Synaptic basis of a sub-second representation of time

A. Barri^{1*}, M. T. Wiechert, M. Jazayeri^{2,3} and D. A. DiGregorio^{1*}

¹ Institut Pasteur, University of Paris, CNRS UMR 3571, Synapse and Circuit Dynamics Laboratory, Paris, France.

² McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA.

³ Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA.

*Correspondence: alessandrobarri@gmail.com, david.digregorio@pasteur.fr

Abstract

Temporal sequences of neural activity are essential for driving well-timed behaviors, but the underlying cellular and circuit mechanisms remain elusive. We leveraged the well-defined architecture of the cerebellum, a brain region known to support temporally precise actions, to explore theoretically whether the experimentally observed diversity of short-term synaptic plasticity (STP) at the input layer could generate neural dynamics sufficient for subsecond temporal learning. Simulated synaptic input generated a diverse set of transient, firing patterns in granule cells (GCs) that provided a temporal basis set for learning precisely timed pauses of Purkinje cell activity associated with delayed eyelid conditioning and Bayesian interval estimation. The learning performance across time intervals was influenced by the temporal bandwidth of the GC basis, which was determined by the input layer synaptic properties. The ubiquity of STP throughout the brain positions it as a general, tunable cellular mechanism for sculpting neural dynamics and fine-tuning behavior.

Introduction

The neuronal representation of time on the sub-second timescale is a fundamental requisite for the perception of time-varying sensory stimuli, generation of complex motor plans, and cognitive anticipation of action¹⁻⁴. But how neural circuits acquire specific temporal contingencies to drive precisely timed behaviors remains elusive. A progressive increase in firing rate ("ramping") towards a threshold can represent different elapsed times by altering the slope of the ramping behavior. Elapsed time can also be encoded by a population of neurons that fire in a particular sequence ("time cells")⁵⁻⁸. Sequential synaptic connections between neurons (synfire chains) can explain the neural sequences representing bird song⁹ and contribute to time delays necessary to cancel self-generated sensory stimuli in the electrosensory lobe of mormyrid fish¹⁰. Temporal dynamics of neural population activity can also be reproduced by training recurrent neural network models¹¹⁻¹³. Nevertheless, the search for a candidate neural mechanism for generating a temporal reference (biological timer) for neural dynamics is an ongoing challenge.

STP, the rapid change in synaptic strength occurring over 10's of milliseconds to seconds, transforms patterned activity in the time domain¹⁴. Depression and facilitation of synaptic strength can act as low-and high-pass filters, respectively¹⁵, and synaptic depression can mediate gain modulation^{16,17}. Within recurrent neural networks, the long timescales of cortical synaptic facilitation provide the substrate for working memory¹⁸. Finally, low-gain recurrent networks that are supplemented by STP can also generate neural representations of time¹⁹. However, experimental evidence of STP-dependent circuit computations is rare, and is largely associated with sensory adaptation²⁰.

The cerebellar cortex (CC) is a prototypical microcircuit known to be important for generating temporally precise motor²¹ and cognitive behaviors²²⁻²⁵ on the sub-second timescale. Its largely feed-forward circuitry has been proposed to learn the temporal contingencies required for prediction from neural sequences across the population of granule cells (GCs) within the input layer²⁶ that excite the sole output of the CC, Purkinje cells (PCs). The synapses between mossy fibers (MFs) and GCs, which convey contextual information to GCs, are highly variable in their synaptic strength and STP time course²⁷. Therefore, we hypothesized that STP of MF-GC synapses could be used as internal timers for a population clock within the cerebellar cortex to generate neural dynamics necessary for temporal learning.

To elaborate this hypothesis, we modeled the CC as a two-layer perceptron network that includes realistic MF-GC connectivity and STP dynamics. The model reproduces learned PC activity associated with a well-known temporal learning task: delayed eyelid conditioning²⁸. The timescales of STP determined the temporal characteristics of the GC population activity, which in turn defined the temporal window of PC temporal learning. The width of PC pauses scaled proportionally with the learned time intervals, similarly to experimentally observed scalar variability of the eyelid conditioning behavior²⁹. Additionally, we found that STP driven GC activity was well suited to implement a Bayesian estimator of time intervals³⁰. We propose that within feed-forward circuits, dynamic synapses serve as tunable clocks that determine the bandwidth of neural circuit dynamics and enable learning temporally precise behaviors.

Results

Cerebellar Cortex model with STP

The CC can be modeled as a two-layer perceptron that performs pattern separation of static input patterns^{31–34}. When temporal processing is considered, CC models are generally augmented by an additional mechanism that generates temporally varying activity patterns in the CC's GC layer^{10,26,35,36}. To test whether heterogeneous MF-GC STP is sufficient to support temporal learning, we implemented STP of the MF-GC synapse in a CC model, hereafter named CCM_{STP}, that deliberately omits all other potential sources of temporal dynamics. In particular, the first set of simulations did not consider recurrent connectivity (**Fig 1b**). STP was simulated using a parallel vesicle pool model of the MF-GC synapse, similar to ref³⁷. It comprises two readily-releasable and depletable vesicle pools, synaptic facilitation, and postsynaptic desensitization. To reproduce the observed functional synaptic diversity, we set vesicle fusion probabilities (p_v), synaptic pool sizes (N), and synaptic facilitation to match the relative strengths, paired-pulse ratios, and transient behaviors across five different types of synapses that were previously characterized²⁷ (**Fig 1a2–a6**). Importantly, the longest timescale in CCM_{STP} is associated with a 2 s vesicle refilling time constant of the slow vesicle pool ($\tau_{ref} = 2s$, **Fig. 1a1**). To capture depression over long timescales^{37,38}, we introduced a phenomenological parameter ($p_{ref} = 0.6$) that effectively mimics a simplified form of activity-dependent recovery from depression (see Methods).

The CCM_{STP} consisted of firing rate units representing MFs, GCs, a single PC, and a single molecular layer interneuron (MLI), i.e. each neuron's activity was represented by a single continuous value corresponding to an instantaneous firing rate. Each GC received 4 MF synapses, randomly selected from the different synapse types according to their experimentally characterized frequency of occurrence²⁷. Importantly, we associated different synapse types with different MF firing rates (**Fig. 1b**, left panels, see Methods). Concretely, we associated synapses with high p_v with MF inputs with comparatively high average firing rates (primary sensory groups 1, 2) and low p_v synapses with MF inputs with comparatively low average firing rates (secondary/processed sensory groups 3, 4, 5), according to experimental observations^{39,40}. We will reconsider this relationship below.

Figure 1 Cerebellar cortex model with dynamic input layer synapses (CCM_{STP}) (**a₁**) Synaptic model scheme. (**a₂₋₆**) Properties of five synapse types. Left: Schemes show differences in presynaptic parameters; the postsynaptic side is identical for all groups. Right: average synaptic weights in response to repetitive 100 Hz stimulation as in³⁸. Insets: First five responses with paired-pulse ratio (PPR) roughly mimic results from Ref.²⁷. Color code for synapse groups is the same as in Ref.²⁷. (**b**) Scheme of CCM_{STP} . MFs are classified according to the groups in **a**. Percentages indicate relative frequency of MF groups. Insets: firing rate

distributions for different MF groups. **(c)** Simulation of CCM_{STP} with randomly drawn J_E weights. First panel: 5 sample MFs. Every second, MF activity is re-drawn from distributions in **b**. Second panel: Normalized weights of 10 MF-GC synapses. Third panel: activity of 10 sample GCs. Last panel: PC activity with different shades of gray indicating different E/I ratios onto the PC. **(d)** Same as **c** but without STP transient dynamics. Low amplitude GC and PC firing rate transients result from 10 ms GC integration time constant. **(e)** Example simulation in which correlated (black symbols) and uncorrelated (red symbols) MF patterns were presented to the network in alternation. The correlation coefficient for sequential patterns was ≈ 0.85 . Firing rates are color-coded. **(f₁)** Steady-state subtracted GC responses from simulation in **e** for uncorrelated (left) and correlated MF pattern switches (right). **(f₂)** Same as **f₁** but for PC. **(g)** Normalized standard deviation of PC transient amplitudes for switches between MF patterns of differing levels of correlation.

To examine CCM_{STP} network dynamics, input MF activity patterns were sampled every second from respective firing rate distributions shown in **Fig. 1b**. Each change in MF patterns evoked transient changes in MF-GC synaptic weights, which in turn generated transient GC firing rate responses that decayed at different rates to a steady-state (**Fig. 1c**). Similar to experimentally recorded PC responses to sensory stimuli *in vivo*⁴¹, switches between different MF patterns also generated heterogeneous transient changes in the PC firing rate, whose directions and magnitudes were controlled by the ratio of the average excitatory to inhibitory weight (**Fig. 1c, bottom**). In contrast, when MF-GC STP was removed, the transient GC and PC responses disappeared (**Fig. 1d**). The amplitude of the firing rate transients increased as the difference from one MF pattern to the next increased, similar to previous theoretical work¹⁶. Sequential delivery of uncorrelated MF firing patterns in CCM_{STP} (**Fig. 1e**) generated GC and PC transients with broadly distributed amplitudes (**Fig. 1f_{1,2}**), which were progressively reduced as the relative change in MF rate decreased (**Fig. 1g**). Thus, dynamic MF-GC synapses allow both GCs and PCs to represent the relative changes in sensory stimuli.

Simulating PC pauses during eyelid conditioning

We next explored whether MF-GC STP diversity permits learning of precisely timed PC pauses associated with the prototypical example of a CC-dependent behavior: delayed eyelid conditioning. In this behavior, animals learn to use a conditioned stimulus (CS) to precisely time a blink in anticipation of an aversive unconditioned stimulus (US). The eyelid closure is driven by a preceding decrease in PC firing rates^{28,42} (**Fig. 2a**). As the CS is typically constant until the time of the US and a precisely timed eyelid response can be learned even if the CS is replaced by direct and constant MF stimulation^{43,44}, we modeled CS delivery in the CCM_{STP} by an instantaneous switch to a novel MF input pattern that persists over the duration of the CS (**Fig. 2a**). Most GC activity transients exhibited a characteristic rapid increase or decrease in firing rate, followed by an exponential-like decay in firing rate (**Fig. 2c**). In contrast to other models of eyelid conditioning²⁶, the activity of most GCs in the CCM_{STP} peaked once during the transient period, occurring shortly (< 50 ms) after the CS onset (**Fig. 2c**). However, the distribution of GC firing rate decay times across the population was highly variable with a fraction of GCs showing decay times to 10% of the transient peak as long as 700 ms (**Fig. 2c and 2d**).

Figure 2 Simulating PC pauses during eyelid conditioning **(a)** Scheme of eyelid conditioning. CS: conditioned stimulus (red). US: unconditioned stimulus (violet). After experiencing CS and US delivered at a fixed temporal contingency over many trials, the animal learns to close its eyelid before the US is delivered (green). A trough in PC activity (blue) precedes the eyelid closure (target time, gray dashed line). **(b)** Representation of the CS by an instantaneous change in MF firing rate. Top: 100 MF sorted according to synaptic types. MF firing rates are color-coded and drawn according to the distributions shown in Fig. 1b. Bottom: two sample MF rates per synaptic group. Colors as in Fig. 1. **(c)** Model GC responses to the CS. Top: 1000 GCs sorted according to average firing rate after CS onset. Firing rates are color-coded. Bottom: steady-state subtracted and individually normalized GC transient responses. **(d)** Pdf of the distribution of GC activity decay times to 10% of the transient peak. **(e)** Example of delayed eyelid conditioning over the course of 4000 learning steps for a 200 ms delay. Dashed line represents the target time used in the supervised learning procedure. Without STP-induced GC transients, no PC pause could be learned (pink line). **(f)** Conditioned eyelid responses for different target times. Different colors indicate PC responses after 4000 learning trials for each target time (dashed lines).

To test whether the GC population dynamics could act as a basis set for learning the precisely timed PC firing rate pauses known to drive the eyelid response, we subjected the GC-PC synaptic weights to a gradient-descent based supervised learning rule. The rule's target

signal consisted of a delta peak (zero firing rate at a specific time bin) at the designated time of the PC firing rate pause (**Fig. 2e**, dotted line). In the course of learning, there was a progressive acquisition of a pause in the PC firing rate (**Fig. 2e**). However, without MF-GC STP, the PC pause was not acquired (**Fig. 2e**, pink). We tested learning of different delay intervals ranging from 25 ms to 700 ms and found that PC pauses could be generated for all delays. The PC pause amplitude and temporal precision (time and width) decreased with increasing CS-US delays (**Fig. 2f**), reminiscent of the shape of PC simple-spike pauses recorded during eyelid conditioning²⁸.

Why might the learned PC pause amplitude and temporal precision be reduced for longer CS-US delays? The parameters associated with the learning algorithm (e.g. the number of iterations) are identical for each CS-US delay. The state of the GC population activity, in contrast, changes throughout the CS. Once all GC activity dynamics reach steady-state firing rates, temporal discrimination by PCs is no longer possible, and interval learning becomes impossible. In other words, for temporal learning to be effective, changes in GC firing rates must be prominent over the relevant timescale. Indeed, in delayed eyelid conditioning simulations where slow or fast GCs were removed, the efficiency of generating PC pauses for short and long intervals was reduced (**Fig. S2**). CCM_{STP} simulations thus demonstrate that a GC temporal basis generated by MF-GC STP is sufficient to reproduce the CC computation underlying delayed eyelid conditioning and suggests that the timescale of GC dynamics influences the timescale of behavioral learning.

Analysis of the synaptic mechanism underlying GC transient responses using a reduced model

PC temporal learning requires GC transient responses, which in our model can only arise from STP at the MF-GC synapse^{27,38}. How are synaptic and GC dynamics determined by quantal and firing rate parameters? The complexity of the full CCM_{STP} with many interacting parameters makes it difficult to assess the effect of each synaptic parameter. To overcome this challenge, we developed a reduced MF-GC synapse model which was analytically solvable for an instantaneous and persistent switch of MF rates. This allowed us to identify the key computational building blocks of CCM_{STP} and understand how they control the overall behavior of the model. Specifically, we omitted short-term facilitation and postsynaptic desensitization and reduced the synaptic model to a single population of high p_v synapses (“drivers”²⁷) and a single population of low p_v synapses (“supporters”²⁷), each with a fast and a slow refilling ready-releasable pool (**Fig. 3b**), thus obtaining a model where STP results from vesicle depletion only. Each GC received exactly two driver and two supporter MF inputs with random and pairwise distinct identities (**Fig. 3a**).

In this reduced model, an instantaneous and persistent switch of MF rate change generates an average postsynaptic current ($I_{syn}(t)$) for each vesicle pool that is remarkably simple. It features a sharp transient change, followed by a mono-exponential decay to a steady-state synaptic current amplitude, A_s , (**Fig 3c**) and can be generally expressed as:

$$I_{syn}(t) = A_s + A_t e^{-\frac{t}{\tau_{syn}}}. \quad (1)$$

Here, A_s is time-invariant, and $A_t e^{-\frac{t}{\tau_{syn}}}$ is a time-dependent component with synaptic relaxation time constant τ_{syn} (**Fig. 3c**) and amplitude A_t . This transient component determines the synapse's ability to encode the passage of time

The solution of the synaptic dynamics model reveals the crucial dependence of τ_{syn} and A_t on the presynaptic and firing rate parameters (see Methods):

$$\tau_{syn} = \frac{\tau_{ref}}{1 + \alpha p_v m}. \quad (2)$$

Here, $\alpha = \tau_{ref}(1 - p_{ref})$, m is the MF firing rate persisting during the CS, and the synaptic parameters p_v , τ_{ref} , and p_{ref} are defined as above. Equation (2) shows that τ_{syn} is inversely related to the MF rate during the CS and the release probability, p_v (**Fig. 3d**). Intuitively, this is because higher p_v and/or m lead to a higher rate of synaptic vesicle fusion, and hence depletion, driving the synaptic response amplitude to steady-state faster. Conversely, slow time constants arise from low p_v and/or low m with the maximum τ_{syn} being equal to the vesicle recovery time τ_{ref} .

The transient amplitude A_t is given by:

$$A_t = \frac{N p_v m}{1 + \alpha p_v m} \frac{\alpha p_v (m - m_{pre})}{1 + \alpha p_v m_{pre}}. \quad (3)$$

Here, N is the number of release sites. Importantly, and in contrast to τ_{syn} , A_t also depends on the presynaptic MF firing rate before the CS, m_{pre} , and on the difference between the MF firing rates before and during the CS. In particular, for both rates sufficiently high, A_t becomes a linear function of the normalized difference between m and m_{pre} , i.e. $A_t \propto (m - m_{pre})/m_{pre}$ (**Fig. 3e**). A_t is sensitive to the relative and not the absolute change in presynaptic rate, as observed previously¹⁶.

Figure 3 MF-GC synaptic time constants and their relative weight determine the time course of GC responses **(a)** Scheme of GC inputs with the simplified synaptic model. Each GC receives exactly two distinct driver (red) and supporter MFs (blue). **(b)** Schemes of the reduced synaptic model of high p_v (red) and low p_v synapses (blue). **(c)** Left: scheme of single pool response with the time constant τ_{syn} (blue line) to a firing rate switch during CS presentation (black solid line). The dashed black line separates the transient (A_t) from the steady-state amplitude (A_s). Right: equations determining the synaptic time constant and synaptic input. **(d)** Slow vesicle pool time constant (τ_{syn}) versus presynaptic MF firing rate. Different shades of gray indicate different release probabilities. **(e)** Driver synapse transient amplitude (A_t) versus relative firing rate change for a baseline firing rate of 80 Hz ($(m - m_{pre})/m_{pre}$, $m_{pre} = 80$ Hz). A negative A_t corresponds to a transient decrease in firing rate. Same color code as in **d**. **(f)** Sample fast GC. Left: driver and supporter MF firing rates drawn from

*thresholded normal distributions ($\mathcal{N}_{thr}(200\text{Hz}, 15\text{Hz})$ and $\mathcal{N}_{thr}(25\text{Hz}, 15\text{Hz})$, respectively) and the corresponding synaptic responses. For clarity, only the τ_{syn} of the respective slow pool is indicated. Upper right panel: GC threshold (dashed line), total synaptic input (black line), total driver input (red line), and total supporter input (blue line). The transient response is dominated by the driver input (red). Lower right panel: resulting GC firing rate response. (**g**) Like **f** but for a sample slow GC. The transient response is dominated by the supporter input (blue).*

The transient GC activity results from the sum of eight synaptic transient current components, (i.e. four inputs, each with two pools). To illustrate the interplay between the A_t and τ_{syn} , we compared the behavior of each synaptic input for a selected fast and slow GC (**Figs. 3f and 3g**). Generally, synaptic inputs from supporters display longer transient currents than synaptic inputs from drivers (**Figs. 3f and 3g**, middle panels) due to their lower firing rates (**Figs. 3f and 3g**, left panels) and low p_v (**Fig. 3b**). A_t is largely determined by the relative change in the respective presynaptic MF rates, $(m - m_{pre})/m_{pre}$ (**Figs 3f and 3g, left panels**). Thus, as driver synapses are stronger than supporter synapses, "fast" GCs are generated when driver inputs exhibit large relative changes in firing rates (**Fig.3f**). "Slow" GCs are generated from synapses with a small relative change in driver firing rates, but large relative supporter rate changes paired with low supporter rates during the CS (**Fig.3g**). Taken together, in the reduced model τ_{syn} and A_t determine the effective timescales of the GC firing rate responses and are explicitly influenced by quantal parameters, synaptic time constants, and the diversity of MF firing rates.

The explicit influence of synaptic parameters on temporal learning

Our simulations suggest that delayed eyelid conditioning across multiple delays necessitates GC population dynamics spanning multiple timescales (**Fig. 2, Fig. S2**). Since individual GC firing rate dynamics depend on the A_t and τ_{syn} of their synaptic inputs (**Fig. 3**), this implies that, firstly, the spectrum of τ_{syn} available to the network should cover the relevant timescales, and secondly, the A_t , which can be understood as the relative weights of synaptic transient components, should be of comparable magnitude across τ_{syn} . To illustrate these points, we used the reduced CCM_{STP} to simulate eyelid response learning with different firing rate properties and examined the relationship between τ_{syn} , A_t , the GC temporal basis, and learning outcome. Importantly, since A_t and τ_{syn} are not independent, the quantity of interest is their joint distribution. We initially set up a reference simulation by choosing MF firing rate distributions such that the diversity of GC transient responses and the temporal learning performance (**Fig. 4a**) were comparable to the CCM_{STP} with native synapses (**Fig. 2f**). For this case, the joint distribution shows that A_t decreased with increasing τ_{syn} . Note that A_t is maximal when the MF firing rates increased from zero m_{pre} to a finite m upon CS onset, maximizing $m - m_{pre}$ (eqn. 3, see also **Fig. S3b,c**). We quantified learning accuracy by calculating an error based on 1) the PC response amplitude, 2) its full width at half maximum and 3) the temporal deviation of its minimum from the target delay (**Fig. 4a**, fifth panel, **Fig. S2a**, see Methods). Importantly, the degradation in temporal precision of the learned PC pauses for longer CS-US delays was concomitant with the reduction of the A_t associated with

longer τ_{syn} (**Fig. 4a**). This suggests that inspection of the joint distribution of τ_{syn} and A_t can provide insight into the temporal learning performance of the network.

Figure 4 Learning performance depends on MF rate distributions (**a**) First panel: Driver and supporter MF firing rate pdfs ($\mu_D = 200$ Hz, $\mu_S = 25$ Hz, $\sigma_D = \sigma_S = 15$ Hz). Second panel: Resulting joint A_t and τ_{syn} distribution, featuring four partially overlapping clusters, corresponding to the slow and fast pools for driver and supporter synapses, and marginal distributions. The color code of the joint distribution scales logarithmically. Colors of marginal distributions indicate driver (red) and supporter (blue) components. Third panel: Normalized GC transient responses to CS. Inset: pdf of the distribution of decay times to 10% of the transient peak. Fourth panel: PC eyelid response learning. Color code delay intervals as in (**Fig. 2e**). Fifth panel: Error for each delay, calculated based on PC response amplitude, full width at half maximum and temporal deviation (**Fig. S2a**). (**b**) Same as **a**, but with $\mu_S = 70$ Hz. Inset: black line is the pdf for simulation with $\mu_S = 70$ Hz and gray line is the pdf from **a** for $\mu_S = 25$ Hz

comparison. Fifth panel: errors relative to the ones in a. (c) Same as a, but with $\mu_D = 100\text{Hz}$ and $\sigma_D = 50\text{Hz}$. (d) Same as a, but without driver inputs.

When changing only the mean firing rate of *supporter* MFs (μ_S) from 25 Hz to 70 Hz, the synaptic time constants were shortened due to the inverse relationship between τ_{syn} and the mossy-fiber firing rate m (**Fig. 4b**, second panel). Consequently, and expectedly, the distribution of GC firing rate decay times was shifted to shorter values, and learning performance was degraded for all CS-US intervals, except the 25 ms delay (**Fig. 4b**). Lowering the mean firing rate of *driver* MFs (μ_D) from 200 Hz to 100 Hz and increasing the standard deviation (σ_D) from 15 Hz to 50 Hz, led to an overall increase of the time constants contributed by driver synapses, as well as an increase in their relative weight (A_t ; **Fig. 4c**, second panel, marginals). As a result the joint probability distribution shows a shift towards faster weighted time constants. It also follows that GC transients are accelerated, and learning precision is decreased. Removing synaptic currents originating from driver synapses only disrupted learning PC pauses for the shortest CS-US interval (**Fig. 4d**). Reduced model simulations with systematic parameter scans across a wide range of mossy fiber firing rate distributions for both synapse types suggested that good synaptic regimes for temporal learning are achieved when driver synaptic weights are comparable or smaller than those of the slow supporting synapses (**Fig. S4**).

All the results taken together suggest that optimal learning occurs when the spectrum of τ_{syn} available to the network covers behaviorally relevant timescales with balanced relative weights (A_t). Synaptic and GC activity timescales can therefore be tuned by simultaneously modulating p_v and the absolute scale of m to provide the necessary distribution of τ_{syn} , whereas the relative change of MF firing can be used to tune the weights (A_t) of τ_{syn} .

Firing rate and synaptic parameters that improve temporal learning performance

Thus far, we used the reduced model to explore how MF firing rates and synaptic properties influenced the timescales of GC activity and the temporal precision of learned PC pauses. The model, however, was constrained by 1) the use of only two synapse types, 2) fixed release probabilities (p_v), 3) MF firing rates that were consistently higher for high p_v synapses than their low p_v counterparts, and 4) an equal number of driver and supporter synapses. We next considered how the relaxation of these assumptions and specific parameter combinations could influence the precision of learned PC pauses. In particular, we simulated reduced models where, in addition to MF firing rates, p_v was sampled from continuous distributions.

Equation 2 suggests that a positive correlation between p_v and m should broaden the distribution of τ_{syn} and broaden the time window of learning. Specifically, we expect learning performance to improve when high(low) firing rate MFs are, on average, paired with high(low) p_v synapses. We chose uniformly distributed p_v and MF firing rates and split both of these equally into two contiguous groups (**Fig. 5a**). We performed training simulations in which we paired high p_v (driver) synapses with high firing rates, or we paired low p_v

(supporter) synapses and high MF firing rates, and vice versa (**Fig. 5b**). Formally, this is equivalent to adjusting the rank correlation (c_{rk}) between the p_v category (supporter or driver) and the m category (high or low, Fig. 5b). We found better learning performance when p_v and m were positively correlated (**Fig. 5c, Fig. S5**), consistent with the experimentally observed high firing rates for the driver-like synapses arising from primary vestibular afferents^{27,39} and low firing rates for the supporter-like secondary vestibular afferents^{27,40}.

Inspired by the number of synapse types observed experimentally²⁷, we augmented the number of synapse groups from 2 to 5 without changing the p_v and firing rate distributions (**Fig. 5d**). We reasoned that the introduction of a larger number of MG-GC synapse types would in principle permit a stronger *linear* correlation between p_v and m to occur (**Fig. 5e**), leading to a broader τ_{syn} spectrum (not shown) and an improved learning of PC pauses. Indeed, for high c_{rk} , the learning performance of the five group CCM_{STP} was better than that of the two-group CCM_{STP} (compare **Fig. 5c** and **Fig. 5f, Fig. S5**). These simulation results suggest that good temporal learning performance of CCM_{STP} can be achieved not simply by generating variability in parameters, but by structuring, or tuning, the relationship between p_v and m .

Figure 5 Correlating release probability and MF rates improves learning performance **(a)** Top: distribution of MF firing rates (m) used to drive the network, divided into low (supporter, green) and high (driver, yellow) rates. Bottom: Distribution of synaptic release probabilities (p_v), divided into low (light gray) and high (dark gray) probabilities. **(b)** Top: p_v versus m for 500 sample synapses for a network with strong negative rank-correlation between the m category (supporter or driver) and the p_v category (high or low). Bottom: same as top, but for strongly positive correlated m and p_v . **(c)** Resulting PC eyelid response learning for low (left) and high (right) correlation, with delay intervals color-coded as in **(Fig. 2f)**. Each curve is the average of 20 network realizations. **(d-f)** Same as **a**, but for distributions divided into five groups. **(g)** Left: MF rates and release probabilities for 5 groups where the average group firing rate is as in **d** but the firing rate variance progressively decreases with the average rate. Right: resulting PC eyelid response learning for high correlation. **(h)** Same as **g** but with zero rate MFs added to the lowest rate distribution. Dashed lines indicate additionally the case when the count of lowest rates and release probabilities is doubled.

Equipped with an understanding of how the synaptic and MF rate parameters can generate different synaptic time constants, we set out to further improve the temporal learning for longer CS-US delays by adjusting the variance of the clustered MF rate

distributions. To increase the weighting of long τ_{syn} , we inversely scaled the variance of the MF firing rate distributions with respect to the mean firing rate (**Fig. 5g**), thereby increasing A_t (**Fig. 4c**). As expected, PC pause learning was better than when using equal-width MF groups (**Fig. 5g, Fig. S5**). An additional enhancement of learning performance could be achieved by adding a small fraction of zero-rate MFs to the lowest group (**Fig. 5g**, 6% zero MFs, same fraction as in **Fig. 4a**), which provide maximal A_t (see **Fig. 4**). Finally, taking into account the experimental finding that low p_v synapses are more frequent than high p_v synapses²⁷, we doubled the fraction of MFs and release probabilities in the lowest group, resulting in the best performance of all versions of CCM_{STP} tested here (**Fig. 5g**). These simulations show that positive correlations between vesicle release probability and presynaptic firing rate broaden the temporal bandwidth of circuit dynamics and improve temporal learning.

STP permits learning optimal estimates of time intervals

Humans and animals have an unreliable sense of time. When estimating or producing time intervals, behavioral response variability scales linearly with the base interval⁴⁵. Previous work has found that humans seek to optimize their timing behavior and minimize the effect of scalar variability. A canonical example of this optimization is evident in the so-called ready-set-go-task⁴⁶ (**Fig. 6c**). In this task, subjects must estimate and subsequently reproduce different time intervals. When the intervals are drawn from a previously learned probability distribution (i.e., prior), subjects generate biased responses that are in accordance with an optimal Bayesian agent. Bayesian integration predicts that when the prior distribution is uniform, time interval production is biased towards the mean of the prior distribution, and that this bias is larger for longer intervals.

A recent study developed a CC circuit model called TRACE that could emulate the Bayesian computations during the ready-set-go task³⁰. In TRACE, learning is mediated by long-term synaptic plasticity of the GC-PC synapses in subsets of GCs that are active at the time of “set.” After learning, PC firing rates undergo prior-dependent pauses that enable the downstream dentate nucleus neurons (DNs) to represent a Bayesian estimate of the time interval. An essential feature of TRACE is that the GCs form a temporal basis set that exhibits temporal scaling. This temporal scaling, which is inherited by PCs, enables DNs to generate estimates that take into account the inherent scalar variability of timing.

In our analysis of eyelid conditioning (**Fig. 2**), we showed that CCM_{STP} generates PC firing rate pauses whose width and amplitude scale with time (**Fig. 6a**). Therefore, we reasoned that CCM_{STP} might also be able to generate biased responses consistent with Bayesian integration. To test this possibility, we analyzed the behavior of CCM_{STP} in the ready-set-go task. We treated the “ready” signal as an instantaneous switch of the MF input rates that persisted over the course of a trial (**Fig. 6c**), similar to the CS in the eyelid simulations. To learn a distribution of ready-set intervals, we generated target signals with intervals that varied from iteration to iteration of our learning algorithm (**Fig. 6c**).

Figure 6 STP generated temporal basis enables the computation of Bayesian time-interval estimates (**a**) Full width at half maximum (cyan) and normalized amplitudes (magenta) of learned PC eyelid responses versus delay interval from the experimentally constrained CCM_{STP} (see **Fig. 2**). Solid lines are linear fits. (**b**) Schemes of native synapse CCM_{STP} (left, see **Fig. 1**) and reduced model (right) with added dentate nucleus cell. (**c**) Scheme of ready-set-go task. The interval between ready and set (sample interval) is drawn from a uniform distribution at every learning iteration (gray box). The ready signal is represented by an instantaneous and persistent switch in MF firing rates (colored lines). The target signal is changed according to the set interval (dotted lines). (**d**) PC responses after 12000 iterations of RSG learning, averaged over 20 network realizations. Shaded area indicates standard deviation. Different colors represent learning of different uniform sample interval distributions. (**e**) Same as **d**, but for DN cell activity. (**f**) Rescaled DN cell activity for different learnt interval distributions (colored) and fitted theoretical Bayesian least squares (BLS) estimator (solid black line), with $w_{weber} = 0.11$ resulting from fit. (**g**) Squared deviation of rescaled DN activity from the BLS estimator for all tested intervals. (**h-k**) Same as **d-g**, but for the reduced model. The reduced model firing rate parameters were $\mu_D = 200\text{Hz}$, $\mu_S = 20\text{Hz}$, $\sigma_D = 10\text{Hz}$ and $\sigma_S = 15\text{Hz}$ and resulted in DN activity was consistent with a Bayesian least squares model with $w_{weber} = 0.08$.

We tested CCM_{STP} with five different uniform distributions of ready-set intervals (25-150 ms, 50-200 ms, 100-300 ms, 200-400 ms, 300-500 ms), resulting in PC pauses that broadened for longer interval distributions, and integrated DN activity that could easily

match the Bayesian least-square model³⁰ by adjusting a single parameter, the Weber fraction w_{weber} (see Methods; **Fig. 6d,h**). Overall, the reduced model interval estimates were closer to the Bayesian optimal estimates than for CCM_{STP} with native synaptic parameters, especially for the 200-400 ms and 300-500 ms intervals (**Fig. 6h-k**). These simulations show that a GC basis generated by MF-GC STP is sufficient for driving Bayesian-like learning of time interval estimates on timescales of several hundreds of milliseconds. It should be noted that our GC temporal basis was not explicitly constructed to accommodate scalar properties. Nevertheless, as in the TRACE model, we observed that interval estimates were biased towards the mean and that these biases were larger for longer intervals. These results suggest that a GC basis set generated from the diverse properties of native MF-GC synapses likely exhibits a scalar property necessary for generating optimal timed behaviors.

Discussion

In order to generate temporally precise behaviors, the brain must establish an internal representation of time. This theoretical study posits that the diversity of synaptic dynamics is a fundamental mechanism for encoding sub-second time in neural circuits. By using eyelid conditioning as a benchmark task for the CCM_{STP}, we elucidated the conditions under which the variability in MF-GC synaptic dynamics generates a GC temporal basis that represents elapsed time and that is sufficient for temporal learning on a subsecond scale. According to David Marr's levels of analysis of information processing systems⁴⁷, our study connects all three levels, from the circuit computation (learning timed PC pauses) to its underlying algorithm (learning with a temporal basis set), and the fundamental biological mechanism (STP).

STP diversity as a timer for neural dynamics

Cerebellar adaptive filter models posit that GCs act as a heterogeneous bank of filters that decompose MF activity into various time-varying activity patterns - or temporal basis functions - which are selected and summed by a synaptic learning rule at the PC to produce an output firing pattern that generates behaviors that minimize error signals arriving via climbing fibers^{35,36}. CCM_{STP} can be viewed straightforwardly as an adaptive filter in which MF-GC synapses act as non-linear elements whose filter properties are determined by the experimentally defined synaptic parameters and modulated by the presynaptic MF firing rates. Because of this filter bank, GCs and PCs in CCM_{STP} respond with exponential-like activity patterns to step-like changes in MF firing rates. Interestingly, hippocampal neural sequences can be described using timing models that use a bank of leaky integrators and produce a family of transient neural activity changes that exhibit variable exponential decays in response to instantaneous changes in input firing rates⁴⁸.

The use of an instantaneous and persistent change in MF activity was motivated by the fact that eyelid conditioning can be achieved if the CS is replaced with a constant MF stimulation^{43,44,49}. Recent evidence from pons recordings during reaching, suggests that MF activity can be persistent with little dynamics⁵⁰. For dynamic changes in MF rates STP is likely to generate outputs that are phase-shifted and/or the derivatives of their input⁵¹. Using heterogeneity of MF-GC STP as a mechanism for adaptive filtering, even time-varying inputs will effectively be diversified within the GC layer and improve the precision of temporal learning.

Synapses within the prefrontal cortex⁵² and at thalamocortical connections⁵³ exhibit diverse firing rate inputs⁵⁴ and release probability, and would generate synaptic dynamics that could drive complex neural dynamics. Reminiscent of PC firing rate pauses during eyeblink conditioning, hippocampal time cells are thought to be generated by a linear combination of exponentially decaying input activity patterns from upstream entorhinal cortical neurons⁶. We, therefore, speculate that STP diversity could underlie the generation of temporal context-like cells in the cerebellum (GCs) and other brain regions.

Timing mechanisms in the cerebellar cortex

In addition to MF-GC STP, the cerebellar cortex is equipped with multiple mechanisms potentially enabling temporal learning⁵⁵. Indeed, time-varying MF inputs could directly provide a substrate for learning of elapsed time, but whether the observed diversity of MF firing is sufficient to mediate temporally precise learning is unknown and merits further exploration. Within the cerebellar cortex, unipolar brush cells are thought to provide delay lines to diversify GC activity patterns^{10,56,57}, but these cell types are rare outside the mammalian vestibular cerebellum. The diversity of GC STP⁵⁸ could add to the diversity of the effective GC-layer basis set. Computer simulations have shown that recurrent GC-GoC-cell network dynamics could also generate a rich GC temporal basis set^{12,26}. However, experimental support of such circuit dynamics is lacking. When we included recurrent GC-GoC synapses in our simulations, more GCs showed a delayed onset firing (**Fig. S6**), but the synaptic time constant distribution and learning rate remained unaltered. These simulations are consistent with a “common mode” GoC cell population activity⁵⁹. Consistent with the importance of MF-GC STP, delayed eyelid conditioning was selectively altered as a result of the loss of fast EPSCs in AMPAR KO mice⁶⁰. Simulations including realistic NMDA, and spillover dynamics⁶¹ can further enrich the temporal scales available to the network⁶². MF-GC STP and other timing mechanisms described above are not mutually exclusive. They could act in concert with the diverse intrinsic properties of GCs⁶³ and PCs⁶⁴ in order to cover different timescales or increase mechanistic redundancy.

Predictions of the CCM_{STP}

Our theory makes several testable predictions. The transient response amplitude of PCs, which is proportional to the relative change in firing rate, can serve as a detector of rapid changes in MF firing patterns (novelty) and thus amplify pattern discrimination similar to that demonstrated for synapse-dependent delay coding²⁷. Consistent with this prediction, single whisker deflections have been shown to generate transient PC activity⁴¹.

The CCM_{STP} is one of the few network models to directly link quantal synaptic parameters and presynaptic activity dynamics to population activity dynamics and temporal learning. In Figures 3 and 4, we showed that the relative weight and temporal span of synaptic time constants dictate the distribution of GC firing rate decay times and, in turn, the timescales of temporal learning. Analytical solutions for simple synapse models (eqn. 3) provide insight into how synaptic parameters influence STP. For example, high levels of correlation between p_v and m , coupled with balanced relative weights of the synaptic time constants, generated a learning performance superior to the native synapses (**Fig. 5d**). Therefore, CCM_{STP} predicts that MFs forming driver synapses (high p_v) would have high baseline and stimulated firing rates, while MFs forming supporter synapses (low p_v) would exhibit low baseline and stimulated firing rates, albeit with large relative changes in firing rates. Indeed, vestibular neurons, which have been shown to exhibit high firing rates^{65,66}, produced MF-GC synapses with high release probability²⁷. In the C3 zone of the anterior lobe in cats, specific firing rates were associated with different MF types⁶⁷. It is tempting to hypothesize that nature tunes presynaptic activity and synaptic dynamics (perhaps by homeostatic or activity dependent mechanisms) in order to preconfigure the window of temporal associations required for a particular behavior.

Synaptic implementation of a Bayesian computation

Bayesian theories of behavior provide an attractive framework for understanding how organisms, including humans, optimize time perception and precise actions despite the cumulative uncertainty in sensory stimuli, neural representations, and generation of actions^{46,68}. We found that CCM_{STP} was capable of generating biased time estimates consistent with Bayesian computations. In general, the magnitude of biases for a Bayesian agent depends on the magnitude of timing variability (i.e., Weber fraction). In our simulations, model parameters that correspond to native synapses from the vestibular cerebellum produced biases that were optimal for a typical weber fraction of 0.11. However, CCM_{STP} is flexible and can be adjusted to generate optimal biases for a wide range of weber fractions. The exact relationship between model parameters and w_{weber} is an important question for future research. We note that the timescales of synaptic properties observed empirically in the vestibular cerebellum²⁷ are only suitable for generating optimal estimates for relatively short time intervals. Therefore, it remains to be seen whether the synaptic mechanisms that underlie CCM_{STP} could accommodate timing behavior for longer timescales. In this respect, one intriguing hypothesis is that synaptic parameters in different regions of the cerebellum are tuned to generate optimal estimates for different ranges of time intervals, similar to the timing variability observed for cerebellar long-term synaptic plasticity rules⁶⁹.

References

1. Broome, B. M., Jayaraman, V. & Laurent, G. Encoding and decoding of overlapping odor sequences. *Neuron* **51**, 467–482 (2006).
2. Crowe, D. A., Averbeck, B. B., Chafee, M. V. & Georgopoulos, A. P. Dynamics of Parietal Neural Activity during Spatial Cognitive Processing. *Neuron* **47**, 885–891 (2005).
3. Harvey, C. D., Coen, P. & Tank, D. W. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484**, 62–68 (2012).
4. Sauerbrei, B. A. *et al.* Cortical pattern generation during dexterous movement is input-driven. *Nature* **577**, 386–391 (2020).
5. Zhou, S., Masmanidis, S. C. & Buonomano, D. V. Neural Sequences as an Optimal Dynamical Regime for the Readout of Time. *Neuron* **108**, 651–658.e5 (2020).
6. Bright, I. M. *et al.* A temporal record of the past with a spectrum of time constants in the monkey entorhinal cortex. *Proc. Natl. Acad. Sci.* **117**, 20274–20283 (2020).
7. MacDonald, C. J., Lepage, K. Q., Eden, U. T. & Eichenbaum, H. Hippocampal “Time Cells” Bridge the Gap in Memory for Discontiguous Events. *Neuron* **71**, 737–749 (2011).
8. Pastalkova, E., Itskov, V., Amarasingham, A. & Buzsaki, G. Internally Generated Cell Assembly Sequences in the Rat Hippocampus. *Science* **321**, 1322–1327 (2008).
9. Long, M. A., Jin, D. Z. & Fee, M. S. Support for a synaptic chain model of neuronal sequence generation. *Nature* **468**, 394–399 (2010).
10. Kennedy, A. *et al.* A temporal basis for predicting the sensory consequences of motor commands in an electric fish. *Nat. Neurosci.* **17**, 416–422 (2014).
11. Laje, R. & Buonomano, D. V. Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nat. Neurosci.* **16**, 925–933 (2013).
12. Yamazaki, T. & Tanaka, S. The cerebellum as a liquid state machine. *Neural Netw.* **20**, 290–297 (2007).
13. Toyoizumi, T. & Abbott, L. F. Beyond the edge of chaos: amplification and temporal integration by recurrent networks in the chaotic regime. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **84**, 051908 (2011).
14. Dittman, J. S., Kreitzer, A. C. & Regehr, W. G. Interplay between facilitation, depression, and residual calcium at three presynaptic terminals. *J. Neurosci.* **20**, 1374–1385 (2000).
15. Abbott, L. F. & Regehr, W. G. Synaptic computation. *Nature* **431**, 796–803 (2004).
16. Abbott, L. F., Varela, J. A., Sen, K. & Nelson, S. B. Synaptic Depression and Cortical Gain Control. *Science* **275**, 220–224 (1997).
17. Rothman, J. S., Cathala, L., Steuber, V. & Silver, R. A. Synaptic depression enables neuronal gain control. *Nature* **457**, 1015–1018 (2009).
18. Mongillo, G., Barak, O. & Tsodyks, M. Synaptic Theory of Working Memory. *Science* **319**, 1543–1546 (2008).
19. Buonomano, D. V. & Maass, W. State-dependent computations: spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* **10**, 113–125 (2009).
20. Chadderton, P., Schaefer, A. T., Williams, S. R. & Margrie, T. W. Sensory-evoked synaptic integration in cerebellar and cerebral cortical neurons. *Nat. Rev. Neurosci.* **15**, 71–83 (2014).
21. Popa, L. S., Hewitt, A. L. & Ebner, T. J. Predictive and Feedback Performance Errors Are Signaled in the Simple Spike Discharge of Individual Purkinje Cells. *J. Neurosci.* **32**, 15345–15358 (2012).

22. Burguière, E. *et al.* Spatial navigation impairment in mice lacking cerebellar LTD: a motor adaptation deficit? *Nat. Neurosci.* **8**, 1292–1294 (2005).
23. Moberget, T., Gullesen, E. H., Andersson, S., Ivry, R. B. & Endestad, T. Generalized Role for the Cerebellum in Encoding Internal Models: Evidence from Semantic Processing. *J. Neurosci.* **34**, 2871–2878 (2014).
24. Gao, Z. *et al.* A cortico-cerebellar loop for motor planning. *Nature* **563**, 113–116 (2018).
25. Chabrol, F. P., Blot, A. & Masic-Flogel, T. D. Cerebellar Contribution to Preparatory Activity in Motor Neocortex. *Neuron* **103**, 506–519.e4 (2019).
26. Medina, J. F. & Mauk, M. D. Computer simulation of cerebellar information processing. *Nat. Neurosci.* **3**, 1205–1211 (2000).
27. Chabrol, F. P., Arenz, A., Wiechert, M. T., Margrie, T. W. & DiGregorio, D. A. Synaptic diversity enables temporal coding of coincident multisensory inputs in single neurons. *Nat. Neurosci.* **18**, 718–727 (2015).
28. Halverson, H. E., Khilkevich, A. & Mauk, M. D. Relating Cerebellar Purkinje Cell Activity to the Timing and Amplitude of Conditioned Eyelid Responses. *J. Neurosci.* **35**, 7813–7832 (2015).
29. White, N. E. & Kehoe, E. J. Coefficients of variation in timing of the classically conditioned eyeblink in rabbits. *J. Neurosci.* **20**, 5000–5009 (2000).
30. Narain, D., Remington, E. D., Zeeuw, C. I. D. & Jazayeri, M. A cerebellar mechanism for learning prior distributions of time intervals. *Nat. Commun.* **9**, (2018).
31. Marr, D. A theory of cerebellar cortex. *J. Physiol.* **202**, 437–470 (1968).
32. Albus, J. S. A Theory of Cerebellar Function. *Math. Biosci.* **10**, 25–61 (1971).
33. Litwin-Kumar, A., Harris, K. D., Axel, R., Sompolsky, H. & Abbott, L. F. Optimal Degrees of Synaptic Connectivity. *Neuron* (2017) doi:10.1016/j.neuron.2017.01.030.
34. Cayco-Gajic, N. A., Clopath, C. & Silver, R. A. Sparse synaptic connectivity is required for decorrelation and pattern separation in feedforward networks. *Nat. Commun.* **8**, (2017).
35. Fujita, M. Adaptive filter model of the cerebellum. *Biol. Cybern.* **45**, 195–206 (1982).
36. Dean, P., Porrill, J., Ekerot, C.-F. & Jörntell, H. The cerebellar microcircuit as an adaptive filter: experimental and computational evidence. *Nat. Rev. Neurosci.* **11**, 30–43 (2010).
37. Hallermann, S. *et al.* Bassoon Speeds Vesicle Reloading at a Central Excitatory Synapse. *Neuron* **68**, 710–723 (2010).
38. Saviane, C. & Silver, R. A. Fast vesicle reloading and a large pool sustain high bandwidth transmission at a central synapse. *Nature* **439**, 983–987 (2006).
39. Park, H. J., Lasker, D. M. & Minor, L. B. Static and dynamic discharge properties of vestibular-nerve afferents in the mouse are affected by core body temperature. *Exp. Brain Res.* **200**, 269–275 (2010).
40. Arenz, A., Silver, R. A., Schaefer, A. T. & Margrie, T. W. The Contribution of Single Synapses to Sensory Representation in Vivo. *Science* **321**, 977–980 (2008).
41. Bosman, L. W. J. *et al.* Encoding of whisker input by cerebellar Purkinje cells: Whisker encoding by Purkinje cells. *J. Physiol.* **588**, 3757–3783 (2010).
42. Ohmae, S. & Medina, J. F. Climbing fibers encode a temporal-difference prediction error during cerebellar learning in mice. *Nat. Neurosci.* **18**, 1798–1803 (2015).
43. Steinmetz, J. E., Lavond, D. G. & Thompson, R. F. Classical conditioning of the rabbit eyelid response with mossy fiber stimulation as the conditioned stimulus. *Bull. Psychon. Soc.* **23**, 245–248 (1985).

44. Khilkevich, A., Zambrano, J., Richards, M.-M. & Mauk, M. D. Cerebellar implementation of movement sequences through feedback. *eLife* **7**, (2018).
45. Gibbon, J. Scalar Expectancy Theory and Weber's Law in Animal Timing. *Psychol. Rev.* **84**, 47 (1977).
46. Jazayeri, M. & Shadlen, M. N. Temporal context calibrates interval timing. *Nat. Neurosci.* **13**, 1020–1026 (2010).
47. Marr, D. *Vision: a computational investigation into the human representation and processing of visual information*. (MIT Press, 1982).
48. Shankar, K. H. & Howard, M. W. A Scale-Invariant Internal Representation of Time. *Neural Comput.* **24**, 134–193 (2012).
49. Albergaria, C., Silva, N. T., Pritchett, D. L. & Carey, M. R. Locomotor activity modulates associative learning in mouse cerebellum. *Nat. Neurosci.* **21**, 725–735 (2018).
50. Guo, J.-Z. *et al.* Disrupting cortico-cerebellar communication impairs dexterity. *eLife* **10**, e65906 (2021).
51. Puccini, G. D., Sanchez-Vives, M. V. & Compte, A. Integrated Mechanisms of Anticipation and Rate-of-Change Computations in Cortical Circuits. *PLoS Comput. Biol.* **3**, e82 (2007).
52. Wang, Y. *et al.* Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nat. Neurosci.* **9**, 534–542 (2006).
53. Diaz-Quesada, M., Martini, F. J., Ferrati, G., Bureau, I. & Maravall, M. Diverse Thalamocortical Short-Term Plasticity Elicited by Ongoing Stimulation. *J. Neurosci.* **34**, 515–526 (2014).
54. Buzsáki, G. & Mizuseki, K. The log-dynamic brain: how skewed distributions affect network operations. *Nat. Rev. Neurosci.* **15**, 264–278 (2014).
55. Gao, Z., van Beugen, B. J. & De Zeeuw, C. I. Distributed synergistic plasticity and cerebellar learning. *Nat. Rev. Neurosci.* **13**, 619–635 (2012).
56. Zampini, V. *et al.* Mechanisms and functional roles of glutamatergic synapse diversity in a cerebellar circuit. *eLife* **5**, e15872 (2016).
57. Guo, C., Huson, V., Macosko, E. Z. & Regehr, W. G. Graded heterogeneity of metabotropic signaling underlies a continuum of cell-intrinsic temporal responses in unipolar brush cells. *Nat. Commun.* **12**, 5491 (2021).
58. Dorgans, K. *et al.* Short-term plasticity at cerebellar granule cell to molecular layer interneuron synapses expands information processing. *eLife* **8**, e41586 (2019).
59. Gurnani, H. & Silver, R. A. Multidimensional population activity in an electrically coupled inhibitory circuit in the cerebellar cortex. *Neuron* S0896627321001975 (2021) doi:10.1016/j.neuron.2021.03.027.
60. Kita, K. *et al.* *GluA4 enables associative memory formation by facilitating cerebellar expansion coding*. 2020.12.04.412023 <https://www.biorxiv.org/content/10.1101/2020.12.04.412023v1> (2020) doi:10.1101/2020.12.04.412023.
61. DiGregorio, D. A., Nusser, Z. & Silver, R. A. Spillover of glutamate onto synaptic AMPA receptors enhances fast transmission at a cerebellar synapse. *Neuron* **35**, 521–533 (2002).
62. Yamazaki, T. & Tanaka, S. A spiking network model for passage-of-time representation in the cerebellum: Cerebellar passage-of-time representation. *Eur. J. Neurosci.* **26**, 2279–2292 (2007).

63. Straub, I. *et al.* Gradients in the mammalian cerebellar cortex enable Fourier-like transformation and improve storing capacity. *eLife* **9**, (2020).
64. Johansson, F., Jirenhed, D.-A., Rasmussen, A., Zucca, R. & Hesslow, G. Memory trace and timing mechanism localized to cerebellar Purkinje cells. *Proc. Natl. Acad. Sci.* **111**, 14930–14934 (2014).
65. Sadeghi, S. G., Chacron, M. J., Taylor, M. C. & Cullen, K. E. Neural Variability, Detection Thresholds, and Information Transmission in the Vestibular System. *J. Neurosci.* **27**, 771–781 (2007).
66. Medrea, I. & Cullen, K. E. Multisensory integration in early vestibular processing in mice: the encoding of passive vs. active motion. *J. Neurophysiol.* **110**, 2704–2717 (2013).
67. Bengtsson, F. & Jorntell, H. Sensory transmission in cerebellar granule cells relies on similarly coded mossy fiber inputs. *Proc. Natl. Acad. Sci.* **106**, 2389–2394 (2009).
68. Remington, E. D., Narain, D., Hosseini, E. A. & Jazayeri, M. Flexible Sensorimotor Computations through Rapid Reconfiguration of Cortical Dynamics. *Neuron* **98**, 1005–1019.e5 (2018).
69. Suvrathan, A., Payne, H. L. & Raymond, J. L. Timing Rules for Synaptic Plasticity Matched to Behavioral Function. *Neuron* **92**, 959–967 (2016).

Online Methods

MF-GC synapse model

The synaptic weight between the j th MF and the i th GC is denoted by W_{ij} . The firing rate of the j th MF is represented by $m_j(t)$ and the average current per unit time transmitted by the synapse between GC i and MF j is:

$$I_{syn,ij}(t) = W_{ij}(t) \cdot m_j(t). \quad (1)$$

Time-dependent MF-GC synaptic weights were modeled using two ready-releasable vesicle pools³⁷, each according to the general form established by Tsodyks and Markram⁷⁰. A similar model was shown to describe accurately STP at the MF-GC synapse³⁷. Accordingly, one vesicle pool was comparatively small with high release probability and a low rate of recovery from vesicle depletion ($0.5s^{-1}$) while the other is comparatively large, with low release probability and a high rate of recovery from depletion ($20ms^{-1}$)³⁷. We refer to these pools as 'slow' and the 'fast', respectively. In the Hallermann model³⁷, the slow pool is refilled by vesicles from the fast pool. Here, for the sake of mathematical tractability, we modeled the pools as being refilled independently (see scheme in **Fig. 1**).

To model vesicle depletion we use the variables x^{slow} and x^{fast} , denoting the fractions of neurotransmitter available at the slow and fast vesicle pool, respectively. The state of the pools between GC i and MF j at time t are then described by:

$$\begin{aligned} \dot{x}_{ij}^{slow}(t) &= \frac{1 - x_{ij}^{slow}(t)}{\tau_{ref}^{slow}} - u_{ij}^{slow}(t) \cdot (1 - p_{ref}) \cdot x_{ij}^{slow}(t) \cdot m_j(t) \\ \dot{x}_{ij}^{fast}(t) &= \frac{1 - x_{ij}^{fast}(t)}{\tau_{ref}^{fast}} - u_{ij}^{fast}(t) \cdot x_{ij}^{fast}(t) \cdot m_j(t), \end{aligned} \quad (2)$$

where, τ_{ref}^{slow} and τ_{ref}^{fast} are the time-constants of recovery from vesicle depletion; these are identical across all synapses. The variables $u_{ij}^{slow}(t)$ and $u_{ij}^{fast}(t)$ denote the pools' respective release probabilities at time t . Experimental data show that, in response to trains of action potentials, MF-GC synapses approach synaptic steady state transmission with a long time-constant^{37,38}, a feature that can be captured with a serial pool model³⁷ (see scheme in **Fig. S7**). In order to capture this behavior with a parallel pool model, we added the phenomenological parameter p_{ref} to the slow pool's dynamical equation. In mechanistic terms, p_{ref} can be thought of the probability of immediately refilling a synaptic docking site after the release of a vesicle. This mechanism effectively mimics a simplified form of activity-dependent recovery from depression. The release probabilities $u_{ij}^{slow}(t)$ and $u_{ij}^{fast}(t)$ are modulated by synaptic facilitation according to:

$$\begin{aligned} \dot{u}_{ij}^{slow}(t) &= \frac{p_{v,slow}^{\alpha} - u_{ij}^{slow}(t)}{\tau_F^{\alpha}} + p_{v,slow}^{\alpha} \cdot (1 - u_{ij}^{slow}(t)) \cdot m_j(t) \\ \dot{u}_{ij}^{fast}(t) &= \frac{p_{v,fast}^{\alpha} - u_{ij}^{fast}(t)}{\tau_F^{\alpha}} + p_{v,fast}^{\alpha} \cdot (1 - u_{ij}^{fast}(t)) \cdot m_j(t). \end{aligned} \quad (3)$$

Here, $p_{v,fast}^{\alpha}$ and $p_{v,slow}^{\alpha}$ denote the release probabilities for the fast and slow pools, respectively, and τ_F^{α} is the facilitation time constant. The index α denotes different synaptic groups and runs from 1 to 5. With the above, the average number of vesicles released at any time t can be written as:

$$\begin{aligned} n_{ij}^{slow}(t) &= N_{slow}^{\alpha} \cdot u_{ij}^{slow}(t) \cdot x_{ij}^{slow}(t) \\ n_{ij}^{fast}(t) &= N_{fast}^{\alpha} \cdot u_{ij}^{fast}(t) \cdot x_{ij}^{fast}(t). \end{aligned} \quad (4)$$

Post-synaptic receptor desensitization induces an additional component of depression of phasic MF-GC synaptic transmission. As both pools share the same post-synaptic side, we model desensitization via the modulation of a single variable $q_{ij}(t)$ which represents the synaptic quantal size and which is influenced by the total number of vesicles released from both pools:

$$\dot{q}_{ij}(t) = \frac{q_0 - q_{ij}(t)}{\tau_D} - \Delta_D \cdot q_{ij}(t) \cdot \frac{n_{ij}^{slow}(t) + n_{ij}^{fast}(t)}{N_{tot}} \cdot m_j(t) \quad (5)$$

where $N_{tot}^{\alpha} = N_{slow}^{\alpha} + N_{fast}^{\alpha}$, τ_D is the time constant of recovery from desensitization, q_0 is the quantal size in absence of ongoing stimulation and Δ_D is a proportionality factor that determines the fractional reduction of $q_{ij}(t)$. As explained below, we set $q_0=1$, i.e. $q_{ij}(t)$ is normalized. Both τ_D and Δ_D are identical across all synapses. Finally, the total synaptic weight is equal to the sum of the contributions from both vesicle pools:

$$W_{ij}(t) = q_{ij}(t) \cdot (n_{ij}^{slow}(t) + n_{ij}^{fast}(t)), \quad (6)$$

Synaptic parameters for generating diverse synaptic strength and dynamics

We set the synaptic parameters of our model such as to reproduce the average behavior of the 5 MF-GC synapse groups which were determined in ref²⁷ based on unitary response current amplitudes, pair pulse ratios and response coefficients of variation.

The vesicle pool refilling time constants τ_{ref}^{slow} and τ_{ref}^{fast} were set to the values measured at the MF-GC synapse in ref.³⁷ and were identical for all synapse groups. The time constant of

facilitation τ_F^α for groups 1-4 was taken from ref.³⁸. The time constant of recovery from desensitization, τ_D , was set equal to the value reported in ref.³⁷ for all groups, and the parameters Δ_D was chosen so as to obtain the relative reduction in quantal size reported in the same reference³⁷. To qualitatively account for the slow approach to steady state transmission observed in MF-GC synapses^{37,38} we set p_{ref} to a value of 0.6 for all synapse types.

To set the presynaptic quantal parameters we firstly required that the model quantal parameters, q_0 , N and p , match the average of those measured in ref.²⁷ for each synapse group. The estimation of the experimental values $q_{0,exp}^\alpha$, N_{exp}^α and $p_{v,exp}^\alpha$ was carried out via multiple-probability fluctuation analysis²⁷, which assumes a single vesicle pool. To constrain the corresponding parameters of our two-pool model, we assumed:

$$\begin{aligned} N_{exp}^\alpha &= N_{tot}^\alpha = N_{slow}^\alpha + N_{fast}^\alpha \\ p_{v,exp}^\alpha &= \frac{N_{slow}^\alpha p_{v,slow}^\alpha + N_{fast}^\alpha p_{v,fast}^\alpha}{N_{tot}^\alpha} \end{aligned} \quad (7)$$

while keeping $p_{v,slow}^\alpha > p_{v,fast}^\alpha$. Since the quantal size did not significantly differ between groups²⁷, we set $q_0 = 1$ for all groups for simplicity. As group 4 featured almost no STP, we modeled these synapses without slow pool.

The above equations do not have a unique solution. In order to constrain the synaptic parameters further, we additionally required that the relative unitary response current amplitudes between synapse groups and their pair pulse ratios approximately equal the experimentally measured ones. To account for the fact that group 5's pair pulse ratio is larger than one, we set $\tau_F = 30$ ms for this group, as in ref.²⁷.

Finally, we extracted the relative occurrence of each synapse type from ref.²⁷.

A set of synaptic parameters that reproduces the behavior of the 5 synapse groups from ref.²⁷ which we used in **Figs. 1, 2** and **6** is summarized in the following table:

	Group 1	Group 2	Group 3	Group 4	Group 5	Ref.
N_{slow}	4	3	4	-	3	27
N_{fast}	16	12	6	10	12	27
$p_{v,slow}$	0.9	0.8	0.4	-	0.4	27
$p_{v,fast}$	0.72	0.55	0.35	0.3	0.15	27
τ_{ref}^{slow} [ms]	2000	2000	2000	-	2000	37
τ_{ref}^{fast} [ms]	20	20	20	20	20	37
τ_F [ms]	12	12	-	12	30	27,38
p_{ref}	0.6	0.6	0.6	-	0.6	-
Δ_D	0.1	0.1	0.1	0.1	0.1	37

τ_D [ms]	100	100	100	100	100	37
occurrence	6%	16%	36%	24%	16%	27

MF firing rate parameters

MF firing rate distributions of the full CCM_{STP} were set according to the broad range described in the literature^{71,72,65,66,73,40,74–77}. MFs forming synapse types 1 and 2, which convey primary sensory information, were set to high firing frequencies according to experimental observations^{40,72} (see Fig. 1b, left panels), while those firing rates for the other synapses types were lower^{75,76}. For the full model, this led to synapses with high p_v being associated with MF inputs with comparatively higher average firing rates (primary sensory groups 1, 2) and synapses with low p_v being associated with MF inputs with comparatively lower average firing rates (secondary/processed sensory groups 3, 4, 5). We chose to describe MF firing rate distributions by Gaussian distributions whose negative tails were set to zero. Means and standard deviations of the Gaussian distributions were set such that the means and standard deviations of the resulting thresholded distributions resulted in the values summarized in the following table:

	Group 1	Group 2	Group 3	Group 4	Group 5
μ [Hz]	200	200	20	20	20
σ [Hz]	20	20	20	20	20

Cerebellar cortical circuit model

The standard cerebellar cortex model with STP (CCM_{STP}) consists of firing rate units representing $M = 100$ MFs, $N = 3000$ GCs, a single PC, and a single molecular layer interneuron (MLI), i.e. each neuron's activity is represented by a single continuous value corresponding to an instantaneous firing rate. The PC is spontaneously active (40 Hz) and linearly sums excitatory inputs from GCs and inhibition from the MLI. Each GC received four MF synapses, randomly selected from the different synapse types according to their experimentally characterized frequency of occurrence²⁷. The synaptic inputs to the GCs and their firing rates are given by:

$$\begin{aligned}
 I_{gc,i}(t) &= \sum_{j \in K} I_{syn,ij}(t) = \sum_{j \in K} W_{ij}(t) m_j(t) \\
 \tau_g \dot{g}c_i(t) &= -gc_i(t) + \alpha_i \cdot \max(I_{gc,i}(t) - \theta_i, 0),
 \end{aligned} \tag{8}$$

Where the granule cell membrane time constant $\tau_g = 10$ ms. In the above equation, K is a set of four indices, randomly drawn from the indices of all MF with the constraint that at least one of these stems from a MF belonging to group 1, 2 or 5, as observed experimentally²⁷. The gain α_i and threshold θ_i are set individually for each GC i as explained below.

MLI activity is assumed to represent the average rate of the GC population, thus allowing each GC to have a net excitatory or inhibitory effect depending on the difference between the MLI-PC inhibitory weight and the respective GC-PC excitatory weight:

$$mli(t) = \frac{1}{N} \sum_{i=1}^N g c_i(t), \quad (9)$$

The synaptic weights between the i th GC and the PC and between the MLI and PC were defined as $J_{E,i}$ and J_I , respectively. The total synaptic input to the PC is thus given by:

$$\begin{aligned} I_{pc}(t) &= \sum_{i=1}^N \frac{J_{E,i}}{N} g c_i(t) - J_I mli(t) + I_{spont} \\ &= \frac{1}{N} \sum_{i=1}^N (J_{E,i} - J_I) g_i(t) + I_{spont}. \end{aligned} \quad (10)$$

Here, I_{spont} is an input that maintains the spontaneous firing of the PC at 40 Hz.

Finally, the PC firing rate is given by:

$$pc(t) = \max(I_{pc}(t), 0). \quad (11)$$

In **Fig. 1**, the GC-PC weights $J_{E,i}$ were drawn from an exponential distribution with mean equal to 1. In order to decrease or increase the ratio of the average excitatory to inhibitory weight in **Fig. 1c** and **1d**, we set $J_I = 1.025$ and $J_I = 0.975$ respectively. The synaptic, full CC model, and the reduced model described below were numerically integrated using the Euler method with step size 0.5 ms.

GC Threshold and gain adjustment: Changing the statistics of the MF firing rate distributions changes the fraction of active GCs at any given time and the average GC firing rates. To avoiding the confounding impact that co-varying these quantities has on learning performance when comparing different MF parameter sets we adjusted GC thresholds, θ_i and gains α_i such that, at steady state, the fraction of active GCs and the average GC firing rates were identical for all MF parameter choices. Specifically, we drew 1000 random MF patterns from the respective firing rate distributions and we calculated the steady inputs values of the synaptic dynamics as follows:

$$\begin{aligned} (u_{ij}^{slow,\mu})^* &= p_{v,slow}^\alpha \cdot \frac{1 + \tau_F^\alpha \cdot m^\mu}{1 + p_{v,slow}^\alpha \cdot \tau_F^\alpha \cdot m_j^\mu} \\ (u_{ij}^{fast,\mu})^* &= p_{v,fast}^\alpha \cdot \frac{1 + \tau_F^\alpha \cdot m^\mu}{1 + p_{v,fast}^\alpha \cdot \tau_F^\alpha \cdot m_j^\mu} \end{aligned} \quad (12)$$

$$\begin{aligned} (x_{ij}^{slow,\mu})^* &= \frac{1}{1 + (u_{ij}^{slow,\mu})^* \cdot \tau_{ref}^{slow} \cdot (1 - p_{ref}) \cdot m_j^\mu} \\ (x_{ij}^{fast,\mu})^* &= \frac{1}{1 + (u_{ij}^{fast,\mu})^* \cdot \tau_{ref}^{fast} \cdot m_j^\mu} \end{aligned} \quad (13)$$

$$(q_{ij}^\mu)^* = \frac{N_{tot}}{N_{tot} + \Delta_D \cdot \tau_D \cdot \left((n_{ij}^{slow,\mu})^* + (n_{ij}^{fast,\mu})^* \right) \cdot m_j^\mu} \quad (14)$$

With these, we obtained, for each individual GC, the distribution of steady state inputs and firing rates:

$$\begin{aligned} (I_{gc,i}^\mu)^* &= \sum_{j \in K} (W^\mu)_{ij}^* m_j^\mu(t) \\ (gc_i^\mu)^* &= \alpha_i \cdot \max((I_{gc,i}^\mu)^* - \theta_i, 0), \end{aligned} \quad (15)$$

We then adjusted α_i and θ_i such that, for each GC, the average steady state GC rate was 5Hz across all patterns and the fraction of patterns with non-zero GC activity was 0.2 which is within the range of experimental observation^{77,78}. Throughout the article, this adjustment was carried out every time we changed synaptic parameters (Fig. 5), the parameters of the MF firing rate distributions (Fig. 4) or the MF to synapse connectivity (Fig. 5).

Supervised learning rule: Eyelid conditioning learning was implemented by adjusting the $J_{E,i}$ via a supervised learning rule. For a given delay t_{target} we used the following target signal to produce PC pauses:

$$I_{target}(t) = I_{spont} \cdot [1 - \delta(t - t_{delay})]. \quad (16)$$

where $\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ otherwise. The target PC firing rate $I_{target}(t)$ is a Dirac pulse in which the PC rate is zero at a particular time t_{target} following the start of the CS. We quantify the deviation of the PC firing rate from the target rate by the least-squares loss E which is to be minimized during learning:

$$\begin{aligned}
 E &= \frac{1}{2} \int_{-T_{pre}}^{T_{CS}} dt \tilde{w}_{err}^2(t) \epsilon^2(t) \\
 &= \frac{1}{2} \int_{-T_{pre}}^{T_{CS}} dt \tilde{w}_{err}^2(t) (I_{pc}(t) - I_{target}(t))^2.
 \end{aligned} \tag{17}$$

Here, T_{CS} is the time interval after CS onset during which we require the PC to follow the target signal and T_{pre} a time interval before CS onset during which the PC should fire at its spontaneous rate. $\epsilon(t)$ denotes the deviation between target and actual PC output at time t . $\tilde{w}_{err}$ is a weight factor that we use to increase the impact of the target time, given by:

$$\begin{aligned}
 \tilde{w}_{err}(t) &= \frac{w_{err}(t)}{\int_{-T_{pre}}^{T_{CS}} dt' w_{err}(t')} \\
 w_{err}(t) &= \begin{cases} 3.5 & \text{if } t = t_{target} \\ 1 & \text{else} \end{cases}
 \end{aligned} \tag{18}$$

In all main figures we used $T_{CS} = 1.4s$ and $T_{pre} = 0.1s$.

GC-PC weights $J_{E,i}$ were modified during learning using gradient descent to reduce the error E at each step of the learning algorithm:

$$\begin{aligned}
 J_i &\leftarrow J_i + \Delta J_i \\
 \Delta J_i &= \eta \frac{\partial E}{\partial J_i} \\
 &= \frac{\eta}{N} \int_{-T_{pre}}^{T_{CS}} dt \tilde{w}_{err}^2(t) \cdot \epsilon(t) \cdot gc_i(t).
 \end{aligned} \tag{19}$$

Here, η is a learning rate. For our simulations, we modified this basic rule in two ways. Firstly, similar to ref.⁷⁹, we explicitly simulated a climbing fiber (CF) rate, cf , that is modulated by the error signal $\epsilon(t) = I_{pc}(t) - I_{target}(t)$ according to:

$$cf(t) = \max(cf_{spont} + \beta \epsilon(t), 0), \tag{20}$$

where cf_{spont} is the spontaneous CF rate and β a proportionality factor. The CF rate was then used to obtain the following modified the weight update:

$$\Delta J_i = \frac{\eta}{N} \int_{-T_{pre}}^{T_{CS}} dt \tilde{w}_{err}^2(t) \cdot (cf_{spont} - cf(t)) \cdot gc_i(t). \tag{21}$$

where we also set $J_{E,i} = 0$ when a learning iteration resulted in a negative weight. As the CF rate is required to be positive or zero, this formulation limits the error information that is transmitted to the PC as compared to the simple gradient rule. Note that this learning rule yields synaptic long-term depression when CF and GC are simultaneously active and long-term potentiation when GC are active alone, consistent with experimental data on GC-PC synaptic plasticity⁵⁵.

Furthermore, recent experimental findings suggest that temporal properties of GC-PC plasticity rules are tuned such as to compensate for the typical delays expected for the time for error information to arrive in the cerebellar cortex⁶⁹. Here, we did not explicitly model CF error information delays, and for the sake of simplicity, directly modeled the timing of PC activity to show that the GC basis set is sufficient to generate an appropriately timed PC pause.

To increase the learning speed, we added a Nesterov acceleration scheme to equation (21)⁸⁰, introducing a momentum term to the gradient, making weight updates during a given iteration of the algorithm dependent on the previous iteration. The implementation we chose additionally features an adaptive reset of the momentum term, improving convergence properties⁸⁰. This addition does not reflect biological mechanisms and is used mainly for practical convenience.

We used simulated GC activity that was subsampled by a factor of 10 and set $\eta = 0.0025, \beta = 0.5$ and the initial distribution of weights to $J_{E,i} = J_I = 10$ for all i . For all eyelid response learning simulations, we chose $cf_{spont} = 1\text{Hz}$ (**Figs. 2, 4, 5**).

Error measure of learned Purkinje cell pause: We defined the error between the PC pause and the I_{target} (see Figs 4, S3, S4 and S5) in the following way:

$$\epsilon_{tot} = \left(1 - \frac{\epsilon_{amp}}{h_{spont}}\right) + \frac{\epsilon_{fwhm}}{s} + 5 \cdot \frac{\epsilon_t}{s}. \quad (22)$$

The first term depends on the amplitude of the PC pause relative to baseline firing, yielding a small error when the amplitude goes to zero. The second term corresponds to the normalized width of the PC pause. Finally, the third term is the normalized deviation of the pause's minimum from the target time, ϵ_t . To increase the importance of this term we scaled it by a factor 5. The error measure in Figs. S4 and S5 is the sum of ϵ_{tot} over all tested delays.

Reduced CC model

In the reduced synaptic model, we neglect facilitation and desensitization yielding constant release probabilities and normalized quantal size:

$$\begin{aligned} u_{ij}^{slow}(t) &= p_{v,slow}^{\alpha} \\ u_{ij}^{fast}(t) &= p_{v,fast}^{\alpha} \\ q_{ij}(t) &= 1. \end{aligned} \quad (23)$$

We obtain for the vesicle pool dynamics:

$$\begin{aligned} \dot{x}_{ij}^{slow}(t) &= \frac{1 - x_{ij}^{slow}(t)}{\tau_{ref}^{slow}} - p_{v,slow}^{\alpha}(1 - p_{ref})x_{ij}^{slow}(t) \cdot m_j(t) \\ \dot{x}_{ij}^{fast}(t) &= \frac{1 - x_{ij}^{fast}(t)}{\tau_{ref}^{fast}} - p_{v,fast}^{\alpha}x_{ij}^{fast}(t) \cdot m_j(t). \end{aligned} \quad (24)$$

and the total synaptic weight becomes:

$$W_{ij}(t) = N_{slow}^{\alpha} \cdot p_{v,slow}^{\alpha} \cdot x_{ij}^{slow}(t) + N_{fast}^{\alpha} \cdot p_{v,fast}^{\alpha} \cdot x_{ij}^{fast}(t). \quad (25)$$

Here the index α denotes membership in the driver or supporter category. The synaptic currents of the reduced model are computed as in the full model. Each GC receives exactly two driver and two supporter MF inputs with random and pairwise distinct identities. To eliminate any non-synaptic dynamics from the reduced model, we removed the GC membrane time constant yielding GC dynamics that follow the synaptic input instantaneously:

$$gc_i(t) = \alpha_i \cdot \max(I_{gc,i}(t) - \theta_i, 0). \quad (26)$$

Finally, GC threshold and gain adjustments were carried out as for the full CCM_{STP} where instead of equations (12) and (14) we used equations (23).

Synaptic parameters of the reduced model: The parameters of the reduced model we set to create two synapse types that capture the essence of the experimentally observed synaptic behavior: a strong and fast driver synapse, and a weak and slow supporter synapse. All synaptic parameters of the model used in **Fig. 3, 4** and **6** are summarized in the following table:

	Drivers	Supporters
N_{slow}	3.5	4
N_{fast}	14	6
$p_{v,slow}$	0.8	0.4

$p_{v,fast}$	0.6	0.2
τ_{ref}^{slow} [ms]	2000	2000
τ_{ref}^{fast} [ms]	20	20
p_{ref}	0.6	0.6
occurrence	50%	50%

In **Fig. 5**, firing rates and release probabilities were randomly drawn from uniform distributions. In detail, the release probabilities of the slow pool, $p_{v,slow}$, were drawn from distributions with a lower and upper bound of 0.1 and 0.9, respectively, (**Fig. 5a, d, g and h**), and the corresponding release probabilities of the fast pool were calculated according to $p_{v,fast} = \frac{2}{3}p_{v,slow}$, keeping them strictly lower. Lower and upper bounds of the distribution of firing rates used in panels **a** and **d** were 5 Hz and 270 Hz, resulting in firing rate standard deviations of $\sigma_{rate} \approx 38.2$ Hz for the two-groups case (**Fig. 5a**) and $\sigma_{rate} \approx 15.3$ Hz for the five-groups case (**Fig. 5d**). The bounds of the distributions in panels **g** and **h** were chosen such as to produce average group firing rates equal to those in panel **d** and firing rate standard deviations that increased with the group index, i.e. $\sigma_{rate} \approx \{5.0, 7.6, 10.2, 12.7, 15.3\}$ Hz for groups 1 to 5, respectively. Finally, the sizes of the slow vesicle pool were fixed at $N_{slow} = 4$ and the size of the fast vesicle pools were set to decrease with the group index, i.e. $N_{fast} = \{16, 6\}$ for the two-groups case, and $N_{fast} = \{16, 12, 8, 6, 6\}$ for the five-groups case. Finally, the desired rank correlation between p_v identities and MF identities was achieved by creating a Gaussian copula reflecting their statistical dependency and reordering the marginal p_v and MF distributions accordingly.

Derivation of τ_{syn} and A_t :

In the reduced model, we derived an analytical solution to the synaptic current driving a GC in response to the CS. Since the equations describing slow and fast vesicle pool dynamics are formally very similar, we describe the derivation for a single slow pool only. Additionally, we suppress all indices for the sake of readability. We assume that the MF rate $m(t)$ switches instantaneously from v_{preCS} to v_{CS} at time $t' = 0$. Integration of equations (24) from $t' = 0$ to t yields:

$$x(t) = (x_{preCS}^* - x_{CS}^*) \exp\left(-\left(\frac{1}{\tau_{ref}} + p_v(1 - p_{ref})v_{CS}\right)t\right) + x_{CS}^*, \quad (27)$$

Here, x_{preCS}^* and x_{CS}^* denote the steady state values of x before (preCS) and after (CS) the firing rate switch. They are given by:

$$x_{\gamma}^* = \frac{1}{1 + \alpha p_v v_{\gamma}}, \quad (28)$$

with

$$\alpha = \begin{cases} \tau_{ref}(1 - p_{ref}) & \text{for slow pool} \\ \tau_{ref} & \text{for fast pool} \end{cases} \quad (29)$$

From equation (27) we can read off the synaptic time constant that governs the speed of transition from steady state a value before the CS to a steady state during the CS:

$$\tau_{syn} = \tau_{ref} \cdot x_{CS}^* = \frac{\tau_{ref}}{1 + \alpha p_v v_{CS}} \quad (30)$$

This equation is similar to one derived previously^{51,81}. The total synaptic current per unit time for a single pool during the CS is given by:

$$I_{syn}(t) = N p_v x(t) v_{CS} \quad (31)$$

Combining equations (27) and (31) we get:

$$\begin{aligned} I(t) &= \frac{N p_v v_{CS}}{1 + \alpha p_v v_{CS}} \left[1 + \frac{\alpha p_v (v_{CS} - v_{preCS})}{1 + \alpha p_v v_{preCS}} \exp\left(-\frac{t}{\tau_{syn}}\right) \right] \\ &= \underbrace{A_s}_{\text{steady state}} + \underbrace{A_t}_{\text{transient amplitude}} \exp\left(-\frac{t}{\tau_{syn}}\right). \end{aligned} \quad (32)$$

Thus, the transient amplitude for a single vesicle pool is:

$$A_t = \frac{N p_v v_{CS}}{1 + \alpha p_v v_{CS}} \frac{\alpha p_v (v_{CS} - v_{preCS})}{1 + \alpha p_v v_{preCS}}. \quad (33)$$

For a single synapse, the total transient amplitude is the sum of the individual fast pool and slow pool transients:

$$A_t^{tot} = A_t^{slow} + A_t^{fast}. \quad (34)$$

To generate the surface plots in **Fig. 4** and **Fig S3** we generated 10^5 firing rates from the driver and supporter MF rate distributions, respectively, and used equations (30), (33) and (34) to calculate the corresponding values of the A_t and τ_{syn} . From these, the plots of the joint A_t and τ_{syn} distribution and the marginal distributions were generated using a two- or one-dimensional kernel density estimator, respectively⁸². Note that, formally, τ_{syn} is maximal when $v_{CS} = 0$. In that case, however, there is no synaptic transmission as $A_t^{tot} = A_t^{slow} =$

$A_t^{fast} = 0$. When plotting the joint A_t - τ_{syn} distribution in **Fig. 4** and **Fig S3**, we therefore omitted time-constants and transient amplitudes corresponding to $\nu_{CS} = 0$.

Bayesian estimation of time intervals

In CCM_{STP} simulations, the “ready” signal was treated as an instantaneous switch of the MF input rates that persisted over the course of a trial (**Fig. 6c**), similar to the CS in the eyelid learning simulations. To learn a distribution of ready-set intervals, we generated target signals within intervals that varied from iteration to iteration of our learning algorithm (**Fig. 6c**). We performed simulations with the following five different uniform distributions of ready-set intervals: 25-150 ms, 50-200 ms, 100-300 ms, 200-400 ms, 300-500 ms. Learning was carried out separately for each interval and for 12000 iterations. We found that in order to achieve the correct biases for the two longest intervals, we had to introduce a higher CF baseline firing rate, $cf_{spont} = 5$ Hz, possibly indicating a higher importance of LTP for this task. The other learning parameters were kept the same as in the eyelid learning simulations.

In keeping with ref.³⁰, we modeled the DN neuron an integrator, whose rate was calculated according to:

$$dn(t) = \int (I_{ext} - J_{pc} pc(t)) dt, \quad (35)$$

where the J_{pc} is the weight of the inhibitory PC-DN synapse and $I_{ext} = \langle pc \rangle$ is an external excitatory input to DN. It was set equal to the average PC firing rate during the ready-set period in order to ensure that excitation and inhibition onto the DN are of comparable size. For simplicity we set $J_{pc} = 1$.

In order to map the DN rate to a time axis (**Fig. 6f,j**), we rescaled every individual DN output curve according to:

$$\widehat{dn}(t) = (t_{s,max} - t_{s,min}) \frac{dn(t) - dn_{min}}{dn_{max} - dn_{min}} + t_{s,min}, \quad (36)$$

where $t_{s,max}$ and $t_{s,min}$ are the maximum and minimum of the respective ready-set interval and dn_{max} and dn_{min} are the maximum and minimum values of the DN firing rate. Since the transformation (36) is linear, the essential features exhibited by the DN firing rate (i.e. its biases) are preserved.

To show how the theoretical Bayesian least squares (BLS) interval estimate can be obtained, we follow the reasoning from ref.³⁰. It is assumed that, in the ready-set-go task, subjects perform a noisy measurement, t_m , of the ready-set interval, t_s , according to:

$$p(t_m|t_s) = \frac{1}{\sqrt{2\pi(w_{weber}t_s)^2}} e^{-\frac{(t_s-t_m)^2}{2(w_{weber}t_s)^2}}. \quad (37)$$

Note that the standard deviation of the estimate of t_m increases with the length of the interval t_s with proportionality factor w_{weber} , which is the weber fraction. Given the distribution of time intervals in the ready-set-go task, $\Pi(t_s)$, the Bayesian estimate of t_s given t_m is:

$$p(t_s|t_m) \propto \Pi(t_s)p(t_m|t_s). \quad (38)$$

The BLS estimate is the expected value of the previous expression:

$$t_e = E[p(t_s|t_m)]. \quad (39)$$

We performed a least-squares fit of the BLS model to the CCM_{STP} outputs (from all five interval distributions simultaneously) with w_{weber} as a single free parameter.

Recurrent Golgi cell inhibition

To probe the effect of recurrent inhibition in the reduced CCM_{STP}, we added one Golgi cell (GoC) that received excitatory inputs from all GCs and formed inhibitory synapses onto all GCs. For simplicity, we assumed that the GoC fires with a rate goc equal to the average GC firing rate, similarly to the MLI, and that all GoC to GC synapses have identical weights, J_{goc} :

$$\begin{aligned} goc(t) &= \frac{1}{N} \sum_{i=1}^N gc_i(t) = \langle gc(t) \rangle \\ I_{gc,i}(t) &= \sum_{j \in K} W_{ij}(t) m_j(t) - J_{goc} \cdot goc(t) \\ gc_i(t) &= \alpha_i \cdot \max(I_{gc,i}(t) - \theta_i, 0). \end{aligned} \quad (40)$$

The above equations imply that, in this configuration, the GoC acts as an activity-dependent GC threshold.

To ensure that the overall GC activity level in the reduced CCM_{STP} with GoC inhibition is comparable to the case without, we require the same criterion as above: an average GC rate of 5Hz and a fraction of activated GCs of 0.2 in steady state. Since the average GC input now depends on the average GC firing rate itself, manual adjustment of GC thresholds, θ_i , and gains, α_i , carried out above, is not feasible.

Instead, a steady-state solution of equations (40) satisfying our requirements has to be found numerically. We first set up the CC network without the GoC and adjusted GC thresholds, θ_i , and gains, α_i , according to the procedure described above. Note that in the reduced model, due to every GC receiving the same combination of inputs (i.e. 2 supporter and two driver inputs), both θ_i and α_i are similar across GCs. We thus made the additional simplification of setting $\theta = E(\theta_i)$ and $\alpha = E(\alpha_i)$ for all GCs. We then reduced GC thresholds by 10% and introduced the GoC.

To obtain the average steady state GC firing-rate we assumed that the synaptic currents of a single GC are normally distributed across MF input patterns, or, equivalently, across GCs. Mean and variance of the GC inputs are:

$$\begin{aligned}\langle I_{gc}^* \rangle &= E(I_{gc,i}^*) = E\left(\sum_{j \in K} W_{ij}^* \cdot m_j\right) - J_{goc} \cdot \langle gc^* \rangle \\ \sigma_{I^*}^2 &= \text{Var}(I_{gc,i}^*) = \text{Var}\left(\sum_{j \in K} W_{ij}^* \cdot m_j\right)\end{aligned}\quad (41)$$

We can then express the average GC firing rate in the $N \rightarrow \infty$ limit as:

$$\langle gc^* \rangle = \alpha \int_{-\infty}^{+\infty} \max(\langle I_{gc}^* \rangle + \sigma_{I^*} \cdot \xi - \tilde{\theta}, 0) \exp\left(-\frac{\xi^2}{2}\right) \frac{d\xi}{\sqrt{2\pi}} \quad (42)$$

where $\tilde{\theta} = 0.9 \theta$. The fraction of active GCs f can be written as:

$$f = \frac{1}{2} \text{erfc}\left(\frac{\theta - \langle I_{gc}^* \rangle}{\sqrt{2} \sigma_{I^*}}\right) \quad (43)$$

We can now impose that

$$\begin{aligned}\langle gc^* \rangle &= 5 \text{ Hz} \\ f &= 0.2\end{aligned}\quad (44)$$

and find a self-consistent solution of equations (41), (42) and (43) by adjusting the parameters J_{goc} and α . To do so we used the hybrid numerical root-finder from the GNU scientific library⁸³ with default step size.

Software implementation

Figures were generated using Matlab. All simulations were performed using custom C++ code and are available on a GitHub repository ().

Methods References

70. Markram, H., Wang, Y. & Tsodyks, M. Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. Natl. Acad. Sci.* **95**, 5323–5328 (1998).
71. Van Kan, P. L., Gibson, A. R. & Houk, J. C. Movement-related inputs to intermediate cerebellum of the monkey. *J. Neurophysiol.* **69**, 74–94 (1993).

72. Park, H. J., Lasker, D. M. & Minor, L. B. Static and dynamic discharge properties of vestibular-nerve afferents in the mouse are affected by core body temperature. *Exp. Brain Res.* **200**, 269–275 (2010).
73. Beraneck, M. & Cullen, K. E. Activity of Vestibular Nuclei Neurons During Vestibular and Optokinetic Stimulation in the Alert Mouse. *J. Neurophysiol.* **98**, 1549–1565 (2007).
74. Dale, A. & Cullen, K. E. The nucleus prepositus predominantly outputs eye movement-related information during passive and active self-motion. *J. Neurophysiol.* **109**, 1900–1911 (2013).
75. Van Dijk, G. *et al.* Probabilistic Identification of Cerebellar Cortical Neurones across Species. *PLoS ONE* **8**, e57669 (2013).
76. Muzzu, T., Mitolo, S., Gava, G. P. & Schultz, S. R. Encoding of locomotion kinematics in the mouse cerebellum. *PLoS ONE* **26** (2018).
77. Chen, S., Augustine, G. J. & Chadderton, P. Serial processing of kinematic signals by cerebellar circuitry during voluntary whisking. *Nat. Commun.* **8**, 232 (2017).
78. Giovannucci, A. *et al.* Cerebellar granule cells acquire a widespread predictive feedback signal during motor learning. *Nat. Neurosci.* (2017) doi:10.1038/nn.4531.
79. Clopath, C., Badura, A., De Zeeuw, C. I. & Brunel, N. A Cerebellar Learning Model of Vestibulo-Ocular Reflex Adaptation in Wild-Type and Mutant Mice. *J. Neurosci.* **34**, 7203–7215 (2014).
80. O'Donoghue, B. & Candes, E. Adaptive restart for accelerated gradient schemes. *Found. Comput. Math.* **15**, 715–732 (2015).
81. Goldman, M. S., Maldonado, P. & Abbott, L. F. Redundancy reduction and sustained firing with stochastic depressing synapses. *J. Neurosci.* **22**, 584–591 (2002).
82. Botev, Z. I., Grotowski, J. F. & Kroese, D. P. Kernel density estimation via diffusion. *Ann. Stat.* **38**, (2010).
83. Galassi, M. & Theiler, J. *GNU Scientific Library Reference Manual (3rd Ed.)*.

Supplementary Figures

Supplementary Figure 1 Behavior of model synapses subjected to different stimulation protocols. Each row represents one of the five synapse types from **Fig. 1**. First column: average synaptic weight in response to 300 Hz train followed by increasing intervals ranging from 25 ms to 5 s as in ref.³⁷. Inset: zoom on 300 Hz train. Second and third column: average synaptic weight in response to trains of 26 stimuli at 20 Hz and 100 Hz, respectively, similar to ref.³⁸

Supplementary Figure 2 (a) Definition of error components used to assess learning performance. **(b)** Learning error as a function of MLI-PC inhibition strength. The total error is computed from the individual components according to the formula shown. Black line represents the average over 20 realizations of the network in **Fig.2**. Error bars are SEM. **(c)** Same as **b** for learning error as a function of number of learning iterations. **(d₁)** As in **Fig.2e**, but without GCs whose transients decayed to 10% of their peak values in more than 150 ms. **(d₂)** Learning performance when using the temporal basis from panel **d₁**. **(e₁)** As in **Fig.2e**, but without GCs whose transients decayed to 10% of their peak values in less than 150 ms. **(e₂)** Learning performance when using the temporal basis from panel **e₁**. Note that learning of short delays was not completely abolished because our manipulations did not affect the rising phases of the GC transients.

Supplementary Figure 3 Additional examples of MF rate distributions and their impact on learning **(a)** Same as Fig. 4a (i.e. reference case). **(b)** Without zero firing rates. **(c)** Simulations in which supporter inputs are set to zero before the CS and switch to finite rates during the CS. **(d)** Simulations in which GC thresholds are set as in **a** (blue and dashed firing rate distributions), but eyelid conditioning is carried out with lower driver inputs (blue and red firing rate distributions).

Supplementary Figure 4 Scan of MF firing rate parameters show distinct roles for driver and supporter inputs **(a)** Definition of error components used to assess learning performance. The total error is computed from the individual components according to the formula shown. **(b)** Definition of the MF firing rate parameters. μ_D , μ_S and σ_D , σ_S denote means and standard deviations of the MF firing rate distributions of drivers (red) and supporters (blue) respectively. **(c)** Scan over driver parameters while keeping supporter parameters fixed. **(d)** Scan over supporter parameters while keeping driver parameters fixed. In **c** and **d** the total learning error is color coded and parameter configurations corresponding to rows in **Fig. 4** are indicated by black dots.

Supplementary Figure 5 Average total error over seven delay intervals for varying rank-correlation between the m category and the p_v category for all scenarios shown in **Fig. 5**.

Supplementary Figure 6 Eyelid learning is not significantly affected by simple GoC feedback (**a₁**) Scheme of CC input layer with GoC inhibition that acts in a purely feed-forward manner. This configuration is functionally identical to the reduced model used in the main text. (**a₂**, **a₃**) GC temporal basis (as in **Fig. 2c**). White line in **a₂** indicates GoC activity. (**a₄**) PC eyelid response learning. (**b₁₋₄**) Same as row **a** but with GC-GoC feedback connections. (**c₁₋₄**) Same as row **b** but without MF-GC STP.

Supplementary Figure 7 Scheme comparing the parallel pool model used here with the serial pool model from ref.³⁷. The symbols k_0^+ , k_0^- , k_1^+ , k_1^- and k_2 are rate constants described in ref.³⁷.