Article:

Gilmer, JI, Farries, MA, Kilpatrick, Z et al. (2023) An emergent temporal basis set robustly supports cerebellar time-series learning. Journal of Neurophysiology, 129 (1). pp. 159-176. ISSN: 0022-3077

https://doi.org/10.1152/jn.00312.2022

1   **An emergent temporal basis set robustly supports cerebellar time-series learning**

2   Jesse I. Gilmer[1,2], Michael A. Farries[3], Zachary Kilpatrick[4], Ioannis Delis[5], Jeremy Cohen[6], and Abigail L.
3   Person[2]
4
5   1.   Neuroscience Graduate Program, University of Colorado School of Medicine, Aurora CO
6   2.   Dept. Physiology and Biophysics, University of Colorado School of Medicine, Aurora CO
7   3.   Knoebel Institute for Healthy Aging, University of Denver, Denver CO
8   4.   Department of Applied Mathematics, University of Colorado Boulder, Boulder CO
9   5.   School of Biomedical Sciences, University of Leeds, Leeds UK
10   6.   University of North Carolina Neuroscience Center, Chapel Hill NC

11   **ABSTRACT**
12   The cerebellum is considered a 'learning machine' essential for time interval estimation underlying motor
13   coordination and other behaviors. Theoretical work has proposed that the cerebellum's input recipient
14   structure, the granule cell layer (GCL), performs pattern separation of inputs that facilitates learning in
15   Purkinje cells (P-cells). However, the relationship between input reformatting and learning has remained
16   debated, with roles emphasized for pattern separation features from sparsification to decorrelation. We
17   took a novel approach by training a minimalist model of the cerebellar cortex to learn complex time-series
18   data from time varying inputs, typical during movements. The model robustly produced temporal basis
19   sets from these inputs, and the resultant GCL output supported better learning of temporally complex
20   target functions than mossy fibers alone. Learning was optimized at intermediate threshold levels,
21   supporting relatively dense granule cell activity, yet the key statistical features in GCL population activity
22   that drove learning differed from those seen previously for classification tasks. These findings advance
23   testable hypotheses for mechanisms of temporal basis set formation and predict that moderately dense
24   population activity optimizes learning. NEW AND NOTEWORTHY: During movement, mossy fiber
25   inputs to the cerebellum relay time-varying information with strong intrinsic relationships to ongoing
26   movement. Are such mossy fibers signals sufficient to support Purkinje signals and learning? In a model,
27   we show how the GCL greatly improves Purkinje learning of complex, temporally dynamic signals
28   relative to mossy fibers alone. Learning-optimized GCL population activity was moderately dense, which
29   retained intrinsic input variance while also performing pattern separation.
30

31   **INTRODUCTION**
32   The cerebellum refines movement and maintains calibrated sensorimotor transformations by learning to
33   predict outcomes of behaviors through error-based feedback (Ito, 1972; Herzfeld et al., 2015; Medina
34   2000; Mauk and Buonomano, 2004; Raymond et al., 1996). A major site of cerebellar learning is in the
35   cerebellar cortex, where Purkinje cells (P-cells) receive sensorimotor information from parallel fibers
36   (Huang et al. 2013) whose synaptic strengths are modified by the conjunction of presynaptic (parallel
37   fiber) activity and climbing fiber inputs to P-cells thought to convey instructive feedback (McCormick et
38   al., 1982; Yang and Lisberger, 2014; Mauk et al., 1986; De Zeeuw et al., 1998). P-cell activity is
39   characterized by rich temporal dynamics during movements, representing putative computations of
40   internal models of the body and the physics of the environment (Wolpert et al., 1998; Shadmehr and
41   Mussa-Ivaldi 1994). Parallel fibers are the axons of cerebellar granule cells (GCs), a huge neuronal
42   population (comprising roughly half of the neurons in the entire brain; Herculano-Houzel 2010), which

43 are the major recipient of extrinsic inputs to the cerebellum. Thus, understanding the output of the GCL is
44 key in determining the encoding capacity and information load of incoming activity projected to the
45 cerebellum. Inputs to GCs arise from mossy fibers (MFs), which convey sensorimotor information for P-
46 cell computations (Rancz et al., 2007; Ishikawa et al., 2015). There are massively more GCs than MFs
47 and each GC typically receives input from just 4 MFs (Palkovits et al., 1971), such that the information
48 carried by each MF is spread among many GCs, but each GC samples from only a tiny fraction of total
49 MFs (Jakab and Hamori 1988; Eccles et al., 1967).
50
51 The GCL has been the focus of theoretical work spanning decades, which has explored the computational
52 advantages of the unique feedforward architecture of the structure. Notably, early studies of the cerebellar
53 circuit by Marr (1969) and Albus (1971) proposed that a key component of the cerebellar algorithm is the
54 sparse representation of MF inputs by GCs. In this view, the cerebellum often must discriminate between
55 overlapping, highly correlated patterns of MF activity with only subtle differences distinguishing them
56 (Bengsston and Jorntell 2009). Sparse recoding of MF activity in a much larger population of GCs
57 ("expansion recoding") increases the dimensionality of population representation and transforms
58 correlated MF activity into independent activity patterns among a subset of GCs (Litwin-Kumar et al.,
59 2017; Cayco-Gajic et al., 2017; Gilmer and Person 2018). These decorrelated activity patterns are easier
60 to distinguish by learning algorithms operating in P-cells, leading to better associative learning and credit
61 assignment (Cayco-Gajic et al., 2017; Sanger et al., 2020).
62
63 The machine learning perspective of the Marr-Albus theory tends to assume that the cerebellum is
64 presented with a series of static input patterns that must be distinguished and categorized. However,
65 during movements, neuronal population dynamics are rarely, if ever, static. Mauk and Buonomano (2004)
66 revisited cerebellar expansion recoding in the context of temporal encoding, a necessary computation for
67 the cerebellar-dependent task of delay eyelid conditioning. They proposed that a static activity pattern in
68 MFs could be recoded in the GC layer as a temporally evolving set of distinct activity patterns, termed a
69 temporal basis set. P-cells could learn to recognize the GC activity pattern present at the correct delay and
70 initiate an eyeblink to avert the "error" signal representing the air puff to the eye. This transformative
71 theory has given rise to an emerging literature exploring mechanisms of basis set formation. A variety of
72 mechanisms have been proposed for how such time-varying population activity might emerge, including
73 local inhibition, short-term synaptic plasticity, diverse unipolar brush cell properties and varying GC
74 excitability (Chabrol et al., 2015; Duguid et al, 2012; Crowley et al., 2009; Rudolph et al., 2015;
75 Buonomano and Mauk 1994; Kanichay and Silver 2008; Simat et al., 2007; Mapelli et al., 2009; Rossi et
76 al., 1996; Gall et al., 2005; Armano et al., 2000; Rizwan et al. 2016; Tabuchi et al., 2019; D'Angelo and
77 De Zeeuw 2009; Kennedy et al., 2014, Guo et al., 2021; Dino et al. 2000).
78
79 Despite these promising avenues, the problem of learning more complex movements presents a distinct
80 set of questions about how the cerebellum processes and uses time variant inputs to learn complex P-cell
81 signals, a type of timeseries. Therefore, to test how expansion recoding of time-varying input contributes
82 to learning, we used a simple model of the GCL and a time-series prediction task to explore the effect of
83 putative GCL filtering mechanisms on expansion recoding and learning. Similar to previous models, this
84 simplified model made GC activity sparser relative to MF inputs (Marr 1969; Albus 1971) and increased
85 the dimensionality of the input activity (Litwin-Kumar et al., 2017) while preserving information
86 (Billings et al., 2014). The model greatly enhanced learning accuracy and speed by P-cells on a difficult

87 time series prediction task when compared to MF inputs alone. Together, these results suggest that the
88 cerebellar GCL provides a rich basis for learning in downstream Purkinje cells, providing a mixture of
89 lossless representation (Billings et al., 2014) and enhanced spatiotemporal representation (Litwin-Kumar
90 et al. 2017) that are selected for by associative learning to support the learning of diverse outputs that
91 support adaptive outputs in a variety of tasks (Fujita 1982; Dean and Porrill 2008).
92
93 **METHODS**
94 **Model construction**
95 The model presented here incorporated only the dominant features of the granule cell layer (GCL) circuit
96 anatomical organization and physiology. The features chosen for the model were the sparse sampling of
97 inputs (GCs have just 4 synaptic input branches in their segregated dendrite complexes on average),
98 which was reflected in the connectivity matrix between the input pool and the GCs, where each GC
99 received 4 inputs with weights of ¼4th (i.e. 1 divided by the number of inputs; 1/M) of the original input
100 strength, summing to a total weight of 1 across all inputs. The other features were thresholding,
101 representing inhibition from local inhibitory Golgi neurons and intrinsic excitability of the GCs. The
102 degree of inhibition and intrinsic excitability (threshold) was a free parameter of the model, and the
103 dynamics were normalized to the z-score of the summated inputs. This feature reflects the monitoring of
104 inputs by Golgi cells while maintaining simplicity in their mean output to GCs. While this model
105 simplifies many aspects of previous models of the GCL, it recreated many of the important features of
106 those models, suggesting that the sparse sampling and firing are the main components dictating GCL
107 functionality.
108 The model, in total, uses the following formulas to determine GC output:
109

$$\text{Eq 1: } GC_i(t) \ = [(\sum_{k_1}^{k_M} \frac{MF_k(t)}{M}) -- \theta \ ]_+$$

110

111
112 where k is a random selection of M MFs from the MF population. The inputs are summed and divided by
113 the total number of MF inputs to the GC, M, so that their total weight is equal to 1. Unless noted as a
114 variable, we used M = 4, reflecting the mean connectivity between MFs and GCs, and the optimal ratio
115 for expansion recoding (Litwin-Kumar et al. 2017), and the point of best input variance retention (Fig. 5).
116 This function is then linearly rectified, i.e. $[x]_+ = x$ if $x > 0$ and 0 otherwise so that there are no negative
117 rates present in the GC activity. The θ function which determines the threshold, estimating the effects of
118 intrinsic excitability and feedforward inhibition, was formulated as:
119

$$\text{Eq 2: } \theta \ = \ \overline{MF} \ + \ (z \ * \ \sigma(MF))$$

120

121
122 Here, z sets the number of standard deviations from the MF mean. z is the only free parameter,
123 which determines the minimum value below which granule cell activity is suppressed. Therefore
124 we report z as the 'threshold'. Note that the summated MF inputs are divided by the number of
125 inputs per GC (M) in Eq. 1 such that their received activity relative to θ is proportional to the
126 input size, M. Since the input to GCs is Gaussian in our model, the summed activity integrated by the
127 GCs is Gaussian as well. For that reason, we found it convenient to define the GC thresholding term in
128 terms of a z-score. Thus, a GC with a threshold of "zero" has its threshold set at the mean value of its MF
129 inputs; such a GC would be silent 50% of the time on average because the Gaussian presynaptic input

130 would be below the mean value half the time. This makes it possible to discuss functionally similar
131 thresholds across varying network architectures (e.g., a GC with a threshold of zero would discard half of
132 its input on average regardless of whether it received 2 or 8 MF inputs).
133

134 **OU input construction**
135 To provide a range of inputs with physiological-like temporal properties that could be parameterized, we
136 used a class of randomly generated signals called Ornstein-Uhlenbeck Processes (OU), defined by the
137 following formula:

138 $$Eq\ 3: OU(t) = \left(OU(t - \Delta t) * e^{\left(-\frac{\Delta t}{\tau}\right)}\right) + \left(\sigma * \sqrt{1 - e^{-2*\frac{\Delta t}{\tau}}} * R\right)$$

139

140 Here t is the time point being calculated, Δt is the time interval (the time base is in *ms* and Δt is 1 *ms*). σ is
141 the predetermined standard deviation of the signal, and R is a vector of normally distributed random
142 numbers. This process balances a decay term, the exponential with e raised to $-\Delta t/\tau$, and an additive term
143 which introduces random fluctuations. Without the additive term, this function decays to zero as time
144 progresses. For all simulations, unless noted otherwise, tau was 100 ms. This resulted in a mean
145 autocorrelation tau of 502+/- 52 ms, which was intermediate between pontine neurons and reach-related
146 electromyograms autocorrelation taus of 351+/-120 ms and 567 +/- 151 ms, used below as model inputs,
147 respectively. After the complete function has been calculated, the desired mean is added to the timeseries
148 to set the mean to a predetermined value.
149

150 The vector R can also be drawn from a matrix of correlated numbers, as was the case in Fig. 7 – figure
151 supplement 1 B & C. These numbers were produced with the MATLAB functions randn() for normal
152 random numbers, and mvnrnd() for matrices with a predetermined covariance matrix supplied to the
153 function. The covariance matrix used for these experiments was always a 1-diagonal with a constant,
154 predetermined, covariance value on the off-diagonal coordinates.
155

156 **Introduction of noise to input and GCL population**
157 To test whether fluctuations riding on input signals influence GCL basis set formation, we introduced
158 Gaussian noise that was re-calculated trial to trial and added it to the MF input population. The amplitude
159 of the introduced noise was scaled to the amplitude of the input, so that the proportion of the signal that is
160 noise could be described with a percentage: % Noise = 100 * Noise Amp. / (Signal Amp. + noise Amp.).
161 For example, if the amplitude of the noise was equal to the amplitude of the input, the % noise would be
162 equivalent to 1 / (1 + 1) = .5, or 50% noise.
163

164 To determine the stability of representations in the MF and GCL populations with introduced noise, we
165 measured the displacement of the temporal location where peak firing occurred between noiseless and
166 noisy activity patterns at threshold 0 (unless noted). This measurement was rectified to obtain the absolute
167 displacement of peak firing time.
168

169 **Learning accuracy and speed assay**
170 To understand how the GCL contributed to learning, we constructed an artificial Purkinje cell (P-cell)
171 layer. The P-cell unit learned to predict a target function through a gradient descent mechanism, such that
172 the change in weight for each step was:

173

174 $$Eq\ 4:\ Err(t) = |P(t) - TF(t)|$$

175

176 $$Eq\ 5:\ \Delta W_i = W_i - (Err(t) * GC_i(t) * \eta)$$

177

178 Where P(t) is the output of the P-cell at time t, TF(t) is the target function at time t, W_i is the weight
179 between the Purkinje cell and the ith GC, and η is a small scalar termed the 'step size'. η was 1E-3 for
180 GCs, and 1E-5 for MF alone in simulations shown in this study where the step size was held fixed, which
181 was chosen to maximize learning accuracy and stability of learning for both populations. While not
182 strictly physiological because of membrane time constant temporal filtering and variable eligibility
183 windows for plasticity, this form of learning is widely applied in neural models, including cerebellar (e.g.
184 Bouvier et al., 2018). Physiological equivalents of negative weights found by gradient descent could be
185 achieved by molecular layer interneuron feedforward inhibition to P-cells. The learning process in Eq. 4
186 and 5 was repeated for T trials at every time point in the desired signal. The number of trials was chosen
187 so that learning reached asymptotic change across subsequent trials. Typically, 1000 trials were more than
188 sufficient to reach asymptote, so that value was used for the experiments in this study.

189

190 The overall accuracy of this process was determined by calculating the mean squared error between the
191 predicted and desired function:

192

193 $$Eq\ 6:\ MSE\ = \frac{1}{T}\sum_{t=1}^{T}(P(t) - TF(t))^2$$

194

195 The learning speed was determined by fitting an exponential decay function to the MSE across every trial
196 and taking the tau of the decay (See methods: Model output metrics, Time decay). A percentage form of
197 this measure is given in Fig. 3B, bottom, to give intuition to the total decrease in MSE when comparing
198 MFs to GCL output. %MSE = $(MSE_{MF} - MSE_{GCL})/MSE_{MF}$.

199

200 **GCL output metrics**
201 To assay the properties of the GCL output that influence learning, we measured the features of GCL
202 output across a spectrum of metrics that have theoretically been associated with GCL functions like
203 pattern separation or expansion, as well as optimization or cost-related metrics developed for this paper.
204 These included: dimensionality, spatiotemporal sparseness, contributing principal components, spatial
205 sparseness (mean population pairwise correlation), temporal sparseness (mean unit autocovariance
206 exponential decay), population variance, temporal lossiness, population lossiness, and temporal cover.

207

208 We considered three forms of lossiness here, two related to the dimensions of sparseness considered
209 above, time and space, and one that is a measure of sparseness on the individual GC level. Temporal
210 lossiness is a measure of the percentage of time points that are not encoded by any members of the GCL
211 population, essentially removing the ability of P-cells to learn at that time point and producing no output
212 at that time in the final estimation of the target function. Increases in the value are guaranteed to degrade
213 prediction accuracy for any target function that does not already contain a zero value at the lossy time
214 point.

215 Eq 7:

$$Temp.\,Lossiness = \frac{1}{T}\sum_{t=1}^{T} x_t \text{ where } x_t \begin{cases} (\sum_{i=1}^{N} GC_i(t)) \leq 0 = 1 \\ \\ else = 0 \end{cases}$$

216
217 Here, T is the total number of points in the encoding epoch, the bracketed portion of the formula is a
218 summation of inputs from all GCs (N = population size) at that timepoint. When all GCs are silent, the
219 sum is 0, and the temporal lossiness is calculated as 1, and when all time points are covered by at least
220 one GC, total temporal lossiness is 0.
221
222 Spatial lossiness, or population lossiness, is the proportion of GCs in the population that are silent for the
223 entirety of the measured epoch. This is thought to reduce total encoding space and deprive downstream P-
224 cells of potential information channels and could potentially impact learning efficacy. It is defined as:
225
226   Eq 8:

$$Pop.\,Lossiness = \frac{1}{N}\sum_{i=1}^{N} x_i \text{ where } x_i \begin{cases} (\sum_{t=1}^{T} GC_t) \leq 0 = 1 \\ \\ else = 0 \end{cases}$$

227
228 Here, N is the total population size of the GCL, and the bracketed portion of the formula is a sum of the
229 activity of GCs across all timepoints, such that if a GC is silent across all timepoints x_i is calculated as 1,
230 indicating the 'loss' of that GC unit's contribution. When all GCs are silent, population lossiness is 1, and
231 when all GCs are active for at least one time point, population lossiness is 0.
232
233 Additionally, we looked at the mean sparseness of activity across the population by measuring the
234 'coverage' or proportion of time points each GC was active during, defined as:
235   Eq 9:

$$Coverage = \frac{1}{N}\sum_{i=1}^{N} (\frac{1}{T}\sum_{t=1}^{T} x_i \text{ where } x_i \begin{cases} GC_i(t) > 0 = 1 \\ else = 0 \end{cases})$$

236
237 As before, N is the number of cells in the population and T is the total length of the epoch. The bracketed
238 function counts the number of time points where GC_i is active, and divides that by the total time period
239 length to get the proportion of time active. This value is summed across all GCs and divided by N to
240 calculate the average coverage in the population. This value has strong synonymy with population
241 variance, so it was not used for fitting assays in later experiments (Fig. 6), but reflects the effect of
242 thresholding on average activity in the GCL population.
243
244 Dimensionality is a measure of the number of independent dimensions needed to describe a set of signals,
245 similar in concept to the principal components of a set of signals. This measure is primarily influenced by
246 covariance between signals, and when dimensionality approaches the number of signals included in the
247 calculation (n), the signals become progressively independent. The GCL has previously been shown to
248 enhance the dimensionality of input sets and does so in the model presented here too. Dimensionality is
249 calculated with:

$$\text{Eq 10: } Dim = \left(\sum_{i=1}^{n} \lambda_i\right)^2 / \left(\sum_{i=1}^{n} \lambda_i{}^2\right)$$

251

252 Provided by Litwin-Kumar, et al, 2016. This is the ratio of the squared sum of the eigenvalues to the sum
253 of the squared eigenvalues of the covariance matrix of the signals.

254

255 Spatiotemporal Sparseness (STS) was a calculated cost function meant to measure the divergence of GC
256 population encoding from a 'perfect' diagonal function where each GC represents one point in time and
257 does not overlap in representation with other units. This form of representation is guaranteed to produce
258 perfect learning, and transformations between the diagonal and any target function can be achieved in a
259 single learning step, making this form of representation an intriguing form of GCL representation, if it is
260 indeed feasible. We calculated the cost as:

261

262 $$\text{Eq 11: } STS = (1 - L_t) * \left(\frac{1}{T}\right) * \left(\frac{W}{GC_w}\right)$$

263

264 Where $(1 - L_t)$ is the cost of temporal lossiness, defined above (Eq. 7), and T is the total length of the
265 epoch. W is the number of unique combinations (termed 'words', akin to a barcode of activity across the
266 population), of GCs across the epoch at each point of discrete time, and GC_w is the average number of
267 words each GC is active at all within the time-bins chosen (e.g. a binary representation of GC activity).
268 The intuition used here is that when there is no temporal lossiness, all points in time are represented,
269 leading the $1 - L_t$ term to have no effect on the STS equation, and when W, the number of unique
270 combinations of GC activities is equal to T, then each point in time has a unique 'word' associated with it.
271 Finally, when GC_w is 1, W/GC_w is equal to W, which only occurs when each GC contributes to a single
272 word. When these conditions are met, STS = 1, otherwise when GCs contribute to more than one word,
273 GC_w increases and W is divided by a number larger than 1, decreasing STS. Alternately, when there are
274 not many unique combinations, such as when every GC has the exact same output, W/GC_w is equal to
275 (1/T), decreasing STS. Finally, because lossiness causes the occurrence of a 'special', but non-associable,
276 word, we multiplied the above calculations by $(1 - L_t)$ to account for the effect of the unique non-encoding
277 word (i.e. all GCs inactive) on distance from the ideal diagonal matrix.

278

279 Mean temporal decay, i.e. temporal sparseness, is a measure of variance across time for individual
280 signals, where a low value would indicate that the signals coherence across time is weak, meaning that the
281 signal varies quickly, whereas a high value would mean that trends in the signal persist for long periods of
282 time. This value is extracted by fitting an exponential decay function to the autocovariance of each unit's
283 signal and measuring the tau of decay in the function:

284

285 $$\text{Eq 12: } y = a * e^{(-x/\tau)}$$

286

287 This is converted to the ms form by taking the ratio of $1000/\tau$. y here τ is a description of the
288 autocovariance of the activity of a MF or GC signal, so when the descriptor τ is a large number, the decay
289 in autocovariance is longer, or slower, when τ is a small number, the autocovariance across time decays
290 more quickly, making the change in activity faster.

291

292 While dimensionality and STS are metrics rooted in a principled understanding of potentially desirable
293 properties of population encoding, the gradient descent algorithm can extract utility from population
294 statistics that are much noisier and correlated than the ideal populations that dimensionality and STS
295 account for. To measure a more general pattern separation feature in GCL output that could still be
296 associated with the complex target function, we turned to principal component analysis (PCA) with the
297 intuition that components which explain variance in the GCL output could be utilized by the downstream
298 Purkinje cell units to extract useful features from the input they receive (Lanore et al., 2021). We
299 parameterized the utility of this measure by taking the proportion of the PCs derived from the GCL output
300 which explained variance (of the GCL output) in that population by more than or equal to 1/N, where N is
301 the number of GCs, suggesting that they explain more variance than would be expected from chance.

303 Population correlation, was measured by taking the mean correlation between all pairwise combinations
304 of GCs using the corr() function in MATLAB and excluding the diagonal and top half of the resultant
305 matrix.

307 Population aggregate variance is a measure related to the expansion or collapse of total space covered by
308 the encoding done by a population, and higher or expanded values in this metric are thought to assist in
309 pattern separation and classification learning.

311 $$\text{Eq 13: Pop. Var} = \sum_{n=1}^{N}(x_n - \boldsymbol{\mu})^2$$

313 As shown in Cayco-Gajic et al. (2017). Here x is the activity of one of n cells across a measured epoch,
314 and μ is the mean of that activity. This value is reported relative to the number of GC units, such that Pop.
315 Var reported in Fig. 6 is normalized to Pop. Var / N.

318 **Variance retained assay**
319 To test the recovery of inputs by a feedforward network with a granule cell layer (GCL), we used
320 explained variance, R^2 , to quantify the quality of recovery of a sequence of normal random variables
321 (Fig. 5) across $N_w = 1000$ numerical experiments. To distinguish this metric from the MSE and R^2
322 metrics to evaluate other models in the study, we rename this 'variance retained'. Within each numerical
323 experiment i, at each time point, a vector of inputs $\boldsymbol{x_t}$ of length M (representing the mossy fiber, MF,
324 inputs) was drawn from an M-dimensional normal distribution with no correlations, $\boldsymbol{x_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I_M})$. This
325 vector is then left-multiplied by a random binary matrix W with N rows and M columns with n 1's per
326 row and the rest zeros, followed by a threshold linearization to obtain the GCL output, $\boldsymbol{y_t} = [W\boldsymbol{x_t} - \boldsymbol{z}]_+$
327 with threshold. This process is then repeated $T = 1000$ times and a downstream linear readout was fit to
328 optimally recover $\boldsymbol{x_t}$ from $\boldsymbol{y_t}$. It can be shown multivariate linear regression (MATLAB's regress()
329 function, employing least squares to minimize mean squared error) solves this problem, identifying for
330 each MF input stream $\boldsymbol{x_{1:T}^j}$, the optimal weighting $B_{1:T}$ from the GCL to estimate $\hat{\boldsymbol{x}}_{1:T}^j = B_{j,1:N}\boldsymbol{y_{1:T}}$.
331 Across time $t = 1:T$, we then computed the squared error across the vector, $MSE_i = \sum_{t=1}^{T}\sum_{j=1}^{M}(\hat{x}_t^j -$
332 $x_t^j)^2$, as well as the summed variance of the actual input, $Var_i = \frac{1}{MT}\sum_{j=1}^{M}\sum_{t=1}^{T}(x_t^j - \bar{x}^j)^2$, where
333 $\bar{x}^j = \frac{1}{T}\sum_{t=1}^{T}x_t^j$ is the mean of the jth MF input stream. Lastly, to compute variance explained, we take

334    $R^2 = 1 - \frac{\sum_{i=1}^{N_w} MSE_i}{\sum_{i=1}^{N_w} Var_i}$, so the higher the relative mean squared error is, the lower the variance explained will

335   be. To generate the panels in Fig. 5, we always kept the number of timepoints and experiments the same,

336   but varied (Fig. 5B) the threshold along the axis and the number of inputs n per GC output; (Fig. 5C) the

337   total number of GC outputs N and input per output n; (Fig. 5D) number of inputs M and outputs N; and

338   finally (Fig. 5E) the number of inputs per GC output n along with the total number of outputs N.

339

340   **Generation of GCL output with defined statistical structure**

341   To determine if the sparseness measures had inherent benefits for learning, we supplemented the GCL

342   output with OU processes with known temporal and correlational properties to examine their effect on

343   learning accuracy (Figure 7 figure supplement 1). We varied the temporal properties by systematically

344   varying the tau value in the exponential decay function. To vary population correlation, the random draw

345   function in the OU process was replaced with a MATLAB function, mvnrnd(), which allowed for preset

346   covariance values to direct the overall covariance between random samples. We used a square matrix with

347   1s on the diagonal and the desired covariance on all off-diagonal locations for this process and varied the

348   covariance to alter the correlation between signals. The OU outputs from this controlled process were

349   then fed into model P cells with randomized OU targets, as per the normal learning condition described

350   above. To vary the effect of the input population size, the size of the supplemented population varied

351   from 10 to 3000 using a tau of 10 and drawing from normal random numbers.

352

353   To measure the effects of STS on learning, a diagonal matrix was used at the input to a Purkinje unit,

354   which represented population activity with an STS of 1 (see Eq 11 in Model output metrics). To degrade

355   the STS metric, additional overlapping activity was injected either by expanding temporal representation

356   or at random, for example, adding an additional point of activity causes inherent overlap in the diagonal

357   matrix, increasing the GC_w denominator of Eq 11 to (1 + 2/N) because the overlapping and overlapped

358   units now each contribute to 1 additional neural word. This process was varied by increasing the amount

359   of overlap to sample STS from 0 to 1.

360

361   **Statistics of GCL output metrics and learning**

362   To estimate the properties of GCL output that contribute to enhanced learning of time series, we used

363   multiple linear regression to find the fit between measures of GCL population activity and observed MSE

364   in learning. Because there are large inherent correlations between the metrics used (dimensionality,

365   spatiotemporal sparseness, explanatory principal components of the GC population, population

366   variability, mean pairwise GC correlation, temporal sparseness, temporal lossiness, population lossiness,

367   and input variance retained) we used two linear regression normalization techniques: LASSO and RIDGE

368   regression. For Figure 7, LASSO was used to isolate the 'top' regressors, while RIDGE was used in

369   Figure 8 to preserve small contributions from regressors. The RIDGE regression method was then used to

370   compare resultant regression slopes (beta coefficients) to changes in task parameters (see Methods on

371   Simulation of cerebellar tasks).

372

373   Regressions were performed using the fitrlinear() function in MATLAB, with LASSO selected by using

374   the 'SpaRSA' (Sparse Reconstruction by Separable Approximation; Wright et al., 2009) solver, and

375   RIDGE selected with the 'lbfgs' (Limited-memory BFGS; Nodecal and Wright 2006) solver techniques.

376   The potential spread of MSE in the models was determined using a K-fold validation technique, with 10

377 'folds' used, as well as for determining the range of absolute slopes shown in Figure 8C, of which the
378 mean and standard deviation of cross-validation trials are plotted with solid lines and shaded polygons,
379 respectively. Models were selected by choosing the model with the least complex fitting parameters (i.e.
380 the model with the highest Lamba) while still falling within the bounds of the model with the minimized
381 MSE plus the standard error (a standard '1SE' method).
382
383 We reasoned that interactions between explanatory GCL statistical features might account for observed
384 learning accuracy to some degree. A standard method for selecting potential interactions while
385 constraining the regression model to a reasonable number of parameters is through selection by Bayesian
386 information criteria (BIC) stepwise regression. We used the MATLAB stepwiselm() function with the
387 BIC method to select from our 9 statistical features and allowed the regression function to select potential
388 interactions between them. The output of the regression listed which linear and interacting components
389 best fit the model. Although this output also included the Beta values of the fits, they were not regularized
390 in a way that was intuitively interpretable, so we therefore transferred the BIC selected parameters to a
391 RIDGE regressor to get the final Beta values and fit.
392
393 To convey the overall contribution of regressors to the above models of MSE, the slope relative to the
394 magnitude of all slopes were used as plotted metrics (Fig. 8C).
395
396 **Pontine neuron activity patterns**
397 To investigate the properties of GCL filtering on physiological inputs to the cerebellar cortex, we
398 extracted recordings of pontine neurons, a primary source of mossy fibers, from the work of Guo,
399 Sauerbrei and colleagues (Guo et al., 2021b) during a reaching task in mice. We used the first 50 neurons
400 for the recording to keep MF counts similar to the modelled OU population, and applied a 100 ms
401 Gaussian filter to the raw spiking data, aligned to reach onset, to obtain the estimated firing rate. The
402 firing rate values were range normalized for display and filtering (Fig. 1B,E) and are shown in order of
403 their peak firing rate time.
404
405 **Simulation of cerebellar tasks**
406 To simulate the transformation between motor commands and kinematic predictions, we used human
407 EMG as a proxy for a motor command-like input signal to the GCL. 30 muscles from 15 bilateral target
408 muscles were used (Delis et al., 2018; Hilt et al., 2018). The target function was a kinematic trajectory
409 recorded simultaneously with the recordings of EMG used for the study. Although many body parts and
410 coordinate dimensions were recorded of the kinematics, we opted to use the kinematic signal with the
411 largest variance to simplify the experiment to a single target function.
412
413 **Code Availability**
414 All computer code and simulation data is freely available at https://github.com/jesse-
415 gilmer/2022-GCL-Paper. Supplementary Figures are available at 10.6084/m9.figshare.20361849
416
417 **RESULTS**
418 **Temporal basis set formation as emergent property of GCL filtering of time-varying inputs**

419   In many motor tasks, both mossy fibers and P-cells show highly temporally dynamic activity patterns,
420   raising the question of how GCL output supports timeseries learning using time varying inputs, a
421   divergence from traditional classification tasks used in most cerebellar models (Fig. 1; Izawa et al. 2012).

423   We used a simple model, similar to previously published architectures (Cayco-Gajic et al., 2017; Litwin-
424   Kumar et al., 2017; Billings et al., 2014), capturing the dominant circuit features of the GCL: sparse
425   sampling of mossy fiber (MF) inputs by postsynaptic granule cells (GCs) and coincidence detection
426   regulated by cellular excitability and local feedforward inhibition (Fig. 1A; Eq.1,2; Marr 1969; Albus
427   1971; Palkovits et al., 1971; Chabrol et al., 2015). GC output is generated by summing MF inputs and
428   thresholding the resultant sum; anything below threshold is set to zero while suprathreshold summed
429   activity is passed on as GC output (Fig. 1A, center). The GC threshold level represents both intrinsic
430   excitability and the effect of local inhibition on regulating GC activity.

432   We fed two naturalistic sources of cerebellar inputs to the model: recordings from the mouse pontine
433   nucleus (PN, Fig. 1B, previously published in Guo, et al., 2021b) and electromyograms measured during
434   reaching tasks (EMG, Fig. 1C, from Delis et al. 2018). In both cases, the GCL enhanced the
435   spatiotemporal representation of input activity. To parameterize such time-varying inputs, we next
436   generated artificial MF activity using Ornstein-Uhlenbeck (OU) stochastic processes. These signals
437   provide a statistically tractable ensemble that was rich enough to capture the dynamic nature of
438   naturalistic inputs while remaining analytically tractable and easily parameterized, fully characterized by
439   just three parameters: correlation time, mean, and standard deviation. Example OU input functions are
440   shown in Fig. 1D (top). Importantly, OU functions preserve autocorrelations typical of physiological
441   signals, such that they are not random from moment-to-moment (Fig. 1D, tau of 100 ms). All OU MFs
442   had the same tau and were not correlated with one another. As with the naturalistic inputs, the model GCL
443   spatiotemporally diversified OU processes Fig. 1D (explored more thoroughly below). The emergence of
444   sparse spatiotemporal representation under the simplistic constraints of the model suggests that the
445   cerebellum's intrinsic circuitry is sufficient to produce spatiotemporal separation when given sufficiently
446   time-varying inputs. Below, we refer to the transformation of information between GCL inputs and
447   outputs as "GCL filtering".

449   **GCL temporal basis is robust to noise**
450   By relying on coincident peaks in time-varying mossy fibers, this mechanism of spatiotemporal
451   sparsening raised the question of whether such temporal basis sets were robust to noise. To address
452   whether noise degrades spatiotemporal representation, we ran repeated simulations, adding Gaussian
453   noise that changed from trial to trial to fixed OU functions, and compared the resultant GCL basis sets
454   (Fig. 2). We modeled trial-over-trial noise variance by superimposing a Gaussian fluctuation such that the
455   overall proportion of the total signal that was noise ranged from 25%-50%.

457   GCL population activity was generally stable across noise levels (Fig. 2B). To quantify stability, we
458   measured the shift in the time of peak rate for each GC over 100 trials at threshold of 0. 50% of granule
459   cells shifted 10 ms or less in the 25% noise condition (Fig. 2C, left) and 50% shifted less than 30 ms
460   when 50% of the signal was unstable noise (Fig. 2C, right). Thus, while the basis set structure is not
461   perfectly resistant to noise, the primary temporally correlated OU signal dominates the population's
462   temporal structure. The effect of high noise on the stability of the temporal basis was dependent on

463   threshold: higher thresholds coupled with higher noise degraded temporal stability. At a threshold of 0,
464   the mean time shift was 136 ms. While at a threshold of 1, the mean time shift was 305 ms.
465
466   **GCL improves time series learning accuracy**
467   If mossy fiber activity is naturally time-varying it raises the question of whether it, by itself, is
468   intrinsically suited to support timeseries learning, obviating a role for the GCL (Markanday et al., 2022;
469   Fig S1). To address this question, we tested whether GCL population activity assisted learning beyond
470   the temporal representations inherent in the mossy fibers. We devised a task where P-cells learned to
471   generate a specific time-varying signals (OU process with 10 ms autocorrelation time) using gradient
472   descent (Equations 4-5, Methods). Inputs to P-cells were either MFs or GCL populations. Initially, P-cell
473   output was distinct from the target function, but over repeated trials P-cell output converged towards the
474   target function (Fig. 4A). We quantified the convergence of the P-cell output to the target function and
475   compared performance between instances when MF activity was sent directly to P-cells ("MFs alone")
476   versus GCL activity was used as P-cell input. Finally, we examined performance of these learning
477   simulations across different thresholds, expressed in terms of a z-score, such that a threshold of "zero"
478   indicates the threshold is at the mean of MF input.
479
480   The model achieved excellent learning with either MFs or GCL inputs. Notably, the GCL markedly
481   enhanced the convergence to a target function at thresholds between –1 and 1 (Fig. 3A), achieving a
482   minimized mean squared error (MSE) of roughly 0.005, outperforming learning using MFs alone (MSE
483   0.02; normalized to a range of [0,1]). To establish an intuition into the practical difference of the range of
484   MSEs achieved with GCL or MFs alone, we tasked the model with learning a timeseries which could be
485   rendered as a recognizable image to human viewers (Fig 3B). This function had an identical range of
486   target function values ([0,1], Fig. 3B). GCL inputs facilitated P-cell timeseries learning that recapitulated
487   the recognizable image (Fig. 3B, bottom; MSE 0.002). By contrast P-cells that received MFs alone
488   generated a timeseries that rendered an unrecognizable image, despite the seemingly excellent MSE of
489   0.02. Thus, the small errors of MF-driven output accumulated along the timeseries to degrade
490   performance, while GCL-driven P-cell output yielded an easily recognizable image (Fig. 3B top right vs
491   three thresholds, bottom). Importantly, this was not a consequence of the large population expansion
492   between MFs and GCs, as increasing the number of MFs alone did not improve performance to the levels
493   observed in the model GCL (Fig. S2A-B). Nevertheless, a sufficiently large GCL population is required
494   to improve learning (Fig. S2B).
495
496   **GCL model speeds time series learning**
497   Having found that the GCL improves the match between predicted output and target output over a range
498   of thresholds, we next examined whether the GCL also increased the speed of convergence. We examined
499   the MSE between the model output and the target function on each trial as training progresses (Fig. 4C,
500   *red circles*) and found that output usually converged rapidly at first then more slowly in later stages of
501   training (Fig. 4A). The reduction in MSE over training in our model was reasonably well fit by a double
502   exponential (Fig. 4B, *red curve*), of the form
503

$$( ) = \ _{1}{}^{(- \ _{1}\ )} + \ _{2}{}^{(- \ _{2}\ )} + $$

504

505 where n is the trial number. We measured the convergence speed of a simulation by the rate constants k_1
506 and k_2. In the vast majority cases, one of these rate constants was 5-50 times larger than the other; we
507 denote the larger constant k_{fast} and the other k_{slow}. For most parameter values, k_{fast} accounts for more than
508 80% of learning.

510 We next examined the influence of several key model parameters on convergence speed, such as
511 threshold and gradient descent steps size. First, we looked at the effect of the GC threshold. Learning was
512 fastest for GCL thresholds near a z-score of zero (Fig. 4C, *red circles*), the level that filters out half of the
513 input received by a GC. Convergence in networks that lack a GCL (MFs directly innervating P-cells) was
514 consistently slower (Fig. 4C, *blue line*) than networks with a GCL. Convergence was also sped up by
515 increasing the size of the parameter jumps in synaptic weight space during gradient descent (the "step
516 size"), but only to a limited degree (Fig. S3A). Indeed, at a GCL threshold of 0, convergence speed
517 *decreased* as the step size increased beyond $\sim 10^{-6}$ (au). We speculated that this trade-off was a
518 consequence of a failure to converge in a subset of simulations. To test this, we looked at the fraction of
519 simulations that converged towards a low MSE as a function of the update magnitude. We found that the
520 fraction of simulations that converged ("fraction successful") decreased with increasing step size (Fig.
521 S3B); in simulations that did not converge, the MSE increased explosively and synaptic weights diverged.
522 In such cases, we assume the large weight updates made it impossible to descend the MSE gradient; each
523 network weight update drastically changed the cost function such that local MSE minima were overshot.
524 When larger step sizes did permit convergence, progress was nevertheless slowed, likely because the
525 relatively large learning rates led to inefficient progress towards the MSE minimum.

527 Although larger step sizes eventually cause learning to slow and then fail entirely at a given GCL
528 threshold, higher thresholds permitted larger step sizes before failures predominated (Fig. S3B). Since
529 higher thresholds permit larger step sizes before convergence failure sets in, convergence speed might be
530 maximized by jointly optimizing step size and GCL threshold. We tested this by systematically raising
531 step sizes at each threshold until convergence success fell to 50%. We defined the "maximum
532 convergence rate" for a given threshold as the maximum convergence rate (derived from fitting the MSE
533 trajectory with a double exponential) yielding successful convergence at least 50% of the time. We found
534 that the threshold giving the fastest convergence was indeed higher when step size was also optimized
535 (Fig. S3B) than when step size was fixed (Fig. 4C). Thus, increased GCL thresholding can allow the
536 network to trade learning accuracy for increased speed of learning.

538 **Recovering GCL input from GCL output**
539 Having established a framework for studying GCL processing of time varying inputs, we wanted to
540 understand to what extent thresholding GCL activity led to the loss of information supplied by MF inputs,
541 which potentially contains useful features for learning. In other words, would Purkinje neurons be
542 deprived of behaviorally relevant mossy fiber information if these inputs are severely filtered by the
543 GCL? To assess this issue, we used a metric of information preservation called *explained variance*,
544 (Achen 1982); however, in this special case, we use the term '*variance retained*', because this metric
545 represents the preservation of information about the input after being subjected to filtering in the GCL
546 layer and we wanted to avoid confusing when describing linear regression results below. Let x_t denote the
547 MF input at time t. If the GCL activity preserves the information present in x_t, then it should be possible
548 to reconstruct the activity of MFs from GCL activity (see Methods for details on how this reconstruction

549 was performed). The variance retained is then the mean squared error between the actual MF input x_t and
550 the reconstructed input, normalized by the MF input variance:

$$R^2 = 1 - \frac{\sum_{t=1}^{T}\langle(\hat{x}_t - x_t)^2\rangle}{\sum_{t=1}^{T}Var[x_t]}$$

553 Our primary finding is that the GCL transmits nearly all of the information present in the MF inputs even
554 at fairly high thresholds, but only if the GCL is sufficiently large relative to the MF population. The
555 threshold, feedforward architecture, and relative balance of MF inputs and GC outputs all affect the
556 quality of the reconstruction. Variance retained by the reconstruction layer decreased with the GC layer
557 threshold, since it masked some subthreshold input values (Fig. 5B). Allowing more MF inputs per GC
558 recovered some of this masked information, since some subthreshold values are revealed through
559 summing with sufficiently suprathreshold values. However, these gains cease beyond a few MF inputs
560 per GC, since the exponential growth of MF combinations rapidly exceeds the number that the GCs can
561 represent (Marr 1969; Gilmer and Person 2017).

563 To disentangle the information contained in the summed inputs, many different combinations of inputs
564 must be represented to disambiguate the contributions of each MF input. Increasing the number of GCs
565 generally increases the variance retained, since more combinations of MF inputs are represented, and
566 reveal subthreshold input values (Fig. 5C). Interestingly, variance retained by the network varied non-
567 monotonically with the number of MF inputs (M) when the number of GCs (N) was fixed. This is because
568 having too few MF inputs means there may not be a sufficient number of combinations so that
569 subthreshold values can be revealed (by summing them with suprathreshold inputs) but having too many
570 saturates the information load of the GC layer (Fig. 5D). Lastly, when fixing the number of MF inputs
571 and GCs, there is an optimal number of MF inputs to each GC, which aligns with the anatomical
572 convergence factor of 4 MF/GC (Fig. 5E), related to previous findings that suggest the best way to
573 maximize dimensionality in the GC output layer is to provide sparse input from the mossy fibers (Litwin-
574 Kumar et al., 2017; Cayco-Gajic et al., 2017). Thus, there are two key features that shape the information
575 transferred to the GCL from the MF inputs. First, the way in which MF inputs are combined to form the
576 total input to each GC determines how much information about subthreshold inputs can be transferred
577 through the nonlinearity. Second, the total number of GC outputs determines how many MF input
578 combinations can be represented, so that, ultimately, the random sums of MFs can be disentangled by the
579 downstream reconstruction layer. Together, information transfer requires a combined summation and
580 downstream decorrelation process accomplished by the three-layer feedforward network.

582 **General statistical features of GCL population activity**
583 We were ultimately interested in which features of GCL signal processing account for learning. As a first
584 step, we examined a variety of population metrics across threshold levels, which had previously been
585 proposed to support perceptron learning. The first set of metrics related to pattern separation: (1)
586 dimensionality (Dim), (2) the number of explanatory principal components (PCs), (3) spatiotemporal
587 sparseness (STS), and (4) population variability (See methods for details). Most of these pattern
588 separation metrics, (Dim, PCs, and STS) showed non-monotonic relationships with threshold and peaked
589 at thresholds ranging between 0.5 and 1.5 (Fig. 6 A, B). Population variability, however, decreased with
590 increasing thresholds (Fig. 6C). Intuitively, this relationship captures the effect of low thresholds allowing

591   GC activity to relay the mean input, with no pattern separation occurring. With increasing threshold, GC
592   activity is driven by coincidence detection, leading to higher dimensional population output. At high
593   thresholds, inputs rarely summate to threshold, leading to lost representation that drives a roll-off in
594   pattern separation within the population. Notably, Dim, PCs, and STS peaked at higher thresholds than
595   peak learning performance, which was best at threshold zero, thus none of these three pattern separation
596   metrics alone map directly to learning performance. Population variability (i.e. GCL variance per unit) is
597   thought to aid classification and separability of GCL output (Cayco-Gajic et al., 2017). This metric's
598   decrease with increasing threshold was likely due to the decrease in overall representation by each unit
599   due to sparsening and diminishing the dynamic range of GC rates due to threshold subtraction (Fig. 2A
600   top, Fig. 6C).

602   The second set of metrics are related to sparseness: (1) temporal sparseness and (2) spatial sparseness.
603   Temporal sparseness – defined by the exponential decay of GC autocovariance, where smaller values
604   typify signals that change quickly with time–- decreased as a function of threshold because of sparsened
605   representation at higher thresholds (Fig. 6D). Spatial sparseness – defined as the mean pairwise GC
606   correlation – shared a drop-off after a threshold of 0, but increased again at high thresholds because only a
607   few MF signals were retained at high threshold and thus were highly correlated (Fig. 6E). By
608   experimental design, decorrelation was already maximized in OU inputs. Similar to the pattern separation
609   metrics, these sparseness metrics did not show an obvious relationship to the U-shaped learning
610   performance seen in Fig. 3A, bottom.

612   Finally, we examined three metrics of lossiness defined to quantify (1) the fraction of the total epoch with
613   no activity in any GC unit (e.g. with "temporal lossiness" of 0.1, 10% of the total epoch has no activity in
614   any GCs) (2) the proportion of granule cells with any activity over the entire epoch ("population
615   lossiness"); and (3) the mean fraction of the epoch in which each granule cell is active ("temporal cover").
616   Not surprisingly, each lossiness metric increased with high thresholds (Fig. 6F). However, despite
617   diminishing activity in individual GCs with increasing threshold, (the blue curve Fig. 6F), each GC was
618   resistant to becoming completely silent (green curve drop, Fig. 6F), owing to a few dominant inputs.

620   Notably, none of these metrics alone obviously tracked the U-shaped learning performance (Fig. 3A).
621   However, collectively, these descriptive statistics of model GCL population activity set the stage for
622   analyzing how information preprocessing by the basic GCL architecture relates to learning time series,
623   explored below.

625   **Improved learning with GCL transformations**
626   With the knowledge that thresholding drives changes both in learning time series (Fig. 3, 4) and in GC
627   population metrics that are theorized to modulate learning (Fig. 5, 6), we next directly investigated the
628   relationships of these metrics to learning performance. To test this, we used LASSO regression to identify
629   variables driving learning performance, taken from the metrics described in Figures 5 and 6 (Fig. 7A,
630   C). We found that a three-term model using the most explanatory variables–- STS, the number of
631   explanatory PCs and variance retained (Fig. 7B, C, D)–- accounted for 91% of learning variance. The
632   three-term model performance is plotted against the observed MSE over a range of thresholds in Fig. 7D,
633   showing strong similarity.

634

635   Although this model accounted for most learning, its performance was notably poorer at threshold ranges
636   where the GCL-based learning was best. We reasoned that interactions between GCL statistical metrics
637   might account for this deviation. To select potential metric interactions while constraining a regression
638   model, we used Bayesian information criteria (BIC) stepwise regression to identify variables that
639   accounted for learning (See Methods for normalization methods; Fig 7E). This model produced a better
640   approximation of learning (Fig. 7F). We found that a handful of competing variables (i.e. pattern
641   separation competing with retention of lossless representation) provided a small but crucial representation
642   of learning, which offset the poor learning between thresholds of -1 to 1 in the purely linear model (Fig.
643   7D vs Fig. 7F). While these interaction components were necessary to produce the best fitting model for
644   learning, the interactions were not the dominant regressors, as indicated by their relatively small Beta
645   values, and PCs and population variance remained top features explaining learning, similar to the linear
646   model.

648   These results were somewhat surprising given prior studies showing benefits of population sparseness or
649   decorrelation to learning. We noted that with the GCL filter model we could not clamp specific
650   population metrics to determine their contribution to learning, thus to interrogate this seeming disparity,
651   we constructed fictive GCL population activity that had specific statistical features and used these as
652   inputs to P-cells. Consistent with previous reports, decorrelation and temporal sparseness improved
653   learning accuracy, with complete decorrelation and temporally sparse supporting the best performance
654   (Fig S4; Cayco-Gaijic et al., 2018). Thus, on their own, population, temporal and idealized spatiotemporal
655   sparseness do modulate learning when their contribution is independent. Nevertheless, these features did
656   not emerge as features driving learning using GCL output from OU inputs to learn timeseries. This
657   discrepancy raises the possibility that the pattern separation metrics that drive learning may be dependent
658   on MF input statistics.

660   **GCL properties that enhance learning in naturalistic tasks**
661   Together, these models suggest that the GCL can reformat inputs in ways that support rapid and accurate
662   timeseries learning. We next asked whether the GCL metrics that drive best learning change when inputs
663   were inherently matched to outputs. This question is motivated by the topographical modules that
664   characterize the real cerebellum, each with associated specialized afferents (Apps and Garwicz, 2005; De
665   Zeeuw, 2020). Might these specialized afferents with specific statistical structure be especially suited to
666   support P-cell tuning for specific behaviors?

668   To examine whether statistical features that drive learning are sensitive to intrinsic input-output
669   relationships, we tested whether model inputs with naturalistic, behaviorally correlated statistics, derived
670   from electromyogram (EMG) signaling could support learning movement kinematics. In this assay, "MF
671   inputs" were EMG signals from human subjects performing a point-to-point reaching task. We tested
672   whether the model could learn associated limb kinematics from this input (Fig. 8A, B; Delis et al. 2018;
673   Tseng et al. 2007; Miall and Wolpert 1996; Wolpert et al., 1998).

675   Consistent with our previous observations, model P-cell output better learned kinematic target functions
676   when EMG inputs were preprocessed by the model GCL rather than fed directly to P-cells (Fig. 8A).
677   Moreover, thresholds that supported best learning were comparable to those using OU functions as inputs
678   (Fig. 8A vs Fig. 3A) and the accuracy of the learned outputs strongly resembled the recorded kinematic

679   positions (Fig. 8B). We observed a slight negative shift in thresholds supporting best performance using
680   EMG, suggesting that GCL population statistics that retain more of their inherent relationship to
681   kinematics (i.e. that the EMG alone predicted kinematics well), facilitated by lower threshold, might be
682   beneficial to learning kinematics. However, some EMG-kinematic pairings had stronger intrinsic
683   relationships than others. We used this variability to assay whether the strength of the intrinsic
684   relationship influenced which population metrics supported best learning. We first identified which
685   population statistics drove learning using RIDGE regression, which preserves even small contributions of
686   regressor variables to the model. We then looked at the slope of regressors that predicted learning as a
687   function of the MSE of MFs alone. We found that when the P-cell MSE was already low with direct MF
688   inputs, the information retention (Fig. 5) emerged as a key predictor of learning (i.e. GCL MSE, Fig. 8C,
689   green). Conversely, when MF based learning was poor (high MSE) a pattern separation metric, number of
690   explanatory PCs, became a more important driver of learning (Fig. 8C, orange). This observation is
691   captured in the metric "Regression Coeff. Ratio" Fig. 8C, which quantifies the coefficient of the variance
692   retained or explanatory PC regressor divided by the sum of all regressor coefficients computed in the
693   RIDGE regression. In effect, this method shows the normalized size of their impact on the regression.
694   Together this suggests that different population statistical features of GCL reformatting may serve
695   learning under different conditions: When intrinsic relationships are strong, the GCL's preservation of
696   MF input variance (variance retained) is an important population statistical feature; when MF activity is
697   more arbitrary relative to what the P-cell needs to encode, explanatory PCs (a pattern separation feature)
698   are more valuable for learning. Thus, "pattern separation" by the GCL is not one universal transform that
699   has broad utility. This observation raises the possibility that regional circuit specializations within the
700   cerebellar cortex, such as density of unipolar brush cells (Dino et al. 2000; Guo et al., 2021), Golgi cells,
701   or neuromodulators could bias GCL information reformatting to be more suitable for learning of different
702   tasks.
703
704   **DISCUSSION**
705   Here we asked a simple question: how does the cerebellar granule layer support temporal learning? The
706   question of the function of GCL architecture has captivated theorists for decades, leading to a hypothesis
707   of cerebellar learning that posits that the GCL reformats information to best suit associative learning in
708   Purkinje cells. Recent work has called many of these foundational ideas into question, however, including
709   whether GCL activity is sparse; high dimensional; and what properties of 'pattern separation' best support
710   learning (Wagner et al., 2017; Giovannucci et al., 2017; Knogler et al., 2017; Cayco-Gajic et al., 2017;
711   Gilmer and Person 2017). To reconcile empirical observations with theory, we hypothesized that input
712   statistics and task structures influence how the GCL supports learning. Here, we used naturalistic and
713   artificial time-varying inputs to a model GCL and identified pattern separation features that supported
714   learning time series, with an arbitrary but temporally linked input-output mapping, recapitulating
715   important features of physiological cerebellar learning tasks (Buonomano and Mauk 1994, Mauk and
716   Donnegan, 1997, Kennedy et al., 2014). Here we attempt to bridge these findings by examining
717   naturalistic challenges faced by the real circuit. Several important observations stemmed from these
718   simulations: (1) with naturalistic input statistics, the GCL produces temporal basis sets akin to those
719   hypothesized to support learned timing with minimal assumptions; (2) this reformatting is highly
720   beneficial to learning at intermediate thresholds; (3) maximal pattern separation does not support the best
721   learning; (4) rather, tradeoffs between loss of information and reformatting favored best learning at
722   intermediate network thresholds; and finally (5) different learning tasks are differentially supported by

723   diverse GCL population statistical features. Together these findings provide insight into the granule cell
724   layer as performing pattern separation of inputs that transform information valuable for gradient descent-
725   like learning.
726
727   *Emergence of spatiotemporal representation and contribution to learning*
728   A perennial question in cerebellar physiology is how the granule cell layer produces temporally varied
729   outputs that could support learned timing (Mauk and Buonomano 2004). While cellular and synaptic
730   properties have been shown to contribute (Chabrol et al., 2015; Duguid et al, 2012; Guo et al., 2021a;
731   Crowley et al., 2009; Rudolph et al., 2015; Buonomano and Mauk 1994; Kanichay and Silver 2008;
732   Simat et al., 2007; Mapelli et al., 2009; Rossi et al., 1996; Gall et al., 2005; Armano et al., 2000; Rizwan
733   et al. 2016; Tabuchi et al., 2019; D'Angelo and De Zeeuw 2009), we observed that with naturalistic
734   inputs, temporal basis set formation is a robust emergent property of the feedforward architecture of the
735   cerebellum coupled with a threshold-linear input-output function of granule cells receiving multiple
736   independent time-varying inputs (Fig. 1B-D). But is this reformatting beneficial to learning? We
737   addressed this question by comparing learning of a complex time-series in model Purkinje cells receiving
738   either mossy fibers alone or reformatted output from the GCL. We found that indeed the GCL
739   outperformed MFs alone in all tasks (Figs. 3, 4, 7). Nevertheless, we wondered what features of the
740   population activity accounted for this improved learning. While sparseness, decorrelation, dimensionality
741   and lossless encoding have been put forward as preprocessing steps supporting learning, we found that
742   none of these alone accounted for the goodness of model performance. Rather, disparate pattern
743   separation metrics appear to strike a balance between maximizing sparseness without trespassing into
744   lossy encoding space that severely, and necessarily, degrades learning of time-series.
745
746   These observations are interesting in light of a long history of work on granule layer function. Marr,
747   Albus, and others proposed that the granule cell layer performs pattern separation useful for classification
748   tasks. In this framework, sparseness is the key driver of performance, and could account for the vast
749   number of granule cells. Nevertheless, large-scale GCL recordings unexpectedly showed high levels of
750   correlation and relatively non-sparse activity (Wagner et al., 2017; Giovannucci et al., 2017; Knogler et
751   al., 2017). Despite methodological caveats, alternate recording methods seem to support the general
752   conclusion that sparseness is not as high as originally thought (Lanore et al. 2021; Kita et al., 2021;
753   Gurgani and Silver 2021). Indeed, subsequent theoretical work showed that sparseness has deleterious
754   properties (Cayco-Gajic et al., 2017; Billings et al., 2014), also observed in the present study, that may
755   explain dense firing patterns seen *in vivo*. Here we found that the best learning occurred when individual
756   granule cell activity occupied around half of the observed epoch (Fig. 6F, blue trace), achieved with
757   intermediate thresholding levels. We also observed temporal organization that is consistent with the firing
758   patterns observed *in vivo*. While these findings seem to suggest that sparseness is not the 'goal' of GCL
759   processing, our findings and others (Litwin-Kumar et al., 2016; Cayco-Gajic et al., 2017) suggest that
760   pattern separation broadly is a positive modulator of GCL support of learning processes.
761
762   Previous work proposed that time-series prediction was possible with access to a diverse set of geometric
763   functions represented in the GC population (Sanger et al., 2020). However, that study left open the
764   question of how such a diverse collection of basis functions would emerge. The GCL model used here
765   minimized free parameters by incorporating very few independent circuit elements, suggesting that a
766   single transform is sufficient to produce a basis set which is universally able to learn arbitrary target

767 functions. We used a simple threshold-linear filter with a singular global threshold that relied on sparse-
768 sampling to produce spatiotemporally varied population outputs. This simple function worked to support
769 learning at a broad range of inputs and thresholding values, ultimately allowing the Purkinje cells
770 downstream to associate the spatiotemporally sparser inputs with feedback to learn arbitrary and complex
771 target functions. The emergence of this basis set is remarkable given the very simple assumptions applied,
772 but is also physiologically realistic, given the simple and well characterized anatomical properties of the
773 MF divergence and convergence patterns onto GCs, which are among the simplest neurons in the brain
774 (Jakab and Hamori, 1988; Palay and Chan-Palay, 1974; Palkovits et al., 1971). Although we suggest that
775 the key regulator of thresholding in the system is the feedforward inhibition from Golgi cells, many
776 factors may regulate the transformation between input and GC output in the network, allowing for
777 multiple levels and degrees of control over the tuning of the filter or real mechanism that controls the
778 outcomes of GCL transformations. Golgi cell dynamics may prove critical for enforcing the balance
779 between pattern separation metrics and lossy encoding (Hull 2020) thus are critical players in mean
780 thresholding found here to optimize learning. Additional mechanistic considerations may also play a role,
781 including short-term synaptic plasticity (Chabrol et al. 2015) network recurrence (Gao et al. 2016; Houck
782 and Person 2014; 2015; Judd et al., 2021), and UBCs (Dino et al. 2000), allowing for a more nuanced and
783 dynamic regulatory system than the one shown here.
784
785 *Recapturing input information in the filtered GCL output*
786 Two schools of thought surround what information is relayed to Purkinje cells through GCs. Various
787 models assume that Purkinje cells inherit virtually untransformed MF information capable of explaining
788 kinematic tuning in P-cells (Markanaday et al. 2022; Herzfeld et al., 2020; Krauzlis and Lisberger, 1991).
789 This view is in contrast to suggestions of Marr and Albus, where the GCL sparsens information to such a
790 degree that Purkinje cells receive only a small remnant of the sensorimotor information present in mossy
791 fiber signals. These divergent views have never been reconciled to our knowledge. We addressed this
792 disconnect by determining the fraction of MF input variance recoverable in GCL output. Interestingly, the
793 GCL population retains sufficient information to recover more than 90% the input variance despite
794 filtering out 50% or more of the original signal (Fig. 5). This information recovery is achieved at the
795 population level and thus requires sufficient numbers of granule cells so that the subset of signals that are
796 subthreshold are also super-threshold in other subsets of GCs through probabilistic integration with other
797 active inputs. While variance recovery is not a true measure of mutual information, it is indicative of the
798 utility that the intersectional filtering performed by the GCL. The expansion of representations in the
799 GCL population achieved by capturing the coincidence of features in the input population creates a
800 flexible representation in the GCL output that has many beneficial properties, including the preservation
801 of information through some degree of preserved mutual information between the GCL and its inputs. Yet
802 despite this retention of input variance by the GCL, its transformations nevertheless greatly improve
803 learning.
804
805 *Enhanced learning speed*
806 Our model not only improved learning accuracy, but also speed, compared to MFs alone (Fig. 4). Both
807 learning speed and accuracy progressed in tandem: threshold parameter ranges that enhanced overall
808 learning speed also minimized mean squared error, suggesting that speed and accuracy are enhanced by
809 similar features in GCL output. Learning speed was well described by a double exponential function with
810 a slow and fast component. This dual time course in the model with only one learning rule is interesting

in light of observations of behavioral adaptation that also follow dual time courses (Herzfeld et al., 2014; Smith et al., 2006). Some behavioral studies have postulated that these time courses suggest multiple underlying learning processes (Yang and Lisberger, 2014). Our model indicates that even with a single learning rule and site of plasticity, multiple time-courses can emerge, presumably because when error becomes low, update rates also slow down.

Another observation stemming from simulations studying learning speed was that the behavior of the model varied as a function of the learning 'step size' parameter of the gradient descent method (Fig. S3). The step size–- ie. the typically small, scalar regulating change in the weights between GCs and P-cells following an error–- determined the likelihood of catastrophically poor learning: when the step size was too large, it led to extremely poor learning because the total output 'explodes' and fails to converge on a stable output. Nevertheless, the model tolerated large steps and faster learning under some conditions, since the threshold also influenced the likelihood of catastrophic learning. Generally, higher thresholds prevented large weight changes from exploding, suggesting that sparse outputs may have an additional role in speeding learning by supporting larger weight changes in Purkinje cells. Indeed, appreciable changes in simple spike rates occur on a trial-by-trial basis, gated by the theorized update signals that Purkinje cells receive, climbing fiber mediated complex spikes. These plastic changes in rate could reflect large weight updates associated with error. Moreover, graded complex spike amplitudes that alter the size of trial-over-trial simple spike rate changes suggest that update sizes are not fixed (Najafi et al., 2014; Herzfeld et al., 2020; Medina and Raymond 2018). Thus, although gradient descent is not wholly physiological, this finding predicts that the amplitude of synaptic weight changes following a complex spike might be set by tunable circuitry in the molecular layer to optimize learning speed relative to the statistics of the GCL output.

Together, this study advances our understanding of how the GCL may diversify time-varying inputs and informs interpretation of empirical results. For instance, the timecourse of learning varies widely across tasks. Eyeblink conditioning paradigms require hundreds of trials to learn (Millenson 1997; Khilkevich et al., 2016; Lincoln et al., 1982), while saccade adaptation and visuomotor adaptation of reaches (Raymond and Lisberger; Martin et al. 1996), require just tens of trials (Tseng et al., 2007; Shadmehr and Mussa-Ivaldi 1994; Ruttle et al., 2021; Calame et al., 2021). A prediction from our study is that the temporal diversity of the GCL basis set during a behavior influences learning speed. Time-invariant cues such as those seen in EBC would be difficult, if not impossible, for our model GCL to reformat and sparsen, as they are incompatible with thresholding-based filtering of input signals. Supportive of this view, recent work showed that EBC learning was faster if the animal is locomoting during training (Albergaria et al., 2018). We hypothesize that naturalistic time-variant signals associated with ongoing movements entering the cerebellum support robust temporal pattern separation in the GCL, enhancing learning accuracy and speed, while time invariant associative signals used in typical classical conditioning paradigms result in an impoverished 'basis', making learning more difficult, despite other circuit elements that may contribute to the GCL basis formation.

859

860   **REFERENCES**
861   Achen CH (1982) Interpreting and Using Regression. Sage University Paper Series on Quantitative
862   Applications in the Social Sciences, vol. 29.

863

864   Albergaria C, Silva NT, Pritchett DL, Carey MR (2018) Locomotor activity modulates associative
865   learning in mouse cerebellum. Nature Neuroscience 21, 725–735. doi:10.1038/s41593-018-0129-x

866

867   Albus JS (1971) A theory of cerebellar function. Mathematical Biosciences 10, 25–61. doi:10.1016/0025-
868   5564(71)90051-4

869

870   Albus JS (1975) Data Storage in the Cerebellar Model Articulation Controller (CMAC). ASME. J. Dyn.
871   Sys., Meas., Control. 97(3): 228–233.

872

873   Armano S, Rossi P, Taglietti V, D'Angelo E (2000) Long-term potentiation of intrinsic excitability at the
874   mossy fiber–granule cell synapse of rat cerebellum. J Neurosci 20:5208–5216, pmid:10884304.

875

876   Apps R, Garwicz M (2005) Anatomical and physiological foundations of cerebellar information
877   processing. Nat Rev Neurosci 6:297-311

878

879   Arenz A, Silver RA Schaefer AT, Margrie TW (2008) The contribution of a single synapse to sensory
880   representation in vivo. 321:977-80

881

882   Bengtsson F, Jorntell H (2009) Sensory transmission in cerebellar granule cells relies on similarly coded
883   mossy fiber inputs. Proceedings of the National Academy of Sciences 106, 2389–2394.
884   doi:10.1073/pnas.0808428106

885

886   Billings G, Piasini E, Lőrincz A, Nusser Z, Silver RA (2014) Network Structure within the Cerebellar
887   Input Layer Enables Lossless Sparse Encoding. Neuron 83, 960–974. doi:10.1016/j.neuron.2014.07.020

888

889   Buonomano DV, Mauk MD (1994) Neural Network Model of the Cerebellum: Temporal Discrimination
890   and the Timing of Motor Responses. Neural Computation 6, 38–55. doi:10.1162/neco.1994.6.1.38

891

892   Calame DJ, Becker MI, Person AL (2021) Associative learning underlies skilled reach adaptation.
893   Biorxiv 2021.12.17.473247

894

895   Cayco-Gajic NA, Clopath C, Silver RA (2017) Sparse synaptic connectivity is required for decorrelation
896   and pattern separation in feedforward networks. Nature Communications 8. doi:10.1038/s41467-017-
897   01109-y

898

899 Cayco-Gajic NA, Silver RA (2019) Re-evaluating Circuit Mechanisms Underlying Pattern Separation.
900 Neuron 101, 584–602. doi:10.1016/j.neuron.2019.01.044

901

902 Chabrol FP, Arenz A, Wiechert MT, Margrie TW, Digregorio DA (2015) Synaptic diversity enables
903 temporal coding of coincident multisensory inputs in single neurons. Nature Neuroscience 18, 718–727.
904 doi:10.1038/nn.3974

905

906 Crowley JJ, Fioravante D, Regehr WG (2009) Dynamics of fast and slow inhibition from cerebellar Golgi
907 cells allow flexible control of synaptic integration. Neuron 63:843–853.
908 doi:10.1016/j.neuron.2009.09.004 pmid:19778512

909

910 D'Angelo E, De Zeeuw CI (2009) Timing and plasticity in the cerebellum: focus on the granular layer.
911 Trends Neurosci 32:30–40. doi:10.1016/j.tins.2008.09.007 pmid:18977038

912

913 De Zeeuw CI, Simpson JI, Hoogenraad CC, Galjart N, Koekkoek SK, Ruigrok TJ (1998) Microcircuitry
914 and function of the inferior olive. Trends Neurosci. 21: 391–400

915

916 De Zeeuw CI (2020) Bidirectional learning in upbound and downbound microzones of the cerebellum.
917 Nat Rev Neurosci 22: 92-110

918

919 Dean P, Porrill J (2008) Adaptive-filter Models of the Cerebellum: Computational Analysis. The
920 Cerebellum 7, 567–571. doi:10.1007/s12311-008-0067-3

921

922 Delis I, Hilt PM, Pozzo T, Panzeri S, Berret B (2018) Deciphering the functional role of spatial and
923 temporal muscle synergies in whole-body movements. Scientific Reports 8. doi:10.1038/s41598-018-
924 26780-z

925

926 Hilt PM, Delis I, Pozzo T, Berret B (2018) Space-by-Time Modular Decomposition Effectively Describes
927 Whole-Body Muscle Activity During Upright Reaching in Various Directions. Front Comput Neurosci.
928 2018;12:20. doi:10.3389/fncom.2018.00020

929

930 Dino MR, Schuerger RJ, Liu Y, Slater NT, Mugnaini E (2000) Unipolar brush cell: a potential
931 feedforward excitatory interneuron of the cerebellum. Neuroscience 98:625–636.

932

933 Duguid I, Branco T, London M, Chadderton P, Hausser M (2012) Tonic Inhibition Enhances Fidelity of
934 Sensory Information Transmission in the Cerebellar Cortex. The Journal of Neuroscience 32, 11132–
935 11143. doi:10.1523/jneurosci.0460-12.2012

936

937 Eccles JC, Ito M, Szentágothai J (1967) The Cerebellum as a Neuronal Machine, Springer, New York
938 (1967)
939 Eriksson, JL, Robert A (1999) The representation of pure tones and noise in a model of cochlear nucleus
940 neurons. The Journal of the Acoustical Society of America 106, 1865–1879. doi:10.1121/1.427936

941

942 Fujita M (1982) Adaptive filter model of the cerebellum. Biological Cybernetics 45, 195–206.
943 doi:10.1007/bf00336192
944
945 Gall D, Prestori F, Sola E, D'Errico A, Roussel C, Forti L, Rossi P, D'Angelo E (2005) Intracellular
946 calcium regulation by burst discharge determines bidirectional long-term synaptic plasticity at the
947 cerebellum input stage. J Neurosci 25:4813–4822, doi:10.1523/JNEUROSCI.0410-05.2005,
948 pmid:15888657.
949
950 Gao Z, Proietti-Onori M, Lin Z, Ten Brinke MM, Boele HJ, Potters JW, Ruigrok TJ, Hoebeek FE, De
951 Zeeuw CI (2016) Excitatory cerebellar nucleocortical circuit provides internal amplification during
952 associative conditioning. Neuron 89:645–657. 10.1016/j.neuron.2016.01.008
953
954 Gilmer JI, Person AL (2017) Morphological Constraints on Cerebellar Granule Cell Combinatorial
955 Diversity. The Journal of Neuroscience 37, 12153–12166. doi:10.1523/jneurosci.0588-17.2017
956
957 Gilmer JI, Person AL (2018) Theoretically Sparse, Empirically Dense: New Views on Cerebellar Granule
958 Cells. Trends in Neurosciences 41, 874–877. doi:10.1016/j.tins.2018.09.013
959
960 Giovannucci A, Badura A, Deverett B, Najafi F, Pereira TD, Gao Z, Ozden I, Kloth AD, Pnevmatikakis
961 E, Paninski L, De Zeeuw CI, Medina JF, Wang SS-H (2017) Cerebellar granule cells acquire a
962 widespread predictive feedback signal during motor learning. Nature Neuroscience 20, 727–734.
963 doi:10.1038/nn.4531
964
965 Guo C, Huson V, Macosko EZ, Regehr WG (2021) Graded heterogeneity of metabotropic signaling
966 underlies a continuum of cell-intrinsic temporal responses in unipolar brush cells. Nat Comm 12:5491
967
968 Guo J-Z, Sauerbrei BA, Cohen JD, Mischiati M, Graves AR, Pisanello f, Branson KM, Hantman AW
969 (2021) Disrupting cortico-cerebellar communication impairs dexterity eLife 10:e65906.
970
971 Gurnani H, Silver RA (2021) Multidimensional population activity in an electrically coupled inhibitory
972 circuit in the cerebellar cortex. Neuron 109, 1739–1753.e8. doi:10.1016/j.neuron.2021.03.027
973
974 Herculano-Houzel S (2010) Coordinated scaling of cortical and cerebellar numbers of neurons. Front
975 Neuroanat 4:12. doi:10.3389/fnana.2010.00012 pmid:20300467
976
977 Herzfeld DJ, Hall NJ, Tringides M, Lisberger, SG (2020) Principles of operation of a cerebellar learning
978 circuit. eLife 9. doi:10.7554/elife.55217
979
980 Herzfeld DJ, Kojima Y, Soetedjo R, Shadmehr R (2015) Encoding of action by the Purkinje cells of the
981 cerebellum. Nature 526, 439–442. doi:10.1038/nature15693
982
983 Houck BD, Person AL (2014) Cerebellar loops: a review of the nucleocortical pathway. Cerebellum
984 13:378–385. 10.1007/s12311-013-0543-2
985

986 Houck BD, Person AL (2015) Cerebellar premotor output neurons collateralize to innervate the cerebellar
987 cortex. J Comp Neurol 523:2254–2271. 10.1002/cne.23787

988

989 Huang CC, Sugino K, Shima Y, Guo C, Bai S, Mensh BD, Nelson SB, Hantman AW (2013)
990 Convergence of pontine and proprioceptive streams onto multimodal cerebellar granule cells. Elife
991 2:e00400. doi:10.7554/eLife.00400 pmid:23467508

992

993 Hull C (2020) Prediction signals in the cerebellum: beyond supervised motor learning. Elife. 9:e54073.
994 doi: 10.7554/eLife.54073. PMID: 32223891; PMCID: PMC7105376.

995

996 Ishikawa T, Shimuta M, Häusser, M (2015) Multimodal sensory integration in single cerebellar granule
997 cells in vivo. eLife 4. doi:10.7554/elife.12916

998

999 Ito M, Shiida T, Yagi N, Yamamoto M (1974) Visual influence on rabbit horizontal vestibulo-ocular
1000 reflex presumably effected via the cerebellar flocculus. Brain Res 65:170 –174.

1001

1002 Izawa J, Criscimagna-Hemminger SE, Shadmehr R (2012) Cerebellar Contributions to Reach Adaptation
1003 and Learning Sensory Consequences of Action. The Journal of Neuroscience 32, 4230–4239.
1004 doi:10.1523/jneurosci.6353-11.2012

1005

1006 Jakab RL, Hamori J (1988) Quantitative morphology and synaptology of cerebellar glomeruli in the rat.
1007 Anatomy and Embryology 179, 81–88. doi:10.1007/bf00305102

1008

1009 Judd EN, Lewis SM, Person AL (2021) Diverse inhibitory projections from the cerebellar interposed
1010 nucleus. Elife. 10:e66231. doi: 10.7554/eLife.66231. PMID: 34542410; PMCID: PMC8483738.

1011

1012 Kalmbach BE, Voicu H, Ohyama T, Mauk MD (2011) A Subtraction Mechanism of Temporal Coding in
1013 Cerebellar Cortex. The Journal of Neuroscience 31, 2025–2034. doi:10.1523/jneurosci.4212-10.2011

1014

1015 Kanichay RT, Silver RA (2008) Synaptic and Cellular Properties of the Feedforward Inhibitory Circuit
1016 within the Input Layer of the Cerebellar Cortex. The Journal of Neuroscience 28, 8955–8967.
1017 doi:10.1523/jneurosci.5469-07.2008

1018

1019 Kennedy A, Wayne G, Kaifosh P, Alviña K, Abbott LF, Sawtell NB (2014) A temporal basis for
1020 predicting the sensory consequences of motor commands in an electric fish. Nature Neuroscience 17,
1021 416–422. doi:10.1038/nn.3650

1022

1023 Khilkevich A, Halverson HE, Canton-Josh JE, Mauk MD (2016) Links Between Single-Trial Changes
1024 and Learning Rate in Eyelid Conditioning. The Cerebellum 15, 112–121. doi:10.1007/s12311-015-0690-8

1025

1026 Kita K, Albergaria C, Machado AS, Carey MR, Müller M, Delvendahl I (2021) GluA4 facilitates
1027 cerebellar expansion coding and enables associative memory formation. eLife 10. doi:10.7554/elife.65152

1028

1029 Knogler LD, Markov DA, Dragomir EI, Štih V, Portugues R (2017) Sensorimotor Representations in
1030 Cerebellar Granule Cells in Larval Zebrafish Are Dense, Spatially Organized, and Non-temporally
1031 Patterned. Current Biology 27, 1288–1302. doi:10.1016/j.cub.2017.03.029

1033 Krauzlis RJ, Lisberger SG (1991) Visual motion commands for pursuit eye movements in the cerebellum.
1034 Science 253:568-71

1036 Lanore F, Cayco-Gajic NA, Gurnani H, Coyle D, Silver, RA (2021) Cerebellar granule cell axons support
1037 high-dimensional representations. Nature Neuroscience 24, 1142–1150. doi:10.1038/s41593-021-00873-x

1039 Lincoln JS, Mccormick DA, Thompson RF (1982) Ipsilateral cerebellar lesions prevent learning of the
1040 classically conditioned nictitating membrane/eyelid response. Brain Research 242, 190–193.
1041 doi:10.1016/0006-8993(82)90510-8

1043 Litwin-Kumar A, Harris KD, Axel R, Sompolinsky H, Abbott LF (2017) Optimal Degrees of Synaptic
1044 Connectivity. Neuron 93, 1153–1164.e7. doi:10.1016/j.neuron.2017.01.030

1046 Liu, Y, Tiganj, Z, Hasselmo, ME, Howard, MW (2019) A neural microcircuit model for a scalable scale-
1047 invariant representation of time. Hippocampus. 29: 260– 274.

1049 Mapelli L, Rossi P, Nieus T, D'Angelo E (2009) Tonic activation of GABAB receptors reduces release
1050 probability at inhibitory connections in the cerebellar glomerulus. J Neurophysiol 101:3089–3099.
1051 doi:10.1152/jn.91190.2008 pmid:19339456

1053 Markanday A, Hong S, Inoue J, De Schutter E, Their P (2022) Multidimensional cerebellar computations
1054 for flexible kinematic control of movements. bioRxiv 2022.01.11.475785

1056 Marr D (1969) A theory of cerebellar cortex. The Journal of Physiology 202, 437–470.
1057 doi:10.1113/jphysiol.1969.sp008820

1059 Martin TA, Keating JG, Goodkin HP, Bastian AJ, Thach WT (1996) Throwing while looking through
1060 prisms: I. Focal olivocerebellar lesions impair adaptation. Brain 119, 1183–1198.
1061 doi:10.1093/brain/119.4.1183

1063 Mauk MD, Buonomano DV (2004) THE NEURAL BASIS OF TEMPORAL PROCESSING. Annual
1064 Review of Neuroscience 27, 307–340. doi:10.1146/annurev.neuro.27.070203.144247

1066 Mauk MD, Donegan NH. A model of Pavlovian eyelid conditioning based on the synaptic organization of
1067 the cerebellum. Learn Mem. 1997 May-Jun;4(1):130-58. doi: 10.1101/lm.4.1.130. PMID: 10456059.

1069 Mauk MD, Steinmetz JE, Thompson RF (1986) Classical conditioning using stimulation of the inferior
1070 olive as the unconditioned stimulus. Proc Natl Acad Sci USA, 83 pp. 5349-5353
1071

1072 Mccormick DA, Clark GA, Lavond DG, Thompson RF (1982) Initial localization of the memory trace for
1073 a basic form of learning. Proceedings of the National Academy of Sciences 79, 2731–2735.
1074 doi:10.1073/pnas.79.8.2731
1075 Medina J (2000) Mechanisms of cerebellar learning suggested by eyelid conditioning. Current Opinion in
1076 Neurobiology 10, 717–724. doi:10.1016/s0959-4388(00)00154-9
1077
1078 Miall RC, Wolpert DM (1996) Forward Models for Physiological Motor Control. Neural Netw. 9:1265–
1079 1279.
1080
1081 Millenson JR, Kehoe EJ, Gormezano I (1977) Classical conditioning of the rabbit's nictitating membrane
1082 response under fixed and mixed CS–US intervals. Learn Motiv, 8 pp. 351-366
1083
1084 Najafi F, Giovannucci A, Wang SS, Medina JF (2014) Coding of stimulus strength via analog calcium
1085 signals in Purkinje cell dendrites of awake mice. Elife. 2014;3:e03663. doi:10.7554/eLife.03663
1086
1087 Nocedal J, Wright SJ (2006) Numerical Optimization, 2nd ed., New York: Springer.
1088
1089 Palay S, Chan-Palay V (1974) Cerebellar cortex: Cytology and Organization, pp 100–132. Berlin-
1090 Heidelberg: Springer-Verlag.
1091
1092 Palkovits M, Magyar P, Szentágothai J (1971) Quantitative histological analysis of the cerebellar cortex
1093 in the cat. Brain Research 32, 15–30. doi:10.1016/0006-8993(71)90152-1
1094
1095 Rancz EA, Ishikawa T, Duguid I, Chadderton P, Mahon S, Häusser M (2007) High-fidelity transmission
1096 of sensory information by single cerebellar mossy fibre boutons. Nature 450, 1245–1248.
1097
1098 Raymond JL, Lisberger SG (1998) Neural Learning Rules for the Vestibulo-Ocular Reflex. The Journal
1099 of Neuroscience 18, 9112–9129. doi:10.1523/jneurosci.18-21-09112.1998
1100
1101 Raymond JL, Medina JF (2018) Computational Principles of Supervised Learning in the Cerebellum.
1102 Annu Rev Neurosci. 41:233-253. doi: 10.1146/annurev-neuro-080317-061948. PMID: 29986160;
1103 PMCID: PMC6056176.
1104
1105 Rizwan AP, Zhan X, Zamponi GW, Turner RW (2016) Long-Term Potentiation at the Mossy Fiber–
1106 Granule Cell Relay Invokes Postsynaptic Second-Messenger Regulation of Kv4 Channels. The Journal of
1107 Neuroscience 36, 11196–11207. doi:10.1523/jneurosci.2051-16.2016
1108
1109 Rossi P, D'Angelo E, Taglietti V (1996) Differential long-lasting potentiation of the NMDA and non-
1110 NMDA synaptic currents induced by metabotropic and NMDA receptor coactivation in cerebellar granule
1111 cells. Eur J Neurosci 8:1182–1189, doi:10.1111/j.1460-9568.1996.tb01286.x, pmid:8752588.
1112
1113 Rudolph S, Hull C, Regehr WG (2015) Active dendrites and differential distribution of calcium channels
1114 enable functional compartmentalization of Golgi cells. J Neurosci 35:15492–15504.
1115 doi:10.1523/JNEUROSCI.3132-15.2015 pmid:26609148

1116 Ruttle JE, Marius 't Hart B, Henriques DYP (2021) Implicit motor learning within three trials. Sci Rep.
1117 11(1):1627. doi: 10.1038/s41598-021-81031-y.
1118
1119 Sanger TD, Yamashita O, Kawato M (2020) Expansion coding and computation in the cerebellum: 50
1120 years after the Marr–Albus codon theory. The Journal of Physiology 598, 913–928. doi:10.1113/jp278745
1121
1122 Saviane C, Silver RA (2006) Fast vesicle reloading and a large pool sustain high bandwidth transmission
1123 at a central synapse. Nature 439:983–987. doi:10.1038/nature04509 pmid:16496000
1124
1125 Shadmehr R, Mussa-Ivaldi F (1994) Adaptive representation of dynamics during learning of a motor task.
1126 The Journal of Neuroscience 14, 3208–3224. doi:10.1523/jneurosci.14-05-03208.1994
1127
1128 Simat M, Parpan F, Fritschy JM (2007) Heterogeneity of glycinergic and gabaergic interneurons in the
1129 granule cell layer of mouse cerebellum. J Comp Neurol 500:71–83. doi:10.1002/cne.21142
1130 pmid:17099896
1131
1132 Smith MA, Ghazizadeh A, Shadmehr R (2006) Interacting adaptive processes with different timescales
1133 underlie short-term motor learning. PLoS Biol 4: e179, 2006. doi:10.1371/journal.pbio.0040179
1134
1135 Solinas S, Nieus T, D'Angelo E (2010) A realistic large-scale model of the cerebellum granular layer
1136 predicts circuit spatio-temporal filtering properties. Front Cell Neurosci. 2010 May 14;4:12. doi:
1137 10.3389/fncel.2010.00012. PMID: 20508743; PMCID: PMC2876868.
1138
1139 Tabuchi S, Gilmer JI, Purba K, Person AL (2019) Pathway-Specific Drive of Cerebellar Golgi Cells
1140 Reveals Integrative Rules of Cortical Inhibition. The Journal of Neuroscience 39, 1169–1181.
1141 doi:10.1523/jneurosci.1448-18.2018
1142
1143 Tseng YW, Diedrichsen J, Krakauer JW, Shadmehr R, Bastian AJ. (2007) Sensory prediction errors drive
1144 cerebellum-dependent adaptation of reaching. J Neurophysiol. 2007 Jul;98(1):54-62. doi:
1145 10.1152/jn.00266.2007. Epub 2007 May 16. PMID: 17507504.
1146
1147 Tyrrell T, Willshaw D (1992) Cerebellar cortex: its simulation and the relevance of Marr's theory. Philos
1148 Trans R Soc Lond B Biol Sci. 29;336(1277):239-57.
1149
1150 Van Kan PL, Gibson AR, Houk JC (1993) Movement-related inputs to intermediate cerebellum of
1151 monkey. 69:74-94
1152
1153 Wagner MJ, Kim TH, Savall J, Schnitzer MJ, Luo L (2017) Cerebellar granule cells encode the
1154 expectation of reward. Nature 544, 96–100. doi:10.1038/nature21726
1155
1156 Williams A, Roberts PD, Leen TK (2003) Stability of negative-image equilibria in spike-timing-
1157 dependent plasticity. Phys Rev E 68, 021923
1158

1159 Wolpert DM, Miall RC, Kawato, M (1998) Internal models in the cerebellum. Trends in Cognitive
1160 Sciences 2, 338–347. doi:10.1016/s1364-6613(98)01221-2

1161

1162 Wright SJ, Nowak RD, Figueiredo MAT (2009) Sparse Reconstruction by Separable Approximation.
1163 Trans. Sig. Proc., Vol. 57, No 7: 2479–2493.

1164

1165 Yang Y, Lisberger SG (2014) Role of Plasticity at Different Sites across the Time Course of Cerebellar
1166 Motor Learning. The Journal of Neuroscience 34, 7077–7090. doi:10.1523/jneurosci.0017-14.2014

1167

1168 Zhou S, Masmanidis SC, Buonomano DV (2020) Neural Sequences as an Optimal Dynamical Regime for
1169 the Readout of Time. Neuron. 108(4):651-658.e5. doi: 10.1016/j.neuron.2020.08.020. Epub 2020 Sep 17.
1170 PMID: 32946745; PMCID: PMC7825362.

1171

1172 **Figure Legends**

1173

1174 **Figure 1: Model architecture and effects of thresholding on GCL population activity**
1175 **A.** Diagram of algorithm implementation. Left shows Ornstein-Uhlenbeck processes (see Methods) as
1176 proxies for mossy fiber (MFs, blue) inputs to granule cell units (GCs, red), with convergence and
1177 divergence of MFs to GCs noted beneath MFs. GCs employ threshold-linear filtering shown beneath the
1178 red parallel fibers. GC outputs are then transmitted to downstream Purkinje cells (P-cells). P-cells learn to
1179 predict target functions by reweighting GC inputs. Differences between the prediction and true target are
1180 transmitted as an 'error', which updates synaptic weights to P-cells. **B.** Emergence of temporal basis sets
1181 in model GCLs using inputs derived from pontine neuron recordings. Top, Pontine recordings in mice
1182 during pellet reaching task, aligned to reach onset at 0 ms. Bottom, Model GCL output using PN activity
1183 as input **C.** Same as B, but using EMGs as MF inputs. Top, electromyogram (EMG) recordings from
1184 human subject in point-to-point reaching task (EMG, top). Bottom, model GCL output using EMG as
1185 input **D**. Same as B, but using OU functions as MF inputs. Bottom, model GCL output using OU
1186 processes as inputs. The model GCL enhanced spatiotemporal representation for all three input types (B-
1187 D).

1188

1189

1190 **Figure 2: Effect of increased input noise on GCL peak activity timing**
1191 **A**. Example MF input modeled as an OU process without noise (left), and with (right).
1192 **B.** Example of a GCL population with stable OU process as input (noiseless; left), and the population
1193 with addition of noise (middle and right). The GC population is ordered by timing of peak rate in the
1194 noiseless condition. **C.** Cumulative distribution of peak rate time shift between 'no noise' and 25% noise
1195 (left) or 50% noise (right), with MFs in black and GCs in red. X-axis is bounded to capture ~85% of
1196 population. CDF step length is 1 ms.

1197

1198

1199 **Figure 3: Enhanced time series learning using GCL mode.**
1200 **A**: Top: GCL output at different threshold levels. Bottom, relationship of threshold level to learning
1201 accuracy (MSE) for P-cells fed MFs directly (blue) or the output of the GCL (orange; error, standard
1202 deviation). **B**. Top left: P-cells were tasked with learning a complex timeseries that could be rendered as

1203 an image recognizable to humans, a cat with superimposed text. Top right: If P-cells were fed MF input
1204 directly, their best learning output was not recognizable as a cat, despite seemingly low MSE of 0.02.
1205 Bottom row: If P-cells were fed GCL output, they learned timeseries that rendered a matching image,
1206 with MSEs dependent on threshold, but varying between 0.0078 and 0.0016. This figure provides an
1207 intuitive sense of the practical difference between MSE of 0.02 and 0.0016, achieved with P-cells learning
1208 using MFs directly or with the support of GCL preprocessing.
1209
1210
1211

1212 **Figure 4. Learning speed increases with GCL**
1213 **A**. Example of learned predictions after 1,5, and 50 trials of learning, with predictions in red and target
1214 function in black. **B.** Exam4ple learning trajectory of MSE fit with a double exponential. Black circles:
1215 MSE of network output on each trial. Red line: double exponential fit MSE during learning. Here, step
1216 size was 10^{-6} and z-scored GCL threshold was 0. We use the exponents k from the exponential fit to
1217 measure learning speed. **C**. Learning speed as a function of GCL threshold (red dots). Blue line: learning
1218 speed in networks lacking GCL, i.e. mossy fibers directly innervate output Purkinje unit, gradient descent
1219 step size was 10^{-6}.
1220

1221 **Figure 5: Recovering inputs with an optimal linear readout**
1222 **A.** Network model schematic. Granule cell (GC, red, center) layer thresholds the sum of (4 here)
1223 randomly chosen mossy fiber (MF, black, left) inputs, which are then fed into a reconstruction layer
1224 which uses the optimal weighting from all N GCs to approximate each of the M inputs (compare blue
1225 readouts to grey inputs). **B.** Increasing the threshold of the GC layer (N=500 outputs) decreases the
1226 explained variance (i.e. variance retained) of the best reconstruction layer (M=50), but the effect is
1227 reduced with an intermediate number of MF inputs per GC. **C.** Variance retained increases with the ratio
1228 of GCs per MF but gains from increasing the number of inputs to each GC are limited (max at 4 inputs).
1229 Here there are M=50 MF Inputs at the threshold = 0. **D.** For a fixed number of GC outputs N, there is an
1230 optimal number of MF inputs (M) for which the variance retained of the reconstruction layer is
1231 maximized. **E. i.** For a fixed number of GC outputs N and MF inputs M=50, there is an optimal number
1232 of inputs per G (around 4) for maximizing variance retained. **ii.** Same as i, but with each value normalized
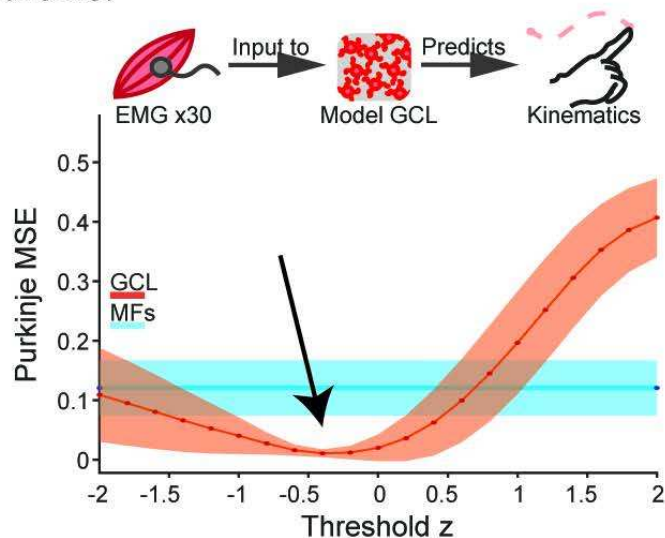1233 to its maximum to show maximized values at inputs = 4.
1234

1235 **Figure 6: Statistical features of GCL output A.** GCL dimensionality (red) and MF dimensionality
1236 (blue) as a function of threshold. Note peak near a threshold of 1 for the GCL. **B.** Two metrics of pattern
1237 separation in GCL output–- STS (light orange) and PCs (dark orange)–- as a function of threshold. Note
1238 peaks near 1.5 and 0.5, respectively. **C.** The sum of GCL variance produced by the model as a function of
1239 threshold. Note monotonic decrease with threshold. **D**. Temporal sparseness as a function of thresholding.
1240 Note monotonic decrease in GCL with thresholding. **E**. Mean pairwise correlation of the population
1241 plotted as a function of threshold. Note trough near 1. **F**: Three forms of lossiness in GCL output as a
1242 function of threshold. Each metric had differential sensitivity to thresholding but note that all decrease
1243 with increasing threshold. Across metrics, function maxima and minima ranged widely and were not
1244 obviously related to thresholds of optimized learning.
1245

Figure 7. Relationship between GCL population statistics and MSE A. LASSO regression was used to identify GCL population metics that predicted learning performance. **A** shows the model selection as a function of progression of the Lambda parameter (which is the penalty applied to each regressor). The following potential regressors were used: dimensionality (Dim.), spatiotemporal sparseness (STS), explanatory principal components of the GC population (PCs), population variability (Pop. Var.), spatial sparseness (S. Sparse.), temporal sparseness (T. Sparse.), temporal lossiness (T. Loss.), population lossiness (P. Loss), and input variance retained (Var. Ret; Figure 5). Arrow shows selection point of LASSO regression MSE using "1SE" (1 standard error) method (see Methods, purple lines, black dot and arrow indicating the selected model, with red arrow showing selection point in the parameter reduction plot, red). **B.** Relationship between LASSO model (predicted relative MSE) against the observed relative MSE (ratio of GC MSE to MF alone MSE) with fit line and variance explained by regression ($R^2 = 0.91$) **C.** Regression slopes of the selected LASSO model from A, showing that STS, PCs, and Input Variance Retained are the selected regressors, with Var. Ret. being the largest contributing factor. All factors normalized to a normal distribution for comparison. **D.** The output of the selected model and the observed MSE plotted against threshold for a comparison of fits, demonstrating high accuracy in the 0-2 range, but less accuracy in the -2-0 range. **E-F,** Similar to C-D except using Bayesian information criteria stepwise regression model to select metrics that explain learning.
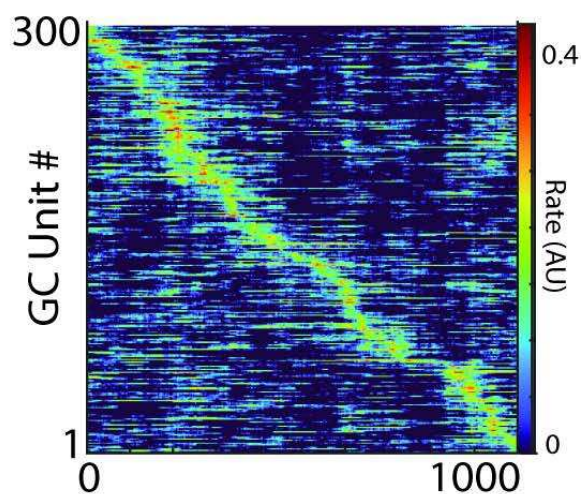
Figure 8. Relationship of MF input to learned output influences how GCL supports learning A. Top**,** Schematic of model task, using recorded EMGs as an input to the model GCL to predict kinematics. Bottom, MSE of model as a function of threshold when using EMG alone (MFs; blue) or GCL (red) as input to model P-cell. At a range of thresholds, P-cells that receive GCL input outperform P-cells receiving MFs alone. **B**. Example of learned kinematic position after training for MF Alone (blue line) and GCL network (red) showing good metric fit by the GCL model. **C**. Plot showing the strength of different GCL population statistical features driving learning that vary as a function of how well MFs intrinsically support learned P-cell output (MF Alone MSE). When MFs are already excellent predictors, information retention (variance retained) has a high regression slope (RIDGE regression method). When MFs are poorer intrinsic predictors, the number of explanatory PCs (a pattern separation metric) emerges as a stronger driver of learning performance. Goodness of fit (R^2) was between 0.83-0.95 across all MF-Alone MSEs used.
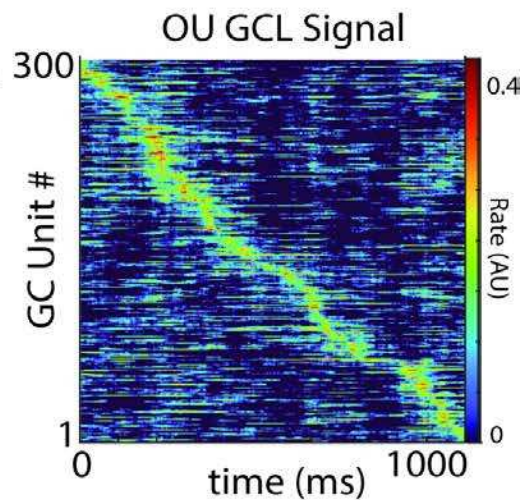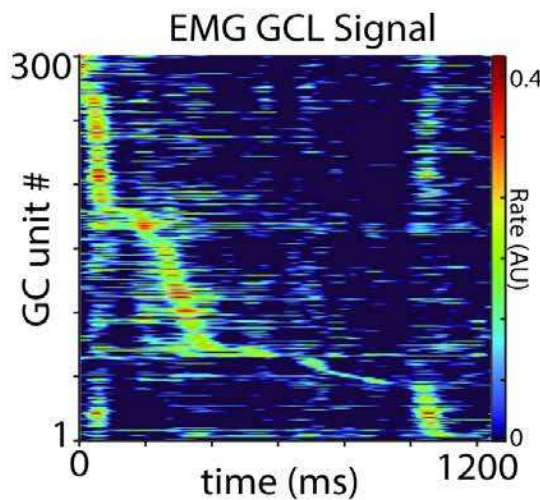
We all know that pattern separation supports classification learning, but what about temporal learning?

1. At intermediate thresholds, a model cerebellar granule layer leads to better learning in Purkinje units that mossy fibers alone.

2. Why? Basic GCL architecture improves temporal pattern separation of temporally varying inputs typical of movements.

A

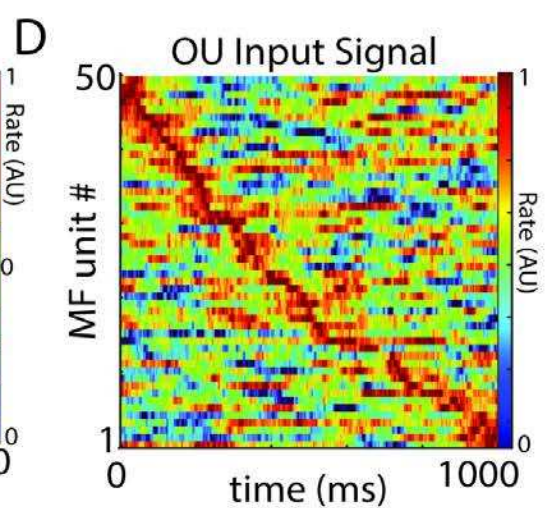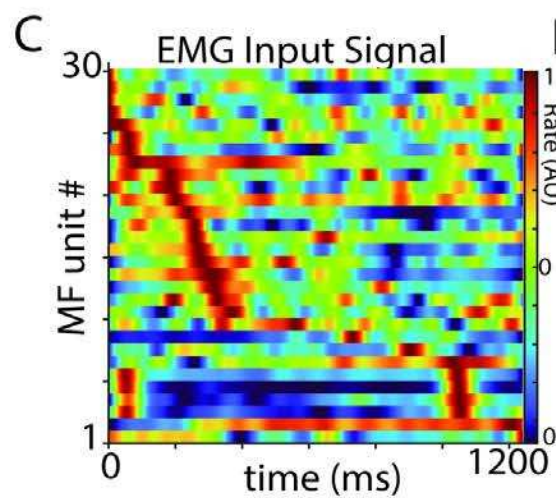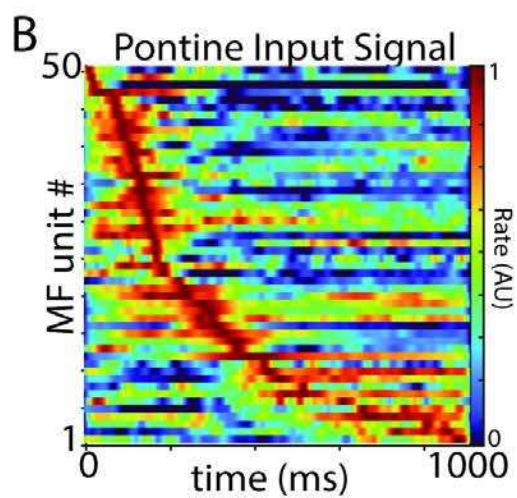Mossy Fibers (Inputs) Granule Cells Purkinje Cells

Threshold

Granule Cell Output

20
10
0

10 20 30
Input Strength

Div ≤ 1:60
Con ≥ 4:1

Target
Learned Output

Error

Signal

time (ms)

B Pontine Input Signal

50

MF unit #

1
0   time (ms)   1000

Pontine GCL Signal

300

GC unit #

1
0   time (ms)   1000

C EMG Input Signal

30

MF unit #

1
0   time (ms)   1200

EMG GCL Signal

300

GC unit #

1
0   time (ms)   1200

D OU Input Signal

50

MF unit #

1
0   time (ms)   1000

OU GCL Signal

300

GC Unit #

1
0   time (ms)   1000

A

Original Input

\+ Low Amplitude Noise (25%) =

\+ High Amplitude Noise (50%) = Noisy Inputs to GCL

B

Noiseless   25% Noise   50% Noise

GC unit #   GC unit #   GC unit #

Rate (AU)   Rate (AU)   Rate (AU)

time (ms)   time (ms)   time (ms)

C

Cumulative probability — 25% Noise

GC / MF

Time shift (ms)

Cumulative probability — 50% Noise

GC / MF

Time shift (ms)

A

Threshold z = -1   Threshold z = 0   Threshold z = 1

GC Unit #

Rate (AU)

time (ms)

GCL

MFs

MSE

Threshold z

B

Target Function
MSE: 0

MF Alone
MSE: 2e-2

Thr z: -0.75   Thr z: 0   Thr z: 0.75
MSE: .0076   MSE: .0016   MSE: .0078

A

Target
Prediction

1 Trial

5 Trials

50 Trials

Signal magnitude (AU)

Time (ms)

B

MSE

$k_{fast} = 0.385$

$k_{slow} = 0.018$

Trial number

C

fast learning component

fast k

slow learning component

slow k
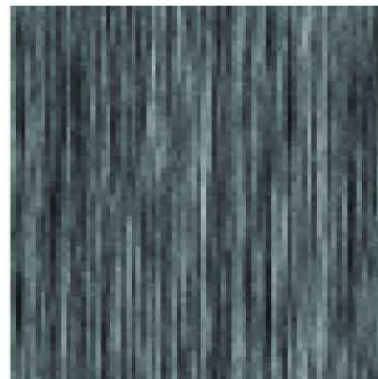
Threshold

A

mossy fiber (MF)
input layer
(M neurons)

x

granule cells (GC)
mixed layer
(N neurons)

y

reconstruction
layer
(M neurons)

x̂

ii

R²/Max(R²)

1

0.9

0.8

4 inputs

0   5   10

no. inputs per GC

B

Explained Variance

1

0.8

0.6

0.4

0.2

1
2   input(s) to
4   each GC
8

-2   0   2

Threshold

C

Explained Variance

1

0.9

0.8

0.7

0.6

0.5

1
2   input(s) to
4   each GC
8

5   10   15

GC:MF ratio

D

Explained Variance

0.8

0.6

0.4

0.2

50
100   GC outputs
500

10¹   10²   10³

MF input count

E i

Explained Variance

1

0.8

0.6

0.4

750
500
400
200

100

50

GC outputs

0   5   10

no. inputs per GC

A

EMG x30 → Input to → Model GCL → Predicts → Kinematics

B

C