

Connecting Biological Detail with Neural Computation:
Application to the Cerebellar Granule-Golgi Microcircuit

Andreas Stöckel
University of Waterloo

Terrence C. Stewart
National Research Council of Canada

Chris Eliasmith
University of Waterloo

Author Note

Andreas Stöckel, Centre for Theoretical Neuroscience, University of Waterloo; Terrence C. Stewart, National Research Council of Canada, University of Waterloo Collaboration Centre; Chris Eliasmith, Centre for Theoretical Neuroscience, University of Waterloo.

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

Correspondence concerning this article should be addressed to Andreas Stöckel, Centre for Theoretical Neuroscience, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

E-mail: astoecke@uwaterloo.ca

Abstract

Neurophysiology and neuroanatomy constrain the set of possible computations that can be performed in a brain circuit. While detailed data on brain microcircuits is sometimes available, cognitive modellers are seldom in a position to take these constraints into account. One reason for this is the intrinsic complexity of accounting for biological mechanisms when describing cognitive function. In this paper, we present multiple extensions to the Neural Engineering Framework (NEF) that simplify the integration of low-level constraints such as Dale's principle and spatially constrained connectivity into high-level, functional models. We focus on a model of eyeblink conditioning in the cerebellum, and, in particular, on systematically constructing temporal representations in the recurrent Granule-Golgi microcircuit. We analyse how biological constraints impact these representations and demonstrate that our overall model is capable of reproducing key properties of eyeblink conditioning. Furthermore, since our techniques facilitate variation of neurophysiological parameters, we gain insights into why certain neurophysiological parameters may be as observed in nature. While eyeblink conditioning is a somewhat primitive form of learning, we argue that the same methods apply for more cognitive models as well. We implemented our extensions to the NEF in an open-source software library named "NengoBio" and hope that this work inspires similar attempts to bridge low-level biological detail and high-level function.

Keywords: biologically plausible spiking neural networks; Dale's principle; Neural Engineering Framework; Legendre delay network; cerebellum; eyeblink conditioning

Connecting Biological Detail with Neural Computation:
Application to the Cerebellar Granule-Golgi Microcircuit

Introduction

Human cognition is ultimately grounded in neurophysiological processes. As suggested by Marr's "levels of analysis" (Marr & Poggio, 1976), cognitive scientists tend to implement models of cognition at algorithmic and computational levels, without explicitly taking limitations of the underlying neural substrate into account (Eliasmith & Kolbeck, 2015).

Depending on the hypothesis that is being explored, ignoring biological detail can be reasonable. Yet, a closer look at biology may help in two complementary ways. First, we can *validate* hypotheses about cognition by determining whether it is possible to implement a particular algorithm under the constraints of the biological network in question. Second, we can *generate* new hypotheses by asking what class of algorithms a certain network could support.

We believe that cognitive modeling must ultimately embrace a combination of top-down and bottom-up modeling to narrow down the vast space of possible cognitive science theories and to direct research attention within that space. However, a central roadblock to the adaptation of such methods is the availability of modeling tools that make it possible to specify detailed biological constraints (e.g., neural response curves, spike rates, synaptic time-constants, connectivity patterns) while still being abstract enough to facilitate the specification of high-level cognitive function.

One approach designed to help bridge this gap is the Neural Engineering Framework (NEF; Eliasmith & Anderson, 2003), in conjunction with the related Semantic Pointer Architecture (SPA; Eliasmith, 2013). Up until recently however, it has been unclear how to incorporate certain biological constraints that are often described in the neuroscience literature into NEF networks. For example, and despite initial progress in this direction (Parisien, Anderson, & Eliasmith, 2008), accounting for Dale's principle (i.e., each neuron in the network acting either purely excitatorily or inhibitorily), as well as incorporating spatial connectivity constraints, has been relatively challenging with the existing NEF-based software tool, Nengo (Bekolay et al., 2014).

Furthermore, there have been no studies on how adding these additional constraints influences the high-level function of NEF networks. In this paper, we attempt to address both issues.

Specifically, we describe recent advances in modeling techniques that partially alleviate the shortcomings of the NEF mentioned above. We then use these methods to construct a model of eyeblink conditioning in the cerebellum, with a focus on the generation of temporal basis function representations in the recurrent Granule-Golgi circuit.

Building a model of temporal representation in the cerebellum is of particular interest, since it remains unclear how exactly the cerebellum manages to learn and reproduce the precise timings observed in eyeblink conditioning. Furthermore, the mechanisms underlying eyeblink conditioning are potentially exploited by cognitive processes as well. 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).

Our experiments suggest that—at least under the constraints we consider—the Granule-Golgi circuit is well-suited to encode temporal information using basis functions. This representation can be used in the context of a spiking neural network to learn delays by modulating synaptic weights in the granular-to-Purkinje projection. Furthermore, we generate hypotheses as to why various biological parameters (such as the sparse connectivity patterns and the time-constants of the neurotransmitters) are as observed.

This paper builds on Stöckel, Stewart, & Eliasmith, 2020a, 2020b, and is structured as follows. We first review the high-level function we hypothesize could be implemented in the Granule-Golgi circuit, followed by an overview of the eyeblink conditioning task, the particular neurophysiological constraints of the cerebellum, and high-level theories of cerebellar function. We then discuss five neural network implementations with an increasing amount of biological

detail, along with the corresponding extensions to the NEF. We perform a series of experiments that explore the impact of individual parameters on the performance of the increasingly realistic system. We extend our model to perform the complete eyeblink conditioning task by incorporating the remaining cerebellar microcircuitry and compare the model behavior to empirical data. Finally, we provide a quick overview of “NengoBio,” the open-source addon to Nengo we developed to encode the aforementioned biological constraints.¹

Background

In order to explore the consequences of adding biological details to a neural system, we need to choose the desired computation that the neural system should ideally perform. In machine learning, the simplest artificial neural networks are purely feed-forward, i.e., they possess no backwards-directed or *recurrent* connections. It is well-known that feed-forward neural networks are universal function approximators (Hornik, Stinchcombe, & White, 1989). That is, given enough neurons, any function $f(x) = y$ can be implemented as a neural network by simply having a single hidden layer of neurons that receives x as an input (via a set of input weights) and produces y as an output (via a set of readout weights).

Neurobiological systems differ from the artificial neural networks mentioned above in two key aspects. First, they are intrinsically dynamical systems, i.e., input and output are functions over time. Second, they often include recurrent connections. As has been shown by Eliasmith and Anderson (2003), adding recurrent connections along with their associated synaptic dynamics allows for the creation of neural networks that approximate any differential equation of the form $d\mathbf{m}/dt = \dot{\mathbf{m}} = f(\mathbf{m}, u, t)$, where f is an arbitrary function describing the dynamics, \mathbf{m} is the system state, u is an external input, and t is time. Again, with a sufficient number of neurons and the corresponding connectivity, such differential equations can be approximated to any desired degree of accuracy. Our primary goal in this paper is to explore how well such computations can be performed in the presence of biological constraints. Once we have established this

¹ See <https://github.com/astoeckel/nengo-bio>.

computational basis, we extend our network to a complete model of the eyeblink conditioning task and compare our simulation results to empirical data.

The Legendre Delay Network

As a benchmark for evaluating the impact of incorporating biological detail into a functional description, we choose the following linear differential equation $\theta \dot{\mathbf{m}} = \mathbf{A}\mathbf{m} + \mathbf{B}u$, where

$$(\mathbf{A})_{ij} = \begin{cases} (2i+1)(-1) & \text{if } i < j, \\ (2i+1)(-1)^{i-j+1} & \text{if } i \geq j, \end{cases} \quad (\mathbf{B})_i = (2i+1)(-1)^i. \quad (1)$$

Here, \mathbf{m} is a q -dimensional vector describing the current state of the linear system, \mathbf{A} is a $q \times q$ matrix describing the state evolution, and \mathbf{B} is a $q \times 1$ matrix describing how a scalar input u influences the state. This particular equation has been derived by taking the Padé approximants of the continuous-time delay $F(s) = e^{-\theta s}$ in the Laplace domain. As such, this differential equation stores a compressed version of the past θ seconds of its inputs in the state variable \mathbf{m} (Voelker & Eliasmith, 2018; Stöckel, 2021b). That is, given \mathbf{m} at any point in time t , it is possible to recover an approximate value of u at a previous point in time $\hat{u}(t - \theta')$ for $0 \leq \theta' \leq \theta$:

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

The function \tilde{P}_{ℓ} is the shifted Legendre polynomial of degree ℓ .

Another way to think of this system is that it encodes the past history of its inputs using a set of *temporal basis functions*. The temporal basis functions that happen to form the representational basis of the above linear system are approximations of the Legendre polynomials \tilde{P}_{ℓ} (see Stöckel, 2021a for a derivation). Of course, some information is lost in this process of compressing the history of u into \mathbf{m} . The quality with which past u can be reconstructed from \mathbf{m} is controlled by the number of state dimensions q , or, more precisely, the ratio of q and θ . As q increases for a constant θ , more details (i.e., higher frequencies) about the past are represented in \mathbf{m} . See Fig. 1 for a schematic overview.

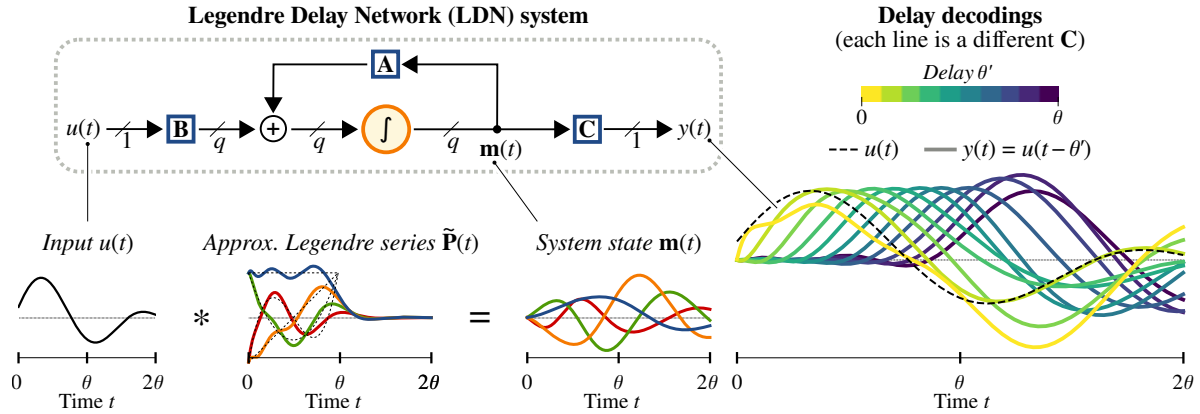


Figure 1. Overview of the Legendre Delay Network (LDN). Feeding a signal $u(t)$ into the LDN system implicitly convolves $u(t)$ with approximations of the first q shifted Legendre polynomials $\tilde{\mathbf{P}}$. Hence, at each point in time t , the system state $\mathbf{m}(t)$ is the generalized Fourier spectrum of the past θ seconds of the input $u_{[t-\theta, t]}$ with respect to the Legendre basis. Linear combinations of the state $\mathbf{C}\mathbf{m}(t) = y(t)$ form functions over $u_{[t-\theta, t]}$, including, but not limited to, delays $\hat{u}(t - \theta')$.

The neural implementation of this computation is called a “Legendre Delay Network” (LDN). Crucially, arbitrary functions over the time interval $[t - \theta, t]$ can be decoded from the neurons representing the network state \mathbf{m} , and not just delays. In other words, the LDN forms the filter bank of an adaptive filter, a hypothesized functional description of the Granule-Golgi circuit in the cerebellum (Fujita, 1982). The LDN has previously been used to model cognitive timing behavior in humans (de Jong, Voelker, van Rijn, Stewart, & Eliasmith, 2019) and is also the core component of a novel machine learning algorithm known as the Legendre Memory Unit (LMU), which has been shown to outperform LSTMs on several benchmark tasks (Voelker, Kajić, & Eliasmith, 2019).

If we use biologically constrained neural networks to approximate the above linear system, then the actual computation performed by the system, and hence the quality of the time window encoded in \mathbf{m} , may be different. We can thus use this system as a benchmark. In the following, after discussing the cerebellum in more detail, we build various approximations of the LDN using different biological constraints, systematically provide them with inputs, and evaluate their performance in terms of how well the input history can be recovered. Finally, we integrate the

LDN into a model of the cerebellum performing the eyeblink conditioning task. To create the model, we compute optimal input and recurrent connection weights independently for each target population to approximate functional descriptions such as eq. (1) given biological constraints. This means we are not modeling the entire developmental process for the cerebellum, but are rather directly creating a model of a mature cerebellum. For the full eyeblink conditioning task, we then also apply a local biologically plausible learning rule to learn the readout weights.

Eyeblink Conditioning

The cerebellum is well-studied and highly regular in its structure, and there are reasons to believe that the cerebellum does compute something akin to the operation performed by the Legendre Delay Network (Fujita, 1982). Behaviorally, the cerebellum is known to be vital for some delay conditioning tasks, such as eyeblink conditioning.

In 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), for example a short tone. The CS precedes the US by a constant time offset Δt . Over time the subject forms a conditioned response (CR) to the previously neutral CS, maintaining the fixed delay Δt . In other words, the subject will learn to have its eye closed Δt seconds after the tone, even if the puff is absent (cf. Heiney, Wohl, Chettih, Ruffolo, & Medina, 2014). Experiments indicate that the formation of this conditioned response critically depends on the cerebellum; previously learned CRs are absent once the cerebellum is ablated (McCormick et al., 1981).

Cerebellar Microcircuitry

Before we continue to discuss the two prevalent theories on how the cerebellum could learn the conditioned response, it is worthwhile to review the cerebellar microcircuitry depicted in Fig. 2. Afferent nerves from pre-cerebellar (PC) nuclei (brainstem and cerebral cortex carrying sensory signals) project as “mossy fibers” onto granule cells in the cerebellar cortex. Cerebellar

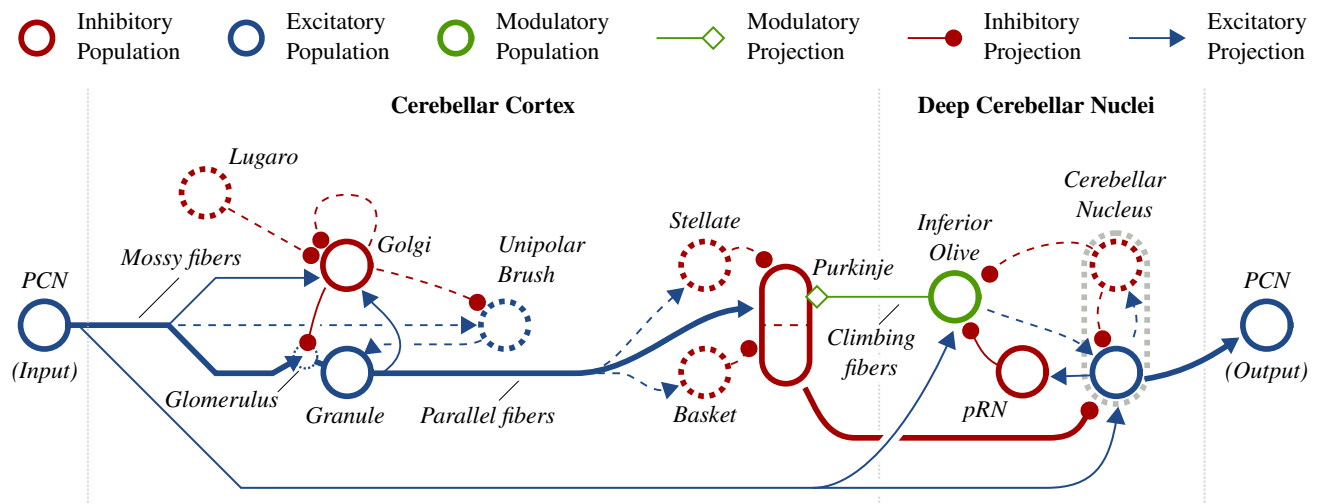


Figure 2. Schematic of the cerebellar microcircuit. Dashed projections and populations are not included in our model. Cerebellar nucleus afferents affect both the excitatory and inhibitory sub-populations. The main feed-forward pathway is highlighted in bold. *PCN* $\hat{=}$ pre-cerebellar neurons/nuclei. *pRN* $\hat{=}$ parvicellular Red Nucleus. Data from Ito, 2010; Llinás, 2010.

granule cells are tiny and account for the majority of neurons in the mammalian brain. Thus, very few PC neurons connect onto a very large number of granule cells.

Granule cells also have inter-neurons interspersed amongst them, known as Golgi cells, forming an inhibitory feedback loop with the granule cells. That is, granule cells excite Golgi cells, and Golgi cells inhibit granule cells (Ito, 2010). Notably, the connections from Golgi and PC cells to granule cells are formed through so called glomeruli. Each granule cells extends to on average four glomeruli and, at each glomerulus, receives input from one PC neuron through a mossy fiber terminal, as well as one or more Golgi cells (Palkovits, Magyar, & Szentágothai, 1972; Jakab & Hátori, 1988; Chadderton, Margrie, & Häusser, 2004). Furthermore, the connectivity between Golgi and granule cells is spatially constrained, i.e., Golgi cells only connect to granule cells in their vicinity (D’Angelo et al., 2013). The ratio of granule to Golgi cells is about 400:1 (Korbo, Andersen, Ladefoged, & Møller, 1993).

Granule cell axons, the so called “parallel fibers,” project onto the Purkinje cells, which inhibit neurons in the cerebellar nucleus, and in turn project back onto the brainstem and cerebral

cortex (Ito, 2010; Llinás, 2010). So called “climbing fibers” project from Inferior Olive neurons in the deep cerebellar nucleus onto Purkinje cells. Evidence suggests (Ito, 2010) that activity in the climbing fibers is responsible for modulating the synaptic strength of granule to Purkinje projections. Climbing fiber activity can thus be interpreted as an “error signal” responsible for driving learning in the cerebellar cortex.

Hypothesized Mechanisms Supporting Delay Learning

There is no current consensus on what the exact mechanism is that supports delay learning in the cerebellum. The classical view is the aforementioned adaptive filter theory (Fujita, 1982). As originally pointed out by Marr (1969), the massive divergence (number of post-neurons for a single pre-neuron) and small convergence (number of pre-neurons for a single post-neuron) in the PC to granule projection suggests that granular cells are tuned to specific patterns of activity in the PC neurons. The adaptive filter theory can be interpreted as extending this idea toward temporal tuning. Granule cell activity does not only depend on the current PC activities, but on their time-course. Such a temporal tuning could be the result of the recurrent granule-to-granule connections mediated by the inhibitory Golgi cells. If this temporal tuning is diverse enough, i.e., granule cells are sensitive to different time-courses, this will form a suitable temporal basis from which arbitrary delays can be decoded. Learning a specific decoding is driven by the climbing fiber error signal, modulating the synaptic weights between the granule and the Purkinje cells. As mentioned above, this is similar in principle to the Legendre Delay Network.

A more recent theory is that responses observed in tasks such as eyeblink conditioning inherently rely on intrinsic properties of the Purkinje cells. In other words, temporal properties of the granule cells play a lesser role. Instead, climbing fiber input triggers processes within the Purkinje cell and their dendritic structures that are responsible for the formation of a delayed output. Indeed, Johansson et al. (2014) find that bypassing the granule cells and directly injecting signals into the parallel fibers still evokes a previously learned delayed response from the cerebellum, albeit a weaker one.

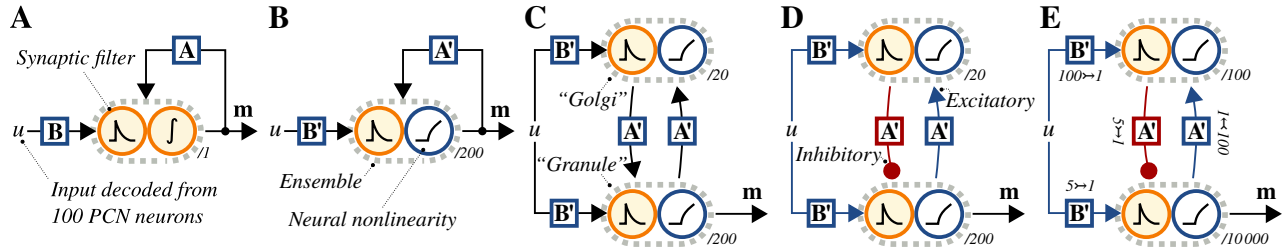


Figure 3. Network types used in our experiments. (A) “Direct” implementation with an optimal integrator. (B) Using the synaptic filter of a single population of spiking neurons for temporal integration. (C) Inter-neuron population in the recurrent path. (D) Same as C, but accounting for Dale’s principle. (E) Same as D, but with more detailed biological constraints (see text).

We think that these two theories are not inherently contradictory. Both temporal tuning of the granular cells and intrinsic temporal properties of the Purkinje cells could play a role in delay learning. The findings we present in this paper provide a strong argument that the biology of the Granule-Golgi circuit is at least well suited for implementing some kind of temporal basis function generation, and as such confirms the results of previous studies (cf. Dean, Porrill, Ekerot, & Jörntell, 2010; Rössert, Dean, & Porrill, 2015). Still, our work only takes a fraction of the available data on cerebellar neurophysiology into account and as such should not be seen as strong evidence for or against one of the two proposed mechanisms. Instead, we are interested in describing methods for adding biological constraints while determining their consequences for higher-level function.

Levels of Biological Detail

To demonstrate our approach of adding biological detail, we first focus on a model of temporal basis function generation in the Granule-Golgi circuit discussed above. We present five models of increasing complexity—the first model is merely an abstract implementation of eq. (1), the final model respects spatial sparsity, convergence, tuning curves, and, to a degree, neurotransmitter constraints. All models are depicted in Fig. 3. In all cases, the scalar input u is received from one hundred spiking Leaky Integrate-and-Fire (LIF) neurons with randomly chosen

tuning curves, representing input provided by the PCN (see Model B for details).

Model A: “Direct” Implementation. For this model, we directly solve the differential equation in eq. (1) by integration. That is, we have a single layer of “neurons” that are pure integrators (i.e., no non-linearity). The matrix \mathbf{A} describes the recurrent connection weights, and \mathbf{B} the input connection weights. This model does not distinguish between the granule and Golgi cells, and does not include details such as individual neurons or spikes. Instead, it focuses on the high-level theory of what the system is computing.

Model B: Single Population. We replace the integrators with a single layer of 200 spiking Leaky Integrate-and-Fire (LIF) neurons. These neurons form a distributed representation of \mathbf{m} using a population code. Each neuron i is parametrized by a randomly chosen preferred stimulus vector \mathbf{e}_i (for *encoder*), gain α_i and bias current J_i^{bias} , resulting in a desired response (i.e., tuning curve) for each neuron:

$$a_i(\mathbf{m}) = G[J_i(\mathbf{m})] = G[\alpha_i(\mathbf{e}_i \cdot \mathbf{m}) + J_i^{\text{bias}}], \quad (3)$$

where G is the neural response curve of the LIF neuron model. The parameters α and J_{bias} are randomly chosen from a distribution that ensures a maximum firing rate of 50 Hz to 100 Hz, consistent with biological recordings of granule cells (Chadderton et al., 2004). We then use least-squares to solve for optimal input and recurrent connection weights that result in these desired tuning curves while implementing the equivalent calculation as in Model A. Importantly, when solving for the recurrent connection weights, we also take into account the synaptic filter, which we model here as a decaying exponential (i.e., a low-pass). This is the standard process in the NEF (Eliasmith & Anderson, 2003).

Model C: Inter-neurons. As a next step, we separate the single layer of neurons into two populations corresponding to the Golgi and granule cells, reflecting the actual biology of the cerebellum (see above). This introduces two synaptic filters which need to be taken into account when solving for the connection weights that best approximate eq. (1). Furthermore, to at least approximate the fact that there are far fewer Golgi cells than granule cells, we use 20 Golgi cells and 200 granule cells.

Model D: Inhibition and Excitation. So far, we have not accounted for Dale’s principle, i.e., Golgi cells being purely inhibitory, and granule cells being purely excitatory. We handle this by switching to the non-negative least-squares problem described in Stöckel & Eliasmith, 2021.

For each post-neuron i we minimize

$$\min_{\mathbf{w}_i^+, \mathbf{w}_i^-} \sum_{k=1}^N \left(\mathbf{w}_i^+ \cdot \mathbf{a}_k^+ - \mathbf{w}_i^- \cdot \mathbf{a}_k^- - J_i(\mathbf{m}_k) \right)^2 \text{ w.r.t. } \mathbf{w}_i^+, \mathbf{w}_i^- \geq 0,$$

where \mathbf{a}_k^+ , \mathbf{a}_k^- are the excitatory and inhibitory pre-activities for sample k , \mathbf{w}^+ , \mathbf{w}^- are the connection weights for excitatory and inhibitory pre-neurons, and $J_i(\mathbf{m}_k)$ is the current required to represent the desired value \mathbf{m}_k as defined in eq. (3).

Model E: Sparse connectivity and activity. For this model, we add in realistic constraints on how connected the neurons are. The previous models used all-to-all connections, whereas for this model, we only allow a subset of those connections to be non-zero. This applies to both the input to the Granule-Golgi system and for the recurrent connections within the granule cells. In particular, we account for the granule cell convergence numbers by randomly selecting five PCN and five Golgi cells as pre-neurons—this number is slightly larger than the number reported above, since, as we discuss below, the number of pre-neurons places a strict upper limit on the connectivity. Given this extremely sparse connectivity, we increase the number of neurons in the simulation to 10 000 granule and one hundred Golgi cells, which is closer to the ratio observed in nature.

To account for spatially imposed connectivity constraints, we assign a location \mathbf{x} in $[-1, 1]^2$ to each neuron. The probability p_{ij} of a post-neuron i to receive a connection from a pre-neuron j is proportional to $\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/\sigma^2)$ (Fig. 4).

Finally, the input representation in the PC neurons was made more temporally sparse by adjusting the gain and bias parameters in eq. (3). Data reported by Chadderton et al. (2004) indicate that granule cell excitatory event rates—which are driven by PCN activity—drop from an average of 40 s^{-1} with a stimulus being present to only 8.5 s^{-1} when there is no stimulus. We adjusted the PCN tuning curves to match these statistics in the final network.

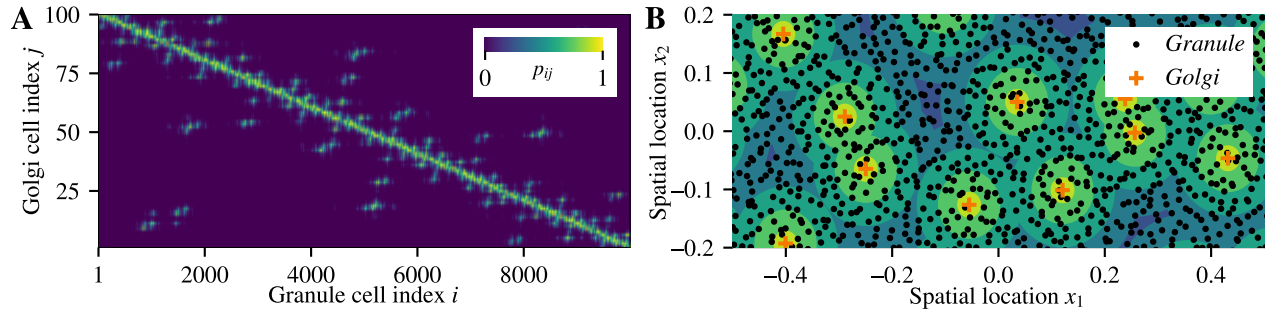


Figure 4. Spatial connectivity constraints. **(A)** Normalized connection probabilities p_{ij} for $\sigma = 0.25$. **(B)** Spatial organisation of the Golgi and granule cells. The background depicts the cumulative density of the Golgi to granule connection probability for a virtual Granule cell at each location (same colors as in **A**).

Experiments and Results

In the following, we discuss three experiments. First, we analyze the temporal representations produced by each of the increasingly constrained models described above. Second, we systematically vary model parameters to gain a better understanding of why certain parameters are as observed. Third, we extend our model to incorporate the remaining components of the cerebellar microcircuit discussed above, resulting in a complete model of eyeblink conditioning.

Influence of biological constraints on temporal representations

To evaluate the effects of these biological details, we systematically generate two different types of input $u(t)$, feed those into the network and record the resulting activity. In particular, we present results with a periodic pulse input of varying pulse width t_{on} and band-limited white noise of varying bandwidth B . These are meant to depict typical sorts of inputs that may arise in experimental situations (pulses) and more real-world situations (band-limited white noise). We then determine how accurately the past history of u over the window θ can be recovered from the resulting network activity via optimal linear readout weights. We use $\theta = 0.4$ s in all simulations; each individual experiment simulates the network for $T = 10$ s. The error is the RMSE of the

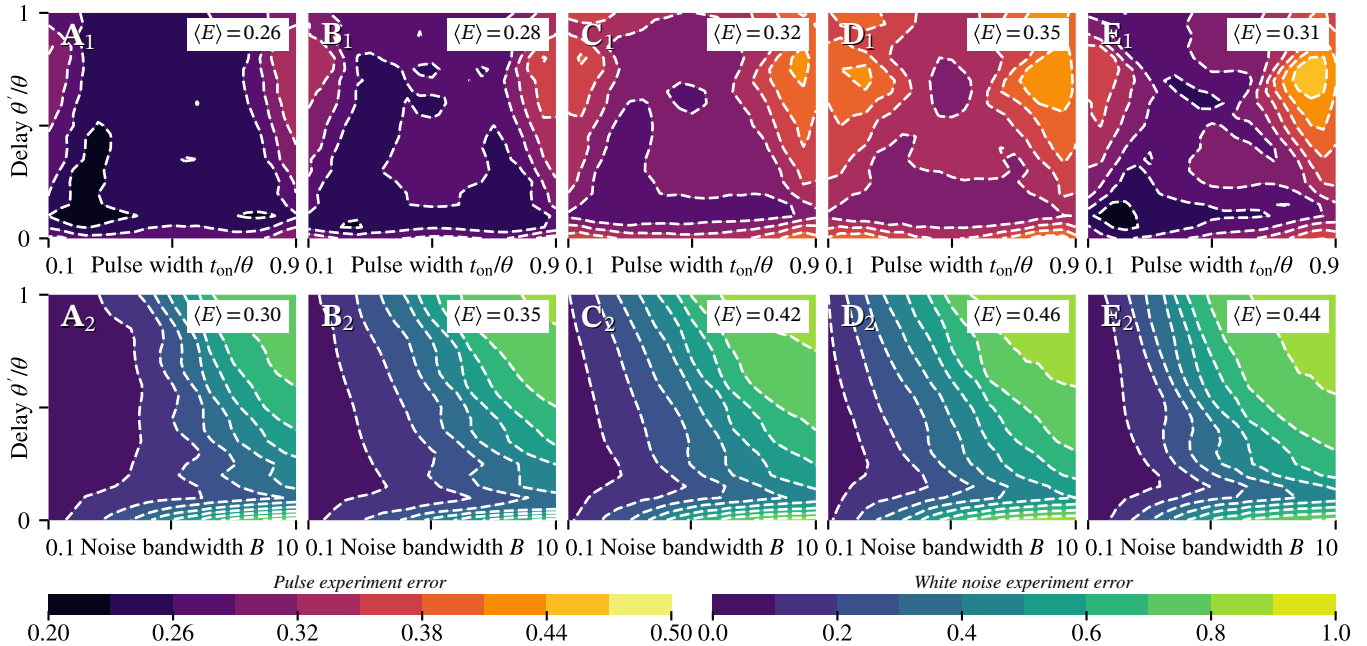


Figure 5. Delayed signal reconstruction error for different types of input signals, delays, and network types. All error values are expressed as RMSE divided by the average RMS of the input signal over ten trials. Columns correspond to the network types in Fig. 3 above. *Top row:* Reconstruction error for rectangle pulse signals of varying width. *Bottom row:* Reconstruction error for a band-limited white noise input signal with varying band-limit.

reconstruction divided by the RMS power of the input itself (normalized RMSE, or NRMSE).

The overall results for all five models are shown in Fig. 5. This shows the average reconstruction error for varying inputs (horizontal axis) and for varying time delays (vertical axis) over ten trials. An example run of Model E (the model with the most biological detail) is shown in Fig. 6. The different decoded output lines (bottom) are all based on the same neural activity (middle), but with different readout weights. These approximate the input value u with various time delays over the entire time window, from the immediate input right now ($\theta'/\theta = 0$) to θ seconds ago ($\theta'/\theta = 1$).

As can be seen in Fig. 5, the network successfully functions as a method for encoding the temporal pattern of input data over the desired window of time θ . Adding more biological detail

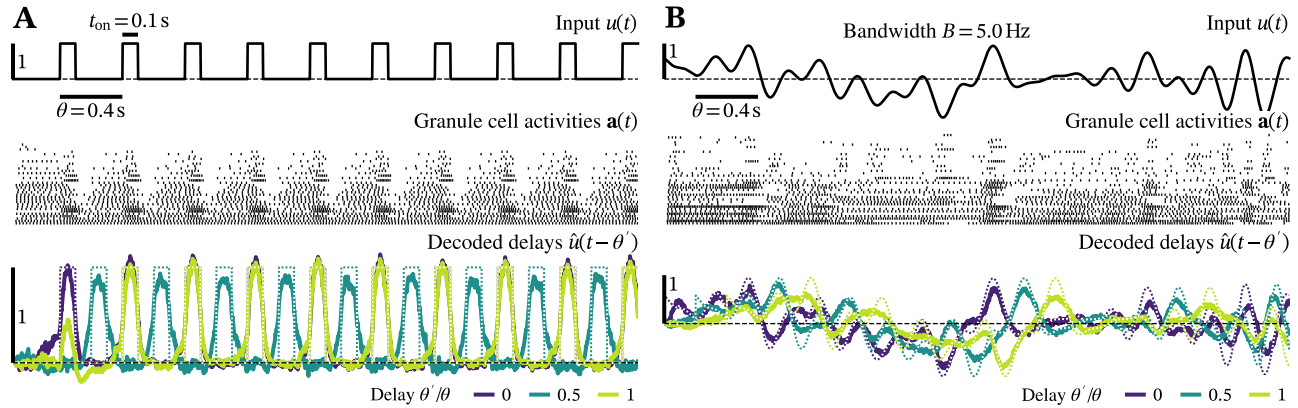


Figure 6. Examples showing the delayed input signals decoded from the granule layer in the detailed model (Fig. 3E). *Top row:* Input signal (rectangle pulses in **A**, white noise in **B**). *Middle row:* Spike raster for 40 randomly selected granule cells. *Bottom row:* Delays decoded from one thousand randomly selected granule cells. Dotted lines correspond to an optimal delay.

generally decreases the accuracy with which the system approximates eq. (1), but most of the information is still encoded. A small improvement in performance can be observed for the most detailed model. We think that this is mostly due to the increased number of neurons.

The pulse input (Fig. 6A) shows that the reconstruction is a smoother version of the input; the system is not good at representing sudden changes, and this is the main source of noise in the reconstruction. This is as expected from using smooth Legendre polynomials as temporal basis functions. Furthermore, we note that there is a peak in accuracy when decoding data from $\theta' = 60$ ms in the past ($\theta'/\theta = 0.15$), and this peak is more pronounced as more biological detail is added. This corresponds to the neurotransmitter time-constant $\tau = 60$ ms we use for all connections, and which is based on a first-order low-pass fit to the NMDA Granule-Golgi dynamics reported in Dieudonné (1998). Note that the reported time-constants of other connections in the system are significantly shorter (Kanichay & Silver, 2008). Techniques described by (Voelker & Eliasmith, 2018) could be used to compute the connections weights for these heterogeneous time-constants. Yet, for the sake of simplicity, here we assume homogeneous time-constants.

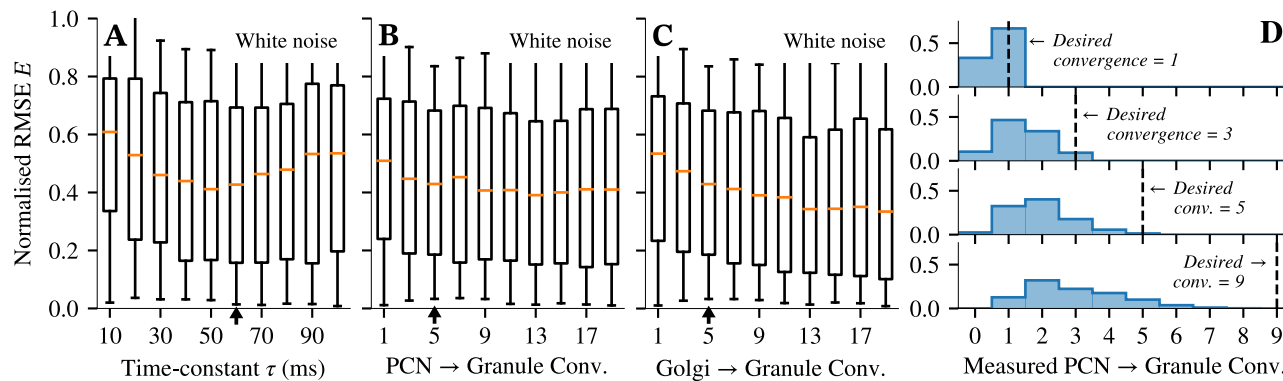


Figure 7. Parameter exploration. **(A-C)** Effects of varying parameters on the delay reconstruction error (arrows indicate default parameters). The box plots show the median, lower and upper quartile of all the data depicted as a contour plot in Fig. 5; whiskers are the min/max. The total number of data-points in each bar plot is $n = 441$. **(D)** Histograms showing the frequency of measured PCN to granule convergences.

Parameter exploration

Notably, we can use our model to determine what the accuracy would be like if we changed individual parameters, such as the aforementioned synaptic time-constant. This is shown in Fig. 7A. Interestingly, the performance of the system is best for a filter time-constant of 50 ms, which is closed to the value we used based on measured Granule-Golgi dynamics.

As discussed above, a striking feature of the cerebellar microcircuitry is the low granule cell convergence. One possible hypothesis is that these numbers are a trade-off between minimizing connectivity and the overall performance of the resulting system. In our model, we can test this hypothesis by systematically varying the number of pre-neurons the solver has access to. Results are shown in Figs. 7B and C.

For the PCN to Granule convergence (Fig. 7B) the performance of the system does improve with larger limits, yet plateaus at still relatively small convergence numbers. Importantly, as mentioned above, the specified desired convergence solely controls the number of *potential* pre-neurons. Since the neural weight solver may still set a weight to zero, these convergence numbers are strict upper limits. Measuring the actual convergence in the PCN to granule

connections (Fig. 7D), we see a peak at one to three PC neurons connecting to each granule cell, and this peak is only weakly affected by the desired convergence number. Importantly, in nature, PC neurons connect to on average four granule cells (see Palkovits et al., 1972, for the complete statistics), whereas we only measure an average convergence of two in our model (for a desired convergence of five). We can match the observed average in our model by increasing the desired convergence to 13. This slightly increases the performance of the system and coincides with the optimal desired convergence depicted in Fig. 7B. Nevertheless, since the gain in performance is relatively small, we decided not to alter the desired convergence in our model setup.

Fig. 7C indicates that the performance of the system improves up to a desired Golgi to granule convergence number of thirteen, with a measured average convergence of 4.8 in the final network. This is a little higher than what is commonly assumed in the neuroscience literature, although, notably, there is some uncertainty regarding this number. The original study by Jakab and Hámori (1988) on the physiology of individual glomeruli (the sites receiving Golgi cell axons and granule cell dendrites; cf. Fig. 2) counts up to 145 Golgi axon synapses in one glomerulus. However, it is unclear whether these synapses originate from a single or different Golgi cells. Most researchers assume at most one Golgi cell connecting to each glomerulus, and, as explained above, each granule cell in turn connecting to on average of four glomeruli (Palkovits et al., 1972). Our model predicts that if the Granule-Golgi circuit were to optimally generate a temporal basis, then multiple Golgi cells should sometimes connect to the same glomerulus.

Extension Toward a Model of Eyeblink Conditioning

Given the most detailed model above for the Golgi and granule cells, we can now introduce learning into the model to test the behavior of the system in an eyeblink conditioning task. As discussed above, the error signal $\varepsilon(t)$ originates from the Inferior Olive (IO). This signal needs to represent the difference between the conditioned response (CR; i.e., what the model has learned so far), and the unconditioned response (UR; i.e., the motor response produced by the innate eyeblink reflex). These are the two inputs to the IO shown in Figure 2. 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 described in the second version of the Granule-Golgi circuit. That is, we train a single-hidden-layer neural network for each of the IO, CN, and pRN using least-squares, and then combine the **D** and **E** matrices to form connection weights. This is a standard technique when building NEF models.

To adjust the connection weights between the Granule cells and the Purkinje cells, we use the Prescribed Error Sensitivity (PES) rule defined by MacNeil and Eliasmith (2011), a biologically plausible variant of the classic delta learning rule. Let $w_{ij}^{\text{Gr} \rightarrow \text{Pu}}$ be the connection weight between the j th granule cell and the i th Purkinje cell, η a learning rate parameter, $a_j^{\text{Gr}}(t)$ the j th post-synaptic granule cell activity filtered by a low-pass filter with time-constant τ_{learn} , and $e_i^{\text{Pu}} \in \{-1, 1\}$ the encoder of the i th Purkinje cell, determining whether this cell is an on- or off-neuron. The delta learning rule is given as

$$\frac{d}{dt} w_{ij}^{\text{Gr} \rightarrow \text{Pu}}(t) = -\eta \varepsilon(t) e_i^{\text{Pu}} a_j^{\text{Gr}}(t). \quad (4)$$

This learning rule can be derived from gradient descent in a single-layer network. For a small enough η , the weights are guaranteed to converge to the least-squares solution used in the previous experiments. The PES rule is given by rewriting eq. (4) in terms of local weights and activities

$$\frac{d}{dt} w_{ij}^{\text{Gr} \rightarrow \text{Pu}}(t) = -\eta \sum_k w_{ik}^{\text{IO} \rightarrow \text{Pu}} a_k^{\text{IO}}(t) a_j^{\text{Gr}}(t),$$

where $a_k^{\text{IO}}(t)$ is the activity of the k th IO cell and $w_{ik}^{\text{IO} \rightarrow \text{Pu}}$ are the synaptic weights between the k th IO cell and the i th Purkinje cell. These weights are the product of the Purkinje cell encoder e_i^{Pu} and a decoding vector \mathbf{d}_k^{IO} that linearly decodes the error signal $\varepsilon(t)$ from IO cell activity.

Figure 8 shows the behavior of a typical run of the detailed version of our model performing the eyeblink conditioning task over 500 trials. The “tone” (CS) is modeled as a rectangle pulse with $t_{\text{on}} = 100$ ms. The “puff” (US) occurs 250 ms after the CS onset. The model learns to have the eye closed when the puff occurs. Notably, individual instances of the network show slight differences in learning speed, just as individual experimental animals do.

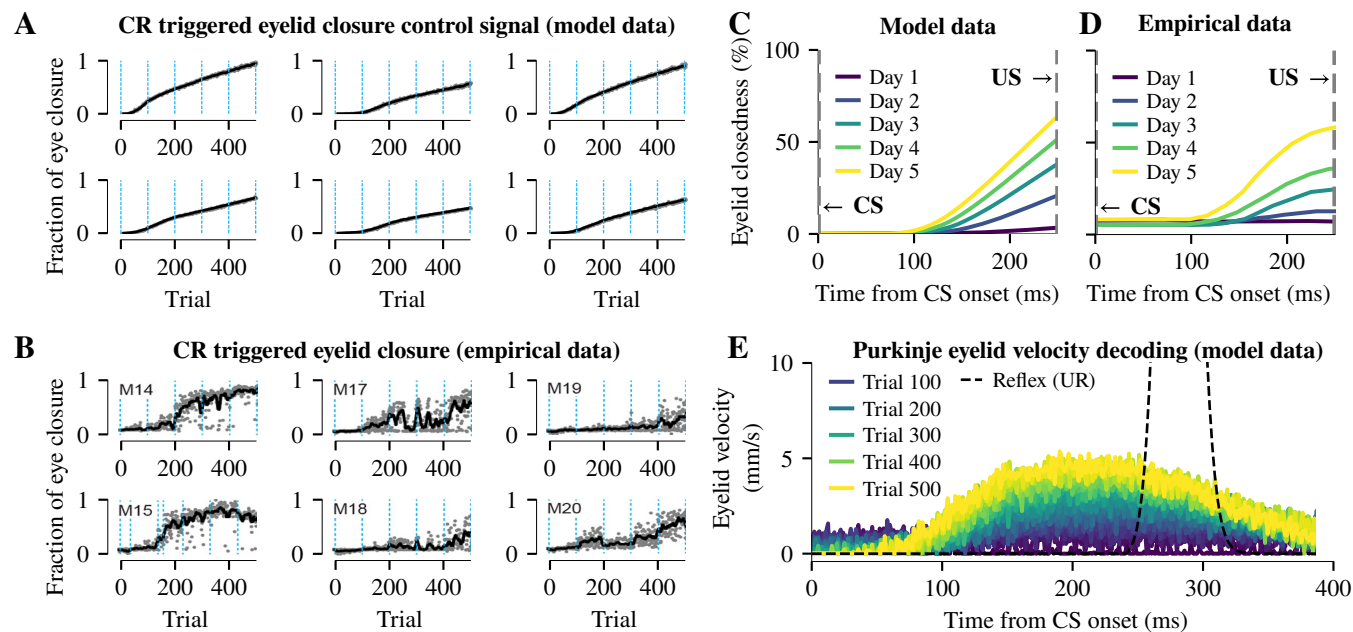


Figure 8. Model and experimental data for the eyeblink conditioning task. **(A, B)** Maximum CR triggered eyelid closure over 500 trials for six random instances of the model/experimental animals. Gray dots correspond to the eyelid closure at the time of the US. Black line is the moving average over 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).

While our model reproduces key features of eyeblink conditioning, its behavior differs from the experimental data in some aspects. Foremost, the model does not reproduce the increase in learning speed over time; to the contrary, learning slows down as the model converges to the optimal set of parameters. Furthermore, our model shows far less inter-trial variance compared to experimental animals. We think that the reasons for this are twofold. First, the 10 000 granule cells used in our experiments provide an on average very stable temporal basis from which we can decode the motor control signal. Second, we do not model external systems that might interfere with the cerebellar motor commands, such as signals originating from motor cortex, or the physics of the eyelid itself. Since these processes may be a significant source of inter-trial

variance, it is not unsurprising that our model produces a relatively noise-free output. Still, more research in these areas will be required in the future.

Most parameters were set based on biological data; the synaptic time-constant $\tau = 5$ ms except in the Granule-Golgi microcircuit, as described above. The only free parameters are the learning rate $\eta = 140 \times 10^{-6}$, $\tau_{\text{pRN}} = 100$ ms for the connections involving the pRN, and $\tau_{\text{learn}} = 60$ ms for filtering the granule cell activity in 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 $v_{\text{th}} = 2 \text{ mm s}^{-1}$ are counted as zero.

The low-pass filter τ_{learn} on the learning connection is required to reproduce the observation that the animal will close the eye slightly *before* the puff occurs as learning progresses. Without the additional filter 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. We note that this is a situation where adding biological detail (the synaptic filtering) improves the performance of the model. If the model were to learn to *exactly* reproduce the UR given the CS, then the eyeblink would occur *after* the puff of air, which would be a less useful response.

Overview of Our Modeling Tool

While it may seem as if our methods were tailored to the cerebellum model discussed above, this is decidedly not the case. We integrated all extensions made to the NEF into an open-source library called “NengoBio” that extends the neural network simulator Nengo.

An example demonstrating the practical usage of NengoBio is given in Fig. 9. In addition to

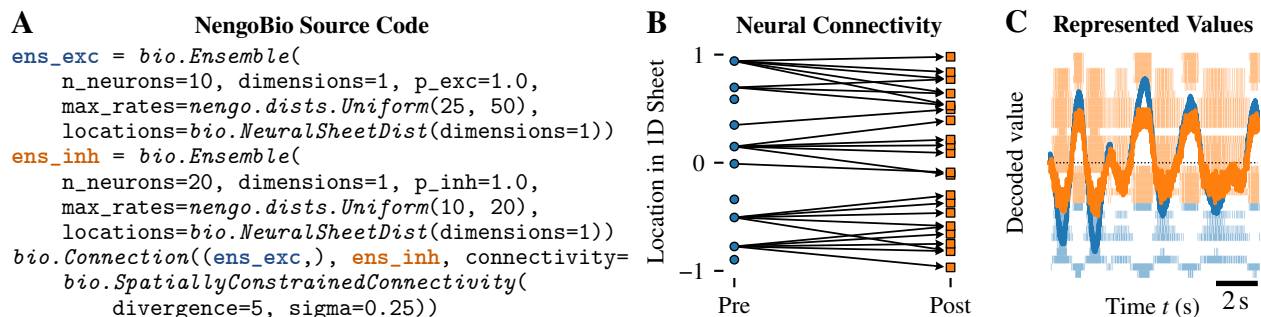


Figure 9. Example demonstrating NengoBio. **(A)** Code describing a communication channel between an excitatory and an inhibitory neuron population. **(B)** Neurons are assigned a one-dimensional location; connection probabilities are based on spatial proximity. Arrows indicate actual connections. **(C)** Decoding the represented values from the spiking activity (spike raster in the background) of the two neuron ensembles (filtered with a 100 ms low-pass). Although the software solves for optimal connection weights, the realized communication channel is not perfect due to the biological constraints. Such deviations could inform hypotheses about the high-level function of a brain circuit.

constraints already supported by Nengo itself, such as maximum firing rates and tuning properties, individual neural ensembles can be marked as being purely excitatory or inhibitory. This information is automatically taken into account when solving for synaptic weights that realize a desired mathematical function in a connection between ensembles. Furthermore, NengoBio can assign a spatial location to each neuron. Connections can be configured to select pre- and post-neuron pairs based on their spatial proximity and to respect maximum convergence and divergence numbers. Being able to quickly incorporate these constraints into network models facilitates exploring how high-level function could be realized in a specific brain circuit.

Discussion

We successfully mapped a high-level, mathematical function onto a brain microcircuit while incorporating biological constraints. This process was simplified by the ability of our modeling tool NengoBio to automatically account for Dale’s principle, spatial constraints, as well

as convergence numbers.

Our results show that the Granule-Golgi circuit could in principle implement a temporal basis function representation, which is in agreement with existing hypotheses about cerebellar function. Measurements from our model could be used to generate hypotheses about the kind of electrophysiological data we would expect to find, if this function was indeed realized in the brain. Having access to low-level biological parameters *in silico* furthermore facilitates the exploration of physiological changes that are difficult to achieve experimentally *in vivo*. As discussed above with respect to the synaptic time-constants and convergence numbers, this allows us to investigate why certain parameters are as observed.

A key difference of our approach to existing models of the Granule-Golgi circuit (such as Rössert et al., 2015), is that our modeling techniques are more general with respect to the high-level function that is being mapped onto the underlying circuit. Instead of relying on random connectivity, we directly specify the high-level function we would like the system to perform. We encourage cognitive modelers to view this particular model as an example; the techniques we present here are in principle compatible with all NEF models, including models of cognitive phenomena using the Semantic Pointer Architecture (SPA; Eliasmith, 2013).

While our model of eyeblink conditioning is concerned with a relatively low-level task, the techniques presented here for mapping function onto brain microcircuits are applicable to models of higher-level cognitive function as well, beyond what was already possible with the NEF and Nengo. 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). Future work will focus on incorporating additional biological detail into the model (such as separate biological time-constants for all synapses), as well as applying our techniques to more complex models.

The source code and software environment for our experiments is available online at <https://osf.io/fmpd2/>.

References

- Bekolay, T., Bergstra, J., Hunsberger, E., DeWolf, T., Stewart, T. C., Rasmussen, D., . . . Eliasmith, C. (2014). Nengo: A python tool for building large-scale functional brain models. *Frontiers in Neuroinformatics*, 7(48).
- Buckner, R. L. (2013). The Cerebellum and Cognitive Function: 25 Years of Insight from Anatomy and Neuroimaging. *Neuron*, 80(3).
- Chadderton, P., Margrie, T. W., & Häusser, M. (2004). Integration of quanta in cerebellar granule cells during sensory processing. *Nature*, 428(6985), 856–860.
- Dean, P., Porrill, J., Ekerot, C.-F., & Jörntell, H. (2010). The cerebellar microcircuit as an adaptive filter: Experimental and computational evidence. *Nature Reviews Neuroscience*, 11(1).
- de Jong, J., Voelker, A. R., van Rijn, H., Stewart, T. C., & Eliasmith, C. (2019). Flexible timing with delay networks – the scalar property and neural scaling. In *17th annual meeting of the international conference on cognitive modelling (iccm)*. Cognitive Science Society.
- 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).
- D'Angelo, E., Solinas, S., Mapelli, J., Gandolfi, D., Mapelli, L., & Prestori, F. (2013). The cerebellar Golgi cell and spatiotemporal organization of granular layer activity. *Frontiers in Neural Circuits*, 7, 93.
- 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. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford University Press.
- Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. MIT Press.
- Eliasmith, C., & Kolbeck, C. (2015). Marr's Attacks: On Reductionism and Vagueness. *TopiCS*,

1–13.

Fujita, M. (1982). Adaptive filter model of the cerebellum. , *45*(3), 195–206.

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).

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366.

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

Jakab, R. L., & Hámori, J. (1988). Quantitative morphology and synaptology of cerebellar glomeruli in the rat. , *179*(1), 81–88.

Johansson, F., Jirenhed, D.-A., Rasmussen, A., Zucca, R., & Hesslow, G. (2014). Memory trace and timing mechanism localized to cerebellar Purkinje cells. , *111*(41), 14930–14934. doi: 10.1073/pnas.1415371111

Kanichay, R. T., & Silver, R. A. (2008). Synaptic and cellular properties of the feedforward inhibitory circuit within the input layer of the cerebellar cortex. , *28*(36), 8955–8967.

Korbo, L., Andersen, B. B., Ladefoged, O., & Møller, A. (1993). Total numbers of various cell types in rat cerebellar cortex estimated using an unbiased stereological method. , *609*(1), 262–268.

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

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).

Marr, D., & Poggio, T. (1976). From Understanding Computation to Understanding Neural Circuitry. *MIT AI Memo*(357).

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).
- 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.
- Palkovits, M., Magyar, P., & Szentágothai, J. (1972). Quantitative histological analysis of the cerebellar cortex in the cat. IV. Mossy fiber-purkinje cell numerical transfer. , 45(1), 15–29. doi: 10.1016/0006-8993(72)90213-2
- Parisien, C., Anderson, C. H., & Eliasmith, C. (2008). Solving the problem of negative synaptic weights in cortical models. *Neural Computation*, 20, 1473–1494.
- 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).
- Stöckel, A. (2021a). Constructing dampened LTI systems generating polynomial bases. *arXiv preprint arXiv:2103.00051*. Retrieved from <https://arxiv.org/abs/2103.00051>
- Stöckel, A. (2021b). Discrete function bases and convolutional neural networks. *arXiv preprint arXiv:2103.05609*. Retrieved from <https://arxiv.org/abs/2103.05609>
- Stöckel, A., & Eliasmith, C. (2021). Passive nonlinear dendritic interactions as a computational resource in spiking neural networks. *Neural Computation*, 33(1), 96-128. doi: 10.1162/neco_a_01338
- Stöckel, A., Stewart, T. C., & Eliasmith, C. (2020a). A biologically plausible spiking neural model of eyeblink conditioning in the cerebellum. In *42nd annual meeting of the cognitive science society* (pp. 1614–1620). Cognitive Science Society.
- Stöckel, A., Stewart, T. C., & Eliasmith, C. (2020b). Connecting biological detail with neural computation: Application to the cerebellar granule-golgi microcircuit. In *18th annual meeting of the international conference on cognitive modelling* (pp. 277–282). Society for Mathematical Psychology.

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).

Voelker, A. R., Kajić, I., & Eliasmith, C. (2019). Legendre memory units: Continuous-time representation in recurrent neural networks. In *Advances in NeurIPS*.