A Biologically Plausible Spiking Neural Model of Eyeblink Conditioning in the Cerebellum

Andreas Stöckel (astoecke@uwaterloo.ca)[1]
Terrence C. Stewart (terrence.stewart@nrc-cnrc.gc.ca)[2]
Chris Eliasmith (celiasmith@uwaterloo.ca)[1]
[1]Centre for Theoretical Neuroscience, University of Waterloo
[2]National Research Council of Canada, University of Waterloo Collaboration Centre
200 University Avenue West, Waterloo, ON N2L 3G1 Canada

Abstract

The cerebellum is classically described in terms of its role in motor control. Recent evidence suggests that the cerebellum supports a wide variety of functions, including timing-related cognitive tasks and perceptual prediction. Correspondingly, deciphering cerebellar function may be important to advance our understanding of cognitive processes. In this paper, we build a model of eyeblink conditioning, an extensively studied low-level function of the cerebellum. Building such a model is of particular interest, since, as of now, it remains unclear how exactly the cerebellum manages to learn and reproduce the precise timings observed in eyeblink conditioning that are potentially exploited by cognitive processes as well. We employ recent advances in large-scale neural network modeling to build a biologically plausible spiking neural network based on the cerebellar microcircuitry. We compare our simulation results to neurophysiological data and demonstrate how the recurrent Granule-Golgi subnetwork could generate the dynamics representations required for triggering motor trajectories in the Purkinje cell layer. Our model is capable of reproducing key properties of eyeblink conditioning, while generating neurophysiological data that could be experimentally verified.

Keywords: cerebellum; classical conditioning; biologically plausible spiking neural network; Neural Engineering Framework; delay network

Introduction

Traditional clinical evidence portrays the cerebellum as supporting the coordination of precise movements. In particular, the cerebellum receives motor commands from cerebral cortex that are combined with proprioceptive feedback, resulting in a corrective signal that is relayed back to the motor pathways in the brainstem and cerebral cortex (Splittgerber, 2018).

Recent evidence – ranging from studies in functional connectivity, neuronal tracing, clinical pathology, to evolutionary physiology – suggests that the tasks supported by the cerebellum are not restricted to motor control alone. The cerebellum may instead be recruited by various brain regions as a "co-processor" to support brain functions related to higher-order cognition, such as language-based thinking, working memory, perceptual prediction, and tasks requiring precise timings in general (Sullivan, 2010; Buckner, 2013; O'Reilly et al., 2008; E et al., 2014). Understanding cerebellar function may thus be crucial for gaining a better understanding of human cognition.

Fortunately, the cerebellar microcircuitry is extremely well-studied (Ito, 2010; Llinás, 2010). It is found to be highly regular, while primarily performing feed-forward processing. The main feed-forward pathway is depicted in Figure 1: afferent nerves from pre-cerebellar nuclei project as "mossy fibres" onto granule cells in the cerebellar cortex. Cerebellar granule cells account for the majority of neurons in the mammalian brain – thus, the divergence of PC neurons onto this sheer number of granule cells results in a large repertoire of stimulus representations. Granule cell axons, the so called "parallel fibres," project onto the Purkinje cells, which inhibit neurons in the cerebellar nucleus, that in turn project back onto the brainstem and cerebral cortex. In addition to this primary pathway, research has focused on recurrent connections between the Golgi and granule cells, as well as the "climbing fibres" originating from the Inferior Olive, which enable synaptic weight modulation in the Granule-Prurkinje projection.

This circuitry has been interpreted by Marr (1969) as a (in modern terms) "supervised learning machine" that relies on the divergence of mossy fibres onto the granular layer for pattern extraction, and olivary-modulated plasticity of the granule-Purkinje projection for learning. The concepts explored in this seminal work form the foundation of most models of cerebellar function (Strata, 2009; Raymond & Medina, 2018).

A class of experiments that explores the role of the cerebellum in motor learning in its most simple form is eyeblink conditioning. A puff of air is directed at the eye (unconditioned stimulus; US), triggering the eyeblink reflex (unconditioned response; UR). The US is paired with a neutral, conditioned stimulus (CS), such as a tone or a flashing light, preceding the US by a constant time offset Δt. The subject learns an association between the UR and the CS and will, over time, form a conditioned response (CR) to the previously neutral CS. Experiments indicate that the formation of the CR critically depends on the cerebellum; previously learned CRs are absent once the cerebellum is ablated (e.g., McCormick et al., 1981).

Central for the purposes of this paper is the observation that the conditioned stimulus-response pair maintains timing. That is, the CR is triggered a period Δt *after* the CS, preserving the original time offset between the conditioned and unconditioned stimulus. Despite detailed knowledge about cerebellar microcircuitry, it is still unclear how the cerebellum learns and replays these relatively precise timings.

In this paper, we construct a biologically plausible spiking neural network model capable of learning timings. Our model is based on the so called "delay network," which we map onto the recurrent Granule-Golgi subnetwork. We first discuss related work, continue with a discussion of our model, and finally compare our simulation results to empirical data.

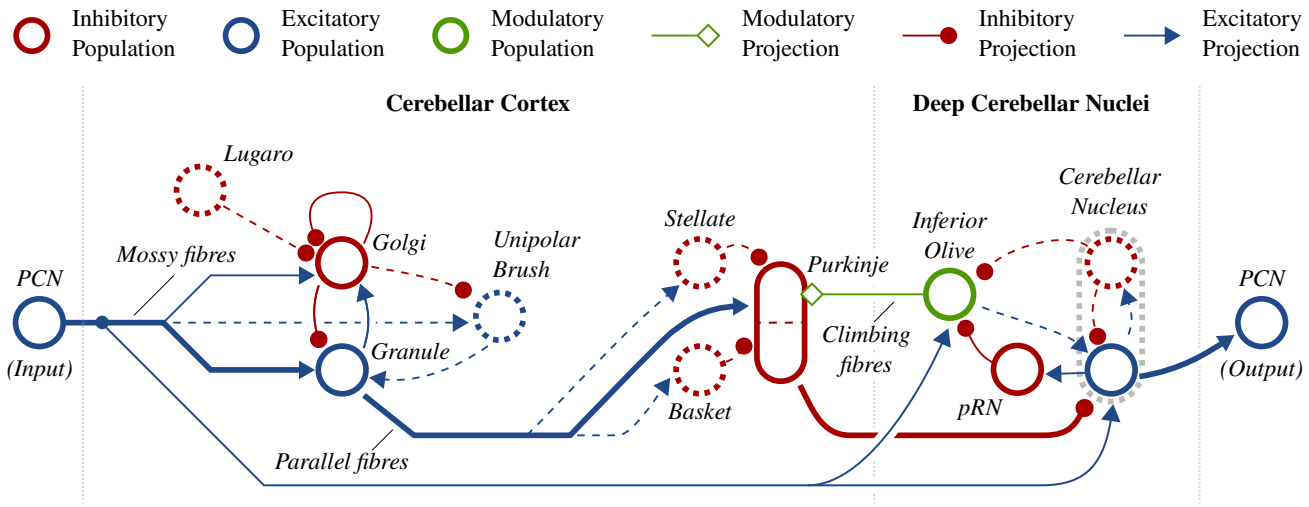

Figure 1: Schematic of the cerebellar microcircuit. Dashed projections and populations are not included in our model. Cerebellar nucleus afferents affect both the excitatory and inhibitory sub-populations. The main feed-forward pathway is highlighted in bold. *PCN* ≙ pre-cerebellar neurons/nuclei. *pRN* ≙ parvicellular Red Nucleus. Data from Ito, 2010; Llinás, 2010.

Related Work

There are two predominant types of computational models of cerebellar timing: models assuming that afferent signals are transformed by the cerebellar microcircuit into a "dynamics representation" (Medina & Mauk, 2000), and models assuming that the reproduction of timings is performed by mechanisms intrinsic to individual Purkinje cells (Lusk et al., 2016). While there is little neurophysiological data which conclusively confirms or rules out any of these hypotheses, we focus on the first idea, dynamics representations.

The idea of dynamics representation is that short bursts of mossy fibre inputs – in particular, the signals corresponding to the CS – are turned into a diverse set of prolonged activities in the granular layer. Choosing the right synaptic weights between the granule and Purkinje cells approximates any desired function over the recent history of the stimulus, for example a function mapping from the CS onto the UR.

Classic hypotheses for dynamics representations include delay lines, in which individual neurons represent a delayed version of the input signal, spectral timing, in which the neural activity follows a set of bell-shaped functions with different offsets and widths, and, finally, neural activities corresponding to oscillations with diverse phases and frequencies. While successful in top-down models, the mechanisms that might underlie these responses are unclear (Medina & Mauk, 2000).

More recent spiking models of cerebellar function (Rössert et al., 2015) exploit the recurrent connection between the Golgi and granule cells as a means to generate pseudo-random dynamics, akin to neural reservoirs. In particular, rather than postulating an intra-neural mechanism that creates a temporal representation, this approach demonstrates that a temporal representation can be created purely through the synaptic connections between neurons.

Instead of relying on random dynamics, we present a systematic approach that approximates a mathematically optimal dynamics representation in the spiking Granule-Golgi network. By incorporating neurophysiological constraints into our model we are able to qualitatively match experimental data on eyeblink conditioning experiments in head-fixed mice (Heiney et al., 2014) while only tuning four parameters.

Representing Time

As discussed above, our model builds on the existing hypothesis that the granule cells in the cerebellum create a representation of the recent history of their inputs; it should be possible to reconstruct the granule cell input from the recent past, given their current activity.

We describe three different versions of the model, at different levels of abstraction. First, we give an abstract mathematical model where the "neurons" are ideal (non-spiking) integrators. Next, we replace these "neurons" with recurrently connected leaky integrate-and-fire (LIF) spiking neurons. Finally, we separate these neurons into excitatory granule cells and inhibitory Golgi cells with realistic connectivity patterns.

Version 1: Ideal Mathematical Model

Voelker and Eliasmith (2018) derive an ideal "delay network" memory by taking the Padé approximate of the continuous-time delay $F(s) = e^{-\theta s}$. This results in a linear dynamical system of dimension q, where the state $\mathbf{m}$ forms a compressed representation of the past history of the input u over the previous θ seconds:

$$\begin{aligned} &\dot{\mathbf{m}} = \mathbf{A}\mathbf{m} + \mathbf{B}u \\ &\theta\mathbf{A} = a_{ij} \in \mathbb{R}^{q\times q}, \; a_{ij} = \begin{cases} (2i+1)(-1) & i < j, \\ (2i+1)(-1)^{i-j+1} & i \geq j, \end{cases} \\ &\theta\mathbf{B} = b_i \in \mathbb{R}^q, \qquad b_i = (2i+1)(-1)^i. \end{aligned} \tag{1}$$

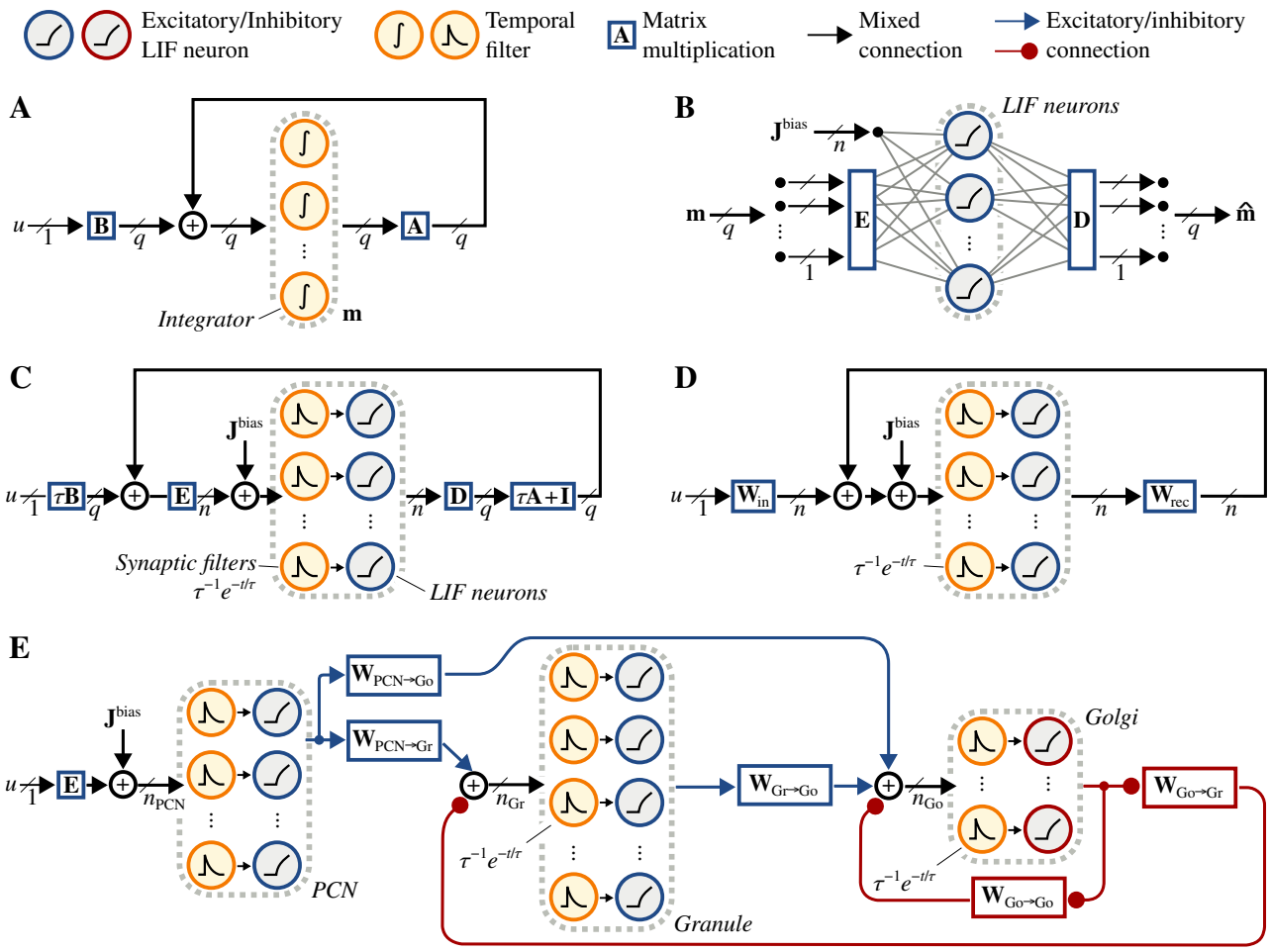

Figure 2: Three variants of the delay network. **(A)** Ideal mathematical implementation with integrators. **(B-D)** Constructing an all-to-all recurrent LIF model of granule cells that approximates the ideal behaviour. **(B)** Communication channel that uses a hidden layer of LIF neurons to represent $\mathbf{m}$. **(C)** Adding an exponential synaptic filter and connecting the output back to the input via the connection weight matrix $\tau\mathbf{A}+\mathbf{I}$. This approximates the desired mathematical function (see text). **(D)** Combining the weight matrices from **(C)** into a single recurrent weight matrix $\mathbf{W}_{\text{rec}}$ and input weight matrix $\mathbf{W}_{\text{in}}$. **(E)** Biologically plausible implementation separating the recurrent connection into an excitatory granular and an inhibitory Golgi cell layer.

Importantly, the state at a previous point in time $t-\theta'$ (where $0 \leq \theta' \leq \theta$) can be approximated by computing the dot product between $\mathbf{m}$ and a decoding vector $\mathbf{d}(\theta')$:

$$u(t-\theta') = \sum_{\ell=0}^{q-1} m_\ell d_\ell(\theta'), \qquad \text{where } d_\ell = \tilde{P}_\ell\left(\frac{\theta'}{\theta}\right), \qquad (2)$$

and $\tilde{P}_\ell$ is the shifted Legendre polynomial of degree ℓ.

This suggests the abstract model for the granule cells depicted in Figure 2A. If there was a separate cell implementing a perfect integrator for each element in $\mathbf{m}$ with the corresponding recurrent weights $\mathbf{A}$ and input weights $\mathbf{B}$, the resulting system should implement the delay network. The response of this system to a pulse input is shown in Figure 3A. Notice that the output is only an approximation of the input – in particular, the pulse is spread out into a "bump". The accuracy of the representation depends on q, the number of dimensions in $\mathbf{m}$.

Version 2: Recurrent All-to-All Network

Evidently, the assumptions made in the previous model are nowhere close to being biologically plausible. As a first step towards more biological realism, we replace the ideal integrators with recurrently connected spiking leaky integrate-and-fire (LIF) neurons with a simple synapse model while maintaining a close approximation of the mathematical model.

Since the granule cells are meant to represent $\mathbf{m}$, we start with a single-hidden-layer network with q inputs and q outputs (Figure 2B). The hidden layer neurons are LIF neurons with a membrane time constant of 20 ms and a refractory period of 2 ms, with each neuron i receiving a bias input J_i^{bias}.

Such a network can be trained to behave as a communication channel; that is, the input and output weights can be set such that output of the network $\hat{\mathbf{m}}$ will be an approximation of the input $\mathbf{m}$. We call the input weights $\mathbf{E}$ (for encoder) and the

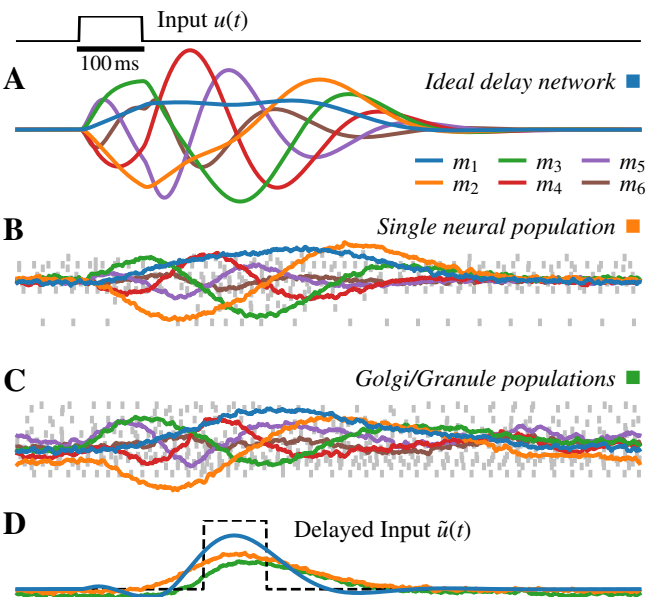

Figure 3: Temporal representation in response to a 100 ms pulse for the three versions of the delay network with $q = 6$, $\theta = 0.4$. **(A)** Ideal mathematical implementation. **(B)** Single population of 200 LIF neurons. Maximum firing rates between 50 and 100 Hz, $\tau = 60$ ms. The decoded value is filtered by a 100 ms low-pass filter. Gray dots correspond to spikes from 20 representative neurons. **(C)** Same as **(B)**, but for the Granule-Golgi circuit with 200 granule and 20 Golgi cells. **(D)** Decoding a 200 ms delay using eq. (2).

output weights $\mathbf{D}$ (for decoder). While there are many methods for generating such weights, here we randomly generate $\mathbf{E}$ and $\mathbf{J}^{\text{bias}}$ such that the neurons exhibit diverse tuning properties with maximum firing rates between 50 and 100 Hz. We then use least-squares minimization over the space of possible $\mathbf{m}$ to solve for weights $\mathbf{D}$ that minimize the average error $\|\mathbf{m} - \hat{\mathbf{m}}\|^2$.

Given this standard feed-forward network, we can construct a recurrent network that will approximate eq. (1). As depicted in Figure 2C, we start by adding a synapse model before every LIF neuron. This synapse model is the standard exponential post-synaptic current model $\tau^{-1}e^{-t/\tau}$, where τ is the post-synaptic current decay time-constant (Roth & van Rossum, 2009). Additionally, we multiply the input by $\tau\mathbf{B}$, the output $\hat{\mathbf{m}}$ by $\tau\mathbf{A} + \mathbf{I}$, and connect this result back to the input $\mathbf{m}$.

To analyze the behaviour of the resulting system, we first note that the current-based synapse model is a linear operator with unit DC-gain, so even though it is being applied at the input to each neuron, it has the same effect as if it were applied before the application of the $\mathbf{E}$ matrix and the bias $\mathbf{J}^{\text{bias}}$. If the original feed-forward network is a good communication channel, i.e., $\hat{\mathbf{m}} \approx \mathbf{m}$, then the addition of the synapse means that the network will now produce an output that is approximately $\mathbf{m}(t) * \tau^{-1}e^{-t/\tau}$, where "$*$" is the convolution operator.

The overall system with its recurrent connection will thus have the following dynamics:

$$\mathbf{m}(t) = (u(t)\tau B + \mathbf{m}(t)(\tau\mathbf{A} + \mathbf{I})) * \tau^{-1}e^{-t/\tau}.$$

Converting into the Laplace domain and noting that the Laplace transform of $\tau^{-1}e^{-t/\tau}$ is $1/(1+s\tau)$, we get

$$\begin{aligned} \mathbf{M}(s) &= \big(U(s)\tau\mathbf{B} + \mathbf{M}(s)\big)\frac{\tau\mathbf{A} + \mathbf{I}}{1+s\tau} \\ \Leftrightarrow \mathbf{M}(s)(1+s\tau) &= U(s)\tau\mathbf{B} + \mathbf{M}(s)\tau\mathbf{A} + \mathbf{M}(s) \\ \Leftrightarrow \mathbf{M}(s)s &= U(s)\mathbf{B} + \mathbf{M}(s)\mathbf{A}. \end{aligned} \tag{3}$$

Converting back to the time domain yields $\dot{\mathbf{m}} = \mathbf{A}\mathbf{m} + \mathbf{B}u$, the linear dynamical system from eq. (1). Figure 3B illustrates that our model indeed approximates the ideal dynamics.

Since there are no non-linearities other than the LIF neurons, the various connection weight matrices can be multiplied together to form matrices $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{rec}}$ that will produce the identical behaviour (Figure 2D).

This general method for converting any dynamical system into a recurrent neural network is the basis of the Neural Engineering Framework (NEF; Eliasmith & Anderson, 2003), and has been used for modelling cognitive phenomena such as working memory (Singh & Eliasmith, 2006) and in the creation of Spaun, the largest existing functional brain model (Eliasmith et al., 2012).

Version 3: Golgi and Granule Cells

While the previous version of our model is certainly an improvement when it comes to biological plausibility, it still ignores fundamental features of cerebellar physiology.

First of all, granule cells have no lateral connections. That is, there is no correlate of the reccurent weight matrix $\mathbf{W}_{\text{rec}}$ in nature. Instead, recurrent connections are mediated via inhibitory interneurons, the Golgi cells (Ito, 2010). Second, the previous version does not distinguish between excitatory and inhibitory neurons; $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{rec}}$ contain both positive and negative weights. In biology, granule cells exclusively evoke excitatory currents and Golgi cells exclusively evoke inhibitory currents. This is a specific instance of Dale's principle (Strata & Harvey, 1999). Finally, the previous version assumes that each neuron i receives a constant bias current J_i^{bias}, which has no clear biological correlate.

The last two issues have been addressed recently by Stöckel and Eliasmith (2019). In short, one can still define the neural tuning in terms of encoders $\mathbf{E}$ and bias currents $\mathbf{J}^{\text{bias}}$. However, these are merely normative, defining the expected behaviour of individual cells when representing a value $\mathbf{m}$. Instead of solving for decoders $\mathbf{D}$ minimizing the error $\|\mathbf{m} - \hat{\mathbf{m}}\|^2$, one can directly solve for full weights matrices $\mathbf{W}$ in "current space", that have the same effect as applying a bias current and an encoding vector. To see this, consider that each neuron i in the post-population is supposed to receive an input current

$$J_i(\mathbf{m}) = \mathbf{m} \cdot \mathbf{e}_i + J_i^{\text{bias}}. \tag{4}$$

In addition, we know that the input current of a neuron within a spiking neural network is approximately given as

$$J_i(\mathbf{a}_{\text{pre}}) = \mathbf{a}_{\text{pre}} \cdot \mathbf{w}_i \,, \tag{5}$$

where $\mathbf{a}_{\text{pre}}$ is a vector combining all pre-population activities.

Combining eqs. (4) and (5) results in a new optimization problem that allows us to solve for synaptic weights $\mathbf{w}_i$ for each post neuron i. These equations can be extended to approximate arbitrary functions $f(\mathbf{m})$ of values $\mathbf{m}$ represented by a pre-population. Adding a non-negativity constraint further allows us to account for Dale's principle, resulting in a non-negative least squares problem.

While we now know how to build networks that do not rely on bias currents and mixed-sign weight matrices, we still need to imprint the dynamics from eq. (1) onto a network that spans multiple neural ensembles. In particular, we have to account for two sets of synaptic filters, as well as the fact that both the granule and Golgi cells receive mossy fibre input (fig. 1). Assuming that both sets of synapses have the same time-constant, one can show similarly to the proof in eq. (3) that choosing the same transformations as above yields the desired dynamics (proof omitted due to space constraints). That is, the input weight matrices $\mathbf{W}_{\text{PCN}\to\text{Gr}}$, $\mathbf{W}_{\text{PCN}\to\text{Go}}$ are approximating the transformation $\tau\mathbf{B}$, and the weight matrices $\mathbf{W}_{\text{Go}\to\text{Gr}}$, $\mathbf{W}_{\text{Gr}\to\text{Go}}$ approximate the transformation $\tau\mathbf{A}+\mathbf{I}$.

Our final network model is depicted in Figure 2E. Notice that we account for lateral inhibitory Golgi cell connections $\mathbf{W}_{\text{Go}\to\text{Go}}$ (Hull & Regehr, 2012). While this connection does not have any effect on the high-level computation that is being performed (the function being computed is $f(\mathbf{m}) = 0$), it ensures that the Golgi cells receive a self-regulatory inhibitory input current that moves the cells towards their desired working regime. The temporal representation produced by this network is shown in Figure 3C.

Learning

Given the above model for the Golgi and granule cells, we can now introduce learning into the model. The training signal comes from the Inferior Olive (IO), and so it needs to represent the difference between the CR and the UR. These are the two inputs to the IO shown in Figure 1. The CR is the inhibitory input from the Cerebellar Nucleus (CN), and the UR is the excitatory input from the PCN.

To create a neural version of this, we use a similar approach as in Version 2 of the granule/Golgi model, but without the recurrence. That is, we train a single-hidden-layer neural network for the IO, CN, and pRN, and then combine $\mathbf{D}$ and $\mathbf{E}$ matrices to form connection weights.

To adjust the connection weights between the Granule cells and the Purkinje cells, we use the following local learning rule, where ω_{ij} is the connection weight between the ith Granule cell and the jth Purkinje cell, γ is a learning rate parameter, a_i is the spiking activity of the ith Granule cell, a_k is the spiking activity of the kth Inferior Olive cell, ω_{jk} is the connection weight from the kth Inferior Olive cell and the jth Purkinje cell. $\mathbf{W}_{jk}$ is the full matrix of these weights, computed by multiplying the decoder $\mathbf{D}$ from the Inferior Olive and the encoder $\mathbf{E}$ from the Purkinje cells. That is,

$$\Delta\omega_{ij} = \gamma a_i \Big(\sum_k a_k \omega_{jk}\Big) \,, \qquad \mathbf{W}_{jk} = \mathbf{E}_{\text{Purkinje}}\mathbf{D}_{\text{IO}} \,.$$

This learning rule is the Prescribed Error Sensitivity (PES) rule defined by MacNeil and Eliasmith (2011), based on the classic delta learning rule.

Experiments and Results

Figure 4 shows the behaviour of a typical run of the detailed version of our model performing the eye-blink conditioning task over 500 trials. The model learns to have the eye closed when the puff occurs. Most parameters were set based on biological data; $\tau = 5\,\text{ms}$ except in the recurrent granule/Golgi connections, where $\tau_{\text{Gr,Go}} = 60\,\text{ms}$ (Dieudonné, 1998). We simulate $n_{\text{Gr}} = 200$ granule cells and $n_{\text{Go}} = 20$ Golgi cells. The only free parameters are the learning rate $\gamma = 0.00025$, $\tau_{\text{pRN}} = 100\,\text{ms}$ for the connections involving the pRN, and $\tau_{\text{learn}} = 200\,\text{ms}$ for the activity used for the learning rule. The learning rate was adjusted to match the number of trials typically needed for mice to learn the task (~ 300 trials). Velocity commands smaller than $v_{\text{th}} = 2\,\text{mm}\,\text{s}^{-1}$ are counted as zero.

While a detailed analysis of this model is still ongoing, two interesting phenomena have been observed in the model so far. First, the model does not work if the synaptic time constant of the neurotransmitters in the Golgi/granule system significantly greater than the actual biological measured value of 70ms. We are still analyzing exactly why this happens, but it only happens for the fully detailed version of the model, not the simpler recurrent all-to-all version.

Second, the model only learns to close the eye *before* the puff actually happens if the two neurotransmitter time constants τ_{pRN} and τ_{learn} are quite long ($\sim 100\,\text{ms}$). Otherwise, the system learns to close the eye soon *after* the puff. This is because it is trying to do exactly what we have asked it to: learn to re-create the same motor pattern as is produced by the unconditioned reflex. But, the unconditioned reflex closes the eye *after* the puff happens, which is too late. However, by slowing the passage of information from the Cerebellar Nucleus back to the Inferior Olive (where the comparison between the UR and CR occurs), we are effectively comparing the reflex at one point in time to the generated output from the cerebellum at an *earlier* point in time. This allows the new learned reflex to occur slightly earlier than the unconditioned response, and thus the eye closes before the air puff.

Discussion and Future Work

First, we have shown how to take an abstract model of cerebellar function and iteratively add more detail. This general methodology will allow us to continue to refine this model, adding details and determining how these details affect the overall performance of the model.

We are currently building on this work in two directions. First, we are performing a more detailed analysis of how dif-

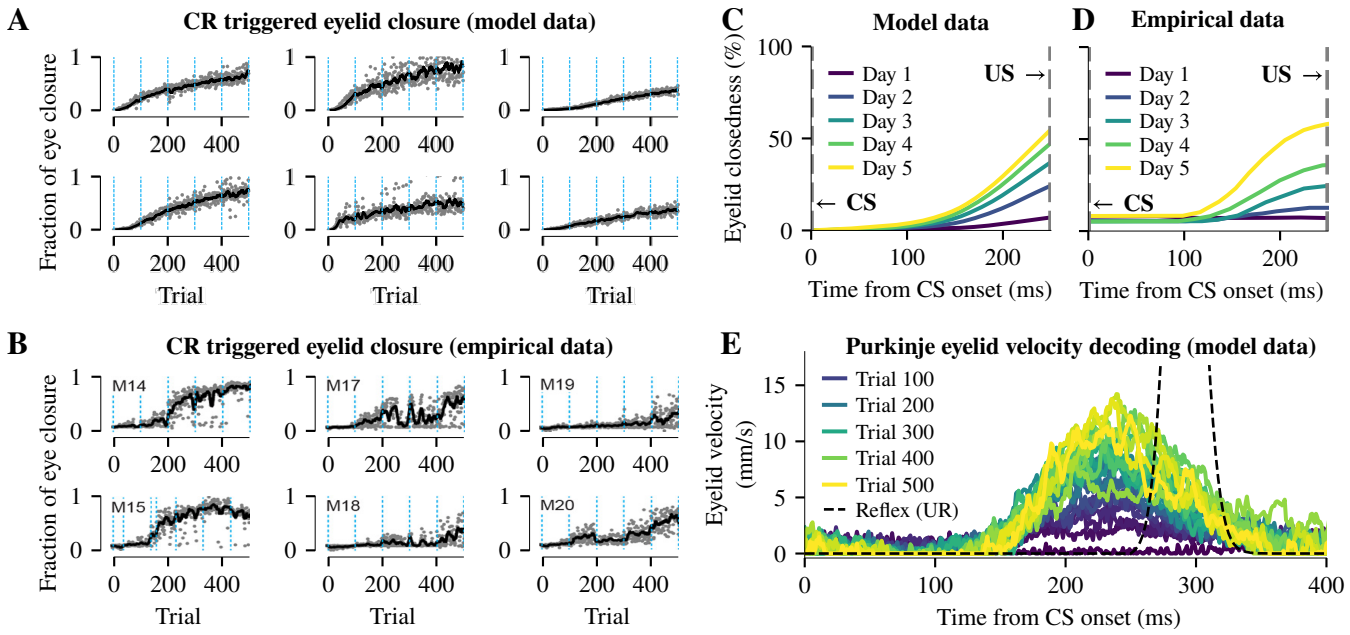

Figure 4: Model and experimental data for the eyeblink conditioning task. **(A, B)** Maximum CR triggered eyelid closure over 500 trials for six random instances of the model/experimental animals. Gray dots correspond to the eyelid closure at the time of the US. Black line is the moving average over 11 trials. Blue dotted lines correspond to an experimental day (100 trials). **(C, D)** Eyelid closure trajectory averaged over one experimental day and all six models/animals. **(E)** Eyelid velocity signal decoded at the Purkinje cells compared to the reflex-triggered velocity signal. Data for **(B, D)** adapted from Heiney et al. (2014).

ferent parameters affect (or do not affect) model performance. Second, we are adding more detailed types of connectivity constraints to our model. For example, granule cells only get input from three to five mossy fibres, and are only connected to Golgi cells in their immediate spatial vicinity.

While our model of eyeblink conditioning is concerned with a relatively low-level task, we hope that the techniques presented here for mapping function onto brain microcircuits are applicable to models of higher-level cognitive function as well. In particular, it would be interesting to see whether our model of the Granule-Golgi circuit in conjunction with the Purkinje cell's plasticity could serve as a supervised learner for timings in cognitive and perceptual tasks, as suggested by various studies (O'Reilly et al., 2008; E et al., 2014).

Acknowledgments

This work was supported by CFI and OIT infrastructure funding as well as the Canada Research Chairs program, NSERC Discovery grant 261453, and AFOSR grant FA9550-17-1-0026.

References

Buckner, R. L. (2013). The Cerebellum and Cognitive Function: 25 Years of Insight from Anatomy and Neuroimaging. *Neuron*, *80*(3).

Dieudonné, S. (1998). Submillisecond kinetics and low efficacy of parallel fibre-Golgi cell synaptic currents in the rat cerebellum. *The Journal of Physiology*, *510*(3).

E, K.-H., Chen, S.-H. A., Ho, M.-H. R., & Desmond, J. E. (2014). A meta-analysis of cerebellar contributions to higher cognition from PET and fMRI studies. *Human brain mapping*, *35*(2).

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. MIT Press.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science*, *338*.

Heiney, S. A., Wohl, M. P., Chettih, S. N., Ruffolo, L. I., & Medina, J. F. (2014). Cerebellar-dependent expression of motor learning during eyeblink conditioning in head-fixed mice. *Journal of Neuroscience*, *34*(45).

Hull, C., & Regehr, W. G. (2012). Identification of an inhibitory circuit that regulates cerebellar golgi cell activity. *Neuron*, *73*(1).

Ito, M. (2010). Cerebellar Cortex. In G. Shepherd & S. Grillner (Eds.), *Handbook of Brain Microcircuits* (1st ed., pp. 293–300). Oxford University Press.

Llinás, R. R. (2010). Olivocerebellar System. In G. Shepherd & S. Grillner (Eds.), *Handbook of Brain Microcircuits* (1st ed.). Oxford University Press.

Lusk, N. A., Petter, E. A., MacDonald, C. J., & Meck, W. H. (2016). Cerebellar, hippocampal, and striatal time cells. *Current Opinion in Behavioral Sciences*, *8*.

MacNeil, D., & Eliasmith, C. (2011). Fine-tuning and the stability of recurrent neural networks. *PLOS ONE*, *6*(9).

Marr, D. (1969). A theory of cerebellar cortex. *The Journal*

of Physiology, *202*(2).

McCormick, D. A., Lavond, D. G., Clark, G. A., Kettner, R. E., Rising, C. E., & Thompson, R. F. (1981). The engram found? Role of the cerebellum in classical conditioning of nictitating membrane and eyelid responses. *Bulletin of the Psychonomic Society*, *18*(3).

Medina, J. F., & Mauk, M. D. (2000). Computer simulation of cerebellar information processing. *Nat. Neurosci.*, *3*(11).

O'Reilly, J. X., Mesulam, M. M., & Nobre, A. C. (2008). The cerebellum predicts the timing of perceptual events. *Journal of Neuroscience*, *28*(9), 2252–2260.

Raymond, J. L., & Medina, J. F. (2018). Computational principles of supervised learning in the cerebellum. *Annual Review of Neuroscience*, *41*(1).

Roth, A., & van Rossum, M. C. W. (2009). Modeling Synapses. In E. D. Schutter (Ed.), *Computational Modeling Methods for Neuroscientists.* The MIT Press.

Rössert, C., Dean, P., & Porrill, J. (2015). At the edge of chaos: How cerebellar granular layer network dynamics can provide the basis for temporal filters. *PLOS Computational Biology*, *11*(10).

Singh, R., & Eliasmith, C. (2006). Higher-dimensional neurons explain the tuning and dynamics of working memory cells. *Journal of Neuroscience*, *26*.

Splittgerber, R. (2018). Cerebellum and its Connections. In *Snell's Clinical Neuroanatomy* (8th ed.).

Strata, P. (2009). David Marr's theory of cerebellar learning: 40 years later. *The Journal of Physiology*, *587*.

Strata, P., & Harvey, R. (1999). Dale's principle. *Brain Research Bulletin*, *50*(5).

Stöckel, A., & Eliasmith, C. (2019). Passive nonlinear dendritic interactions as a general computational resource in functional spiking neural networks. *arXiv*.

Sullivan, E. V. (2010). Cognitive functions of the cerebellum. *Neuropsychology review*, *20*(3).

Voelker, A. R., & Eliasmith, C. (2018). Improving Spiking Dynamical Networks: Accurate Delays, Higher-Order Synapses, and Time Cells. *Neural Computation*, *30*(3).