Adaptive Cerebellar Feedback Enables Multi-Task Learning Through Flexible Reorganization of Cortical Dynamics

Anonymous authors
Paper under double-blind review

Abstract

Accumulating evidence suggests that high-dimensional cortical population activity is organized along low-dimensional latent dynamics that support flexible representations and computations. However, existing theoretical models often rely on explicit architectural constraints, such as low-rank recurrent connectivity, to induce structured dynamics, leaving unresolved the question of how such low-dimensional organization may emerge through biologically plausible mechanisms. Inspired by experimental advances in cortico-cerebellar interactions, we propose a cerebellar adaptive feedback mechanism that enables the emergence of low-dimensional cortical dynamics without imposing constraints on recurrent connectivity. We develop a cerebro-cerebellar coordination framework in which a frozen full-rank cortical recurrent neural network (RNN) preserves high-dimensional representational capacity, while an adaptive cerebellar pathway provides task-dependent feedback to regulate cortical dynamics through biologically plausible local plasticity. This interaction reorganizes cortical activity into low-dimensional and separable neural manifolds, reducing representational interference during multi-task learning. We evaluate this framework on cognitive rule learning and handwritten-letter trajectory prediction tasks, demonstrating that adaptive cerebellar feedback effectively improves multi-task learning performance. These findings reveal a biologically grounded computational principle for organizing flexible neural dynamics.

1 Introduction

Extensive neural recordings indicate that population activity during cognition and motor control is often governed by low-dimensional latent dynamics embedded in a much higher-dimensional neural state space (Churchland et al., 2012; Mante et al., 2013; Gallego et al., 2017; Cunningham & Yu, 2014; Shenoy et al., 2013; Pandarinath et al., 2018; Duncker & Sahani, 2021; Vyas et al., 2020; Jazayeri & Ostojic, 2021). Such low-dimensional organization confers several computational advantages. By separating task-relevant latent variables from neuron-specific variability, it can reduce representational redundancy, increase robustness to noise, and enable stable downstream readout (Perich et al., 2025; Langdon et al., 2023; Gallego et al., 2020; Stavisky et al., 2017; Rule et al., 2020). In addition, the geometry of low-dimensional organization can facilitate efficient learning by reducing interference between competing representations and segregating them into approximately orthogonal subspaces (Vyas et al., 2020; Wakhloo et al., 2026; Weber et al., 2023; Flesch et al., 2022; Goudar et al., 2023). Similar phenomena are observed in artificial neural networks: object-manifold geometry constrains classification capacity, and low-rank recurrent connectivity can yield interpretable low-dimensional dynamics even in large models(Duncker & Sahani, 2021; Mao et al., 2024; Cohen et al., 2020; Wakhloo et al., 2026; Pezon et al., 2026).

Despite these insights, many computational models achieve low-dimensional dynamics by imposing engineered architectural constraints, such as low-rank connectivity or explicit manifold regularization (Costacurta et al., 2024; Beiran et al., 2023; Pezon et al., 2026; Pollock & Jazayeri, 2020). While these approaches are analytically valuable and clarify why low-dimensional dynamics can support memory representation and downstream readout in

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

multi-task learning, they leave a central question unresolved: can such low-dimensional control emerge through flexible, adaptive mechanisms?

A longstanding view in systems neuroscience is that a region's dynamics are not determined solely by its local circuitry, but are continuously shaped by inputs from other brain regions (Rosen & Freedman, 2026; Gozel & Doiron, 2024; Allen & Morishita, 2024; Kim et al., 2023; Shen et al., 2022). In this context, reciprocal projections between the cerebellum and multiple cortical areas have received increasing attention (Kelly & Strick, 2003; Bostan et al., 2013; Buckner et al., 2011). Functionally, beyond its established role in motor control and learning, accumulating electrophysiological and fMRI evidence implicates the cerebellum in cognitive functions (Middleton & Strick, 1994; Ito, 2008; Buckner, 2013; Prati et al., 2024; Zang & De Schutter, 2023). Through widespread bidirectional connectivity, the cerebellum can monitor, transform, and feed back cortical activity, thereby influencing cortical dynamics (Gao et al., 2018; Zhu et al., 2023; Israely et al., 2025; Wagner & Luo, 2020; Bansal et al., 2024; Hur et al., 2024; Bostan & Strick, 2018). Prior work has emphasized cerebellar contributions to predictive feedback, error correction, and stabilization of cortical activity (Raymond et al., 1996; Sokolov et al., 2017; Bracco et al., 2026; Pemberton et al., 2024; Boven et al., 2023; Li & Mrsic-Flogel, 2020). Recent studies further suggest that cortico-cerebellar coupling supports flexible cognition and that cerebellum-inspired principles enhance robust generalization through structured representation transformation (Bao et al., 2026; Zhang et al., 2026). Nonetheless, we hypothesize that its conserved circuit architecture and well-characterized plasticity mechanisms may implement a general computation: adaptively transforming high-dimensional cortical states into structured feedback signals that scaffold task-relevant latent dynamics.

The main contributions of our work are as follows:

- Adaptive cortico-cerebellar feedback enables biologically plausible regulation of cortical dynamics. We introduce a cortico-cerebellar coupling framework in which a frozen full-rank cortical recurrent neural network is coupled with an adaptive cerebellar pathway that processes cortical activity and returns structured feedback. Rather than imposing fixed low-dimensional constraints on cortical connectivity, the cerebellar pathway learns through circuit motifs and local plasticity rules to extract task-relevant structure and dynamically reshape cortical activity, providing a biologically plausible mechanistic account of flexible multi-task representation and learning.
- Cerebellar feedback induces emergent low-dimensional cortical dynamics and reduces multi-task interference. We show that adaptive cerebellar feedback reduces the effective dimensionality and cross-task overlap of cortical representations, thereby mitigating interference and improving multi-task learning. The resulting dynamics retain the expressive capacity of full-rank recurrent computation while acquiring the robustness and efficiency of low-dimensional latent control, linking cortical recurrent dynamics, cerebellar feedforward adaptation, and cortico-cerebellar feedback within a unified mechanism for multi-task learning.

2 Related work and our contributions

Cortico-cerebellar coupling. The cerebellum has long been conceptualized as an adaptive system that supports learning by providing predictive, error-driven feedback (Raymond & Medina, 2018). Recent computational work has revisited this idea in cortico-cerebellar architectures where a cerebellar module predicts future feedback or teaching signals to address temporal credit assignment in the presence of delayed, sparse, or intermittent supervision (Pemberton et al., 2024). By learning to anticipate future errors or outcomes, these models can substantially accelerate cortical learning and improve performance when external feedback is delayed. Related frameworks also propose that cortico-cerebellar loops facilitate task acquisition, context-dependent control, and stabilization of recurrent cortical dynamics by supplying predicted feedback to cortex (Boven et al., 2023). Our contribution differs in objective and mechanism. In our model, cerebellar feedback is not primarily a predictor of delayed supervision nor an error proxy. Instead, it acts as a structured transformation

pathway that reshapes the signals transmitted within the cortical network, reorganizing cortical representations across tasks to reduce interference and improve compositionality.

Cerebellar signal transformation and filtering. A complementary line of work views the cerebellum as an adaptive signal-processing system that transforms input streams to support control and learning (Raymond et al., 1996; Marr, 1969; Albus, 1971; Dean et al., 2010; Sokolov et al., 2017). Classical Marr-Albus theory proposes that mossy-fiber inputs are expanded and sparsified by granule cells, yielding a high-dimensional basis in which contextual and sensorimotor variables can be separated and recombined; Purkinje cells then learn task-specific readouts via plastic parallel fiber synapses modulated by climbing-fiber instructive signals (Lanore et al., 2021). In this view, the cerebellum implements an adaptive filter or supervised learning system: mossy fibers convey task state and context, granule cells generate a rich intermediate representation, Purkinje cells implement learned readouts, and cerebellar nuclei convert these computations into outputs that influence cortical and motor circuits.

Our contribution leverages—but repurposes—this transformation view. We model the cerebellum as a pathway that converts cortical population states into task-dependent feedback. Crucially, the module does not merely relay cortical activity; it applies a low-rank, gain-modulated transformation that sculpts cortical trajectories, promoting separable, task-relevant latent manifolds. In Section 3, we show that this structured feedback is essential for reducing representational overlap and improving learning across multiple cognitive and sensorimotor tasks under biologically plausible plasticity.

3 Method Overview

3.1 Cognitive and motor task.

We evaluate the model on complementary cognitive and motor tasks. In the cognitive task, models perform four context-dependent sensorimotor tasks requiring rule selection, temporal processing, and memory maintenance, where sensory cues are transformed into two-dimensional responses under Pro/Anti and Delay/Memory contexts. In the motor task, models learn 26 handwritten-letter trajectories in a continual-learning setting and generate two-dimensional temporal trajectories for both new and previously acquired letters. Full task details are provided in the Appendix A.

3.2 Model Architectures

To examine the role of cerebellar adaptation in shaping cortical dynamics, we compare a cortex-only model with a cortico-cerebellar coupled model. The coupled model introduces an adaptive cerebellar feedback pathway that modulates cortical activity, while both models share identical cortical state spaces and baseline parameters, isolating the contribution of cerebellar coupling (Figure 1).

3.2.1 Cortex-Only RNN

We model the cortex as a continuous-time, full-rank RNN with hidden state $\mathbf{h}(t) \in \mathbb{R}^N$:

$$\begin{aligned} \tau \frac{\mathrm{d}\mathbf{h}(t)}{dt} &= -\mathbf{h}(t) + \phi(W_{\mathrm{rec}}\mathbf{h}(t) + W_{\mathrm{in}}\mathbf{u}(t) + \xi(t)) \\ \mathbf{y}(t) &= W_{\mathrm{out}}\mathbf{h}(t) + \mathbf{b} \end{aligned} \tag{1}$$

where $\mathbf{u}(t) \in \mathbb{R}^I$ is the external input, and $\xi(t) \sim \mathcal{N}(0, \sigma_\xi^2 I)$ represents intrinsic neural noise. The activation function $\phi(\cdot)$ is tanh. The recurrent and input weights $W_{\mathrm{rec}} \in \mathbb{R}^{N\times N}$ and $W_{\mathrm{in}} \in \mathbb{R}^{N\times I}$ are initialized as fixed full-rank matrices. The output is generated by a linear readout with parameters $W_{\mathrm{out}} \in \mathbb{R}^{Y\times N}$ and $\mathbf{b} \in \mathbb{R}^Y$.

During training, only the readout parameters W_{out} and b are optimized with Adam, while W_{rec} and W_{in} remain fixed.

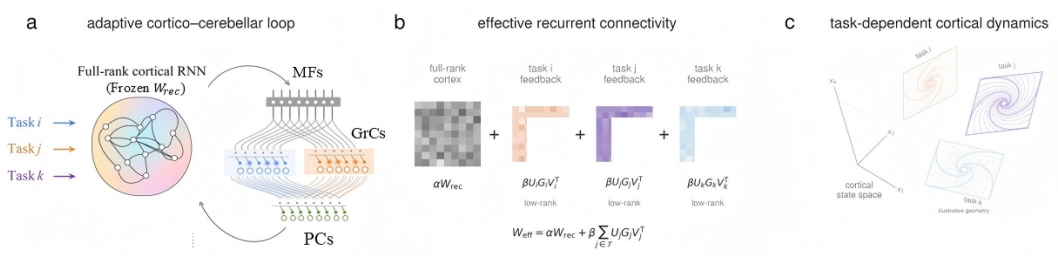

Figure 1: Schematic of the cortico-cerebellar coupled model. A full-rank cortical RNN receives task-dependent inputs and generates high-dimensional activity. Cortical dynamics are modulated by an adaptive cerebellar pathway, which transforms cortical states and returns structured, task-dependent low-rank feedback signals. This feedback reorganizes entangled cortical trajectories into task-specific latent manifolds, enabling separable cortical dynamics across tasks.

3.2.2 Cortico-cerebellar coupled model

The cortico-cerebellar model augments the cortical RNN with an adaptive cerebellar feedback loop comprising two stages: (i) a read-in pathway from mossy fibers to granule cells (MFs → GrCs) that transforms high-dimensional cortical activity into compact, decorrelated representations; and (ii) a read-out pathway from granule cells to Purkinje cells (GrCs → PCs) that learns error-modulated feedback signals to regulate cortical dynamics.

(i) Cerebellar read-in pathway

The read-in stage extracts latent features from cortical states. For task k, cortical activity $\mathbf{h}(t)$ serves as MF input. The read-in matrix $V \in \mathbb{R}^{M \times N}$ of MFs → GrCs projects the N-dimensional cortical state into an M-dimensional GrC activity:

$$\boldsymbol{c}(t) = \boldsymbol{h}(t)V \tag{2}$$

To model activity-dependent MF–GrC synaptic plasticity, we update V_k using a local Oja-type rule (Oja, 1982):

$$\Delta V^{\text{Hebb}} = \bar{\mathbf{h}}(t)\mathbf{c}(t)^T - V\mathbf{c}(t)^2 \tag{3}$$

where the first term implements Hebbian learning and the second provides activity-dependent normalization to prevent weight divergence.

To promote non-redundant feature extraction, we introduce an anti-Hebbian lateral term to model Golgi-cell-mediated decorrelation in the granule layer.

$$\Delta V^{\text{Lat}} = -V\left(\boldsymbol{c}(t)\boldsymbol{c}(t)^T - \text{diag}\left(\boldsymbol{c}(t)^2\right)\right). \tag{4}$$

which suppresses correlations among GrC features and promotes sparse, decorrelated representations.

Therefore, the read-in update is:

$$\widetilde{V} = V + \eta_V \Delta V^{\text{Hebb}} + \eta_V^{\text{lat}} \Delta V^{\text{lat}} \tag{5}$$

where η_V and η_V^{lat} are learning rates for the Hebbian and lateral components, respectively.

(ii) Cerebellar read-out stage

While the read-in stage of MF-GrC pathway serves to sculpt cortical representation, the GrC-PC pathway implements an error-modulated linear readout. Let $U \in \mathbb{R}^{N \times K}$ denote the read-out matrix. To avoid biologically implausible backpropagation-through-time, we use a three-factor plasticity rule to update U driven by (i) presynaptic activity, (ii) postsynaptic/target-related activity, and (iii) a modulatory error signal. The presynaptic activity is the context-gated GrC activity $\mathbf{z}(t) = \mathbf{g} \odot \tilde{V}\mathbf{h}(t-1)$, where $\mathbf{g}$ represents a

context-dependent modulatory gain conveyed by CF input. In parallel, the error-dependent modulatory signal for synaptic plasticity at GrC-PC synapses is given by $B\mathbf{e}_t = B(y_t - \hat{y}_t)$, where B is a fixed random feedback matrix. To stabilize the feedback basis, we incorporate an Oja-like normalization term. The local update rule for U is:

$$\Delta U = \langle B\mathbf{e}(t)\mathbf{z}(t)^T P(t)\rangle_t - U\,\mathrm{diag}\left(\langle \mathbf{z}(t)^2\rangle_t\right). \tag{6}$$

The read-out update is:

$$\tilde{U} = U + \eta_U\,\Delta U. \tag{7}$$

where η_U denote the learning rates.

(iii) Coupled cortical dynamics with cerebellar feedback

The cerebellar module introduces low-rank adaptive feedback into cortical dynamics:

$$\tau\frac{\mathrm{d}\mathbf{h}(t)}{dt} = -\mathbf{h}(t) + \phi\left((\alpha W_{\mathrm{rec}} + \beta\sum_{j\in T} U_j \mathrm{diag}(g_j) V_j^{\top})\mathbf{h}(t) + W_{\mathrm{in}}\mathbf{u}(t) + \xi(t)\right) \tag{8}$$

Where α rescales recurrent gain, β controls cerebellar feedback strength, and T denotes the task set. For each task, the cerebellar feedback term $U_j \mathrm{diag}(g_j) V_j^{\top}$ provides a low-rank modification of cortical dynamics, with rank bounded by the GrC dimensionality M, while the cortical recurrent matrix W_{rec} remains full-rank. Thus, cerebellar feedback selectively reshapes cortical activity within constrained latent subspaces while preserving the expressive capacity of the cortical network.

3.3 Experiment settings

Models were trained using Adam optimization. The cortex-only model optimized only the readout parameters, whereas the coupled model additionally updated cerebellar read-in/read-out parameters through local plasticity rules. All results are reported as mean $\pm$ standard deviation across multiple random seeds. Detailed training settings and hyperparameters are provided in the Appendix C.

4 Result

4.1 Dynamic flexibility of the cortico-cerebellar model

The cortico-cerebellar dynamics are governed by the interaction between cortical recurrence strength α and cerebellar feedback strength β (Eq. 8). Without task inputs, varying these parameters induces distinct dynamical regimes (Figure. 2A). Strong cortical recurrence promotes irregular activity, whereas low-rank cerebellar feedback constrains trajectories within a low-dimensional subspace and stabilizes fixed-point attractors. Increasing feedback rank expands the dimensionality of this subspace, yielding richer attractor structures. Figure. 2B further shows that cerebellar feedback reshapes the stability landscape by controlling transitions between decaying and oscillatory dynamics under constant input. Together, these results show that low-rank cerebellar feedback regulates high-dimensional cortical dynamics by reshaping the stability landscape and organizing activity into stable, structured latent manifolds.

4.2 Both cortex-only and cortico-cerebellar models support single-task learning

Our dynamical analysis shows that cerebellar feedback reshapes full-rank cortical RNN dynamics into structured, low-dimensional regimes. We next assess the representational capacity and learning performance of cortex-only and cortico-cerebellar models during single-task learning. We use cognitive benchmark introduced by Duncker et al. (Duncker et al., 2020), which evaluates context-dependent sensorimotor remapping and working memory through four tasks: DelayPro, DelayAnti, MemoryPro, and MemoryAnti. Each trial provides five input channels encoding fixation, context, timing, and stimulus angle. The model reports

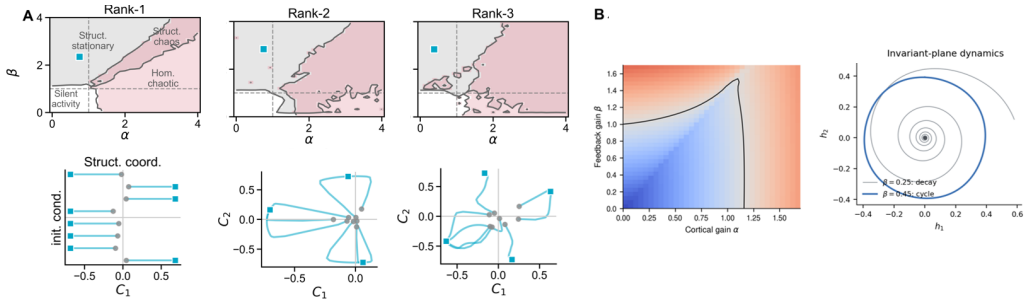

Figure 2: Cortical dynamics in the coupled cortico-cerebellar system. (A) Phase diagrams for rank-1, rank-2, and rank-3 cerebellar feedback across cortical recurrence α and feedback strength β. Colored markers denote representative conditions, with bottom panels showing the state-space trajectories projected onto low-dimensional coordinates. (B) Feedback-dependent stability transitions. The stability landscape and representative trajectories illustrate the control of decaying and oscillatory regimes by cortical and cerebellar gains.

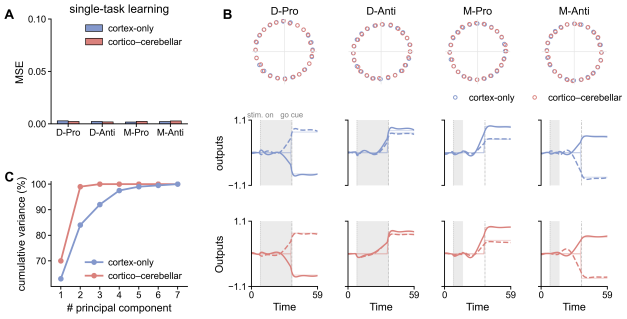

Figure 3: Single-task learning performance of cortex-only and cortico-cerebellar coupled models on four-task benchmark. (A) MSE performance across tasks. (B) Task-specific outputs. Top: target and model-predicted angular outputs across stimulus conditions. Middle and bottom: representative output trajectories over time for the cortex-only and coupled models, respectively. Gray indicates targets; blue and red indicate cortex-only and coupled models. (C) Dimensionality of neural dynamics measured by cumulative explained variance across principal components.Only the mean across multiple random seeds is shown here; the corresponding variance is reported in the Appendix.

either the presented stimulus angle in Pro tasks or the opposite angle in Anti tasks, while Memory tasks additionally require maintaining the stimulus representation across a delay period. Full task details are provided in Appendix A.

When trained on a single task, both models achieved high performance with comparable MSE values (Figure 3A) and accurately reproduced target response angles across stimulus conditions (Figure 3B). This indicates that the cortex-only model can solve the task when cross-task interference is absent. Despite similar performance, the coupled model exhibited more lower-dimensional dynamics (Figure 3C). In the DelayPro task, the first two principal components captured nearly all population variance in the coupled model, whereas the cortex-only model required substantially more components. Similar results were observed across other tasks (Appendix D). These findings suggest that cerebellar low-rank feedback reorganizes cortical activity into lower-dimensional representations, providing a structured substrate for reducing interference during multi-task learning.

4.3 Multi-task learning relies on low-rank cerebellar feedback

We next evaluated whether the two models could acquire the full four-task set under sequential learning. Tasks were trained in the representative order DelayPro → DelayAnti

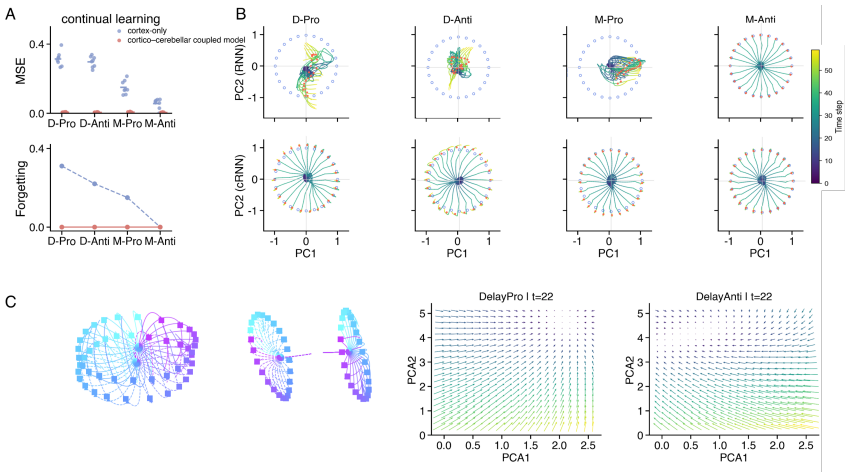

Figure 4: Performance on the four-task learning for the cortex-only and cortico–cerebellar models. Tasks were introduced sequentially in the order: DelayPro (D-Pro) → DelayAnti (D-Anti) → MemoryPro (M-Pro) → MemoryAnti (M-Anti). (A) Final performance and forgetting measured after completion of the full training sequence. (B) Task-specific output behavior and neural dynamics after sequential learning. For each task, we plot model output trajectories and target angles across stimulus conditions, and visualize low-dimensional neural manifolds. (C) Visualize low-dimensional neural manifolds (via projection of population activity onto the Purkinje/cerebellar latent coordinates) for the Pro and Anti tasks. The PCA-plane vector fields reveal task-dependent dynamical flows.

Table 1: Ablation analysis of adaptive cerebellar feedback components

Model	Feedback formulation	MSE
No Feedback	–	0.3160 ± 0.0022
Fully Shared Feedback	$U\mathrm{diag}(g)V^T$	0.0195 ± 0.0059
Shared U, V + Task Gain	$U\mathrm{diag}(g_k)V^T$	0.0147 ± 0.0045
Task-specific U_k + Shared V	$U_k\mathrm{diag}(g_k)V^T$	0.0065 ± 0.0022
Shared U + Task-specific V_k	$U\mathrm{diag}(g_k)V_k^T$	0.0054 ± 0.0028
Task-specific U_k, V_k + Shared Gain	$U_k\mathrm{diag}(g)V_k^T$	0.0035 ± 0.0002
Full Task-specific Feedback (ours)	$U_k\mathrm{diag}(g_k)V_k^T$	$\mathbf{0.0029 \pm 0.0001}$

→ MemoryPro → MemoryAnti, with each task learned once without revisiting previous tasks. Consistent trends were observed across different task orders and network initializations.The cortex-only model exhibited strong order-dependent interference (Figure 4A), with the earliest learned tasks showing the largest performance degradation. This impairment was reflected in output behavior and neural dynamics (Figure 4B), including scattered final responses, distorted trajectories, and overlapping low-dimensional task representations.

In contrast, the cortico-cerebellar model maintained accurate performance across all four tasks after sequential learning. Its outputs remained aligned with target angles, trajectories preserved angle-specific dynamics, and neural activity formed more regular and separated low-dimensional manifolds. Notably, DelayPro and DelayAnti trajectories formed distinct clusters ordered smoothly by stimulus angle (Figure 4C), indicating separation of Pro/Anti rule identity while preserving continuous stimulus angular representations within each task-specific submanifold. This organization indicates that cerebellar feedback reshapes cortical population dynamics into rule-specific manifolds that retain the circular geometry of the stimulus space. Consistent with this interpretation, PCA-plane vector fields show distinct task-dependent flow patterns for DelayPro versus DelayAnti, implying that the two rules are implemented by different cortical dynamical landscapes. Mechanistically, rule identity is encoded by manifold separation, whereas stimulus direction is encoded by smooth ordering within each manifold.

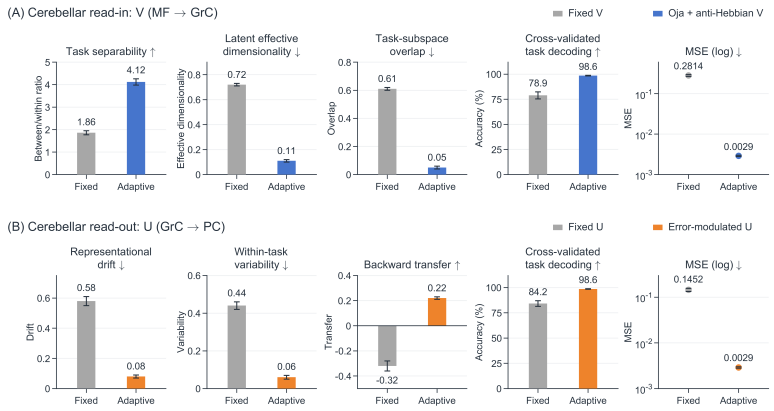

Figure 5: (A) Fixed versus adaptive MF–GrC read-in pathways evaluated by representation structure, task overlap, decoding accuracy, and performance. (B) Fixed versus error-modulated GrC–PC read-out pathways evaluated by representational drift, variability, transfer, decoding accuracy, and performance.

To identify the component contribution of cerebellar feedback, we progressively constrained the feedback pathway in Table 1. Removing cerebellar feedback severely impaired performance (MSE: 0.3160 ± 0.0022), indicating increased task interference. Sharing task-specific projections or gain modulation also degraded performance, whereas the full task-specific feedback model achieved the lowest error (0.0029 ± 0.0001). These results demonstrate that adaptive modulation of both feedback representations and task-dependent gains is critical for robust multi-task learning. Further mechanism analysis indicates adaptive MF–GrC read-in plasticity promoted compact and separable task representations, reducing effective dimensionality and cross-task overlap while improving task decoding (Figure 5A). In contrast, adaptive GrC–PC read-out plasticity stabilized task mappings by reducing representational drift and improving transfer across tasks (Figure 5B). These complementary mechanisms allow cerebellar feedback to reshape representations and preserve task-specific computations during continual learning.

We next examined the contribution of biologically plasticity rules to adaptive cerebellar feedback (Table 2). Replacing both local learning mechanisms with gradient-based optimization led to the largest performance degradation, while removing either MF–GrC or GrC–PC adaptation produced partial impairments. The combination of Oja/anti-Hebbian read-in and three-factor read-out learning achieved the best performance, suggesting that complementary plasticity mechanisms promote stable multi-task learning. We also compared architectural variants to assess the role of adaptive low-rank cerebellar feedback (Table 3). The cortex-only model showed severe performance degradation, whereas adaptive cortico-cerebellar coupling achieved the lowest error. Fixed or randomly perturbed low-rank connectivity provided limited improvements, highlighting that task-specific adaptive feedback, rather than low-rank structure alone, is essential for robust multi-task learning.

Table 2: Ablation of biological learning rules in cerebellar feedback

Model	Update V_k	Update U_k	MSE
All-Gradient	Adam	Adam	0.1231 ± 0.0215
Gradient Read-In	Adam	three-factor	0.0513 ± 0.0062
Gradient Read-Out	Oja + anti-Hebbian	Adam	0.0786 ± 0.0071
Full Model(ours)	Oja + anti-Hebbian	three-factor	$\mathbf{0.0029 \pm 0.0001}$

Table 3: Comparison of different model architectures

Model	MSE ↓
Cortex-only	0.3160 ± 0.0022
Cortico-cerebellar (ours)	$\mathbf{0.0029 \pm 0.0001}$
Cortex-only + low-rank perturbation to full-rank connectivity	0.2519 ± 0.0112
Cortex-only + low-rank connectivity	0.2207 ± 0.0026
Cerebellar learned task-specific low-rank	0.0331 ± 0.0015
Cerebellar setting low-rank	0.0745 ± 0.0026

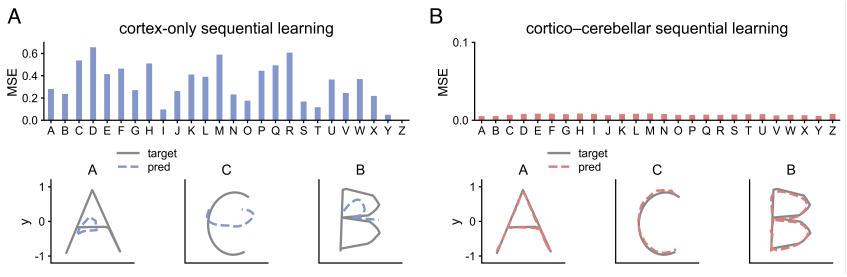

Figure 6: Sequential learning of the cortex-only (A) and cortico-cerebellar models (B) in handwritten-letter trajectory benchmark. Only the mean across multiple random seeds is shown here; the corresponding variance is reported in the Appendix.

4.4 Continuous learning of complex trajectories of the cortico-cerebellar model

Although the four-task benchmark evaluates context-dependent computation, its output dynamics are relatively simple. We therefore introduced a handwritten-letter trajectory benchmark to further probe learning capacity and resistance to interference, where models sequentially learn to generate two-dimensional time-varying trajectories for 26 letter-generation tasks (Appendix A). This benchmark expands the task repertoire and imposes stronger demands on temporal dynamics and continual learning. After sequential training on all letters, the cortico-cerebellar model maintained low generation error across the alphabet, whereas the cortex-only model exhibited rapid performance degradation after learning a small subset of tasks (Figure 6). These results demonstrate that adaptive cerebellar low-rank feedback improves memory retention and supports scalable multi-task learning beyond simple task settings.

5 Discussion

We introduced a cortico-cerebellar coupling network for robust multi-task representation and learning, aimed at capturing the synergy between cortical recurrent dynamics and cerebellar regulation. The framework retains the expressive capacity of a full-rank cortical RNN while augmenting it with an adaptive cerebellar pathway that transforms cortical population activity and feeds back structured signals to shape subsequent cortical dynamics. Unlike approaches that enforce low-dimensionality through engineered architectural constraints (e.g., low-rank recurrent connectivity), our results suggest that cerebellar feedback can act as a flexible circuit-level mechanism for sculpting low-dimensional, task-relevant cortical dynamics. Several limitations remain. The cerebellar pathway is modeled at a functional level as an abstraction of cortico-cerebellar interactions. Incorporating more detailed biological mechanisms could sharpen mechanistic interpretability and enable closer comparisons to experimental data. Future work should test whether analogous modulatory signals can be derived from realistic sensory prediction errors and/or reward-based feedback.

AI use statement

In this work, we used generative AI tools for language polishing and simulation-code debugging. The authors reviewed the text, code, results and cited sources and take responsibility for the final content. The experimental design, simulations and scientific claims are the authors' own work.

Ethics statement

This work does not involve human participants, animal experiments, personally identifiable information, or the collection and release of new datasets. The proposed computational framework is developed and evaluated using synthetic benchmark tasks and does not introduce direct risks related to privacy, safety, discrimination, or harmful applications. We have followed principles of research integrity and transparency throughout the study.

Reproducibility statement

We have made efforts to ensure the reproducibility of our results. Detailed descriptions of benchmark task generation, learning rules, training procedures, and evaluation protocols are provided in the Appendix.

References

James S Albus. A theory of cerebellar function. Mathematical biosciences, 10(1-2):25–61, 1971.

Samuel J Allen and Hirofumi Morishita. Local and long-range input balance: A framework for investigating frontal cognitive circuit maturation in health and disease. Science Advances, 10(38):eadh3920, 2024.

Kanika Bansal, Zaira Cattaneo, Viola Oldrati, Chiara Ferrari, Emily D Grossman, and Javier O Garcia. A modulator of cognitive function: Cerebellum modifies human cortical network dynamics via integration. bioRxiv, pp. 2024–09, 2024.

Yeyao Bao, Liao Yu, Liangfu Lu, Zhuoqin Yang, and Yunliang Zang. Cerebellar microcircuits enable robust evidence-based decisions through cortico–cerebellar coupling. Proceedings of the National Academy of Sciences, 123(35):e2616911123, 2026.

Manuel Beiran, Nicolas Meirhaeghe, Hansem Sohn, Mehrdad Jazayeri, and Srdjan Ostojic. Parametric control of flexible timing through low-dimensional neural manifolds. Neuron, 111(5):739–753, 2023.

Andreea C Bostan and Peter L Strick. The basal ganglia and the cerebellum: nodes in an integrated network. Nature Reviews Neuroscience, 19(6):338–350, 2018.

Andreea C Bostan, Richard P Dum, and Peter L Strick. Cerebellar networks with the cerebral cortex and basal ganglia. Trends in cognitive sciences, 17(5):241–254, 2013.

Ellen Boven, Joseph Pemberton, Paul Chadderton, Richard Apps, and Rui Ponte Costa. Cerebro-cerebellar networks facilitate learning through feedback decoupling. Nature communications, 14(1):51, 2023.

Martina Bracco, Varsha Vasudevan, Vridhi Rohira, Quentin Welniarz, Mihoby Razafinimanana, Alienor Richard, Christophe Gitton, Sabine Meunier, Antoni Valero-Cabré, Denis Schwartz, et al. Cerebellar-cortical beta oscillations emerge as a predictive signal facilitating the stability of behavioral performance. Current Biology, 36(2):522–531, 2026.

Randy L Buckner. The cerebellum and cognitive function: 25 years of insight from anatomy and neuroimaging. Neuron, 80(3):807–815, 2013.

Randy L Buckner, Fenna M Krienen, Angela Castellanos, Julio C Diaz, and BT Thomas Yeo. The organization of the human cerebellum estimated by intrinsic functional connectivity. Journal of neurophysiology, 106(5):2322–2345, 2011.

Mark M Churchland, John P Cunningham, Matthew T Kaufman, Justin D Foster, Paul Nuyujukian, Stephen I Ryu, and Krishna V Shenoy. Neural population dynamics during reaching. Nature, 487(7405):51–56, 2012.

Uri Cohen, SueYeon Chung, Daniel D Lee, and Haim Sompolinsky. Separability and geometry of object manifolds in deep neural networks. Nature communications, 11(1):746, 2020.

Julia C Costacurta, Shaunak Bhandarkar, David Zoltowski, and Scott W Linderman. Structured flexibility in recurrent neural networks via neuromodulation. Advances in Neural Information Processing Systems, 37:1954–1972, 2024.

John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. Nature neuroscience, 17(11):1500–1509, 2014.

Paul Dean, John Porrill, Carl-Fredrik Ekerot, and Henrik Jörntell. The cerebellar microcircuit as an adaptive filter: experimental and computational evidence. Nature Reviews Neuroscience, 11(1):30–43, 2010.

Lea Duncker and Maneesh Sahani. Dynamics on the manifold: Identifying computational dynamical activity from neural population recordings. Current opinion in neurobiology, 70:163–170, 2021.

Lea Duncker, Laura Driscoll, Krishna V Shenoy, Maneesh Sahani, and David Sussillo. Organizing recurrent network dynamics by task-computation to enable continual learning. Advances in neural information processing systems, 33:14387–14397, 2020.

Timo Flesch, Keno Juechems, Tsvetomira Dumbalska, Andrew Saxe, and Christopher Summerfield. Orthogonal representations for robust context-dependent task performance in brains and neural networks. Neuron, 110(7):1258–1270, 2022.

Juan A Gallego, Matthew G Perich, Lee E Miller, and Sara A Solla. Neural manifolds for the control of movement. Neuron, 94(5):978–984, 2017.

Juan A Gallego, Matthew G Perich, Raeed H Chowdhury, Sara A Solla, and Lee E Miller. Long-term stability of cortical population dynamics underlying consistent behavior. Nature neuroscience, 23(2):260–270, 2020.

Zhenyu Gao, Courtney Davis, Alyse M Thomas, Michael N Economo, Amada M Abrego, Karel Svoboda, Chris I De Zeeuw, and Nuo Li. A cortico-cerebellar loop for motor planning. Nature, 563(7729):113–116, 2018.

Vishwa Goudar, Barbara Peysakhovich, David J Freedman, Elizabeth A Buffalo, and Xiao-Jing Wang. Schema formation in a neural population subspace underlies learning-to-learn in flexible sensorimotor problem-solving. Nature neuroscience, 26(5):879–890, 2023.

Olivia Gozel and Brent Doiron. Between-area communication through the lens of within-area neuronal dynamics. Science Advances, 10(42):eadl6120, 2024.

Sung Won Hur, Karen Safaryan, Long Yang, Hugh T Blair, Sotiris C Masmanidis, Paul J Mathews, Daniel Aharoni, and Peyman Golshani. Correlated signatures of social behavior in cerebellum and anterior cingulate cortex. Elife, 12:RP88439, 2024.

Sharon Israely, Hugo Ninou, Ori Rajchert, Lee Elmaleh, Ran Harel, Firas Mawase, Jonathan Kadmon, and Yifat Prut. Cerebellar output shapes cortical preparatory activity during motor adaptation. Nature Communications, 16(1):2574, 2025.

Masao Ito. Control of mental activities by internal models in the cerebellum. Nature Reviews Neuroscience, 9(4):304–313, 2008.

Mehrdad Jazayeri and Srdjan Ostojic. Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. Current opinion in neurobiology, 70:113–120, 2021.

Roberta M Kelly and Peter L Strick. Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. The Journal of neuroscience, 23(23):8432–8444, 2003.

Jaekyung Kim, Abhilasha Joshi, Loren Frank, and Karunesh Ganguly. Cortical–hippocampal coupling during manifold exploration in motor cortex. Nature, 613(7942): 103–110, 2023.

Christopher Langdon, Mikhail Genkin, and Tatiana A Engel. A unifying perspective on neural manifolds and circuits for cognition. Nature Reviews Neuroscience, 24(6):363–377, 2023.

Frederic Lanore, N Alex Cayco-Gajic, Harsha Gurnani, Diccon Coyle, and R Angus Silver. Cerebellar granule cell axons support high-dimensional representations. Nature neuroscience, 24(8):1142–1150, 2021.

Nuo Li and Thomas D Mrsic-Flogel. Cortico-cerebellar interactions during goal-directed behavior. Current opinion in neurobiology, 65:27–37, 2020.

Valerio Mante, David Sussillo, Krishna V Shenoy, and William T Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. nature, 503(7474): 78–84, 2013.

Jialin Mao, Itay Griniasty, Han Kheng Teoh, Rahul Ramesh, Rubing Yang, Mark K Transtrum, James P Sethna, and Pratik Chaudhari. The training process of many deep networks explores the same low-dimensional manifold. Proceedings of the National Academy of Sciences, 121(12):e2310002121, 2024.

David Marr. A theory of cerebellar cortex. The Journal of physiology, 202(2):437–470, 1969.

Frank A Middleton and Peter L Strick. Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. Science, 266(5184):458–461, 1994.

Erkki Oja. Simplified neuron model as a principal component analyzer. Journal of mathematical biology, 15(3):267–273, 1982.

Chethan Pandarinath, Daniel J O'Shea, Jasmine Collins, Rafal Jozefowicz, Sergey D Stavisky, Jonathan C Kao, Eric M Trautmann, Matthew T Kaufman, Stephen I Ryu, Leigh R Hochberg, et al. Inferring single-trial neural population dynamics using sequential auto-encoders. Nature methods, 15(10):805–815, 2018.

Joseph Pemberton, Paul Chadderton, and Rui Ponte Costa. Cerebellar-driven cortical dynamics can enable task acquisition, switching and consolidation. Nature Communications, 15(1):10913, 2024.

Matthew G Perich, Devika Narain, and Juan A Gallego. A neural manifold view of the brain. Nature Neuroscience, 28(8):1582–1597, 2025.

Louis Pezon, Valentin Schmutz, and Wulfram Gerstner. Linking neural manifolds to circuit structure in recurrent networks. Neuron, 114(9):1682–1694, 2026.

Eli Pollock and Mehrdad Jazayeri. Engineering recurrent neural networks from task-relevant manifolds and dynamics. PLoS computational biology, 16(8):e1008128, 2020.

José Mário Prati, André Pontes-Silva, and Anna Carolyna Lepesteur Gianlorenço. The cerebellum and its connections to other brain structures involved in motor and non-motor functions: A comprehensive review. Behavioural Brain Research, 465:114933, 2024.

Jennifer L Raymond and Javier F Medina. Computational principles of supervised learning in the cerebellum. Annual review of neuroscience, 41(1):233–253, 2018.

Jennifer L Raymond, Stephen G Lisberger, and Michael D Mauk. The cerebellum: a neuronal learning machine? Science, 272(5265):1126–1131, 1996.

Matthew C Rosen and David J Freedman. How distributed is the brain-wide network that is recruited for cognition? Nature Reviews Neuroscience, 27(2):138–150, 2026.

Michael E Rule, Adrianna R Loback, Dhruva V Raman, Laura N Driscoll, Christopher D Harvey, and Timothy O'Leary. Stable task information from an unstable neural population. elife, 9:e51121, 2020.

Shan Shen, Xiaolong Jiang, Federico Scala, Jiakun Fu, Paul Fahey, Dmitry Kobak, Zhenghuan Tan, Na Zhou, Jacob Reimer, Fabian Sinz, et al. Distinct organization of two cortico-cortical feedback pathways. Nature communications, 13(1):6389, 2022.

Krishna V Shenoy, Maneesh Sahani, and Mark M Churchland. Cortical control of arm movements: a dynamical systems perspective. Annual review of neuroscience, 36(1): 337–359, 2013.

Arseny A Sokolov, R Chris Miall, and Richard B Ivry. The cerebellum: adaptive prediction for movement and cognition. Trends in cognitive sciences, 21(5):313–332, 2017.

Sergey D Stavisky, Jonathan C Kao, Stephen I Ryu, and Krishna V Shenoy. Motor cortical visuomotor feedback activity is initially isolated from downstream targets in output-null neural state space dimensions. Neuron, 95(1):195–208, 2017.

Saurabh Vyas, Matthew D Golub, David Sussillo, and Krishna V Shenoy. Computation through neural population dynamics. Annual review of neuroscience, 43(1):249–275, 2020.

Mark J Wagner and Liqun Luo. Neocortex–cerebellum circuits for cognitive processing. Trends in neurosciences, 43(1):42–54, 2020.

Albert J Wakhloo, Will Slatton, and SueYeon Chung. Neural population geometry and optimal coding of tasks with shared latent structure. Nature neuroscience, 29(3):682–692, 2026.

Jan Weber, Gabriela Iwama, Anne-Kristin Solbakk, Alejandro O Blenkmann, Pal G Larsson, Jugoslav Ivanovic, Robert T Knight, Tor Endestad, and Randolph Helfrich. Subspace partitioning in the human prefrontal cortex resolves cognitive interference. Proceedings of the National Academy of Sciences, 120(28):e2220523120, 2023.

Yunliang Zang and Erik De Schutter. Recent data on the cerebellum require new models and theories. Current opinion in neurobiology, 82:102765, 2023.

Yaoduo Zhang, Jian Zhang, and Yunliang Zang. Cerebellum-inspired kernel for efficient out-of-distribution detection. bioRxiv, pp. 2026–03, 2026.

Jia Zhu, Hana Hasanbegović, Liu D Liu, Zhenyu Gao, and Nuo Li. Activity map of a cortico-cerebellar loop underlying motor planning. Nature neuroscience, 26(11):1916–1928, 2023.

A Task structure and input-output protocol

A.1 Cognitive task

We evaluated the model on a family of four context-dependent sensorimotor tasks designed to dissociate rule selection, temporal structure, and memory maintenance. The task set consisted of DelayPro (D-Pro), DelayAnti (D-Anti), MemoryPro (M-Pro), MemoryAnti (M-Anti) trials. Each trial required the network to transform a sensory cue, represented by an angle θ, into a two-dimensional response vector. The Pro/Anti context determined whether the response should match or oppose the sensory cue, while the Delay/Memory context determined whether the sensory signal remained available until the response period or had to be maintained internally after a brief presentation.

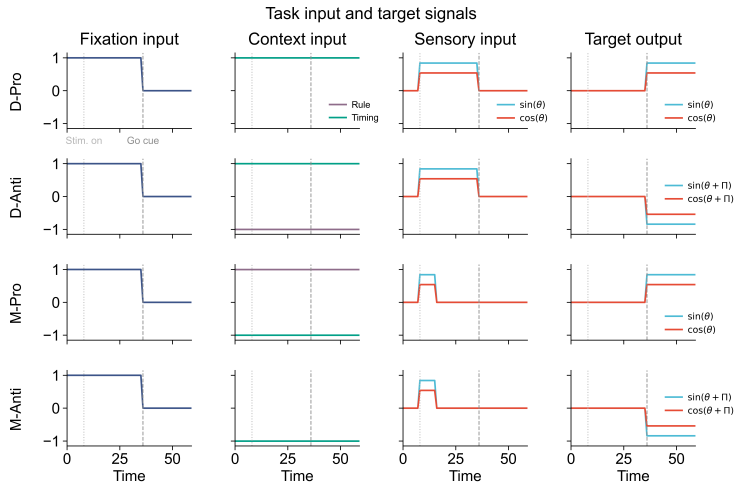

Figure 7: Figure S.1: Task input and target-output structure. Schematic of the four context-dependent sensorimotor tasks used in the experiments: Delay-Pro, Delay-Anti, Memory-Pro, and Memory-Anti. Each trial consists of fixation input, rule and timing context inputs, sensory input, and target output. The fixation signal remains active before the go cue and is released at response onset. Context inputs specify the required rule, Pro versus Anti, and temporal condition, Delay versus Memory. Sensory cues are represented as $\sin(\theta)$ and $\cos(\theta)$. In Delay tasks, the sensory cue is maintained until the go cue, whereas in Memory tasks it is presented briefly and must be retained internally. During the response epoch, Pro trials require the model to output $[\sin(\theta), \cos(\theta)]$, while Anti trials require the opposite-direction response $[\sin(\theta + \pi), \cos(\theta + \pi)]$. Dotted and dashed vertical lines indicate stimulus onset and go cue, respectively.

Each trial lasted $T = 60$ discrete time steps. The sensory cue appeared at $t_{\text{stim}} = 8$, and the response period began at $t_{\text{go}} = 36$. In Delay tasks, the sensory cue remained present from t_{stim} until t_{go}. In Memory tasks, the cue was only presented for a short duration of $T_{\text{mem}} = 8$ time steps, after which the network had to maintain the relevant information until the go cue.

At each time step t, the model received a five-dimensional input vector

$$x_t = \left[x_t^{\text{fix}}, x_t^{\text{rule}}, x_t^{\text{time}}, x_t^{\sin}, x_t^{\cos}\right].$$

The first input dimension encoded fixation. It was set to 1 before the response period and 0 after the go cue:

$$x_t^{\text{fix}} = \begin{cases} 1, & t < t_{\text{go}}, \\ 0, & t \geq t_{\text{go}}. \end{cases}$$

The second input dimension encoded the behavioral rule. It was set to $+1$ for Pro tasks and -1 for Anti tasks:

$$x^{\text{rule}} = \begin{cases} +1, & \text{Pro task}, \\ -1, & \text{Anti task}. \end{cases}$$

The third input dimension encoded the temporal context. It was set to $+1$ for Delay tasks and -1 for Memory tasks:

$$x^{\text{time}} = \begin{cases} +1, & \text{Delay task}, \\ -1, & \text{Memory task}. \end{cases}$$

The final two input dimensions encoded the sensory cue using a circular representation:

$$s(\theta) = [\sin(\theta), \cos(\theta)].$$

For Delay tasks, the sensory input was present from t_{stim} to t_{go}:

$$\left[x_t^{\sin}, x_t^{\cos}\right] = \begin{cases} [\sin(\theta), \cos(\theta)], & t_{\text{stim}} \le t < t_{\text{go}}, \\ [0, 0], & \text{otherwise.} \end{cases}$$

For Memory tasks, the same sensory input was only presented during a short cue epoch:

$$\left[x_t^{\sin}, x_t^{\cos}\right] = \begin{cases} [\sin(\theta), \cos(\theta)], & t_{\text{stim}} \le t < t_{\text{stim}} + T_{\text{mem}}, \\ [0, 0], & \text{otherwise.} \end{cases}$$

The target output was two-dimensional and remained zero before the response period. After the go cue, the network was required to output either the original sensory direction in Pro tasks or the opposite direction in Anti tasks:

$$y_t = \begin{cases} [0, 0], & t < t_{\text{go}}, \\ \sigma_k \left[\sin(\theta), \cos(\theta)\right], & t \ge t_{\text{go}}, \end{cases}$$

where $\sigma_k = +1$ for Pro tasks and $\sigma_k = -1$ for Anti tasks. This design requires the model to combine three computational demands: context-dependent rule selection, temporally structured cue processing, and maintenance of sensory information across a delay period. Thus, successful performance cannot be achieved by a purely reactive mapping from current sensory input to output; the network must use contextual signals to determine the appropriate transformation and, in Memory tasks, preserve task-relevant information after the cue disappears.

A.2 Motor task: Handwritten-letter trajectory prediction task

We evaluated the model on a sequential handwritten-letter trajectory generation benchmark. The task set consisted of 26 tasks, corresponding to the uppercase letters A through Z. Each task required the network to generate a two-dimensional continuous trajectory representing the target letter. The tasks were learned in a strict continual-learning sequence $A \to B \to C \to \cdots \to Z$, so that after learning each new letter, the model was evaluated on both the newly acquired task and previously learned letters. Each letter was represented by a canonical two-dimensional trajectory. We first defined each letter as a continuous polyline template in a normalized coordinate system. The templates were centered and scaled to a common spatial range, and then resampled to a fixed number of time points corresponding to the response epoch. This produced a canonical target trajectory

$$\tau_k = \{p_{k,1}, p_{k,2}, \ldots, p_{k,T_{\text{resp}}}\}, \quad p_{k,t} \in \mathbb{R}^2,$$

for each letter task $k \in \{1, \ldots, 26\}$.

Each trial lasted $T = 64$ discrete time steps. The first $T_{\text{prep}} = 8$ steps were treated as a preparatory epoch. The response period began at $T_{\text{go}} = 20$, and lasted $T_{\text{resp}} = T - T_{\text{go}} = 44$ time steps. Before the go cue, the target output was zero. During the response period, the target output followed the letter trajectory.

At each time step, the network received a 28-dimensional input vector,

$$x_t \in \mathbb{R}^{28},$$

composed of a 26-dimensional one-hot task cue and a 2-dimensional phase signal. The one-hot cue specified the identity of the current letter and was provided throughout the entire trial:

$$x_{t,k}^{\text{letter}} = \begin{cases} 1, & \text{if the current task is letter } k, \\ 0, & \text{otherwise.} \end{cases}$$

The final two input dimensions encoded the progression of the writing movement using a circular phase signal. The phase signal was inactive before the response epoch and evolved only after the go cue:

$$x_t^{\text{phase}} = \begin{cases} [0, 0], & t < T_{\text{go}}, \\ [\sin(\phi_t), \cos(\phi_t)], & t \ge T_{\text{go}}, \end{cases}$$

where $\phi_t \in [0, 2\pi]$ was linearly spaced across the response epoch. Thus, the task cue determined which letter should be generated, while the phase input provided a continuous temporal index specifying where the model was along the writing trajectory.

The target output was a two-dimensional coordinate sequence,

$$y_t \in \mathbb{R}^2.$$

For a trial of letter k, the target was defined as

$$y_t = \begin{cases} [0, 0]\,, & t < T_{\text{go}}, \\ p_{k,t-T_{\text{go}}+1}, & t \geq T_{\text{go}}. \end{cases}$$

Therefore, the model had to remain silent before the response epoch and then generate the full two-dimensional trajectory of the cued letter after the go cue.

B Dynamical analysis of the cortico–cerebellar model

We characterize the intrinsic dynamical properties of the cortico–cerebellar architecture, The cortical state $h \in \mathbb{R}^N$ evolves according to

$$\begin{aligned} \tau \dot{h} &= -h + \phi(Ah + I), \\ A &= \alpha W + \beta K, \\ K &= \sum_j U_j G_j V_j^T. \end{aligned} \tag{9}$$

Here, W represents intrinsic cortical recurrence, while K captures adaptive cortico–cerebellar feedback through multiple low-dimensional pathways. Each pathway projects cortical activity into a latent cerebellar representation through V_j^T, applies adaptive gain modulation G_j, and feeds the transformed representation back through U_j. The effective rank of each feedback pathway satisfies

$$\text{rank}(U_j G_j V_j^T) \leq \min(N, M_j, \text{rank}(G_j)),$$

indicating that cerebellar feedback introduces structured low-dimensional modifications to cortical dynamics when the total effective feedback rank is small relative to the cortical dimensionality.

B.1 Fixed points and local stability

For constant input and frozen parameters, equilibria satisfy

$$h^* = \tanh(Ah^* + I).$$

Linearizing the system around an equilibrium yields the Jacobian

$$J_h(h^*) = \tau^{-1}(-I_N + D^* A),$$

where

$$D^* = \text{diag}[\text{sech}^2(Ah^* + I)]$$

denotes the local neuronal gain matrix. Local stability is determined by the spectral abscissa

$$s(J) = \max_k \text{Re}(\lambda_k(J)).$$

A fixed point is locally asymptotically stable when

$$s(J_h) < 0,$$

or equivalently,

$$\max_k \mathrm{Re}(\lambda_k(D^* A)) < 1.$$

This criterion describes local perturbation decay around a specific equilibrium and does not imply global convergence of all trajectories.

B.2 Global contraction and bounded dynamics

Because the nonlinear activation function is bounded and globally Lipschitz, the system admits globally bounded solutions. For two trajectories receiving identical inputs, their distance satisfies

$$\frac{d}{dt}\|h_1 - h_2\|_2 \leq -\frac{1 - \|A\|_2}{\tau}\|h_1 - h_2\|_2.$$

Therefore, $\|A\|_2 < 1$ provides a sufficient condition for global contraction, guaranteeing convergence toward a unique equilibrium under constant input. However, this condition is conservative and is not necessary for stability. More generally, stability can also be assessed through the logarithmic norm of the Jacobian along the state trajectory.

B.3 Effect of low-rank cerebellar feedback on cortical dynamics

The contribution of cerebellar feedback can be expressed as a low-rank perturbation of cortical dynamics:

$$J = J_0 + \frac{\beta}{\tau} DUGV^T,$$

where

$$J_0 = \tau^{-1}(-I_N + D\alpha W).$$

Using the matrix determinant lemma, the characteristic equation can be reduced to an M-dimensional determinant:

$$\det(\lambda I - J) = \det(\lambda I - J_0) \det\left[I_M - \frac{\beta}{\tau} V^T(\lambda I - J_0)^{-1} DUG\right].$$

This formulation shows that low-dimensional cerebellar feedback can selectively reshape cortical eigenmodes through structured feedback pathways. Importantly, although the connectivity modification is low-rank, the resulting change in network dynamics can extend beyond the feedback subspace because altered operating points modify the nonlinear gain matrix D^*. Thus, low-rank feedback can induce distributed changes in cortical dynamics without requiring full-rank connectivity changes.

B.4 Reduced latent dynamics under feedback-dominated conditions

When intrinsic cortical recurrence is removed ($\alpha = 0$), the system admits an exact reduced representation. Defining the latent variable

$$\kappa = V^T h,$$

the dynamics become

$$\tau\dot{\kappa} = -\kappa + V^T \tanh(\beta UG\kappa + I).$$

This reduction provides an exact low-dimensional description of the feedback-driven dynamics without requiring mean-field approximations. The full cortical state can be reconstructed through the corresponding linear filtering process. However, this reduced formulation does not imply that cortical activity is confined to a linear subspace, since nonlinear activation and temporal integration can generate higher-dimensional embeddings.

B.5 Slow modes and dynamical timescales

For stable equilibria, the slowest local dynamical timescale is determined by the eigenvalue with the largest real part:

$$T_{\text{mode}} = -\frac{1}{s(J)}, \qquad s(J) < 0.$$

This quantity characterizes asymptotic modal decay but does not directly define nonlinear settling time. In non-normal recurrent networks, transient amplification can occur even when all eigenvalues indicate stability. Therefore, slow dynamical modes should be distinguished from claims of nonlinear slow manifolds, which additionally require approximate invariance and transverse stability of the corresponding low-dimensional structure.

B.6 Oscillatory dynamics and bifurcation analysis

Potential oscillatory regimes can be analyzed through Hopf bifurcation theory. Along an equilibrium branch, a Hopf bifurcation occurs when a complex conjugate eigenvalue pair crosses the imaginary axis:

$$\lambda_{\pm} = \pm i\omega_H,$$

with a nonzero crossing velocity and appropriate nonlinear criticality conditions. The emergence of stable periodic solutions requires additional verification, including the sign of the first Lyapunov coefficient and numerical validation through return residuals and Floquet multipliers. A closed trajectory observed in a projection space is insufficient evidence for a true periodic attractor or ring-like continuous attractor structure.

B.7 Interpretation and validation requirements

The above analysis establishes mathematical properties of the frozen cortico–cerebellar dynamical system, including feedback rank constraints, local stability conditions, and exact reduced dynamics under specific assumptions. These analytical results describe possible mechanisms through which cerebellar feedback can reshape cortical dynamics, but task-dependent properties such as compositional generalization, memory capacity, or attractor organization require direct evaluation using trained models and controlled perturbation experiments.

C Detailed Experimental Setting

C.1 Environments

All experiments were implemented in Python with PyTorch. Tasks and trials were generated procedurally without external datasets. Training trials were sampled online during optimization, whereas evaluation trials were generated from fixed held-out task combinations and kept identical across models within each split. Each experiment was repeated across multiple random initializations, and reported results represent mean performance across

independent runs. Random seeds were controlled across Python, NumPy, and PyTorch for reproducibility. Experiments were executed on CUDA-enabled GPUs when available.

C.2 Training

We evaluated continual learning performance under two sequential learning settings: a four-task context-dependent sensorimotor benchmark and a larger handwritten-trajectory generation benchmark. Tasks were presented sequentially, with optimization performed on one task at a time. During each phase, mini-batches contained samples exclusively from the current task. After each learning phase, the model was evaluated on held-out trials from all tasks encountered so far to measure task retention and interference during sequential learning.

For the context-dependent sensorimotor benchmark, the task sequence was DelayPro $\rightarrow$ DelayAnti $\rightarrow$ MemoryPro $\rightarrow$ MemoryAnti. Each task was trained for 400 epochs with batch size 128. Trial angles θ were sampled uniformly from $[-\pi, \pi]$. Each trial lasted $T = 60$ time steps, with stimulus onset at $t_{\text{stim}} = 8$ and response onset at $t_{\text{go}} = 36$. The input dimension was 5, consisting of fixation, rule context, timing context, $\sin(\theta)$, and $\cos(\theta)$. The output dimension was 2. Delay tasks retained the sensory cue until the go cue, whereas Memory tasks presented the cue only for an 8-step stimulus epoch. Pro tasks required outputting $[\sin(\theta), \cos(\theta)]$, while Anti tasks required the opposite response, equivalently $[\sin(\theta + \pi), \cos(\theta + \pi)]$. Gaussian input noise with standard deviation 0.04 was added during training, while evaluation used noiseless inputs with 256 trials per task.

For this benchmark, we compared an uncoupled cortex-only RNN with the coupled cortex-cerebellum model. Both models were optimized using Adam with an identical learning rate of 2×10^{-3}, with only the readout layers updated during sequential learning. All recurrent dynamics, input weights, and cortico-cerebellar coupling parameters were kept fixed after initialization. The coupled model used a low-rank feedback module with hidden size $H = 192$, feedback rank $R = 10$, feedback scale 0.65, and leak parameter $\alpha = 0.2$.

The coupled model additionally used local plasticity updates for the task-specific low-rank factors. The V_k factors were updated using an Oja/Hebbian rule with lateral anti-Hebbian decorrelation, and the U_k factors were updated using a three-factor error-modulated rule. The corresponding learning rates were set to $\eta_V = 0.05$ and $\eta_U = 0.03$, respectively. These local updates were performed once per epoch starting from stimulus onset and were applied throughout sequential learning.

For the handwritten-letter trajectory benchmark, the model learned 26 tasks corresponding to uppercase letters A through Z in the strict order $A \rightarrow B \rightarrow \cdots \rightarrow Z$. Each letter was defined by a continuous two-dimensional polyline template, normalized to a common spatial scale and resampled to the response duration. Each trial lasted $T = 64$ time steps, with a preparatory period of $T_{\text{prep}} = 8$ and response onset at $T_{\text{go}} = 20$, yielding $T_{\text{resp}} = 44$ response steps. The input dimension was 28, consisting of a 26-dimensional one-hot letter cue and a 2-dimensional phase signal. The one-hot cue was provided throughout the trial, while the phase signal $[\sin(\phi_t), \cos(\phi_t)]$ advanced only during the response epoch. The target was zero before the go cue and followed the corresponding two-dimensional letter trajectory during the response period.

Each letter task was trained for 100 epochs with batch size 96 and Adam learning rate 2×10^{-3}. During training, trajectories were augmented with small affine-like variations: amplitude jitter with standard deviation 0.05, random rotation with standard deviation 0.12 radians, translation noise with scale 0.04, coordinate noise with standard deviation 0.01, and input noise with standard deviation 0.01. Evaluation used canonical trajectories without jitter or noise, with 128 trials per task. Both benchmarks used a masked mean-squared error objective,

$$\mathcal{L}_{\text{task}} = \frac{1}{BT} \sum_{b,t} m_t \, \|\hat{y}_{b,t} - y_{b,t}\|_2^2 .$$

For the context-dependent sensorimotor tasks, the mask weight was $m_t = 0.10$ before the go cue and $m_t = 1.5$ during the response epoch. For the handwritten-letter benchmark,

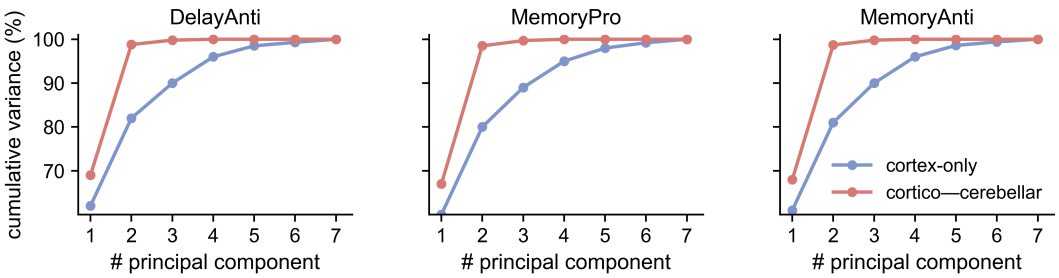

Figure 8: Cumulative variance by principal components for the cortex-only and cortico–cerebellar coupled models.

the pre-response weight was $m_t = 0.05$, and the response weight was again $m_t = 1.5$. This weighting encouraged near-zero outputs before movement onset while emphasizing accurate generation during the response period. After each phase, we recorded new-task performance, old-task retention, full task-wise MSE across phases, and sequential forgetting, defined as the increase in MSE between the phase in which a task was first learned and the final evaluation.

D Supplementary results

D.1 Cumulative Variance Explained by Principal Components Across Four Cognitive Tasks

We illustrated the cumulative variance explained by the principal components of population activity for both the cortex-only and cortico–cerebellar coupled models across the four cognitive tasks, as shown in Figure 6. Notably, the coupled model (red) captures near-complete variance using substantially fewer principal components than the cortex-only model in all tasks. This demonstrates that cerebellar feedback significantly compresses cortical dynamics.

D.2 Learning Speed Across Four Cognitive Tasks

We analyzed the training dynamics—specifically, the mean squared error (MSE) curves—for both the cortex-only and cortico–cerebellar coupled models. As illustrated in the top panels of Figure 7, although both models eventually achieve a low MSE, the coupled model converges substantially faster than the cortex-only model across all tasks. Quantification of this effect using fast and slow exponential fits in bottom panels reveals that the coupled model exhibits consistently larger decay values than the cortex-only baseline. This demonstrates that cerebellar feedback significantly accelerates both the rapid initial reduction in error and the slower refinement phase of learning.

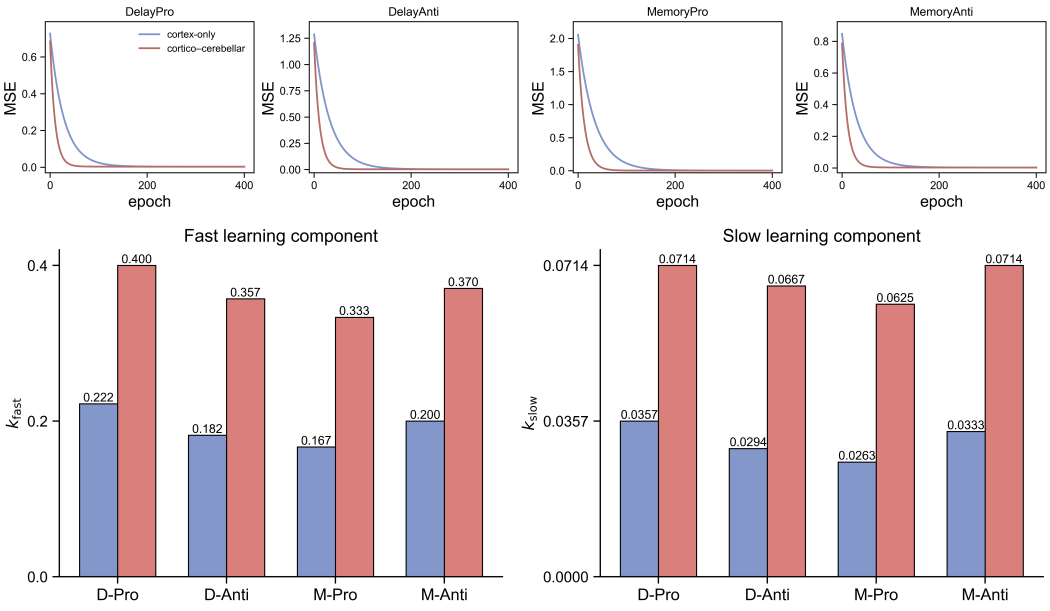

Figure 9: Comparison of training dynamics and learning speeds between the cortex-only and cortico–cerebellar coupled models.