PREFERENCE CONDITIONED STATE PERSISTENCE FOR SPIKING SEQUENTIAL RECOMMENDATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Sequential recommendation relies on recurrent states that retain interaction history while remaining responsive to the current preference context. Adaptive spiking neurons can modulate state persistence, but existing formulations primarily characterize retention coefficients rather than how those coefficients are assigned to recurrent state groups. We introduce Preference-Conditioned Dynamics (PCD), which predicts groupwise retention from the current preference representation and recurrent history. The resulting recurrent transition is determined not only by the retention coefficients but also by their assignment to fixed state groups. To isolate this assignment, we replay learned membrane-retention sequences and permute coefficients across groups while preserving the coefficient multiset and mean at every event. On Amazon Video Games, reassignment produces measurable changes in PCD ranking metrics and substantially smaller responses in a Dynamic Gated Neuron comparator with the same recurrent-cell parameter count. MovieLens 1M reproduces this contrast for normalized discounted cumulative gain and mean reciprocal rank at cutoff 10. Controlled training variants and complementary CIFAR-10 experiments further show that the learned retention structure depends on conditioning-feature correspondence and on gradient access to that feature. These results show that marginal retention statistics are insufficient to characterize the functional state-persistence structure learned by recurrent Spiking Neural Networks.

1 INTRODUCTION

Sequential recommendation requires recurrent states to preserve useful interaction history while allowing the influence of that history to vary with the current preference context. A single retention coefficient imposes a common carryover on the recurrent state, whereas groupwise retention permits different state components to evolve over different effective timescales and to adapt those timescales to context. Heterogeneous and adaptive timescales have long been used to distribute memory across recurrent units (Alpay et al., 2016; Quax et al., 2020).

Modern sequential recommenders already condition the influence of historical interactions on the prediction context. Attention and temporal encoding reweight past interactions according to current content or elapsed time, while continuous-time and state-space formulations modify the state transition itself (Kang & McAuley, 2018; Li et al., 2020; Neupane et al., 2024; Xiao et al., 2025). SSR realizes sequential recommendation with spike representations and temporal filters (Di Yu et al., 2025). Selective state-space models further make transition parameters input dependent, allowing current content to regulate state propagation and forgetting (Gu & Dao, 2023).

Spiking Neural Networks (SNNs) likewise support adaptive state persistence through neuronal dynamics. Learnable membrane time constants, gated or multi-compartment neurons, heterogeneous timescales, fractional-order memory, and activity-dependent conductance distribute temporal integration across membrane, dendritic, synaptic, or conductance variables (Fang et al., 2021; Yao et al., 2022; Zhang et al., 2024; Zheng et al., 2024; Ge et al., 2026; Bai et al., 2026). External or input-dependent signals can further modulate thresholds or membrane time constants (Helfer et al., 2023; Chen et al., 2023; Luca et al., 2026; Zhang et al., 2025a; 2026), and the Dynamic Gated Neuron (DGN) makes conductance, and hence temporal retention, depend on neuronal activity. These mechanisms establish that spiking state persistence can be heterogeneous and context dependent, but

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.

they characterize how retention values are generated rather than how realized values are coupled to fixed recurrent state groups.

Once groupwise retention coefficients are realized, the recurrent transition depends not only on their values but also on which fixed state groups receive them. We refer to this event-specific coefficient-to-state correspondence as persistence assignment. In the recurrent update considered here, each coefficient acts through its group's state discrepancy, the difference between the carried state and the current recurrent input. The same coefficient multiset can therefore induce different transitions under different assignments. The resulting question is whether training learns a task-dependent coefficient-to-state correspondence with functional consequences beyond those implied by the coefficient values alone. Because coefficient values and their assignment normally change together during training, isolating this relation requires holding the realized coefficients fixed while changing only their correspondence with recurrent groups.

To study this question in sequential recommendation, we introduce Preference-Conditioned Dynamics (PCD). PCD partitions recurrent channels into fixed groups and predicts groupwise retention from the current preference representation, recurrent history, and elapsed time. We isolate persistence assignment with frozen-sequence replay: the membrane-retention sequence is recorded on the unperturbed trajectory and replayed under prescribed group permutations, preserving the coefficient values while changing their correspondence with recurrent groups during nonlinear rollout. For the matched recommendation comparison, DGN-CM retains the same recurrent-cell parameter count as PCD. Figure 1 illustrates preference-conditioned persistence and the replay intervention.

In recommendation, coefficient-preserving replay changes ranking under PCD and produces larger reassignment responses than the matched DGN-CM comparator on Amazon Video Games and MovieLens 1M. Training controls further show that the learned retention structure changes with conditioning-feature correspondence and gradient access while task performance remains similar. Complementary CIFAR-10 experiments provide a controlled cross-domain test of the same dependence.

Our main contributions are summarized as follows:

- We formulate Preference-Conditioned Dynamics for spiking sequential recommendation and formalize persistence assignment as the correspondence between realized retention coefficients and fixed recurrent state groups.
- We derive the exact state change induced by coefficient-preserving permutation and use frozen-sequence replay to isolate its functional consequence at fixed coefficient values; PCD exhibits direct ranking responses and larger paired reassignment effects than DGN-CM across two recommendation datasets.
- We show through controlled changes in conditioning-feature correspondence and gradient access, together with complementary CIFAR-10 experiments, that learned retention structure depends on the information available to the persistence predictor during training.

2 Related Work

Context-Dependent State Evolution. Sequential models already vary state evolution with content and time. Attention recommenders reweight historical items or elapsed intervals, while continuous-time and state-space recommenders alter user state transitions directly (Kang & McAuley, 2018; Li et al., 2020; 2024; Neupane et al., 2024; Xiao et al., 2025; Zhang et al., 2025b; Wang et al., 2025). TiM4Rec inserts temporal modulation into structured state-space duality for recommendation (Fan et al., 2025). Beyond recommendation, Mamba makes state-space parameters depend on the input (Gu & Dao, 2023), and recent analysis shows that the latent modes carrying output can migrate with input (Bhattacharya, 2026).

Heterogeneous Intrinsic Spiking Dynamics. Temporal adaptation in SNNs has been introduced at several loci of intrinsic neuronal dynamics. Within the leaky integrate-and-fire (LIF) family, the Parametric LIF (PLIF) neuron learns membrane decay, whereas the Gated LIF (GLIF) neuron combines gated neuronal behaviors (Fang et al., 2021; Yao et al., 2022). Two-Compartment LIF (TC-LIF) neurons and temporal dendritic heterogeneity distribute temporal integration across somatic and dendritic states (Zhang et al., 2024; Zheng et al., 2024). HetSyn assigns time constants to

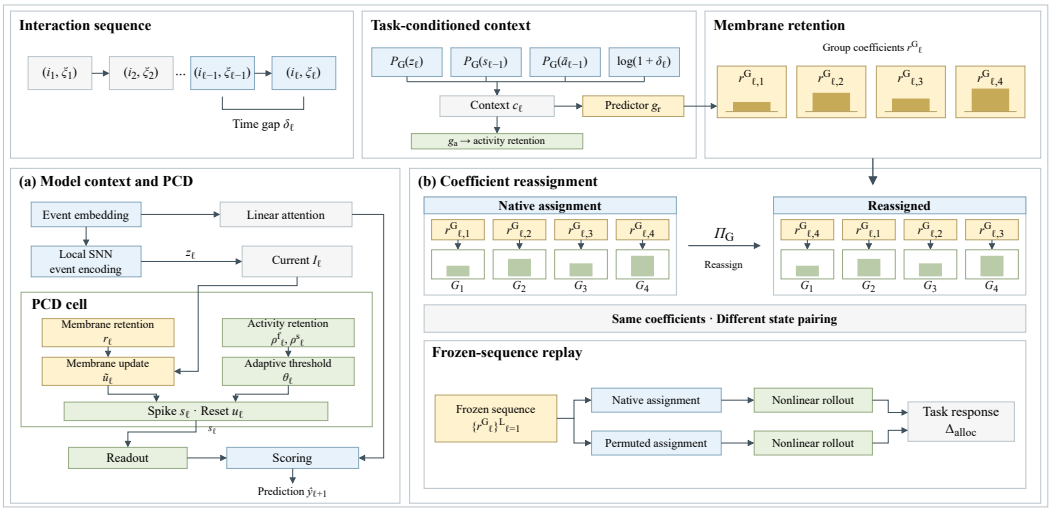

Figure 1: Persistence assignment and frozen-sequence replay in PCD. (a) Group membrane retention is conditioned on the preference representation, recurrent history, and elapsed time. (b) Replay preserves the captured coefficient values and permutes only their assignment to recurrent groups before nonlinear rollout.

individual synapses and moves heterogeneous integration into synaptic current (Deng et al., 2026). Fractional-order neurons extend the temporal kernel beyond first-order Markov carryover (Ge et al., 2026). These methods place heterogeneous temporal integration in membrane, dendritic, synaptic, or fractional-order dynamics.

Conditioned Intrinsic Spiking Dynamics. Other work makes intrinsic dynamics depend on neuronal activity or external information. Context modulation and the Contextual Embedding LIF (CE-LIF) neuron alter circuit or threshold dynamics (Helfer et al., 2023; Chen et al., 2023). The brain-inspired adaptive LIF (BA-LIF) neuron predicts membrane time constants from the current input; a Temporal Focus-based SNN (TF-SNN) and PAS-Net regulate thresholds from sample- or context-dependent signals (Zhang et al., 2025a; 2026; Zheng et al., 2026). DGN makes conductance activity-dependent (Bai et al., 2026), while learned global perturbations adapt intrinsic parameters to robustness demands (Luca et al., 2026). These mechanisms determine how retention or threshold values are generated. Persistence assignment concerns the subsequent correspondence between realized retention coefficients and fixed recurrent state groups.

Spiking Sequence Models and Control. Spiking sequence models alter representation, temporal coding, or recurrent state evolution for long-range dependencies (Lv et al., 2024b;a; Bal & Sengupta, 2025); Signal-SGN and MK-SGN extend this direction to structured skeleton dynamics (Zheng et al., 2025a;b), and SSR to sequential recommendation (Di Yu et al., 2025). In control, Proxy Target and CaRe-BN target SNN-specific training and normalization, while Cross-Step Residual Potential Initialization (CRPI) carries residual membrane potentials across decision steps (Xu et al., 2026a;b;c). These transformations operate on representation, optimization, normalization, or state carryover. Persistence assignment addresses the coefficient-to-fixed-state-group correspondence after coefficient realization.

3 Task-Conditioned Persistence Assignment

Consider a recurrent spiking state updated once per behavioral event ℓ. The event index ℓ carries persistent state, whereas τ indexes only the local neural steps used to encode the current event. We first define persistence assignment independently of application; the experiments instantiate task-conditioned group retention as PCD in recommendation and TCD in vision.

3.1 Event-Indexed Reference Dynamics

Let $\mathbf{h}_\ell \in \mathbb{R}^H$ be the contextualized feature of the current event and $\mathbf{s}^{\text{in}}_{\ell,\tau} \in \{0,1\}^H$ the output of a local LIF encoder. Averaging its T_s inner steps defines the event representation

$$\mathbf{z}_\ell = \frac{1}{T_s} \sum_{\tau=1}^{T_s} \mathbf{s}^{\text{in}}_{\ell,\tau} \in [0,1]^H.$$

The encoder is stateless along ℓ. Persistent state begins in the recurrent cell. Let $\mathbf{u}_{\ell-1} \in \mathbb{R}^H$ and $\mathbf{s}_{\ell-1} \in \{0,1\}^H$ denote membrane and spike state carried from the preceding valid event. Using a scalar coefficient r gives the fixed-retention reference

$$\begin{aligned} \mathbf{I}_\ell &:= \kappa(W_z \mathbf{z}_\ell + \mathbf{b}_z + W_s \mathbf{s}_{\ell-1}), \\ \widetilde{\mathbf{u}}^{\text{ref}}_\ell &:= r\, \mathbf{u}_{\ell-1} + (1-r)\, \mathbf{I}_\ell - \boldsymbol{\gamma} \odot \mathbf{s}_{\ell-1}. \end{aligned} \tag{1}$$

Here $W_z, W_s \in \mathbb{R}^{H\times H}$, $\mathbf{b}_z \in \mathbb{R}^H$, $\kappa > 0$, $\boldsymbol{\gamma} \in \mathbb{R}^H_+$, and $r \in (0,1)$. The scalar r interpolates between the carried membrane state and the new recurrent current, while the final term is feedback from the preceding spike. Throughout the paper, r denotes retention over one event. The reset at the same event is defined in Eq. (12).

3.2 Persistence Assignment with Fixed Recurrent Groups

Partition the H recurrent channels into G equal groups $\{\mathcal{G}_g\}_{g=1}^G$. Let $\mathcal{P}_G : \mathbb{R}^H \to \mathbb{R}^G$ average within groups and $\mathcal{E}_G : \mathbb{R}^G \to \mathbb{R}^H$ broadcast a group value to its member channels, so $\mathcal{P}_G(\mathcal{E}_G(\mathbf{v})) = \mathbf{v}$. The recurrent-channel basis and equal-size partition remain fixed throughout. We call the event-specific correspondence between realized coefficients and these groups persistence assignment. Let $\mathbf{r}^G_\ell \in (r_{\min}, r_{\max})^G$ denote membrane retention for each group and $\mathbf{r}_\ell = \mathcal{E}_G(\mathbf{r}^G_\ell)$. Replacing the scalar r in Eq. (1) yields

$$\begin{aligned} \widetilde{\mathbf{u}}_\ell &:= \mathbf{r}_\ell \odot \mathbf{u}_{\ell-1} + (\mathbf{1} - \mathbf{r}_\ell) \odot \mathbf{I}_\ell - \boldsymbol{\gamma} \odot \mathbf{s}_{\ell-1} \\ &= \mathbf{I}_\ell + \mathbf{r}_\ell \odot \mathbf{d}_\ell - \boldsymbol{\gamma} \odot \mathbf{s}_{\ell-1}, \qquad \mathbf{d}_\ell := \mathbf{u}_{\ell-1} - \mathbf{I}_\ell. \end{aligned} \tag{2}$$

We refer to $\mathbf{d}_\ell$ as the state discrepancy. A retention coefficient affects the transition through the state discrepancy it multiplies. Within the configured retention range, the reference dynamics are recovered exactly when $\mathbf{r}^G_\ell = r\mathbf{1}$.

To change assignment without changing coefficient values, let $\Pi_G \in \{0,1\}^{G\times G}$ be a group permutation and set $\mathbf{r}^\Pi_\ell = \mathcal{E}_G(\Pi_G \mathbf{r}^G_\ell)$. Holding the state before intervention and the recurrent current fixed,

$$\Delta_\Pi \widetilde{\mathbf{u}}_\ell = (\mathbf{r}^\Pi_\ell - \mathbf{r}_\ell) \odot \mathbf{d}_\ell. \tag{3}$$

Writing $\mathbf{d}_{\ell,g} = (\mathbf{d}_\ell)_{\mathcal{G}_g}$, the disjoint channel support of the recurrent groups gives the exact decomposition

$$\|\Delta_\pi \widetilde{\mathbf{u}}_\ell\|_2^2 = \sum_{g=1}^G (r^G_{\ell,\pi(g)} - r^G_{\ell,g})^2 \, \|\mathbf{d}_{\ell,g}\|_2^2 . \tag{4}$$

The permutation changes only group ordering: the coefficient multiset, its mean, and every symmetric marginal statistic are identical. A histogram or marginal distribution of $\mathbf{r}_\ell$ records which coefficient values are present but discards their pairing with $\mathbf{d}_\ell$. The same retention multiset can produce different transitions when its coefficients are paired with different state discrepancies. Averaging Eq. (4) over a uniformly random group permutation gives:

$$\mathbb{E}_\pi \left[\|\Delta_\pi \widetilde{\mathbf{u}}_\ell\|_2^2 \right] = \sum_{g=1}^G \|\mathbf{d}_{\ell,g}\|_2^2 \left[\sigma^2_{r,\ell} + (r^G_{\ell,g} - \bar{r}^G_\ell)^2 \right], \tag{5}$$

where $\bar{r}^G_\ell = G^{-1} \sum_g r^G_{\ell,g}$ and $\sigma^2_{r,\ell} = G^{-1} \sum_g (r^G_{\ell,g} - \bar{r}^G_\ell)^2$. Expected displacement vanishes for homogeneous group persistence or zero state discrepancy. Later spikes and resets propagate this one-event perturbation through the nonlinear recurrent rollout. Appendix A.3 derives the permutation expectation and its first-order loss extension.

3.3 Task-Conditioned Retention and Gradient Access

For the PCD recurrent realization, the group coefficients are predicted from the current task representation and recurrent history. Let $\mathbf{a}^f_{\ell-1}, \mathbf{a}^s_{\ell-1} \in [0,1]^H$ be two activity traces from the preceding event, $\bar{\mathbf{a}}_{\ell-1} = \frac{1}{2}(\mathbf{a}^f_{\ell-1} + \mathbf{a}^s_{\ell-1})$, and δ_ℓ the recorded gap from the preceding valid event. The input to the retention predictor is

$$\mathbf{c}_\ell = \big[\mathcal{P}_G(\mathbf{z}_\ell), \mathcal{P}_G(\mathbf{s}_{\ell-1}), \mathcal{P}_G(\bar{\mathbf{a}}_{\ell-1}), \log(1+\delta_\ell)\big] \in \mathbb{R}^{3G+1}. \tag{6}$$

The predictor combines current task content with preceding spike state, accumulated activity, and recorded event spacing. A membrane branch $g_r : \mathbb{R}^{3G+1} \to \mathbb{R}^G$ and an activity branch $g_a : \mathbb{R}^{3G+1} \to \mathbb{R}^{2G}$ map this context to bounded group coefficients. Writing $[\mathbf{v}^f_\ell, \mathbf{v}^s_\ell] = \mathrm{split}_{G,G}(g_a(\mathbf{c}_\ell))$,

$$\begin{aligned}
\mathbf{r}^G_\ell &= r_{\min}\mathbf{1} + (r_{\max} - r_{\min})\sigma(g_r(\mathbf{c}_\ell)),\\
\boldsymbol{\rho}^{f,G}_\ell &= \rho^f_{\min}\mathbf{1} + (\rho^f_{\max} - \rho^f_{\min})\sigma(\mathbf{v}^f_\ell),\\
\boldsymbol{\rho}^{s,G}_\ell &= \rho^s_{\min}\mathbf{1} + (\rho^s_{\max} - \rho^s_{\min})\sigma(\mathbf{v}^s_\ell),\\
\mathbf{r}_\ell &= \mathcal{E}_G(\mathbf{r}^G_\ell), \qquad \boldsymbol{\rho}^p_\ell = \mathcal{E}_G(\boldsymbol{\rho}^{p,G}_\ell), \quad p \in \{f,s\}.
\end{aligned} \tag{7}$$

Each branch outputs G group-specific retention coefficients, which are broadcast over the fixed recurrent channel groups.

Gradients from the loss reach the task representation through the same retention predictor. Let $\mathbf{q}_\ell = \mathcal{P}_G(\mathbf{z}_\ell)$, $\mathbf{d}_{\ell,g} = (\mathbf{d}_\ell)_{\mathcal{G}_g}$, and $\mathbf{v}^{\mathrm{sg}}_{\ell,g} = \nabla^{\mathrm{sg}}_{\widetilde{\mathbf{u}}_{\ell,\mathcal{G}_g}} \mathcal{J}$ for the adopted surrogate backward graph. Then

$$\begin{aligned}
\chi_{\ell,g} &:= \left\langle \mathbf{v}^{\mathrm{sg}}_{\ell,g}, \mathbf{d}_{\ell,g} \right\rangle = \frac{\partial^{\mathrm{sg}}\mathcal{J}}{\partial r^G_{\ell,g}},\\
\nabla^{\mathrm{ret,sg}}_{\mathbf{q}_\ell}\mathcal{J} &= J^\top_{\mathbf{r}^G_\ell,\mathbf{q}_\ell}\boldsymbol{\chi}_\ell, \qquad \boldsymbol{\chi}_\ell = (\chi_{\ell,1}, \ldots, \chi_{\ell,G})^\top.
\end{aligned} \tag{8}$$

The first line is the signed coefficient sensitivity induced by the state discrepancy; the second composes it with the Jacobian of the retention predictor with respect to the task representation. Applying stop-gradient to $\mathbf{q}_\ell$ leaves its forward values unchanged while removing the gradient from retention to that representation.

At the same event, the current feature has two distinct routes into the pre-reset membrane. Holding the prior recurrent state fixed,

$$\left.\frac{d\widetilde{\mathbf{u}}_\ell}{d\mathbf{z}_\ell}\right|_{\mathbf{u}_{\ell-1},\mathbf{s}_{\ell-1}} = \mathrm{Diag}(\mathbf{1} - \mathbf{r}_\ell)\,\kappa W_z + \mathrm{Diag}(\mathbf{d}_\ell)\frac{\partial \mathbf{r}_\ell}{\partial \mathbf{z}_\ell}. \tag{9}$$

The first term is the direct recurrent-current path and the second is the retention-mediated path. Feature reassignment changes the source supplied to the retention predictor while leaving the direct current path fixed; stop-gradient preserves the predictor's forward feature values but removes the retention-mediated backward route. Appendix A.2 gives the grouped derivative form.

The same signed coefficient sensitivity yields a local first-order ordering criterion for assignment. Because the experiments do not estimate this local covariance or pairwise ordering directly, we keep the finite nonlinear replay as the empirical object and give the first-order permutation, covariance, and swap relations in Appendix A.3.

3.4 History-Dependent Threshold Adaptation and Spiking Transition

The recurrent state also contains two activity traces for history-dependent threshold adaptation. Their updates are

$$\begin{aligned}
\mathbf{a}^f_\ell &:= \boldsymbol{\rho}^f_\ell \odot \mathbf{a}^f_{\ell-1} + (\mathbf{1} - \boldsymbol{\rho}^f_\ell) \odot \mathbf{s}_{\ell-1},\\
\mathbf{a}^s_\ell &:= \boldsymbol{\rho}^s_\ell \odot \mathbf{a}^s_{\ell-1} + (\mathbf{1} - \boldsymbol{\rho}^s_\ell) \odot \mathbf{s}_{\ell-1}.
\end{aligned} \tag{10}$$

Each update is a convex interpolation between the carried trace and the preceding spike, so traces initialized at zero remain in $[0,1]^H$. Larger $\rho^{f,s}_{\ell,h}$ places more weight on carried activity and less on the new spike.

The updated traces set the firing threshold,

$$\boldsymbol{\theta}_\ell := \boldsymbol{\theta}_0 + \boldsymbol{\beta}_f \odot \mathbf{a}_\ell^f + \boldsymbol{\beta}_s \odot \mathbf{a}_\ell^s, \tag{11}$$

where $\boldsymbol{\theta}_0, \boldsymbol{\beta}_f, \boldsymbol{\beta}_s \in \mathbb{R}_+^H$. Positive gains make accumulated activity raise threshold and reduce immediate firing propensity. The neuron then emits a binary spike vector and applies a multiplicative zero reset at the same event,

$$\mathbf{s}_\ell = \mathbb{I}[\widetilde{\mathbf{u}}_\ell - \boldsymbol{\theta}_\ell \geq 0], \qquad \mathbf{u}_\ell = \widetilde{\mathbf{u}}_\ell \odot (\mathbf{1} - \mathbf{s}_\ell). \tag{12}$$

At event ℓ, retention is predicted from $\mathbf{z}_\ell$ and states available through $\ell - 1$. The activity traces are then updated, the threshold is formed, and the current spike is emitted. Recorded spacing changes the coefficients without inserting recurrent events, while τ remains confined to the local encoder. Appendix A.1 specifies the exact pooling, retention parameterization, surrogate reset, elapsed gap, and auxiliary terms.

4 Experiments

We evaluate coefficient reassignment on Video and ML-1M, and task-feature source and gradient access on Video and CIFAR-10. Alternative state and temporal realizations are reported in Appendix D.1.

4.1 Evaluation Protocols

On Video, the PCD–DGN-CM comparison uses 20 paired training runs with the same chronological split by user and full ranking evaluation. We report normalized discounted cumulative gain at cutoff 10 (NDCG@10), mean reciprocal rank at cutoff 10 (MRR@10), and Recall@10. DGN-CM generates adaptive membrane persistence from neuronal variables with the same recurrent-cell parameter count as PCD and excludes the current task feature from the retention predictor. Repeated reassignments after training are averaged within each model before uncertainty is estimated across independently trained models. A five-run comparison either reassigns the task feature during training or moves modulation by the current event from intrinsic retention to the representation. Protocol and integrity details appear in Appendices B.2 and B.3.

ML-1M repeats replay of fixed membrane retention coefficients with 10 paired seeds, longer histories, and sampled leave-one-out evaluation. CIFAR-10 (Krizhevsky et al., 2009) uses the same spiking ResNet-19 architecture and the same retention predictor parameter count to vary feature matching and gradient access over five paired runs. Dataset statistics and recommendation protocols are summarized in Table B1; optimization and recurrent-architecture details are provided in Appendices A.1 and D.2.

4.2 Coefficient-Preserving Reassignment

PCD and DGN-CM generate adaptive persistence from different inputs. PCD conditions retention on the current task representation together with recurrent firing, adaptation history, and recorded spacing. Full ranking metrics are reported in Table B2.

Feature reassignment substitutes only the task feature entering the retention predictor; recurrent history, recorded spacing, the main recommendation feature path, and all model parameters are unchanged. Coefficient reassignment is the group-permutation intervention of Section 3.2, applied to the captured membrane-retention sequence.

Task feature reassignment affects all three retention branches. The membrane branch shows no detectable global-mean shift, whereas the activity branches do; only membrane-coefficient reassignment produces a detectable ranking change. Subsequent coefficient reassignment uses the membrane branch. Appendix B.2 reports the complete branch estimates and confidence intervals.

For each trained model, we record the membrane-retention sequence on the unperturbed trajectory and use that sequence exogenously during replay. For each model or seed, Δ denotes the original-assignment metric minus the reassigned metric.

Table 1: Direct and paired recommendation responses to coefficient reassignment.

Dataset	Metric	PCD Δ	DGN-CM Δ	Interaction [95% CI]	W/T/L
Video	NDCG@10	+0.000796	+0.000037	+0.000759 [+0.000491, +0.001028]	17/0/3
	MRR@10	+0.000556	+0.000044	+0.000512 [+0.000305, +0.000718]	16/0/4
	Recall@10	+0.001590	+0.000018	+0.001572 [+0.001067, +0.002077]	18/0/2
ML-1M	NDCG@10	+0.015604	+0.006243	+0.009362 [+0.003742, +0.014982]	9/0/1
	MRR@10	+0.016400	+0.005965	+0.010435 [+0.004684, +0.016185]	9/0/1
	Recall@10	+0.012397	+0.006970	+0.005426 [-0.000123, +0.010975]	7/0/3

At every recorded event, the prescribed group permutation is applied to the captured coefficients while the recurrent trajectory evolves under the resulting nonlinear rollout. Table 1 reports the direct mean response in each state law and the paired interaction $\Delta_{\mathrm{PCD}} - \Delta_{\mathrm{DGN\text{-}CM}}$; positive interaction indicates a larger response in PCD. Paired 95% CIs and win/tie/loss counts (W/T/L) apply to the interaction. Figure 1(b) illustrates the replay protocol.

On Video, the direct PCD response is much larger than the DGN-CM response across all three ranking metrics; for NDCG@10, PCD changes by $+0.000796$ (95% CI $[+0.000555, +0.001037]$), whereas DGN-CM changes by $+0.000037$ with an interval spanning zero. All three paired interaction intervals lie above zero. On ML-1M, the direct changes are again larger for PCD, with NDCG@10 and MRR@10 interactions above zero and Recall@10 unresolved. Every nonzero replay permutation preserves the sorted retention multiset, and unpermuted replay reproduces the stored metrics and coefficient traces. Full integrity checks appear in Appendix B.3.

4.3 Training Dependence of Persistence Assignment

Three variants, each evaluated over five runs, change the task feature used during training or the location of current-event modulation while keeping the predictor parameter count constant. They compare intrinsic conditioning with the native feature, intrinsic conditioning trained with reassigned task features, and representation-level modulation. In the last condition, the intrinsic retention predictor receives no current task feature.

Here ΔNDCG@10 is the original-feature score minus the feature-reassigned score, and retention displacement is the paired mean absolute difference over event, time, and recurrent group entries.

Ranking performance under original features remains similar across the three variants, but the reassignment-induced ranking change decreases from aligned intrinsic conditioning to reassigned-feature training and is smallest under representation-level modulation (Table 2). Paired confidence intervals support both the ranking and membrane-displacement contrasts; Appendix B.4 reports the corresponding estimates.

We also pair each target feature bijectively with a near or far feature donor in representation space. The two assignments produce different cosine similarities, but the confidence interval for their difference in ranking change includes zero. Across the tested bijective assignments, ranking change is not monotonic in pairwise feature distance. Appendix B.4 specifies the construction and the comparison with a non-bijective donor assignment.

4.4 Feature-Source Organization under Gradient Access

On CIFAR-10, we vary task-feature source and gradient access from retention to that feature. Before the factorial training comparison, we vary the information available to retention under original-feature evaluation. DGN-CM uses summaries of neuronal state, PLIF uses globally learned persistence, HistoryOnly uses spike and adaptation history, FeatureOnly uses the current visual feature, and Task-Conditioned Dynamics (TCD) combines the current feature with activity history. Variants with a retention predictor use the same predictor parameter count. Original-feature Top-1 across these persistence inputs is summarized as mean ± standard deviation over five runs in Table 3.

HistoryOnly, FeatureOnly, and TCD achieve similar original-feature accuracy despite using different retention inputs (Table 3). We cross feature matching (aligned or shuffled) with gradient access (open

Table 2: Video controls for persistence assignment.

	Ranking		Retention displacement		
Condition	**Original NDCG@10**	**Reassignment ΔNDCG@10**	**Membrane**	**Fast**	**Slow**
Aligned intrinsic	0.08536	0.000177	0.0262	0.0723	0.01050
Reassigned feature training	0.08576	0.000066	0.0094	0.0389	0.01220
Representation level	0.08573	0.000029	0.0026	0.0018	0.00052

Table 3: Persistence inputs and accuracy on CIFAR-10.

Dynamics	**Persistence information**	**Original Top-1 (%)**
DGN-CM	summaries of neuronal state	89.454 ± 0.314
PLIF	globally learned persistence	90.260 ± 0.420
HistoryOnly	spike + adaptation history	90.716 ± 0.186
FeatureOnly	current task feature	90.490 ± 0.229
TCD	feature + activity history	90.870 ± 0.246

or detached). Detachment preserves the forward task feature while blocking its backward route, whereas shuffling changes which sample provides that feature during training. This factorial design yields TCD, TCD-Detach, TCD-Shuffle, and TCD-Shuffle-Detach. All four training conditions use each sample's own feature for task-accuracy evaluation.

After training, feature reassignment leaves the image and prediction pathway unchanged while replacing the feature supplied to the retention predictor. A same-class donor supplies a feature from another image with the same class, whereas a cross-class donor supplies a feature from a different class. We define the same-to-cross accuracy gap as Top-1 accuracy under same-class reassignment minus Top-1 accuracy under cross-class reassignment for the same trained model and evaluation order. For source-class robustness, donor classes are selected by cyclic offsets $+1, +3, +7, +9$ modulo ten, and repeated source-image realizations are averaged within each model.

For source indexing, each reassigned sample is indexed by the class that supplied its donor feature before computing the groupwise retention correlation with retention under original-feature evaluation. Target indexing and source indexing operate on the same reassigned samples and differ only in the class index. Writing $G_{a,g}$ for the same-to-cross gap under feature factor $a \in \{\text{aligned}, \text{shuffled}\}$ and gradient factor $g \in \{\text{open}, \text{detached}\}$, the alignment-by-gradient interaction is

$$\Delta\Delta = (G_{\text{aligned,open}} - G_{\text{aligned,detached}}) - (G_{\text{shuffled,open}} - G_{\text{shuffled,detached}}).$$

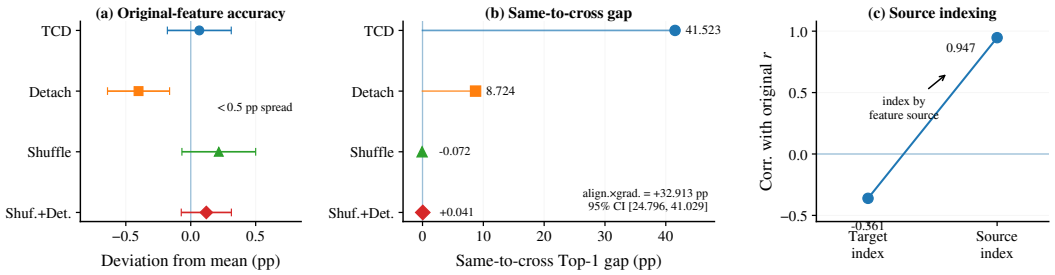

Figure 2: Feature-source organization of retention. (a) Original-feature Top-1. (b) Same-to-cross gaps under feature matching and gradient detachment. (c) Membrane-retention correlation under target- and donor-class indexing.

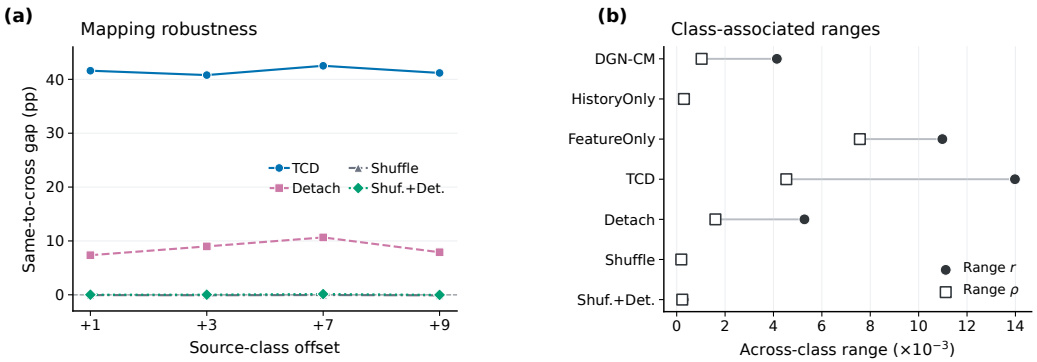

Figure 3: CIFAR-10 source robustness. (a) Same-to-cross gaps by donor-class offset. (b) Membrane and activity retention ranges across conditions.

Original-feature accuracy is similar across conditions, while gradient detachment sharply reduces the same-to-cross gap (Figure 2(a,b)). Donor-class indexing reveals source-associated membrane retention with little global shift (Figure 2(c)).

The same-to-cross gap varies little across the tested donor-class offsets (Figure 3(a)). Random pairing preserves a positive gap for TCD and TCD-Detach, whereas the shuffled-feature intervals include zero. Class-associated retention ranges are largest for TCD and contract under gradient detachment or shuffled-feature training; random pairing compresses them further (Figure 3(b)). Removing task context remains detrimental. Conditions with larger class-associated retention ranges also show larger reassignment gaps, while global retention level and original task accuracy remain comparatively stable. Appendix C.3 reports the complete control estimates.

5 Discussion

Persistence heterogeneity and assignment are distinct properties of vector-valued recurrence. Heterogeneity specifies which coefficient values are present; assignment specifies which recurrent state discrepancies they scale. A coefficient-preserving permutation leaves marginal coefficient statistics unchanged while altering this correspondence, and Eq. (5) makes its joint dependence on coefficient heterogeneity and state discrepancy explicit. Recommendation replay shows that the learned correspondence affects task outputs at fixed coefficient values, while training controls show that its organization changes with task-feature source and gradient access.

Means, ranges, and histograms describe available persistence values but discard which recurrent coordinates receive them. Similar task performance and marginal operating points can coexist with different responses to assignment or feature-source interventions. Thus coefficient-to-state correspondence can carry functional information absent from marginal summaries.

Persistence assignment is event indexed rather than a static memory-length parameter. Subsequent state-dependent spikes and resets propagate a local reassignment, and frozen-sequence replay measures the resulting finite nonlinear consequence. The surrogate sensitivity analysis gives a possible local training route rather than an empirical diagnostic. We keep the recurrent partition fixed; multi-event coefficient products and homogeneous horizon diagnostics are reported in Appendix A.1, while learned partitions and other vector-gated dynamics remain outside scope.

6 Conclusion

Vector-valued persistence contains an assignment degree of freedom beyond its realized coefficient values. Coefficient-preserving replay shows functional consequences, while controlled recommendation and vision interventions show that its organization changes with task-feature source and gradient access. Marginal coefficient statistics therefore do not fully describe the realized recurrent transition.

AI USE STATEMENT

Generative artificial intelligence assisted literature synthesis, result interpretation and manuscript editing. The authors independently verified all assisted content and take responsibility for the final work.

REPRODUCIBILITY STATEMENT

Experimental protocols, implementation details, and theoretical derivations are documented in Appendices A–D. The anonymized supplementary code provides the corresponding experiment and analysis scripts.

REFERENCES

Tayfun Alpay, Stefan Heinrich, and Stefan Wermter. Learning multiple timescales in recurrent neural networks. In *International conference on artificial neural networks*, pp. 132–139. Springer, 2016.

Qianyi Bai, Wang Wang, and Qiang Yu. A brain-inspired gating mechanism unlocks robust computation in spiking neural networks. In *International Conference on Learning Representations*, volume 2026, pp. 156097–156129, 2026.

Malyaban Bal and Abhronil Sengupta. P-spikessm: Harnessing probabilistic spiking state space models for long-range dependency tasks. In *International Conference on Learning Representations*, volume 2025, pp. 73912–73927, 2025.

Raktim Bhattacharya. An exact instrument for state usage in selective state-space models, and the input-driven migration it reveals. *arXiv preprint arXiv:2607.11796*, 2026.

Xinyi Chen, Jibin Wu, Huajin Tang, Qinyuan Ren, and Kay Chen Tan. Unleashing the potential of spiking neural networks for sequential modeling with contextual embedding. *arXiv preprint arXiv:2308.15150*, 2023.

Zhichao Deng, Zhikun Liu, Junxue Wang, Shengqian Chen, Xiang Wei, and Qiang Yu. Hetsyn: Versatile timescale integration in spiking neural networks via heterogeneous synapses. volume 38, pp. 40915–40946, 2026.

Changze Lv Di Yu, Xin Du, Linshan Jiang, Qing Yin, Wentao Tong, Xiaoqing Zheng, and Shuiguang Deng. Cost-effective on-device sequential recommendation with spiking neural networks. 2025.

Hao Fan, Mengyi Zhu, Yanrong Hu, Hailin Feng, Zhijie He, Hongjiu Liu, and Qingyang Liu. Tim4rec: An efficient sequential recommendation model based on time-aware structured state space duality model. *Neurocomputing*, pp. 131270, 2025.

Wei Fang, Zhaofei Yu, Yanqi Chen, Timothée Masquelier, Tiejun Huang, and Yonghong Tian. Incorporating learnable membrane time constant to enhance learning of spiking neural networks. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2641–2651. IEEE, 2021.

Chengjie Ge, Yufeng Peng, Zihao Li, Qiyu Kang, Xueyang Fu, Xuhao Li, Qixin Zhang, Ren Junhao, and Zheng-Jun Zha. Fractional-order spiking neural network. In *International Conference on Learning Representations*, volume 2026, pp. 79651–79696, 2026.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. 2023.

Peter Helfer, Corinne Teeter, Aaron Hill, Craig M Vineyard, James B Aimone, and Dhireesha Kudithipudi. Context modulation enables multi-tasking and resource efficiency in liquid state machines. In *Proceedings of the 2023 International Conference on Neuromorphic Systems*, pp. 1–9, 2023.

Mark Horowitz. 1.1 computing's energy problem (and what we can do about it). In *2014 IEEE international solid-state circuits conference digest of technical papers (ISSCC)*, pp. 10–14. Ieee, 2014.

Wang-Cheng Kang and Julian McAuley. Self-attentive sequential recommendation. In *2018 IEEE international conference on data mining (ICDM)*, pp. 197–206. IEEE, 2018.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. Technical report, 2009.

Jiacheng Li, Yujie Wang, and Julian McAuley. Time interval aware self-attention for sequential recommendation. In *Proceedings of the 13th international conference on web search and data mining*, pp. 322–330, 2020.

Nian Li, Xin Ban, Cheng Ling, Chen Gao, Lantao Hu, Peng Jiang, Kun Gai, Yong Li, and Qingmin Liao. Modeling user fatigue for sequential recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 996–1005, 2024.

Sarah Luca, T Patrick Xiao, Frances S Chance, Sapan Agarwal, Corinne Teeter, and G William Chapman. Learned adaptive properties for mitigation of weight perturbations in embedded spiking networks. *Frontiers in Neuroscience*, 20:1766765, 2026.

Changze Lv, Dongqi Han, Yansen Wang, Xiaoqing Zheng, Xuanjing Huang, and Dongsheng Li. Advancing spiking neural networks for sequential modeling with central pattern generators. volume 37, pp. 26915–26940, 2024a.

Changze Lv, Yansen Wang, Dongqi Han, Xiaoqing Zheng, Xuanjing Huang, and Dongsheng Li. Efficient and effective time-series forecasting with spiking neural networks. 2024b.

Krishna P Neupane, Ervine Zheng, and Qi Yu. Evidential stochastic differential equations for time-aware sequential recommendation. volume 37, pp. 67556–67574, 2024.

Jing Pei, Lei Deng, Sen Song, Mingguo Zhao, Youhui Zhang, Shuang Wu, Guanrui Wang, Zhe Zou, Zhenzhi Wu, Wei He, et al. Towards artificial general intelligence with hybrid tianjic chip architecture. *Nature*, 572(7767):106–111, 2019.

Silvan C Quax, Michele D'asaro, and Marcel AJ Van Gerven. Adaptive time scales in recurrent neural networks. *Scientific reports*, 10(1):11360, 2020.

Stefan Scholze, Johannes Partzsch, Sebastian Höppner, Florian Kelber, Andreas Dixius, Marco Stolba, Sirine Arfa, Marc Berthel, Georg Ellguth, Jim Garside, et al. The spinnaker2 chip: a many-core platform for flexible and scalable brain-inspired computing. *IEEE Open Journal of Circuits and Systems*, 2026.

Guanrui Wang, Songchen Ma, Yujie Wu, Jing Pei, Rong Zhao, and Luping Shi. End-to-end implementation of various hybrid neural networks on a cross-paradigm neuromorphic chip. *Frontiers in neuroscience*, 15:615279, 2021.

Maolin Wang, Sheng Zhang, Ruocheng Guo, Wanyu Wang, Xuetao Wei, Zitao Liu, Hongzhi Yin, Yi Chang, and Xiangyu Zhao. Star-rec: Making peace with length variance and pattern diversity in sequential recommendation. In *Proceedings of the 48th international ACM SIGIR conference on research and development in information retrieval*, pp. 1530–1540, 2025.

Wei Xiao, Huiying Wang, Qifeng Zhou, and Qing Wang. Ss4rec: Continuous-time sequential recommendation with state space models. *arXiv preprint arXiv:2502.08132*, 2025.

Zijie Xu, Tong Bu, Zecheng Hao, Jianhao Ding, and Zhaofei Yu. Proxy target: Bridging the gap between discrete spiking neural networks and continuous control. volume 38, pp. 159158–159184, 2026a.

Zijie Xu, Zihan Huang, Yiting Dong, Kang Chen, Wenxuan Liu, and Zhaofei Yu. Error amplification limits ann-to-snn conversion in continuous control. 2026b.

Zijie Xu, Xinyu Shi, Yiting Dong, Zihan Huang, and Zhaofei Yu. Care-bn: Precise moving statistics for stabilizing spiking neural networks in reinforcement learning. In *International Conference on Learning Representations*, volume 2026, pp. 50749–50772, 2026c.

Man Yao, Jiakui Hu, Zhaokun Zhou, Li Yuan, Yonghong Tian, Bo Xu, and Guoqi Li. Spike-driven transformer. volume 36, pp. 64043–64058, 2023.

Xingting Yao, Fanrong Li, Zitao Mo, and Jian Cheng. Glif: A unified gated leaky integrate-and-fire neuron for spiking neural networks. volume 35, pp. 32160–32171, 2022.

Jiqing Zhang, Malu Zhang, Yuanchen Wang, Qianhui Liu, Baocai Yin, Haizhou Li, and Xin Yang. Spiking neural networks with adaptive membrane time constant for event-based tracking. *IEEE Transactions on Image Processing*, 34:1009–1021, 2025a.

Sheng Zhang, Maolin Wang, Wanyu Wang, Jingtong Gao, Xiangyu Zhao, Yu Yang, Xuetao Wei, Zitao Liu, and Tong Xu. Glint-ru: Gated lightweight intelligent recurrent units for sequential recommender systems. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pp. 1948–1959, 2025b.

Shimin Zhang, Qu Yang, Chenxiang Ma, Jibin Wu, Haizhou Li, and Kay Chen Tan. Tc-lif: A two-compartment spiking neuron model for long-term sequential modelling. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 16838–16847, 2024.

Zhen Zhang, Jie Guo, Nan An, JunXiang Wu, Mingjin Zhang, Shuiying Xiang, and Yunsong Li. Tf-snn: Temporal focus-based dynamic neuron regulation framework for spiking neural networks. *Expert Systems with Applications*, pp. 132981, 2026.

Hanle Zheng, Zhong Zheng, Rui Hu, Bo Xiao, Yujie Wu, Fangwen Yu, Xue Liu, Guoqi Li, and Lei Deng. Temporal dendritic heterogeneity incorporated with spiking neural networks for learning multi-timescale dynamics. *Nature Communications*, 15(1):277, 2024.

Naichuan Zheng, Yuchen Du, Hailun Xia, and Zeyu Liang. Signal-sgn: A spiking graph convolutional network for skeleton action recognition via learning temporal-frequency dynamics. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 1539–1548, 2025a.

Naichuan Zheng, Hailun Xia, Zeyu Liang, and Yuchen Du. Mk-sgn: A spiking graph convolutional network with multimodal fusion and knowledge distillation for skeleton-based action recognition. *Neurocomputing*, pp. 131796, 2025b.

Naichuan Zheng, Hailun Xia, Zepeng Sun, Weiyi Li, and Yinzhe Zhou. Towards green wearable computing: A physics-aware spiking neural network for energy-efficient imu-based human activity recognition. *arXiv preprint arXiv:2604.10458*, 2026.

A Recurrent-State Semantics and Assignment Geometry

A.1 Temporal Axes and Recurrent-State Semantics

The recurrent state advances along the behavioral-event index ℓ, whereas τ is confined to local encoding. Tables A1 and A2 summarize the model/training constants and recurrent-state structure used in the reported configuration. At $H = 64$, PCD core and DGN-CM each contain 16,496 recurrent cell trainable parameters. Storage is arithmetic weight storage derived from this parameter count in 32-bit floating point (FP32) and 16-bit floating point (FP16), not measured device memory.

For a contextualized event feature $\mathbf{h}_\ell$, the local spike encoder uses

$$\mathbf{j}_\ell^{\text{in}} = \alpha_e(W_e\mathbf{h}_\ell + \mathbf{b}_e), \qquad \alpha_e = 4.0,$$

and repeats this projected current over T_s local LIF steps before averaging the emitted binary spikes to form $\mathbf{z}_\ell$ as defined in Section 3.1. This local encoder follows standard charge–fire–reset LIF dynamics and is reset before each forward call. For $T_s = 1$, the encoder executes one LIF update on a tensor of shape $[B, L, H]$; for $T_s > 1$, the projected event current is repeated along the local neural axis and the resulting spikes are averaged. Encoder state is reset before each forward call, so T_s controls only the local simulation window and does not create recurrence across behavioral events.

Table A1: PCD and Video settings.

Setting	Value
Hidden width / channel groups	$H = 64$ / $G = 8$
Inner neural horizon	$T_s = 2$
Encoder / recurrent current scale	$\alpha_e = 4.0$ / $\kappa = 2.0$
Retention predictor hidden width	50
Membrane retention interval	$[0.05, 0.98]$
Fast / slow activity retention intervals	$[0.05, 0.85]$ / $[0.75, 0.995]$
Video optimization	200 epochs; batch 4096; Adam; learning rate 10^{-3}
Firing regularizer	weight 0.001; target 0.15

Table A2: PCD recurrent-state structure.

Quantity	Reported value	Structural implication
State carried across events	$\{\mathbf{u}, \mathbf{s}, \mathbf{a}^f, \mathbf{a}^s\}$	four H-channel state vectors indexed by event
Retention groups	$G = 8$	eight channels per retention group
Retention context	$3G + 1 = 25$	grouped input conditioned on the current event
Dynamic retention outputs	$3G = 24$	group-level coefficients per event
Recurrent-cell parameters	16,496	64.4 KiB FP32 / 32.2 KiB FP16 weights

The recorded elapsed gap is computed in each dataset's timestamp units and enters the model as a numerical covariate, not as a physical membrane time constant. For a valid event,

$$\delta_\ell = \begin{cases} \xi_\ell - \xi_{\ell-1}, & \ell \geq 2 \text{ and } \xi_\ell - \xi_{\ell-1} > 0, \\ 1, & \text{otherwise}, \end{cases}$$

while padding is masked. If timestamp lists are unavailable, every valid event receives $\delta_\ell = 1$. When elapsed time is disabled, the same unit gap is used at every valid event, so the temporal predictor input remains fixed at $\log 2$ without changing predictor dimensionality. The control that shuffles time gaps across the batch permutes complete gap rows across users, whereas the within-user control preserves each user's gap multiset and permutes only its association with event position.

A.2 Grouped Retention and Local Dynamics

Grouped retention parameterization. For H divisible by G, let $q = H/G$ be the number of channels per group and let $\mathcal{G}_g$ denote the gth group. The operators used in Eq. (6) are group averaging and broadcast,

$$[\mathcal{P}_G(\mathbf{x})]_g = \frac{1}{q} \sum_{h \in \mathcal{G}_g} x_h, \qquad [\mathcal{E}_G(\mathbf{v})]_h = v_g \quad \text{for } h \in \mathcal{G}_g.$$

With $H = 64$ and $G = 8$, each group contains eight channels. The operators satisfy $\mathcal{P}_G(\mathcal{E}_G(\mathbf{v})) = \mathbf{v}$ for $\mathbf{v} \in \mathbb{R}^G$, whereas $\mathcal{E}_G(\mathcal{P}_G(\mathbf{x}))$ is the blockwise-constant projection of $\mathbf{x}$ onto the grouped subspace. The two retention maps in Section 3.3 are two-layer multilayer perceptrons with Sigmoid Linear Unit (SiLU) activation and hidden width 50,

$$g_r(\mathbf{c}) = W_{r,2}\,\mathrm{SiLU}(W_{r,1}\mathbf{c} + \mathbf{b}_{r,1}) + \mathbf{b}_{r,2},$$
$$g_a(\mathbf{c}) = W_{a,2}\,\mathrm{SiLU}(W_{a,1}\mathbf{c} + \mathbf{b}_{a,1}) + \mathbf{b}_{a,2}.$$

The retention intervals are listed in Table A1. The overlap between the two activity intervals leaves their ordering at each event unconstrained.

Local feature derivatives. For the local task feature sensitivity, define $\mathbf{q}_\ell = \mathcal{P}_G(\mathbf{z}_\ell)$ and let $p \in \{f, s\}$. Holding $\mathbf{I}_\ell$ and the prior recurrent state fixed, the direct derivatives are

$$\frac{\partial \widetilde{\mathbf{u}}_\ell}{\partial \mathbf{q}_\ell} = \mathrm{Diag}(\mathbf{u}_{\ell-1} - \mathbf{I}_\ell)\frac{\partial \mathbf{r}_\ell}{\partial \mathbf{q}_\ell}, \qquad \frac{\partial \mathbf{a}_\ell^p}{\partial \mathbf{q}_\ell} = \mathrm{Diag}(\mathbf{a}_{\ell-1}^p - \mathbf{s}_{\ell-1})\frac{\partial \boldsymbol{\rho}_\ell^p}{\partial \mathbf{q}_\ell}. \tag{13}$$

These partial derivatives are mediated by the retention predictor with $\mathbf{I}_\ell$ held fixed. Combining $\partial \mathbf{I}_\ell / \partial \mathbf{z}_\ell = \kappa W_z$ with the membrane derivative in Eq. (13) yields the two-path decomposition

in Eq. (9). The direct-current and retention-mediated terms can both propagate through the later recurrent trajectory.

Threshold and auxiliary item path. Positive threshold and feedback parameters are represented through unconstrained raw variables,

$$\boldsymbol{\theta}_0 = \mathrm{softplus}(\widehat{\boldsymbol{\theta}}_0), \quad \boldsymbol{\beta}_f = \mathrm{softplus}(\widehat{\boldsymbol{\beta}}_f), \quad \boldsymbol{\beta}_s = \mathrm{softplus}(\widehat{\boldsymbol{\beta}}_s), \quad \boldsymbol{\gamma} = \mathrm{softplus}(\widehat{\boldsymbol{\gamma}}). \tag{14}$$

Their configured initial values are 0.35, 0.50, 0.20, and approximately zero, respectively. The full model auxiliary item offset is

$$\epsilon(i) = \alpha_i\big(\mathrm{softplus}(e_i) - \log 2\big), \qquad \alpha_i = 0.05, \tag{15}$$

where e_i is a learned scalar embedding initialized to zero. Centering by $\log 2$ makes this path neutral at initialization. When this auxiliary full model path is enabled, it enters the threshold as

$$\boldsymbol{\theta}_\ell^{\mathrm{full}} = \boldsymbol{\theta}_0 + \boldsymbol{\beta}_f \odot \mathbf{a}_\ell^f + \boldsymbol{\beta}_s \odot \mathbf{a}_\ell^s + \epsilon(i_\ell)\mathbf{1}.$$

The default PCD core comparisons disable item excitability, so $\epsilon(i_\ell) = 0$ there. Popularity-based excitability is also disabled in the reported configuration.

Spike surrogate, reset, and recurrent readout. The PCD threshold uses a hard forward spike and a triangular surrogate only in backpropagation. For scalar threshold argument $x = \widetilde{u} - \theta$ and support parameter $\lambda_s = 1$,

$$\Theta(x) = \mathbb{I}[x \geq 0], \qquad D^{\mathrm{sg}}\Theta(x) := \frac{1}{\lambda_s}\max\left(1 - \frac{|x|}{\lambda_s}, 0\right). \tag{16}$$

Here $D^{\mathrm{sg}}\Theta$ denotes the surrogate backward operator, not the classical derivative of the hard threshold. In the PCD recurrent implementation, the same surrogate spike value remains in the multiplicative reset path $\mathbf{u}_\ell = \widetilde{\mathbf{u}}_\ell \odot (\mathbf{1} - \mathbf{s}_\ell)$ during backpropagation. The reset mask is therefore not detached. Surrounding SpikingJelly modules instead use the arctangent surrogate with scale 2.0 and detached reset. All recurrent states and the readout trace are initialized to zero at the start of a forward pass. The recurrent readout uses a separate trace that is excluded from the retention context in Eq. (6),

$$\mathbf{h}_\ell^{\mathrm{tr}} = \lambda \mathbf{h}_{\ell-1}^{\mathrm{tr}} + (1-\lambda)\mathbf{s}_\ell, \qquad \mathbf{y}_\ell^{\mathrm{rec}} = W_o\,\mathrm{Dropout}(\mathbf{h}_\ell^{\mathrm{tr}}) + \mathbf{b}_o, \tag{17}$$

with $\lambda = 0.85$. The resulting recurrent readout is gated by spikes and temporally convolved before fusion with the unchanged linear attention branch and downstream scoring path.

Retention and local simulation horizons. For a frozen coefficient sequence, define the event-domain coefficient products

$$\mathbf{R}^{u}_{\ell\to\ell+k} = \bigodot_{j=\ell+1}^{\ell+k} \mathbf{r}_j, \qquad \mathbf{R}^{a,p}_{\ell\to\ell+k} = \bigodot_{j=\ell+1}^{\ell+k} \boldsymbol{\rho}_j^p, \quad p \in \{f, s\}. \tag{18}$$

For a locally constant scalar coefficient $\eta \in (0,1)$, the homogeneous event-domain e-folding horizon is $-1/\log\eta$. Applying this diagnostic to the configured intervals gives approximately 0.33–49.5 events for membrane retention, 0.33–6.15 events for the fast activity trace, and 3.48–199.5 events for the slow activity trace. For the activity traces, $\mathbf{R}^{a,p}$ gives exact homogeneous coefficient carryover. For membrane state, $\mathbf{R}^{u}$ is the pre-reset retention factor, while realized post-reset propagation also depends on spike/reset masks and the resulting nonlinear trajectory.

Video models are selected by validation NDCG@10. The training objective is cross-entropy plus a binary-PCD firing-rate regularizer with weight 0.001 and target activity 0.15 (Table A1). This auxiliary penalty is separate from the recurrent state dynamics. The three NDCG@10 values in Table A3 are tightly clustered across the tested local simulation windows, while reported fitting memory increases with T_s. The reported model uses $T_s = 2$.

A.3 Assignment Geometry and Surrogate Sensitivity

We distinguish forward reassignment identities, identities under the adopted surrogate backward graph, and first-order approximations at one event. Frozen-sequence replay uses the forward identities, while the first-order terms characterize local reassignment around the reference trajectory.

Table A3: Inner-horizon check.

T_s	**NDCG@10**	**Fitting memory (GB)**
1	0.0860	2.183
2	0.0861	2.300
4	0.0862	2.535

Proposition A.1 (Exact reassignment geometry at one event). Let $r^G_{\ell,g}$ denote the scalar membrane retention coefficient broadcast to group $\mathcal{G}_g$, and let

$$\mathbf{d}_{\ell,g} = (\mathbf{u}_{\ell-1} - \mathbf{I}_\ell)_{\mathcal{G}_g}.$$

For a permutation π of the G groups, Eq. (3) yields the exact group displacement at one event

$$\Delta_\pi \widetilde{\mathbf{u}}_{\ell,\mathcal{G}_g} = (r^G_{\ell,\pi(g)} - r^G_{\ell,g})\mathbf{d}_{\ell,g}.$$

Summing its squared norm over disjoint group supports gives Eq. (4).

Proposition A.2 (Expected displacement under uniform reassignment). Let $\bar{r}^G_\ell = G^{-1}\sum_g r^G_{\ell,g}$ and $\sigma^2_{r,\ell} = G^{-1}\sum_g (r^G_{\ell,g} - \bar{r}^G_\ell)^2$. If π is uniformly distributed over all group permutations, then $r^G_{\ell,\pi(g)}$ is uniform over the group coefficients for every fixed g. Expanding the squared difference in Eq. (4) and taking this expectation yields Eq. (5).

Corollary A.1 (Null assignment conditions). Eq. (5) vanishes if all group coefficients are equal or if $\mathbf{u}_{\ell-1} = \mathbf{I}_\ell$ in every group. Hence nonzero expected displacement requires coefficient heterogeneity together with a nonzero state discrepancy. The statement concerns the pre-reset membrane at one event. Spike/reset changes can alter the subsequent recurrent trajectory.

Proposition A.3 (Surrogate coefficient sensitivity). We next connect the forward geometry to a scalar loss $\mathcal{J}$ minimized by training. Under the triangular/arctangent surrogate rules specified in Appendix A.1, let $\mathbf{v}^{\mathrm{sg}}_{\ell,g} = \nabla^{\mathrm{sg}}_{\widetilde{\mathbf{u}}_{\ell,\mathcal{G}_g}} \mathcal{J}$. Since $r^G_{\ell,g}$ enters the pre-reset membrane as $r^G_{\ell,g}\mathbf{d}_{\ell,g}$, the surrogate coefficient sensitivity is exactly the first line of Eq. (8). This is a chain-rule identity for the specified surrogate backward graph; the hard-threshold forward map remains nondifferentiable in the classical sense.

Remark A.1 (Local assignment ordering). For a permutation applied at a single event ℓ, define the coefficient perturbation $\delta r^G_{\ell,g} = r^G_{\ell,\pi(g)} - r^G_{\ell,g}$. A first-order surrogate expansion at the reference trajectory gives

$$\Delta\mathcal{J}^{(1)}_{\ell,\pi} = \sum_{g=1}^{G} \chi_{\ell,g}\big(r^G_{\ell,\pi(g)} - r^G_{\ell,g}\big). \tag{19}$$

If the same permutation is applied at multiple events, the sequence-level first-order term is the sum of these eventwise terms; the finite replay remains the experimentally measured quantity.

For a uniformly random group permutation, $\mathbb{E}_\pi[r^G_{\ell,\pi(g)}] = \bar{r}^G_\ell$ for every fixed g, so

$$\mathbb{E}_\pi\left[\Delta\mathcal{J}^{(1)}_{\ell,\pi}\right] = -G\ \mathrm{Cov}_g\big(r^G_{\ell,g}, \chi_{\ell,g}\big)\,. \tag{20}$$

The covariance is a local ordering diagnostic around the reference trajectory; finite nonlinear behavior is measured by replay. For a transposition of groups i and j,

$$\Delta\mathcal{J}^{(1)}_{\ell,i\leftrightarrow j} = (r^G_{\ell,j} - r^G_{\ell,i})(\chi_{\ell,i} - \chi_{\ell,j}). \tag{21}$$

If $r^G_{\ell,i} > r^G_{\ell,j}$, the original assignment is locally no worse than the swapped assignment when $\chi_{\ell,i} \le \chi_{\ell,j}$. The ordering is defined by the signed sensitivity χ, not by its magnitude alone.

Retention-mediated feature route. If $\mathbf{q}_\ell = \mathcal{P}_G(\mathbf{z}_\ell)$ supplies the task feature to the retention predictor, the membrane branch contributes the retention-mediated route in Eq. (8), while Eq. (9) separates that route from the direct recurrent-current path. The activity retention branches have analogous chain-rule terms. Detaching $\mathbf{q}_\ell$ preserves its forward values and leaves the predictor parameters trainable while removing the gradient route from retention to the task representation; feature reassignment instead changes the source supplied to that route.

B Recommendation Protocols and Assignment Evidence

B.1 Recommendation Data and Evaluation Protocols

For the compact sequence-model context in Table B3, GLINT-RU denotes Gated Lightweight Intelligent Recurrent Units, and PCD-LIF denotes SpikeGLINTRec with PCD recurrent dynamics and the LIF event encoder. Full PCD on Video is reported in Table B2; all values remain protocol specific.

For every recommendation intervention, the independently trained model or paired seed is the statistical unit. Repeated deterministic reassignments are averaged within that unit before uncertainty is computed across units. Unless otherwise stated, paired 95% confidence intervals use two-sided Student-t intervals over the model- or seed-wise differences after within-unit averaging. Mean $\pm$ standard deviation summarizes independent runs. Metrics are interpreted only within a common dataset and ranking protocol. Beauty and ML-1M use the RecBole 1.2.1 protocol used for the GLINT-RU comparison, whereas Video follows the SSR dataset statistics and full-ranking split structure. Published SSR results provide protocol-specific prior-work context.

Hyperparameters of the recurrent state are selected on validation NDCG@10 and then held fixed for the reported test comparisons. DGN-CM uses adaptive neuron dynamics in which membrane persistence is generated from neuronal variables and the current task feature block is excluded from the retention predictor. The outer recommendation architecture is unchanged, and its recurrent-cell parameter count is the same as PCD. The DGN-CM search over $r_0 \in \{0.50, 0.70, 0.85, 0.95\}$ selects $r_0 = 0.95$, which is used for Video and transferred unchanged to Beauty.

B.2 Task-Feature Source and Realized Retention

The reassignment study after training evaluates 20 independently trained PCD core models. The retention context is $[\mathcal{P}_G(\mathbf{z}_\ell), \mathcal{P}_G(\mathbf{s}_{\ell-1}), \mathcal{P}_G(\bar{\mathbf{a}}_{\ell-1}), \log(1+\delta_\ell)]$. Reassignment changes the current task feature block $\mathcal{P}_G(\mathbf{z}_\ell)$ before either retention branch is evaluated, while recurrent spike history, adaptation history, recorded spacing, the main recommendation feature path, and all parameters remain fixed. Three deterministic cyclic assignments are averaged within model before paired uncertainty is computed across models.

All three cyclic assignments produce similar NDCG@10 changes. Zeroing the task feature block produces a larger but more heterogeneous change than feature reassignment. Five trained PCD core models are evaluated under aligned, reassigned, and zero task features. Predictor outputs are transformed through their runtime ranges before comparison. Retention displacement is the paired event/time/group mean absolute difference, with global mean shift reported separately.

Task feature reassignment changes all three retention branches in all five models. For membrane retention, the 95% CI for the global mean shift includes zero, whereas the fast and slow activity branches show positive mean shifts. Zero context produces a larger displacement than reassignment in every branch.

Branch-specific analysis shows a detectable ranking change for membrane retention. Reassigning membrane retention by itself or all three branches lowers NDCG@10 across the five analyzed models, while the intervals for the fast and slow activity branches span zero and the paired all-branch-versus-membrane-only interval includes zero under fixed coefficients. Subsequent reassignment experiments use membrane retention across recurrent groups.

Table B1: Recommendation datasets and protocols.

Dataset	Scale			Sequence		Evaluation
	Interactions	Users	Items	Avg. hist.	Max. L	
Beauty	198,502	22,363	12,101	8.88	100	100-popular-item sampled leave-one-out
Video	496,904	55,144	17,286	9.01	20	full ranking; 72% train, 8% validation, 20% test
ML-1M	1,000,209	6,040	3,706	165.60	200	100-popular-item sampled leave-one-out

Table B2: Recommendation ranking performance.

Dataset	Method	Runs	Ranking metrics		
			Recall@10	MRR@10	NDCG@10
Video	CE-LIF-style, $r = 0.50$	5	0.13688 ± 0.00254	0.06976 ± 0.00086	0.08544 ± 0.00126
	CE-LIF-style, $r = 0.95$	5	0.13652 ± 0.00186	0.06994 ± 0.00082	0.08548 ± 0.00105
	DGN-CM, $r_0 = 0.95$	20	0.13770 ± 0.00100	0.06930 ± 0.00073	0.08530 ± 0.00065
	PCD core	20	0.13832 ± 0.00138	0.07023 ± 0.00076	0.08612 ± 0.00080
	Full PCD	5	0.13866 ± 0.00102	0.07004 ± 0.00106	0.08604 ± 0.00067
Beauty	DGN-CM, $r_0 = 0.95$	3	0.39517 ± 0.00115	0.20583 ± 0.00076	0.25040 ± 0.00070
	PCD core	3	0.39857 ± 0.00354	0.20620 ± 0.00110	0.25150 ± 0.00157

Table B3: Sequence-model NDCG@10 by dataset.

Dataset	Model	Runs	NDCG@10
Video	SASRec	3	0.08687 ± 0.00045
Video	GLINT-RU-style dense model	5	0.08610 ± 0.00062
ML-1M	SASRec	3	0.48060 ± 0.00157
ML-1M	LightSANs	3	0.47873 ± 0.00169
ML-1M	SpikeGLINTRec (PCD-LIF)	3	0.46920 ± 0.00197

Table B4: Retention changes under task feature reassignment.

Retention branch	Feature-reassignment response	Global-mean shift (95% CI)
Membrane	displaced in all 5 models	$[-0.00760,\ 0.00659]$
Fast activity	displaced in all 5 models	$[0.02477,\ 0.07272]$
Slow activity	displaced in all 5 models	$[0.00089,\ 0.00888]$

B.3 Persistence Assignment with Fixed Coefficients

For each model, the unperturbed membrane and activity retention sequences are recorded and then reused independently of the perturbed recurrent trajectory. Four nonzero cyclic reassignments are applied to the membrane groups and averaged within model before inference. At one event, reassignment acts according to Eq. (3) on the state available through $\ell - 1$, changing which state discrepancies receive the captured coefficients while leaving the coefficient multiset and mean invariant. The identical replay and group permutations are applied to the DGN-CM models with the same recurrent-cell parameter count. After averaging permutations within model, the direct paired interaction is $I_s = \Delta^{\text{assign}}_{\text{PCD},s} - \Delta^{\text{assign}}_{\text{DGN},s}$ over the 20 model pairs.

The unpermuted assignment reproduces the original metrics and coefficient traces, and every nonzero reassignment preserves the sorted coefficient multiset. On Video, PCD has direct mean responses of $+0.00079621$, $+0.00055600$, and $+0.00159003$ for NDCG@10, MRR@10, and Recall@10, whereas the corresponding DGN-CM means remain near zero. For NDCG@10, the PCD 95% CI is $[+0.00055539, +0.00103703]$ and the DGN-CM interval is $[-0.00002619, +0.00010026]$. All three paired interactions favor the larger PCD response under the same recurrent-cell parameter count.

Replication on ML-1M. We repeat reassignment of the fixed coefficient sequence on ten paired seeds under the ML-1M protocol in Table B1, averaging offsets $+1, +2, +3, +4$ within seed before inference. In Table B5, interaction is PCD minus DGN-CM; confidence intervals and W/T/L counts use the paired model units on Video and paired seed units on ML-1M.

Appendix B.4 analyzes the role of the task feature within PCD.

Unpermuted replay reproduces the stored metrics and coefficient traces. Section 4 reports the replay-integrity tolerance for both recommendation datasets. On ML-1M, the 95% CIs for the NDCG@10 and MRR@10 interactions lie above zero, whereas the interval for Recall@10 includes zero. The NDCG direction is consistent across offsets and seeds.

B.4 Training Dependence of Persistence Assignment

Five runs compare aligned intrinsic conditioning, intrinsic conditioning trained with reassigned task features, and conditioning applied to the representation with the same predictor parameter count. Task metrics use aligned context at evaluation. Each trained model is then evaluated with aligned, reassigned, and zero task features, with repeated reassignments averaged within model before paired inference.

Ranking performance remains similar across the three conditions, whereas the NDCG change under task feature reassignment differs. Intrinsic conditioning shows the largest change, conditioning applied to the representation the smallest, and training with reassigned features reduces the displacement of membrane retention across groups. The differences for the fast and slow activity branches remain within uncertainty.

The two bijective near/far donor assignments produce different cosine similarities, but the confidence interval for their difference in ranking change includes zero. A non-bijective nearest/farthest donor assignment produces a graded ranking change but also changes donor coverage and multiplicity. Across the tested bijective assignments, ranking change is not monotonic in embedding distance.

C Controlled Vision Analysis of Persistence Assignment

C.1 Vision Protocol and Persistence Inputs

The CIFAR-10 study uses the official 50,000-image training set with a fixed 5,000-image validation subset and the untouched 10,000-image test set. All models share the same spiking ResNet-19 architecture, four neural steps, data augmentation, optimizer schedule, and model selection on the validation set. Eight variants are evaluated over five independent runs. Table 3 reports the information available to persistence in each variant. DGN-CM and all variants with a retention predictor contain 4,358,490 total parameters and 25,232 predictor parameters, whereas PLIF contains 4,330,506 parameters. The factorial training comparison crosses task-feature matching with gradient access, while aligned task context is used for the common comparison of task accuracy. Feature reassignment after training is reported separately in Appendix C.3.

Source-offset robustness and coefficient ranges associated with classes are summarized in Figure 3. Appendices C.2 and C.3 give the corresponding reassignment construction and robustness checks.

C.2 Feature Matching and Gradient Access

Figure 2 summarizes original-feature accuracy, same-to-cross gaps, and source-indexed retention correlation; Figure 3(a) gives source-offset robustness. Table C1 provides the original-feature operating points for the four training conditions. The donor construction is specified below. Evaluations after training reuse each model selected on validation data and replace the exogenous feature block of the retention predictor while the image and prediction pathway remain

Table C1: CIFAR-10 aligned operating points over five runs (mean ± standard deviation).

Variant	Aligned Top-1 (%)
TCD	90.870 ± 0.246
TCD-Detach	90.402 ± 0.239
TCD-Shuffle	91.020 ± 0.284
TCD-Shuffle-Detach	90.924 ± 0.193

Table B5: Reassignment results on Video and ML-1M.

(a) Video

Metric	PCD Δ_{alloc}	DGN-CM Δ_{alloc}	Interaction [95% CI]	W/T/L
NDCG@10	$+0.00079621$	$+0.00003703$	$+0.00075918$ [$+0.00049065$, $+0.00102770$]	$17/0/3$
MRR@10	$+0.00055600$	$+0.00004421$	$+0.00051179$ [$+0.00030541$, $+0.00071816$]	$16/0/4$
Recall@10	$+0.00159003$	$+0.00001815$	$+0.00157188$ [$+0.00106724$, $+0.00207652$]	$18/0/2$

(b) ML-1M

Metric	PCD Δ_{alloc}	DGN-CM Δ_{alloc}	Interaction [95% CI]	W/T/L
NDCG@10	$+0.015604$	$+0.006243$	$+0.009362$ [$+0.003742$, $+0.014982$]	$9/0/1$
MRR@10	$+0.016400$	$+0.005965$	$+0.010435$ [$+0.004684$, $+0.016185$]	$9/0/1$
Recall@10	$+0.012397$	$+0.006970$	$+0.005426$ [-0.000123, $+0.010975$]	$7/0/3$

Table B6: Paired Video controls.

Paired contrast	Difference	95% CI
Intrinsic vs. representation, ΔNDCG@10	$+0.000147$	[0.000052, 0.000243]
Aligned vs. training with reassigned features, membrane displacement	0.01677	[0.01014, 0.02339]
Near vs. far donors, cosine-similarity difference	0.06963	[0.05618, 0.08308]

fixed. Original-feature evaluation uses each sample's own pooled feature context and produces the task-accuracy values in Table C1. Same-class reassignment uses a feature from another image with the same CIFAR-10 label. Cross-class reassignment uses a source class offset of $+1, +3, +7$, or $+9$ modulo ten. Within each source-class offset, three partner realizations randomize the particular source image. The label selects the donor class for reassignment and is not provided to the model. Repeated evaluations of one trained model are averaged within model before uncertainty across five runs is computed. Restoring each sample's own context under the same evaluation order blocked by class reproduces the original-feature top-1 and top-5 results. With evaluation order held fixed, the same-to-cross accuracy gap changes with feature assignment. Across the four source-class offsets, the TCD same-to-cross gap remains between 40.79 pp and 42.51 pp.

C.3 Feature-Source Indexing and Reassignment Robustness

Under source-class reassignment, the global mean shift remains below 5×10^{-4} for both membrane and activity retention. Re-indexing the same samples by donor class changes the membrane-retention correlation from -0.361 to 0.947; the activity-retention correlation under source indexing is 0.793.

TCD has the largest class-associated membrane-retention range and same-to-cross accuracy gap; detachment reduces both, and variants trained with shuffled features drive the accuracy gap near zero. Figure 4 summarizes this descriptive relation across the four training conditions.

Random partner assignment removes the class-preserving donor constraint. Three pairing realizations are averaged within each trained model before the aligned-to-random comparison. TCD and TCD-Detach retain nonzero aligned-to-random accuracy gaps, while both intervals for variants trained with shuffled features include zero. Random pairing also compresses the retention ranges associated with classes. Zero context changes accuracy for both shuffled-feature variants.

The aligned-to-random confidence intervals are reported in Section 4.4. Additional random-partner and zero-context statistics are:

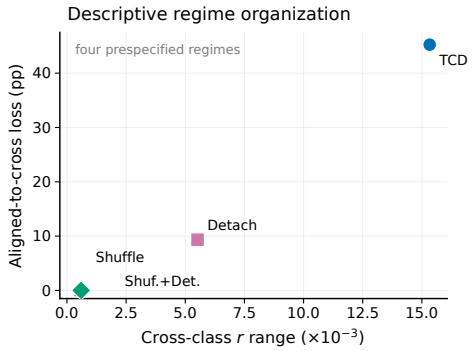

Figure 4: CIFAR-10 regime organization under feature-source and gradient controls.

Table C2: Random donor and zero-context controls.

Statistic	**Result**
TCD membrane range, class-preserving → random	$0.01399 \rightarrow 0.00092$
TCD activity range, class-preserving → random	$0.00451 \rightarrow 0.00035$
Zero-context loss, TCD-Shuffle / TCD-Shuffle-Detach	3.650 pp / 1.204 pp

D Recurrent Realizations and Computational Analysis

D.1 Alternative Recurrent Realizations

Table D1 compares alternative activity state, temporal input, auxiliary item, and static retention realizations on Video. The five-run comparisons use the default PCD core, while the static-retention study uses dynamic PCD with the same recurrent configuration over three runs.

Table D1: Alternative recurrent realizations on Video.

Control axis	**Realization**	**Runs**	**NDCG@10**
Default PCD core	Two activity states, feedback on; time in retention predictor; item offset off	5	0.08588 ± 0.00118
Activity state	Two-state, feedback off	5	0.08614 ± 0.00035
	Single-state, feedback on	5	0.08648 ± 0.00019
	Single-state, feedback off	5	0.08652 ± 0.00057
Temporal input	Unit gap	5	0.08568 ± 0.00033
	Representation-level time	5	0.08658 ± 0.00027
	Within-user shuffled gaps	5	0.08546 ± 0.00134
Auxiliary item path	Full PCD with item offset	5	0.08604 ± 0.00067
Static retention	Fixed $r = 0.70$	3	0.08583 ± 0.00051
	Dynamic full PCD	3	0.08613 ± 0.00035

Single-state variants match or slightly exceed the default two-state realization. Representation-level time gives NDCG@10 comparable to retention-predictor time, and the auxiliary item offset changes NDCG@10 little. Fixed and dynamic retention also produce similar NDCG@10 under the same recurrent configuration. Fixed-retention rows remove that prediction route, whereas single-state variants change state structure and predictor parameterization. These rows characterize task performance; assignment replay is confined to the primary recurrent configuration.

D.2 Model Scale and Arithmetic Assumptions

Table B3 consolidates the predictive metrics; the quantities below describe model scale and the arithmetic assumptions used for computational interpretation. The full PureSpike model in the event domain and the simplified Shift-PCD realization provide a secondary representation comparison. On Video, NDCG@10 increases from 0.06753 for PureSpike to 0.07480 for Shift-PCD, while the hybrid PCD model remains higher. SSR follows a different representation strategy by converting dense item representations into spike codes and filtering them in the spike domain (Di Yu et al., 2025); the present model retains dense item computation.

Dimensions of the recurrent state and the input/output sizes of the retention predictor are consolidated with the model constants in Table A2. Once the current event representation $\mathbf{z}_\ell$ is available, the recurrent update advances $\{\mathbf{u}_\ell, \mathbf{s}_\ell, \mathbf{a}_\ell^f, \mathbf{a}_\ell^s\}$ from their preceding values. Dense embedding/context, linear attention, fusion, and final scoring remain dense computations. Per event, the retention predictor receives 25 grouped context values and produces 24 retention coefficients. These dimensions describe the recurrent state update, not the full recommendation computation.

The $16{,}496$ recurrent parameters require approximately 64.4 KiB at FP32 or 32.2 KiB at FP16. For scale, SpiNNaker2 reports 128 KiB of local static random-access memory (SRAM) per processing element together with an ARM core, numerical accelerators, and event routing (Scholze et al., 2026). These storage figures cover recurrent weights only; deployment additionally requires code, recurrent state, buffers, routing metadata, and quantization. Tianjic provides a separate cross-paradigm example with unified communication for mixed Artificial Neural Network and Spiking Neural Network modules (Pei et al., 2019; Wang et al., 2021).

The efficiency analysis uses an arithmetic proxy for each operation under two execution assumptions. Following conventional 45-nm, 32-bit accounting (Horowitz, 2014; Yao et al., 2023; Di Yu et al., 2025), the synaptic-operation (SOP) count $\mathrm{SOP}_m = f_m T_m F_m$ serves as an arithmetic proxy for an event-eligible operator m, where F_m is its dense operation count per neural step, f_m is the measured firing activity, and T_m is the neural horizon. If every spike-eligible transform can skip inactive events, the observed activity implies approximately 57.05% fewer estimated encoder operations and 34.05% fewer estimated end-to-end operations. Under the current hybrid execution boundary, rate-coded and continuous transforms remain dense unless event-skippable, reducing the estimates to 4.33% and 2.69%. These arithmetic proxies exclude latency, memory/data movement, and hardware overheads.