Variations on the Theme of Synaptic Filtering: A Comparison of Integrate-and-Express Models of Synaptic Plasticity for Memory Lifetimes

Terry Elliott[1]

Department of Electronics and Computer Science,

University of Southampton,

Highfield,

Southampton, SO17 1BJ,

United Kingdom.

Running Title: Comparison of Integrate-and-Express Models.

June 26, 2016.

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

Abstract

Integrate-and-express models of synaptic plasticity propose that synapses integrate plasticity induction signals before expressing synaptic plasticity. By discerning trends in their induction signals, synapses can control destabilising fluctuations in synaptic strength. In a feedforward, perceptron framework with binary-strength synapses for associative memory storage, we have previously shown that such a filter-based model outperforms other, non-integrative, "cascade"-type models of memory storage in most regions of biologically-relevant parameter space. Here, we consider some natural extensions of our earlier filter model, including one specifically tailored to binary-strength synapses and one that demands a fixed, consecutive number of same-type induction signals rather than merely an excess before expressing synaptic plasticity. With these extensions, we show that filter-based models outperform non-integrative models in all regions of biologically-relevant parameter space except for a small sliver in which all models encode memories only weakly. In this sliver, which model is superior depends on the metric used to gauge memory lifetimes (whether a signal-to-noise ratio or a mean first passage time). After comparing and contrasting these various filter models, we discuss the multiple mechanisms and timescales that underlie both synaptic plasticity and memory phenomena, and suggest that multiple, different filtering mechanisms may operate at single synapses.

1 Introduction

Palimpsest models of associative memory store new memories by forgetting old ones (Nadal *et al.*, 1986; Parisi, 1986), overcoming the catastrophic forgetting of a Hopfield network at a critical memory loading (Hopfield, 1982). Palimpsests achieve this by imposing bounds on synaptic strength. Some limited experimental evidence provides support for binary- (Petersen *et al.*, 1998; O'Connor *et al.*, 2005) or even ternary-strength synapses (Montgomery & Madison, 2002, 2004), with such small numbers of strength states automatically limiting synaptic strengths. Several models of palimpsest memory systems based on binary-strength and more general, discrete, multi-level synapses exist (for example, Tsodyks, 1990; Amit & Fusi, 1994, Fusi *et al.*, 2005, Leibold & Kempter, 2006, 2008; Rubin & Fusi, 2007; Fusi & Abbott, 2007; Barrett & van Rossum, 2008; Huang & Amit, 2010, 2011). These palimpsest models are typically non-integrative, randomly expressing synaptic plasticity without regard for whether the resulting changes in synaptic strength deleteriously affect the recall of already-stored memories. The result is that the fidelity of recall of a memory falls monotonically, often exponentially fast over time. Much effort has been devoted to slowing down this monotonic decay in order to extend memory lifetimes.

In order to control destabilising fluctuations in synaptic strength in both a developmental (Elliott, 2008, 2011a,b; Elliott & Lagogiannis, 2009) and memory (Elliott & Lagogiannis, 2012; Elliott, 2014) context, we have proposed that synapses should integrate synaptic plasticity induction signals before expressing synaptic plasticity. In these "integrate-and-express" models, a synapse seeks to discern the trend in its induction signals before expressing synaptic plasticity. In doing so, it then tries not to express a change in synaptic strength that may not be mandated by the recent history of its induction signals. In a feedforward setting with a single perceptron, we showed that an integrative, filter-based synapse outperforms non-integrative models in terms of memory lifetimes in most regions of biologically-relevant parameter space (Elliott & Lagogiannis, 2012). Furthermore, the ongoing synaptic plasticity induced by the storage of further memories, leading to the monotonic fall of a memory's trace in non-integrative models, actually causes an initial rise in the fidelity of recall of a memory in our model (Elliott & Lagogiannis, 2012).

Here, we extend our previous work to consider a range of natural extensions of the particular filter model that we considered before (Elliott & Lagogiannis, 2012), comparing and contrasting various filter-based models in terms of memory performance. The model that we considered before was not particular to binary-strength synapses but specifically allowed immediate generalisation to multi-state synapses. However, a version of the model exists that is tailored to saturated upper and lower strength states and is thus expected to be best-suited to binary-strength synapses where there are no intermediate strength states between the two saturated states. We also consider extensions of this and our earlier model in which the filter injection protocol is modified in order to gauge the sensitivity of our results to this protocol. All four variants, although controlling fluctuations, are nevertheless still subject to them. We therefore also consider a model that expresses plasticity only after a fixed, consecutive number of same-type induction signals arise, rather than some excess number that can occur in any order as in the other models. Such a consecutive sequence can arise via a fluctuation, but its probability is vastly suppressed.

Our paper is organised as followed. In section 2 we introduce our general approach and discuss in some detail our earlier work, which enables us to set up the machinery required later. We then introduce, in section 3, the five filter models that we compare. Next, in section 4, we derive the mean memory signal in all five models, using as general a method as possible. Then we compare memory lifetimes in all five filter models in section 5. In section 6 we discuss Lahiri & Ganguli's (2013) derivation of bounds on memory signal envelopes in relation to our current work. Finally, in section 7, we discuss our results and evaluate the overall approach to memory storage based on palimpsests.

2 General Approach and Earlier Results

Before proceeding with a derivation of results for the extensions of the filter model that we examined earlier (Elliott & Lagogiannis, 2012), for orientation and to define our general approach, we first provide a detailed summary of our earlier work. This allows us to set up the basic machinery, of which we will then make extensive use in later sections.

2.1 Perceptron Formulation in Continuous Time

A single perceptron with N synapses of binary strengths $S_i(t) \in \{-1, +1\}$, where $i = 1, \ldots, N$ indexes the synapses and t is time, is required to store a sequence of "memories" $\boldsymbol{\xi}^\alpha$, $\alpha = 0, 1, 2, \ldots$. The components ξ_i^α take values ± 1 with probability $\frac{1}{2}$ independent of both i and α, so that the ξ_i^α are uncorrelated both across synapses and between memories. In a discrete time formalism, memory α is stored at time $t = \alpha$. Biologically, however, discrete-time storage is unrealistic and instead memories are more plausibly stored as a continuous-time process. The simplest continuous-time process to consider is the Poisson process, so we shall take the memories to be stored as a Poisson process of rate r. For the models that we consider here, we may without loss of generality take $r = 1$ Hz, as r can be reinstated in formulae with the simple replacement $t \to r\,t$, but we retain r in places for the purposes of clarity. Although memories are stored as a Poisson process, we take the first memory $\boldsymbol{\xi}^0$ always to be stored at $t = 0^-$ s. We use $t = 0^-$ s rather than $t = 0$ s so that we may refer to the time immediately after the storage of $\boldsymbol{\xi}^0$ as $t = 0$ s.

The memory $\boldsymbol{\xi}^0$ is the "tracked" memory and we are interested in the fidelity of recall of this tracked memory by the perceptron as the later memories $\boldsymbol{\xi}^\alpha$, $\alpha \geq 1$, are stored. Changes in the synaptic strengths $S_i(t)$ induced by this ongoing, subsequent memory storage will over time degrade and eventually wash out the tracked memory, $\boldsymbol{\xi}^0$. We gauge the fidelity of recall of $\boldsymbol{\xi}^0$ by tracking the perceptron's activation in response to the tracked memory. With input $\boldsymbol{x}$, $x_i \in \{-1, +1\}$, to its N synapses, the perceptron's activation is of the standard form,

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

We are interested only in this activation and not its thresholding leading to the perceptron's two-level output: we only need to know whether the perceptron is above or below firing threshold. We define the tracked memory signal to be the perceptron's activation upon re-presentation of memory $\boldsymbol{\xi}^0$, so just $h_{\boldsymbol{\xi}^0}(t)$. We refer to this for simplicity as just the memory signal and write $h(t) = h_{\boldsymbol{\xi}^0}(t)$. With uncorrelated memories and binary strengths ±1, memory $\boldsymbol{\xi}^0$ is still stored by the perceptron at some later time t provided that $h(t) > 0$. We are not, however, interested in the dynamics of $h(t)$ for any particular realisation of the sequence of memories $\boldsymbol{\xi}^\alpha$, $\alpha \geq 0$, but rather in the statistics of $h(t)$ averaged over all possible sequences. We define the mean and variance of $h(t)$ to be

$$\mu(t) = \mathsf{E}[h(t)], \tag{2.2}$$

$$\sigma(t)^2 = \mathsf{Var}[h(t)], \tag{2.3}$$

where $\mathsf{E}[\cdot]$ and $\mathsf{Var}[\cdot]$ denote the expectation value and variance, respectively, and the ratio $\mu(t)/\sigma(t)$ is the signal-to-noise ratio (SNR) of the perceptron's activation. We will discuss this definition of the perceptron's SNR in relation to the SNR based on the "ideal observer" approach in section 2.3.

The statistics of $h(t)$ may be used to define the lifetime of memory $\boldsymbol{\xi}^0$. Many alternative approaches to gauging memory lifetimes exist (see, for example, Tsodyks, 1990; Amit &Fusi, 1994; Leibold & Kempter, 2006; Huang & Amit, 2010; Elliott, 2014). The SNR $\mu(t)/\sigma(t)$ (Tsodyks, 1990) is in many respects the easiest to use, but a mean first passage time (MFPT) is superior although analytically much harder to study (Elliott, 2014). The SNR memory lifetime is defined as the time $\tau_{\rm snr}$ at which the SNR reaches unity, $\mu(\tau_{\rm snr})/\sigma(\tau_{\rm snr}) = 1$, and because of its simplicity, we shall almost exclusively use this definition here. For an MFPT definition of memory lifetimes, we consider the first passage time for any particular realisation of the memory signal $h(t)$ to fall below firing threshold, which here is taken to be zero. The mean over all such realisations then defines the MFPT memory lifetime, $\tau_{\rm mfpt}$ (Elliott, 2014). We shall use this definition occasionally in order to compare $\tau_{\rm mfpt}$ and $\tau_{\rm snr}$.

With input $\boldsymbol{x} = \boldsymbol{\xi}^\alpha$ for some α, the input to the perceptron is of zero mean and variance $1/N$. We may therefore regard a synapse with strength $+1$ as a "strong" synapses and one with strength -1 as "weak" (and not as excitatory and inhibitory) because we can always add an overall constant to these strengths and change the perceptron's firing threshold to compensate.

It is not necessary to specify a target perceptron output for memory $\boldsymbol{\xi}^\alpha$ in a feedforward framework because we may without loss of generality consider instead storing $-\boldsymbol{\xi}^\alpha$ instead of $+\boldsymbol{\xi}^\alpha$. With this convention, ξ_i^α is the synaptic plasticity induction signal to synapse i upon the storage of memory α. With $\xi_i^\alpha = +1$, the synapse should potentiate (strengthen) and with $\xi_i^\alpha = -1$, it should depress (weaken).

2.2 Filter-Based Synaptic Plasticity

We now discuss how synapses response to these synaptic plasticity induction signals for the specific filter model that we studied earlier (Elliott & Lagogiannis, 2012). We proposed that a synapse integrates these signals by modifying an internal filter state, with synaptic plasticity being expressed only when the filter reaches threshold. Potentiating induction signals increment the filter state and potentiation is expressed only when the filter state reaches its upper threshold, which we denote by $+\Theta_+$, with $\Theta_+ > 0$. Correspondingly, depressing induction signals decrement the filter state and depression is expressed only when the filter state reaches its lower threshold, which we denote by $-\Theta_-$, with $\Theta_- > 0$. When the filter state reaches threshold, it is returned to the zero state in this model. For reasons that we will explain in relation to other filter models, we call this particular filter model the A0 filter. Let filter states be labelled by letters such as I and J, with $I, J \in \{-(\Theta_- - 1), \ldots, +(\Theta_+ - 1)\}$. We do not include $\pm\Theta_\pm$ in this set because when a filter reaches these threshold values, it immediately leaves them. We may summarise the transitions in filter state as follows:

$$\left.\begin{array}{l} \xi_i^\alpha = +1 \quad \Rightarrow \quad \begin{cases} I \mapsto I+1 & \text{for } I < +(\Theta_+ - 1), \\ I \mapsto 0 \;\&\; \Uparrow & \text{for } I = +(\Theta_+ - 1), \end{cases} \\ \\ \xi_i^\alpha = -1 \quad \Rightarrow \quad \begin{cases} I \mapsto I-1 & \text{for } I > -(\Theta_- - 1), \\ I \mapsto 0 \;\&\; \Downarrow & \text{for } I = -(\Theta_- - 1). \end{cases} \end{array}\right\} \tag{2.4}$$

The arrows $\Uparrow$ and $\Downarrow$ indicate the expression of synaptic plasticity, either potentiation ($\Uparrow$) or depression ($\Downarrow$), upon reaching filter threshold. Of course, a strong synapse cannot express potentiation and a weak synapse cannot express depression, so in these cases no change in synaptic strength occurs although for this filter model, the filter state is nevertheless returned to the $I = 0$ state. Because we employ balanced, equiprobable induction signals, $\mathsf{Prob}[\xi_i^\alpha = \pm 1] = \frac{1}{2}$, we may consider only symmetric filters, with the upper and lower thresholds equidistant from the $I = 0$ state, or $\Theta_\pm = \Theta$, where Θ refers to the common upper and lower threshold values. Fig. 1A shows a $\Theta = 4$ filter as a continuous-time Markov process with potentiating and depressing induction signals arising with rates $r_\pm$, which here we take to be $r_\pm = \frac{1}{2} r$, i.e. the Poisson rate r of memory storage multiplied by $\mathsf{Prob}[\xi_i^\alpha = \pm 1] = \frac{1}{2}$.

Thinking of the filter state as represented by a $(2\,\Theta - 1)$-dimensional vector, let the $(2\,\Theta - 1) \times (2\,\Theta - 1)$ matrix $\mathbb{S}^+$ be the matrix that increments the filter state but without implementing the upper threshold process, and let the matrix $\mathbb{T}^+$ implement only this upper threshold process. For $\Theta = 3$, for example,

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

The matrix $\mathbb{S}^+$ is just a shift operator without wrap-around, while $\mathbb{T}^+$ sends the $I = +(\Theta - 1)$ state to $I = 0$. Let $\mathbb{S}^-$ and $\mathbb{T}^-$ be the corresponding matrices for decrementing filter states and implementing only the lower threshold process. For $\Theta = 3$,

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

Clearly $\mathbb{S}^- = (\mathbb{S}^+)^{\mathrm{T}}$, where the superscript T denotes the transpose, and $\mathbb{T}^-$ sends the $I = -(\Theta - 1)$ state to $I = 0$. The joint strength and filter state of a synapse is represented by a $2(2\,\Theta - 1)$-dimensional vector, and we define the first block of $(2\,\Theta - 1)$ entries of such a vector to correspond to the filter state when the synapse is weak and the second block of $(2\,\Theta - 1)$ entries to correspond to the filter state when the synapse is strong. Let the $2(2\,\Theta - 1) \times 2(2\,\Theta - 1)$ matrices $\mathbb{M}_2^{\pm}$ implement, respectively, potentiating and depressing steps. We use a subscript "2" on such $2(2\,\Theta - 1) \times 2(2\,\Theta - 1)$ matrices so that they are easily distinguished from their $(2\,\Theta - 1) \times (2\,\Theta - 1)$ submatrices. Then, schematically, we have

$$\mathbb{M}_2^+ = \left(\begin{array}{c|c} \mathbb{S}^+ & \mathbb{O} \\ \hline \mathbb{T}^+ & \mathbb{S}^+ + \mathbb{T}^+ \end{array}\right) \quad \text{and} \quad \mathbb{M}_2^- = \left(\begin{array}{c|c} \mathbb{S}^- + \mathbb{T}^- & \mathbb{T}^- \\ \hline \mathbb{O} & \mathbb{S}^- \end{array}\right), \qquad (2.5)$$

where $\mathbb{O}$ denotes a $(2\,\Theta - 1) \times (2\,\Theta - 1)$ matrix with entries of zero everywhere. The submatrix $\mathbb{T}^+$ in the lower left block of $\mathbb{M}_2^+$ indicates a change in strength from weak to strong via the upper filter threshold process, while its appearance in the lower right block indicates an upper filter threshold process but without a change in strength because the synapse is already strong. Similarly for $\mathbb{T}^-$ in $\mathbb{M}_2^-$.

Let the matrix

$$\mathbb{M}_2 = \tfrac{1}{2}\left(\mathbb{M}_2^+ + \mathbb{M}_2^-\right) \tag{2.6}$$

represent a potentiating event with probability $\mathsf{Prob}[\xi_i^\alpha = +1] = \frac{1}{2}$ and a depressing event with probability $\mathsf{Prob}[\xi_i^\alpha = -1] = \frac{1}{2}$. We have that, schematically,

$$\mathbb{M}_2 = \frac{1}{2}\left(\begin{array}{c|c} \mathbb{S}^+ + \mathbb{S}^- + \mathbb{T}^- & \mathbb{T}^- \\ \hline \mathbb{T}^+ & \mathbb{S}^+ + \mathbb{S}^- + \mathbb{T}^+ \end{array}\right). \tag{2.7}$$

The matrix $\mathbb{M}_2$ averages over both induction signals at any given synapse and is therefore the relevant matrix for averaging over the later, non-tracked memories $\boldsymbol{\xi}^\alpha$, $\alpha \geq 1$. As $t \to \infty$, the joint probability distribution of filter and strength states approaches the equilibrium (or stationary or asymptotic) distribution defined by the (right) eigenvector of $\mathbb{M}_2$ with unit eigenvalue. Defining first the $(2\,\Theta - 1)$-dimensional vector $\boldsymbol{A}$ with components[2]

$$A_I = \frac{1}{\Theta^2}\left(\Theta - |I|\right), \tag{2.8}$$

then the equilibrium eigenvector of $\mathbb{M}_2$, normalised to a probability distribu-

[2] We index $(2\,\Theta - 1)$-dimensional vector components for convenience according to the filter state $I = -(\Theta-1), \ldots, +(\Theta-1)$. Similarly, $(2\,\Theta-1)\times(2\,\Theta-1)$ matrix elements are also indexed in this way.

tion, is just, schematically,

$$\boldsymbol{A}_2 = \tfrac{1}{2}\big(\boldsymbol{A}^{\mathrm{T}}\big|\boldsymbol{A}^{\mathrm{T}}\big)^{\mathrm{T}}, \tag{2.9}$$

so that both strength states are equiprobable in equilibrium and the probability distribution of filter states in equilibrium is governed by $\boldsymbol{A}$ regardless of the strength state. Again, we use a subscript "2" on such $2(2\,\Theta-1)$-dimensional vectors to distinguish them clearly from the $(2\,\Theta-1)$-dimensional vectors that form their first or second blocks of entries.

It is against the background of the equilibrium distribution $\boldsymbol{A}_2$ that the tracked memory $\boldsymbol{\xi}^0$ is stored at time $t = 0^-$ s. For synapses initially experiencing $\boldsymbol{\xi}_i^0 = +1$, their probability distribution at $t = 0$ s is $\mathbb{M}_2^+\boldsymbol{A}_2$, while for synapses with $\boldsymbol{\xi}_i^0 = -1$, their distribution at $t = 0$ s is $\mathbb{M}_2^-\boldsymbol{A}_2$. Let $\boldsymbol{n} = (1, \ldots, 1)^{\mathrm{T}}$ be a $(2\,\Theta-1)$-dimensional vector with all components unity, and let $\boldsymbol{\Omega}_2 = \big(-\boldsymbol{n}^{\mathrm{T}}\big|+\boldsymbol{n}^{\mathrm{T}}\big)^{\mathrm{T}}$ represent the vector of synaptic strengths associated with the joint distribution of filter and strength states. The mean initial perceptron signal $\mu(0)$ is then just

$$\mu(0) = \tfrac{1}{2}\,\boldsymbol{\Omega}_2^{\mathrm{T}}\big(\mathbb{M}_2^+ - \mathbb{M}_2^-\big)\boldsymbol{A}_2, \tag{2.10}$$

since $\mathsf{Prob}[\xi_i^\alpha = \pm 1] = \frac{1}{2}$. For the subsequent storage of memories $\boldsymbol{\xi}^\alpha$, $\alpha \geq 1$, the joint distribution of filter and strength states evolves according to the superposed matrix $\mathbb{M}_2$. After the storage of memory α, $\alpha \geq 1$, $\xi_i^0 = \pm 1$

synapses are distributed as $\mathbb{M}_2^\alpha \, \mathbb{M}_2^\pm \boldsymbol{A}_2$, where, to be clear, $\mathbb{M}_2^\alpha$ is $\mathbb{M}_2$ raised to the power of α. In discrete time, with $t = \alpha$ only, we would then have

$$\mu(\alpha) = \tfrac{1}{2} \, \boldsymbol{\Omega}_2^{\mathrm{T}} \mathbb{M}_2^\alpha \big(\mathbb{M}_2^+ - \mathbb{M}_2^- \big) \boldsymbol{A}_2, \tag{2.11}$$

while in continuous time, we have

$$\mu(t) = \tfrac{1}{2} \, \boldsymbol{\Omega}_2^{\mathrm{T}} \left[\sum_{\alpha=0}^{\infty} \frac{(rt)^\alpha}{\alpha!} e^{-rt} \, \mathbb{M}_2^\alpha \right] \big(\mathbb{M}_2^+ - \mathbb{M}_2^- \big) \boldsymbol{A}_2 \tag{2.12a}$$

$$= \tfrac{1}{2} \, \boldsymbol{\Omega}_2^{\mathrm{T}} \big[\exp \left(rt \, \mathbb{G}_2 \right) \big] \big(\mathbb{M}_2^+ - \mathbb{M}_2^- \big) \boldsymbol{A}_2 \tag{2.12b}$$

$$= \tfrac{1}{2} \, \boldsymbol{\Omega}_2^{\mathrm{T}} \, \mathbb{P}_2(t) \big(\mathbb{M}_2^+ - \mathbb{M}_2^- \big) \boldsymbol{A}_2, \tag{2.12c}$$

where $\mathbb{P}_2(t) = \exp \left(rt \, \mathbb{G}_2 \right)$ is the joint filter and strength state probability transition matrix over time t, with $\mathbb{G}_2 = \mathbb{M}_2 - \mathbb{I}_2$, where $\mathbb{I}_2$ is the $2(2\,\Theta - 1) \times 2(2\,\Theta - 1)$ identity matrix.

With symmetric filters (for which $\Theta_\pm = \Theta$) and equiprobable induction signals, the two distributions $\mathbb{M}_2^\pm \boldsymbol{A}_2$ are mirror images of each other, by which we mean that the components of $\mathbb{M}_2^+ \boldsymbol{A}_2$ read in reverse order are identical to the components of $\mathbb{M}_2^- \boldsymbol{A}_2$ read is standard order. For example, for $\Theta = 3$, we have

$$\mathbb{M}_2^+ \boldsymbol{A}_2 = \tfrac{1}{18} (0, 1, 2, 3, 2 \,|\, 0, 1, 4, 3, 2)^{\mathrm{T}},$$

$$\mathbb{M}_2^- \boldsymbol{A}_2 = \tfrac{1}{18} (2, 3, 4, 1, 0 \,|\, 2, 3, 2, 1, 0)^{\mathrm{T}}.$$

Defining the matrix $\mathbb{R}_2$ to be the operator that reverses the order of the components of such vectors,

$$\mathbb{R}_2 = \left(\begin{array}{c|c} \mathbb{O} & \mathbb{R} \\ \hline \mathbb{R} & \mathbb{O} \end{array} \right), \tag{2.13}$$

where the $(2\,\Theta - 1) \times (2\,\Theta - 1)$ matrix $\mathbb{R}$ has entries of unity on the anti-diagonal rather than the diagonal and zeros elsewhere,

$$\mathbb{R} = \begin{pmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{pmatrix}, \tag{2.14}$$

we have that $\mathbb{R}_2 \mathbb{M}_2^+ \boldsymbol{A}_2 = \mathbb{M}_2^- \boldsymbol{A}_2$. Of course, we also have that $\mathbb{R}_2^2 = \mathbb{I}_2$. Now, the distribution $\mathbb{M}_2^+ \boldsymbol{A}_2$ arises from synapses experiencing $\xi_i^0 = +1$, while $\mathbb{M}_2^- \boldsymbol{A}_2$ arises from those experiencing $\xi_i^0 = -1$. Synapses with $\xi_i^0 = +1$ contribute positively to $h(t) = \frac{1}{N} \sum_{i=1}^{N} \xi_i^0 \, S_i(t)$, while synapses with $\xi_i^0 = -1$ contribute negatively. For synapses with $\xi_i^0 = -1$, the roles of strong and weak synaptic strength states are reversed in $h(t)$. However, the mirror-image structure of $\mathbb{M}_2^- \boldsymbol{A}_2$ also ensures that the roles of strong and weak synapses are reversed, and also filter states are reversed, with upper and lower thresholds swapping around. The reversal of strong and weak synaptic strengths in their contribution to $h(t)$ for $\xi_i^0 = -1$ is therefore exactly cancelled out by the mirror-image structure of $\mathbb{M}_2^- \boldsymbol{A}_2$ compared to $\mathbb{M}_2^+ \boldsymbol{A}_2$ for synapses with $\xi_i^0 = +1$. Therefore, defining the variables $\widetilde{S}_i(t) = \xi_i^0 S_i(t)$, all N variables $\widetilde{S}_i(0)$ are identically-distributed random variables. We may see this explicitly be computing $\mathsf{Prob}[\widetilde{S}_i(0) = \pm 1]$. For $\xi_i^0 = +1$, we have

$$\begin{aligned}\mathsf{Prob}[S_i(0) = +1] &= \frac{1}{2}\left(1 + \frac{1}{\Theta^2}\right),\\ \mathsf{Prob}[S_i(0) = -1] &= \frac{1}{2}\left(1 - \frac{1}{\Theta^2}\right),\end{aligned}$$

so that $\mathsf{Prob}[\widetilde{S}_i(0) = \pm 1] = \frac{1}{2}\left(1 \pm \frac{1}{\Theta^2}\right)$, while for $\xi_i^0 = -1$, we have

$$\begin{aligned}\mathsf{Prob}[S_i(0) = +1] &= \frac{1}{2}\left(1 - \frac{1}{\Theta^2}\right),\\ \mathsf{Prob}[S_i(0) = -1] &= \frac{1}{2}\left(1 + \frac{1}{\Theta^2}\right),\end{aligned}$$

so that $\mathsf{Prob}[\widetilde{S}_i(0) = \pm 1] = \frac{1}{2}\left(1 \pm \frac{1}{\Theta^2}\right)$, and thus the distribution of $\widetilde{S}_i(0)$ is independent of the sign of ξ_i^0, as stated. Moreover, because the same superposed transition matrix $\mathbb{M}_2 = \frac{1}{2}\left(\mathbb{M}_2^+ + \mathbb{M}_2^-\right)$ is applied to all synapses for the storage of subsequent memories $\boldsymbol{\xi}^\alpha$, $\alpha \geq 1$, all N variables $\widetilde{S}_i(t)$ remain for all times $t \geq 0$ identically-distributed random variables. The memory signal $h(t) = \frac{1}{N}\sum_{i=1}^N \widetilde{S}_i(t)$ is therefore the sum over N identically-distributed random variables, so has mean and variance,

$$\mu(t) = \mathsf{E}\big[\widetilde{S}(t)\big], \tag{2.15}$$

$$\begin{aligned}\sigma(t)^2 &= \frac{1}{N}\mathsf{Var}\big[\widetilde{S}(t)\big] + \left(1 - \frac{1}{N}\right)\mathsf{Cov}[\widetilde{S}_i(t), \widetilde{S}_j(t)]\\ &= \frac{1}{N}\left[1 - \mu(t)^2\right] + \left(1 - \frac{1}{N}\right)\mathrm{Cov}(t),\end{aligned} \tag{2.16}$$

where $\mathsf{E}\big[\widetilde{S}(t)\big]$ and $\mathsf{Var}\big[\widetilde{S}(t)\big]$ are the mean and variance, respectively, of any one of the $\widetilde{S}_i(t)$ variables, and $\mathsf{Cov}[\widetilde{S}_i(t), \widetilde{S}_j(t)] = \mathrm{Cov}(t)$ is the covariance between any (distinct) pair of them.

We may see explicitly that $\mu(t)$ is the sum over N identically-distributed random variables by transforming $-\boldsymbol{\Omega}_2^{\mathrm{T}}\mathbb{M}_2^{\alpha}\,\mathbb{M}_2^{-}\boldsymbol{A}_2$ into $+\boldsymbol{\Omega}_2^{\mathrm{T}}\mathbb{M}_2^{\alpha}\,\mathbb{M}_2^{+}\boldsymbol{A}_2$ using the $\mathbb{R}_2$ matrix:

$$\begin{aligned}
-\boldsymbol{\Omega}_2^{\mathrm{T}}\mathbb{M}_2^{\alpha}\,\mathbb{M}_2^{-}\boldsymbol{A}_2 &= -\boldsymbol{\Omega}_2^{\mathrm{T}}\,\mathbb{R}_2^2\,\mathbb{M}_2^{\alpha}\,\mathbb{R}_2^2\,\mathbb{M}_2^{-}\boldsymbol{A}_2 \\
&= +\boldsymbol{\Omega}_2^{\mathrm{T}}\,\mathbb{R}_2\,\mathbb{M}_2^{\alpha}\,\mathbb{R}_2\,\mathbb{M}_2^{+}\boldsymbol{A}_2 \\
&= +\boldsymbol{\Omega}_2^{\mathrm{T}}\,(\mathbb{R}_2\mathbb{M}_2\mathbb{R}_2)^{\alpha}\,\mathbb{M}_2^{+}\boldsymbol{A}_2 \\
&= +\boldsymbol{\Omega}_2^{\mathrm{T}}\,\mathbb{M}_2^{\alpha}\,\mathbb{M}_2^{+}\boldsymbol{A}_2,
\end{aligned}$$

where we have used $\mathbb{R}_2\boldsymbol{\Omega}_2 = -\boldsymbol{\Omega}_2$, $\mathbb{R}_2\mathbb{M}_2^{-}\boldsymbol{A}_2 = \mathbb{M}_2^{+}\boldsymbol{A}_2$, and we have the key property

$$\mathbb{R}_2\mathbb{M}_2\mathbb{R}_2 = \mathbb{M}_2, \tag{2.17}$$

for $\mathbb{M}_2 = \frac{1}{2}\left(\mathbb{M}_2^{+}+\mathbb{M}_2^{-}\right)$ with $\mathsf{Prob}[\xi_i^0 = \pm 1] = \frac{1}{2}$. This crucial property follows from the fact that potentiation and depression processes are treated identically and symmetrically for $\mathsf{Prob}[\xi_i^0 = \pm 1] = \frac{1}{2}$, so that both synaptic strength states and all filter states are treated identically and symmetrically. Thus,

$$\mu(\alpha) = \tfrac{1}{2}\,\boldsymbol{\Omega}_2^{\mathrm{T}}\mathbb{M}_2^{\alpha}\left(\mathbb{M}_2^{+}-\mathbb{M}_2^{-}\right)\boldsymbol{A}_2 \equiv +\boldsymbol{\Omega}_2^{\mathrm{T}}\mathbb{M}_2^{\alpha}\,\mathbb{M}_2^{+}\boldsymbol{A}_2 \equiv -\boldsymbol{\Omega}_2^{\mathrm{T}}\mathbb{M}_2^{\alpha}\,\mathbb{M}_2^{-}\boldsymbol{A}_2, \tag{2.18}$$

or

$$\mu(t) = \tfrac{1}{2}\,\boldsymbol{\Omega}_2^{\mathrm{T}}\mathbb{P}_2(t)\big(\mathbb{M}_2^{+}-\mathbb{M}_2^{-}\big)\boldsymbol{A}_2 \equiv +\boldsymbol{\Omega}_2^{\mathrm{T}}\,\mathbb{P}_2(t)\,\mathbb{M}_2^{+}\boldsymbol{A}_2 \equiv -\boldsymbol{\Omega}_2^{\mathrm{T}}\,\mathbb{P}_2(t)\,\mathbb{M}_2^{-}\boldsymbol{A}_2. \tag{2.19}$$

But $\pm\boldsymbol{\Omega}_2^{\mathrm{T}}\mathbb{P}_2(t)\,\mathbb{M}_2^{\pm}\boldsymbol{A}_2 \equiv \mathsf{E}[\widetilde{S}(t)]$, so $\mu(t)$ has indeed reduced to the mean of a single $\widetilde{S}(t)$ variable from a sum over N of them. This equivalence argument between synapses with $\xi_i^0 = +1$ and $\xi_i^0 = -1$ in terms of their contributions to the tracked memory signal $h(t)$ when $\mathsf{Prob}[\xi_i^0 = \pm 1] = \frac{1}{2}$ means that we may simply consider $\xi_i^0 \equiv +1\ \forall i$ and use only $\mathbb{M}_2^{+}$ as the transition matrix applied to the equilibrium distribution $\boldsymbol{A}_2$ for the initial storage of $\boldsymbol{\xi}^0$.

In discrete time, because the induction signals ξ_i^α are uncorrelated between synapses and across memories, the covariance term in Eq. (2.16) vanishes identically and we are left with $\sigma(\alpha)^2 = \left[1 - \mu(\alpha)^2\right]/N$ at times $t = \alpha$. However, in continuous time, despite the ξ_i^α begin uncorrelated, driving memory storage as a continuous-time random process induces correlations between synaptic strengths so that in general $\mathrm{Cov}(t)$ does not vanish. In order to compute

$$\mathrm{Cov}(t) = \mathsf{E}\big[\widetilde{S}_i(t)\,\widetilde{S}_j(t)\big] - \mu(t)^2 \tag{2.20}$$

for any (distinct) pair of synapses i and j, we must employ the probability transition matrix that describes the joint evolution of the probability distribution for any pair of synapses. We therefore need the tensor product, and we may write this two-synapse transition matrix $\mathbb{P}_{2\otimes 2}(t)$ as

$$\begin{aligned}\mathbb{P}_{2\otimes 2}(t) &= \sum_{\alpha=0}^{\infty} \frac{(rt)^\alpha}{\alpha!} e^{-rt} \left(\mathbb{M}_2 \otimes \mathbb{M}_2\right)^\alpha \\ &= \exp\left[rt\left(\mathbb{M}_2 \otimes \mathbb{M}_2 - \mathbb{I}_2 \otimes \mathbb{I}_2\right)\right]. \end{aligned} \tag{2.21}$$

Because $\mathbb{P}_{2\otimes 2}(t) \neq \mathbb{P}_2(t) \otimes \mathbb{P}_2(t)$, we have that $\mathrm{Cov}(t) \neq 0$ in continuous time. In fact, $\mathrm{Cov}(t) > 0$ for $t > 0$. Driving memory storage in continuous time therefore increases $\sigma(t)^2$ and so decreases the SNR $\mu(t)/\sigma(t)$, although $\mathrm{Cov}(t) \to 0$ as $t \to \infty$ so the decreased SNR $\mu(t)/\sigma(t)$ only significantly affects memory lifetimes when they are short. Nevertheless, $\mathrm{Cov}(t) > 0$ is an important additional source of noise that is not captured in a discrete time framework.

Before proceeding to summarise our earlier derivation of an explicit formula for $\mu(t)$, we are now in a position to compare our definition of the perceptron SNR with the ideal observer SNR.

2.3 Perceptron versus Ideal Observer Approach

We have defined the SNR $\mu(t)/\sigma(t)$ on the basis of the perceptron's activation when (re-)presented with the tracked memory $\boldsymbol{\xi}^0$,

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

Although driving memory storage as a continuous-time process is, we believe, more realistic that using a discrete-time process, the non-zero (and positive) covariance term in $\sigma(t)^2$ in continuous time essentially forces us to consider continuous time, since otherwise we would be led to under-estimate or ignore noise terms and so over-estimate memory lifetimes in biologically-relevant scenarios. By focusing on the perceptron's activation statistics, we are also naturally led to consider an MFPT definition of memory lifetimes rather than one based on SNRs (Elliott, 2014). Specifically, we may consider the first passage time for $h(t)$ to fall below firing threshold for any particular realisation of the memories $\boldsymbol{\xi}^\alpha$, and thus the MFPT, which is averaged over all possible realisations, and define memory lifetimes accordingly. Such a definition has several advantages over the SNR $\mu(t)/\sigma(t)$, including the fact that MFPT memory lifetimes are identical is discrete and continuous time, while SNR memory lifetimes are not (Elliott, 2014).

In the ideal observer approach, it may appear that a different viewpoint is taken. We follow the elegant and exceptionally clear presentation by Lahiri & Ganguli (2013), but adapted to our own notation here. Upon the storage of $\boldsymbol{\xi}^0$ at time $t = 0^-$ s, there is an "ideal" set of synaptic strengths, call them $\widehat{S}_i$, corresponding to $\widehat{S}_i = +1$ for those synapses experiencing a potentiating induction signal and $\widehat{S}_i = -1$ for those experiencing a depressing induction signal. Letting $\widehat{\boldsymbol{S}}$ and $\boldsymbol{S}(t)$ be vectors with components of the ideal strengths $\widehat{S}_i$ and the actual strengths at some later time $S_i(t)$, respectively, then the overlap $\widehat{\boldsymbol{S}} \cdot \boldsymbol{S}(t)$ measures the "quality" of the storage of the memory $\boldsymbol{\xi}^0$, where the dot "$\cdot$" denotes the dot product. The SNR in the ideal observer approach is then defined, in discrete time, by

$$\mathcal{SNR}(\alpha) = \frac{\mathsf{E}\big[\widehat{\boldsymbol{S}} \cdot \boldsymbol{S}(\alpha)\big] - \mathsf{E}\big[\widehat{\boldsymbol{S}} \cdot \boldsymbol{S}(\infty)\big]}{\sqrt{\mathsf{Var}\big[\widehat{\boldsymbol{S}} \cdot \boldsymbol{S}(\infty)\big]}}, \tag{2.22}$$

where $S_i(\infty)$ means the equilibrium strength, i.e. $\lim_{t\to\infty} S_i(t)$. Here, because we take balanced processes, $\mathsf{Prob}[\xi_i^0 = \pm 1] = \frac{1}{2}$, we have that $\mathsf{E}[S_i(\infty)] = 0$, but in general $\mathsf{E}[S_i(\infty)] \neq 0$ for unbalanced processes with $\mathsf{Prob}[\xi_i^0 = \pm 1] = g_\pm$, where $g_+ + g_- = 1$. Of course, a continuous-time form of Eq. (2.22) can be written down, simply as

$$\mathcal{SNR}(t) = \frac{\mathsf{E}\big[\widehat{\boldsymbol{S}} \cdot \boldsymbol{S}(t)\big] - \mathsf{E}\big[\widehat{\boldsymbol{S}} \cdot \boldsymbol{S}(\infty)\big]}{\sqrt{\mathsf{Var}\big[\widehat{\boldsymbol{S}} \cdot \boldsymbol{S}(\infty)\big]}}. \tag{2.23}$$

Lahiri & Ganguli (2013) then show that

$$\mathcal{SNR}(t) = 2g_+g_-\sqrt{N}\,\boldsymbol{\Omega}_2^{\mathrm{T}}\big[\exp\left(rt\,\mathbb{G}_2\right)\big]\left(\mathbb{M}_2^+ - \mathbb{M}_2^-\right)\boldsymbol{A}_2, \tag{2.24}$$

where the denominator in the SNR has been replaced by $\sqrt{N}$. Note carefully the structural similarity of Eq. (2.24) to $\mu(t)$ in Eq. (2.12b). For $g_\pm = \frac{1}{2}$, we have precisely that $\mathcal{SNR}(t) = \sqrt{N}\,\mu(t)$.

This near-equivalence is, in fact, hardly surprising. Critically, the ideal strengths $\widehat{S}_i$ are precisely equal to the induction signals ξ_i^0, so that $\widehat{\boldsymbol{S}} \equiv \boldsymbol{\xi}^0$.

Thus,

$$\begin{aligned}\mathsf{E}\big[\widehat{\boldsymbol{S}}\cdot\boldsymbol{S}(t)\big] &= \mathsf{E}\big[\boldsymbol{\xi}^0\cdot\boldsymbol{S}(t)\big]\\ &= \mathsf{E}\Big[\sum_{i=1}^{N}\xi_i^0\,S_i(t)\Big]\\ &\equiv N\,\mathsf{E}\big[h(t)\big]\\ &= N\mu(t). \qquad (2.25)\end{aligned}$$

The signal numerator in Eq. (2.23) is therefore, up to the overall factor of N, just $\mu(t)-\mu(\infty)$, and we can rewrite Eq. (2.23) as

$$\mathcal{SNR}(t) = \frac{\mathsf{E}[h(t)] - \mathsf{E}[h(\infty)]}{\sqrt{\mathsf{Var}[h(\infty)]}} = \frac{\mu(t)-\mu(\infty)}{\sqrt{\left[1-\mu(\infty)^2\right]/N}}, \qquad (2.26)$$

expressed purely in terms of the perceptron's mean activation in response to the tracked memory. The numerator $\mu(t)-\mu(\infty)$ is just the perceptron's excess mean activation above its baseline or equilibrium mean activation. For $g_\pm = \frac{1}{2}$, $\mu(\infty) = 0$ and $\mathsf{Var}[h(\infty)] = \left[1-\mu(\infty)^2\right]/N \equiv 1/N$, so for balanced potentiation and depression, $\mathcal{SNR}(t) = \sqrt{N}\mu(t)$ holds exactly. For unbalanced processes,

$$\mu(t) = \boldsymbol{\Omega}_2^{\mathrm{T}}\left[\exp\left(rt\,\mathbb{G}_2\right)\right]\left(g_+\mathbb{M}_2^+ - g_-\mathbb{M}_2^-\right)\boldsymbol{A}_2, \qquad (2.27)$$

and, following Lahiri & Ganguli's derivation of $\mathcal{SNR}(t)$, we obtain

$$\mu(t)-\mu(\infty) = 2g_+g_-\boldsymbol{\Omega}_2^{\mathrm{T}}\left[\exp\left(rt\,\mathbb{G}_2\right)\right]\left(\mathbb{M}_2^+ - \mathbb{M}_2^-\right)\boldsymbol{A}_2, \qquad (2.28)$$

which leads precisely to Eq. (2.24) when the variance $\sigma(\infty)^2$ is replaced by $1/N$. For such unbalanced processes, in an MFPT approach, we would naturally be led to consider the MFPT for $h(t)$ to reach $\mu(\infty)$ or some threshold in excess of $\mu(\infty)$.

Because our SNR $\mu(t)/\sigma(t)$ (for balanced processes) or $\left[\mu(t) - \mu(\infty)\right]/\sigma(t)$ (for unbalanced processes) and the ideal observer SNR, $\mathcal{SNR}(t)$, are ratios, overall scale factors are irrelevant as they cancel out. We can then redefine the ideal observer SNR as

$$\mathcal{SNR}(t) = \frac{\mu(t) - \mu(\infty)}{\sigma(\infty)}. \tag{2.29}$$

We thus see the principal difference between our SNR, $\left[\mu(t) - \mu(\infty)\right]/\sigma(t)$, and the ideal observer SNR, $\left[\mu(t) - \mu(\infty)\right]/\sigma(\infty)$: we use the noise in the current signal while the ideal observer approach uses the noise in the equilibrium signal. Asking which is correct is meaningless because we may make definitions however we wish, but we may certainly ask which is more useful, which generalises more readily, and which captures more naturally and faithfully the key dynamics limiting memory lifetimes? We have that $\sigma(t) \to 1/\sqrt{N}$ as $t \to \infty$ and for all but short memory lifetimes, we may safely replace $\sigma(t)$ by $1/\sqrt{N}$, whether in discrete time or in continuous time. Thus, pragmatically speaking, the difference between $\left[\mu(t) - \mu(\infty)\right]/\sigma(t)$ and $\left[\mu(t) - \mu(\infty)\right]/\sigma(\infty)$ is largely moot, in terms of defining memory lifetimes. In principle, however, both $h(t)$ and $h(\infty)$ are subject to noise and we wish to determine whether $\mu(t)$ can be discriminated from $\mu(\infty)$ at, say, the level of one standard deviation. In continuous time there is *always* more noise in $h(t)$ than in $h(\infty)$, because $\mathrm{Cov}(t) > 0$. Therefore, $\sigma(t)$ provides a stronger level of discrimination than $\sigma(\infty)$: the use of $\left[\mu(t) - \mu(\infty)\right]/\sigma(\infty)$ may lead us to conclude that $\mu(t)$ can be discriminated from $\mu(\infty)$, i.e. that $\mu(t) > \mu(\infty) + \sigma(\infty)$, when in fact the use of $\left[\mu(t) - \mu(\infty)\right]/\sigma(t)$ may lead us to conclude that $\mu(t)$ cannot be discriminated from $\mu(\infty)$, i.e. that $\mu(t) < \mu(\infty) + \sigma(t)$. On these grounds, it appears better, and safer, to use $\left[\mu(t) - \mu(\infty)\right]/\sigma(t)$ rather than $\left[\mu(t) - \mu(\infty)\right]/\sigma(\infty)$. Furthermore, by thinking in terms of the current signal statistics rather than the equilibrium signal statistics, we are naturally led to consider generalising the SNR $\left[\mu(t) - \mu(\infty)\right]/\sigma(t)$ to an MFPT approach to memory lifetimes. Since we have shown that $\mathcal{SNR}(t)$ is basically just the perceptron's activation, it makes sense to think in terms of a perceptron's current activation because the perceptron does not have access to its equilibrium state. Indeed, the equilibrium state itself may drift as the perceptron's input statistics drift over time, in which case the only dynamical quantity available is $h(t)$ and its relation to firing threshold. Moreover, with multi-level synapses with more than two strength states, our perceptron-based SNR (or MFPT) generalises immediately while for an ideal observer, it is unclear how $\widehat{\boldsymbol{S}}$ should be defined: should $\widehat{S}_i$ be set at its upper or lower limit depending on the sign of ξ_i^0; or should it be incremented or decremented after initialising from a random distribution, and by how many steps; or should it be defined in some other way? Such issues do not arise when we work purely with the perceptron's activation $h(t)$. Overall, then, we consider that the SNR $\left[\mu(t) - \mu(\infty)\right]/\sigma(t)$ is theoretically better motivated, sounder and more natural than $[\mu(t) - \mu(\infty)] / \sigma(\infty)$, and affords greater scope for generalisation.

2.4 Renewal Theory Approach to $\mu(t)$

We may now turn to a recapitulation of the calculation of $\mu(t)$ for the filter model studied earlier (Elliott & Lagogiannis, 2012). We have that

$$\mu(t) = \boldsymbol{\Omega}_2^{\mathrm{T}} \mathbb{P}_2(t)\, \mathbb{M}_2^{+} \boldsymbol{A}_2, \tag{2.30}$$

where $\mathbb{P}_2(t) = \exp(rt\,\mathbb{G}_2)$. In principle, the calculation of $\mathbb{P}_2(t)$ as a matrix exponential is straightforward and we may, if necessary, resort to numerical matrix methods. Unfortunately, for $\Theta > 2$ the stochastic matrix $\mathbb{M}_2$ (and therefore $\mathbb{G}_2 = \mathbb{M}_2 - \mathbb{I}_2$) is defective, so that it lacks a complete basis of eigenvectors. We cannot therefore compute $\exp(rt\,\mathbb{G}_2)$ by a standard eigen-expansion of $\mathbb{G}_2$, although we could use generalised eigenvectors and the Jordan normal form. Furthermore, because $\mathbb{M}_2$ is not a simple, tridiagonal matrix, its spectrum is not, in general, easy to compute analytically.

We may circumvent these problems associated with a direct attack on $\mathbb{P}_2(t)$ by instead using renewal theory (Cox, 1962). First, we decompose $\mathbb{P}_2(t)$ into the sub-blocks

$$\mathbb{P}_2(t) = \left(\begin{array}{c|c} \mathbb{P}^{-|-}(t) & \mathbb{P}^{-|+}(t) \\ \hline \mathbb{P}^{+|-}(t) & \mathbb{P}^{+|+}(t) \end{array} \right), \tag{2.31}$$

where the submatrices are $(2\,\Theta - 1) \times (2\,\Theta - 1)$, and the indices A and B on

$\mathbb{P}^{A|B}(t)$ with $A, B \in \{-, +\}$ denote a change in strength state from strength B to strength A. The elements of these submatrices are $p_{I|J}^{A|B}(t)$, where the lower indices indicate a change in filter state from state J to state I. Thus, $p_{I|J}^{A|B}(t)$ is the probability of a transition from strength state B and filter state J to strength state A and filter state I in a time t. Writing,

$$\mathbb{M}_2^{+}\boldsymbol{A}_2 = \big(\boldsymbol{B}^{-\,\mathrm{T}}\big|\boldsymbol{B}^{+\,\mathrm{T}}\big)^{\mathrm{T}}, \tag{2.32}$$

and recalling that $\boldsymbol{\Omega}_2^{\mathrm{T}} = \big(-\boldsymbol{n}^{\mathrm{T}}\big|+\boldsymbol{n}^{\mathrm{T}}\big)$, we have that

$$\mu(t) = \boldsymbol{n}^{\mathrm{T}}\big[\mathbb{P}^{+|+}(t) - \mathbb{P}^{-|+}(t)\big]\boldsymbol{B}^{+} + \boldsymbol{n}^{\mathrm{T}}\big[\mathbb{P}^{+|-}(t) - \mathbb{P}^{-|-}(t)\big]\boldsymbol{B}^{-}. \tag{2.33}$$

If we denote the (transposed) rows of the submatrices $\mathbb{P}^{A|B}(t)$ by the vectors $\boldsymbol{p}_I^{A|B}(t)$, which carry a filter index I labelling the matrix row, then we may instead write

$$\mu(t) = \sum_I \Big\{ \big[\boldsymbol{p}_I^{+|+}(t) - \boldsymbol{p}_I^{-|+}(t)\big] \cdot \boldsymbol{B}^{+} + \big[\boldsymbol{p}_I^{+|-}(t) - \boldsymbol{p}_I^{-|-}(t)\big] \cdot \boldsymbol{B}^{-} \Big\}, \tag{2.34}$$

because the $\boldsymbol{n}^{\mathrm{T}}$ in Eq. (2.33) merely sums over the final filter state I. We shall use this result below.

We may explicitly compute the two terms in $\mu(t)$ in Eq. (2.34) by decomposing the transitions $\mathbb{P}^{A|B}(t)$ into those in which filter thresholds are not reached in time t and those in which they are. Consider, therefore, transitions in filter state without reaching either threshold. Such a process is a random walk between two absorbing boundaries, because if any particular realisation of the process reaches a threshold, it is removed from the ensemble. Let $f_{I|J}(t)$ denote the probability of a transition between initial filter state J and final filter state I in time t without touching the absorbing boundaries at $\pm\Theta$. We define

$$\mathbb{S} = \tfrac{1}{2}\big(\mathbb{S}^{+}+\mathbb{S}^{-}\big), \tag{2.35}$$

and write $\mathbb{J} = \mathbb{S} - \mathbb{I}$, where $\mathbb{I}$ is the $(2\,\Theta - 1) \times (2\,\Theta - 1)$ identity matrix. Then if $\mathbb{F}(t)$ is the matrix with elements $f_{I|J}(t)$, we have

$$\mathbb{F}(t) = \exp\big(rt\,\mathbb{J}\big). \tag{2.36}$$

Because escape through the filter thresholds is inevitable, we must have that $\mathbb{F}(t) \to \mathbb{O}$ as $t \to \infty$. Consider also the densities for reaching the two filter thresholds, starting from filter state J, in time t. We denote these densities by $G^{+}_{J}(t)$ and $G^{-}_{J}(t)$ for the upper $(+\Theta)$ and lower $(-\Theta)$ thresholds, respectively. They satisfy the equation

$$\frac{dG^{\pm}_{J}(t)}{dt} = \tfrac{1}{2}r\,\big[G^{\pm}_{J+1}(t) + G^{\pm}_{J-1}(t)\big] - r\,G^{\pm}_{J}(t), \tag{2.37}$$

subject to the absorbing boundary conditions $G^{+}_{-\Theta}(t) = 0$ and $G^{+}_{+\Theta}(t) = \delta(t)$ for $G^{+}_{J}(t)$, and $G^{-}_{-\Theta}(t) = \delta(t)$ and $G^{-}_{+\Theta}(t) = 0$ for $G^{-}_{J}(t)$; $\delta(t)$ is the Dirac delta function. Let the Laplace transform of a function $g(t)$ be denoted by $\widehat{g}(s)$ with transformed variable s. Then, Laplace transforming Eq. (2.37), we have the recurrence relation

$$\widehat{G}^{\pm}_{J+1}(s) - 2(1+s/r)\widehat{G}^{\pm}_{J}(s) + \widehat{G}^{\pm}_{J-1}(s) = 0, \tag{2.38}$$

using $G^{\pm}_{J}(0) \equiv 0$ for $|J| < \Theta$, with solutions

$$\widehat{G}^{\pm}_{J}(s) = \frac{\left[\Phi_{+}(s)\right]^{\Theta \pm J} - \left[\Phi_{-}(s)\right]^{\Theta \pm J}}{\left[\Phi_{+}(s)\right]^{2\Theta} - \left[\Phi_{-}(s)\right]^{2\Theta}}, \tag{2.39}$$

where

$$\Phi_{\pm}(s) = (1+s/r) \pm \sqrt{(1+s/r)^2 - 1}. \tag{2.40}$$

The sum $G^{+}_{J}(t) + G^{-}_{J}(t)$ is the probability density function for escaping through either threshold, starting from state J, at time t. The probability

$$H_J(t) = 1 - \int_0^t d\tau \left[G^{+}_{J}(\tau) + G^{-}_{J}(\tau)\right] \tag{2.41a}$$

is therefore the probability of having not escaped through either threshold, starting from state J, in time t. The sum $\sum_{I=-(\Theta-1)}^{+(\Theta-1)} f_{I|J}(t)$ is also the probability that starting from state J, the system has not escaped through either threshold in time t, so we must also have

$$H_J(t) = \sum_{I=-(\Theta-1)}^{+(\Theta-1)} f_{I|J}(t), \tag{2.41b}$$

which will be useful later. Finally, we note that, say, $f_{+(\Theta-1)|J}(t)$ is the probability of reaching state $+(\Theta-1)$ from state J in time t. A single, potentiating step of rate $\frac{1}{2}r$ will take the system from $+(\Theta-1)$ to Θ and thus escape through the upper boundary. Thus, we have that

$$G_J^{\pm}(t) = \tfrac{1}{2} r f_{\pm(\Theta-1)|J}(t), \tag{2.42}$$

which we shall use below.

With the transition probabilities $f_{I|J}(t)$ and escape densities $G_J^{\pm}(t)$ in hand, we may now write down the system of renewal equations for $p_{I|J}^{A|B}(t)$ governing the evolution of states in the filter model studied earlier (Elliott & Lagogiannis, 2012):

$$\begin{aligned}
p_{I|J}^{+|+}(t) &= f_{I|J}(t) + \int_0^t d\tau \big[p_{I|0}^{+|+}(t-\tau)\, G_J^{+}(\tau) + p_{I|0}^{+|-}(t-\tau)\, G_J^{-}(\tau)\big], && \text{(2.43a)}\\
p_{I|J}^{-|+}(t) &= \phantom{f_{I|J}(t) +} \int_0^t d\tau \big[p_{I|0}^{-|+}(t-\tau)\, G_J^{+}(\tau) + p_{I|0}^{-|-}(t-\tau)\, G_J^{-}(\tau)\big], && \text{(2.43b)}\\
p_{I|J}^{+|-}(t) &= \phantom{f_{I|J}(t) +} \int_0^t d\tau \big[p_{I|0}^{+|+}(t-\tau)\, G_J^{+}(\tau) + p_{I|0}^{+|-}(t-\tau)\, G_J^{-}(\tau)\big], && \text{(2.43c)}\\
p_{I|J}^{-|-}(t) &= f_{I|J}(t) + \int_0^t d\tau \big[p_{I|0}^{-|+}(t-\tau)\, G_J^{+}(\tau) + p_{I|0}^{-|-}(t-\tau)\, G_J^{-}(\tau)\big]. && \text{(2.43d)}
\end{aligned}$$

The first equation, for $p_{I|J}^{+|+}(t)$, contains an inhomogeneous term $f_{I|J}(t)$ corresponding to a filter transition without any escape process and therefore without any (possible) change in strength. The homogeneous term captures the two possible first-escape processes in the dynamics. For example, the $p_{I|0}^{+|+}(t-\tau)\, G_J^{+}(\tau)$ term represents a first-escape process through the upper filter threshold at time $\tau \in (0,t)$. Following this first-escape process, the filter is reset to zero, but because the initial strength is already $+1$, no change in strength is possible due to saturation. In the remaining time $t-\tau$, the system must evolve from the zero filter state and strength $+1$ to filter state I and strength $+1$, accounting for the $p^{+|+}_{I|0}(t-\tau)$ factor multiplying $G^+_J(\tau)$. As τ is arbitrary, we average over the first-escape time by integrating over the density. The $p^{+|-}_{I|0}(t-\tau)\,G^-_J(\tau)$ term represents a similar process, but with a first-escape process through lower filter threshold, a change in strength from $+1$ to -1 and resetting the filter to zero; the remaining dynamics are then the $p^{+|-}_{I|0}(t-\tau)$ transition. The other three equations have similar interpretations. However, for $p^{-|+}_{I|J}(t)$ and $p^{+|-}_{I|J}(t)$, there are no inhomogeneous terms because the required changes in strength from $+1$ to -1 or *vice versa* are impossible without filter threshold processes.

Because the integrals in Eqs. (2.43a–d) are just Laplace convolutions, we take Laplace transforms, obtaining

$$\widehat{p}^{+|+}_{I|J}(s) = \widehat{f}_{I|J}(s) + \widehat{p}^{+|+}_{I|0}(s)\,\widehat{G}^+_J(s) + \widehat{p}^{+|-}_{I|0}(s)\,\widehat{G}^-_J(s), \tag{2.44a}$$

$$\widehat{p}^{-|+}_{I|J}(s) = \widehat{p}^{-|+}_{I|0}(s)\,\widehat{G}^+_J(s) + \widehat{p}^{-|-}_{I|0}(s)\,\widehat{G}^-_J(s), \tag{2.44b}$$

$$\widehat{p}^{+|-}_{I|J}(s) = \widehat{p}^{+|+}_{I|0}(s)\,\widehat{G}^+_J(s) + \widehat{p}^{+|-}_{I|0}(s)\,\widehat{G}^-_J(s), \tag{2.44c}$$

$$\widehat{p}^{-|-}_{I|J}(s) = \widehat{f}_{I|J}(s) + \widehat{p}^{-|+}_{I|0}(s)\,\widehat{G}^+_J(s) + \widehat{p}^{-|-}_{I|0}(s)\,\widehat{G}^-_J(s). \tag{2.44d}$$

By setting $J=0$, we may explicitly compute $\widehat{p}^{A|B}_{I|0}(s)$ and then obtain $\widehat{p}^{A|B}_{I|J}(s)$ purely in terms of $\widehat{f}_{I|J}(s)$ and $\widehat{G}^\pm_J(s)$. We may then compute $\widehat{p}^{+|\pm}_{I|J}(s) - \widehat{p}^{-|\pm}_{I|J}(s)$, needed in Eq. (2.34), and so determine $\widehat{\mu}(s)$. We defer the details of these calculations until the section 4, where we will perform a single, general calculation for all the filter models discussed in the next section rather than just the filter model that we examined earlier (Elliott & Lagogiannis, 2012).

Before turning to these other filter models and the general calculation, we prove that the matrix elements $p_{I|J}^{A|B}(t)$ defined by the system of renewal equations above, Eqs. (2.43a–d), are a solution of the matrix differential equation

$$\frac{d\mathbb{P}_2(t)}{dt} = r\,\mathbb{P}_2(t)\,\mathbb{G}_2, \tag{2.45}$$

and thus satisfy $\mathbb{P}_2(t) = \exp\left(rt\,\mathbb{G}_2\right)$. We first note that, for example,

$$\begin{aligned} p_{I|0}^{+|+}(t-\tau)\,G_J^{+}(\tau) &= \tfrac{1}{2}r\,p_{I|0}^{+|+}(t-\tau)\,f_{+(\Theta-1)|J}(\tau) \\ &= \tfrac{1}{2}r\sum_{K,L} p_{I|K}^{+|+}(t-\tau)\,\delta_{K,0}\,\delta_{L,+(\Theta-1)}\,f_{L|J}(\tau) \\ &= \tfrac{1}{2}r\sum_{K,L} p_{I|K}^{+|+}(t-\tau)\big(\mathbb{T}^{+}\big)_{KL} f_{L|J}(\tau) \\ &= \tfrac{1}{2}r\Big[\mathbb{P}^{+|+}(t-\tau)\,\mathbb{T}^{+}\,\mathbb{F}(\tau)\Big]_{IJ}, \end{aligned} \tag{2.46}$$

where we have used $\big(\mathbb{T}^{\pm}\big)_{KL} = \delta_{K,0}\,\delta_{L,\pm(\Theta-1)}$, with $\delta_{I,J}$ being the Kronecker delta function, and with the matrices $\mathbb{T}^{\pm}$ being defined above. Thus, we may write the four renewal equations, Eqs. (2.43a–d), as

$$\mathbb{P}^{+|+}(t) = \mathbb{F}(t) + \tfrac{1}{2}r\int_0^t d\tau \big[\mathbb{P}^{+|+}(t-\tau)\,\mathbb{T}^{+} + \mathbb{P}^{+|-}(t-\tau)\,\mathbb{T}^{-}\big]\mathbb{F}(\tau), \tag{2.47a}$$

$$\mathbb{P}^{-|+}(t) = \tfrac{1}{2}r\int_0^t d\tau \big[\mathbb{P}^{-|+}(t-\tau)\,\mathbb{T}^{+} + \mathbb{P}^{-|-}(t-\tau)\,\mathbb{T}^{-}\big]\mathbb{F}(\tau), \tag{2.47b}$$

$$\mathbb{P}^{+|-}(t) = \tfrac{1}{2}r\int_0^t d\tau \big[\mathbb{P}^{+|+}(t-\tau)\,\mathbb{T}^{+} + \mathbb{P}^{+|-}(t-\tau)\,\mathbb{T}^{-}\big]\mathbb{F}(\tau), \tag{2.47c}$$

$$\mathbb{P}^{-|-}(t) = \mathbb{F}(t) + \tfrac{1}{2}r\int_0^t d\tau \big[\mathbb{P}^{-|+}(t-\tau)\,\mathbb{T}^{+} + \mathbb{P}^{-|-}(t-\tau)\,\mathbb{T}^{-}\big]\mathbb{F}(\tau). \tag{2.47d}$$

Defining the matrices

$$\mathbb{F}_2(t) = \left(\begin{array}{c|c} \mathbb{F}(t) & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{F}(t) \end{array}\right) \quad \text{and} \quad \mathbb{T}_2 = \frac{1}{2}\left(\begin{array}{c|c} \mathbb{T}^{-} & \mathbb{T}^{-} \\ \hline \mathbb{T}^{+} & \mathbb{T}^{+} \end{array}\right) \tag{2.48}$$

and combining the submatrices $\mathbb{P}^{A|B}(t)$ into the full matrix $\mathbb{P}_2(t)$, we may rewrite Eqs. (2.47a–d) to obtain the single $2(2\,\Theta-1)\times 2(2\,\Theta-1)$ matrix renewal equation

$$\mathbb{P}_2(t) = \mathbb{F}_2(t) + r\int_0^t d\tau\,\mathbb{P}_2(t-\tau)\,\mathbb{T}_2\,\mathbb{F}_2(\tau). \tag{2.49}$$

We now differentiate this equation with respect to t. As $d\mathbb{F}(t)/dt = r\,\mathbb{J}\,\mathbb{F}(t) = r\,\mathbb{F}(t)\mathbb{J}$ from Eq. (2.36), we define the matrix $\mathbb{S}_2$ by

$$\mathbb{S}_2 = \left(\begin{array}{c|c} \mathbb{S} & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{S} \end{array}\right), \tag{2.50}$$

and also define $\mathbb{J}_2 = \mathbb{S}_2 - \mathbb{I}_2$ in analogy with $\mathbb{J} = \mathbb{S} - \mathbb{I}$, so that $\mathbb{F}_2(t) = \exp(rt\,\mathbb{J}_2)$ and thus $d\mathbb{F}_2(t)/dt = r\,\mathbb{J}_2\,\mathbb{F}_2(t) = r\,\mathbb{F}_2(t)\,\mathbb{J}_2$. Then, differentiating Eq. (2.49), we have

$$\begin{aligned}
\frac{d\mathbb{P}_2(t)}{dt} &= r\,\mathbb{F}_2(t)\,\mathbb{J}_2 + r\,\mathbb{T}_2\,\mathbb{F}_2(t) + r\int_0^t d\tau \frac{d\mathbb{P}_2(t-\tau)}{dt}\,\mathbb{T}_2\,\mathbb{F}_2(\tau) \\
&= r\,\mathbb{F}_2(t)\,\mathbb{J}_2 + r\,\mathbb{T}_2\,\mathbb{F}_2(t) \\
&\quad - r\Big[\mathbb{P}_2(t-\tau)\,\mathbb{T}_2\,\mathbb{F}_2(\tau)\Big]\Bigg|_{\tau=0}^{\tau=t} + r\left[\mathbb{P}_2(t) - \mathbb{F}_2(t)\right]\mathbb{J}_2, \qquad (2.51)
\end{aligned}$$

where we have used $d\mathbb{P}_2(t-\tau)/dt = -d\mathbb{P}_2(t-\tau)/d\tau$ and then integrated by parts, replacing the resulting integral with $\mathbb{P}_2(t) - \mathbb{F}_2(t)$. Thus,

$$\frac{d\mathbb{P}_2(t)}{dt} = r\,\mathbb{P}_2(t)\big(\mathbb{S}_2 + \mathbb{T}_2 - \mathbb{I}_2\big) \equiv r\,\mathbb{P}_2(t)\,\mathbb{G}_2, \qquad (2.52)$$

since

$$\mathbb{G}_2 = \mathbb{M}_2 - \mathbb{I}_2 = \frac{1}{2}\left(\begin{array}{c|c} \mathbb{S}^+{+}\mathbb{S}^-{+}\mathbb{T}^- & \mathbb{T}^- \\ \hline \mathbb{T}^+ & \mathbb{S}^+{+}\mathbb{S}^-{+}\mathbb{T}^+ \end{array}\right) - \left(\begin{array}{c|c} \mathbb{I} & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{I} \end{array}\right), \qquad (2.53)$$

or $\mathbb{M}_2 = \mathbb{S}_2 + \mathbb{T}_2$. Thus, $\mathbb{P}_2(t)$ defined by the renewal equations in Eqs. (2.43a–d) does indeed satisfy the backward equation $d\mathbb{P}_2(t)/dt = r\,\mathbb{P}_2(t)\mathbb{G}_2$, which has solution $\mathbb{P}_2(t) = \exp(rt\,\mathbb{G}_2)$, and so also satisfies the forward equation $d\mathbb{P}_2(t)/dt = r\,\mathbb{G}_2\mathbb{P}_2(t)$.

3 Variations on Synaptic Filtering

The filter model that we studied earlier (Elliott & Lagogiannis, 2012), and described above in section 2 and depicted in Fig. 1A, is defined by three principal characteristics:

1. a one-step random walk on filter states between two boundaries, at upper and lower filter thresholds;

2. both boundaries are absorbing, with the filter resetting or injection dynamics being independent of synaptic strength and, in particular, independent of whether or not synaptic plasticity can actually be expressed, because of possible saturation;

3. injection upon threshold is always to the zero filter state.

The first two characteristics are embodied by the step matrices $\mathbb{S}^{\pm}$ and thus by $\mathbb{F}(t)$ and specifically by $\mathbb{F}_2(t)$ in Eq. (2.49). The third characteristic is captured by the threshold matrices $\mathbb{T}^{\pm}$ and so by $\mathbb{T}_2$ in Eq. (2.48). Because this filter model is essentially defined as a random walk between two absorbing boundaries with injection into the zero filter state, we refer to it as the A0 filter model, with "A" for absorbing and "0" for zero injection.

We can vary the A0 model in a variety of different ways. Specifically, we can modify the nature of the thresholds, the filter injection or resetting process upon reaching threshold, and even the underlying one-step random walk itself. Before motivating and describing extensions of the A0 model, we first generalise our earlier definitions of the $\mathbb{S}^{\pm}$ and $\mathbb{T}^{\pm}$ matrices. Previously, $\mathbb{S}^{\pm}$ were defined as shift operators without wrap-around, incrementing or decrementing the filter state but without considering the threshold processes. Their components satisfy $\left(\mathbb{S}^{\pm}\right)_{KL} = \delta_{K,L\pm1}$. We now define the two matrices $\mathbb{S}^{\pm}(\rho_{\pm})$ as having components

$$\left[\mathbb{S}^{\pm}(\rho_{\pm})\right]_{KL} = \delta_{K,L\pm1} + \rho_{\pm}\, \delta_{K,\pm(\Theta-1)}\, \delta_{L,\pm(\Theta-1)}, \tag{3.1}$$

for parameters $\rho_{\pm} \in \{0, 1\}$, so that $\mathbb{S}^{+}(\rho_{+})$ has ρ_{+} as the last entry on its diagonal and $\mathbb{S}^{-}(\rho_{-})$ has ρ_{-} as the first entry on its diagonal. For example, for $\Theta = 3$,

$$\mathbb{S}^{+}(\rho_{+}) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & \rho_{+} \end{pmatrix} \quad \text{and} \quad \mathbb{S}^{-}(\rho_{-}) = \begin{pmatrix} \rho_{-} & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Of course, $\mathbb{S}^{\pm}(0) \equiv \mathbb{S}^{\pm}$. However, the matrix $\mathbb{S}^{+}(1)$ increments all filter states but implements a reflecting boundary at threshold $+\Theta$, so that the state $I = +(\Theta - 1)$ is reflected from the boundary and stays at $+(\Theta - 1)$. Similarly, the matrix $\mathbb{S}^{-}(1)$ decrements all filters states but implements a reflecting boundary at threshold $-\Theta$, reflecting $I = -(\Theta - 1)$ back to $-(\Theta - 1)$. The parameters ρ_{+} and ρ_{-} therefore control the nature of the upper and lower thresholds, respectively, with $\rho = 0$ giving an absorbing boundary and $\rho = 1$ a reflecting boundary. Analogous to the definition of $\mathbb{S}$, we now also define the generalised form,

$$\mathbb{S}(\rho_+, \rho_-) = \tfrac{1}{2}\big[\mathbb{S}^+(\rho_+) + \mathbb{S}^-(\rho_-)\big]. \tag{3.2}$$

For the matrices $\mathbb{T}^\pm$, we previously had $\big(\mathbb{T}^\pm\big)_{KL} = \delta_{K,0}\,\delta_{L,\pm(\Theta-1)}$. These matrices determine how the filter state is reset or injected when a threshold process occurs. The presence of $\delta_{K,0}$ indicates that injection is always to the $K = 0$ filter state; the presence of $\delta_{L,\pm(\Theta-1)}$ indicates that threshold processes can only occur if the filter is in the $\pm(\Theta - 1)$ states, otherwise further plasticity induction signals are required. We generalise these matrices so that they depend on the $(2\,\Theta - 1)$-dimensional vector $\boldsymbol{\sigma}$,

$$\big[\mathbb{T}^\pm(\boldsymbol{\sigma})\big]_{KL} = \sigma_K\,\delta_{L,\pm(\Theta-1)}, \tag{3.3}$$

so that the injection process is defined by $\boldsymbol{\sigma}$, with $\boldsymbol{n}\cdot\boldsymbol{\sigma} = 1$, with σ_K being the probability for injection into filter state K upon threshold. Schematically, we may write

$$\mathbb{T}^+(\boldsymbol{\sigma}) = \begin{pmatrix} | & & | & | \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\sigma} \\ | & & | & | \end{pmatrix} \quad\text{and}\quad \mathbb{T}^-(\boldsymbol{\sigma}) = \begin{pmatrix} | & | & & | \\ \boldsymbol{\sigma} & \mathbf{0} & \cdots & \mathbf{0} \\ | & | & & | \end{pmatrix},$$

showing the columns of $\mathbb{T}^\pm(\boldsymbol{\sigma})$, being either the vector $\boldsymbol{\sigma}$ or the $(2\,\Theta - 1)$-dimensional zero vector, $\mathbf{0}$. We define the vector $\boldsymbol{\Delta}$ to have components $\Delta_K = \delta_{K,0}$ and the vector $\boldsymbol{m} = \boldsymbol{n}/(2\,\Theta - 1)$, so that $m_K = 1/(2\,\Theta - 1)$. Below we will be particularly interested in the two injection distributions $\boldsymbol{\sigma} = \boldsymbol{\Delta}$ and

Filter name	Boundary types	Injection type	Step type
A0	Absorbing + Absorbing	Zero	Single
Ar	Absorbing + Absorbing	Random	Single
R0	Absorbing + Reflecting	Zero	Single
Rr	Absorbing + Reflecting	Random	Single
S	Absorbing + Absorbing	Zero	Variable

Table 1. The five alternative filter models, characterised by their boundary types, injection type and step type.

$\boldsymbol{\sigma} = \boldsymbol{m}$.

3.1 Model Versions and the Distributions $\boldsymbol{A}_2$ and $\mathbb{M}^+\boldsymbol{A}_2$

The various models that we discuss continue to be defined by the matrix renewal equation in Eq. (2.49),

$$\mathbb{P}_2(t) = \mathbb{F}_2(t) + r\int_0^t d\tau \, \mathbb{P}_2(t-\tau) \, \mathbb{T}_2 \, \mathbb{F}_2(\tau), \tag{2.49}$$

where the matrices $\mathbb{F}_2(t) = \exp\left[rt\left(\mathbb{S}_2 - \mathbb{I}_2\right)\right]$ and $\mathbb{T}_2$ will be suitably modified for each version. In Table 1 we provide a convenient summary of the five filter models that we will consider here and that we now discuss.

3.1.1 A0 Model

For completeness, we collect together here the defining properties of the A0 model in terms of the generalised matrices $\mathbb{S}^{\pm}(\rho_{\pm})$ and $\mathbb{T}^{\pm}(\boldsymbol{\sigma})$. We have that

$$\mathbb{S}_2 = \left(\begin{array}{c|c} \mathbb{S}(0,0) & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{S}(0,0) \end{array}\right) \quad \text{and} \quad \mathbb{T}_2 = \frac{1}{2}\left(\begin{array}{c|c} \mathbb{T}^{-}(\boldsymbol{\Delta}) & \mathbb{T}^{-}(\boldsymbol{\Delta}) \\ \hline \mathbb{T}^{+}(\boldsymbol{\Delta}) & \mathbb{T}^{+}(\boldsymbol{\Delta}) \end{array}\right), \tag{3.4}$$

with $\mathbb{M}_2 = \frac{1}{2}\left(\mathbb{M}_2^{+}+\mathbb{M}_2^{-}\right) = \mathbb{S}_2 + \mathbb{T}_2$ and the matrix $\mathbb{M}_2^{+}$ is just twice the potentiating part of $\mathbb{M}_2$. The equilibrium distribution, satisfying $\mathbb{M}_2\boldsymbol{A}_2 = \boldsymbol{A}_2$, is $\boldsymbol{A}_2 = \frac{1}{2}\left(\boldsymbol{A}^{\mathrm{T}}\middle|\boldsymbol{A}^{\mathrm{T}}\right)^{\mathrm{T}}$, with

$$A_I = \frac{\Theta - |I|}{\Theta^2}. \tag{3.5}$$

The vectors $\boldsymbol{B}^{\pm}$ defined by $\mathbb{M}_2^{+}\boldsymbol{A}_2 = \left(\boldsymbol{B}^{-\,\mathrm{T}}\middle|\boldsymbol{B}^{+\,\mathrm{T}}\right)^{\mathrm{T}}$ and required in Eq. (2.34) have components

$$B_I^{-} = \tfrac{1}{2}A_{I-1}, \tag{3.6a}$$

$$B_I^{+} = \tfrac{1}{2}\left[A_{I-1} + 2A_{+(\Theta-1)}\Delta_I\right]. \tag{3.6b}$$

The presence of A_{I-1} in both $B_I^{\pm}$ indicates the action of the step operator $\mathbb{S}^{+}(0)$ on states, while the remaining term in B_I^{+} arises because of the two threshold processes due to the action of the upper threshold matrix $\mathbb{T}^{+}(\boldsymbol{\Delta})$.

3.1.2 R0 Model

The A0 model generalises immediately to any model in which synapses may take any number of discrete values of synaptic strength and not just binary strengths. We may, however, also consider an alternative filter model specific to binary-strength synapses. If a binary-strength synapse is strong (weak) and reaches the upper (lower) filter threshold, the A0 filter will inject to the zero filter state but a change in synaptic strength is not possible because of saturation. Instead, we may suppose that a filter threshold changes from absorbing to reflecting under these circumstances. That is, a strong (weak) synapse that reaches the upper (lower) filter threshold is returned or reflected back to filter state $I = +(\Theta - 1)$ $[I = -(\Theta - 1)]$ because no change in strength is possible. Such a filter is shown in Fig. 1B, where we have explicitly shown for clarity the transitions between and within both strength states. We stress, however, that there is still only a single filter per synapse: only the nature of the boundary states change, in a strength-dependent manner. Because changes in strength are still associated with injection into the zero filter state, we refer to this filter as the R0 filter, with "R" for reflecting. The R0 filter constitutes a random walk between an absorbing boundary and a reflecting boundary, where these boundaries swap around under a change in synaptic strength.

For the R0 model, for a weak synapse the relevant step operator is $\mathbb{S}(\rho_+, \rho_-)$ with $\rho_+ = 0$ (upper threshold is absorbing) and $\rho_- = 1$ (lower threshold is reflecting), while for a strong synapses these thresholds must be reversed so we have $\rho_+ = 1$ (upper threshold is reflecting) and $\rho_- = 0$ (lower threshold is absorbing). Thus, the two defining matrices are

$$\mathbb{S}_2 = \left(\begin{array}{c|c} \mathbb{S}(0,1) & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{S}(1,0) \end{array}\right) \quad \text{and} \quad \mathbb{T}_2 = \frac{1}{2}\left(\begin{array}{c|c} \mathbb{O} & \mathbb{T}^-(\boldsymbol{\Delta}) \\ \hline \mathbb{T}^+(\boldsymbol{\Delta}) & \mathbb{O} \end{array}\right), \tag{3.7}$$

and the matrices $\mathbb{M}_2^\pm$ are therefore

$$\mathbb{M}_2^+ = \left(\begin{array}{c|c} \mathbb{S}^+(0) & \mathbb{O} \\ \hline \mathbb{T}^+(\boldsymbol{\Delta}) & \mathbb{S}^+(1) \end{array}\right) \quad \text{and} \quad \mathbb{M}_2^- = \left(\begin{array}{c|c} \mathbb{S}^-(1) & \mathbb{T}^-(\boldsymbol{\Delta}) \\ \hline \mathbb{O} & \mathbb{S}^-(0) \end{array}\right). \tag{3.8}$$

The matrix $\mathbb{T}_2$ contains only $\mathbb{O}$ submatrices in its upper left and lower right sub-blocks because the absorbing boundary dynamics have been replaced by the reflecting boundaries present in $\mathbb{S}_2$. Because the two non-zero sub-blocks in $\mathbb{S}_2$ differ, we write

$$\mathbb{F}_2(t) = \left(\begin{array}{c|c} \mathbb{F}^+(t) & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{F}^-(t) \end{array}\right), \tag{3.9}$$

where $\mathbb{F}^+(t) = \exp\{rt\,[\mathbb{S}(0,1) - \mathbb{I}]\}$ and $\mathbb{F}^-(t) = \exp\{rt\,[\mathbb{S}(1,0) - \mathbb{I}]\}$, and $\mathbb{F}^+(t)$ refers to transitions in filter state without escape through the upper threshold (the lower one being reflecting for weak synapses) and $\mathbb{F}^-(t)$ refers to transitions in filter state without escape through the lower threshold (the upper one being reflecting for strong synapses). We denote the components of $\mathbb{F}^\pm(t)$ by $f^\pm_{I|J}(t)$. The equilibrium distribution for the R0 model is strength-dependent, so we write

$$\boldsymbol{A}_2 = \tfrac{1}{2}\left(\boldsymbol{A}^{-\,\mathrm{T}}|\boldsymbol{A}^{+\,\mathrm{T}}\right)^{\mathrm{T}}, \tag{3.10}$$

with $\boldsymbol{A}^{\pm}$ being the (conditional) filter probability distributions for strong and weak synapses, with $\boldsymbol{n} \cdot \boldsymbol{A}^{\pm} = 1$. Because of the underlying symmetry, however, $\boldsymbol{A}^{\pm}$ are mirror images, with $A^{+}_{+J} = A^{-}_{-J}$ or $\mathbb{R}\boldsymbol{A}^{+} = \boldsymbol{A}^{-}$. By explicitly computing the equilibrium eigenvector of $\mathbb{M}_2$, we find that

$$A^{+}_{I} = \frac{2}{\Theta(3\,\Theta - 1)} \times \begin{cases} \Theta & \text{for } I \geq 0 \\ \Theta + I & \text{for } I < 0 \end{cases}, \tag{3.11}$$

and the vectors $\boldsymbol{B}^{\pm}$ have components

$$B^{-}_{I} = \tfrac{1}{2}A^{-}_{I-1} \equiv \tfrac{1}{2}A^{+}_{-I+1}, \tag{3.12a}$$

$$\begin{aligned} B^{+}_{I} &= \tfrac{1}{2}\left[A^{+}_{I-1} + A^{-}_{+(\Theta-1)}\Delta_I + A^{+}_{+(\Theta-1)}\delta_{I,+(\Theta-1)}\right] \\ &\equiv \tfrac{1}{2}\left[A^{+}_{I-1} + A^{+}_{-(\Theta-1)}\Delta_I + A^{+}_{+(\Theta-1)}\delta_{I,+(\Theta-1)}\right], \end{aligned} \tag{3.12b}$$

where the $\delta_{I,+(\Theta-1)}$ term in B^{+}_{I} is due to the reflection rather than absorption process at the upper filter threshold. We have used the symmetry $A^{-}_{+J} = A^{+}_{-J}$ to express these distributions purely in terms of $\boldsymbol{A}^{+}$.

3.1.3 Ar and Rr Models

Injection into the zero filter state upon reaching filter threshold is a very precise requirement for filter dynamics that may not be achievable in a biological setting where filter states may be encoded in the conformational configurations or phosphorylation states of perhaps rather small ensembles of large macromolecules (Elliott, 2011a). It is possible, for example, that the injection upon threshold could be smeared around the zero state according to some distribution. For simplicity and to avoid excessive parameter dependence, we consider the worst case scenario: injection upon threshold is completely random, so that the synapse injects into any filter state with probability $1/(2\,\Theta - 1)$. Thus, we consider versions of the A0 and R0 filters that we refer to as the Ar and Rr filters, with "r" for random injection. Considering such extreme cases permits us to determine to what extent the results for the A0 and R0 filters are dependent on precise injection processes.

The various defining matrices for the Ar and Rr models are identical to those for the A0 and R0 models above except that the zero injection distribution $\boldsymbol{\sigma} = \boldsymbol{\Delta}$ is replaced by the random injection distribution $\boldsymbol{\sigma} = \boldsymbol{m}$. Thus, in Eqs. (3.4) and (3.7) above, we merely replace $\mathbb{T}^{\pm}(\boldsymbol{\Delta})$ by $\mathbb{T}^{\pm}(\boldsymbol{m})$ in the two definitions of $\mathbb{T}_2$. As for the A0 model, the equilibrium distribution $\boldsymbol{A}_2$ of the Ar model can be written symmetrically as $\left(\boldsymbol{A}^{\mathrm{T}}\middle|\boldsymbol{A}^{\mathrm{T}}\right)^{\mathrm{T}}$, and as for the R0 model, we must write the equilibrium distribution of the Rr model as $\left(\boldsymbol{A}^{-\,\mathrm{T}}\middle|\boldsymbol{A}^{+\,\mathrm{T}}\right)^{\mathrm{T}}$, where again $\boldsymbol{A}^{\pm}$ are mirror images. Explicit computation of the equilibrium eigenvector of $\mathbb{M}_2$ for the Ar model shows that

$$A_I = \frac{3\,(\Theta^2 - I^2)}{\Theta\,(4\,\Theta^2 - 1)}, \tag{3.13}$$

with the vectors $\boldsymbol{B}^{\pm}$ given by

$$B_I^- = \tfrac{1}{2} A_{I-1}, \tag{3.14a}$$

$$B_I^+ = \tfrac{1}{2} \left[A_{I-1} + 2A_{+(\Theta-1)} m_I \right]. \tag{3.14b}$$

These forms for $B_I^{\pm}$ differ from those for the A0 model in Eq. (3.6) only in having m_I in place of Δ_I for the injection distribution. For the Rr model, its equilibrium distribution is given by

$$A_I^+ = \frac{3(\Theta + I)(3\,\Theta - 1 - I)}{2\,\Theta(2\,\Theta - 1)(4\,\Theta - 1)}, \tag{3.15}$$

and the vectors $\boldsymbol{B}^{\pm}$ have components,

$$B_I^- = \tfrac{1}{2} A_{I-1}^- \equiv \tfrac{1}{2} A_{-I+1}^+, \tag{3.16a}$$

$$\begin{aligned} B_I^+ &= \tfrac{1}{2} \left[A_{I-1}^+ + A_{+(\Theta-1)}^- m_I + A_{+(\Theta-1)}^+ \delta_{I,+(\Theta-1)} \right] \\ &\equiv \tfrac{1}{2} \left[A_{I-1}^+ + A_{-(\Theta-1)}^+ m_I + A_{+(\Theta-1)}^+ \delta_{I,+(\Theta-1)} \right]. \end{aligned} \tag{3.16b}$$

Again, the forms for these components for the Rr model differ from those for the R0 model in Eq. (3.12) only in having m_I in place of Δ_I.

3.1.4 S Model

The four filter types discussed above are essentially defined by symmetric one-step random walk processes between absorbing or reflecting boundaries. Although such mechanisms suppress fluctuation-induced changes in synaptic strength, such changes are still possible. For example, starting from the zero filter state and with $\Theta = 5$, a sequence of nine potentiating induction signals and four depressing induction signals *in any order* will drive the filter to its upper threshold and a change in strength if the synapse is weak. Although a minimum of Θ induction signals of the same type is required to drive a filter to threshold from the zero state, these same-type signals are not required to be consecutive. If a different-type signal is interposed in this minimal sequence, then a further same-type signal is required to reverse its effect. We therefore also consider a much stronger filtering process than those considered above. This stronger process requires Θ *consecutive* same-type induction signals from the zero state to reach filter threshold. Specifically, if the filter is in a positive (negative) state and a depressing (potentiating) induction signal occurs, then the filter is returned to the zero state. This enforces the requirement of a minimum of Θ consecutive same-type induction signals to reach threshold. We call this type of filter a super or S filter and show its transitions in Fig. 1C. Technically, it is defined by a variable-step random walk between two absorbing boundaries, with zero injection. In principle we could consider the four obvious versions, of either absorbing or reflecting boundaries as appropriate, and either zero or random injections. Clearly, random injections undermine the principle of demanding a minimum number of consecutive same-type induction signals, so we do not consider them. We could consider a binary-synapse-specific reflecting boundary version of the super filter, but in the interests of simplicity, we examine only the absorbing boundary form. The S filter considered here is therefore to be viewed as a variable-step version of the one-step A0 filter. In fact, for the specific cases $\Theta = 1$ and $\Theta = 2$, the A0 and S filters are identical.

The step matrices $\mathbb{S}^{\pm}$ for the A0 and Ar models must be modified to take into account the fact that filter states are set to zero if an induction signal occurs that is opposite in sign to the current filter state. We denote these modified matrices by $\overline{\mathbb{S}}^{\pm}$, which we may think of as rectifying shift operators. They have components given by

$$\left(\overline{\mathbb{S}}^{+}\right)_{KL} = \begin{cases} \delta_{K,L+1} & \text{for } L \geq 0 \\ \delta_{K,0} & \text{for } L < 0 \end{cases}, \tag{3.17}$$

$$\left(\overline{\mathbb{S}}^{-}\right)_{KL} = \begin{cases} \delta_{K,0} & \text{for } L > 0 \\ \delta_{K,L-1} & \text{for } L \leq 0 \end{cases}. \tag{3.18}$$

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

$$\overline{\mathbb{S}}^{+} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \overline{\mathbb{S}}^{-} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.19}$$

With $\overline{\mathbb{S}} = \frac{1}{2}\big(\overline{\mathbb{S}}^{+} + \overline{\mathbb{S}}^{-}\big)$, the S model is then defined by the two matrices

$$\mathbb{S}_2 = \left(\begin{array}{c|c} \overline{\mathbb{S}} & \mathbb{O} \\ \hline \mathbb{O} & \overline{\mathbb{S}} \end{array}\right) \quad \text{and} \quad \mathbb{T}_2 = \frac{1}{2}\left(\begin{array}{c|c} \mathbb{T}^{-}(\boldsymbol{\Delta}) & \mathbb{T}^{-}(\boldsymbol{\Delta}) \\ \hline \mathbb{T}^{+}(\boldsymbol{\Delta}) & \mathbb{T}^{+}(\boldsymbol{\Delta}) \end{array}\right). \tag{3.20}$$

The equilibrium eigenvector of $\mathbb{M}_2$ is symmetric in weak and strong states, and a direct computation shows that

$$A_I = \frac{2^{\Theta-1-|I|}}{3 \times 2^{\Theta-1} - 2}. \tag{3.21}$$

The vectors $\boldsymbol{B}^{\pm}$ have components given by

$$B_I^{-} = \begin{cases} 0 & \text{for } I < 0, \\ \frac{1}{2}\sum_{J<0} A_J & \text{for } I = 0, \\ \frac{1}{2}A_{I-1} & \text{for } I > 0, \end{cases} \tag{3.22a}$$

$$B_I^{+} = \begin{cases} 0 & \text{for } I < 0, \\ \frac{1}{2}\left[\left(\sum_{J<0} A_J\right) + 2A_{+(\Theta-1)}\right] & \text{for } I = 0, \\ \frac{1}{2}A_{I-1} & \text{for } I > 0. \end{cases} \tag{3.22b}$$

The A_{I-1} terms for $I > 0$ indicate that $\mathbb{M}_2^{+}$ for the S filter is a standard step operator on positive filter states. On negative filter states, it sends them to the zero filter state, so the distributions $\boldsymbol{B}^{\pm}$ vanish on these states. The term $\sum_{J<0} A_J$ in $B_0^{\pm}$ is due to this sending of negative filter states to the zero state.

3.1.5 Examples of Equilibrium Distributions

Having defined all five filter models, we may now compare their equilibrium distributions graphically. In Fig. 2, we plot A_I or $A_I^\pm$ against I for all five models, for the specific choice $\Theta = 8$. We see that the equilibrium distributions for the A0 and Ar filters are quite similar, as is the case for the R0 and Rr equilibrium distributions. Random rather than zero injection somewhat smooths out the equilibrium distributions. However, even if we considered random injection for the S filter, its equilibrium distribution would still be very strongly peaked around the zero filter state as it is for zero injection, because of the number of non-threshold processes that return the S filter to the zero state.

We note that for all the various A_I and $A_I^\pm$ above for the different filter models, we define $A_{\pm\Theta} \equiv 0$ or $A_{\pm\Theta}^\pm \equiv 0$ when the expressions above do not automatically satisfy these requirements. We shall use this convention below.

3.2 Escape Densities and Non-Escape Transition Probabilities

We must derive expressions for the escape densities $G_J^\pm(t)$ and non-escape transition probabilities $f_{I|J}(t)$ or $f_{I|J}^\pm(t)$ for the five filter models. For the three models with two absorbing boundaries (A0, Ar and S filters), the two densities $G_J^+(t)$ and $G_J^-(t)$ refer to the two escape processes through the upper and lower filter thresholds, respectively, in either strength state. Similarly, the relevant non-escape transition probability is $f_{I|J}(t)$ as this is independent of synaptic strength for these three models. The probability $H_J(t)$ of not having escaped through either filter threshold from state J in time t in the presence of two absorbing boundaries is given by the two alternative forms in Eqs. (2.41a,b),

$$H_J(t) = 1 - \int_0^t d\tau \big[G_J^+(\tau) + G_J^-(\tau) \big] = \sum_{I=-(\Theta-1)}^{+(\Theta-1)} f_{I|J}(t).$$

For the R0 and Rr models there is only one escape process because the filter has only one relevant, absorbing threshold, with the other being reflecting. For these two models, the density $G_J^-(t)$ and transition probability $p_{I|J}^-(t)$ refer to processes in the presence of an escape process at the lower filter threshold and so are the relevant quantities for strong synapses. Conversely, $G_J^+(t)$ and $f_{I|J}^+(t)$ refer to processes involving the upper filter threshold and so are relevant to weak synapses. In the presence of a single absorbing boundary, we define the probabilities $H_J^\pm(t)$ via the two alternatives,

$$H_J^\pm(t) = \begin{cases} 1 - \displaystyle\int_0^t d\tau \, G_J^\pm(\tau), & \text{(3.23a)} \\ \displaystyle\sum_{I=-(\Theta-1)}^{+(\Theta-1)} f_{I|J}^\pm(t), & \text{(3.23b)} \end{cases}$$

which are the equivalents of Eqs. (2.41a,b) for two absorbing boundaries. Because all four filter models A0, Ar, R0 and Rr are defined by one-step processes, we can derive a single, generic expression for $G_J^\pm(t)$ by using the parameters $\rho_\pm$ to control the type of boundary. However, because the S model is rectifying, we must derive its escape densities via a different method. We set $r = 1$ Hz throughout to avoid unnecessary factors of r in equations.

3.2.1 Derivation for One-Step Filters A0, Ar, R0 and Rr

For convenience we consider a standard symmetric random walk on the states $\{1, \ldots, n\}$ with boundaries at 0 and $n + 1$. We will move back to filter indices at the end of the calculation. The parameters ρ_+ and ρ_- will still control the upper ($n + 1$) and lower (0) boundary types, respectively. The random walk between the two boundaries is defined by the system of equations

$$\frac{df_{1|j}}{dt} = \tfrac{1}{2}\left(\rho_- f_{1|j} + f_{2|j}\right) - f_{1|j}, \tag{3.24a}$$

$$\frac{df_{i|j}}{dt} = \tfrac{1}{2}\left(f_{i-1|j} + f_{i+1|j}\right) - f_{i|j} \quad \text{for } 1 < i < n, \tag{3.24b}$$

$$\frac{df_{n|j}}{dt} = \tfrac{1}{2}\left(f_{n-1|j} + \rho_+ f_{n|j}\right) - f_{n|j}, \tag{3.24c}$$

where we drop the argument on $f_{i|j}(t)$ for convenience. Taking Laplace transforms, we obtain

$$(1 + s)\widehat{f}_{1|j} - \delta_{1,j} = \tfrac{1}{2}\left(\rho_- \widehat{f}_{1|j} + \widehat{f}_{2|j}\right), \tag{3.25a}$$

$$(1 + s)\widehat{f}_{i|j} - \delta_{i,j} = \tfrac{1}{2}\left(\widehat{f}_{i-1|j} + \widehat{f}_{i+1|j}\right) \quad \text{for } 1 < i < n, \tag{3.25b}$$

$$(1 + s)\widehat{f}_{n|j} - \delta_{n,j} = \tfrac{1}{2}\left(\widehat{f}_{n-1|j} + \rho_+ \widehat{f}_{n|j}\right), \tag{3.25c}$$

where we have again dropped the argument on the Laplace transform $\widehat{f}_{i|j}(s)$ for convenience. We define the generating function $\mathcal{H}_j(z, t) = \sum_{i=1}^{n} f_{i|j}(t)\, z^i$

and sum over Eq. (3.25b) multiplied by z^i, taking into account the boundary forms in Eqs. (3.25a) and (3.25c). We obtain

$$\widehat{\mathcal{H}}_j(z,s) = z\,\frac{(1-\rho_- z)\widehat{f}_{1|j}(s) + z^n(z-\rho_+)\widehat{f}_{n|j}(s) - 2\,z^j}{z^2 - 2(1+s)z + 1}, \tag{3.26}$$

where the boundary terms $\widehat{f}_{1|j}(s)$ and $\widehat{f}_{n|j}(s)$ explicitly appear. Now, analogous to Eq. (2.42), we have that $\widehat{G}_j^-(s) = \frac{1}{2}\widehat{f}_{1|j}(s)$ when the lower boundary is absorbing and $\widehat{G}_j^+(s) = \frac{1}{2}\widehat{f}_{n|j}(s)$ when the upper boundary is absorbing. Thus, we may write,

$$\widehat{\mathcal{H}}_j(z,s) = 2\,z\,\frac{(1-\rho_- z)\widehat{G}_j^-(s) + z^n(z-\rho_+)\widehat{G}_j^+(s) - z^j}{z^2 - 2(1+s)z + 1}. \tag{3.27}$$

To determine $\widehat{G}_j^{\pm}(s)$, we observe that $\mathcal{H}(z,t)$ is by definition a finite polynomial in z of degree n with non-negative coefficients. Analyticity therefore requires that the zeros in the denominator of $\mathcal{H}(z,t)$ are cancelled by zeros in its numerator (Cox & Miller, 1965). The zeros in the denominator of $\widehat{\mathcal{H}}(z,s)$ are at the two locations $z = \Phi_\pm(s)$, where $\Phi_\pm(s)$ were defined in Eq. (2.40) (with $r = 1$ Hz). We therefore deduce that

$$\widehat{G}_j^+ = \frac{\Phi_+^{j-1}(\rho_- - \Phi_+) - \Phi_-^{j-1}(\rho_- - \Phi_-)}{\Phi_-^{n-1}(\rho_+ - \Phi_-)(\rho_- - \Phi_-) - \Phi_+^{n-1}(\rho_+ - \Phi_+)(\rho_- - \Phi_+)}, \tag{3.28}$$

$$\widehat{G}_j^- = \frac{\Phi_+^{n-j}(\rho_+ - \Phi_+) - \Phi_-^{n-j}(\rho_+ - \Phi_-)}{\Phi_-^{n-1}(\rho_+ - \Phi_-)(\rho_- - \Phi_-) - \Phi_+^{n-1}(\rho_+ - \Phi_+)(\rho_- - \Phi_+)}. \tag{3.29}$$

Finally, by writing $z^2 - 2(1+s)z + 1 = (1 - \Phi_+ z)\,(1 - \Phi_- z)$ and expanding the denominator of $\widehat{\mathcal{H}}(z,s)$ in a power series in z, we can explicitly determine $\widehat{f}_{i|j}(s)$. We obtain

$$\widehat{f}_{i|j} = 2\left(\frac{\Phi_+^i - \Phi_-^i}{\Phi_+ - \Phi_-} - \rho_-\,\frac{\Phi_+^{i-1} - \Phi_-^{i-1}}{\Phi_+ - \Phi_-}\right)\widehat{G}_j^- - 2\,\frac{\Phi_+^{i-j} - \Phi_-^{i-j}}{\Phi_+ - \Phi_-}\chi_{i\geq j}, \tag{3.30}$$

where $\chi_{i\geq j} = 1$ if $i \geq j$ and 0 otherwise. In fact, $f_{i|j}(t) = f_{j|i}(t)$ for all four choices of $\rho_\pm$, so we can make this symmetry explicit by rewriting the expression for $\widehat{f}_{i|j}$ in terms of $\frac{1}{2}\left(\widehat{f}_{i|j} + \widehat{f}_{j|i}\right)$.

We move back to filter indices and ranges by setting $n = 2\,\Theta - 1$ and writing, for example, $J = j - \Theta$. For the A0 and Ar models, for which $\rho_+ = 0$ and $\rho_- = 0$, we of course obtain identical expressions for $\widehat{G}_J^\pm(s)$ already stated in Eq. (2.39), we which reproduce here from completeness:

$$\widehat{G}_J^\pm(s) = \frac{\big[\Phi_+(s)\big]^{\Theta\pm J} - \big[\Phi_-(s)\big]^{\Theta\pm J}}{\big[\Phi_+(s)\big]^{2\,\Theta} - \big[\Phi_-(s)\big]^{2\,\Theta}}. \tag{2.39}$$

For the R0 and Rr models, we set either $\rho_+ = 0$ and $\rho_- = 1$ for $\widehat{G}_J^+(s)$ or $\rho_+ = 1$ and $\rho_- = 0$ for $\widehat{G}_J^-(s)$. We obtain

$$\widehat{G}_J^\pm(s) = \frac{\big[\Phi_+(s)\big]^{\Theta-1\pm J}\big[1 - \Phi_+(s)\big] - \big[\Phi_-(s)\big]^{\Theta-1\pm J}\big[1 - \Phi_-(s)\big]}{\big[\Phi_+(s)\big]^{2\,\Theta-1}\big[1 - \Phi_+(s)\big] - \big[\Phi_-(s)\big]^{2\,\Theta-1}\big[1 - \Phi_-(s)\big]}. \tag{3.31}$$

We note that $\widehat{G}_{-J}^+(s) = \widehat{G}_{+J}^-(s)$ for both forms of $\widehat{G}_J^\pm(s)$ in Eqs. (2.39) and (3.31), i.e. for all four filters A0, Ar, R0 and Rr. This symmetry reflects the underlying symmetry between the positive and negative filter states for the A0 and Ar models and the underlying symmetry between the positive and negative filter states and the strong and weak synaptic strengths in the R0 and Rr models.

3.2.2 Derivation for Super Filter

We now consider the S filter. Using standard filter indices, it satisfies the system of equations

$$\frac{df_{I|J}}{dt} = \tfrac{1}{2} f_{I-1|J} - f_{I|J} \quad \text{for } I > 0, \tag{3.32a}$$

$$\frac{df_{0|J}}{dt} = \tfrac{1}{2} \sum_{K \neq 0} f_{K|J} - f_{0|J}, \tag{3.32b}$$

$$\frac{df_{I|J}}{dt} = \tfrac{1}{2} f_{I+1|J} - f_{I|J} \quad \text{for } I < 0, \tag{3.32c}$$

or Laplace transforming,

$$(1+s)\widehat{f}_{I|J} - \delta_{I,J} = \tfrac{1}{2} \widehat{f}_{I-1|J} \quad \text{for } I > 0, \tag{3.33a}$$

$$(1+s)\widehat{f}_{0|J} - \delta_{0,J} = \tfrac{1}{2} \sum_{K \neq 0} \widehat{f}_{K|J}, \tag{3.33b}$$

$$(1+s)\widehat{f}_{I|J} - \delta_{I,J} = \tfrac{1}{2} \widehat{f}_{I+1|J} \quad \text{for } I < 0, \tag{3.33c}$$

Unfortunately we cannot use a generating function to solve the system of equations in Eq. (3.33) because of the presence of $\sum_{K \neq 0} \widehat{f}_{K|J}$ in Eq. (3.33b). Fortunately, we can directly solve this system. We observe that

$$\begin{aligned}\sum_{K\neq 0}\widehat{f}_{K|J} &= \sum_{K}\widehat{f}_{K|J}-\widehat{f}_{0|J}\\ &= \widehat{H}_J-\widehat{f}_{0|J}\\ &= \frac{1}{s}\left(1-\widehat{G}^+_J-\widehat{G}^-_J\right)-\widehat{f}_{0|J}\\ &= \frac{1}{s}\left[1-\tfrac{1}{2}\widehat{f}_{+(\Theta-1)|J}-\tfrac{1}{2}\widehat{f}_{-(\Theta-1)|J}\right]-\widehat{f}_{0|J}, \end{aligned} \tag{3.34}$$

so we may write Eq. (3.33b) as

$$\left(\tfrac{3}{2}+s\right)\widehat{f}_{0|J}-\delta_{0,J}=\frac{1}{2\,s}\left[1-\tfrac{1}{2}\widehat{f}_{+(\Theta-1)|J}-\tfrac{1}{2}\widehat{f}_{-(\Theta-1)|J}\right]. \tag{3.35}$$

We may use the simple, one-step recurrence relations in Eq. (3.33) for $I>0$ and $I<0$ to express $\widehat{f}_{+(\Theta-1)|J}$ and $\widehat{f}_{-(\Theta-1)|J}$ in terms of $\widehat{f}_{0|J}$:

$$[2(1+s)]^{\Theta-1}\,\widehat{f}_{+(\Theta-1)|J}=\widehat{f}_{0|J}+2\,[2(1+s)]^{|J|-1}\,\chi_{J>0}, \tag{3.36a}$$

$$[2(1+s)]^{\Theta-1}\,\widehat{f}_{-(\Theta-1)|J}=\widehat{f}_{0|J}+2\,[2(1+s)]^{|J|-1}\,\chi_{J<0}. \tag{3.36b}$$

Hence, we deduce that

$$\widehat{f}_{0|J}(s)=\frac{[2(1+s)]^{\Theta-1}\,(1+2\,s\,\delta_{0,J})+[2(1+s)]^{|J|-1}\,(\delta_{0,J}-1)}{1+[2(1+s)]^{\Theta-1}\,s\,(3+2\,s)}, \tag{3.37}$$

and $\widehat{G}^{\pm}_J(s) = \frac{1}{2}\widehat{f}_{\pm(\Theta-1)|J}(s)$ are given by

$$\widehat{G}^{\pm}_J(s) = \frac{1}{[2(1+s)]^{\Theta-1}} \left\{ [2(1+s)]^{|J|-1} \chi_{J \gtrless 0} + \tfrac{1}{2}\widehat{f}_{0|J}(s) \right\}. \tag{3.38}$$

We may now also write down $\widehat{f}_{I|J}(s)$ in general, but we only require $\widehat{G}^{\pm}_J(s)$. We note again the symmetry $\widehat{G}^{+}_{-J}(s) = \widehat{G}^{-}_{+J}(s)$, so that in fact all five filter models considered here respect it.

3.3 Mean Escape Times

Determining the mean escape times for the filter escape process enables us to gauge how strongly a filter controls fluctuations. The stronger the dependence on Θ, the stronger the control of fluctuations. Because the Laplace transform of a random variable is, up to an overall sign, just its moment generating function (MGF), the functions $\widehat{G}^{\pm}_J(s)$ computed above are therefore the MGFs for the escape densities through the upper and lower filter thresholds. Expanding as a power series in s generates all the moments of the escape processes. For the A0, Ar and S filters, we are interested in the total escape process through either filter threshold process, so we require the sum $\widehat{G}^{+}_J(s) + \widehat{G}^{-}_J(s)$. For the R0 and Rr filters, there is only one threshold through which to escape, and by symmetry we need only consider, say, $\widehat{G}^{+}_J(s)$. We therefore write either $\widehat{G}^{+}_J(s) + \widehat{G}^{-}_J(s) = 1 - s\,\tau_J + \mathcal{O}(s^2)$ or $\widehat{G}^{+}_J(s) = 1 - s\,\tau_J + \mathcal{O}(s^2)$, where the leading $\mathcal{O}(s^0)$ term of unity reflects the inevitability of escape and τ_J is the mean time to escape starting from filter state J. We find that

$$\tau_J = \begin{cases} \Theta^2 - J^2 & \text{for A filters,} \\ (\Theta - J)(3\,\Theta + J - 1) & \text{for R filters,} \\ 3\big(2^{\Theta-1} - 2^{|J|-1}\big) + \big(3 \times 2^{|J|-1} - 2\big)\delta_{J,0} & \text{for S filter.} \end{cases} \tag{3.39}$$

We are particularly interested in the mean escape time from the zero filter state (for the injection distribution $\boldsymbol{\sigma} = \boldsymbol{\Delta}$) and in the average mean escape time, averaged over all filter states (for the injection distribution $\boldsymbol{\sigma} = \boldsymbol{m}$). Writing $\boldsymbol{\tau}$ as a vector containing the components τ_J, the relevant mean escape times are $\boldsymbol{\Delta} \cdot \boldsymbol{\tau}$ and $\boldsymbol{m} \cdot \boldsymbol{\tau}$. For $\boldsymbol{\Delta} \cdot \boldsymbol{\tau}$, we immediately have

$$\boldsymbol{\Delta} \cdot \boldsymbol{\tau} = \begin{cases} \Theta^2 & \text{for A filters,} \\ \Theta(3\,\Theta - 1) & \text{for R filters,} \\ 3 \times 2^{\Theta-1} - 2 & \text{for S filter,} \end{cases} \tag{3.40}$$

while for $\boldsymbol{m} \cdot \boldsymbol{\tau}$, we obtain

$$\boldsymbol{m} \cdot \boldsymbol{\tau} = \begin{cases} \frac{1}{3}\,\Theta(2\,\Theta + 1) & \text{for A filters,} \\ \frac{2}{3}\,\Theta(4\,\Theta - 1) & \text{for R filters,} \\ \dfrac{3 \times 2^{\Theta-1}(2\,\Theta - 3) + 4}{(2\,\Theta - 1)} & \text{for S filter.} \end{cases} \tag{3.41}$$

In the A0/Ar and R0/Rr filters, fluctuations are quadratically suppressed as a function of Θ. However, we see that in the S filter, fluctuations are exponentially suppressed as a function of Θ, indicating that the requirement of

Θ consecutive same-type induction signals is very strong. However, this vast suppression of fluctuations comes at a considerable price, discussed below. In Fig. 3, we plot $\boldsymbol{\Delta} \cdot \boldsymbol{\tau}$ for the A0, R0 and S filters and $\boldsymbol{m} \cdot \boldsymbol{\tau}$ for the Ar and Rr models, so that we plot the mean escape times averaged over the relevant initial filter injection $\boldsymbol{\sigma}$ distribution for the specific model (for $r = 1$ Hz). We see that the average mean escape times for the A0 and Ar models are quite similar, and that those for the R0 and Rr models are extremely similar. For $\Theta \geq 7$, the S model has the largest average mean escape time, while for $\Theta \leq 6$, the R0 model has. For smaller Θ, therefore, reflecting boundary filters provide a stronger control of fluctuations, while for larger Θ, the rectifying dynamics of the S filter are stronger and become vastly more so as Θ increases. We will see later that memory lifetimes in these five filter models respect this ordering of average mean escape times.

4 Derivation of $\widehat{\boldsymbol{\mu}}(\boldsymbol{s})$

We may now derive expressions for $\widehat{\mu}(s)$ in the various models. We initially proceed as generally as possible, deriving a single, general expression for $\widehat{\mu}(s)$. We may then reduce this expression to the form required for any one of the various filter models considered here.

4.1 General Results

With the help of the matrix renewal equation in Eq. (2.49), we first write down a set of general equations that cover all the filter models considered here. For two absorbing boundaries, we had

$$\mathbb{T}_2 = \frac{1}{2}\left(\begin{array}{c|c} \mathbb{T}^-(\boldsymbol{\sigma}) & \mathbb{T}^-(\boldsymbol{\sigma}) \\ \hline \mathbb{T}^+(\boldsymbol{\sigma}) & \mathbb{T}^+(\boldsymbol{\sigma}) \end{array}\right),$$

while for one absorbing boundary,

$$\mathbb{T}_2 = \frac{1}{2}\left(\begin{array}{c|c} \mathbb{O} & \mathbb{T}^-(\boldsymbol{\sigma}) \\ \hline \mathbb{T}^+(\boldsymbol{\sigma}) & \mathbb{O} \end{array}\right),$$

where $\boldsymbol{\sigma}$ is the injection distribution. We therefore introduce a parameter β, where $\beta = 1$ for a filter with two absorbing boundaries and $\beta = 0$ for a filter with one absorbing boundary, and write

$$\mathbb{T}_2 = \frac{1}{2}\left(\begin{array}{c|c} \beta\,\mathbb{T}^-(\boldsymbol{\sigma}) & \mathbb{T}^-(\boldsymbol{\sigma}) \\ \hline \mathbb{T}^+(\boldsymbol{\sigma}) & \beta\,\mathbb{T}^+(\boldsymbol{\sigma}) \end{array}\right). \tag{4.1}$$

This enables us to consider both classes of filter simultaneously. We also write

$$\mathbb{F}_2(t) = \left(\begin{array}{c|c} \mathbb{F}^+(t) & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{F}^-(t) \end{array}\right) = \exp rt\big(\mathbb{S}_2 - \mathbb{I}_2\big), \tag{4.2}$$

with the understanding that if $\beta = 1$, then we set $\mathbb{F}^{\pm}(t) = \mathbb{F}(t)$, and that $f^{\pm}_{I|J}(t)$ or $f_{I|J}(t)$, $G^{\pm}_{J}(t)$, and $H^{\pm}_{J}(t)$ or $H_J(t)$ are the appropriate functions for the particular filter model under consideration. Eq. (2.49) then applies to all the filter models considered above. Of course, we can Laplace transform Eq. (2.49) and directly obtain

$$\widehat{\mathbb{P}}_2(s) = \big[(s+1)\mathbb{I}_2 - \mathbb{S}_2 - \mathbb{T}_2\big]^{-1}, \tag{4.3}$$

or equivalently, $\mathbb{P}_2(t) = \exp\left[rt\left(\mathbb{S}_2 + \mathbb{T}_2 - \mathbb{I}_2\right)\right]$, but we still must compute either the matrix inverse or the matrix exponential.

We do not, in fact, need the full form of $\mathbb{P}_2(t)$, but instead just the two sums $\sum_I \big[\boldsymbol{p}_I^{+|\pm}(t) - \boldsymbol{p}_I^{-|\pm}(t)\big]$ appearing in Eq. (2.34), where the vectors $\boldsymbol{p}_I^{A|B}(t)$ were defined earlier as the (transposed) rows of the matrices $\mathbb{P}^{A|B}(t)$. Unpacking Eq. (2.49) and writing out the four Laplace-transformed equations equivalent to Eqs. (2.44a–d), we obtain

$$\widehat{p}^{+|+}_{I|J}(s) = \widehat{f}^{-}_{I|J}(s) + \beta\,\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{p}}^{+|+}_{I}(s)\,\widehat{G}^{+}_{J}(s) + \boldsymbol{\sigma}\cdot\widehat{\boldsymbol{p}}^{+|-}_{I}(s)\,\widehat{G}^{-}_{J}(s), \tag{4.4a}$$

$$\widehat{p}^{-|+}_{I|J}(s) = \beta\,\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{p}}^{-|+}_{I}(s)\,\widehat{G}^{+}_{J}(s) + \boldsymbol{\sigma}\cdot\widehat{\boldsymbol{p}}^{-|-}_{I}(s)\,\widehat{G}^{-}_{J}(s), \tag{4.4b}$$

$$\widehat{p}^{+|-}_{I|J}(s) = \boldsymbol{\sigma}\cdot\widehat{\boldsymbol{p}}^{+|+}_{I}(s)\,\widehat{G}^{+}_{J}(s) + \beta\,\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{p}}^{+|-}_{I}(s)\,\widehat{G}^{-}_{J}(s), \tag{4.4c}$$

$$\widehat{p}^{-|-}_{I|J}(s) = \widehat{f}^{+}_{I|J}(s) + \boldsymbol{\sigma}\cdot\widehat{\boldsymbol{p}}^{-|+}_{I}(s)\,\widehat{G}^{+}_{J}(s) + \beta\,\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{p}}^{-|-}_{I}(s)\,\widehat{G}^{-}_{J}(s). \tag{4.4d}$$

The interpretation of these equations is straightforward. As before, the (transformed) densities $\widehat{G}^{\pm}_{J}(s)$ signal a first escape process through the indicated threshold, with β controlling whether or not that threshold is actually absorbing; if not, that term is absent. Following the first escape process, the filter state is reset with probability distribution $\boldsymbol{\sigma}$. Specifically, $\boldsymbol{\sigma} \cdot \widehat{\boldsymbol{p}}_I^{A|B}(s) = \sum_K \sigma_K \, \widehat{p}_{I|K}^{A|B}(s)$, so such terms represent subsequent transitions from injection site K, weighted according the probability of injection into that site σ_K, to the final filter state I over the remaining time, summed over K.

We define the two vectors $\boldsymbol{\Gamma}^{\pm}(t) = \sum_I \left[\boldsymbol{p}_I^{+|\pm}(t) - \boldsymbol{p}_I^{-|\pm}(t)\right]$, which have components $\Gamma_J^{\pm}(t) = \sum_I \left[p_{I|J}^{+|\pm}(t) - p_{I|J}^{-|\pm}(t)\right]$. Subtracting Eq. (4.4b) from (4.4a) and Eq. (4.4d) from (4.4c), and summing over I, we obtain the two equations

$$\widehat{\Gamma}_J^{+}(s) = +\widehat{H}_J^{-}(s) + \beta \, \boldsymbol{\sigma} \cdot \widehat{\boldsymbol{\Gamma}}^{+}(s) \, \widehat{G}_J^{+}(s) + \quad \boldsymbol{\sigma} \cdot \widehat{\boldsymbol{\Gamma}}^{-}(s) \, \widehat{G}_J^{-}(s), \tag{4.5a}$$

$$\widehat{\Gamma}_J^{-}(s) = -\widehat{H}_J^{+}(s) + \quad \boldsymbol{\sigma} \cdot \widehat{\boldsymbol{\Gamma}}^{+}(s) \, \widehat{G}_J^{+}(s) + \beta \, \boldsymbol{\sigma} \cdot \widehat{\boldsymbol{\Gamma}}^{-}(s) \, \widehat{G}_J^{-}(s), \tag{4.5b}$$

where we have used $H_J^{\pm}(t) = \sum_I f_{I|J}^{\pm}(t)$. We now take the dot product of these two equations with the injection distribution $\boldsymbol{\sigma}$ in order to obtain explicit expressions for $\boldsymbol{\sigma} \cdot \widehat{\boldsymbol{\Gamma}}^{\pm}(s)$. Let $\boldsymbol{G}^{\pm}(t)$ and $\boldsymbol{H}^{\pm}(t)$ be vectors with components $G_J^{\pm}(t)$ and $H_J^{\pm}(t)$, respectively. Because $G_{+J}^{\pm} = G_{-J}^{\mp}$ for any filter model considered here, if the injection distribution is symmetric about the zero filter state (i.e. any distribution for which $\boldsymbol{\sigma} = \mathbb{R}\boldsymbol{\sigma}$), then $\boldsymbol{\sigma} \cdot \boldsymbol{G}^{+}(t) \equiv \boldsymbol{\sigma} \cdot \boldsymbol{G}^{-}(t)$ and similarly for $\boldsymbol{H}^{\pm}(t)$. We may therefore just write $\boldsymbol{\sigma} \cdot \boldsymbol{G}(t)$ and $\boldsymbol{\sigma} \cdot \boldsymbol{H}(t)$ for symmetric injection distributions. The two injection distributions $\boldsymbol{\sigma} = \boldsymbol{\Delta}$ and $\boldsymbol{\sigma} = \boldsymbol{m}$ that we consider are certainly symmetric, but our results will apply to any symmetric injection distribution $\boldsymbol{\sigma}$. From Eq. (4.5) we therefore obtain

$$\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{\Gamma}}^{\pm}(s) = \pm\frac{\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{H}}(s)}{1+(1-\beta)\,\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{G}}(s)}. \tag{4.6}$$

We can then rewrite Eqs. (4.5a,b) as

$$\widehat{\Gamma}^{+}_{J}(s) = +\widehat{H}^{-}_{J}(s) + \Big[\beta\,\widehat{G}^{+}_{J}(s) - \quad \widehat{G}^{-}_{J}(s)\Big]\,\frac{\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{H}}(s)}{1+(1-\beta)\,\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{G}}(s)}, \tag{4.7a}$$

$$\widehat{\Gamma}^{+}_{J}(s) = -\widehat{H}^{+}_{J}(s) + \Big[\quad \widehat{G}^{+}_{J}(s) - \beta\,\widehat{G}^{-}_{J}(s)\Big]\,\frac{\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{H}}(s)}{1+(1-\beta)\,\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{G}}(s)}. \tag{4.7b}$$

Since $\widehat{\mu}(s) = \boldsymbol{B}^{+}\cdot\widehat{\boldsymbol{\Gamma}}^{+}(s) + \boldsymbol{B}^{-}\cdot\widehat{\boldsymbol{\Gamma}}^{-}(s)$ from Eq. (2.34), we have the general result, for any of the filter models considered here,

$$\begin{aligned}\widehat{\mu}(s) = {}& \Big[\boldsymbol{B}^{+}\cdot\widehat{\boldsymbol{H}}^{-}(s) - \boldsymbol{B}^{-}\cdot\widehat{\boldsymbol{H}}^{+}(s)\Big] \\ &+ \Big[\big(\beta\,\boldsymbol{B}^{+} + \boldsymbol{B}^{-}\big)\cdot\widehat{\boldsymbol{G}}^{+}(s) - \big(\boldsymbol{B}^{+} + \beta\,\boldsymbol{B}^{-}\big)\cdot\widehat{\boldsymbol{G}}^{-}(s)\Big] \\ &\quad\times\frac{\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{H}}(s)}{1+(1-\beta)\,\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{G}}(s)}.\end{aligned} \tag{4.8}$$

For two absorbing boundaries, $\beta = 1$, this reduces to

$$\widehat{\mu}(s) = \big(\boldsymbol{B}^{+} - \boldsymbol{B}^{-}\big)\cdot\widehat{\boldsymbol{H}}(s) + \big(\boldsymbol{B}^{+} + \boldsymbol{B}^{-}\big)\cdot\Big[\widehat{\boldsymbol{G}}^{+}(s) - \widehat{\boldsymbol{G}}^{-}(s)\Big]\,\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{H}}(s), \tag{4.9}$$

and for a reflecting boundary, $\beta = 0$, we have

$$\widehat{\mu}(s) = \Big[\boldsymbol{B}^{+} \cdot \widehat{\boldsymbol{H}}^{-}(s) - \boldsymbol{B}^{-} \cdot \widehat{\boldsymbol{H}}^{+}(s)\Big] \\ - \Big[\boldsymbol{B}^{+} \cdot \widehat{\boldsymbol{G}}^{-}(s) - \boldsymbol{B}^{-} \cdot \widehat{\boldsymbol{G}}^{+}(s)\Big] \, \frac{\boldsymbol{\sigma} \cdot \widehat{\boldsymbol{H}}(s)}{1 + \boldsymbol{\sigma} \cdot \widehat{\boldsymbol{G}}(s)}. \tag{4.10}$$

4.2 Specific Forms

With these general results in hand, we now reduce them to the specific forms for each filter model. We rewrite Eqs. (4.9) and (4.10) in terms of the relevant equilibrium distributions $\boldsymbol{A}$ or $\boldsymbol{A}^{\pm}$, and make use of $G^{\pm}_{+J} = G^{\mp}_{-J}$, $H^{\pm}_{+J} = H^{\mp}_{-J}$, $A_{+J} = A_{-J}$ and $A^{\pm}_{+J} = A^{\mp}_{-J}$. We then obtain for the A0 and Ar filters,

$$\widehat{\mu} = \boldsymbol{\sigma} \cdot \widehat{\boldsymbol{H}} \, \Big[A_{\Theta-1} + \sum_{I>0} \big(A_{I-1} - A_{-I-1}\big)\Big(\widehat{G}^{+}_{I} - \widehat{G}^{-}_{I}\Big)\Big], \tag{4.11}$$

and for the S filter,

$$\widehat{\mu} = \boldsymbol{\sigma} \cdot \widehat{\boldsymbol{H}} \, \Big[A_{\Theta-1} + \sum_{I>0} A_{I-1}\Big(\widehat{G}^{+}_{I} - \widehat{G}^{-}_{I}\Big)\Big], \tag{4.12}$$

which is structurally identical to the result for the A0 and Ar filters, save for the absence of the A_{-I-1} term in the sum. For the R0 and Rr filters, writing the result purely in terms of G^{-}, H^{-} and A^{+} (so for the strong rather than weak strength state), we have

$$\begin{aligned}\widehat{\mu} =& \frac{1}{2}\Big[A^{+}_{-\Theta+1}\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{H}}^{-} + A^{+}_{\Theta-1}\widehat{H}^{-}_{\Theta-1} - \sum_{I}\big(A^{+}_{I+1} - A^{+}_{I-1}\big)\widehat{H}^{-}_{I}\Big] \\ &- \frac{1}{2}\frac{\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{H}}^{-}}{1+\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{G}}^{-}}\Big[A^{+}_{-\Theta+1}\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{G}}^{-} + A^{+}_{\Theta-1}\widehat{G}^{-}_{\Theta-1} - \sum_{I}\big(A^{+}_{I+1} - A^{+}_{I-1}\big)\widehat{G}^{-}_{I}\Big].\end{aligned} \tag{4.13}$$

Using the explicit results for the equilibrium distributions, we then obtain, for the A0, Ar and S filters,

$$\widehat{\mu} = A_{\Theta-1}\,\boldsymbol{\sigma}\cdot\widehat{\boldsymbol{H}}\,\Big[1 + \sum_{I>0} c_I\left(\widehat{G}^{+}_{I} - \widehat{G}^{-}_{I}\right)\Big], \tag{4.14}$$

where

$$c_I = \begin{cases} 2 & \text{for A0 filter} \\ 4\,I/(2\,\Theta-1) & \text{for Ar filter} \\ 2^{\Theta-1-I} & \text{for S filter} \end{cases}, \tag{4.15}$$

and where of course $A_{\Theta-1}$ and $\boldsymbol{\sigma}$ are the appropriate value and distribution, respectively, for the required model. For the R0 filter, we get

$$\widehat{\mu}_{\mathrm{R0}} = A^{+}_{-\Theta+1}\left\{\Big[\Theta\widehat{H}^{-}_{\Theta-1} - \sum_{I<0}\widehat{H}^{-}_{I}\Big] - \frac{\widehat{H}^{-}_{0}}{1+\widehat{G}^{-}_{0}}\Big[\Theta\,\widehat{G}^{-}_{\Theta-1} - \sum_{I<0}\widehat{G}^{-}_{I}\Big]\right\}, \tag{4.16}$$

and finally, for the Rr filter,

$$\begin{aligned}\widehat{\mu}_{\mathrm{Rr}} &= A^{+}_{-\Theta+1}\Bigg\{\Big[\Theta\widehat{H}^{-}_{\Theta-1} + \frac{1}{2\,\Theta-1}\sum_{I} I\,\widehat{H}^{-}_{I} - (\Theta-1)\,\boldsymbol{m}\cdot\widehat{\boldsymbol{H}}^{-}\Big] \\ &\quad - \frac{\boldsymbol{m}\cdot\widehat{\boldsymbol{H}}^{-}}{1+\boldsymbol{m}\cdot\widehat{\boldsymbol{G}}^{-}}\Big[\Theta\,\widehat{G}^{-}_{\Theta-1} + \frac{1}{2\,\Theta-1}\sum_{I} I\,\widehat{G}^{-}_{I} - (\Theta-1)\,\boldsymbol{m}\cdot\widehat{\boldsymbol{G}}^{-}\Big]\Bigg\}.\end{aligned} \tag{4.17}$$

Using the different forms of $\widehat{G}^{\pm}_{J}(s)$ and equilibrium distributions derived earlier, we may explicitly compute the sums in the various forms of $\widehat{\mu}$. The calculations are tedious but routine. We eventually obtain the following results:

$$\widehat{\mu}_{\mathrm{A0}} = \frac{2}{\Theta^2}\frac{\Phi_{+}\,(1+\Phi_{+})}{(1-\Phi_{+})^3}\,\frac{\big(1-\Phi_{+}^{\Theta}\big)^3}{(1+\Phi_{+}^{2\,\Theta})\,(1+\Phi_{+}^{\Theta})}, \tag{4.18}$$

$$\begin{aligned}\widehat{\mu}_{\mathrm{Ar}} &= \frac{6}{\Theta(2\,\Theta+1)(2\,\Theta-1)^2}\,\frac{\Phi_{+}\,(1+\Phi_{+})}{(1-\Phi_{+})^5} \\ &\quad\times\frac{\big[2\,\Theta\,(1-\Phi_{+})\big(1+\Phi_{+}^{2\,\Theta}\big)-(1+\Phi_{+})\big(1-\Phi_{+}^{2\,\Theta}\big)\big]^2}{(1-\Phi_{+}^{4\,\Theta})},\end{aligned} \tag{4.19}$$

$$\begin{aligned}\widehat{\mu}_{\mathrm{R0}} &= \frac{4}{\Theta(3\,\Theta-1)}\Bigg[\frac{\Phi_{+}}{(1-\Phi_{+})^2}+2\frac{\Phi_{+}^2}{(1-\Phi_{+})^3}\,\frac{\big(1-\Phi_{+}^{\Theta-1}\big)}{(1+\Phi_{+}^{\Theta})} \\ &\qquad - 2\,\Theta\frac{(1+\Phi_{+})}{(1-\Phi_{+})^2}\,\frac{\Phi_{+}^{2\,\Theta}}{\big(1+\Phi_{+}^{3\,\Theta-1}\big)\,(1+\Phi_{+}^{\Theta})}\Bigg],\end{aligned} \tag{4.20}$$

$$\begin{aligned}\widehat{\mu}_{\mathrm{Rr}} &= \frac{6}{\Theta(4\,\Theta-1)}\frac{\Phi_{+}\,(1+\Phi_{+})}{(1-\Phi_{+})^3} \\ &\quad\times\frac{\big[1-2\,\Theta\,\Phi_{+}^{2\,\Theta-1}+(2\,\Theta-1)\Phi_{+}^{2\,\Theta}\big]\,\big[(2\,\Theta-1)-2\,\Theta\,\Phi_{+}+\Phi_{+}^{2\,\Theta}\big]}{\big[(2\,\Theta-1)\,(1-\Phi_{+}^{4\,\Theta})-2(\Theta-1)\Phi_{+}\,\big(1-\Phi_{+}^{4\,\Theta-2}\big)\big]},\end{aligned}$$

$$\widehat{\mu}_{\mathrm{S}} = \frac{1}{(3\times 2^{\Theta-1}-2)}\,\widehat{H}_0(s)\left\{1+\frac{1}{s}\left[1-\frac{1}{(1+s)^{\Theta-1}}\right]\right\} \tag{4.21}$$

$$\begin{aligned}&= \frac{1}{(3\times 2^{\Theta-1}-2)}\,\frac{1}{s}\left(1-\frac{(1+2\,s)}{\{1+[2(1+s)]^{\Theta-1}s\,(3+2\,s)\}}\right) \\ &\qquad\times\left\{1+\frac{1}{s}\left[1-\frac{1}{(1+s)^{\Theta-1}}\right]\right\},\end{aligned} \tag{4.22}$$

with the form for $\widehat{\mu}_{\mathrm{A0}}$ having been given before (Elliott & Lagogiannis, 2012). The function Φ_+, or the functions $\Phi_\pm$ with $\Phi_- = 1/\Phi_+$ defined in Eq. (2.40) here with $r = 1$ Hz, appear in all the forms for $\widehat{\mu}$ above except that for $\widehat{\mu}_{\mathrm{S}}$. This is because the S filter dynamics do not take the form of a standard, single-step random walk, for which the functions $\Phi_\pm$ naturally arise. We have given an intermediate form in Eq. (4.21) for $\widehat{\mu}_{\mathrm{S}}$ involving $\widehat{H}_0(s)$ because we will approximate $\widehat{H}_0(s)$ in order to obtain a simple expression for memory lifetimes in the S filter model.

The Laplace transforms $\widehat{\mu}_{\mathrm{A0}}$, $\widehat{\mu}_{\mathrm{Ar}}$ and $\widehat{\mu}_{\mathrm{R0}}$ may be inverted analytically. However, $\widehat{\mu}_{\mathrm{Rr}}$ and $\widehat{\mu}_{\mathrm{S}}$ cannot in general be inverted analytically because the locations of the poles of the Laplace transforms are not available, although good approximations exist. In these two cases, we determine the poles' locations numerically and then invert. For the other three cases, the inversion is routine using the methods described elsewhere (Elliott & Lagogiannis, 2012). In brief, we must find the locations of the poles in s in the expressions above for $\widehat{\mu}_{\mathrm{A0}}(s)$, $\widehat{\mu}_{\mathrm{Ar}}(s)$ and $\widehat{\mu}_{\mathrm{R0}}(s)$. By inspection, we can essentially just read off the locations of the poles as functions of $\Phi_+(s)$ rather than functions of s. But because $\Phi_+(s)\,\Phi_-(s) = 1$ and $\Phi_+(s) + \Phi_-(s) = 2(1+s)$, if $\Phi_+(s) = \omega$ and $\Phi_+(s) = \omega^*$ are a pole and its complex conjugate in Φ_+, then

$$\frac{\Phi_+(s)}{[\Phi_+(s)]^2 - (\omega + \omega^*)\,\Phi_+(s) + 1} = \frac{1}{2}\,\frac{1}{s + 1 - \frac{1}{2}(\omega + \omega^*)}, \tag{4.23}$$

so a pole and its complex conjugate in Φ_+ combine to create a simple pole in s at

$s = \frac{1}{2}\,(\omega + \omega^{*}) - 1$. Thus, complex conjugate pairs of poles in $\Phi_{+}(s)$, which can be read off, combine to create a simple pole in s, and the corresponding residue at the simple s pole can be easily computed. In this way, even eventually obtain

$$\begin{aligned}
\mu_{\mathrm{A0}}(t) &= \frac{1}{\Theta^3} \sum_{l=0}^{\Theta-1} \cot^2 \tfrac{(2\,l+1)\pi}{4\,\Theta}\; \exp\left\{-t\left[1 - \cos\tfrac{(2\,l+1)\pi}{2\,\Theta}\right]\right\} \\
&\quad - \frac{4}{\Theta^3} \sum_{l=0}^{\left\lfloor \frac{\Theta-1}{2} \right\rfloor} \cot^2 \tfrac{(2\,l+1)\pi}{2\,\Theta}\; \exp\left\{-t\left[1 - \cos\tfrac{(2\,l+1)\pi}{\Theta}\right]\right\}, \qquad (4.24) \\
\mu_{\mathrm{Ar}}(t) &= \frac{3}{2\,\Theta^2(2\,\Theta+1)(2\,\Theta-1)^2} \\
&\quad \times \sum_{l=1}^{2\,\Theta} \left\{ \left[1-(-1)^l\right] \cot^4 \tfrac{l\pi}{4\,\Theta} - 4\,\Theta^2 \left[1+(-1)^l\right] \cot^2 \tfrac{l\pi}{4\,\Theta} \right\} \\
&\qquad\qquad \times \exp\left[-t\left(1 - \cos\tfrac{l\pi}{2\,\Theta}\right)\right], \qquad (4.25) \\
\mu_{\mathrm{R0}}(t) &= \frac{4}{(3\,\Theta-1)^2} \sum_{l=0}^{\left\lfloor \frac{3\,\Theta-2}{2} \right\rfloor} (-1)^l \cot \tfrac{(2\,l+1)\pi}{2(3\,\Theta-1)} \frac{\cos \frac{(2\,l+1)\pi}{2(3\,\Theta-1)}}{\cos \frac{(2\,l+1)\Theta\pi}{2(3\,\Theta-1)}} \\
&\qquad\qquad \times \exp\left\{-t\left[1 - \cos\tfrac{(2\,l+1)\pi}{(3\,\Theta-1)}\right]\right\} \\
&\quad - \frac{4(\Theta+1)}{\Theta^2(3\,\Theta-1)} \sum_{l=0}^{\left\lfloor \frac{\Theta-1}{2} \right\rfloor} \cot^2 \tfrac{(2\,l+1)\pi}{2\,\Theta}\, \exp\left\{-t\left[1 - \cos\tfrac{(2\,l+1)\pi}{\Theta}\right]\right\}, \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (4.26)
\end{aligned}$$

where $\lfloor x \rfloor$ is the floor function. The result for $\mu_{\mathrm{A0}}(t)$ has been given before (Elliott & Lagogiannis, 2012).

Below we will require a simple approximation for $\mu_{\mathrm{S}}(t)$ for the S filter. If we examine numerically the location of the poles of $\widehat{G}_0(s)$ for the S filter, which gives us $\widehat{H}_0(s) = \left[1 - 2\,\widehat{G}_0(s)\right]/s$ in Eq. (4.21), then we find one real negative pole very close to $s = 0$ with the remaining poles very nearly lying on a circle in the complex s plane and having more negative real parts than the pole near $s =$

0. The pole close to $s = 0$ gives the slowest decay mode and is well separated from the other poles, so we can safely discard the other poles. Let $-s_0 < 0$ be the location of this pole. We find that $s_0 \approx A_{\Theta-1} = 1/\big(3 \times 2^{\Theta-1} - 2\big)$. Since $2\,G_0(t)$ corresponds to a probability density function, we must write $G_0(t) \approx \frac{1}{2} s_0 \exp(-s_0\, t)$ in this approximation and hence $H_0(t) \approx \exp(-s_0\, t)$ or $\widehat{H}_0(s) \approx 1/(s + s_0)$. The residue at $s = -s_0$ in Eq. (4.21) is then immediate, and since $s_0 \ll 1$, we may write

$$\mu_{\mathrm{S}}(t) \approx \frac{\Theta}{(3 \times 2^{\Theta-1} - 2)} \exp(-s_0\, t), \tag{4.27}$$

where the other poles, associated with $s = -1$ in Eq. (4.21), have also been discarded as they also correspond to transients that decay much more rapidly than the mode associated with the $-s_0$ pole.

5 Comparison of Memory Lifetimes

We may now use our results above to examine perceptron memory lifetimes in the five filter models discussed above. We shall also compare these models to the "cascade" model (Fusi *et al.*, 2005). We have previously derived an expression for $\widehat{\mu}(s)$, call it $\widehat{\mu}_{\mathrm{C}}(s)$, in the cascade model (Elliott & Lagogiannis, 2012), so we do not reproduce those analytical results here. The expression for $\widehat{\mu}_{\mathrm{C}}(s)$ cannot be inverted analytically because the locations of its poles cannot be determined analytically. As with $\widehat{\mu}_{\mathrm{Rr}}(s)$ and $\widehat{\mu}_{\mathrm{S}}(s)$, we therefore locate the poles of $\widehat{\mu}_{\mathrm{C}}(s)$ numerically and then perform the Laplace inversion. A cascade of size n has $2\,n$ states, while a filter with threshold Θ has $2\,\Theta - 1$ states. Taking into account a state for its binary-valued strength, a synapse with a filter of threshold Θ requires $2\,\Theta$ states to represent both strength and filter states. We may therefore directly compare a filter with threshold Θ to a cascade of size $n = \Theta$, since they have identical numbers of states.

The SNR memory lifetime τ_{snr}, the solution of $\mu(\tau_{\mathrm{snr}})/\sigma(\tau_{\mathrm{snr}}) = 1$, will mostly be determined analytically. Although we determined $\widehat{\mu}(s)$ analytically in the previous section for all five filter models, and can either analytically or numerically invert the Laplace transforms to obtain $\mu(t)$, it is very hard to determine $\sigma(t)^2$ analytically except for very small values of Θ. However, provided memory lifetimes are acceptably large, it is usually the case that we can just approximate $\sigma(t)^2$ by $1/N$, which is what we do below. When we do require the exact value of $\sigma(t)^2$, we use numerical matrix methods to compute it.

We confirm our analytical results above for $\widehat{\mu}(s)$ and therefore $\mu(t)$ and our numerical matrix methods for computing $\sigma(t)^2$ by comparing them to simulation results. Full details of our simulation protocols may be found elsewhere (Elliott & Lagogiannis, 2012). By averaging over a sufficiently large number of individual simulations, we can obtain agreement between simulations and analytical results as close as we like. We stress that our analytical results are exact: we have not needed to perform mean-field approximations or $1/N$ expansions to obtain them. Our results hold even for $N = 1$ or $N = 2$ synapses, and we have only to average over a very large ensemble of simulations to obtain agreement at any desired level. We therefore we do not show simulation results for the statistics of $h(t)$ here. We employ simulations also to determine MFPT memory lifetimes, τ_{mfpt}, where required. Details of those simulation protocols may also be found elsewhere (Elliott, 2014). Because exact analytical results are not available for τ_{mfpt} for the models considered here, we can show only the simulation results for τ_{mfpt}.

First, in Fig. 4, we show $\mu(t)$ as a function of t for all models, for six different choices of Θ (or n). For all five filter models, we see that $\mu(t)$ rises, reaches a maximum, and then falls. We have previously observed and explained these dynamics for the A0 filter model (Elliott & Lagogiannis, 2012). The storage of the tracked memory biases those synapses that experienced a potentiating (depressing) induction signal to potentiate (depress) or remain strong (weak) if already strong (weak). This bias leads directly to the memory signal rise. As the bias is removed, the memory signal reaches a maximum. It then starts to fall as the synaptic strengths return to their equilibrium distribution. For all choices of Θ, we see that $\mu_{\mathrm{R0}}(t)$ and $\mu_{\mathrm{Rr}}(t)$ are very similar. This similarity reflects the fact that the equilibrium distributions of filter states are also very similar for these two models, as we saw in Fig. 2, and that their mean escape times are also very similar, as in Fig. 3. We also see that $\mu_{\mathrm{A0}}(t)$ and $\mu_{\mathrm{Ar}}(t)$ approach each other quite closely at larger times, although they are somewhat different at smaller times. This difference reflects the differences in the two models' equilibrium distributions: although of the same qualitative shape, the distribution for Rr is somewhat slightly less pronounced around the zero filter state and somewhat higher at states closer to thresholds compared to the distribution for R0. Overall, the similarity between the dynamics of $\mu(t)$ for zero injection and random injection processes indicates that the precise details of the injection process upon reaching filter threshold are not too important.

Comparing the initial memory signals $\mu(0)$ for the filter models, we have that $\mu_{\mathrm{A0}}(0) = 1/\Theta^2$, $\mu_{\mathrm{Ar}}(0) = 3/[\Theta(2\,\Theta + 1)]$, $\mu_{\mathrm{R0}}(0) = 2/[\Theta(3\,\Theta - 1)]$, $\mu_{\mathrm{Rr}}(0) = 3/[\Theta(4\,\Theta - 1)]$, and $\mu_{\mathrm{S}}(0) = 1/\left(3 \times 2^{\Theta-1} - 2\right)$. For large enough Θ, the four one-step filter models have initial memory signals, in descending order, of $\{\mu_{\mathrm{Ar}}(0), \mu_{\mathrm{A0}}(0), \mu_{\mathrm{Rr}}(0), \mu_{\mathrm{R0}}(0)\} \sim \{3/2, 1, 3/4, 2/3\}/\Theta^2$. The R0 and Rr filters therefore have lower initial signals than the A0 and Ar filters. Conversely, $\mu_{\mathrm{R0}}(t)$ and $\mu_{\mathrm{Rr}}(t)$ are sustained for longer compared to $\mu_{\mathrm{A0}}(t)$ and $\mu_{\mathrm{Ar}}(t)$. This is as expected. By reflecting a filter back to just below threshold rather than injecting it to the zero state, a reflecting boundary filter makes a synapse's saturated strength states more stable. We can see this explicitly by comparing the slowest decaying modes in $\mu_{\mathrm{A0}}(t)$ and $\mu_{\mathrm{R0}}(t)$. For $\mu_{\mathrm{A0}}(t)$, the slowest decay rate is governed by $1 - \cos\frac{\pi}{2\,\Theta} \approx \frac{\pi^2}{8\,\Theta^2}$, while for $\mu_{\mathrm{R0}}(t)$, we have instead $1 - \cos\frac{\pi}{(3\,\Theta-1)} \approx \frac{\pi^2}{2(3\,\Theta-1)^2} \approx \frac{\pi^2}{18\,\Theta^2}$, where the approximate forms follow for Θ large enough. The slowest mode for the R0 filter therefore decays at a rate 4/9 times more slowly than that for the A0 filter.

For the S filter, its initial signal is exponentially suppressed as a function of Θ. Despite this small initial signal, $\mu_{\mathrm{S}}(t)$ rises and is strongly sustained, exhibiting plateau-like dynamics over orders of magnitude of time. This is true even for the larger values of Θ in Fig. 4, although $\mu_{\mathrm{S}}(t)$ remains under the cut-off of 10^{-3} for $\mu(t)$ used to plot these figures: such low values of $\mu(t)$ are biologically irrelevant because any biologically realistic value of N cannot be large enough to generate an adequate SNR. Nevertheless, were we to show values of $\mu(t) < 10^{-3}$, we would see in all cases $\mu_{\mathrm{S}}(t)$ sustained high enough for long enough that it would always eventually exceed the memory signals associated with the four one-step filter models, A0, Ar, R0 and Rr. As with the one-step filters, the S filter exhibits a multiplicative rise in its memory signal from its initial value to its peak value that is of order Θ. In summary, as a model, to be biologically useful the S filter must have a reasonably small filter threshold, perhaps not in excess of 10.

Although the cascade model is designed to have a high initial memory signal $\mu_{\mathrm{C}}(0) = 2/n$, we see from Fig. 4 that all the one-step filters quickly surpass the cascade's signal as it drops monotonically and theirs rise to their peaks. This occurs by at most $t \approx 10/r$ s, and usually much sooner, from the data in these figures for different Θ or n. Even the S filter, with its initially very suppressed signal, catches up with the cascade, for example in Fig. 4C, although somewhat later. Moreover, the filters' memory signals remain higher than the cascade's over ranges of $\mu(t)$ that are biologically relevant.

Having examined the first-order statistic $\mu(t)$ in Fig. 4, we now examine the second-order statistic $\sigma(t)^2$ and the SNR $\mu(t)/\sigma(t)$ in Fig. 5. In this figure, we do not approximate $\sigma(t)^2$, but rather compute it using numerical matrix methods. The covariance between pairs of synapses' strengths contributes to $\sigma(t)^2$, as can be seen in Eq. (2.16). This rather counter-intuitive covariance arises because synaptic modifications are driven by a continuous-time stochastic (here Poisson) process (Elliott & Lagogiannis, 2012). In Figs. 5A, C and E, we plot this covariance (or rather its square root) against time for Θ (or n) equal to 4, 7 and 10, respectively. We have not shown results for the Ar and Rr filters to avoid clutter and because they are very similar to those for the A0 and R0 filters. As we have observed before (Elliott & Lagogiannis, 2012), the cascade's covariance peaks for $t \approx 1/r$ s, and the amplitude of this peak is relatively insensitive to the choice of n. This is because of the presence of cascade states with high probabilities for making strength transitions. In contrast, the filters' covariances scale down rapidly as Θ increases. There is a minimum in any filter's covariance due to the peak in $\mu(t)$, and therefore two peaks in the covariance either side of this minimum. While the one-step filter models' two covariance peaks are of similar amplitudes, the second peak in that for the S filter is increasingly suppressed relative to its first peak as Θ increases. This suppression of the second peak serves to reduce the impact of the S filter's covariance on SNRs and therefore on SNR memory lifetimes.

In Figs. 5B, D and F we show the SNRs for the parameters corresponding to Figs. 5A, C and E, respectively, including those for the Ar and Rr filters, for $N = 10^4$ synapses. The cascade model's SNR is significantly undermined by its initially very high variance due to the high probability states that increase the model's initial signal (Elliott & Lagogiannis, 2012). For the parameters used in this figure, the one-step filter models' SNRs exceed that for the cascade by a time at most $t \approx 1/r$ s, and for smaller Θ, much more quickly. Thus, the cascade model's initially high $\mu(t)$ is swamped by its initially high $\sigma(t)$. The SNR $\mu(t)/\sigma(t)$ is not shown below unity on these graphs because SNR memory lifetimes are defined by the (possibly largest) solutions of $\mu(t)/\sigma(t) = 1$. We can read off or compare memory lifetimes for different models by observing where the lines for $\mu(t)/\sigma(t)$ intercept the abscissae on these graphs. The SNRs for the R0 and Rr filters are extremely similar and for larger times almost indistinguishable. These two filters always have higher SNRs for larger times than the A0 and Ar filters, and so have larger (and almost identical) SNR memory lifetimes than the A0 and Ar filters. The A0 and Ar filters have similar but not identical SNR profiles, with the A0 filter having a larger SNR memory lifetime than the Ar filter in these figures. For $\Theta = 7$ and with $N = 10^4$, in Fig. 5D, the S filter only just has the largest SNR memory lifetime. Because of the S filter's memory signal dynamics, whether its SNR ever reaches unity and exceeds that of the other filters depends very strongly on the choice of N.

In the Θ–N or n–N plane, in Fig. 6, we indicate which of the six models (the five different filters and the cascade) has the largest SNR memory lifetime, using $\sigma(t)^2 \approx 1/N$ for convenience. For SNR lifetimes in excess of $t \approx 10/r$ s, this approximation is good: for longer memory lifetimes, the covariance has died away and does not contribute significantly to $\sigma(t)^2$, and the $\mu(t)^2$ term in the non-covariance contribution to $\sigma(t)^2$ is usually negligible. In the blacked-out region, SNRs never exceed unity for all six models. We see that the filter models are superior to the cascade model in all regions of parameter space except higher Θ and n and smaller N (N under around 700 for $n = 20$). Reflecting boundary filters are superior to one-step double absorbing boundary filters, except in a small sliver of parameter space. The apparent transition from R0 to Rr for fixed Θ as N increases is somewhat misleading: the SNR memory lifetimes of both filters are, as we have seen above, extremely similar, so although the transition is real, both filters' memory performances are essentially identical and any difference is entirely marginal. For a range of Θ between 6 and 12 inclusive and for N large enough, the S filter is superior to the Rr (and R0) filter. These large values of N are required to overcome the S filter's suppressed memory signal. If we were not to consider the S filter in Fig. 6, the Rr filter would have the largest SNR memory lifetimes in the region occupied by the S filter, except for a single line for $n = 11$ and N in excess of approximately 850,000, where the cascade would have only a very marginally larger SNR memory lifetime than the Rr filter. Of course, values of N that high are not, in any event, biologically plausible.

We take cross-sections through the plane in Fig. 6 by plotting SNR memory lifetimes for fixed Θ or n as a function N in Fig. 7. We see in all cases that the R0 and Rr filters' memory lifetimes are virtually identical, except for small N, where small differences in the onsets of lifetimes exist. The sudden onset of (or bifurcation in) filters' memory lifetimes as N increases is due to encoding failure for small N: the SNR does not ever exceed unity, so according to the SNR memory lifetime definition, there is no solution for τ_{snr}. The S filter requires the largest value of N for the onset of its SNR memory lifetimes. Indeed, for $\Theta = 13$ in Fig. 7D, the S filter's results are only just visible for N

very close to 10^6. We see that there are regimes, for Θ or n equal to 10 or 13 in these figures, in which the cascade's memory lifetimes exceed those for the A0 and Ar filters but are lower than those for the R0 and Rr filters, for larger values of N. If we showed results for $n = 11$, there would be a small region for $N \gtrapprox 850,000$ in which the cascade's lifetimes would exceed those of the Rr filter, but would be under those of the S filter.

Fig. 6 indicated a sliver of parameter space in which the cascade model has the highest SNR memory lifetimes, above the blacked-out region where all models fail to encode the tracked memory adequately but below the sliver of parameter space in which the A0 filter has the highest SNR memory lifetimes. From the cross-sections in Fig. 7 we see, however, that for small N in those regions where the cascade model apparently outperforms the various filter models, the cascade's SNR memory lifetimes are in fact minuscule, of order $\tau_{\text{snr}} \approx 1/r$ s. With an increase in N and the onset of filters' SNR memory lifetimes, they essentially jump immediately to substantially higher values than the cascade's. Even for $\Theta = 20$ or $n = 20$, at $N \approx 700$ at the transition between the cascade model and the A0 filter model in Fig. 6, the cascade model's SNR memory lifetime is only $\tau_{\text{snr}} \approx 3/r$ s while the A0 filter's lifetime is nearly $\tau_{\text{snr}} \approx 200/r$ s. The entire region, then, occupied by the cascade in Fig. 6 corresponds to extremely weak memory encoding and very short SNR memory lifetimes. It is true that for smaller N, at the onset of (bifurcation in) SNR memory lifetimes for filter models, the bifurcation point corresponds to a single time point at which the memory signal peak just reaches an SNR of unity. However, even a small increase in N will create more significant regions in time in which the SNR exceeds unity, giving rise to very rapid increases in SNR memory lifetimes. There are no such bifurcation dynamics for the cascade model because its memory signal falls monotonically.

We can derive approximate expressions for $\tau_{\rm snr}$ in the filter models by considering only the slowest decaying modes in the expressions for $\mu(t)$ and then solving $\mu(\tau_{\rm snr}) = 1/\sqrt{N}$. We do this only for the zero injection models. For the A0, R0 and S filters, we obtain

$$\tau_{\rm snr}^{\rm A0} \approx \frac{4\,\Theta^2}{\pi^2} \log_e \frac{256}{\pi^4} \frac{N}{\Theta^2}, \tag{5.1a}$$

$$\tau_{\rm snr}^{\rm R0} \approx \frac{(3\,\Theta-1)^2}{\pi^2} \log_e \frac{256}{3\pi^2} \frac{N}{(3\,\Theta-1)^2}, \tag{5.1b}$$

$$\tau_{\rm snr}^{\rm S} \approx \frac{3}{4}\, 2^{\Theta} \log_e \frac{4}{9} \frac{\Theta^2 N}{2^{2\Theta}}, \tag{5.1c}$$

where strictly these expressions are valid only for large Θ but in practice they give good approximations even for small Θ. We have used the approximate form for $\mu_{\rm S}(t)$ in Eq. (4.27) in deriving $\tau_{\rm snr}^{\rm S}$. These analytical results agree extremely closely with the results shown in Fig. 7, being essentially perfect for larger values of N. There are discrepancies only for smaller N in the vicinities of the onsets of memory lifetimes because these simple analytical results do not exhibit bifurcations as N increases. Nevertheless, pseudo-bifurcations are present in the sense that we demand that $\tau_{\rm snr} > 0$, so we must demand that the arguments of the logarithms exceed unity. Thus, we require $N_{\rm A0} > 0.38\,\Theta^2$, $N_{\rm R0} > 0.12\,(3\,\Theta - 1)^2$ and $N_{\rm S} > 2.25 \times 2^{2\,\Theta}/\Theta^2$. For Θ in excess of around 7,

N_{S} in fact gives an extremely good approximation to the location of the real bifurcation. We may turn these limits around and regard them as limits on filter sizes. If we take the maximum number of synapses possible in real neurons to be around 250,000, then maximum filter sizes for these three models are $\Theta_{\mathrm{A0}} \lessapprox 810$, $\Theta_{\mathrm{R0}} \lessapprox 490$ and $\Theta_{\mathrm{S}} \lessapprox 12$. For the A0 and R0 filters, the biological upper limits on Θ are likely to be very much lower than these theoretical extremes.

Previously we have observed that SNR memory lifetimes can give a very misleading indication of memory performance for small N (Elliott, 2014). While SNR lifetimes suggest a minimum value of N for successful memory encoding, we found that with MFPT memory lifetimes, such was not that case. Only by taking an asymptotically valid, large N form for MFPT memory lifetimes is it possible to observe a minimum value of N similar to that seen for SNR memory lifetimes. However, such a form is by definition valid only for large N, and any such minimum value of N is purely an artifact of the approximation and is not seen in the full form. These results were derived for the simplest possible model of synaptic plasticity applied to memory storage, but they are likely to generalise to the more complicated models considered here. In Fig. 8, we therefore examine MFPT memory lifetimes for the specific choice of Θ or n equal to 10. This figure is therefore analogous to Fig. 7C, although we have run simulations for Fig. 8 only up to $N = 10^5$. In Fig. 8A, we show the MFPT memory lifetimes for all six models. We do not see any regions of encoding failure. All filter models have MFPT memory lifetimes even for $N = 100$, including the S filter model. Indeed, the S filter model has for all plotted N a very large τ_{mfpt}, broadly consistent with its τ_{snr} in Fig. 7C for larger N. The R0 and Rr filter MFPT memory lifetimes are extremely similar, and the A0 and Ar filter MFPT lifetimes are also similar, although not so similar as R0 and Rr. These qualitative features are precisely as in Fig. 7C for SNR lifetimes. To facilitate comparison, we plot τ_{mfpt} and τ_{snr} in Fig. 8B for the A0 and R0 filters and for the cascade model; the other filters are not plotted to avoid clutter. The numerical values for the filters' memory lifetimes according to both metrics are very similar, except for smaller N where SNR lifetimes would indicate memory encoding failure. There appear to be differences between the cascade model's two memory lifetimes, but the numerical differences are not large and are merely magnified in a log-log plot when the absolute values are small. We see that although according to the SNR metric, the cascade model outperforms the filter models in the small N region (although its SNR memory lifetimes are minuscule here), according to the MFPT metric, it is the filters that outperform the cascade in this small N region, despite the cascade being designed to operate well in such regions by having a large initial signal.

What accounts for these differences between SNR and MFPT memory lifetimes for small N? The mean memory signal $\mu(t)$ is precisely that: the *mean* of the random variable $h(t)$. The random variable $h(t)$ has a distribution around its mean $\mu(t)$. While $\sqrt{N}\,\mu(t)$ (the approximated SNR) may be below unity at $t = 0$ s and stay below unity for all time, some realisations of $h(t)$ will strongly encode the initial memory while others will not. A first passage time memory lifetime will assign a lifetime of 0 s to those realisations in which the initial memory is not strongly encoded while those realisations in which it is strongly encoded will be associated with non-zero lifetimes. The MFPT memory lifetime is, by definition, an average over all these possibilities. While an SNR memory lifetime essentially collapses the distribution of $h(t)$ to a single point, an MFPT memory lifetime takes into account the stochasticity in the dynamics of $h(t)$ over all possible realisations. However, if strong encoding of the initial memory becomes increasingly improbable as N decreases, then we would expect the variance in first-passage-time-defined memory lifetimes to increase. We confirm this in Fig. 8C, in which we plot τ_{mfpt} and the one standard deviation region around it. We do this only for the R0 filter model and the cascade model, although we obtain similar results for the other filter models. We see explicitly that for smaller N, τ_{mfpt} is swamped by its variance while for larger N and stronger encoding, τ_{mfpt} can be distinguished from zero at the one standard deviation level. Interestingly, we see that the cascade model's τ_{mfpt} requires a higher value of N to be distinguishable from zero than the R0 filter's τ_{mfpt}, and this is true for all the one-step filter models. This is despite the fact that according to an SNR criterion for $n = 10$, the cascade has non-zero SNR memory lifetimes (and thus successful, strong encoding) over the entire range of N considered here. For the S filter, the standard deviation in τ_{mfpt} swamps it over the entire range of displayed N. This is hardly surprising, given the very weak encoding of the initial memory by the S filter even for sizeable N.

6 Lahiri & Ganguli's "Memory Frontier"

Lahiri & Ganguli (2013) adopt a very powerful and extremely interesting approach to the problem of memory lifetimes in models such as those considered here. They perform a general analysis, argued to be valid for *any* model of synaptic plasticity, that seeks to establish an upper bound on the SNR envelope $\mathcal{SNR}(t)$ and that can therefore in principle provide a theoretically optimal SNR memory lifetime. Their SNR is $\mathcal{SNR}(t)$ rather than our $\mu(t)/\sigma(t)$, but as we saw above in section 2.3, with the approximation that $\sigma(t) \approx 1/\sqrt{N}$, these two SNRs are identical and, up to the overall factor of $\sqrt{N}$, just the perceptron's (perhaps excess) mean activation.

They establish two bounds on $\mathcal{SNR}(t)$, given by

$$\mathcal{SNR}(0) \leq \sqrt{N}, \tag{6.1a}$$

$$\int_0^\infty dt\, \mathcal{SNR}(t) \leq \tfrac{1}{r}\sqrt{N}(M-1), \tag{6.1b}$$

where M is the dimensionality of the state vectors describing synapses' full states, so that, equivalently, the full state transition matrices are $M \times M$. For our models, $M = 2(2\,\Theta - 1)$ and for the cascade model, $M = 2\,n$. The first bound is on the initial SNR and the second bound is on the area under the SNR curve. Discarding the overall scale factor $\sqrt{N}$ and setting $r = 1$ Hz for convenience and without loss of generality, we may alternatively write these bounds in terms of the excess mean perceptron activation, defined to be $\eta(t) =$

$\mu(t) - \mu(\infty)$, as

$$\eta(0) \leq 1, \tag{6.2a}$$

$$\mathcal{A} = \int_0^\infty dt\, \eta(t) \leq M - 1, \tag{6.2b}$$

which defines $\mathcal{A}$. We note that the first bound is essentially trivial: $\mu(t) \leq 1$ by the definition of $h(t)$ and we maximise the difference $\mu(t) - \mu(\infty)$ for balanced processes with $g_\pm = \frac{1}{2}$ for which $\mu(\infty) \equiv 0$. Thus, $\eta(0) = \mu(0) - \mu(\infty) \leq 1$ is essentially automatic, when viewed from our perceptron formulation. We also observe that

$$\lim_{s \to 0} \widehat{\eta}(s) = \lim_{s \to 0} \int_0^\infty dt\, \eta(t)\, \exp(-s\, t) = \int_0^\infty dt\, \eta(t) \equiv \mathcal{A}, \tag{6.3}$$

so that in the limit $s \to 0$, the Laplace transform of $\eta(t)$ reduces identically to the area under the (rescaled) SNR curve. For balanced processes, we have that $\mathcal{A} \equiv \lim_{s \to 0} \widehat{\mu}(s)$. Laplace transforming Eq. (2.28) (with $r = 1$ Hz), we have for a general model that

$$\widehat{\eta}(s) = 2 g_+ g_- \boldsymbol{\Omega}_2^{\mathrm{T}} \big[(1+s)\, \mathbb{I}_2 - \mathbb{M}_2 \big]^{-1} \big(\mathbb{M}_2^+ - \mathbb{M}_2^- \big) \boldsymbol{A}_2. \tag{6.4}$$

We cannot just set $s = 0$ in this equation because the matrix $\mathbb{I}_2 - \mathbb{M}_2 = -\mathbb{G}_2$ is not invertible due to the existence of the equilibrium eigenvector of $\mathbb{M}_2$. However, we may expand $\big[(1+s)\, \mathbb{I}_2 - \mathbb{M}_2\big]^{-1}$ as a power series in $\mathbb{M}_2$, which amounts to defining the matrix exponential in Eq. (2.28) via its power series.

We may then set $s = 0$ to give

$$\lim_{s \to 0} \widehat{\eta}(s) = 2 g_+ g_- \boldsymbol{\Omega}_2^{\mathrm{T}} \sum_{\alpha=0}^{\infty} \mathbb{M}_2^{\alpha} \big(\mathbb{M}_2^{+} {-} \mathbb{M}_2^{-} \big) \boldsymbol{A}_2. \tag{6.5}$$

The sum is convergent because $\mathbb{M}_2^{\alpha} \to \mathbb{A}_2$ as $\alpha \to \infty$, where $\mathbb{A}_2$ is a matrix all of whose columns are the equilibrium distribution $\boldsymbol{A}_2$, or $\mathbb{A}_2 = \boldsymbol{A}_2 \big(\boldsymbol{n}^{\mathrm{T}} \big| \boldsymbol{n}^{\mathrm{T}} \big)$. Following Lahiri & Ganguli, we may write $\mathbb{M}_2^{+} {-} \mathbb{M}_2^{-} = d\mathbb{G}_2/dg_+$ because $\mathbb{G}_2 = g_+ \mathbb{M}_2^{+} + (1 - g_+) \mathbb{M}_2^{-} - \mathbb{I}_2$, and then use $\mathbb{G}_2 \boldsymbol{A}_2 = \boldsymbol{0}$ to write $(d\mathbb{G}_2/dg_+) \boldsymbol{A}_2 = -\mathbb{G}_2 \, d\boldsymbol{A}_2/dg_+$. We then have

$$\lim_{s \to 0} \widehat{\eta}(s) = -2 g_+ g_- \boldsymbol{\Omega}_2^{\mathrm{T}} \sum_{\alpha=0}^{\infty} \mathbb{M}_2^{\alpha} \big(\mathbb{M}_2 - \mathbb{I}_2 \big) \frac{d\boldsymbol{A}_2}{dg_+} \equiv 2 g_+ g_- \boldsymbol{\Omega}_2^{\mathrm{T}} \frac{d\boldsymbol{A}_2}{dg_+}, \tag{6.6}$$

which is Lahiri & Ganguli's result for $\mathcal{A}$, on which they then obtain the bound in Eq. (6.2b).[3] Finally, for an arbitrary model, dropping the subscript "2" for

[3] It may appear that we have written $\sum_{\alpha=0}^{\infty} \mathbb{M}_2^{\alpha}$ as the inverse of $\mathbb{I}_2 - \mathbb{M}_2$ in Eq. (6.6), but the sum is not convergent and the inverse does not exist because of the unit eigenvalue of $\mathbb{M}_2$. Rather, we must strictly regard $\sum_{\alpha=0}^{\infty} \mathbb{M}_2^{\alpha} \big(\mathbb{M}_2 - \mathbb{I}_2 \big)$ as always acting on vectors, and then (absolute) convergence is assured.

convenience just in this set of inequalities, we note that

$$\begin{aligned}
\mathcal{A} &= 2g_+g_-\boldsymbol{\Omega}^{\mathrm{T}}\frac{d\boldsymbol{A}}{dg_+} \\
&\le 2g_+g_-\sum_{i=1}^{M}\left|\frac{dA_i}{dg_+}\right| \\
&= 2g_+g_-\left\{\left[\sum_{i=1}^{M-1}\left|\frac{dA_i}{dg_+}\right|\right]+\left|\sum_{i=1}^{M-1}\frac{dA_i}{dg_+}\right|\right\} \\
&\le 4g_+g_-\sum_{i=1}^{M-1}\left|\frac{dA_i}{dg_+}\right| \\
&\le (M-1)\,4g_+g_-\max_{i\in\{1,\ldots,M\}}\left|\frac{dA_i}{dg_+}\right| \\
&\le (M-1)\sup_{g_+\in[0,1]}\max_{i\in\{1,\ldots,M\}}\left|4g_+(1{-}g_+)\frac{dA_i}{dg_+}\right|,
\end{aligned} \tag{6.7}$$

where, in the third line, we have used the fact that $\sum_{i=1}^{M} dA_i/dg_+ \equiv 0$ to express any one component (which we took as dA_M/dg_+) in terms of the others. Whether it is possible to obtain a bound on the expression in the last line via elementary methods remains to be determined, but we have at least the critical dependence on $M-1$ in Eq. (6.2b) via elementary methods.

By performing an eigen-expansion of the matrix $\mathbb{G}_2$, Lahiri & Ganguli write $\eta(t)$ in the form

$$\eta(t) = \sum_a \mathcal{I}_a \exp(-t/\tau_a), \tag{6.8}$$

where $\lambda_a = -1/\tau_a$ are the eigenvalues of $\mathbb{G}_2$ and the coefficients $\mathcal{I}_a$ can be computed from Eq. (2.28) in terms of the left and right eigenvectors of $\mathbb{G}_2$.

The two bounds in Eq. (6.2) then imply the two constraints,

$$\sum_a \mathcal{I}_a \leq 1, \tag{6.9a}$$

$$\sum_a \mathcal{I}_a \tau_a \leq M - 1. \tag{6.9b}$$

They seek to maximise $\eta(t)$ at any given time with respect to the parameters $\mathcal{I}_a$ and τ_a subject to these two constraints, although they acknowledge that the constraints are likely not complete, in the sense that some choices of $\mathcal{I}_a$ and τ_a that satisfy these constraints may not in fact be realisable for any actual synaptic model. That is, they maximise $\eta(t)$ essentially ignoring the fact that the $\lambda_a = -1/\tau_a$ are the eigenvalues of a generating matrix $\mathbb{G}_2$ of a stochastic process and that the $\mathcal{I}_a$ are determined from its left and right eigenvectors. They obtain a theoretical bound on $\eta(t)$ at each point in time that arises from precisely one non-zero value for some $\mathcal{I}_a$, so that $\eta(t)$ is at each time point just a single exponential. By comparing this theoretical bound on $\eta(t)$ to real models via numerical optimisation or hand searches, they find that their bound on $\eta(t)$ can be achieved at very small times ($rt \ll 1$) and at large times, but that their search does not find actual models that can achieve or indeed even come close to their bound at intermediate times (see their Fig. 4).

Saturation of their bound near $t = 0$ s is essentially trivial because the bound $\eta(0) \leq 1$ is itself essentially trivial, as we saw above. In addition, saturation of their bound at large times is inevitable because $\eta(t) \to 0$ as $t \to \infty$ and the slowest decaying mode in $\eta(t)$, as we have exploited above in section 5, always takes over, leaving a single mode present in $\eta(t)$ for t large enough. These two extremes of small and large t are not especially informative. In many respects, the dynamics of $\eta(t)$, or $\mu(t)$, at intermediate times are precisely those that are of the greatest interest. As we have seen in our own filter models, in these intermediate time regimes, the memory signal need not even be monotonic decreasing, but instead can rise and fall. Unfortunately, it is precisely in this intermediate time regime that Lahiri & Ganguli fail to achieve their theoretical bound in actual, realisable models, and they themselves concede that their bound is likely not achievable in this regime because their two constraints in Eq. (6.9) likely do not bite strongly enough.

Much more important, however, in terms of attempting to apply Lahiri and Ganguli's results to our models, is the major assumption made in the eigenexpansion of $\mathbb{G}_2$, leading to Eq. (6.8). They have assumed that the matrix $\mathbb{G}_2$ is not defective, i.e. they have assumed that $\mathbb{G}_2$ possesses a complete set of eigenvectors. This is not a valid assumption in general, however. For example, the A0 and S filter models considered above possess defective generating matrices. In general, then, we would need to employ generalised eigenvectors and use the Jordan normal form, and thus we would expect the expansion coefficients $\mathcal{I}_a$ in Eq. (6.8) not to be constants but rather to be polynomials in t with their degrees determined by any degeneracy in the eigenvalues of $\mathbb{G}_2$. Despite this expectation, it is striking that the form for $\mu_{\text{A0}}(t)$ in Eq. (4.24) is precisely of the form given in Eq. (6.8) [with $\mu(\infty) = 0$ so that $\eta(t) \equiv \mu(t)$], even though the A0 filter model's generating matrix is defective for $\Theta > 2$. However, we notice that some of the $\mathcal{I}_a$ are negative. Indeed, it is necessarily the case that in a model with a non-monotonic memory signal, some of the $\mathcal{I}_a$ are negative: all the $\mathcal{I}_a$ being non-negative would entail only monotonic-decreasing $\eta(t)$. This is in fact true in all five filter models. It is unclear whether Lahiri & Ganguli have implicitly assumed that all the $\mathcal{I}_a$ are non-negative in their optimisation search. The assumption of a non-defective $\mathbb{G}_2$ is not, however, the only assumption in writing down Eq. (6.8). Another major assumption is that the eigenvalues λ_a are real. Of course, $\mu(t)$ or $\eta(t)$ must be real, but the eigenvalues λ_a and therefore the expansion coefficients $\mathcal{I}_a$ need not be. Indeed, if we explicitly examine the S filter, then we find that not only is its generating matrix defective, but also it possesses some complex eigenvalues. For example, for $\Theta = 3$, we find that $\mu_S(t)$ takes the form

$$\begin{aligned}\mu_{\mathrm{S}}(t) = \big[a_1 \cos(\gamma_1 t) + b_1 \sin(\gamma_1 t)\big] \exp(\gamma_2\, t) \\ + c_1 \exp(\lambda_3\, t) + \big(d_1 + e_1 t\big) \exp(-t),\end{aligned}$$

where the various exponents involve the roots of a cubic and the coefficients involve the roots of a sextic. We see not only the appearance of a first-order polynomial in t due to the defective nature of $\mathbb{G}_2$, but we also see the presence of its complex eigenvalues indicated by the appearance of trigonometric functions. Further, these oscillatory components, while very rapidly decaying and minuscule, are superimposed on the overall rise and fall of the memory signal $\mu_{\mathrm{S}}(t)$. Such oscillations are perhaps to be expected in the presence of the rectifying behaviour present in the S model. With a non-defective matrix $\mathbb{G}_2$, the constraints in Eq. (6.9) must remain valid even in the presence of complex eigenvalues $\lambda_a = -1/\tau_a$ and complex coefficients $\mathcal{I}_a$, but then the optimisation search for models that achieve the bounds in Eq. (6.2) should be extended to include not only the possibility of negative coefficients $\mathcal{I}_a$ but also complex eigenvalues λ_a and complex coefficients $\mathcal{I}_a$. However, with a defective matrix $\mathbb{G}_2$, the eigen-expansion in Eq. (6.8) must be modified to permit the appearance of polynomial coefficients and therefore the constraint in Eq. (6.9b) requires modification. The constraint in Eq. (6.9a) remains essentially unchanged if we interpret $\mathcal{I}_a$ to mean the polynomials evaluated at $t = 0$.

These considerations lead us to suspect that Lahiri & Ganguli's SNR envelope bound at intermediate times is of limited utility because it is almost certainly not achievable by any actual model. Furthermore, we have seen that the assumptions involved in the implementation and indeed formulation of their constraints in Eq. (6.9) are not general enough and do not apply to our filter models. Nevertheless, for the sake of completeness, it is interesting to examine $\lim_{s \to 0} \widehat{\mu}(s) = \mathcal{A}$ in the five filter models above, and in the cascade model. We find that,

$$\lim_{s\to 0} \widehat{\mu}(s) = \begin{cases} \Theta & \text{for A0 model,} \\ \frac{1}{3}(2\,\Theta + 1) & \text{for Ar model,} \\ \frac{(2\,\Theta-1)(7\,\Theta-1)}{3(3\,\Theta-1)} & \text{for R0 model,} \\ \frac{3\,\Theta(2\,\Theta-1)}{4\,\Theta-1} & \text{for Rr model,} \\ \Theta & \text{for S model,} \\ \frac{n^2-n+2}{2\,n} & \text{for cascade model,} \end{cases} \tag{6.10}$$

and if we divide through by $M = 2\,(2\,\Theta - 1)$ or $M = 2\,n$ and take the limit of large Θ or n, we obtain the normalised areas, in order of the models listed in Eq. (6.10), of $\frac{1}{4} = 0.25$, $\frac{1}{6} \approx 0.167$, $\frac{7}{18} \approx 0.389$, $\frac{3}{8} = 0.375$, $\frac{1}{4} = 0.25$, and $\frac{1}{4} = 0.25$. The R0 model therefore has the largest normalised area, with the Rr model following closely behind. The theoretical bound, according to Eq. (6.2b), is unity. It is interesting, however, to compare the A0 and S models, both of which have $\mathcal{A} = \Theta$, while their initial signals $\mu_{\mathrm{A0}}(0) = 1/\Theta^2$ and $\mu_{\mathrm{S}}(0) = 1/(3 \times 2^{\Theta-1} - 2)$ are in general vastly different. That the S model has the same SNR area as the A0 model despite its vastly suppressed initial signal indicates that its memory signal is sustained for very much (indeed, vastly) longer than that for the A0 model. In general, then, the S model has a (vastly) longer SNR memory lifetime (albeit requiring absurdly large values of N) than the A0 model. It therefore appears that the interaction between $\mu(0)$ [or $\eta(0)$ in general] and $\mathcal{A}$ can be very counter-intuitive, especially in the presence of non-monotonic and strongly-sustained memory signals. A consideration of the SNR envelope and its associated area appears predicated on the tacit assumption that the SNR can only fall monotonically.

7 Discussion

We have proposed in earlier work that synapses may implement plasticity induction signal filtering mechanisms that control fluctuations in synaptic strength driven by ongoing synaptic plasticity (Elliott, 2008). These integrate-and-express mechanisms powerfully control fluctuations in developmental processes that destabilise developmentally-relevant patterns of synaptic connectivity (Elliott & Lagogiannis, 2008; Elliott, 2011a,b). In the context of the storage of memories, these mechanisms lead to enhanced memory lifetimes and a rise in the initial memory signal, actually driven by ongoing synaptic plasticity, that is in radical contrast to related but non-integrative models of synaptic plasticity (Elliott & Lagogiannis, 2012). As binary-strength synapses automatically provide the limits on synaptic strengths than turn memory systems into palimpsests (Nadal *et al.*, 1986; Parisi, 1986), many models of memory storage use binary-strength synapses (see, for example, Willshaw *et al.*, 1969; Tsodyks, 1990; Fusi *et al.*, 2005), although multi-level, discrete-state synapses also naturally provide such limits (see, for example, Barrett & van Rossum, 2008; Huang & Amit, 2011). In our earlier work on memory lifetimes with integrate-and-express models, we therefore used binary-strength synapses, but we specifically employed a filter mechanism that was independent of synaptic strength, so that it would generalise immediately to any number of states of synaptic strength (Elliott & Lagogiannis, 2012).

Here, we have compared and contrasted a variety of different filtering mechanisms in the context of memory lifetimes, with some of these mechanisms being tailored to specifically binary-strength synapses. Filters with two absorbing boundaries (A0, Ar and S) generalise to discrete state (and effectively unbounded state) synapses immediately. However, filters with an absorbing boundary and a reflecting boundary are suited to synapses with saturated strength states and best suited to binary-strength synapses. Although a reflecting boundary can always be employed by a general, discrete-state synapse at its upper and lower strength limits, the presence of the reflecting boundaries will exert the greatest influence on synaptic dynamics for binary-strength synapses. If a discrete-state synapse has many (and perhaps even a dynamically variable number of) states of strength, we consider it unlikely that the plasticity signal integration mechanism would be sensitive to a synapse's current strength state, and specifically whether the synapse is saturated. However, if a discrete-state synapse possesses only a few states of strength (perhaps only two or three), then it is entirely possible that the integrative- and strength-change-processes become strongly coupled together in order to optimise functioning.

Considering for the moment only zero injection processes, the R0 filter model is nearly everywhere in parameter space superior to the A0 filter model in terms of SNR memory lifetimes. The reflecting boundaries make weak (lower) and strong (upper) strength states more stable because the escape times through the absorbing boundaries from just below the reflecting boundaries are longer than the escape times through the absorbing boundaries from the zero filter state. This difference in stability is seen even in the escape times from the zero filter state, with the R0 filter taking nearly three times longer on average to reach an absorbing threshold compared to the A0 filter [see Eq. (3.40)]. For the mean memory signal, the slowest decaying mode with the R0 filter has a rate governed, approximately, by $\pi^2/(18\,\Theta^2)$, while for the A0 filter, the slowest mode has a rate governed by around $\pi^2/(8\,\Theta^2)$. This difference does not translate directly into a roughly 2.25-fold difference in SNR memory lifetimes because of differences in the initial and peak memory signals between the two models, but for larger N, the ratio in SNR lifetimes is a little under 2. For more general, discrete-state synapses, we would expect this enhanced stability of the saturated strength states also to be apparent in enhanced memory lifetimes compared to purely absorbing boundaries, but the impact of the saturated states would be expected to be progressively diluted as the total number of strength states increases. In that sense, reflecting boundary filters are expected to be have the greatest impact on binary-strength synapses.

The A0 filter is superior to the R0 filter in terms of SNR memory lifetimes only in a thin sliver of parameter space for small N. Just like the cascade model's sliver of superior parameter space, the A0 filter is better than the R0 filter only because, from an SNR perspective, it encodes memories somewhat more strongly than the R0 filter. This arises directly because the equilibrium distributions of the R0 filter states are shifted somewhat towards the reflecting boundaries and away from the absorbing boundaries. It is the probability of occupation of filter states just below absorbing boundaries that determines the initial memory signal, and hence the strength of the initial encoding. This difference is specific to the SNR memory lifetime metric. With an MFPT memory lifetime metric, the R0 filter is always superior to the A0 filter.

With random rather than zero injection, we found that the R0 and Rr models are virtually identical in terms of their memory performance, whether determined by SNRs or by MFPTs, with any differences between these two filters being entirely marginal. For the A0 and Ar models, differences are discernible but they perform very similarly. Specifically, for larger times, their memory signals converge and for larger N, their SNR memory lifetimes also converge. We considered random injection as a worst-case-scenario, avoiding excessive parameter dependence, in order to determine how robust filtering mechanisms are against the possibility of noise in the zero injection process at filter threshold. It is rather striking that the precise details of the injection process at filter threshold appear to make relatively little and indeed in some cases hardly any difference to memory performance. How can such insensitivity be explained? Provided that the injection distribution is symmetrically distributed around the zero filter state and not skewed towards an absorbing threshold, we would expect that contributions from filter states equidistant from the zero filter state will, roughly speaking, average out to a contribution similar to that from the zero filter state. For example, for a strong synapse, a filter at $I = -1$ will escape slightly more quickly through the lower boundary than it would from $I = 0$, but conversely, a filter at $I = +1$ will escape slightly more slowly than it would from $I = 0$. The contributions from $I = -1$ and $I = +1$ will, roughly speaking, average out to that from the $I = 0$ state. The details of the injection process upon symmetric injections around the zero filter state should not, therefore, be too important. However, if targeting the zero state upon injection exhibits a systematic bias towards the upper or lower filter threshold, then we should expect more sensitivity to the details of the injection process.

The four one-step filters, although controlling fluctuations in synaptic strength, are still subject to them. We might describe these filters as lacking intentionality: they are not “looking” for a particular sequence of induction signals as the cue on which to express synaptic plasticity, but rather will express plasticity when there is a certain excess of induction signals of one sign compared to the other sign. How this excess is arranged in the overall sequence of induction signals is irrelevant, provided that boundaries are not hit in the meantime. The super or S filter, however, may be described as intentional: it insists on a sequence of Θ consecutive induction signals of the same sign; a signal with the opposite sign will reset the filter to the zero state. Essentially, the S filter counts the number of consecutive same-sign induction signals and expresses plasticity if this number reaches Θ. Of course, such a sequence of Θ same-type induction signals can arise purely as a fluctuation, but only with the vastly suppressed probability of $2^{-\Theta}$ for balanced potentiation and depression.

This suppression is reflected in the mean escape time from the zero filter state given in Eq. (3.40) and in the slowest decay mode in the mean memory signal, whose approximate form is given in Eq. (4.27), where the rate is controlled by the factor $1/\left(3 \times 2^{\Theta-1} - 2\right)$. Even for relatively small Θ, a plateau-like feature sustained over several orders of magnitude in time is present in the mean memory signal. Furthermore, the second peak in the S filter's covariance is strongly suppressed, even compared to the overall suppression of all filters' covariances with increasing Θ. This suppression further helps to sustain the plateau-like feature in the S filter's SNR. This strong, sustained feature can increase memory lifetimes compared to the other filter models. Unfortunately, however, the S filter's initial memory signal, and its peak memory signal, are also suppressed, because of the strong and inevitable clustering of the S filter's equilibrium distribution of filter states around the zero state. Although strongly sustained, the S filter's memory signal is therefore initially rather weak compared to the other filters' signals, at least until those filters' signals have decayed below the S filter's sustained signal. The S filter's rather weak signal therefore requires large N for its SNR to exceed unity. Imposing a maximum, biologically plausible limit on N of around 250,000, we saw that the S filter requires Θ under approximately 12 for plausibility. For Θ above this, the model cannot generate a memory signal strong enough to be of use in a real, biological system. In contrast, the other, one-step filters can operate with Θ of order several hundred, although in practice it is likely that this vastly overestimates the available pool of large macromolecules available at single synapses for instantiating any putative synaptic filter (see, for example, Harris & Stevens, 1989; Nusser *et al.*, 1998; Bagal *et al.*, 2005; Miller *et al.*, 2005; Asrican *et al.*, 2007).

We have also compared the various filter models to the cascade model of synaptic plasticity (Fusi *et al.*, 2005). Filters are superior in almost all regions of biologically-relevant parameter space, by which we essentially mean N under somewhere between 10^5 and 10^6: Purkinje cells may have up to perhaps around 2.5×10^5 synapses (Napper & Harvey, 1988) although most neurons have considerably fewer synapses, on the order of 10^3 or 10^4. The cascade model is apparently superior only for larger n and small N, but in practice its SNR memory lifetimes in this region are minuscule. Furthermore, its MFPT memory lifetimes in this region are both small and dominated by their variance, and in fact smaller than the filter models' MFPT memory lifetimes in this region. In this region, no model performs well and comparison is meaningless. Overall, however, the cascade model fails in its own terms. Its state transitions are designed to ensure a high initial memory signal and a sustained memory signal. But, as we have shown before (Elliott & Lagogiannis, 2012), the very states that are introduced into the model to ensure a high initial signal are precisely the same states that induce a large covariance between synapses' strengths, significantly reducing the SNR for earlier times. The high initial signal is therefore swamped by the variance. The sustained signal for longer memory lifetimes is inferior to the signals from filter dynamics in biologically-relevant regions of parameter space. The cascade model only starts to compete with filters for values of n that are so large that the cascade model's memory signal is well below any biologically-relevant level, or correspondingly, requires implausibly large values of N (i.e. $N > 10^6$) to lift the SNR above unity. Finally, although filter-based models generalise immediately to discrete, multi-level synapses, even with reflecting boundaries, the structure of the cascade model appears wedded to binary-strength synapses and does not appear to generalise easily.

Real memory systems are organised in a spectacularly complicated fashion (Eichenbaum & Cohen, 2001) and even the paradigmatic CA3–CA1 hippocampal synapse is of almost infinite complexity compared to modellers' idealisations of it (Andersen *et al.*, 2007). Synaptic plasticity and memory phenomena exhibit processes occurring on multiple timescales, from seconds to minutes to hours to days, with the former believed to provide the mechanistic underpinnings and therefore explanation of the latter (Roberson *et al.*, 1996; Heisenberg, 2003; Reymann & Frey, 2007). Synaptic plasticity is organised into at least three separate phases, with short-term plasticity, early-phase plasticity and late-phase plasticity, being characterised by their dependence on kinase and phosphatase activity or their dependence on protein synthesis (Roberson *et al.*, 1996; for models of the transition to late-phase synaptic plasticity, see Clopath *et al.*, 2008, Barrett *et al.*, 2009, Päpper *et al.*, 2011). Moreover, memory, even in single systems, is under executive control with the integrative actions of adenylyl cyclase and its dependence on neuromodulatory activity being of pivotal importance in providing regulatory information from central systems (see, for example, Davis, 2005; Hawkins *et al.*, 2006; Reymann & Frey, 2007).

Memory is not, therefore, organised as the simple, freely-running, autonomous process of transients written on top of equilibrium distributions that many models assume. This assumption essentially emerged out of the attempt to overcome the Hopfield model's (Hopfield, 1982) catastrophic forgetting dynamics by imposing bounds on synaptic strength, leading to palimpsest models in which newer memories are stored by overwriting older ones (Nadal *et al.*, 1986; Parisi, 1986). This approach leads directly to the notorious plasticity versus stability dilemma (Abraham & Robins, 2005). Existing non-integrative models of associative memory storage exhibit a monotonically falling memory signal, and the thrust of research has essentially been to delay this decline by dilating time via appeals to sparse coding and stochastically (that is, completely randomly) expressing only a small fraction of synaptic plasticity events. Nevertheless, the underlying problem remains: the memory signal falls monotonically as the initial memory signal transient is wiped out by ongoing synaptic plasticity, leading to a return, slowed down or otherwise, to equilibrium. The cascade model does attempt to include memory signal dynamics on multiple timescales, perhaps suggesting an attempt to take the multiple timescales of real synaptic plasticity seriously (Fusi *et al.*, 2005). However, the inclusion of multiple timescales in a single, autonomous process, ignoring the regulatory control of real biochemical cascades in real synapses, leads to disaster: the initial memory signal is dominated by its variance. The attempt to sever the

Gordian knot that ties the initial memory signal to memory lifetimes in these non-integrative models fails. A more recent, far more elaborate incarnation of the original cascade model (Benna & Fusi, 2015) does not appear fundamentally to address these issues. First, the memory signal continues to fall monotonically. Second, slow and fast variables continue to be used to transfer memories from more labile to less labile synaptic states (rather than from one memory system to another, as in biological systems), but these less labile states will inevitably continue to degrade the initial SNR. Third, there is no attempt to capture or represent the executive or regulatory control of memory storage and consolidation processes that is represent in real memory systems. Fourth, results are often presented for N in excess of one billion synapses, which is around four orders of magnitude too large for real neurons. Finally, the model appears to possess favourable memory signal dynamics only for these very large values of N, like the original cascade model, because these dynamics only set in when the memory signal is very small. Although this more recent incarnation appears to have several computational advantages over the original cascade model, it appears not to be possible to regard it as a model of the organisation of real memory systems in biological organisms (Eichenbaum & Cohen, 2001).

To some degree, integrative, filter-based models do sever the Gordian knot tying the initial SNR to memory lifetimes, but only partially. The rise in the initial memory signal is driven precisely by ongoing synaptic plasticity due to the storage of later memories. In other, non-integrative models, this ongoing storage *always* degrades the tracked memory. Nevertheless, even the peak memory signal in a filter-based model is constrained by the initial memory signal, and as the S filter shows, a strongly sustained memory signal is necessarily associated with an overall weak memory signal, even if it initially rises. Fundamentally, while the dynamics of the transient are radically different, filter-based synapses still store memories by inducing a transient on an equilibrium distribution in the absence of any executive control and without considering multiple timescales for plasticity or memory phenomena. There are hints that multiple timescales may emerge naturally in a filter-based approach (Elliott & Lagogiannis, 2012), perhaps providing a partial account of the spacing effect in the transition to late-phase plasticity (Carew *et al.*, 1972; Tully *et al.*, 1994; Beck *et al.*, 2000; Sutton *et al.*, 2002), but neuromodulatory processes are a critical component in this effect (Huang & Kandel, 1995; Sajikumar & Frey, 2004). Even with the repeated presentation of the same memory, whether spaced apart in time or massed together, in the absence of any other mechanisms the memory trace would still relax back to equilibrium in a filter-based model.

Of course, we have not sought to integrate multiple timescales into a single synaptic process and we have not explicitly considered the transition from early-phase to late-phase synaptic plasticity that would sustain the memory trace essentially indefinitely. Intriguingly, however, the comparative study of the different filter-based models considered here suggests that perhaps the multiple timescales associated with synaptic plasticity may arise due to different filter sizes [as we have suggested before (Elliott & Lagogiannis, 2012)], but perhaps more radically, due to entirely different types of synaptic filter. It appears that we must not seek to integrate multiple timescales into a single, unified synaptic process, and indeed the experimental evidence clearly indicates that multiple biochemical pathways operate in parallel. Thus, it seems to be necessary to consider multiple filter-like processes operating in concert at single synapses. We could therefore imagine an A0- or R0-like filter operating at a single synapse to provide a strong initial signal that is sustained during early-phase plasticity. In parallel, an S-like filter could also operate, providing an initially much weaker but growing and more strongly sustained, longer-lived signal that perhaps underlies the transition to late-phase plasticity. Of course, such a signal is still merely a transient in the absence of any other mechanism. But perhaps the strongly tetanising stimulus that characterises late-phase rather than early-phase plasticity specifically targets an S-like filter rather than an A0- or R0-like filter, and the expression of synaptic plasticity based on S-like filter threshold processes leads to locked states of synaptic strength. Whether the interactions between multiple filters operating in parallel as single synapses could lead, for example, to a full understanding of the spacing effect is unclear, but it will be fascinating to pursue these various ideas in later work.

References

Abraham, W., & Robins, A. (2005). Memory retention - the synaptic stability versus plasticity dilemma. *Trends Neurosci.*, **28**, 73–78.

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

Andersen, P., Morris, R., Amaral, D., Bliss, T., & O'Keefe, J., editors (2007). *The Hippocampus Book.* Oxford University Press, Oxford.

Asrican, B., Lisman, J., & Otmakhov, N. (2007). Synaptic strength of individual spines correlates with bound Ca^{2+}-calmodulin-dependent kinase II. *J. Neurosci.*, **27**, 14007–14011.

Bagal, A., Kao, J., Tang, C.-M., & Thompson, S. (2005). Long-term potentiation of exogenous glutamate responses at single dendritic spines. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 14434–14439.

Barrett, A., Billings, G., Morris, R., & van Rossum, M. (2009). State based model of long-term potentiation and synaptic tagging and capture. *PLoS Comput. Biol.*, **5**, e1000259.

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

Beck, C., Schroeder, B., & Davis, R. (2000). Learning performance of normal and mutant *Drosophila* after repeated conditioning trials with discrete stimuli. *J. Neurosci.*, **20**, 2944–2953.

Benna, M.K., & Fusi, S. (2015). Computational principles of biological memory. *arXiv*:1507.07580.

Carew, T., Pinsker, H., & Kandel, E. (1972). Long-term habituation of a defensive withdrawal reflex in Aplysia. *Science*, **175**, 451–454.

Clopath, C., Ziegler, L., Vasilaki, E., Büsing, L., & Gerstner, W. (2008). Tag-trigger-consolidation: A model of early and late long-term-potentiation and depression. *PLoS Comput. Biol.*, **4**, e1000258.

Cox, D., & Miller, H. (1965). *The Theory of Stochastic Processes.* Chapman & Hall/CRC, London.

Cox, D.R. (1962). *Renewal Theory.* Methuen, London.

Davis, R. (2005). Olfactory memory formation in *Drosophila*: From molecular to systems neuroscience. *Annu. Rev. Neurosci.*, **28**, 275–302.

Eichenbaum, H., & Cohen, N. (2001). *From Conditioning to Conscious Recollection.* Oxford University Press, Oxford.

Elliott, T. (2008). Temporal dynamics of rate-based plasticity rules in a stochastic model of spike-timing-dependent plasticity. *Neural Comput.*, **20**, 2253–2307.

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

Elliott, T. (2011b). Stability against fluctuations: Scaling, bifurcations and spontaneous symmetry breaking in stochastic models of synaptic plasticity. *Neural Comput.*, **23**, 674–734.

Elliott, T. (2014). Memory nearly on a spring: A mean first passage time approach to memory lifetimes. *Neural Comput.*, **26**, 1873–1923.

Elliott, T., & Lagogiannis, K. (2009). Taming fluctuations in a stochastic model of spike-timing-dependent plasticity. *Neural Comput.*, **21**, 3363–3407.

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

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

Fusi, S., Drew, P., and Abbott, L. (2005). Cascade models of synaptically stored memories. *Neuron*, **45**, 599–611.

Harris, K., & Stevens, J. (1989). Dendritic spines of CA1 pyramidal cells in the rat hippocampus: serial electron microscopy with reference to their biophysical characteristics. *J. Neurosci.*, **9**, 2982–2887.

Hawkins, R., Kandel, E., & Bailey, C. (2006). Molecular mechanisms of memory storage in *Aplysia*. *Biol. Bull.*, **210**, 174–191.

Heisenberg, M. (2003). Mushroom body memoir: From maps to models. *Nature Rev. Neurosci.*, **4**, 266–275.

Hopfield, 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.

Huang, Y.-Y., & Kandel, E. (1995). D1/D5 receptor agonists induce a protein synthesis-dependent late potentiation in the CA1 region of the hippocampus. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 2446–2450.

Lahiri, S., & Ganguli, S. (2013). A memory frontier for complex synapses. In: pp. 1034–1042 of *Advances in Neural Information Processing Systems 26*, Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., & Weinberger, K.Q. (eds.), MIT Press, Cambridge, MA.

Leibold, C., & Kempter, R. (2006). Memory capacity for sequences in a recurrent network with biological constraints. *Neural Comput.*, **18**, 904–941.

Leibold, C., & Kempter, R. (2008). Sparseness constrains the prolongation of memory lifetime via synaptic metaplasticity. *Cerebral Cortex*, **18**, 67–77.

Miller, P., Zhabotinsky, A., Lisman, J., & Wang, X.-J. (2005). The stability of a stochastic CaMKII switch: Dependence on the number of enzyme molecules and protein turnover. *PLoS Biology*, **3**, 0705.

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

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

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

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

Nusser, Z., Lujan, R., Laube, G., Roberts, J., Molnar, E., & Somogyi, P. (1998). Cell type and pathway dependence of synaptic AMPA receptor number and variability in the hippocampus. *Neuron*, **21**, 545–559.

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

Päpper, M., Kempter, R., & Leibold, C. (2011). Synaptic tagging, evaluation of memories, and the distal reward problem. *Learn. and Mem.*, **18**, 58–70.

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

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

Reymann, K., & Frey, J. (2007). The late maintenance of hippocampal LTP: Requirements, phases, 'synaptic tagging', 'late-associativity' and implications. *Neuropharm.*, **52**, 24–40.

Roberson, E., English, J., & Sweatt, J. (1996). A biochemist's view of long-term potentiation. *Learn. and Mem.*, **3**, 1–24.

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

Sajikumar, S., & Frey, J. (2004). Late-associativity, synaptic tagging, and the role of dopamine during LTP and LTD. *Neurobiol. Learn. and Mem.*, **82**, 12–25.

Sutton, M., Ide, J., Masters, S., & Carew, T. (2002). Interaction between amount and pattern of training in the induction of intermediate- and long-term memory for sensitization in Aplysia. *Learn. and Mem.*, **9**, 29–40.

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

Tully, T., Preat, T., Boynton, S., & Del Vecchio, M. (1994). Genetic dissection of consolidated memory in *Drosophila*. *Cell*, **79**, 35–47.

Willshaw, D., Duneman, O., & Longuet-Higgins, H. (1969). Nonholographic associative memory. *Nature*, **222**, 960–962.

Figure Captions

Figure 1: The three principal (zero injection) synaptic filtering variants considered here. Synaptic filter states are represented by the circled numbers. Transitions between states labelled $r_+ \uparrow$ ($r_- \downarrow$) indicate potentiating (depressing) induction signals of Poisson rate r_+ (r_-) that lead to changes in the filter state as indicated. Transitions labelled $r_+ \Uparrow$ ($r_- \Downarrow$) indicate the expression of synaptic plasticity, leading to an increase (decrease) in synaptic strength if possible. For simplicity, we have depicted filters with $\Theta = 4$, so that synapses may exist in filter states $I = -3, \ldots, +3$, with the states $I = \pm 4$ being boundary states. (A) The standard, doubly absorbing filter, A0. Induction signals increment or decrement the filter state. A potentiating (depressing) induction signal in state $I = +\Theta - 1$ ($I = -\Theta + 1$) leads to the filter state being reset to $I = 0$ and a change in synaptic strength being expressed, if possible. (B) A filter with one reflecting boundary and one absorbing boundary, R0. Induction signals increment or decrement the filter state, as for the standard, absorbing filter. We have denoted the filter transitions in each strength state ($S = +1$ for strong and $S = -1$ for weak; each strength state denoted by a gray rectangle) for clarity, but there is nevertheless only a single synaptic filter, regardless of the strength state. When a synapse is strong (weak), a potentiating (depressing) induction signal in filter state $I = +\Theta - 1$ ($I = -\Theta + 1$) cannot lead to a change in synaptic strength for a binary-strength synapse. In these cases, instead of resetting the filter to $I = 0$, the synapse remains in its current filter state (i.e. the boundaries reflect the transition). However, when a synapse is weak (strong), a potentiating (depressing) induction signal in filter state $I = +\Theta - 1$ ($I = -\Theta + 1$) can lead to a change in synaptic strength. In these cases, the boundaries absorb the transition, the filter is reset to $I = 0$, and a change in strength is expressed. (C) A super or S filter is a doubly absorbing filter, but the transitions between filter states associated with induction signals differ from those of the standard, A0 absorbing filter. A potentiating (depressing) induction signal in non-negative (non-positive) filter states leads to an increment (decrement) in filter state, as usual. However, a potentiating (depressing) induction signal in negative (positive) filter states immediately resets the filter to the state $I = 0$.

Figure 2: Equilibrium distributions of synaptic filter states, for $\Theta = 8$. The upper two figures show the distributions for the standard, doubly absorbing filters A0 and Ar; the middle two figures show them for the reflecting boundary filters R0 and Rr; the lower two show them for the super or S filter. Distributions for both strength states are shown, as indicated, but for the A0, Ar and S filters, the distributions are independent of synaptic strength. Solid lines show the distributions for the zero injection filters (A0 and R0), while dashed lines show them for the random injection filters (Ar and Rr). We have not shown the distributions for the S filter with random injections because they are virtually indistinguishable from the zero injection version.

Figure 3: Mean filter escape times $\boldsymbol{\Delta} \cdot \boldsymbol{\tau}$ or $\boldsymbol{m} \cdot \boldsymbol{\tau}$ as a function of filter size, Θ. For the A0, R0 and S filters, we show $\boldsymbol{\Delta} \cdot \boldsymbol{\tau} = \tau_0$, the mean escape time starting from the zero filter state, into which these three filters are injected upon threshold. For the Ar and Rr filters, we show $\boldsymbol{m} \cdot \boldsymbol{\tau}$, the mean escape time averaged over all filter states, since these two filters are injected randomly (uniformly) into any filter state upon threshold.

Figure 4: Mean memory signal as a function of time for the various models, as indicated, for different of choices of Θ (filter size) or n (cascade size), as indicated in each separate figure. In the legends, "A0" ("Ar") denotes an absorbing filter with zero (random) injections; "R0" a reflecting filter with zero (random) injections; "S" a super filter (only with zero injections); "C" the cascade model. We do not plot $\mu(t)$ below 10^{-3} because such small values would require (roughly) $N > 10^6$ synapses for a corresponding SNR in excess of unity. If (E) and (F), lines for the S filter are absent because its memory signal does not exceed this limit. In (D), the line for the S filter is only just visible above the rt-axis.

Figure 5: Covariance between pairs of synapses' strengths and the SNR $\mu(t)/\sigma(t)$ for the tracked memory as a function of time for the various models, as indicated, for different choices of Θ (filter size) or n (cascade size), as indicated in each separate figure. Legends are as in Fig. 4. (A), (C) and (E) show the covariance for Θ (or n) = 4, 7 and 10, respectively. We have not shown

results for the random injection variants for clarity, because they are extremely similar to the zero injection versions. (B), (D) and (F) show the corresponding SNRs for the same parameter choices in (A), (C) and (E), respectively. These SNRs are generated for $N = 10^4$ synapses. In (F), the SNR for the S filter remains below unity. A larger value of N would be required to bring its SNR above unity.

Figure 6: The Θ–N (or n–N) plane, indicating which model has the largest SNR memory lifetime. In each enclosed and labelled region (same conventions as Fig. 4), the indicated model has the largest SNR memory lifetime. In the blacked-out region, all models' SNRs do not exceed unity, so SNR memory lifetimes do not exist in this region. In almost all regions of biologically-relevant parameter space, synaptic filters out-perform the cascade model. The cascade out-performs filters only for smaller N, but only because filters do not successfully encode memories for such small N, at least according to an SNR criterion.

Figure 7: Cross-sections through the Θ–N (or n–N) plane in Fig. 6 for fixed Θ (or n), as indicated in each separate figure, explicitly plotting SNR memory lifetimes $r\tau_{\text{snr}}$ against N, for the various models. Legends are as in Fig. 4. SNR memory lifetimes for the reflecting boundary filter with either zero or random injections (R0 and Rr filters) are nearly indistinguishable, except for smaller N. Although the cascade model out-performs filter models for very small N for Θ (or n) greater than 7, we see that in fact its SNR memory lifetimes in

this parameter region are minuscule. No model performs well in this small N region, at least according to an SNR criterion.

Figure 8: Mean first passage time memory lifetimes for Θ or n equal to 10. Legends are as in Fig. 4. (A) MFPT lifetimes for the various models, showing non-zero lifetimes for small N. (B) A direct comparison of SNR and MFPT memory lifetimes for three models. (C) MFPT memory lifetimes and the one standard deviation regions around them. We have displaced the cascade model's error bars slightly rightwards so that they can be clearly distinguished from the R0 filter model's error bars. In these figures, data are obtained in simulations by averaging over a total of $10^8/N$ simulations. For larger N there is more self-averaging so we need not average over as many simulations in order to obtain good statistics.

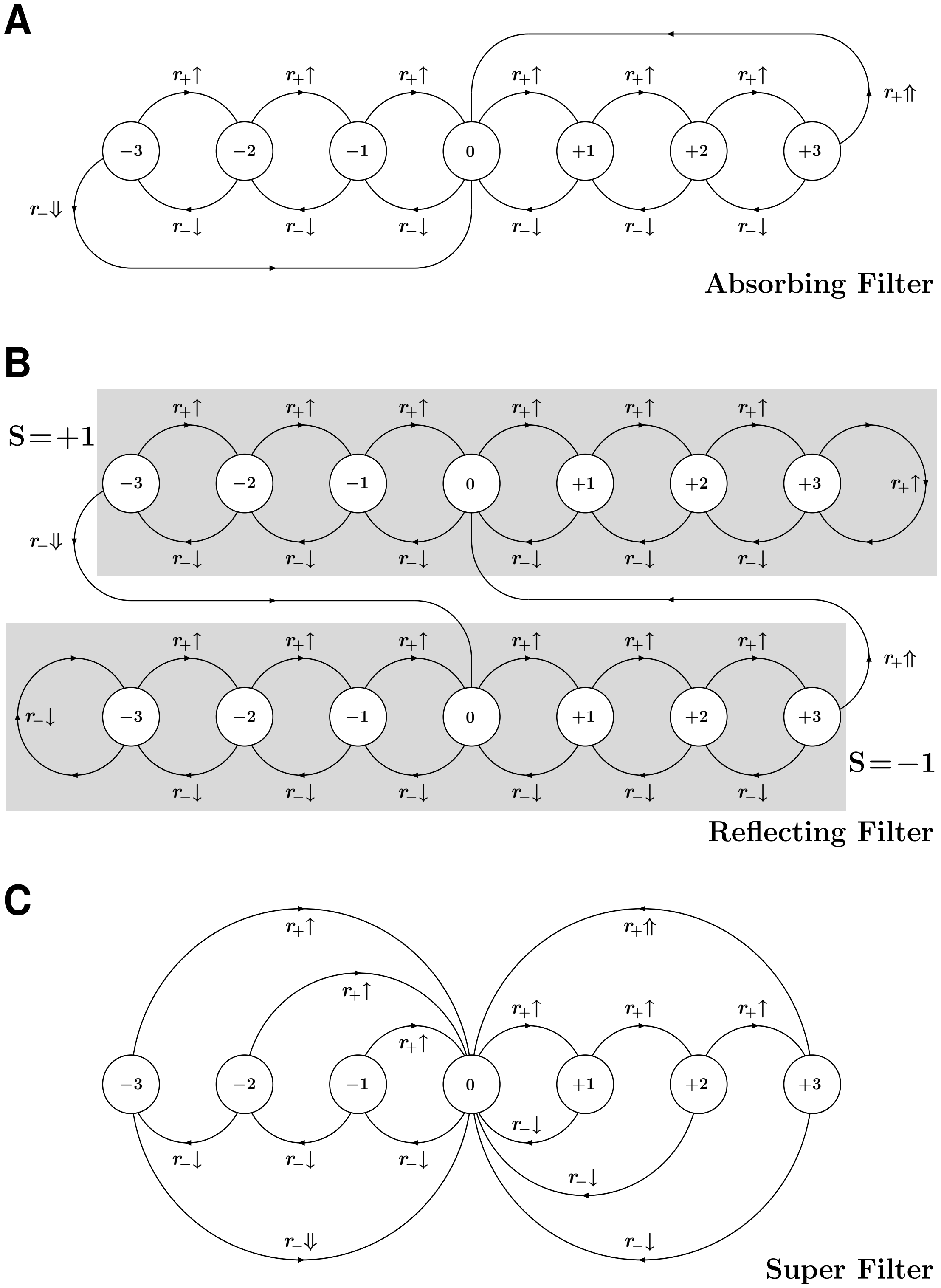

A
-3 -2 -1 0 $+1$ $+2$ $+3$
$r_+\upharpoonleft$
$r_-\downharpoonleft$
$r_+\Uparrow$
$r_-\Downarrow$
Absorbing Filter
B
S=+1
S=−1
-3 -2 -1 0 $+1$ $+2$ $+3$
$r_+\upharpoonleft$
$r_-\downharpoonleft$
$r_+\Uparrow$
$r_-\Downarrow$
Reflecting Filter
C
-3 -2 -1 0 $+1$ $+2$ $+3$
$r_+\upharpoonleft$
$r_-\downharpoonleft$
$r_+\Uparrow$
$r_-\Downarrow$
Super Filter

Figure 1

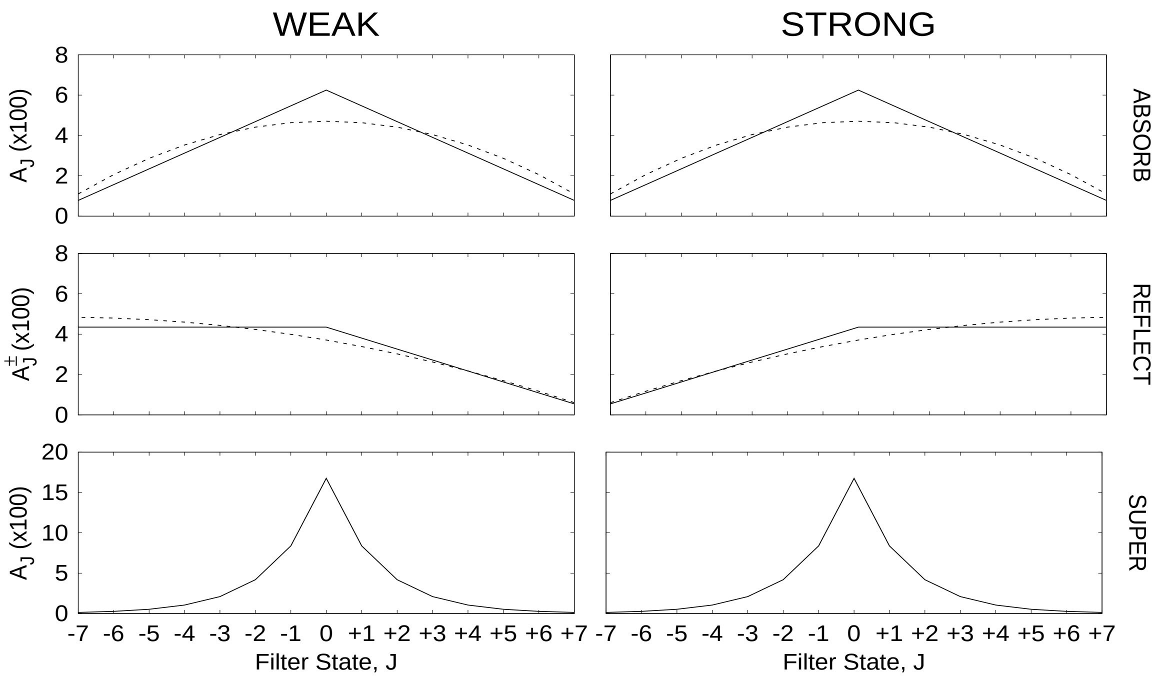
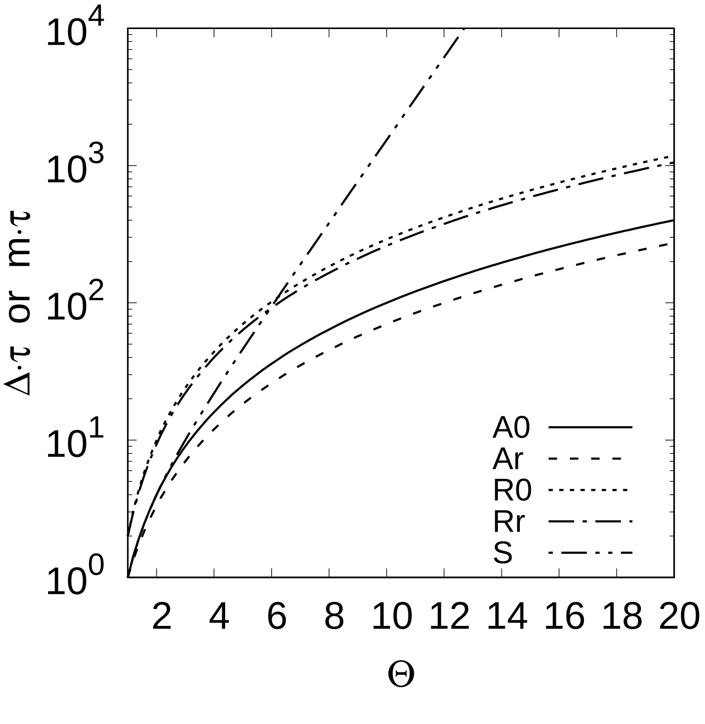

WEAK
STRONG
A_J (x100)
$A_J^{\pm}$ (x100)
A_J (x100)
ABSORB
REFLECT
SUPER
0
2
4
6
8
0
5
10
15
20
-7 -6 -5 -4 -3 -2 -1 0 +1 +2 +3 +4 +5 +6 +7
Filter State, J

Figure 2



10^4
10^3
10^2
10^1
10^0
$\Delta\cdot\tau$ or $m\cdot\tau$
2
4
6
8
10
12
14
16
18
20
Θ
A0
Ar
R0
Rr
S

Figure 3

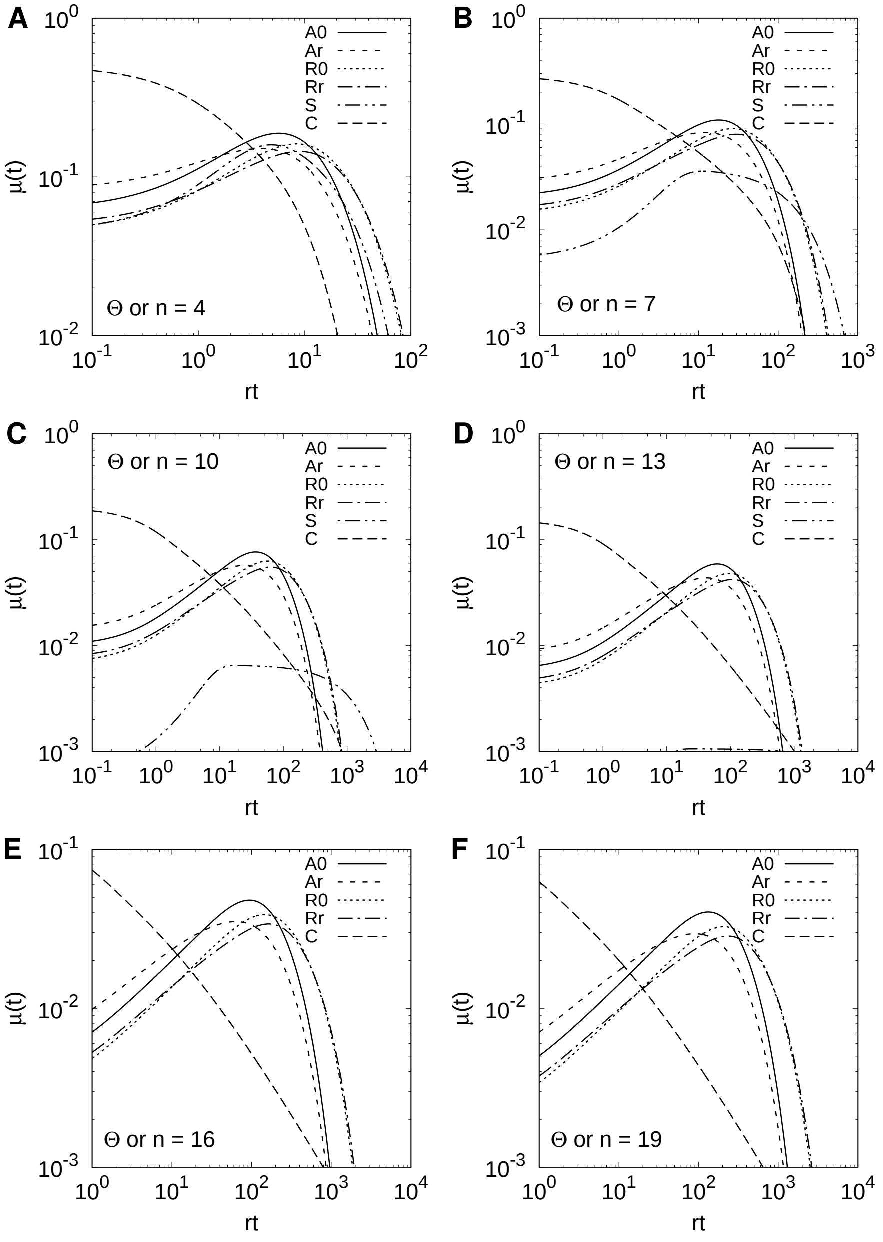

A
B
C
D
E
F
μ(t)
rt
A0
Ar
R0
Rr
S
C
Θ or n = 4
Θ or n = 7
Θ or n = 10
Θ or n = 13
Θ or n = 16
Θ or n = 19
10^0
10^-1
10^-2
10^-3
10^1
10^2
10^3
10^4

Figure 4

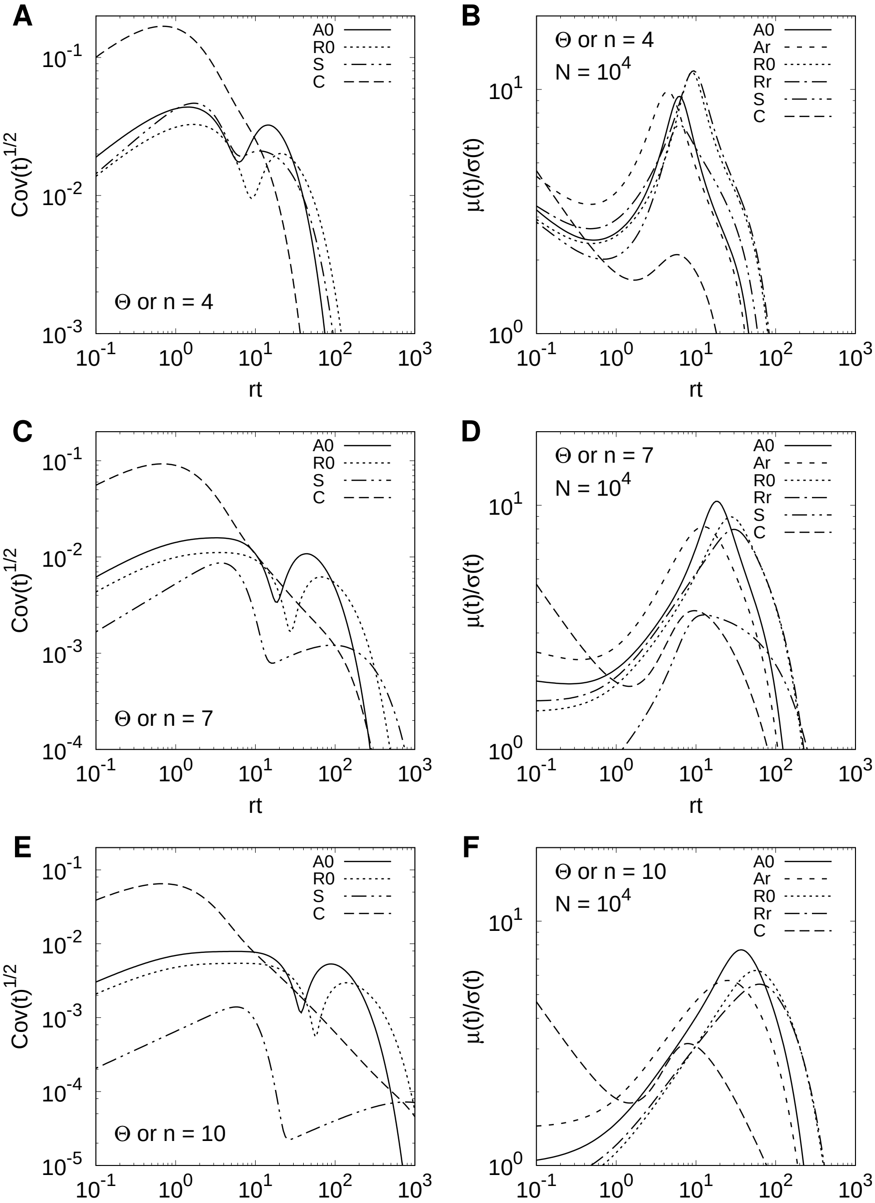

A
A0
R0
S
C
Cov(t)$^{1/2}$
10^{-1}
10^{-2}
10^{-3}
Θ or n = 4
10^{-1} 10^{0} 10^{1} 10^{2} 10^{3}
rt
B
Θ or n = 4
N = 10^4
A0
Ar
R0
Rr
S
C
μ(t)/σ(t)
10^{1}
10^{0}
rt
C
A0
R0
S
C
Cov(t)$^{1/2}$
10^{-1}
10^{-2}
10^{-3}
10^{-4}
Θ or n = 7
rt
D
Θ or n = 7
N = 10^4
A0
Ar
R0
Rr
S
C
μ(t)/σ(t)
rt
E
A0
R0
S
C
Cov(t)$^{1/2}$
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
Θ or n = 10
rt
F
Θ or n = 10
N = 10^4
A0
Ar
R0
Rr
C
μ(t)/σ(t)
rt

Figure 5

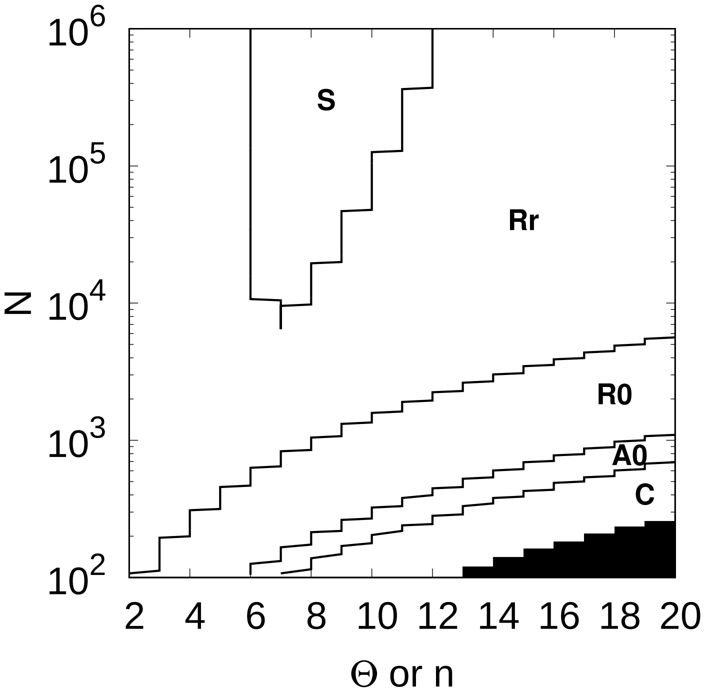


Figure 6

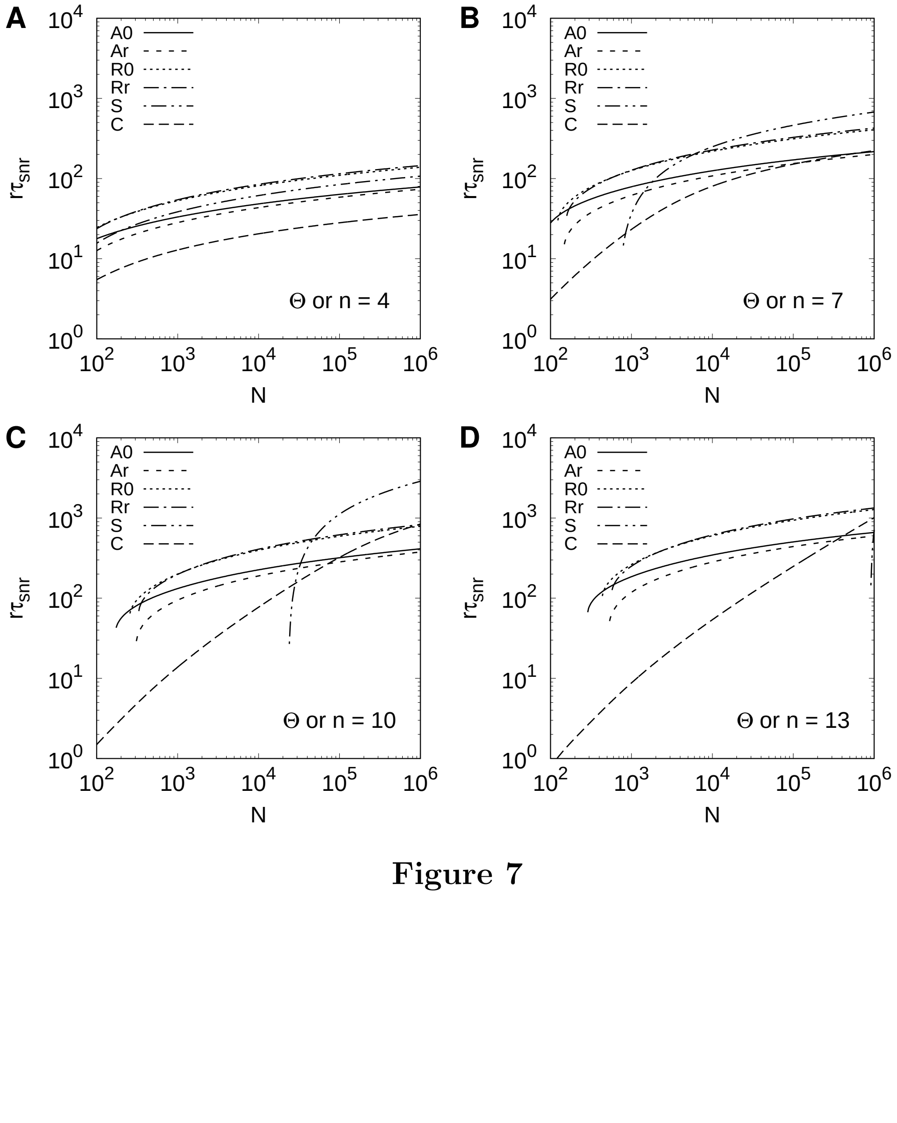

A
B
C
D
A0
Ar
R0
Rr
S
C
$r\tau_{snr}$
N
10^0
10^1
10^2
10^3
10^4
10^5
10^6
Θ or n = 4
Θ or n = 7
Θ or n = 10
Θ or n = 13

Figure 7

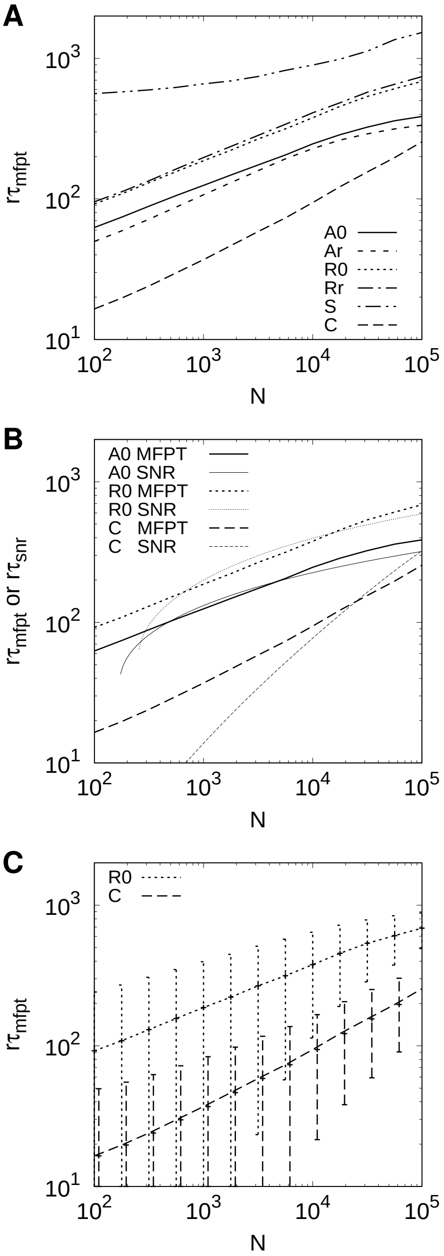

A
$r\tau_{mfpt}$
N
A0
Ar
R0
Rr
S
C
B
$r\tau_{mfpt}$ or $r\tau_{snr}$
A0 MFPT
A0 SNR
R0 MFPT
R0 SNR
C MFPT
C SNR
C
R0
C
10^1
10^2
10^3
10^4
10^5

Figure 8