(t)$  denote the probability distribution of any single synapse's strength and let  $\mathbf{P}_2(t)$  denote the joint probability distribution of any pair of

synapses' strengths. Since  $\mu(t) \equiv \boldsymbol{\Omega}^T \mathbf{P}_1(t)$  and  $\psi(t) \equiv (\boldsymbol{\Omega}^T \otimes \boldsymbol{\Omega}^T) \mathbf{P}_2(t)$ , we obtain

$$\frac{1}{r} \frac{d\mu(t)}{dt} = \boldsymbol{\Omega}^T \mathbb{M} \mathbf{P}_1(t) - \mu(t), \quad (3.11a)$$

$$\frac{1}{r} \frac{d\psi(t)}{dt} = (\boldsymbol{\Omega}^T \otimes \boldsymbol{\Omega}^T) (\mathbb{M} \otimes \mathbb{M}) \mathbf{P}_2(t) - \psi(t). \quad (3.11b)$$

which follow directly from the evolution equations for  $\mathbf{P}_1(t)$  and  $\mathbf{P}_2(t)$ .

For  $\nu = 2$ ,  $\boldsymbol{\Omega} = (-1, +1)^T$  is a left eigenvector of  $\mathbb{M}$  in Eq. (2.5). For  $\nu = 2$ , the right-hand sides of Eq. (3.11) therefore simplify, generating a closed system of equations for  $\mu(t)$  and  $\psi(t)$  from which  $A(x)$  and  $B(x)$  follow. For general  $\nu \geq 2$ , unless  $\boldsymbol{\Omega}$  is a left eigenvector of  $\mathbb{M}$ , the right-hand sides of Eq. (3.11) do not close. To make progress, we must assume that  $\boldsymbol{\Omega}$  is a left eigenvector of  $\mathbb{M}$ ,

$$\boldsymbol{\Omega}^T \mathbb{M} = \lambda_\Omega \boldsymbol{\Omega}^T, \quad (3.12)$$

where  $\lambda_\Omega$  is the eigenvalue of  $\mathbb{M}$  associated with its left eigenvector  $\boldsymbol{\Omega}$ . In the following section we construct models of synaptic plasticity satisfying this eigenvector requirement.

With the exact eigenvector requirement on  $\mathbb{M}$  and  $\boldsymbol{\Omega}$ , we obtain

$$\frac{1}{r} \frac{d\mu(t)}{dt} = -(1 - \lambda_\Omega) \mu(t), \quad (3.13a)$$

$$\frac{1}{r} \frac{d\varphi(t)}{dt} = -(1 - \lambda_\Omega^2) \left[ \varphi(t) - \frac{1}{N} \langle \boldsymbol{\Omega}^2 \rangle \right], \quad (3.13b)$$

with explicit solutions

$$\mu(t) = \mu(0) e^{-r(1-\lambda_\Omega)t}, \quad (3.14a)$$

$$\varphi(t) = \frac{1}{N} \langle \boldsymbol{\Omega}^2 \rangle + \left( 1 - \frac{1}{N} \right) \mu(0)^2 e^{-r(1-\lambda_\Omega^2)t}, \quad (3.14b)$$



where the initial mean memory signal  $\mu(0)$  immediately after the storage of  $\xi^0$  depends on the details of the model of synaptic plasticity. We will write  $\mu_0 = \mu(0)$  throughout for convenience. By comparing Eqs. (3.8) and (3.13) we can read off the expectation values of the jump moments,

$$\mathbb{E}[A(h)] = -(1 - \lambda_\Omega) \mu(t), \quad (3.15a)$$

$$\mathbb{E}[B(h)] = -(1 - \lambda_\Omega^2) \left[ \varphi(t) - \frac{1}{N} \langle \Omega^2 \rangle \right] - 2 \mathbb{E}[hA(h)], \quad (3.15b)$$

from which we finally deduce that

$$A(x) = -(1 - \lambda_\Omega) x, \quad (3.16a)$$

$$B(x) = (1 - \lambda_\Omega)^2 x^2 + \frac{1 - \lambda_\Omega^2}{N} \langle \Omega^2 \rangle. \quad (3.16b)$$

For  $\nu = 2$ ,  $\lambda_\Omega = 1 - p$  and  $\langle \Omega^2 \rangle = 1$ , so these jump moments reduce identically to the coefficients of the MFPT equation in Eq. (2.12) for binary synapses.

Although the eigenvector requirement in Eq. (3.12) may appear to be very strong, in general even if  $\Omega$  is not an exact left eigenvector of  $\mathbb{M}$  but is sufficiently close to one, say  $\mathbf{e}$ , then we would expect to obtain a good approximation by using  $\Omega$  as an approximate eigenvector of  $\mathbb{M}$ . If a symmetric  $\mathbb{M}$  has a complete set of orthonormal eigenvectors  $\hat{\mathbf{e}}^m$ , with say  $\hat{\mathbf{e}}^1 \propto \mathbf{e}$  and  $\Omega$  is close to  $\mathbf{e}$ , then we can write

$$\Omega = \left( \frac{\mathbf{e}^T \Omega}{\mathbf{e}^T \mathbf{e}} \right) \mathbf{e} + \sum_{m=2}^{\nu} \left( \hat{\mathbf{e}}^{mT} \Omega \right) \hat{\mathbf{e}}^{mT}. \quad (3.17)$$

If the contribution from the first term involving  $\mathbf{e}$  dominates the contributions from the other eigenvectors, then we can write  $\Omega \approx (\mathbf{e}^T \Omega / \mathbf{e}^T \mathbf{e}) \mathbf{e}$  with  $\Omega^T \mathbb{M} \approx \lambda_e \Omega^T$ , where  $\lambda_e$  is the eigenvalue associated with the closest eigenvector  $\mathbf{e}$ .

In general, then, provided that  $\mathbb{M}$  has an eigenvector close enough to the actual vector of possible synaptic strengths  $\mathbf{\Omega}$ , we would expect to obtain good quantitative agreement between our analytical results below, where we assume that  $\mathbf{\Omega}$  is an exact left eigenvector of  $\mathbb{M}$ , and numerical or simulation-based results, for which we may relax this assumption.

### 3.3 Extraction of FPT Distribution

Although it is possible to obtain the FPT distribution for the full forms of  $A(x)$  and  $B(x)$  in Eq. (3.16), we may consider a simpler form for  $B(x)$  and obtain extremely good agreement with the full results. Specifically, we write  $B(x)$  only to first order in  $1 - \lambda_\Omega$ , so that

$$B(x) \approx \frac{2(1 - \lambda_\Omega)}{N} \langle \mathbf{\Omega}^2 \rangle. \quad (3.18)$$

This is equivalent to considering an OU limit; for  $\nu = 2$ ,  $\lambda_\Omega = 1 - p$ , so this is just the small  $p$  limit. Below we frequently refer to dynamics for large enough  $N$ . By this we only mean  $N$  large compared, say, to 100, but not so large that the simpler form for  $B(x)$  is invalidated. Biologically, large  $N$  means  $N$  of order  $10^4$  or  $10^5$ ; larger values are irrelevant. In this OU limit, Eq. (3.7) becomes

$$\frac{s}{r(1 - \lambda_\Omega)} \widehat{G}(h_0, s) = -h_0 \frac{d}{dh_0} \widehat{G}(h_0, s) + \frac{\langle \mathbf{\Omega}^2 \rangle}{N} \frac{d^2}{dh_0^2} \widehat{G}(h_0, s). \quad (3.19)$$

The parameter  $\nu$  enters in two ways. First, through a rescaling of the rate  $r$  via  $1 - \lambda_\Omega$  as  $\lambda_\Omega$  will depend on  $\nu$ . Both  $r$  and  $1 - \lambda_\Omega$  rescale the time  $t$ ; we define  $\rho = r(1 - \lambda_\Omega)$  and write  $s' = s/\rho$ . Second, through a rescaling of  $N$  via the quantity  $\langle \mathbf{\Omega}^2 \rangle$  to generate an effective number of synapses  $N' = N/\langle \mathbf{\Omega}^2 \rangle$ .

We can then rewrite Eq. (3.19) as

$$s' \widehat{G}(h_0, s') = -h_0 \frac{d}{dh_0} \widehat{G}(h_0, s') + \frac{1}{N'} \frac{d^2}{dh_0^2} \widehat{G}(h_0, s'). \quad (3.20)$$

The solution of this equation, subject to the boundary conditions at  $h_0 = \theta$  and  $h_0 = \theta^* > \theta$  and taking  $\theta^* \rightarrow \infty$ , is

$$\widehat{G}(h_0, s') = \frac{H_{-s'}(h_0 \sqrt{N'/2})}{H_{-s'}(\theta \sqrt{N'/2})}, \quad (3.21)$$

where  $H_{-a}(y)$  is a Hermite polynomial of possibly non-integer order.

In general, we cannot invert Eq. (3.21), although we can expand as a power series in  $s'$  to obtain the moments of the FPT distribution. However, for the particular case of  $\theta = 0$  we can explicitly write down the solution of the original Fokker-Planck equation in the OU limit satisfying the absorbing boundary condition  $P(\theta, t|h_0, 0) = 0$ . If  $f(h, h_0; t)$  is the standard OU solution of the Fokker-Planck equation in the absence of an absorbing boundary, then a standard image construction gives  $P(h, t|h_0, 0) = f(h, +h_0; t) - f(h, -h_0; t)$  with  $h_0 > 0$  as the solution for  $h \geq 0$  satisfying the boundary condition. From Eq. (3.4) we then obtain

$$G(h_0, t) = \rho \sqrt{\frac{2N'}{\pi}} \frac{h_0 e^{-\rho t}}{(1 - e^{-2\rho t})^{3/2}} \exp \left[ -\frac{N'}{2} \frac{(h_0 e^{-\rho t})^2}{(1 - e^{-2\rho t})} \right] \quad (3.22)$$

as an explicit form of the FPT distribution for  $\theta = 0$ .

Expanding Eq. (3.21) to second order in  $s'$  we obtain expressions for the lowest order statistics of the FPT distribution. Defining  $\tau(x)$  via

$$\rho \tau(x) = \frac{1}{2} \left[ \pi \operatorname{erfi}(x \sqrt{N'/2}) - N' x^2 {}_2F_2(\{1, 1\}, \{3/2, 2\}, N' x^2/2) \right], \quad (3.23)$$

where  $\operatorname{erfi}(x)$  is the imaginary error function and  ${}_2F_2$  is a hypergeometric function, the MFPT is

$$\tau_{\text{mfpt}}(h_0) = \tau(h_0) - \tau(\theta). \quad (3.24)$$

We have derived this form before for  $\nu = 2$  (Elliott, 2014), but Eq. (3.24) generalises the result to  $\nu \geq 2$ . For the mean squared FPT, and thence the variance denoted by  $\sigma_{\text{fpt}}^2(h_0)$ , we can obtain exact results, but they are in general very messy and we do not reproduce them here. However, for  $h_0\sqrt{N'}$  large enough, the results for both  $\tau_{\text{mfpt}}(h_0)$  and  $\sigma_{\text{fpt}}^2(h_0)$  simplify dramatically. They differ qualitatively between  $\theta > 0$  and  $\theta = 0$  because  $\mu(t) \rightarrow 0$  as  $t \rightarrow \infty$ , so  $\theta = 0$  is a special case tuned precisely to match the asymptotic mean memory signal (Elliott, 2014). For  $h_0\sqrt{N'}$  large enough and for  $\theta > 0$ , we obtain

$$\rho \tau_{\text{mfpt}}(h_0) \sim \log_e(h_0/\theta), \quad (3.25a)$$

$$\rho^2 \sigma_{\text{fpt}}^2(h_0) \sim 0, \quad (3.25b)$$

while for  $\theta = 0$ ,

$$\rho \tau_{\text{mfpt}}(h_0) \sim \frac{1}{2} [\gamma + \log_e 2 + \log_e(N'h_0^2)], \quad (3.26a)$$

$$\rho^2 \sigma_{\text{fpt}}^2(h_0) \sim \frac{\pi^2}{8}, \quad (3.26b)$$

where  $\gamma \approx 0.577$  is Euler's constant. For  $\theta = 0$ , the behaviour of the MFPT is logarithmic in  $N$  for large enough  $N$ , but for  $\theta > 0$  the  $N$ -dependence drops out. The variance in the FPT for  $\theta > 0$  approaches zero as  $N$  increases; for  $\theta = 0$ , it approaches a non-zero constant. Eqs. (3.25a) and (3.26a) generalise our previous results for  $\nu = 2$ , but we have not derived results for  $\sigma_{\text{fpt}}^2$  before.

The results above are averaged over all realisations of the later memories  $\xi^\alpha$ ,  $\alpha > 0$ , but they have a fixed initial value  $h_0 > \theta$ . For the FPT distribution also

averaged over the tracked memory  $\xi^0$ , we must evaluate  $\langle \widehat{G}(h_0, s') \rangle_{h_0}$ , where we average over the initial distribution of  $h_0$  with  $h_0 > \theta$ . From Eq. (3.14) this distribution has mean  $\mu_0$  and variance  $\sigma(0)^2 = [\langle \Omega^2 \rangle - \mu_0^2] / N$ , and for  $N$  large enough the distribution is Gaussian. For the models of synaptic plasticity that we consider below, we can typically assume that  $\mu_0^2 \ll \langle \Omega^2 \rangle$ . For  $\nu = 2$ , for example, this is just the requirement that  $p^2 \ll 1$ . Thus, it is convenient although not necessary to make the approximation  $\sigma(0)^2 \approx 1/N'$ . We write

$$\begin{aligned} \langle \widehat{G}(h_0, s') \rangle_{h_0} &= \sqrt{\frac{N'}{2\pi}} \int_{\theta}^{\infty} dx \frac{H_{-s'}(x\sqrt{N'/2})}{H_{-s'}(\theta\sqrt{N'/2})} e^{-N'[x-\mu_0]^2/2} \\ &= \sqrt{\frac{1}{2\pi}} \int_{\theta'}^{\infty} dy \frac{H_{-s'}(y/\sqrt{2})}{H_{-s'}(\theta'/\sqrt{2})} e^{-(y-\mu')^2/2}, \end{aligned} \quad (3.27)$$

where we have defined the scaled forms  $\mu' = \mu_0\sqrt{N'}$  and  $\theta' = \theta\sqrt{N'}$ . For  $\theta > 0$ , it does not appear to be possible to evaluate this integral in terms of known functions, so we must resort to numerical or approximation methods in this case. For  $\theta = 0$ , the integral can be evaluated, but we may also explicitly average Eq. (3.22) over the distribution of  $h_0$  with  $h_0 > 0$ , giving

$$\begin{aligned} \langle G(h_0, t) \rangle_{h_0} &= \frac{\rho}{\pi} e^{-\rho t} \left\{ \frac{e^{-\frac{1}{2}\mu'^2}}{\sqrt{1 - e^{-2\rho t}}} \right. \\ &\quad \left. + \sqrt{\frac{\pi}{2}} \mu' e^{-\frac{1}{2}(\mu' e^{-\rho t})^2} \left[ 1 + \operatorname{erf} \left( \mu' \sqrt{\frac{1 - e^{-2\rho t}}{2}} \right) \right] \right\}, \end{aligned} \quad (3.28)$$

where erf is the error function. We may then obtain  $\tau_{\text{mfpt}} = \langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  and the variance in the FTP, which we denote by  $\sigma_{\text{fpt}}^2$  or  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}^2$ , where we use this latter as a convenient shorthand for  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}^2 = \langle t^2(h_0) \rangle_{h_0} - \langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}^2$ , where  $t^2(h_0)$  is the second moment of the FPT distribution for a definite value of  $h_0$ . The full results for  $\theta = 0$  are fairly simple but unenlightening. However, for either small  $\mu'$  or large  $\mu'$ , retaining just the first few terms in the expansions

they reduce to

$$\rho \tau_{\text{mfpt}} \sim \begin{cases} \frac{1}{2} \log_e 2 + \frac{1}{\sqrt{2\pi}} \mu' + \frac{1}{8} \mu'^2 & \text{for } \mu' \text{ small} \\ \frac{1}{2} (\gamma + \log_e 2 + \log_e \mu'^2) & \text{for } \mu' \text{ large} \end{cases}, \quad (3.29)$$

and

$$\rho^2 \sigma_{\text{fpt}}^2 \sim \begin{cases} \frac{1}{24} (\pi^2 + 6 \log_e^2 2) + \frac{1}{\sqrt{2\pi}} (2 - \log_e 2) \mu' & \text{for } \mu' \text{ small} \\ \frac{\pi^2}{8} & \text{for } \mu' \text{ large} \end{cases}. \quad (3.30)$$

These statistics averaged over  $h_0$  for large  $\mu'$  coincide with those for a definite value of  $h_0$  for large enough  $h_0 \sqrt{N'}$  in Eq. (3.26) when we replace  $h_0$  by  $\mu_0$ . This reflects the fact that when  $N$  is large enough,  $\langle G(h_0, t) \rangle_{h_0}$  can be replaced by its mean field form  $G(\langle h_0 \rangle_{h_0}, t) \equiv G(\mu_0, t)$ .

## 4 Simple Synapses Satisfying an Eigenvector Constraint

We now construct two models of synaptic plasticity satisfying the requirement  $\mathbf{\Omega}^T \mathbf{M} = \lambda_{\mathbf{\Omega}} \mathbf{\Omega}^T$ . In the first, we pick  $\mathbf{\Omega}$  to be an eigenvector of  $\mathbf{M}$  where  $\mathbf{M}$  is a generalised form of the transition matrix given in Eq. (2.5) for  $\nu = 2$ . In the second, we modify  $\mathbf{M}$  so that it has as an eigenvector  $\mathbf{\Omega}$  an arrangement of synaptic strengths that is uniformly spaced.

### 4.1 Modifying $\mathbf{\Omega}$

The simplest generalisation of the  $\nu = 2$  stochastic updater synapse is one that expresses plasticity with fixed probability  $p$  regardless of its strength state (unless saturated at its upper or lower value). The  $2 \times 2$  matrices  $\mathbf{M}^{\pm}$  in

Eq. (2.4) then become the  $\nu \times \nu$  matrices

$$\mathbb{M}^+ = \text{diag}\{\underbrace{q, \dots, q}_{\nu-1}, 1\} + \text{diag}_l\{\underbrace{p, \dots, p}_{\nu-1}\}, \quad (4.1a)$$

$$\mathbb{M}^- = \text{diag}\{1, \underbrace{q, \dots, q}_{\nu-1}\} + \text{diag}_u\{\underbrace{p, \dots, p}_{\nu-1}\}, \quad (4.1b)$$

where  $\text{diag}_u$  and  $\text{diag}_l$  refer to the upper and lower diagonals, respectively. The superposed matrix  $\mathbb{M} = \frac{1}{2}(\mathbb{M}^+ + \mathbb{M}^-)$  can then be written as

$$\mathbb{M} = \mathbb{I} + p \mathbb{C}, \quad (4.2)$$

where the matrix  $\mathbb{C}$  is

$$\mathbb{C} = \frac{1}{2}(\text{diag}\{-1, \underbrace{-2, \dots, -2}_{\nu-2}, -1\} + \text{diag}_u\{\underbrace{+1, \dots, +1}_{\nu-1}\} + \text{diag}_l\{\underbrace{+1, \dots, +1}_{\nu-1}\}). \quad (4.3)$$

We use the symbol “ $\mathbb{C}$ ” (for constant) because its defining off-diagonal elements are all the same constant. The spectrum of  $\mathbb{C}$  is standard (e.g., Elliott, 2016a), so we just state its eigenvalues,

$$\lambda_m = \cos \frac{m \pi}{\nu} - 1, \quad m = 0, \dots, \nu - 1, \quad (4.4)$$

and its orthonormal eigenvectors  $\widehat{\mathbf{e}}^m$  with components  $\widehat{e}_i^m$ ,  $i = 1, \dots, \nu$ ,

$$\widehat{e}_i^m = \begin{cases} \sqrt{\frac{1}{\nu}} & \text{for } m = 0 \\ \sqrt{\frac{2}{\nu}} \cos \left[ \frac{m \pi}{2 \nu} (2i - 1) \right] & \text{for } m = 1, \dots, \nu - 1 \end{cases}. \quad (4.5)$$

These eigenvectors of  $\mathbb{C}$  are of course also eigenvectors of  $\mathbb{M}$ , with eigenvalues  $1 + p \lambda_m$ , and as  $\mathbb{M}$  is symmetric, its left and right eigenvectors are identical. The eigenvector with eigenvalue unity, corresponding to  $m = 0$ , is the equilib-

rium eigenvector. Defining the vector  $\boldsymbol{\nu} = (1, \dots, 1)^T$ , a  $\nu$ -dimensional vector, the equilibrium distribution is just  $\mathbf{A} = \frac{1}{\nu} \boldsymbol{\nu}$ .

For a multistate synapse it is standard to consider a uniformly spaced sequence of synaptic strengths. We define the vector  $\mathcal{L}$  (for linear) to have components

$$\mathcal{L}_i = -1 + 2 \frac{i-1}{\nu-1}, \quad (4.6)$$

which are uniformly spaced in  $[-1, +1]$ . Except for  $\nu = 2$  and  $\nu = 3$ , however,  $\mathcal{L}$  is not an eigenvector of  $\mathbb{M}$ . We require instead an eigenvector of  $\mathbb{M}$  whose components monotonically increase (with a change in sign if necessary) and are antisymmetrically arranged around zero. The requisite eigenvector is  $\widehat{\mathbf{e}}^1$  and we define the vector  $\mathcal{S} \propto \widehat{\mathbf{e}}^1$  (for sigmoidal or sinusoidal) to have components

$$\mathcal{S}_i = -\frac{\cos \frac{\pi}{2\nu}(2i-1)}{\cos \frac{\pi}{2\nu}}. \quad (4.7)$$

For  $\nu = 2$  and  $\nu = 3$ ,  $\mathcal{L} \equiv \mathcal{S}$ . Viewed from the middle of the strength range, for  $\nu > 3$  this arrangement is sinusoidal, effecting saturation-like dynamics at the lower and upper ends of its range. In many respects, such dynamics may be considered to be more desirable than uniformly spaced strengths.

For this standard form of  $\mathbb{M}$ , we therefore set  $\boldsymbol{\Omega} = \mathcal{S}$  and we have  $\lambda_{\Omega} = 1 - p(1 - \cos \frac{\pi}{\nu})$ . We find that  $\langle \mathcal{S}^2 \rangle = \frac{1}{2} \sec^2 \frac{\pi}{2\nu}$ . To compute the initial signal  $\mu_0$ , we require  $\mathbb{M}^+ \mathbf{A} = \frac{1}{\nu} (1 - p, 1, \dots, 1, 1 + p)^T$ , in which only the first and last components are modified compared to  $\mathbf{A}$ . Since  $\mathcal{S}^T \mathbf{A} \equiv 0$ , we have the initial signal  $\mu_0 = \mathcal{S}^T \mathbb{M}^+ \mathbf{A} = 2p/\nu$ . We note that because of the structure of  $\mathbb{M}^+ \mathbf{A}$ , the initial signal is  $2p/\nu$  whether the strength vector is  $\mathcal{S}$  or  $\mathcal{L}$ .

We must verify that  $\mathbb{E}[S_1(t)^2] \equiv \langle \boldsymbol{\Omega}^2 \rangle$ , independent of  $t$ . First, we write  $\mathbb{M}^+ \mathbf{A} = \mathbf{A} + \mathbf{B}$ , where  $\mathbf{B}$  is antisymmetric, or  $B_i = -B_{\nu+1-i}$  for any  $i$ . We



then observe that because  $\mathbb{M} = \frac{1}{2}(\mathbb{M}^+ + \mathbb{M}^-)$ , if the vector  $\mathbf{B}$  is antisymmetric, then so is the vector  $\mathbb{M}\mathbf{B}$ . Thus, the distribution of any synapse's strengths at time  $t$  can be written as  $\mathbf{A} + \mathbf{B}(t)$  with  $\mathbf{B}(t)$  antisymmetric. Then

$$\mathbb{E}[S_1(t)^2] = \sum_{i=1}^{\nu} A_i \Omega_i^2 + \sum_{i=1}^{\nu} B_i(t) \Omega_i^2 \equiv \langle \Omega^2 \rangle + \sum_{i=1}^{\nu} B_i(t) \Omega_i^2 = \langle \Omega^2 \rangle. \quad (4.8)$$

This confirms the stated result in Eq. (3.10) for this model.

Finally, we consider relaxing the requirement that  $\Omega = \mathcal{S}$  by examining the overlap between the eigenvectors of  $\mathbb{M}$  and the linear strength vector  $\mathcal{L}$ . Specifically, we compute the overlap  $\mathcal{L}^T \hat{\mathbf{e}}^m$  for  $m \neq 1$  relative to that for  $m = 1$ , corresponding to  $\hat{\mathbf{e}}^1 \propto \mathcal{S}$ . We obtain

$$\frac{\mathcal{L}^T \hat{\mathbf{e}}^m}{\mathcal{L}^T \hat{\mathbf{e}}^1} = \left( \frac{1 - (-1)^m}{2} \right) \left( \frac{\tan \frac{\pi}{2\nu}}{\tan \frac{m\pi}{2\nu}} \right) \left( \frac{\sin \frac{\pi}{2\nu}}{\sin \frac{m\pi}{2\nu}} \right) \rightarrow \left( \frac{1 - (-1)^m}{2} \right) \frac{1}{m^2},$$

where the limit is taken for  $\nu \rightarrow \infty$ . In this limit the maximum relative overlap occurs for  $m = 3$ , giving a factor of  $1/9$ , or about 11%, and the total relative overlap over all  $m \neq 1$  gives  $\frac{\pi^2}{8} - 1$ , or about 23%. These relatively small contributions to the expansion in Eq. (3.17) suggest that using  $\Omega = \mathcal{L}$  instead of the exact eigenvector  $\mathcal{S}$  should incur an error of at most around 25%. In fact, we find that the error is typically much smaller.

## 4.2 Modifying $\mathbb{M}$

Above we retained the standard form of  $\mathbb{M}$  and modified  $\Omega$ , setting  $\Omega = \mathcal{S}$ . Now we consider retaining  $\Omega = \mathcal{L}$  and instead modifying  $\mathbb{M}$ . We write  $\mathbb{M}^\pm = \mathbb{I} + p\mathbb{Q}^\pm$ , where  $p$  is the parameter that controls the overall probability that any given change in synaptic strength is expressed. Then  $\mathbb{M} = \frac{1}{2}(\mathbb{M}^+ + \mathbb{M}^-) = \mathbb{I} + p\mathbb{Q}$ , where we write  $\mathbb{Q} = \frac{1}{2}(\mathbb{Q}^+ + \mathbb{Q}^-)$ . We require a matrix  $\mathbb{Q}$  that treats

potentiation and depression processes symmetrically; that has  $\mathbf{A} = \frac{1}{\nu} \boldsymbol{\nu}$  as a right eigenvector so that the equilibrium distribution of synaptic strengths is uniform; and that also has the vector  $\boldsymbol{\mathcal{L}}$  as a left eigenvector. Writing

$$\mathbb{Q}^+ = -\text{diag}\{a_1, \dots, a_\nu\} + \text{diag}_l\{a_1, \dots, a_{\nu-1}\}, \quad (4.9a)$$

$$\mathbb{Q}^- = -\text{diag}\{a_\nu, \dots, a_1\} + \text{diag}_u\{a_{\nu-1}, \dots, a_1\}, \quad (4.9b)$$

where  $a_\nu = 0$ , the structures of  $\mathbb{Q}^\pm$  ensure that potentiation and depression are treated symmetrically. The vector  $\mathbf{A}$  is always a left eigenvector of  $\mathbb{Q}$  and the easiest way to ensure that it is also a right eigenvector is for  $\mathbb{Q}$  to be symmetric, requiring  $a_i = a_{\nu-i}$ . The simplest non-constant form for  $a_i$  is therefore  $a_i \propto (\nu - i)i + c$  for some constant  $c$ , and  $a_\nu = 0$  forces  $c = 0$ . By setting

$$a_i = \frac{(\nu - i)i}{\nu - 1} \quad (4.10)$$

for  $i = 0, \dots, \nu$ , we may confirm that  $\boldsymbol{\mathcal{L}}$  is also an eigenvector of  $\mathbb{Q}$ . The overall normalisation is chosen so that  $\mathbb{Q} \equiv \mathbb{C}$  for  $\nu = 2$  and  $\nu = 3$  since  $\boldsymbol{\mathcal{L}} \equiv \boldsymbol{\mathcal{S}}$  for these special cases. The off-diagonal elements of  $\mathbb{Q}$  are arranged quadratically, hence our use of the symbol “ $\mathbb{Q}$ ” (for quadratic), in contrast to  $\mathbb{C}$  above.

Let the eigenvalues of  $\mathbb{M}$  be  $\lambda_m$ ,  $m = 1, \dots, \nu$ , with associated un-normalised (but orthogonal) eigenvectors  $\mathbf{e}^m$  having components  $e_i^m$ . We write  $f_m(z) = \sum_{i=1}^{\nu} e_i^m z^i$  for the generating function of these components and then derive an equation for  $f_m(z)$ . Because of the natural boundaries at  $i = 0$  and  $i = \nu$  at which  $a_i = 0$ , we may extend the sum defining  $f_m(z)$  over all  $i$ . After lengthy but straightforward algebra, the eigenvalue equation  $2(\nu - 1)\mathbb{Q} \mathbf{e}^m = \bar{\lambda}_m \mathbf{e}^m$  for the eigenvalues  $\bar{\lambda}_m$  of the scaled matrix  $2(\nu - 1)\mathbb{Q}$  can then be written as the

differential equation

$$\begin{aligned} \bar{\lambda}_m f_m(z) &= \nu [(1-z)^2 f'_m(z) - z^{-1}(1-z)f_m(z)] \\ &\quad - [z(1-z)^2 f''_m(z) - (1-z^2)f'_m(z) + z^{-1}(1-z)f_m(z)], \end{aligned} \quad (4.11)$$

where primes denote differentiation with respect to  $z$ . Demanding a terminating power series solution determines the eigenvalues as  $\bar{\lambda}_m = -m(m-1)$  and hence

$$\lambda_m = 1 - p \frac{m(m-1)}{2(\nu-1)}, \quad (4.12)$$

and the corresponding power series solution is

$$f_m(z) = (\nu^{-1} C_{m-1})^{-1} z (1-z)^{m-1} \sum_{i=0}^{\nu-m} m^{-1+i} C_{m-1}^{\nu-1-i} C_{m-1} z^i. \quad (4.13)$$

Clearly  $f_1(z) = \sum_{i=1}^{\nu} z^i$  so that  $\mathbf{e}^1 \propto \mathbf{A}$ , and explicitly evaluating  $f_2(z)$ , we find  $f_2(z) = -\sum_{i=1}^{\nu} \mathcal{L}_i z^i$ , or  $\mathbf{e}^2 = -\mathcal{L}$ , so that  $\mathcal{L}$  is indeed an eigenvector of  $\mathbb{M}$ , as advertised. Its eigenvalue is  $1 - p/(\nu-1)$ .

Because  $\mathbb{M}$  is a stochastic matrix, its elements must be non-negative. The diagonal elements of  $\mathbb{M}$  take the form  $1 - \frac{1}{2} p (a_i + a_{\nu+1-i}) = 1 - \frac{1}{2} p (a_i + a_{i-1})$  for  $i = 1, \dots, \nu$ . We therefore require

$$\nu \leq \begin{cases} \frac{1}{p} \left( 2 + \sqrt{4 - 4p + 2p^2} \right) \approx \frac{4}{p} - 1 & \text{for even } \nu \\ \frac{4}{p} - 1 & \text{for odd } \nu \end{cases}. \quad (4.14)$$

For any given choice of  $p$ , we must have  $\nu \lesssim 4/p - 1$ , restricting the number of states of strength available to a synapse; conversely, for any given choice of  $\nu$ ,  $p$  cannot exceed an upper limit. From a biological perspective, we can circumvent this bound by imposing a non-linearity on the matrices  $p\mathbb{Q}^{\pm}$  so that the

elements  $\pm p a_i$  are replaced by  $\pm \min(1, p a_i)$ . This is equivalent to potentiation or depression being inevitable in certain synaptic strength states. Mathematically, determining the spectrum of  $\mathbb{M}$  with such a non-linearity would in general be difficult, so for simplicity we restrict to the above bound on  $\nu$  for convenience, but with the understanding that in principle there is no obstacle to larger values.

To calculate the initial signal  $\mu_0$ , we require

$$\mathbb{M}^+ \mathbf{A} = \frac{1}{\nu} (1 - p(a_1 - a_0), 1 - p(a_2 - a_1), \dots, 1 - p(a_\nu - a_{\nu-1}))^T. \quad (4.15)$$

Because  $a_{i-1} - a_i = \mathcal{L}_i$ , we have  $\mathbb{M}^+ \mathbf{A} = \mathbf{A} + \frac{p}{\nu} \mathcal{L}$ , and so

$$\mu_0 = \mathcal{L}^T \mathbb{M}^+ \mathbf{A} = \frac{p}{\nu} \mathcal{L}^T \mathcal{L} = \frac{p}{3} \frac{\nu + 1}{\nu - 1}, \quad (4.16)$$

since  $\langle \mathcal{L}^2 \rangle = (\nu + 1)/[3(\nu - 1)]$ . Unlike  $\mu_0$  for  $\mathbb{M} = \mathbb{I} + p \mathbb{C}$ , here for  $\mathbb{M} = \mathbb{I} + p \mathbb{Q}$ ,  $\mu_0$  does depend on whether the strength vector is  $\mathcal{L}$  or  $\mathcal{S}$ . For  $\mathcal{S}$ ,  $\mu_0$  in Eq. (4.16) would become  $\mu_0 = \frac{p}{\nu} \mathcal{S}^T \mathcal{L} = \frac{p}{\nu(\nu-1)} \text{cosec}^2 \frac{\pi}{2\nu}$ . The antisymmetry of  $\mathcal{L}$  in  $\mathbb{M}^+ \mathbf{A} = \mathbf{A} + \frac{p}{\nu} \mathcal{L}$  also immediately establishes  $\mathbb{E}[S_1(t)^2] = \langle \Omega^2 \rangle$  by the same arguments as above, and as required by Eq. (3.10).

Finally, we examine the relative overlap between the sinusoidal strength vector  $\mathcal{S}$  and the normalised eigenvectors  $\hat{\mathbf{e}}^m$ ,  $m = 1, \dots, \nu$ , of this modified form of  $\mathbb{M}$ , where  $\hat{\mathbf{e}}^2 \propto \mathcal{L}$ . We cannot obtain useful expressions for general even  $m$  (for odd  $m$  the overlap vanishes), so we state results only for small even  $m$ , for which the overlap is greatest. For large  $\nu$ , we obtain

$$\frac{\mathcal{S}^T \hat{\mathbf{e}}^m}{\mathcal{S}^T \hat{\mathbf{e}}^2} \rightarrow \begin{cases} -2\sqrt{21} (10 - \pi^2) / \pi^2 & \text{for } m = 4 \\ +5\sqrt{33} (1008 - 112 \pi^2 + \pi^4) / \pi^4 & \text{for } m = 6 \\ -28\sqrt{5} (308880 - 35640 \pi^2 + 450 \pi^4 - \pi^6) / \pi^6 & \text{for } m = 8 \end{cases} .$$

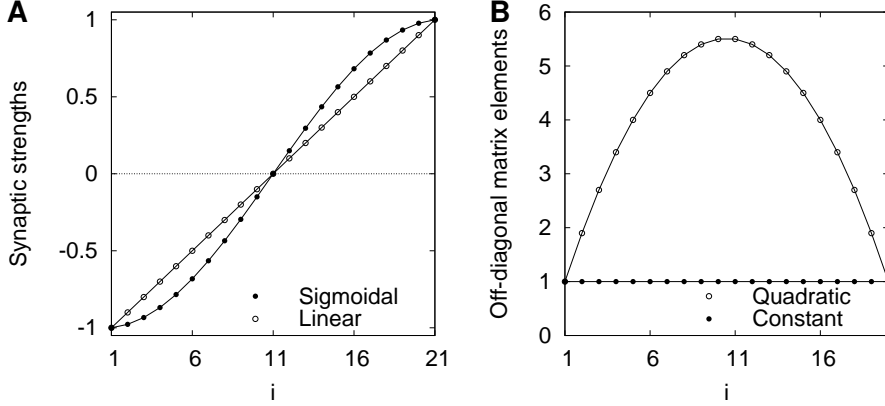
These evaluate to approximately  $-0.121$ ,  $3.95 \times 10^{-3}$  and  $-5.97 \times 10^{-5}$ , respectively, and for  $m = 10$ , we get approximately  $5.21 \times 10^{-7}$ . For  $m = 4$ , the relative overlap is around 12%, for  $m = 6$  under 0.5% and all other values are negligible. Again, then, we expect our analytical results for  $\boldsymbol{\Omega} = \boldsymbol{\mathcal{L}}$  to provide good quantitative agreement with simulation results obtained using  $\boldsymbol{\Omega} = \boldsymbol{\mathcal{S}}$ .

### 4.3 Summary of Both Plasticity Models

	$\mathbb{M} = \mathbb{I} + p\mathbb{C}$	$\mathbb{M} = \mathbb{I} + p\mathbb{Q}$
$\boldsymbol{\Omega}$	$\boldsymbol{\mathcal{S}}$	$\boldsymbol{\mathcal{L}}$
$\lambda_{\boldsymbol{\Omega}}$	$1 - p \left(1 - \cos \frac{\pi}{\nu}\right)$	$1 - p \frac{1}{(\nu-1)}$
$\langle \boldsymbol{\Omega}^2 \rangle$	$\frac{1}{2} \sec^2 \frac{\pi}{2\nu}$	$\frac{1}{3} \frac{\nu+1}{\nu-1}$
$\mu_0$	$\frac{2p}{\nu}$	$\frac{p}{3} \frac{\nu+1}{\nu-1}$
$\nu_{\max}$	$\infty$	$\frac{4}{p} - 1$

**Table 1:** Summary of key quantities for the two models of synaptic plasticity satisfying an eigenvector constraint. Here  $\nu_{\max}$  refers to an upper limit on  $\nu$ .

In Table 1 we assemble for convenience the key quantities in the two models of synaptic plasticity above that satisfy the eigenvector condition  $\boldsymbol{\Omega}^T \mathbb{M} = \lambda_{\boldsymbol{\Omega}} \boldsymbol{\Omega}^T$ . In Fig. 1, we explicitly illustrate the key properties of the vectors  $\boldsymbol{\mathcal{L}}$  and  $\boldsymbol{\mathcal{S}}$  and the matrices  $\mathbb{C}$  and  $\mathbb{Q}$  for the particular choice,  $\nu = 21$  states of synaptic strength. The saturation-like behaviour of  $\boldsymbol{\mathcal{S}}$  is apparent compared to  $\boldsymbol{\mathcal{L}}$ , although in practice these two vectors are quite similar. The quadratic behaviour of the off-diagonal elements of  $\mathbb{Q}$  is transparent, showing that the expression of synaptic plasticity has greatest overall probability for synaptic strengths that are of intermediate sizes, while those at the extremes of the interval  $[-1, +1]$  have lowest overall probability. In contrast, for  $\mathbb{C}$  the probability of the expression of plasticity is independent of synaptic strength.



**Figure 1:** Illustration of major features of synaptic strengths and transition matrices, for  $\nu = 21$ . (A) Synaptic strengths  $\mathcal{L}_i$  and  $\mathcal{S}_i$  plotted against  $i$ . (B) Off-diagonal elements of  $\mathbb{C}$  and  $\mathbb{Q}$ , enumerated down the off-diagonal indexed by  $i$ .

## 5 $\mu(t)$ and $\sigma(t)^2$ for General $\Omega$

Consider any symmetric stochastic matrix  $\mathbb{M}$  that treats potentiation and depression processes symmetrically and that has a complete set of orthonormal eigenvectors  $\hat{\mathbf{e}}^m$  with associated eigenvalues  $\Lambda_m$ . Then because

$$\mu(t) = \Omega^T e^{r(\mathbb{M}-\mathbb{I})t} \mathbb{M}^+ \mathbf{A}, \quad (5.1a)$$

$$\psi(t) = (\Omega^T \otimes \Omega^T) e^{r(\mathbb{M} \otimes \mathbb{M} - \mathbb{I} \otimes \mathbb{I})t} (\mathbb{M}^+ \mathbf{A} \otimes \mathbb{M}^+ \mathbf{A}), \quad (5.1b)$$

the spectral decomposition of  $\mathbb{M}$  allows us to write

$$\mu(t) = \sum_m (\Omega^T \hat{\mathbf{e}}^m) (\hat{\mathbf{e}}^{mT} \mathbb{M}^+ \mathbf{A}) e^{-r(1-\Lambda_m)t}, \quad (5.2a)$$

$$\psi(t) = \sum_{m_1, m_2} (\Omega^T \hat{\mathbf{e}}^{m_1}) (\Omega^T \hat{\mathbf{e}}^{m_2}) (\hat{\mathbf{e}}^{m_1T} \mathbb{M}^+ \mathbf{A}) (\hat{\mathbf{e}}^{m_2T} \mathbb{M}^+ \mathbf{A}) e^{-r(1-\Lambda_{m_1}\Lambda_{m_2})t}. \quad (5.2b)$$

Using  $\psi(t)$  and  $\mu(t)$  we may write down  $\sigma(t)^2$ . We now consider the two explicit forms for  $\mathbb{M}$  above for general  $\boldsymbol{\Omega}$ , and then we may write down SNR memory lifetimes for a variety of models.

### 5.1 Results for $\mathbb{M} = \mathbb{I} + p\mathbb{Q}$

For this form of  $\mathbb{M}$ , we have  $\mathbb{M}^+ \mathbf{A} = \mathbf{A} + \frac{p}{\nu} \boldsymbol{\mathcal{L}}$ , where both  $\mathbf{A}$  and  $\boldsymbol{\mathcal{L}}$  are eigenvectors of  $\mathbb{M}$ . The sums over eigenvectors in Eq. (5.2) therefore collapse to just sums involving  $m = 1$  for  $\hat{\mathbf{e}}^1 \propto \mathbf{A}$  and  $m = 2$  for  $\hat{\mathbf{e}}^2 \propto \boldsymbol{\mathcal{L}}$ , regardless of  $\boldsymbol{\Omega}$ . However,  $\boldsymbol{\Omega}^T \mathbf{A} \equiv 0$  always, since we assume that the vector of possible synaptic strengths  $\boldsymbol{\Omega}$  is antisymmetric, while  $\mathbf{A}$  is symmetric. Hence, the sums collapse to only  $m = 2$ . We are then left with

$$\mu(t) = \frac{p}{3} \frac{\nu + 1}{\nu - 1} \left( \frac{\boldsymbol{\Omega}^T \boldsymbol{\mathcal{L}}}{\boldsymbol{\mathcal{L}}^T \boldsymbol{\mathcal{L}}} \right) e^{-r(1-\Lambda)t}, \quad (5.3a)$$

$$\psi(t) = \mu_0^2 e^{-r(1-\Lambda^2)t}, \quad (5.3b)$$

for *any*  $\boldsymbol{\Omega}$ , where  $\Lambda = 1 - p/(\nu - 1)$ . We have written the expression for  $\mu(t)$  in a form so that we transparently recover  $\mu_0 = \mu(0)$  in Table 1 as the initial memory signal when  $\boldsymbol{\Omega} = \boldsymbol{\mathcal{L}}$ . Strikingly, only a single eigenmode contributes to these statistics, regardless of the vector of possible synaptic strengths  $\boldsymbol{\Omega}$ . This eigenmode is, moreover, the most slowly decaying mode. This remarkable behaviour is entirely due to the very special form of the synaptic configuration immediately after the storage of  $\boldsymbol{\xi}^0$ .

### 5.2 Results for $\mathbb{M} = \mathbb{I} + p\mathbb{C}$

The spectral decomposition of  $\mathbb{C}$  is given explicitly in Eqs. (4.4) and (4.5). Because  $\mathbb{M}^+ \mathbf{A} = \mathbf{A} + \frac{p}{\nu} \mathbf{v}$ , where  $\mathbf{v} = (-1, 0, \dots, 0, +1)^T$  is not an eigenvector

of  $\mathbb{M}$  except for  $\nu = 2$  and  $\nu = 3$ , the sums over eigenvectors in Eq. (5.2) do not collapse for general  $\mathbf{\Omega}$ . However, for the specific choice  $\mathbf{\Omega} = \mathbf{S}$ , the sums do collapse down to just  $m = 1$ , corresponding to  $\hat{\mathbf{e}}^1 \propto \mathbf{S}$ . In general, however, since we are interested in memory lifetimes, we are interested specifically in the large-time behaviour of  $\mu(t)$  and  $\psi(t)$ . We may therefore simplify by considering an approximation that includes just the most slowly decaying eigenmode, which also corresponds to the  $m = 1$  mode and thus  $\mathbf{S}$ . Asymptotically, this approximation becomes exact. We then obtain

$$\mu(t) \underset{\mathbf{S}}{\cong} \frac{2p}{\nu} \left( \frac{\mathbf{\Omega}^T \mathbf{S}}{\mathbf{S}^T \mathbf{S}} \right) e^{-r(1-\Lambda)t}, \quad (5.4a)$$

$$\psi(t) \underset{\mathbf{S}}{\cong} \left[ \frac{2p}{\nu} \left( \frac{\mathbf{\Omega}^T \mathbf{S}}{\mathbf{S}^T \mathbf{S}} \right) \right]^2 e^{-r(1-\Lambda^2)t}, \quad (5.4b)$$

where  $\Lambda = 1 - p \left( 1 - \cos \frac{\pi}{\nu} \right)$  and where we use the symbol “ $\underset{\mathbf{S}}{\cong}$ ” to indicate that we have equality for  $\mathbf{\Omega} = \mathbf{S}$  (exact equality for all times  $t$ ) and asymptotic equality otherwise (asymptotic equality only at large times).

### 5.3 SNR Memory Lifetimes

We use these results to obtain SNR memory lifetimes for either choice of  $\mathbb{M}$  and for either choice of  $\mathbf{\Omega}$ , giving four combinations. As  $\tau_{\text{snr}}$  is the solution of  $\mu(\tau_{\text{snr}}) = \sigma(\tau_{\text{snr}})$ , in general it must be obtained numerically, but for large  $\tau_{\text{snr}}$  we may approximate  $\sigma(t)^2 \approx \langle \mathbf{\Omega}^2 \rangle / N = 1/N'$ . We write Eqs. (5.3a) and (5.4a) in the common form  $\mu(t) \cong E e^{-r(1-\Lambda)t}$ , where  $E$  depends on the choice of  $\mathbb{M}$  and  $\mathbf{\Omega}$ . We have exact equality in this equation for three combinations, for which  $E \equiv \mu_0$ , and asymptotic equality for the remaining combination. We then obtain

$$r(1-\Lambda)\tau_{\text{snr}} \approx \frac{1}{2} \log_e(N'E^2). \quad (5.5)$$



This result is identical, up to additive constants, to the asymptotic form for  $\tau_{\text{mfpt}}(h_0)$  in Eq. (3.26a) for  $\theta = 0$  with  $h_0 = E$ , where  $E \equiv \mu_0$  for three combinations. In Table 2, we give the explicit results for  $r\tau_{\text{snr}}$  for all four combinations, in the full form for any  $\nu$  and for large  $\nu$ . We note that  $768/\pi^4 \approx 7.88 \approx 8$  and therefore also  $32/\pi^4 \approx 0.33 \approx 1/3$ , so these SNR lifetimes are quite insensitive to the choice of  $\Omega$ . We have obtained the results for  $\mathbb{C}$  &  $\mathcal{L}$  before (Elliott, 2016a). The  $\mathbb{C}$  &  $\mathcal{L}$  model has also been extensively studied by previous authors, although purely in an SNR context. For example, Amit & Fusi (1994) derived, up to overall numerical factors, expressions essentially equivalent to our own for  $\tau_{\text{snr}}$ ; see Elliott (2016a) for a fuller discussion.

Model	Full $\nu$ form	Large $\nu$ form
$\mathbb{C}$ & $\mathcal{L}$	$\frac{1}{2p(1-\cos\frac{\pi}{\nu})} \log_e \frac{48Np^2 \cot^4 \frac{\pi}{2\nu}}{\nu^4(\nu^2-1)}$	$\frac{\nu^2}{p\pi^2} \log_e \frac{768Np^2}{\pi^4\nu^2}$
$\mathbb{C}$ & $\mathcal{S}$	$\frac{1}{2p(1-\cos\frac{\pi}{\nu})} \log_e \frac{8Np^2 \cos^2 \frac{\pi}{2\nu}}{\nu^2}$	$\frac{\nu^2}{p\pi^2} \log_e \frac{8Np^2}{\nu^2}$
$\mathbb{Q}$ & $\mathcal{L}$	$\frac{\nu-1}{2p} \log_e \frac{Np^2}{3} \frac{\nu+1}{\nu-1}$	$\frac{\nu}{2p} \log_e \frac{Np^2}{3}$
$\mathbb{Q}$ & $\mathcal{S}$	$\frac{\nu-1}{2p} \log_e \frac{2Np^2 \cot^2 \frac{\pi}{2\nu} \operatorname{cosec}^2 \frac{\pi}{2\nu}}{\nu^2(\nu-1)^2}$	$\frac{\nu}{2p} \log_e \frac{32Np^2}{\pi^4}$

**Table 2:** Results for  $r\tau_{\text{snr}}$  for the four combinations of  $\mathbb{C}$  or  $\mathbb{Q}$  and  $\mathcal{L}$  or  $\mathcal{S}$ , in the full form for any  $\nu$  and for large  $\nu$ .

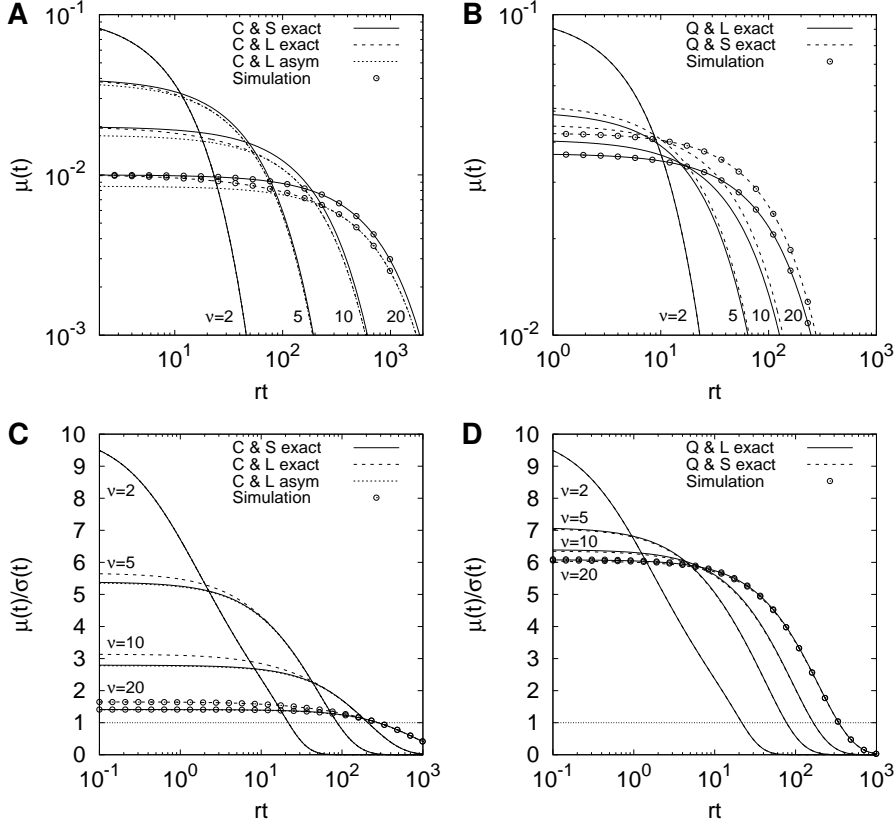
## 6 Results: Analysis and Simulation

We now examine memory lifetimes using the above analytical results. We also compare these results to data obtained from simulation. Details of simulation protocols may be found elsewhere (Elliott & Lagogiannis, 2012; Elliott, 2014). To obtain good statistics, we average simulations over multiple runs. Typically we average over  $10^5$  runs, but for  $N = 10^5$  we use  $10^4$  runs for speed. Simulation

results produce both the mean and variance in the memory signal and the MFPT memory lifetime and its variance.

We first exhibit the mean memory signal  $\mu(t)$  and the SNR  $\sigma(t)/\mu(t)$  in Fig. 2. In Fig. 2A, we plot analytical results for  $\mu(t)$  against time for the  $\mathbb{C} \& \mathcal{L}$  and  $\mathbb{C} \& \mathcal{S}$  models. For the former we show the exact result including all eigenmodes and the asymptotic result with just the slowest eigenmode. We also show simulation data, confirming agreement with analytical results for sufficient runs. The behaviour of the exact form of the  $\mathbb{C} \& \mathcal{L}$  model interpolates between its asymptotic form at large times and the  $\mathbb{C} \& \mathcal{S}$  model at small times. The initial signal for both exact  $\mathbb{C}$  models is  $\frac{2p}{\nu}$ , so as  $\nu$  increases,  $\mu_0$  decreases, and in the limit it drops to zero. Conversely, the dominant eigenmode is governed by  $\frac{p\pi^2}{2\nu^2}$ , so as  $\nu$  increases, this mode is increasingly longer lived. Fig. 2B shows similar results for the  $\mathbb{Q} \& \mathcal{S}$  and  $\mathbb{Q} \& \mathcal{L}$  models, although there is no equivalent of the asymptotic form of the  $\mathbb{C} \& \mathcal{L}$  model. The initial signals of these models are proportional to  $\frac{p}{3} \frac{\nu+1}{\nu-1}$  (with equality for  $\mathbb{Q} \& \mathcal{L}$ ), so as  $\nu$  increases, these signals remain non-zero, in radical contrast to the  $\mathbb{C}$  model. The dominant eigenmode is now governed by  $\frac{p}{\nu}$ , so although this goes to zero more slowly than  $\frac{p\pi^2}{2\nu^2}$  as  $\nu$  increases (i.e. the mode dies faster with  $t$ ), the memory signal is still increasingly longer lived as  $\nu$  increases.

These differences between the initial signals and the decay rates of the relevant eigenmodes in the  $\mathbb{C}$  and  $\mathbb{Q}$  models can be seen in Figs. 2C and 2D, showing the SNR  $\mu(t)/\sigma(t)$ , with  $\sigma(t)^2$  determined exactly. The SNRs for the asymptotic  $\mathbb{C} \& \mathcal{L}$  and exact  $\mathbb{C} \& \mathcal{S}$  models in Fig. 2C are virtually indistinguishable, as are those for the  $\mathbb{Q} \& \mathcal{L}$  and  $\mathbb{Q} \& \mathcal{S}$  models in Fig. 2D. With the



**Figure 2:** Evolution of mean memory signal  $\mu(t)$  and SNR  $\mu(t)/\sigma(t)$  over time. (A) Exact analytical results for  $\mu(t)$  for the  $\mathbb{C} \& \mathcal{S}$  and  $\mathbb{C} \& \mathcal{L}$  models and for the most slowly decaying mode of the  $\mathbb{C} \& \mathcal{L}$  model (“asym”), for different choices of  $\nu$ , as indicated. (B) Same as A, except for the  $\mathbb{Q}$  model; the exact results only contain a single decay mode. In both A and B, all results for  $\nu = 2$  are identical, so only a single line is visible. Simulation results, shown as points, are shown only for  $\nu = 20$  for clarity. (C) and (D) Same as A and B, respectively, but showing  $\mu(t)/\sigma(t)$  instead of  $\mu(t)$ . A thin line corresponding to  $\mu(t)/\sigma(t) = 1$  is also shown. In all panels with have set  $p = 0.1$  and in C and D, we have set  $N = 10^4$ .

approximation  $\sigma(t)^2 \approx \langle \Omega^2 \rangle / N$ , we obtain

$$\frac{[\mu(t)/\sigma(t)]_{\mathbb{Q} \& \mathcal{S}}}{[\mu(t)/\sigma(t)]_{\mathbb{Q} \& \mathcal{L}}} = \frac{[\mu(t)/\sigma(t)]_{\mathbb{C} \& \mathcal{L}}^{\text{asympt}}}{[\mu(t)/\sigma(t)]_{\mathbb{C} \& \mathcal{S}}} = \cos \kappa, \quad (6.1)$$

where  $\kappa$  is the angle between the  $\mathcal{L}$  and  $\mathcal{S}$  vectors, with  $\cos \kappa \geq \frac{4\sqrt{6}}{\pi^2} \approx 0.993$  for  $\nu \geq 2$ . Hence, these ratios are very close to unity, explaining the near-indistinguishability of the SNRs in Figs. 2C and 2D. Also, because the exact form for the  $\mathbb{C} \& \mathcal{L}$  model approaches its asymptotic form, their SNRs must become indistinguishable and so nearly indistinguishable from the  $\mathbb{C} \& \mathcal{S}$  model's SNR. All  $\mathbb{C}$  models will thus have nearly identical SNR memory lifetimes; similarly for both  $\mathbb{Q}$  models. This is clear from the intercepts of the  $\mu(t)/\sigma(t)$  curves with unity, defining  $\tau_{\text{snr}}$ , in Figs. 2C and 2D.

We explore the SNR memory lifetime  $\tau_{\text{snr}}$  in more detail in Fig. 3. We again determine  $\sigma(t)^2$  exactly, but with  $\sigma(t)^2 \approx \langle \Omega^2 \rangle / N$  the results are virtually indistinguishable. In Figs. 3A and 3B we plot  $\tau_{\text{snr}}$  as a function of  $N$ . For the  $\mathbb{C}$  model in Fig. 3A, thick lines show results for the  $\mathbb{C} \& \mathcal{L}$  model including all eigenmodes, while thin lines show results for the  $\mathbb{C} \& \mathcal{S}$  model, which are indistinguishable from the asymptotic form of the  $\mathbb{C} \& \mathcal{L}$  model. For large enough  $N$  all results are indistinguishable, while for smaller  $N$  the results for the exact form of the  $\mathbb{C} \& \mathcal{L}$  model differ slightly from the other results. For the  $\mathbb{Q}$  models in Fig. 3B, both sets of results are indistinguishable. In all cases, we see a bifurcation-like phenomenon with the rapid onset of  $\tau_{\text{snr}} > 0$  for  $N$  above some critical value. Taking the large  $\nu$  forms in Table 2, these critical values can be read off as

$$N > \begin{cases} \frac{\nu^2}{8p^2} & \text{for } \mathbb{C} \& \mathcal{S} \\ \frac{3}{p^2} & \text{for } \mathbb{Q} \& \mathcal{L} \end{cases} \quad (6.2a)$$

$$N > \begin{cases} \frac{\nu^2}{8p^2} & \text{for } \mathbb{C} \& \mathcal{S} \\ \frac{3}{p^2} & \text{for } \mathbb{Q} \& \mathcal{L} \end{cases} \quad (6.2b)$$

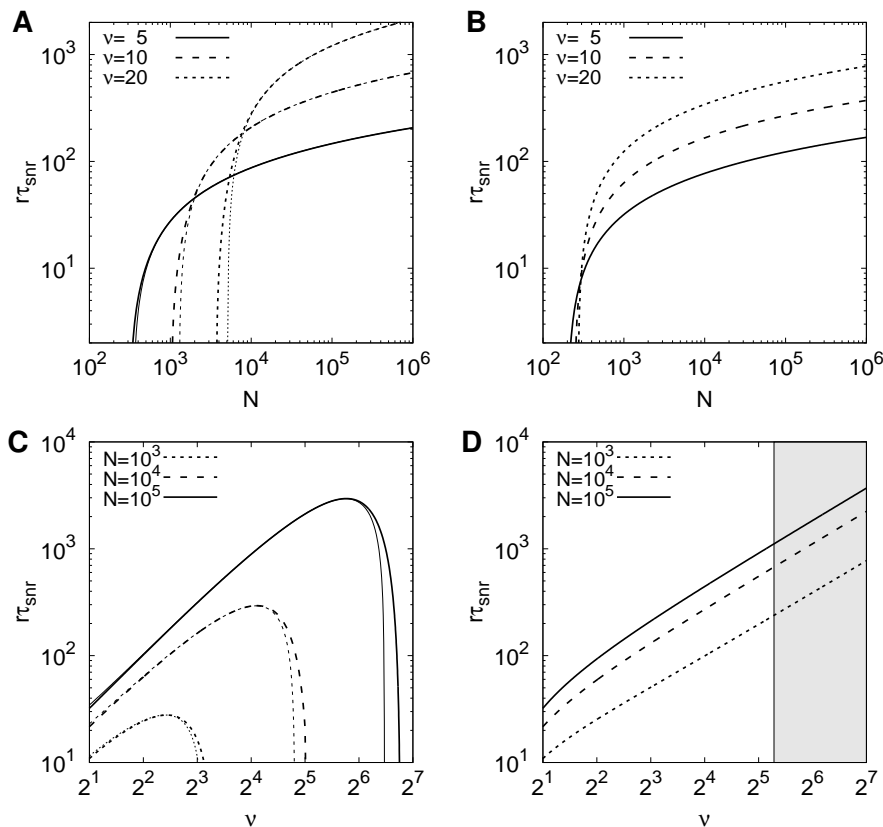
for the two simplest cases. For the  $\mathbb{C}$  model this critical value scales quadratically with  $\nu$ , but for the  $\mathbb{Q}$  model it exhibits no dependence on  $\nu$  in the limit of large  $\nu$ . Away from this limit there is mild sensitivity to  $\nu$ , but even for  $\nu = 10$  and  $\nu = 20$ , the critical values are very similar.

This striking difference between the  $\mathbb{C}$  and  $\mathbb{Q}$  models is further illustrated by plotting  $\tau_{\text{snr}}$  as a function of  $\nu$  rather than  $N$ , in Figs. 3C and 3D. The trade-off between the rapid fall in the initial signal  $\frac{2p}{\nu}$  and the increasingly longer-lived eigenmode governed by  $\frac{p\pi^2}{2\nu^2}$  for the  $\mathbb{C}$  model gives rise to a maximum in  $\tau_{\text{snr}}$  for a particular value of  $\nu$ . Beyond this maximum there are differences in the lifetimes for the various  $\mathbb{C}$  models because for shorter lifetimes and larger  $\nu$ , more eigenmodes in the  $\mathbb{C} \& \mathcal{L}$  model must be considered. For the asymptotic form of the  $\mathbb{C} \& \mathcal{L}$  model or the exact form of the  $\mathbb{C} \& \mathcal{S}$  model, we determine the value of  $\nu$ , call it  $\nu_{\text{opt}}$ , that optimises  $\tau_{\text{snr}}$ , giving

$$\nu_{\text{opt}} = \begin{cases} p \sqrt{\frac{768}{\pi^4 e}} \sqrt{N} & \text{for } \mathbb{C} \& \mathcal{L} \\ p \sqrt{\frac{8}{e}} \sqrt{N} & \text{for } \mathbb{C} \& \mathcal{S}. \end{cases} \quad (6.3a)$$

$$\nu_{\text{opt}} = \begin{cases} p \sqrt{\frac{8}{e}} \sqrt{N} & \text{for } \mathbb{C} \& \mathcal{S}. \end{cases} \quad (6.3b)$$

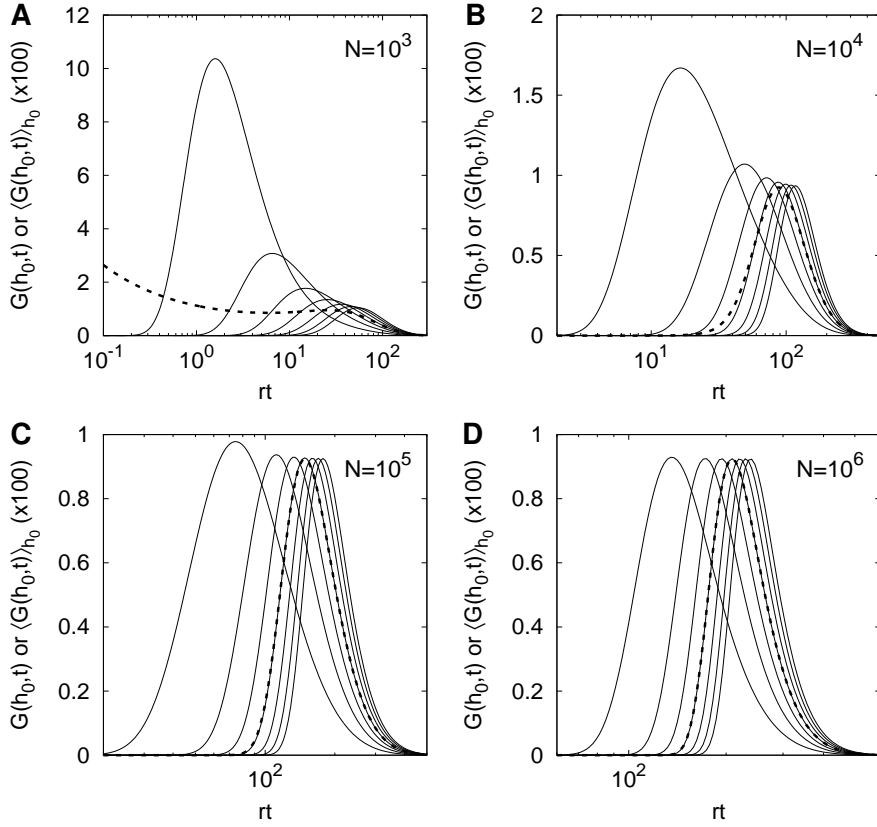
The numerical factors are virtually identical. These give optimal SNR memory lifetimes, respectively, of  $0.294 pN/r$  and  $0.298 pN/r$ . SNR optimality conditions have been studied by previous authors (e.g., Amit & Fusi, 1994); see Elliott (2016a) for a fuller discussion and a critique of the relevance of such conditions. In contrast to the behaviour for the  $\mathbb{C}$  model, in the  $\mathbb{Q}$  model  $\tau_{\text{snr}}$  grows monotonically with  $\nu$  because its initial signal is bounded from below by  $\frac{p}{3}$  rather than collapsing as  $\nu$  increases. Indeed, from the  $\tau_{\text{snr}}$  results in Table 2, and as seen in the shaded region in Fig. 3D, the  $\mathbb{Q}$  model's SNR memory lifetimes would grow unboundedly as  $\nu$  increases, although  $\nu$  is restricted to  $\nu \lesssim \frac{4}{p} - 1$ .



**Figure 3:** SNR memory lifetime  $\tau_{\text{snr}}$  as a function of  $N$  and  $\nu$ , for  $p = 0.1$ . (A) Memory lifetimes for the  $\mathbb{C} \& \mathcal{L}$  (thick lines) and  $\mathbb{C} \& \mathcal{S}$  (thin lines) models for different choices of  $\nu$ , as a function of  $N$ . (B) Same as A, except for the  $\mathbb{Q} \& \mathcal{L}$  model; results for the  $\mathbb{Q} \& \mathcal{S}$  model are virtually identical and are not shown. (C) Same as A, except results are shown for different choices of  $N$ , as a function of  $\nu$ . (D) Same as B, except results are shown for different choices of  $N$ , as a function of  $\nu$ . We have shaded the region for  $\nu > 39$  because for  $p = 0.1$ , the  $\mathbb{Q}$  matrix ceases to be a stochastic matrix here.

Having examined SNR memory lifetimes, we turn to FPT-defined memory lifetimes. Before explicitly considering the FPT statistics, we first plot the FPT distribution itself as a function of time in Fig. 4. For simplicity we set  $\theta = 0$  so that we may use Eqs. (3.22) and (3.28), and we consider only the  $\mathbb{C} \& \mathcal{S}$  model as results for the  $\mathbb{Q} \& \mathcal{L}$  model are qualitatively identical. For each choice of  $N$  and for fixed  $h_0$ ,  $G(h_0, t)$  is roughly-speaking rather log-normal-like in appearance. As  $h_0$  increases, the peak in the FPT distribution shifts to larger  $t$ , with the distribution broadening over a wider range of  $t$ ; the distributions for fixed  $N$  also become increasingly similar as  $h_0$  increases. For  $N$  large enough, the averaged distribution  $\langle G(h_0, t) \rangle_{h_0}$  is essentially identical to its mean field form  $G(\mu_0, t)$  over the shown range of  $t$ , where for the parameters used,  $\mu_0 = 0.04$ . For small  $t$ ,  $\langle G(h_0, t) \rangle_{h_0} \sim \frac{1}{\pi} \sqrt{\frac{\rho}{2t}} e^{-\frac{1}{2}\mu'^2}$ , with the divergence reflecting the presence of the absorbing boundary at  $h_0 = \theta$ . For  $N = 10^3$  in Fig. 4A, we see the onset of this divergence in  $\langle G(h_0, t) \rangle_{h_0}$  for small  $t$ , but as  $N$  increases the factor  $e^{-\frac{1}{2}\mu'^2}$ , where  $\mu' = \mu_0 \sqrt{N'}$ , is increasingly suppressed, requiring very small  $t$  for  $\langle G(h_0, t) \rangle_{h_0}$  to start growing.

We now consider the FPT statistics. In Figs. 5 and 6 for the  $\mathbb{C} \& \mathcal{S}$  and  $\mathbb{Q} \& \mathcal{L}$  models, respectively, we consider the mean  $\tau_{\text{mfpt}}(h_0)$  and standard deviation  $\sigma_{\text{fpt}}(h_0)$  as a function  $h_0$ . In Figs. 5A and 6A, we plot  $\tau_{\text{mfpt}}(h_0)$  for various choices of  $\nu$  for  $\theta = 0$ , showing results using the full form of the jump moments and using their OU form. They differ by no more than 0.5% and are indistinguishable in these figures, so the OU form constitutes an extremely good approximation. We also show the asymptotic form of  $\tau_{\text{mfpt}}(h_0)$  for  $\theta = 0$  given in Eq. (3.26a). This form is valid for either large enough  $N$  or large enough  $h_0$ . For large enough  $h_0$ ,  $\tau_{\text{mfpt}}(h_0)$  behaves logarithmically, but for smaller  $h_0$  it grows initially linearly with  $h_0$ . As the asymptotic form induces a bifurcation-like process completely equivalent to that seen for  $\tau_{\text{snr}}$ , we deduce



**Figure 4:** First passage time distributions  $G(h_0, t)$  and  $\langle G(h_0, t) \rangle_{h_0}$  as functions of time  $t$  for the  $\mathbb{C}$  &  $\mathcal{S}$  model. We set  $\theta = 0$ ,  $p = 0.1$  and  $\nu = 5$ . Solid lines show  $G(h_0, t)$  for  $h_0 = 0.01, \dots, 0.07$  in increments of 0.01 moving left to right in all four panels. Thick dashed lines show  $\langle G(h_0, t) \rangle_{h_0}$ . Each panel, (A) to (D), is distinguished by a particular choice of  $N$ , as indicated.

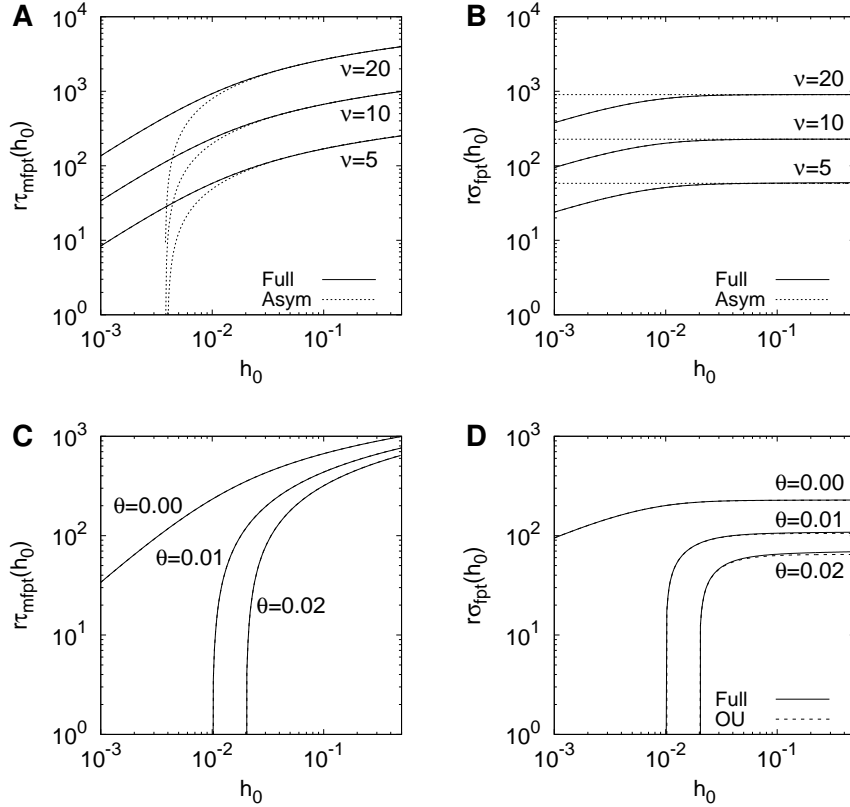


that SNR memory lifetimes are only asymptotically valid, giving misleading results for smaller  $N$ . Further, while the behaviours of SNR memory lifetimes for the  $\mathbb{C}$  and  $\mathbb{Q}$  models are very different, MFPT memory lifetimes for small  $h_0$  (and so for smaller  $N$ ) for these models exhibit no such differences. Hence, not only are SNR memory lifetimes asymptotic approximations, but they exhibit a strong model-dependence that is not seen in MFPT memory lifetimes. Figs. 5B and 6B show the equivalents of Figs. 5A and 6A but for  $\sigma_{\text{fpt}}(h_0)$ . These figures confirm Eq. (3.26b), that for  $\theta = 0$  the variance in the FPT distribution is independent of  $h_0$  for  $h_0$  large enough.

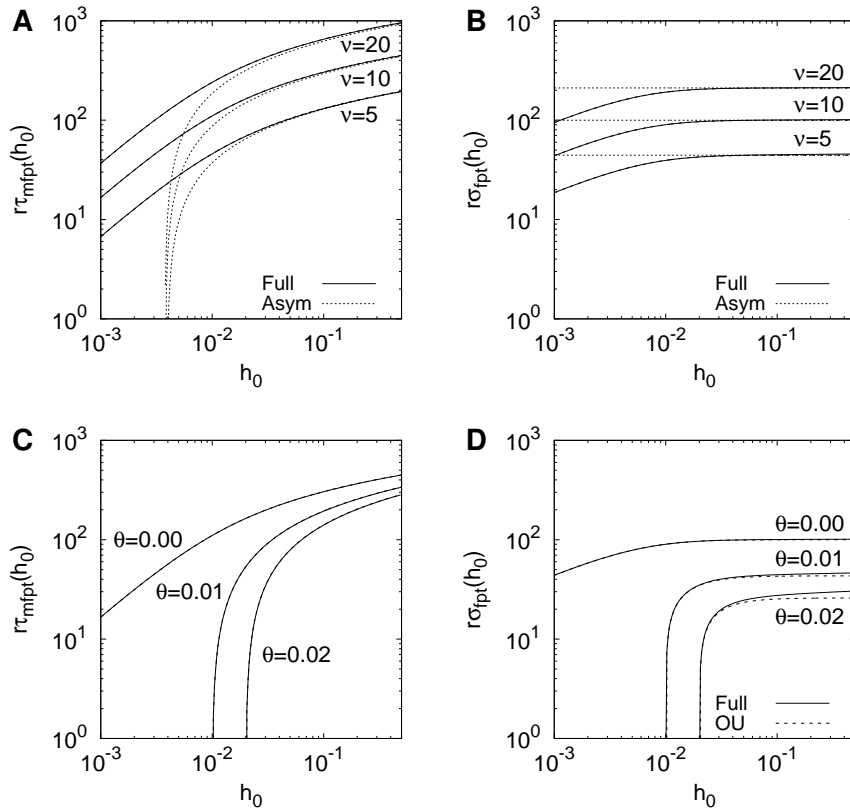
In Figs. 5C and 6C we consider  $\tau_{\text{mfpt}}(h_0)$  for  $\theta > 0$ , and Figs. 5D and 6D consider  $\sigma_{\text{fpt}}(h_0)$ . For  $\tau_{\text{mfpt}}(h_0)$ , the results using the full jump moments are indistinguishable from their OU form, although for  $\sigma_{\text{fpt}}(h_0)$  for  $\theta > 0$  we begin to see small discrepancies at larger values of  $h_0$ , but the OU form remains a qualitatively excellent and quantitatively very good approximation. As  $h_0$  increases, the OU form of  $\sigma_{\text{fpt}}(h_0)$  asymptotes to a constant (the full form exhibits a very slow rise), with this constant reducing as  $\theta$  increases. Furthermore,  $\tau_{\text{mfpt}}(h_0)$  systematically falls as  $\theta$  increases, although for fixed  $N$ , it continues to asymptote to logarithmic behaviour for large  $h_0$ . Again, we see no qualitative difference in the behaviours of the  $\mathbb{C}$  and  $\mathbb{Q}$  models for  $\theta > 0$ .

We now turn to examining the FPT statistics averaged over the distribution of  $h_0$  rather than as a function of  $h_0$ . Because we average over simulations in which the tracked memory varies between simulations, we automatically obtain  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  and  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$ , so we may directly compare simulation results to analytical results. Analytical FPT results are available only for the  $\mathbb{C}$  &  $\mathcal{S}$  and  $\mathbb{Q}$  &  $\mathcal{L}$  models, but we may simulate all four models by implementing either choice of  $\Omega$  for either choice of  $\mathbb{M}$ .

Fig. 7 shows for  $\theta = 0$  both  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  and its mean field form  $\tau_{\text{mfpt}}(\mu_0)$



**Figure 5:** First passage time statistics  $\tau_{\text{mfpt}}(h_0)$  and  $\sigma_{\text{fpt}}(h_0)$  as a function of the initial perceptron output  $h_0$ , for the  $\mathbb{C}$  &  $\mathcal{S}$  model. (A)  $\tau_{\text{mfpt}}(h_0)$  for different choices of  $\nu$  and with firing threshold  $\theta = 0$ , computed exactly (“full”), using the OU form of the FPT distribution (results indistinguishable from full form), or using the asymptotic (logarithmic) form of the OU result for large  $h_0$  (“asym”) in Eq. (3.26a). (B) Same as A, except results are shown for  $\sigma_{\text{fpt}}(h_0)$ , the asymptotic form of which is given in Eq. (3.26b). (C)  $\tau_{\text{mfpt}}(h_0)$  for different choices of firing threshold  $\theta$  and with  $\nu = 20$ . Both the full and OU forms are plotted but are indistinguishable. (D) Same as C, except results are shown for  $\sigma_{\text{fpt}}(h_0)$ . Now the full and OU forms are just distinguishable. In all panels we have set  $p = 0.1$  and  $N = 10^4$ .

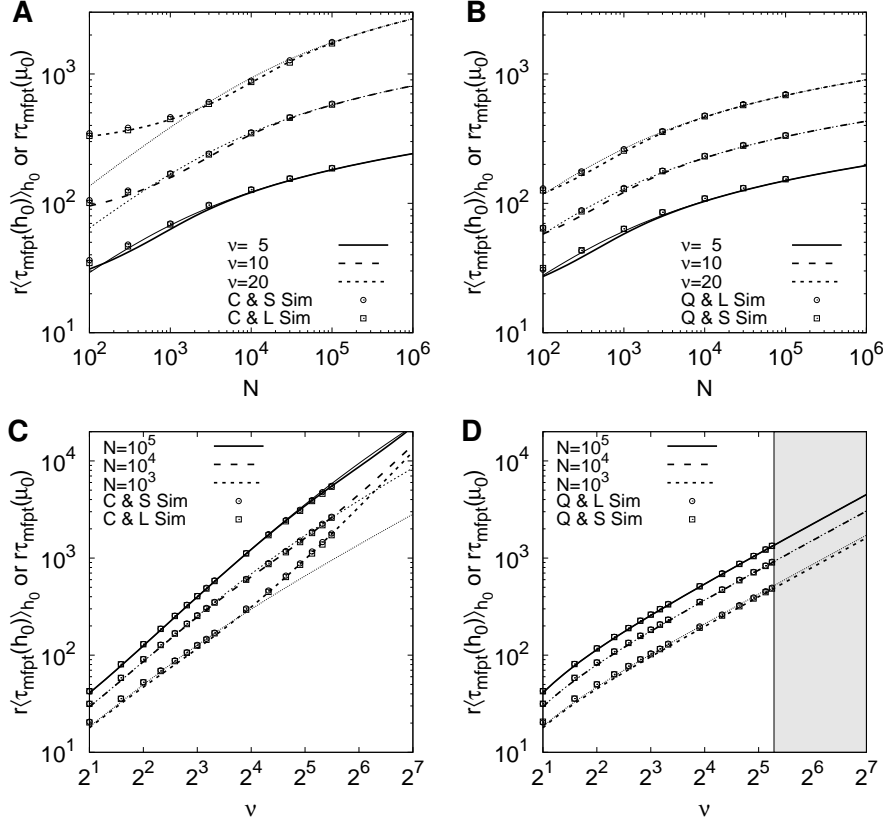


**Figure 6:** First passage time statistics  $\tau_{mfpt}(h_0)$  and  $\sigma_{fpt}(h_0)$  as a function of the initial perceptron output  $h_0$ , for the  $\mathcal{Q}$  &  $\mathcal{L}$  model. The format of this figure is otherwise identical to Fig. 5.

and is directly comparable to Fig. 3 showing  $\tau_{\text{snr}}$ . In Figs. 7A and 7B for the  $\mathbb{C}$  and  $\mathbb{Q}$  models, respectively, we plot memory lifetimes against  $N$ . For larger values of  $N$ ,  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  is indistinguishable from  $\tau_{\text{mfpt}}(\mu_0)$ , but for smaller values, they separate. For the  $\mathbb{C}$  model, the region where the separation occurs as a function of  $N$  depends strongly on the value of  $\nu$ , while for the  $\mathbb{Q}$  model, it is quite insensitive to  $\nu$ . Unlike SNR lifetimes, MFPT lifetimes do not exhibit bifurcation-like dynamics at critical values of  $N$ , and they are substantially non-zero even when  $\tau_{\text{snr}} = 0$ . We note that  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  and  $\tau_{\text{mfpt}}(\mu_0)$  separate in the vicinity of the bifurcation-like dynamics in  $\tau_{\text{snr}}$ .

Plotting memory lifetimes against  $\nu$  instead of  $N$  in Figs. 7C and 7D, for the  $\mathbb{C}$  model  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  and  $\tau_{\text{mfpt}}(\mu_0)$  are indistinguishable for smaller values of  $\nu$  but separate for larger values, with the separation region depending on  $N$ . For the  $\mathbb{Q}$  model  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  and  $\tau_{\text{mfpt}}(\mu_0)$  are indistinguishable for larger choices of  $N$  for all values of  $\nu$ , while for  $N = 10^3$  there is a very small, systematic difference that is essentially independent of  $\nu$ . Unlike SNR lifetimes, the MFPT lifetimes in Fig. 7C do not exhibit an optimal value of  $\nu$  at which MFPT lifetimes are maximised as a function of  $\nu$ . We again note that the separation of  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  and  $\tau_{\text{mfpt}}(\mu_0)$  occurs near the value of  $\nu$  at which SNR lifetimes are optimised for the  $\mathbb{C}$  model. For the  $\mathbb{Q}$  model, neither SNR nor MFPT lifetimes exhibit such optimality.

In Eq. (3.29) we saw that  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  and  $\tau_{\text{mfpt}}(\mu_0)$  coincide for large  $\mu' = \mu_0 \sqrt{N'}$ . To determine what constitutes large  $\mu'$ , we integrate the analytical large  $\mu'$  form of  $\langle G(h_0, t) \rangle_{h_0}$  over time, obtaining  $\text{erf}(\mu'/\sqrt{2})$ . This must be close to unity. Since  $\mu' = 2$  gives 95% agreement, large  $\mu'$  means  $\mu' \gtrsim 2$ . We also saw from Eqs. (3.26) and (5.5) that the expressions for  $\tau_{\text{mfpt}}(\mu_0)$  for large  $\mu'$  and for  $\tau_{\text{snr}}$  are identical up to additive constants. Therefore,  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  and  $\tau_{\text{snr}}$  are similar only when  $\tau_{\text{mfpt}}(\mu_0)$  is similar to  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$ , for  $\mu' \gtrsim 2$ .



**Figure 7:** MFPT memory lifetime  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  as a function of  $N$  and  $\nu$ , for  $p = 0.1$ . Also shown for comparison is  $\tau_{\text{mfpt}}(\mu_0)$ , so with  $\tau_{\text{mfpt}}(h_0)$  evaluated as the mean initial memory signal. The format of this figure is essentially identical to Fig. 3, except that thick lines show  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  while thin lines show  $\tau_{\text{mfpt}}(\mu_0)$ ; analytical results are shown (and available) only for the  $\mathbb{C}$  &  $\mathbb{S}$  and  $\mathbb{Q}$  &  $\mathbb{L}$  models; and simulation results are also shown as points.

For  $\mu' \gtrsim 2$ , qualitative differences between  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$ ,  $\tau_{\text{mfpt}}(\mu_0)$  and  $\tau_{\text{snr}}$  will emerge and become significant for small  $\mu'$ , i.e.  $\mu' < 1$  or  $\mu' \ll 1$ . For the  $\mathbb{C} \& \mathcal{S}$  model,  $\mu' = 2\sqrt{2N} \frac{p}{\nu} \cos\left(\frac{\pi}{2\nu}\right) \sim 2\sqrt{2N} \frac{p}{\nu}$  for large  $\nu$ , while for the  $\mathbb{Q} \& \mathcal{L}$  model,  $\mu' = p \sqrt{\frac{N}{3} \frac{\nu+1}{\nu-1}} \sim p \sqrt{\frac{N}{3}}$ . Computing the parameter values for which  $\mu' \approx 2$  in Fig. 7 we obtain: in Fig. 7A,  $N \approx 1400, 5100$  and  $20,000$  for  $\nu = 5, 10$  and  $20$ , respectively; in Fig. 7B,  $N \approx 800, 1000$  and  $1100$  for  $\nu = 5, 10$  and  $20$ ; and in Fig. 7C,  $\nu \approx 4, 14$  and  $44$  for  $N = 10^3, 10^4$  and  $10^5$ , respectively. In Fig. 7D, for  $N = 10^5$ ,  $\mu' \gtrsim 18.7$  and for  $N = 10^4$ ,  $\mu' \gtrsim 5.9$ , so in excess of 2; and for  $N = 10^3$ ,  $\mu'$  ranges from 3.2 to 1.9 for  $\nu$  between 2 and 39 with  $\mu' = 2$  for  $\nu = 11$ , so  $\mu'$  is either large enough or close to being so. These parameter values are in good agreement with the locations of the separation of  $\tau_{\text{mfpt}}(\mu_0)$  from  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  in Figs. 7A–C and explain their indistinguishability or closeness in Fig. 7D. They therefore also indicate when SNR lifetimes will start to significantly differ from MFPT lifetimes.

For the  $\mathbb{C}$  model,  $\mu' \rightarrow 0$  as  $\nu \rightarrow \infty$ , so that the small  $\mu'$  form for  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  in Eq. (3.29) holds. For the  $\mathbb{Q}$  model,  $\mu'$  is usually large enough that the large  $\mu'$  form holds. In both cases, these results prove that  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  increases indefinitely as  $\nu$  increases, unlike  $\tau_{\text{snr}}$  for the  $\mathbb{C}$  model. We conclude that not only do SNR lifetimes incorrectly predict a minimum value of  $N$  for successful memory storage, but they also incorrectly predict a maximum possible memory lifetime for some particular choice of  $\nu$  in the  $\mathbb{C}$  model.

Comparing analytical and simulation results in Fig. 7, we see very good agreement. First, simulation results for  $\Omega = \mathcal{S}$  and  $\Omega = \mathcal{L}$  for the  $\mathbb{C}$  or  $\mathbb{Q}$  models are typically extremely similar and often indistinguishable. For smaller  $N$ , the difference can be 10%, but for larger  $N$ , the difference reduces. For the  $\mathbb{C}$  model, it can be as much as 4% for larger  $N$  but is typically around 1%; for the  $\mathbb{Q}$  model, it can be up to 2% but is again typically under 1%. These dif-

ferences are better than our estimates above of the errors incurred in using the “wrong” eigenvector, which were around 25% and 12% for the  $\mathbb{C}$  and  $\mathbb{Q}$  models. Thus, while the FPT calculations require that  $\mathbf{\Omega}$  is an appropriate eigenvector of  $\mathbb{M}$ , in general the results are only mildly sensitive to the actual choice of  $\mathbf{\Omega}$ . Second, comparing simulation and analytical results for  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$ , we again see very good agreement, with small discrepancies only for smaller  $N$ . The continuum limit underlying the Fokker-Planck analysis requires larger  $N$ , with some discrepancies arising when binomial distributions are replaced by Gaussian distributions. Also, the Fokker-Planck analysis ignores jump processes, which will be more significant for smaller  $N$  and so for shorter MFPTs.

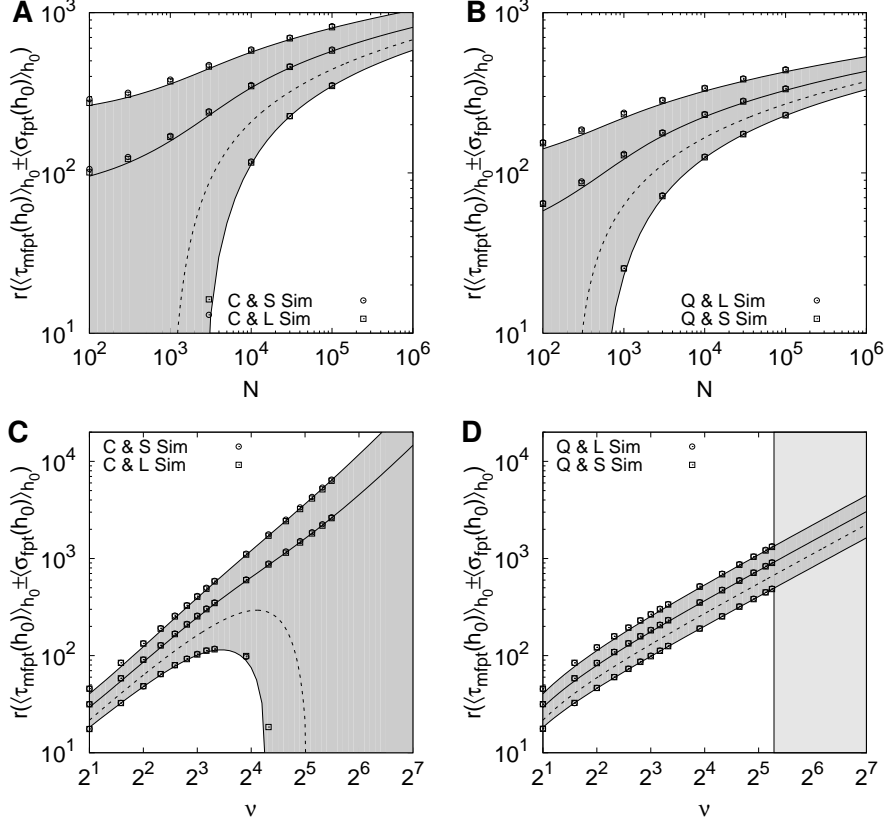
Fig. 7 considers only the MFPT  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$ . In Fig. 8, we examine  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$  by plotting the one standard deviation region around  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$ . To avoid clutter, we do this only for  $\nu = 10$  in Figs. 8A and 8B and only for  $N = 10^4$  in Figs. 8C and 8D. The small and large  $\mu'$  results in Eq. (3.30) provide good agreement with the exact form in their appropriate regions of validity. We also see good agreement between analytical and simulation results, although we shall consider this agreement in more detail later. For comparison, we also plot  $\tau_{\text{snr}}$  in Fig. 8. Although  $\tau_{\text{snr}}$  qualitatively tracks  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0} - \langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$  quite closely, we stress that  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  is the *mean* first passage time and is thus explicitly an average over all possible sources of fluctuations that can drive the perceptron’s activation below the firing threshold. The standard deviation  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$  by definition gives a measure of the range of the FPTs that arise. The FPT distribution can have a positive mean but with a large standard deviation, indicating that memory storage in these parameter regimes is subject to a high degree of variability. In contrast,  $\tau_{\text{snr}} = 0$  merely gives the impression that memory storage is impossible without, as it were, any “nuance”. Fig. 8 indicates that where memory storage is weak, SNR lifetimes are zero,

whereas positive MFPT memory lifetimes are subject to a high degree of variability, reflecting this weak encoding. We have made preliminary observations of this variability in FPT memory lifetimes before, but only in simulations of a filter-based model of synaptic plasticity (Elliott, 2016a).

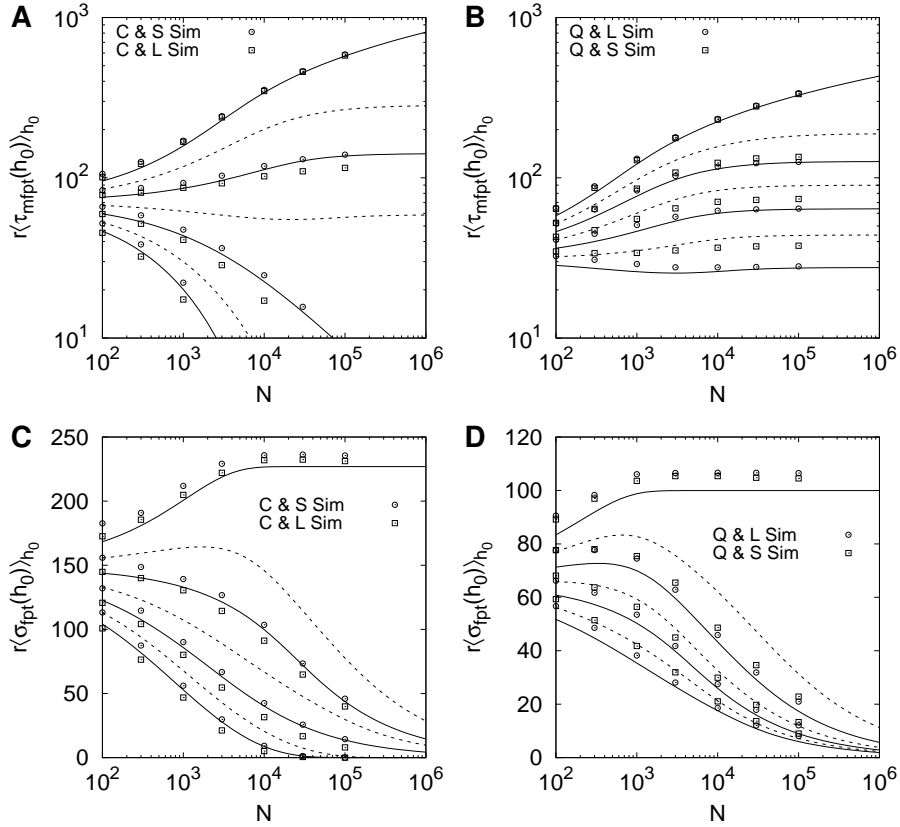
Finally, in Fig. 9 we plot  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  and  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$  separately for different choices of  $\theta$ . For the parameters used, for the  $\mathbb{C}$  model (for both  $\mathcal{S}$  and  $\mathcal{L}$ )  $\mu_0 = 0.02$ ; for the  $\mathbb{Q}$  &  $\mathcal{L}$  model  $\mu_0 \approx 0.041$ ; and for the  $\mathbb{Q}$  &  $\mathcal{S}$  model  $\mu_0 \approx 0.045$ . We vary  $\theta$  from 0 to 0.03 in increments of 0.005. For  $\theta = 0$  in Figs. 9A and 9B,  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  asymptotes to logarithmic growth in  $N$ , reflecting its approach to its mean field form  $\tau_{\text{mfpt}}(\mu_0)$  from Eq. (3.29). In Fig. 9A for  $0 < \theta < \mu_0 = 0.02$  and in Fig. 9B for all choices of  $\theta > 0$ , however,  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  asymptotes to an  $N$ -independent constant as  $N$  increases, with the asymptote coinciding again with the mean field value  $\tau_{\text{mfpt}}(\mu_0)$  in Eq. (3.25a) and being attained for  $N$  between  $10^4$  and  $10^5$ . In contrast, for the  $\mathbb{C}$  model for  $\theta > \mu_0$ ,  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  tends to zero as  $N$  increases. This occurs because the distribution of  $h_0$  becomes tightly focused around  $\mu_0$  as  $N$  increases, and for  $\mu_0 < \theta$  most of the distribution therefore falls below  $\theta$ . Identical behaviour would also occur for the  $\mathbb{Q}$  model when  $\mu_0 < \theta$ . For  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$  in Figs. 9C and 9D, we again see a difference between  $\theta = 0$  and  $\theta > 0$ . For  $\theta = 0$ ,  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$  asymptotes to a non-zero constant independent of  $N$  as  $N$  increases, but for any  $\theta > 0$ ,  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$  asymptotes to zero as  $N$  increases.

Comparing analytical and simulation results in Fig. 9, we see good qualitative agreement, but with numerical differences for increasing  $\theta$  and differences in trends between results for  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  and  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$ . For  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$ , simulation results for the  $\mathbb{C}$  &  $\mathcal{S}$  and  $\mathbb{Q}$  &  $\mathcal{L}$  models agree closely with their respective analytical results, but with expected differences for smaller  $N$ , as discussed above. For the  $\mathbb{C}$  model, simulation results with  $\Omega = \mathcal{L}$  systematically





**Figure 8:** MFPT memory lifetime  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  and one standard deviation (schematically  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$ ) in the FPT distribution around it as a function of  $N$  and  $\nu$ , for  $p = 0.1$  and  $\theta = 0$ . Solid lines show analytical results for FPT statistics while the dashed lines show for comparison the SNR memory lifetime  $\tau_{\text{snr}}$ ; data points correspond to simulation results. The one standard deviation region around  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  is shaded for clarity. (A) Analytical FPT statistics for the  $\mathbb{C}$  &  $\mathcal{S}$  model for  $\nu = 10$  as a function of  $N$ . (B) Same as A, except for the  $\mathbb{Q}$  &  $\mathcal{L}$  model. (C) Analytical FPT statistics for the  $\mathbb{C}$  &  $\mathcal{S}$  model for  $N = 10^4$  as a function of  $\nu$ . (D) Same as C, except for the  $\mathbb{Q}$  &  $\mathcal{L}$  model. The invalid region  $\nu > 39$  is shaded.



**Figure 9:** First passage time statistics  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  and  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$  as functions of  $N$  for different choices of  $\theta$ , for  $\nu = 10$  and  $p = 0.1$ . Moving from top to bottom in all panels, lines correspond to  $\theta = 0.000$ ,  $\theta = 0.005$ ,  $\theta = 0.010$ ,  $\theta = 0.015$ ,  $\theta = 0.020$ ,  $\theta = 0.025$  and  $\theta = 0.030$ ; line styles alternate between solid and dashed for clarity. Simulation data points are shown only for  $\theta = 0.000$ ,  $\theta = 0.010$ ,  $\theta = 0.020$  and  $\theta = 0.030$  and so correspond only to the solid lines. (A) and (B) Analytical results for  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  for, respectively, the C & S and Q & L models. (C) and (D) Same and A and B, except for  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$ .

underestimate those with  $\Omega = \mathcal{S}$ , with increasing differences for increasing  $\theta$ . For the  $\mathbb{Q}$  model, however, simulation results with  $\Omega = \mathcal{S}$  systematically overestimate those with  $\Omega = \mathcal{L}$ , with the same trend in  $\theta$ . From Fig. 9A for  $\theta = 0.03$ , these differences increase as  $N$  increases and so  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  decreases, while in Fig. 9B for  $\theta = 0.03$ , the differences asymptote as  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  asymptotes. These differences therefore reflect differences in the escape dynamics between models with  $\Omega = \mathcal{S}$  and with  $\Omega = \mathcal{L}$  when the escape times are shorter rather than when  $\theta$  is larger *per se*. Models with  $\Omega = \mathcal{L}$  have shorter escape times compared to models with  $\Omega = \mathcal{S}$ .

For  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$  in Figs. 9C and 9D, we again see extremely good qualitative agreement between analytical and simulation results and between the two sets of simulation results. However, for  $\theta = 0$ , we see a roughly 5% discrepancy between analytical results and the corresponding simulation results, even for large  $N$ . This discrepancy reduces as  $\theta$  increases and  $N$  increases, likely because  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$  goes to zero in this regime. Higher-order FPT statistics are inevitably much more sensitive to the precise details of the dynamics that drive escape through the threshold  $\theta$ . The approximations in ignoring jump processes are thus more significant for higher-order statistics. To confirm this, when we run simulations with  $p = 0.01$  (for say  $N = 10^5$  and  $\nu = 10$ ), reducing the affect of jump processes, the discrepancy between analytical and simulation results for  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$  falls to 0.5% (data not shown). Thus, it is not surprising to find these small discrepancies at the level of  $\langle \sigma_{\text{fpt}}(h_0) \rangle_{h_0}$ , especially for larger choices of  $p$ . Nevertheless, we have extremely good qualitative agreement and good quantitative agreement.

## 7 Discussion

In this paper we have continued our study of FPT-defined memory lifetimes in palimpsest memory models by extending our earlier analytical work on simple, binary-strength synapses (Elliott, 2014) to consider simple but general multi-state synapses with  $\nu$  states of synaptic strength. In order to push through these calculations in the general case, we have needed to assume that the vector of possible strengths  $\mathbf{\Omega}$  is a (left) eigenvector of the strength transition matrix  $\mathbb{M}$ , although we have seen that this requirement may be relaxed if  $\mathbf{\Omega}$  is sufficiently close to an eigenvector of  $\mathbb{M}$ . For  $\nu = 2$  and  $\nu = 3$  for standard choices of  $\mathbf{\Omega}$  and  $\mathbb{M}$ , this eigenvector condition is automatically satisfied. In this way, we have been able to derive the full FPT distribution for the perceptron’s activation to fall (to or) below firing threshold, which enables us to examine not only the MFPT memory lifetime but also the variance in the FPT-defined memory lifetime.

Although studying the dynamics of palimpsest memory lifetimes using a FPT approach is analytically considerably harder than the SNR method, the FPT approach has many advantages. From a purely theoretical perspective, the FPT approach avoids much of the arbitrariness that is inevitably associated with an SNR definition of memory lifetimes. Furthermore, FPT-defined lifetimes are independent (up to a trivial scaling factor of the rate,  $r$ ) of whether we consider a discrete time or a continuous time formulation, in contrast to SNR lifetimes, which depend strongly on the formulation. Perhaps most important, for the purposes of defining a measure that gauges memory lifetimes in real, biological systems, FPT memory lifetimes relate directly and immediately to a neuron’s firing or membrane potential properties. In contrast, the SNR represents the view of an “ideal observer” and mixes up statistics of dif-

fering orders. A neuron or indeed any non-ideal observer does not have access to the required statistics. Memory lifetimes defined via MFPTs require only a direct read-out of a neuron's membrane potential: is it above or below firing threshold? This is the only relevant determinant of whether or not a neuron's synaptic connections currently store a memory. The fine details of the relaxation of a neuron's mean memory signal back to equilibrium, especially when this signal is already extremely close to equilibrium, are for all practical purposes entirely irrelevant. This is demonstrated sharply by the catastrophic breakdown of the logarithmic growth of memory lifetimes when a perceptron's firing threshold cuts off such dynamics. Moreover, the membrane potential is in fact maintained in neurons with only a finite number of channels, and thus has only a finite rather than an infinite representational resolution. Single channel opening would typically be expected to generate a change in membrane potential of around 0.5 mV (Sigworth, 1980), so resolution below this limit is impossible. Therefore, any approach that implicitly requires incredibly fine, almost infinitesimal discriminations between a neuron's current membrane potential and its equilibrium potential imposes demands that cannot ever in practice be met. That is, the membrane potential has only a finite, discrete resolution, and this necessarily destroys any approach that tacitly requires an infinite representational capacity.

These issues aside, we have also seen that a FPT definition of memory lifetimes reveals serious limitations in the SNR approach. SNR lifetimes exhibit a pseudo-bifurcation, giving a minimum number of synapses,  $N$ , below which memory storage is (defined to be) unsuccessful. MFPT lifetimes exhibit no such dynamics and memory lifetimes are positive even for small values of  $N$ . Related to this, the logarithmic growth in SNR memory lifetimes is seen to be only asymptotically valid, and so valid only for large  $N$ . Extrapolating to

small  $N$  gives the misleading bifurcation-like behaviour on which the apparent minimum value of  $N$  is based. In addition, the use of SNR memory lifetimes can suggest optimality conditions, giving optimal parameter choices (e.g., an optimal value of  $\nu$ ) that maximise memory lifetimes. This is observed in complex models of synaptic plasticity (Elliott, 2016a) and also for a simple model (the  $\mathbb{C}$  model) considered here. However, with MFPT memory lifetimes, no such optimality is observed. Previously we saw this only in simulation (Elliott, 2016a) and so we could not be certain that the trends seen in simulation data would continue indefinitely. However, our analysis here demonstrates beyond doubt that MFPT memory lifetimes in the  $\mathbb{C}$  model do not exhibit the optimality seen with SNR memory lifetimes. Further, by examining the variance in the FPT-defined memory lifetimes, we see that where SNR memory lifetimes are small or zero, and so memory storage is defined to be unsuccessful, MFPT memory lifetimes are positive (and can be significantly so) but are subject to a high degree of variability. This variability indicates regimes of relatively weak rather than unsuccessful memory encoding, but even in such regions, we have argued before that at least a minimum of 16% of storage events occur robustly (Elliott, 2016a). Finally, as mentioned, a non-zero firing threshold (for balanced potentiation and depression processes) destroys the logarithmic growth in SNR memory lifetimes.

Given the analytical simplicity of the SNR approach to memory lifetimes compared to the considerably harder FPT approach, it is natural to wonder whether there are circumstances under which SNR memory lifetimes, despite all these shortcomings, are acceptable surrogates for FPT lifetimes. Analytically, we have seen that for  $\theta = 0$ ,  $\tau_{\text{snr}}$  and the asymptotic mean field form  $\tau_{\text{mfpt}}(\mu_0)$  of the full MFPT form  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  are identical up to additive constants. For  $\theta > 0$ , we can extend the SNR definition to include an accessibility

criterion (Elliott, 2014), which states that  $\tau_{\text{snr}}$  is defined as the solution of  $\mu(\tau_{\text{snr}}) = \theta + \sigma(\tau_{\text{snr}})$ , so that the mean memory signal at time  $t = \tau_{\text{snr}}$  must be above the standard deviation in the signal by the perceptron’s threshold. Eq. (5.5) is then modified to  $r(1 - \Lambda)\tau_{\text{snr}} = \frac{1}{2} \log_e [N' E^2 / (1 + \theta')^2]$ , where  $\theta' = \theta \sqrt{N'}$ . For  $\theta > 0$ , the asymptotic solution is just  $r(1 - \Lambda)\tau_{\text{snr}} = \log_e (E/\theta)$ , which is precisely  $\tau_{\text{mfpt}}(h_0)$  in Eq. (3.25a) for  $h_0 = \mu_0$ . In general, then,  $\tau_{\text{snr}}$  is an acceptable surrogate for  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  precisely when the full MFPT lifetime  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  is well approximated by its mean field form  $\tau_{\text{mfpt}}(\mu_0)$ . This is the case for  $N$  large enough, when the distribution of  $h_0$  is tightly focused around  $\mu_0$ . However, precisely how large  $N$  must be to be “large enough”, or how close  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  and  $\tau_{\text{mfpt}}(\mu_0)$  must be, are very model- and parameter-dependent. In particular, from Eqs. (3.29) and (3.30), we saw that the relevant parameter that controls the approach of  $\langle G(h_0, t) \rangle_{h_0}$  to its mean field form  $G(\mu_0, t)$  is  $\mu' = \mu_0 \sqrt{N/\langle \Omega^2 \rangle}$ . While  $\langle \Omega^2 \rangle$  does not vary too much for the models considered here,  $\mu_0$  varies significantly for the  $\mathbb{C}$  &  $\mathcal{S}$  model. As  $\mu_0$  drops to zero as  $\nu$  increases,  $N$  must be taken larger and larger for  $\mu'$  to remain large enough for the asymptotic limit to be achieved, although  $N$  cannot be taken much in excess of  $10^5$  and continue to be biologically relevant. However, for the  $\mathbb{Q}$  &  $\mathcal{L}$  model,  $\mu_0$  does not fall to zero as  $\nu$  increases, so  $\mu'$  is typically large enough for most parameters. In summary,  $\tau_{\text{snr}}$  is an acceptable surrogate for  $\langle \tau_{\text{mfpt}}(h_0) \rangle_{h_0}$  when the latter is well approximated by  $\tau_{\text{mfpt}}(\mu_0)$ , and this occurs for large  $N$  in general but for large  $\mu' = \mu_0 \sqrt{N/\langle \Omega^2 \rangle}$  in particular, so that the realisation of this limit for biologically relevant values of  $N$  can be strongly parameter- and model-dependent.

A perennial problem with palimpsest models of memory is that attempts to enhance memory lifetimes have appeared inevitably to lead to a decrease in the strength of the initial memory encoding. This Gordian knot is partially

although by no means completely severed in models based on integrative filtering mechanisms (Elliott & Lagogiannis, 2012; Elliott, 2016a). It is therefore surprising that in constructing the  $\mathbb{Q}$  model above purely for the purposes of writing down a transition matrix  $\mathbb{M} = \mathbb{I} + p\mathbb{Q}$  with eigenvector  $\boldsymbol{\Omega} = \boldsymbol{\mathcal{L}}$  with uniformly spaced synaptic strengths, entirely by accident we also happen to write down a model in which  $\mu_0 = \frac{p}{3} \frac{\nu+1}{\nu-1} > \frac{p}{3}$ . I.e. the initial signal  $\mu_0$  does not collapse to zero as  $\nu$  increases, in radical contrast to the  $\mathbb{C}$  model. Furthermore, memory storage in the  $\mathbb{Q}$  model is always robust as  $\nu$  increases, with low variability as revealed by the variance in the FPT distribution. It may be argued that as a biological model, changes in synaptic strength that are themselves strength-dependent, as required by the  $\mathbb{Q}$  matrix, are unnatural or implausible. However, there is ample evidence that the degree of synaptic plasticity can depend on synaptic strength (see, e.g., van Rossum *et al.* (2000) and references therein). The  $\mathbb{Q}$  matrix does impose an upper limit on  $\nu$  for technical reasons, so  $\nu$  cannot be increased arbitrarily to produce any desired MFPT memory lifetime. For reasons of analytical tractability, we have not considered truncating the elements of  $\mathbb{Q}$  so that it remains a stochastic matrix when  $\nu$  exceeds what would otherwise be its upper limit of around  $\frac{4}{p} - 1$ . It will, however, be fascinating to explore the dynamics of this model as it moves into this regime, in order to determine how memory lifetimes are modified.

The extension of our present analysis from simple to complex multistate synapses is a question of particular interest. It is possible to extend the analysis of simple, binary-strength synapses to complex, binary-strength synapses by integrating out the internal synaptic states and working purely in strength-change processes (Elliott, 2017). This is possible because for binary-strength synapses, we can continue to focus on the transitions in perceptron activation. Critical to our present analysis of simple multistate synapses was the freedom



to modify either  $\Omega$  or  $M$  so that the jump moments for changes in the perceptron activation could be determined even if we cannot reduce these changes for multistate synapses to a Markov process. With complex multistate synapses, the strength transition matrices become memory storage step-dependent and their spectra also become step-dependent. Whether it is possible to modify the approach here and make it work for complex multistate synapses remains to be seen.

**Acknowledgements:** I thank Paul Adams, State University of New York at Stony Brook, for discussions regarding the resolution of the membrane potential. I also thank an anonymous reviewer for insightful comments on the paper, specifically on the relationship between SNRs and MFPTs.

## 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 Comput. Biol.*, **4**, e1000230.
- 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 com-

- parison of integrate-and-express models of synaptic plasticity for memory lifetimes. *Neural Comput.*, **28**, 2393–2460.
- Elliott, T. (2017). Mean first passage memory lifetimes by reducing complex synapses to simple synapses. *Neural Comput.*, **29**, 1468–1527.
- Elliott, T., & Lagogiannis, K. (2012). The rise and fall of memory in a model of synaptic integration. *Neural Comput.*, **24**, 2604–2654.
- Fusi, S., Drew, P.J., & Abbott, L.F. (2005). Cascade models of synaptically stored memories. *Neuron*, **45**, 599–611.
- Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 2554–2558.
- Huang, Y., & Amit, Y. (2010). Precise capacity analysis in binary networks with multiple coding level inputs. *Neural Comput.*, **22**, 660–688.
- Huang, Y., & Amit, Y. (2011). Capacity analysis in multi-state synaptic models: A retrieval probability perspective. *J. Comput. Neurosci.*, **30**, 699–720.
- Lahiri, S., & Ganguli, S. (2013). A memory frontier for complex synapses. *Pages 1034–1042 of: Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., & Weinberger, K.Q. (eds), Advances in Neural Information Processing Systems 26.* Cambridge, MA: MIT Press.
- Leibold, C., & Kempster, R. (2006). Memory capacity for sequences in a recurrent network with biological constraints. *Neural Comput.*, **18**, 904–941.
- Leibold, C., & Kempster, R. (2008). Sparseness constrains the prolongation of memory lifetime via synaptic metaplasticity. *Cereb. Cortex*, **18**, 67–77.

- Nadal, J.P., Toulouse, G., Changeux, J.P., & Dehaene, S. (1986). Networks of formal neurons and memory palimpsests. *Europhys. Lett.*, **1**, 535–542.
- Parisi, G. (1986). A memory which forgets. *J. Phys. A: Math. and Gen.*, **19**, L617–L620.
- Rubin, D.D.B.D., & Fusi, S. (2007). Long memory lifetimes require complex synapses and limited sparseness. *Front. Comput. Neurosci.*, **1**, 7.
- Sigworth, F.J. (1980). The variance of sodium current fluctuations at the node of Ranvier. *J. Physiol.*, **307**, 97–129.
- 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.