CERTIFIED COORDINATE ADAPTATION FOR FINITE-CAPACITY LATENT DYNAMICS LEARNERS

Anonymous authors
Paper under double-blind review

ABSTRACT

A representation can retain sufficient information for prediction yet remain poorly matched to a finite-capacity downstream learner, limiting how effectively that information can be used. To address this mismatch, we introduce Profiled Certified Exact Gauge Adaptation (PCEGA), which learns an invertible coordinate map and refits a dynamics predictor while keeping the encoder and decoder frozen. Invertibility preserves encoded information, and exact inverse compensation preserves the original reconstruction. A computable global bi-Lipschitz bound controls coordinate distortion, while coordinate optimization uses decoded prediction error and accounts for predictor refitting and ridge selection. Theoretically, for any fixed integer $p \geq 1$, we prove that invertible affine maps are exactly the globally bi-Lipschitz transformations that preserve the full class of polynomial dynamics of degree at most p under conjugacy. Experimentally, the complete default global-budget procedure reduces normalized mean squared error by 48.82% and 45.11% on independently trained Lorenz and Rössler charts, respectively. On an analytic-chart benchmark covering four dynamical systems, validation-selected maps from a learned coordinate library reduce normalized mean absolute error by 15.39%–71.11% across linear, quadratic polynomial, random-feature, and radial basis function learners. Both comparisons use independently tuned baselines in the original coordinates. Crossed-learner experiments further reveal learner-specific coordinate preferences, supporting adaptation tailored to the downstream learners.

1 INTRODUCTION

Large-scale pretraining supports representation reuse in protein modeling (Lin et al., 2023), genomics (Nguyen et al., 2024), single-cell transcriptomics (Theodoris et al., 2023), and spatial transcriptomics (Tejada-Lapuerta et al., 2025). A practical strategy freezes the pretrained model and trains task-specific predictors on its embeddings, separating representation learning from downstream modeling. However, pretraining objectives need not align with a downstream learner's inductive bias. Even when an embedding retains task information, its coordinates may make that information difficult for a finite-capacity predictor to exploit. Can an existing embedding be adapted for prediction while preserving the information it contains?

We study this question in latent dynamics prediction, where a frozen encoder–decoder provides an explicit reconstruction criterion (Lusch et al., 2018; Champion et al., 2019; Linot & Graham, 2020). An encoder E maps x to $z = E(x)$, and a decoder D returns latent predictions to physical space. For each invertible map H, we refit $\widehat{g}_H \in \mathcal{G}$ to predict $H(E(x^+))$ from $H(E(x))$. Here $\mathcal{G}$ is the prescribed dynamics family and the subscript H records the learner's dependence on its fitting coordinates; only E and D remain frozen. The resulting paths are

$$\widehat{x}^+ = DH^{-1}\widehat{g}_H HE(x), \qquad DH^{-1}HE(x) = DE(x). \tag{1}$$

Although H and H^{-1} cancel along the reconstruction path, they need not cancel around a restricted dynamics learner. The effective prediction class becomes $H^{-1}\mathcal{G}H$, allowing predictive utility to change without discarding encoded information. Learning H adds supervised capacity to the complete learner, confined to an invertible coordinate transformation.

We implement this approach through *Profiled Certified Exact Gauge Adaptation* (PCEGA). Explicitly invertible couplings provide exact inverse compensation, while a computable global bi-Lipschitz

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.

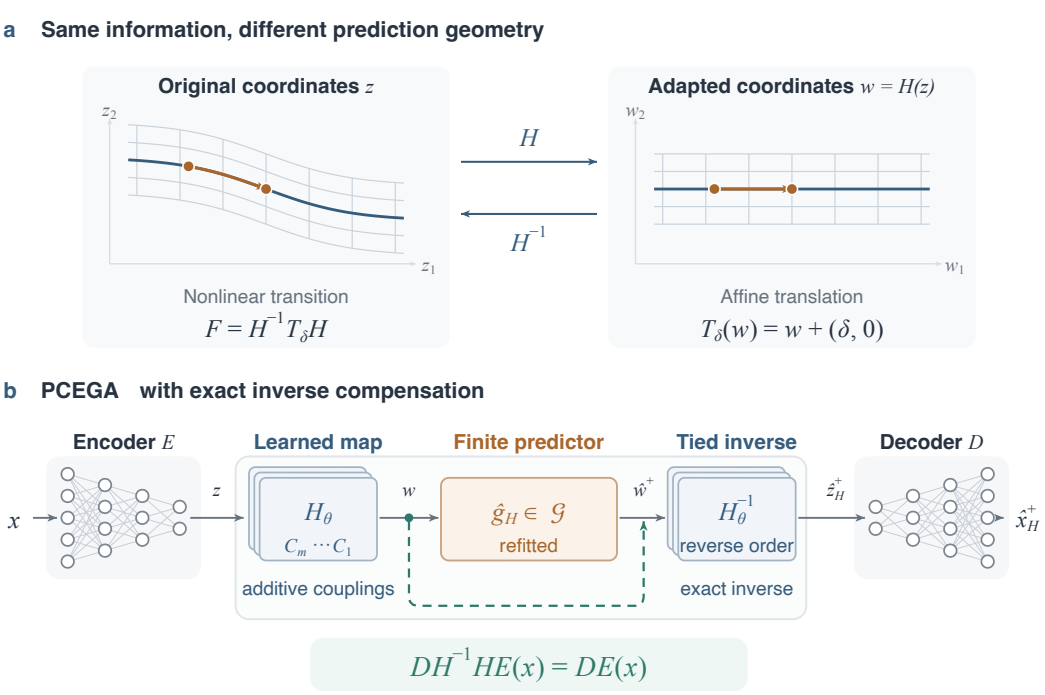

Figure 1: **PCEGA adapts predictive coordinates while preserving a frozen representation. (a)** An illustrative conjugacy $F = H^{-1} \circ T_\delta \circ H$ turns a nonlinear transition into $T_\delta(w) = w + (\delta, 0)$, a shift in the first coordinate. Translation is illustrative, not assumed for the experimental systems. **(b)** Frozen E, D surround a learned coupling stack H_θ, a predictor $\hat{g}_H$ refitted in its coordinates, and the tied inverse applied in reverse layer order. The green bypass gives $DH_\theta^{-1} H_\theta E(x) = DE(x)$, preserving the original reconstruction.

bound controls coordinate distortion. Coordinate search uses decoded prediction error and accounts for predictor fitting and ridge selection. An independent held-out selector retains identity when a candidate fails the acceptance rule. Learned coordinates can depend on both the task and the learner. We therefore compare each learner with and without its own adaptation, with independent tuning in both cases. Figure 1 illustrates the interface. Our contributions are threefold:

Information-preserving adaptation. We combine a frozen encoder–decoder interface, a certified distortion budget, and coordinate optimization through predictor refitting, followed by independent acceptance and evaluation.

An expressivity boundary. We prove that invertible affine maps are exactly the globally bi-Lipschitz transformations preserving a full polynomial dynamics class of fixed finite degree $p \geq 1$, characterizing when adaptation changes that class.

Empirical gains and learner dependence. We evaluate controlled systems, measured input–output systems, and frozen flow representations. Crossed-learner comparisons reveal coordinate preferences, while physical-loss refitting controls show that gains can persist after optimizing the learner parameters directly in physical space.

2 Related Work

Invertible representation adaptation. Diffeomorphic Learning and FineMorphs improve prediction through input deformations and affine–diffeomorphic compositions (Younes, 2020; Lohr & Younes, 2024); Verdière et al. (2025) learn reduced features from components of a diffeomorphism. These methods do not require inverse compensation around a prescribed dynamics learner within a frozen encoder–decoder interface. Algebraic invertibility also need not control inverse sensitivity (Behrmann et al., 2021; Wang et al., 2024). PCEGA combines refitted predictors with tied inverses

under a global distortion budget. Appendix D.2 compares an author-source FineMorphs adaptation on the same frozen charts.

Coordinate learning for dynamics. Koopman, SINDy and consistent Koopman autoencoders jointly learn representations and dynamics (Lusch et al., 2018; Champion et al., 2019; Azencot et al., 2020). Conjugate-mapping models also learn coordinates and latent dynamics (Bramburger et al., 2021). FlowDMD, EDMD-IDL and AIKAE already exploit invertibility for reconstruction (Meng et al., 2024; Jin et al., 2024; Frion et al., 2025), focusing on observables for linear latent evolution. We instead adapt an existing representation to a specified finite-capacity learner family, which may be nonlinear. Freezing E, D and imposing $DH^{-1}HE = DE$ isolates coordinate utility from changes to the base reconstruction.

Usable information and representation symmetry. Predictive $\mathcal{V}$-information and explicitness distinguish encoded information from its accessibility to restricted predictors (Xu et al., 2020; Eastwood et al., 2023). Representation-symmetry analyses connect downstream risk to fitted-predictor invariance (Tinati & Tu, 2026) and characterize coordinate-stable probe spaces (Lee & Kondor, 2026). We study the two-sided conjugacy $H^{-1}\mathcal{G}H$ of vector-valued transition maps and characterize its affine equivalence boundary for polynomial dynamics.

3 Problem Formulation and Profiled Coordinate Adaptation

3.1 Prediction interface and the coordinate bottleneck

Let O be an observation or fixed-observer state, O^+ its value at a prescribed prediction horizon, and $Y = r(O^+)$ the physical target under a fixed readout r (identity for full-state prediction). Frozen E maps O to $z \in \mathbb{R}^d$, and frozen D approximates the readout from this state. For a fitting set S of n_S transitions, let $z_i = E(O_i)$, $z_i^+ = E(O_i^+)$ and $S_H = \{(H(z_i), H(z_i^+))\}_{i=1}^{n_S}$. A fitting algorithm $\mathcal{A}_\lambda$, including feature construction, preprocessing and regularization, returns the downstream predictor $\widehat{g}_{H,S,\lambda} = \mathcal{A}_\lambda(S_H) \in \mathcal{G}$. The complete physical predictor is

$$f_{H,S,\lambda} = D \circ H^{-1} \circ \widehat{g}_{H,S,\lambda} \circ H \circ E. \tag{2}$$

Known exogenous inputs enter the downstream learner directly and bypass H; their arguments are omitted for clarity.

For a bimeasurable bijection H, $\sigma(H(z)) = \sigma(z)$ and $DH^{-1}HE = DE$. Preservation concerns the information retained by E and its original reconstruction, without requiring lossless encoding. In the original latent coordinates, the effective prediction class is

$$\mathcal{G}_H = H^{-1}\mathcal{G}H := \{H^{-1} \circ g \circ H : g \in \mathcal{G}\}. \tag{3}$$

When $\mathcal{G}_H \neq \mathcal{G}$, the realizable latent functions change; finite-sample fitting and regularization can also change performance. We call the mismatch between retained information and its accessibility to a specified downstream learner a *coordinate bottleneck*.

3.2 Explicit inverses and certified distortion

Partition the latent vector into two blocks, $z = (a, b)$. Let θ_j denote the trainable weights and biases of conditioner t_j, and collect them as $\theta = (\theta_1, \ldots, \theta_m)$. Define the additive coupling and its inverse as

$$C_j(a,b) = \big(a,\ b + t_j(a;\theta_j)\big), \qquad C_j^{-1}(a,b) = \big(a,\ b - t_j(a;\theta_j)\big). \tag{4}$$

Alternating the updated block gives $H_\theta = C_m \circ \cdots \circ C_1$, whose inverse applies the explicit layer inverses in reverse order. Zero conditioner output matrices and biases initialize identity. For conditioners with 1-Lipschitz activations, the product of the effective weight matrices' spectral norms supplies a global bound L_j. Both the forward and inverse coupling constants are bounded by $\rho(L_j) = (\sqrt{L_j^2 + 4} + L_j)/2$, see Lemma 1 in Appendix A.2 for a derivation. Composition gives

$$\max\{\mathrm{Lip}(H_\theta), \mathrm{Lip}(H_\theta^{-1})\} \le B(\theta) := \prod_{j=1}^{m} \rho(L_j). \tag{5}$$

PCEGA supports two ways of allocating the distortion budget. The layer-budget (LB) variant constrains each coupling layer separately, whereas the global-budget (GB) variant controls the composed distortion bound and allows layers to share the budget. The following sections describe the default GB procedure; experiments using LB are explicitly identified. We impose $B(\theta) \le \kappa$, defining the certified family $\mathcal{H}_\kappa$ in the declared standardized coordinates. $B(\theta)$ is a sufficient upper bound, not the exact Lipschitz constant. The GB procedure restores feasibility after updates by jointly scaling conditioner output matrices: $L_j \mapsto sL_j$, with $s \in [0,1]$ found by bisection. An internal cap $1 \le \kappa_{\text{int}} \le \kappa$ provides a numerical margin. Layers share the composed budget; rescaling is not an optimal projection (Appendix C.5).

3.3 Profiled fitting and coordinate search

For each candidate coordinate map, we fit the downstream predictor in the transformed coordinates and select its ridge strength using inner development data. We then refit the predictor and evaluate decoded prediction error on outer development data. This error guides the coordinate update, allowing the map to adapt to the predictor's fitting procedure.

For fixed H, S, let $X_{H,S}$ be the feature matrix including an intercept and $Z^+_{H,S}$ the transformed targets, with normalization fitted on S only. We fit the regression head by solving

$$\widehat{W}_{H,S,\lambda} = \arg\min_W \left\{ n_S^{-1} \|X_{H,S}W - Z^+_{H,S}\|_F^2 + \lambda \operatorname{tr}(W^\top P_{H,S} W) \right\}. \tag{6}$$

Here $\widehat{W}_{H,S,\lambda}$ contains the coefficients of regression head. The parameter $\lambda \in \Lambda \subset [0,\infty)$ controls learner regularization, where Λ is a fixed finite ridge grid; θ parameterizes the coordinate map and κ bounds its distortion. Default covariance-aware ridge uses the fitting-set feature covariance with a numerical floor as the nonconstant block of $P_{H,S} \succeq 0$, leaving the intercept unpenalized. Differentiated fits require $X^\top_{H,S}X_{H,S}/n_S + \lambda P_{H,S}$ to be invertible; solver and failure conventions appear in Appendix C.3.

For held-out trajectories V, decoded physical error is

$$\widehat{R}_V(H; S, \lambda) = |V|^{-1} \sum_{\tau \in V} \mathcal{L}_\tau(f_{H,S,\lambda}), \tag{7}$$

where each study fixes its trajectory loss, normalization and evaluation times. Development supplies K rotating splits into disjoint support, inner and outer trajectories. For split r, we select ridge by inner error after fitting on support, then refit on support plus inner data before evaluating outer error:

$$\begin{aligned} &\lambda_r(H) = \arg\min_{\lambda \in \Lambda} \widehat{R}_{S^r_{\text{in}}}(H; S^r_{\text{sup}}, \lambda), \\ &\mathcal{L}_r(H) = \widehat{R}_{S^r_{\text{out}}}(H; S^r_{\text{sup}} \cup S^r_{\text{in}}, \lambda_r(H)), \qquad \min_{H \in \mathcal{H}_\kappa} K^{-1} \sum_r \mathcal{L}_r(H). \end{aligned} \tag{8}$$

Ridge ties use the fixed grid order. Outer data participate in coordinate training and are not the final test set.

Ridge selection is discrete. We differentiate through the selected refit, H, H^{-1} and D, holding the ridge choice and numerical covariance floor fixed. The resulting selected-branch hypergradient, $\widetilde{\nabla}$, is approximate; complete-training controls appear in Appendix C.3. This procedure requires differentiable fits, so black-box learners need a different coordinate-search method. Moreover, using physical error for selection does not make transformed-target regression equivalent to physical-loss fitting, motivating the control in Section 5.3.

3.4 Selection and deployment

The default GB procedure separates four data roles at the trajectory level. **Development** supplies the folds used to optimize H. **Checkpoint** selects the coordinate map and final ridge coefficients. **Selector** accepts or rejects the frozen candidate, and **Report** evaluates the final decision. Coordinate gradients use Development only; Selector and Report do not participate in checkpoint or ridge selection.

Algorithm 1 PCEGA: default global-budget procedure.

Input: Frozen (E, D); fitter $\mathcal{A}_\lambda$; ridge grid Λ;
disjoint Development, Checkpoint, Selector and Report;
K folds, T updates, checkpoint schedule and $1 \le \kappa_{\text{int}} \le \kappa$.
Output: Frozen coordinate–learner pair $(H^\star, \widehat{g}^\star)$.
Initialize $H_\theta = I$ and checkpoint set $\mathcal{C} = \{(0, I)\}$.
for $t = 1, \dots, T$ **do**
  Select λ_r, refit learners and compute $\mathcal{L}_r$ on all K folds using Eq. (8).
  Update θ with $\widetilde{\nabla}_\theta K^{-1} \sum_r \mathcal{L}_r(H_\theta)$.
  Restore $B(\theta) \le \kappa_{\text{int}}$ if needed; save eligible scheduled checkpoints in $\mathcal{C}$.
end for
Select H_c from $\mathcal{C}$ using the prescribed Checkpoint score.
Independently tune ridge for $Q \in \{I, H_c\}$ on Development/Checkpoint.
Refit each $\widehat{g}_Q$ on Development $\cup$ Checkpoint and freeze both pairs.
$(H^\star, \widehat{g}^\star) \leftarrow (I, \widehat{g}_I)$.
if the candidate passes the fixed Selector rule **then**
  $(H^\star, \widehat{g}^\star) \leftarrow (H_c, \widehat{g}_{H_c})$.
end if
return $(H^\star, \widehat{g}^\star)$; evaluate on Report without further fitting or selection.

Scheduled coordinate maps are eligible when $B(\theta) \le \kappa$ and inverse round-trip error meets a fixed tolerance. Checkpoint uses fixed tuning/evaluation splits to select among eligible maps, including identity; ties favor the earlier update. Learners in the candidate and identity coordinates are then independently ridge-tuned by fitting on Development and scoring on Checkpoint, and refitted on their union. This tuning is separate from inner ridge selection.

Selector compares these frozen pairs using fixed thresholds for mean error reduction and the number of improving trajectories. Rejection returns identity and its independently fitted predictor without reopening search or revising thresholds. Algorithm 1 summarizes the procedure; scoring details and the expanded algorithm are in Appendices C.5, C.3 and F.2.1. Historical measured protocols that refit after Selector are distinguished in Table 29. Deployment uses Eq. (2); neither acceptance nor certification guarantees a test gain.

4 Design Guarantees and Expressivity Boundary

Preservation and sensitivity. Together with $DH^{-1}HE = DE$, Eq. (5) gives

$$\kappa^{-1}\|z - z'\| \le \|H(z) - H(z')\| \le \kappa\|z - z'\|. \tag{9}$$

If D is L_D-Lipschitz on the relevant domain, a perturbation e in transformed latent space satisfies

$$\|DH^{-1}(w + e) - DH^{-1}(w)\| \le L_D\kappa\|e\|. \tag{10}$$

This conditional one-step bound controls inverse-coordinate sensitivity, not recursive stability. Norms refer to the standardized coordinates (Appendices C.1 and B.3).

The affine boundary and fitted prediction. Let $\mathcal{P}_p^d$ be all maps $\mathbb{R}^d \to \mathbb{R}^d$ with componentwise polynomials of total degree at most p, including constants.

Theorem 1 (Affine boundary for polynomial dynamics)**.** *For a finite integer $p \ge 1$ and a globally bi-Lipschitz bijection $H : \mathbb{R}^d \to \mathbb{R}^d$,*

$$H^{-1}\mathcal{P}_p^d H = \mathcal{P}_p^d \quad \Longleftrightarrow \quad H \text{ is invertible affine}. \tag{11}$$

The proof uses conjugated translations: their polynomial displacement is bounded and hence constant, forcing H to be affine (Appendix A.1). Nonaffine changes therefore alter the class without changing latent information, but cannot improve all tasks: for every nonaffine H, two globally Lipschitz tasks have strictly opposite approximation preferences under an identity decoder and a full-support distribution with all polynomial moments finite (Corollary 1). This motivates task-specific search with identity retained as a candidate.

Table 1: Test NMAE reduction (%) for the validation-gated predictor in the analytic-chart benchmark. Positive values indicate improvement relative to independently profiled predictors in identity coordinates. Zero denotes identity fallback. Each cell averages three technical-seed risks before comparing twelve paired trajectories.

System	Linear	Q2	ELM	RBF	MLP	Trees
Lorenz–63	64.64	24.87	23.74	33.50	0.00	0.00
Rössler	57.18	32.85	30.41	36.97	0.00	0.00
Duffing	15.39	16.32	24.82	17.05	0.00	0.00
Hénon	57.12	37.32	47.20	71.11	0.00	5.69

The theorem concerns global latent classes. Restriction to the data support and noninjective decoding may erase their distinction in physical predictions. For fitted polynomial learners, affine equivariance additionally requires compatible normalization and transport of the regularization metric; a coordinate-local covariance floor can break it (Proposition 1). RBF and finitely trained neural predictors need not share this invariance. We therefore assess fitted performance with independently tuned predictors and decoded physical error.

5 Experiments

We first evaluate incremental prediction gains, then study coordinate search and learner structure. Original efficacy and structural studies use layer-budget (LB) constraints with separate caps for individual coupling layers. Later learned-chart and flow studies use the global-budget (GB) procedure, allowing layers to share the composed distortion budget. Comparisons are matched within each study's frozen protocol. Synthetic evaluation uses independent trajectories; measured studies use the specified records and temporal splits. Appendix Table 29 summarizes the fitting variants; Appendices F.2 and H give the full data roles and estimators.

5.1 Coordinate Adaptation Benefits Multiple Learner Families

Gains across four learners and four systems. Coordinate adaptation improves prediction for Linear, Q2, random-feature (ELM), and RBF learners across all four analytic-chart systems in Table 1. Each system has a known six-dimensional chart and 40-dimensional observations. Certified coordinate maps are learned separately for the four learner families under the LB protocol. For each evaluated learner, Atlas selects from the resulting coordinate-map library and identity before the independent Selector gate. On twelve test trajectories per system, the four primary learner families achieve NMAE reductions of 15.39%–71.11% relative to independently profiled predictors in identity coordinates. MLP returns identity in all four systems, while Trees improve only on Hénon, indicating that the gains depend on the learner. The full protocol, controls, and absolute errors are reported in Appendix E.4.

Under separate frozen protocols, LB adaptation also improves prediction on L3F, lifted Rössler, Silverbox, and Wiener–Hammerstein (Table 2). Synthetic affine/PCA controls have approximately zero Q2 effect, and random certified maps do not reproduce the learned gains. Measured-system predictors estimate the full next-delay vector, are refitted with Selector data after coordinate selection, and are evaluated by continuous physical free-run error. Exact-history-shift controls appear in Appendix G. Synthetic evaluations use three optimization seeds and independent trajectories. Silverbox and Wiener–Hammerstein each use one continuous record, with 21,638 and 78,750 samples, respectively; these sample counts are not counts of independent replications.

Prediction gains on frozen flow representations. GB adaptation improves prediction on a held-out block of measured cylinder-wake PIV data with $21{,}600$ observed values per frame. With frozen CNN encoder–decoders at latent dimensions 16 and 32 and standardized Q2 predictors with diagonal ridge, MSE decreases by 1.93%–3.04% across four dimension–horizon settings, averaging three optimization seeds. All twelve candidates pass selection and improve held-out error, with distortion bounds below three and reconstruction preserved up to floating-point error (Appendix I.1). Affine

Table 2: Original within-learner results under separate frozen protocols. Units, metrics and evaluation records are specific to each row; effects are not pooled. Synthetic effects use the original trajectory-macro estimator.

System	Metric / evaluation	Identity	LB-PCEGA	Reduction
L3F	NMAE / 10 trajectories	0.097922	0.084769	13.421%
Lifted Rössler	NMAE / 10 trajectories	0.033879	0.031762	6.225%
Silverbox	RMSE (mV) / 1 record	9.0847	4.8716	46.375%
Wiener–Hammerstein	RMSE / 1 record	0.156839	0.140424	10.466%

and physical-loss controls identify fitting and initialization effects, without establishing that nonlinear adaptation is necessary (Appendix J).

Lower amylase MSE with frozen protein embeddings. An input-side extension lowers amylase MSE using all 320 dimensions of frozen ESM-2 8M embeddings (Lin et al., 2023). A certified coordinate map is trained with a linear predictor, then frozen before fitting an MLP. It preserves embedding information under a global bound of 2.99, without decoder compensation or training-time profiling. On the previously explored FLIP2 split (Didi et al., 2026), MSE decreases by 10.34% relative to identity, with five paired wins and all five maps retained by validation (Table 3). These development-stage results are metric- and task-dependent: mean Spearman decreases from 0.5238 to 0.5074, the forward-mutation split shows no consistent raw-adapter gain, and the second protein task returns the constant baseline (Appendix L.3). A separate nested-ridge adaptation study appears in Appendix L.2.

Table 3: Amylase regression on frozen embeddings. NMSE is mean $\pm$ SD over five complete technical repeats; other columns are means. The public split was previously explored, so these are development-stage robustness results. Longer training matches updates, not runtime; the wider MLP approximately matches deployed parameters.

Method	NMSE	MAE	Spearman
Identity MLP	0.9800 ± 0.0467	0.0461	0.5238
Longer-trained MLP	0.9497 ± 0.0466	0.0452	0.5380
Wider MLP	0.9767 ± 0.0210	0.0460	0.5354
Random certified map + MLP	0.9744 ± 0.0492	0.0456	0.5285
Certified map + MLP	0.8786 ± 0.0288	0.0412	0.5074

5.2 Global-Budget Search Improves Prediction on Learned Charts

The complete workflow improves prediction across charts. The default GB procedure reduces NMSE by 48.82% on L3F and 45.11% on Rössler across independently trained charts (Table 4). Each run repeats Algorithm 1, including coordinate search, predictor fitting, and selection with identity fallback. The controlled studies below use separate protocols.

Table 4: Complete default GB workflow (Algorithm 1). Each system uses three independently trained charts and two seeds for coordinate optimization per chart and twenty new Report trajectories shared within each system. Gains are ratios of mean risks; acceptance counts technical runs, not independent charts. Full protocol: Appendix C.2.

System	Identity NMSE	GB NMSE	Reduction	Accepted
L3F	0.007337	0.003756	48.82%	6/6
Rössler	0.001931	0.001060	45.11%	6/6

Search and training choices affect prediction gains. Enforcing the budget during optimization yields larger error reductions than selecting certified checkpoints from an uncapped path in a matched study. LB, GB, and uncapped paths share identical effective initialization and new downstream data on three charts per system. At 192 updates, GB reduces NMSE relative to identity by

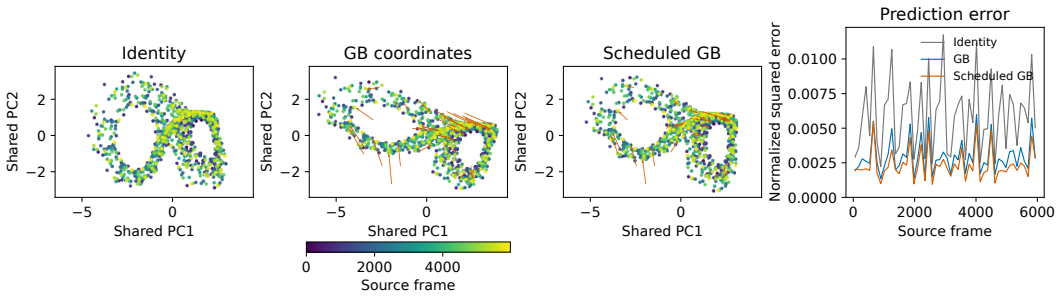

Figure 2: Coordinate adaptation on a frozen L3F chart. All maps share a Development-fitted PCA projection and axes. Color denotes source time; arrows show unscaled projected displacement and errors use 50 fixed time bins. This descriptive Checkpoint trajectory was used for selection and fitting, not independent evaluation. Details and both systems: Appendix F.3.

$50.10\%/48.81\%$ on L3F/Rössler, compared with $31.91\%/33.17\%$ for the best certified uncapped checkpoints (Appendix C.5).

Further gains arise primarily from the learning-rate recipe in a matched ablation. Changing the learning-rate schedule and fixing ridge during optimization reduces NMSE by an additional $13.89\%/18.83\%$ relative to the original 192-update GB recipe. This comparison uses one fixed chart per system and three new joint data–initialization repetitions; all six improve, and all paths select the last allowed update (Appendix C.4). A subsequent matched 2×2 ablation across six frozen charts separates the two changes: profiled and fixed-ridge training yield nearly identical risks under matched learning-rate settings (Table 13). Profiling remains our fitting-aware search implementation; final predictors are independently tuned throughout.

5.3 Coordinate Gains Persist after Physical-Loss Refitting

Matched physical-loss refitting does not remove the coordinate gains at the tested budget. Holding H, E, D fixed, we optimize the Q2 coefficients directly through the decoder for both adapted and identity coordinates. All 18 predictors are frozen before evaluation on ten new trajectories per system. This control evaluates fixed candidates, not the complete acceptance procedure (Appendix F.1).

Refitting reduces NMSE in all six coordinate–system combinations. Predictors using the scheduled GB maps retain $58.73\%/51.37\%$ gains on L3F/Rössler over equally refitted identity predictors (Appendix Figure 3). Reconstructed-target fitting gives similar results. Figure 2 provides a descriptive visualization of coordinate changes and associated prediction errors on a Checkpoint trajectory.

5.4 Coordinate Utility Depends on Learner Family and Structure

Preferred coordinates depend on the learner. Q2 and RBF each prefer maps learned for their own family in the crossed comparisons on L3F and Rössler (Table 5). We evaluate LB maps learned for Q2 and RBF using Q2, RBF, and MLP downstream learners for direct $h = 12$ prediction. The study includes 120 retained candidates, three optimization seeds, and 20 new trajectories per system. Each learner is independently profiled and refitted, with fitting budgets matched within families; no test trajectory selects a map or hyperparameter (Appendix B.1).

All four other-minus-own contrasts have positive prespecified approximate 98.75% trajectory-bootstrap intervals and agree in sign across all three seeds. Intervals average the fixed seed family within trajectories (Appendix E.1). The small Rössler RBF gain over identity is less robust to seed resampling. MLP tests coordinate transfer, not MLP-specific coordinate learning; separate adaptation controls appear in Appendix K.

Two controls support this learner dependence. A prespecified straightener on the analytic Lorenz chart worsens Q2 prediction but improves ELM prediction while preserving reconstruction exactly (Appendix E.5). Input-only controls retain both directed Q2/RBF preferences on 20 additional trajectories per system, with all four adjusted lower endpoints positive. These controls hold the maps

Table 5: Absolute physical NMAE in the crossed experiment. Each entry averages the fixed three-seed family and 20 fresh trajectories. Bold marks the lower of the two learned-coordinate errors for the Q2 and RBF learners.

System	Learner	Identity	Q2 coordinate	RBF coordinate
L3F	Q2	0.097292	**0.085215**	0.096603
	RBF-256	0.006382	0.006216	**0.005742**
	MLP	0.010412	0.011061	0.011363
Rössler	Q2	0.033951	**0.031790**	0.033779
	RBF-256	0.004684	0.004746	**0.004642**
	MLP	0.006132	0.006572	0.005991

fixed and independently tune each learner, retaining coordinate-specific normalization (Appendix E.3, Table 22).

Localization reduces coordinate gains for Q2 learners. On L3F, holding the maps learned for global Q2 fixed while replacing the global model with $K = 4$ and $K = 16$ local Q2 models reduces physical prediction gains over the corresponding identity baselines from 13.09% to 4.99% and 0.33%, respectively. Reoptimizing the maps for $K = 16$ selects identity. These results show that measured coordinate gains depend on learner structure, although localization also changes model size and estimation (Appendix E.2).

6 Discussion and Conclusion

PCEGA makes coordinate adaptation a separate, trainable component of latent dynamics prediction, preserving information and base reconstruction under a global geometric budget. It improves specified finite-capacity downstream learners across controlled dynamics, measured systems and high-dimensional flow observations.

The affine equivalence theorem characterizes transformations preserving the full polynomial dynamics class. Cross-learner experiments, physical-loss refitting and localization show how coordinate utility depends on the learner family, fitting and model structure; flow controls also identify affine fitting effects. Together, these results connect exact representation preservation to its computational usefulness for a specified downstream learner.

Coordinate adaptation complements learner selection. In our common-interface comparison, local Q2, RBF, MLP, and FineMorphs achieve lower prediction errors than GB–Q2; RBF and MLP also have lower batch-one latency and smaller prediction state (Appendix D.2). The gains reported here therefore support adaptation within a specified learner family, while the choice of that family remains consequential.

Our guarantees concern information preservation, reconstruction, and coordinate distortion. Extending these guarantees to decoded robustness and recursive prediction stability, and evaluating adaptation across independent physical runs, are natural directions for future work.

AI use statement. Generative AI tools were used to assist with language polishing and manuscript review, as well as to generate portions of the code. The main research ideas were developed and organized by the authors. The authors take full responsibility for the final manuscript, including its theoretical claims, implementation, and reported results.

REFERENCES

Omri Azencot, N. Benjamin Erichson, Vanessa Lin, and Michael W. Mahoney. Forecasting sequential data using consistent koopman autoencoders. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 475–485. PMLR, 2020.

Jens Behrmann, Paul Vicol, Kuan-Chieh Wang, Roger Grosse, and Joern-Henrik Jacobsen. Understanding and mitigating exploding inverses in invertible neural networks. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130, pp. 1792–1800. PMLR, 2021.

Jason J. Bramburger, Steven L. Brunton, and J. Nathan Kutz. Deep learning of conjugate mappings. *Physica D: Nonlinear Phenomena*, 427:133008, 2021. doi: 10.1016/j.physd.2021.133008.

Kathleen Champion, Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116(45):22445–22451, 2019. doi: 10.1073/pnas.1906995116.

Max D. Champneys, Gerben I. Beintema, Roland Tóth, Maarten Schoukens, and Timothy J. Rogers. Baseline results for selected nonlinear system identification benchmarks. *arXiv preprint arXiv:2405.10779*, 2024. URL https://arxiv.org/abs/2405.10779.

Frederic Chazal, Vin de Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173:193–214, 2014. doi: 10.1007/s10711-013-9937-z.

Gustav Conradie, Nicolas Boullé, Jean-Christophe Loiseau, Steven L. Brunton, and Matthew J. Colbrook. Trustworthy koopman operator learning: Invariance diagnostics and error bounds. *arXiv preprint arXiv:2603.15091*, 2026. URL https://arxiv.org/abs/2603.15091.

Kieran Didi, Sarah Alamdari, Xijie Lu Alex, Bruce Wittmann, Kadina Johnston, Ava Amini, Ali Madani, Maya Czeneszew, Christian Dallago, and Kevin Yang. FLIP2: Expanding protein fitness landscape benchmarks for real-world machine learning applications. Zenodo, version 3, dataset, 2026. URL https://zenodo.org/records/18433203.

Cian Eastwood, Andrei Liviu Nicolicioiu, Julius von Kügelgen, Armin Kekić, Frederik Träuble, Andrea Dittadi, and Bernhard Schölkopf. DCI-ES: An extended disentanglement framework with connections to identifiability. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/pdf?id=462z-gLgSht.

Anthony Frion, Lucas Drumetz, Mauro Dalla Mura, Guillaume Tochon, and Abdeldjalil Aïssa-El-Bey. Augmented invertible koopman autoencoder for long-term time series forecasting. *Transactions on Machine Learning Research*, 2025. URL https://openreview.net/forum?id=o6ukhJLzMQ.

Yuhong Jin, Lei Hou, and Shun Zhong. Extended dynamic mode decomposition with invertible dictionary learning. *Neural Networks*, 173:106177, 2024. doi: 10.1016/j.neunet.2024.106177.

Su Hyeong Lee and Risi Kondor. Deep minds and shallow probes. *arXiv preprint arXiv:2605.11448*, 2026. URL https://arxiv.org/abs/2605.11448.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan Dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. doi: 10.1126/science.ade2574. URL https://doi.org/10.1126/science.ade2574.

Alec J. Linot and Michael D. Graham. Deep learning to discover and predict dynamics on an inertial manifold. *Physical Review E*, 101(6):062209, 2020. doi: 10.1103/PhysRevE.101.062209.

Michele Lohr and Laurent Younes. Finemorphs: Affine-diffeomorphic sequences for regression. *Journal of Machine Learning Research*, 25(245):1–38, 2024. URL https://jmlr.org/papers/v25/23-0824.html.

Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9:4950, 2018. doi: 10.1038/s41467-018-07210-0.

Yuhuang Meng, Jianguo Huang, and Yue Qiu. Koopman operator learning using invertible neural networks. *Journal of Computational Physics*, 501:112795, 2024. doi: 10.1016/j.jcp.2024.112795.

Eric Nguyen, Michael Poli, Matthew G. Durrant, Brian Kang, Dhruva Katrekar, David B. Li, Liam J. Bartie, Armin W. Thomas, Samuel H. King, Garyk Brixi, Jeremy Sullivan, Madelena Y. Ng, Ashley Lewis, Aaron Lou, Stefano Ermon, Stephen A. Baccus, Tina Hernandez-Boussard, Christopher Ré, Patrick D. Hsu, and Brian L. Hie. Sequence modeling and design from molecular to genome scale with Evo. *Science*, 386(6723):eado9336, 2024. doi: 10.1126/science.ado9336. URL https://doi.org/10.1126/science.ado9336.

Feliks Nüske, Sebastian Peitz, Friedrich Philipp, Manuel Schaller, and Karl Worthmann. Finite-data error bounds for koopman-based prediction and control. *Journal of Nonlinear Science*, 33(1):14, 2023. doi: 10.1007/s00332-022-09862-1.

Johan Schoukens, Johan A. K. Suykens, and Lennart Ljung. Wiener–hammerstein benchmark. In *Proceedings of the 15th IFAC Symposium on System Identification*, Saint-Malo, France, 2009. URL https://www.nonlinearbenchmark.org/benchmarks/wiener-hammerstein.

Alejandro Tejada-Lapuerta, Anna C. Schaar, Robert Gutgesell, Giovanni Palla, Lennard Halle, Mariia Minaeva, Larsen Vornholz, Leander Dony, Francesca Drummer, Till Richter, Mojtaba Bahrami, and Fabian J. Theis. Nicheformer: a foundation model for single-cell and spatial omics. *Nature Methods*, 22(12):2525–2538, 2025. doi: 10.1038/s41592-025-02814-z. URL https://doi.org/10.1038/s41592-025-02814-z.

Christina V. Theodoris, Ling Xiao, Anant Chopra, Mark D. Chaffin, Zeina R. Al Sayed, Matthew C. Hill, Helene Mantineo, Elizabeth M. Brydon, Zexian Zeng, X. Shirley Liu, and Patrick T. Ellinor. Transfer learning enables predictions in network biology. *Nature*, 618(7965):616–624, 2023. doi: 10.1038/s41586-023-06139-9. URL https://doi.org/10.1038/s41586-023-06139-9.

Mohammad Tinati and Stephen Tu. On the asymptotics of self-supervised pre-training: Two-stage m-estimation and representation symmetry. In *Proceedings of Thirty Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 6197–6309, 2026. URL https://proceedings.mlr.press/v336/tinati26a.html.

Romain Verdière, Clémentine Prieur, and Olivier Zahm. Diffeomorphism-based feature learning using Poincaré inequalities on augmented input space. *Journal of Machine Learning Research*, 26(139):1–31, 2025. URL https://jmlr.org/papers/v26/23-1707.html.

Ruigang Wang, Krishnamurthy Dvijotham, and Ian Manchester. Monotone, bi-lipschitz, and polyak–lojasiewicz networks. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 50379–50399. PMLR, 2024.

Torbjörn Wigren and Johan Schoukens. Three free data sets for development and benchmarking in nonlinear system identification. In *2013 European Control Conference*, pp. 2933–2938, 2013. doi: 10.23919/ECC.2013.6669201.

Yilun Xu, Shengjia Zhao, Jiaming Song, Russell Stewart, and Stefano Ermon. A theory of usable information under computational constraints. In *International Conference on Learning Representations*, 2020. URL https://arxiv.org/abs/2002.10689.

Laurent Younes. Diffeomorphic learning. *Journal of Machine Learning Research*, 21(220):1–28, 2020. URL https://jmlr.org/papers/v21/18-415.html.

APPENDIX CONTENTS

A Theory and Proofs **13**
A.1 Structural Proof and Scope . . . 13
A.2 Global Coupling Certificate . . . 14
A.3 Affine Equivariance of Profiled Heads . . . 14
A.4 Auxiliary Results and Proofs . . . 15

B Experimental Configuration and Estimation **16**

C Component Construction and Search **19**
C.1 Original Layer-Budget Adaptation Method . . . 19
C.2 Independent Chart and Global-Budget Study . . . 21
C.3 Numerical fitting and complete-training sensitivity . . . 23
C.4 A simpler training recipe with an unchanged component . . . 23
C.5 Matched initialization and eligible-checkpoint selection . . . 24

D Original Effects and Receiver Alternatives **25**

E Receiver Structure and Transfer **28**
E.1 Crossed-Interaction Details . . . 28
E.2 Structural Contraction: Evidence Roles and Metrics . . . 29
E.3 Frozen Cross-Component Evaluation . . . 29
E.4 Controlled Six-Receiver Benchmark . . . 30
E.5 Analytic straightener control . . . 33

F Physical-Loss Refitting and Interface Details **34**

G Coordinate Sensitivity and Measured Rollout Diagnostics **36**

H Study Protocols and Risk Estimators **38**

I Coordinate Adaptation for High-Dimensional Flow Observations **38**

J Affine and Physical-Loss Controls on the Flow Representation **39**

K MLP-specific Training-time Adaptation **40**

L Supervised Coordinate Adaptation for Protein Embeddings **41**
L.1 Task and two-stage protocol . . . 41
L.2 Original-component port with nested profiling . . . 42
L.3 Additional settings and evidence boundaries . . . 43

M Archived Development and Reproducibility **43**

A Theory and Proofs

Proof index. The main expressivity boundary and strict risk separation are proved in Appendix A.1; the coupling certificate and affine fitting equivariance are in Appendices A.2 and A.3. Auxiliary class, risk and topology results follow in Appendix A.4. The exact-orbit and local-fitting identities remain with their definitions in Appendix B.2; fitted-residual transport remains in Appendix C.1. Each proof follows its statement.

A.1 Structural Proof and Scope

Proof of Theorem 1. Step 1: conjugated translations have bounded displacement. Assume $H^{-1}\mathcal{P}_p^d H = \mathcal{P}_p^d$. For any $b \in \mathbb{R}^d$, $T_b(y) = y + b$ belongs to this class because $p \geq 1$. Consequently $Q_b = H^{-1}T_b H$ is polynomial and

$$\|Q_b(x) - x\| = \|H^{-1}(H(x) + b) - H^{-1}(H(x))\| \leq \mathrm{Lip}(H^{-1})\|b\| \qquad (x \in \mathbb{R}^d).$$

Step 2: bounded polynomial displacement is constant. Each component of $Q_b(x) - x$ is a globally bounded real polynomial. A nonconstant polynomial has a nonzero highest-degree homogeneous part; on a line in a direction where that part is nonzero, it is unbounded. Thus $Q_b(x) = x + c_b$ for some vector c_b independent of x.

Step 3: the inverse is affine. Surjectivity of H and Step 2 give $H^{-1}(y + b) - H^{-1}(y) = c_b$ for all y, b. Setting $y = 0$ identifies $c_b = H^{-1}(b) - H^{-1}(0)$. Hence $f(y) = H^{-1}(y) - H^{-1}(0)$ satisfies

$$f(y + b) = f(y) + f(b).$$

Additivity implies rational homogeneity. Continuity of the Lipschitz map f extends it to real homogeneity, so $f(y) = Ay$. Bijection of H^{-1} makes A invertible. Therefore H is affine.

Step 4: affine maps suffice. If $H(x) = Bx + a$ with B invertible, substitution into $H^{-1}gH$ preserves total degree at most p for every $g \in \mathcal{P}_p^d$. This gives one inclusion; applying the same argument to H^{-1} gives the reverse inclusion. Both H and H^{-1} are globally Lipschitz. □

Corollary 1 (Task-dependent approximation preferences)**.** *Fix a finite integer $p \geq 1$, an identity physical decoder, and a full-support input distribution on $\mathbb{R}^d$ with all polynomial moments finite. For every nonaffine globally bi-Lipschitz bijection H, there exist two globally Lipschitz transition tasks with strictly opposite preferences: one has zero optimal squared approximation error in $H^{-1}\mathcal{P}_p^d H$ and positive error in $\mathcal{P}_p^d$; the other has the reverse.*

Proof. Let μ denote the input distribution. *Construct the two tasks.* Steps 1–3 above show more than class nonclosure: if every conjugated translation were polynomial, H would be affine. There is therefore a b for which $F_+ = H^{-1}T_b H$ is nonpolynomial. It is exactly representable in H coordinates because $T_b \in \mathcal{P}_p^d$. Apply the same argument to nonaffine H^{-1} to obtain a c for which HT_cH^{-1} is nonpolynomial. Set $F_- = T_c$, which is exact in identity coordinates. Both physical tasks are globally Lipschitz.

Turn nonmembership into a positive risk gap. Under the moment assumption, $\mathcal{P}_p^d$ is a finite-dimensional, hence closed, subspace of $L^2(\mu; \mathbb{R}^d)$. Full support ensures that a continuous nonpolynomial map cannot equal a polynomial almost everywhere: their difference at any point persists on an open neighborhood of positive measure. Thus F_+ has strictly positive distance from $\mathcal{P}_p^d$, while its gauged error is zero.

For F_-, write $\nu = H_{\#}\mu$ and $G = HT_cH^{-1}$. The pushforward has full support and all polynomial moments because H is a bi-Lipschitz homeomorphism with at most linear growth. The same closed-subspace argument gives $\delta = \inf_{q \in \mathcal{P}_p^d} \|G - q\|_{L^2(\nu)} > 0$. For a common bi-Lipschitz constant κ, pointwise comparison gives

$$\kappa^{-1}\|G(w) - q(w)\| \leq \|H^{-1}G(w) - H^{-1}q(w)\| \leq \kappa\|G(w) - q(w)\|.$$

Square, integrate, and take infima: the gauged physical risk is between $\kappa^{-2}\delta^2$ and $\kappa^2\delta^2$, whereas identity risk is zero. This proves both strict preferences, without asserting a uniform gap. □

The proof uses elementary bounded-polynomial and continuous-Cauchy facts. It is not a priority claim for general representation symmetry or conjugacy. The global ambient result cannot be inferred from finitely many observations, or from a lower-dimensional attractor. Noninjective decoders may identify different latent predictions. Numerical quadrature checks in the artifact test the explicit shear example at degrees one through three; they are not a proof over all dynamics.

A.2 Global Coupling Certificate

Split $u = (u_a, u_b)$ and define an additive coupling

$$C_t(u_a, u_b) = (u_a, u_b + t(u_a)),$$
$$C_t^{-1}(v_a, v_b) = (v_a, v_b - t(v_a)). \tag{12}$$

It is analytically invertible for every shift network t.

Lemma 1 (Tight additive-coupling certificate)**.** *If $t : \mathbb{R}^p \to \mathbb{R}^q$ is globally L-Lipschitz in Euclidean norm, the forward and inverse coupling obey the sharp dimension-free bound*

$$\mathrm{Lip}(C_t), \mathrm{Lip}(C_t^{-1}) \le \rho(L) = \frac{\sqrt{L^2+4}+L}{2}. \tag{13}$$

Proof of Lemma 1. For two inputs, write a for their first-block difference, b for their second-block difference, and $\delta = t(u_a) - t(v_a)$. Then $\|\delta\| \le L\|a\|$. Maximizing $\|a\|^2 + \|b+\delta\|^2$ at fixed $\|a\|^2 + \|b\|^2$ aligns b and δ and reduces, also for rectangular shift Jacobians, to the scalar shear $S_L = \left[\begin{smallmatrix} 1 & 0 \\ L & 1 \end{smallmatrix}\right]$. More explicitly, the squared norm ratio is at most the largest eigenvalue of

$$S_L^\top S_L = \begin{bmatrix} 1+L^2 & L \\ L & 1 \end{bmatrix}, \qquad \lambda_{\max} = \frac{2+L^2+L\sqrt{L^2+4}}{2} = \rho(L)^2.$$

A scalar linear shift $t(a) = La$ with a maximizing eigenvector attains this bound, proving dimension-free sharpness. Replacing δ by $-\delta$ proves the same inverse bound. A tanh-separated shift network has $L \le \prod_j \|W_j\|_2$; channel masks are norm-nonincreasing coordinate projections, and permutations are orthogonal; coupling composition multiplies bounds. The train-only normalizer is a translation and scalar dilation conjugated through the full map, so its scalar factors cancel. This argument holds for a general composed budget. Implementation audits recompute effective float32 operator norms. The historical fixed-ridge precursor used public/construction values $1.25/1.245$; the principal cap-three studies use $3/2.995$. These are implementation margins, not different versions of the mathematical bound. □

A.3 Affine Equivariance of Profiled Heads

Proposition 1 (Ideal affine equivariance of profiled polynomial transition heads)**.** *Let $H(z) = Az + b$ with invertible A, use a complete polynomial dictionary through degree p including a constant term, transform source and target coordinates by the same H, and compensate predictions by H^{-1}. Suppose both coordinates use the corresponding support-fitted affine centering and normalization, the same loss and ridge normalization, and an invertible empirical normal matrix for every profiled λ. If feature covariance is used as the ridge metric without the numerical floor in Eq. (22), then the decoded predictor is unchanged for every fixed λ. Profiling is also affine-equivariant when it uses the same finite ridge grid, decoded validation loss and coordinate-independent tie rule.*

Proof of Proposition 1. After exact support centering and coordinate normalization, the affine map induces an invertible linear map L on both source and target coordinates. Completeness of the degree-p dictionary gives an invertible feature-basis map M with $\phi_p(Lz) = \phi_p(z)M$. In row-feature convention, feature Gram, covariance penalty, and feature–target moment transform as $G' = M^\top G M$, $C' = M^\top C M$, and $Q' = M^\top Q L^\top$. Hence

$$W'_\lambda = [M^\top(G + \lambda C)M]^{-1} M^\top Q L^\top = M^{-1} W_\lambda L^\top. \tag{14}$$

The transformed query prediction is therefore the affine transform of the original prediction; applying H^{-1} makes them identical. The equality holds pointwise for every λ, so the decoded validation scores agree entry by entry. A common tie rule then selects the same grid entry. The additive numerical covariance floor is deliberately excluded because it is not congruence-equivariant under a general nonorthogonal M. □

Corollary 2 (Controlled affine null)**.** *Under Proposition 1, the decoded predictor is also unchanged for a complete state–control polynomial dictionary when an invertible affine gauge acts only on the state and the exogenous control is left fixed.*

Proof of Corollary 2. Apply the preceding proof to the invertible augmented block-affine map $(z, u) \mapsto (Az + b, u)$. The feature-basis map remains invertible and decoder compensation acts only on the transformed state block. □

A.4 Auxiliary Results and Proofs

Proofs are placed immediately after their statements. The remaining results and proofs appear with their original formulation and component definitions in Appendices B.2 and C.1.

Proposition 2 (Generic nonlinear polynomial nonclosure)**.** *Let $p \geq 1$ and let $\mathcal{P}_p$ be all polynomials through degree p. If a component of a polynomial diffeomorphism H has degree $r > 1$, its pullback $H^*\mathcal{P}_p$ generically contains a nonzero degree-pr term and therefore differs from $\mathcal{P}_p$.*

Proof of Proposition 2. Choose a component H_j whose highest nonzero homogeneous term has degree $r > 1$ and take $f(y) = y_j^p$. If P_r is that leading homogeneous term, the leading term of $f \circ H$ is P_r^p. It is nonzero because the real polynomial ring has no zero divisors. Thus $f \circ H$ has degree $pr > p$ and cannot lie in $\mathcal{P}_p$. This establishes nonclosure, not the sign of its risk effect. □

Proposition 3 (Universal-coordinate regret)**.** *For positive weights w_c summing to one, define $R_{\mathrm{uni}} = \inf_H \sum_c w_c R_c(H)$ and $R_{\mathrm{atlas}} = \sum_c w_c \inf_H R_c(H)$. Then $R_{\mathrm{uni}} - R_{\mathrm{atlas}} \geq 0$, with equality when all conditions share a common attained minimizer. The inequality also holds for a frozen finite candidate family.*

Proof of Proposition 3. For every condition c, minimizing over its own gauge cannot have larger risk than evaluating any common gauge H. Weighting and summing these inequalities proves the claim. If the weighted universal objective attains its infimum, equality holds exactly when its minimizer minimizes every positive-weight condition. This applies in particular to a finite candidate library. Without universal attainment, equality only requires a sequence whose excess risk tends to zero for every condition; individual attainment alone does not imply a strictly positive gap. □

Proposition 4 (Decoded error amplification)**.** *Let O^+ be a target observer state in the domain of E, with physical readout $r(O^+)$. Let H^{-1} be κ-Lipschitz and D be L_D-Lipschitz on a domain containing both $E(O^+)$ and $H^{-1}(\widehat{w})$ almost surely. Set $\epsilon_{\mathrm{rec}} = \mathbb{E}\|D(E(O^+)) - r(O^+)\|^2$. For finite expected squared errors, Eq. (26) holds for every $\eta > 0$.*

Proof. Let $v = H(E(O^+))$. Add and subtract the frozen decoder's base readout:

$$DH^{-1}(\widehat{w}) - r(O^+) = \underbrace{DH^{-1}(\widehat{w}) - DH^{-1}(v)}_{a} + \underbrace{D(E(O^+)) - r(O^+)}_{b}.$$

Composition of the two Lipschitz bounds gives $\|a\| \leq L_D\kappa\|\widehat{w} - v\|$. Young's inequality, $2\langle a, b\rangle \leq \eta\|a\|^2 + \eta^{-1}\|b\|^2$, therefore yields

$$\|a + b\|^2 \leq (1 + \eta)L_D^2\kappa^2\|\widehat{w} - v\|^2 + (1 + \eta^{-1})\|D(E(O^+)) - r(O^+)\|^2.$$

Taking expectations proves the claim. The second term is the base readout error (reconstruction error when r is identity); inverse compensation does not make it zero. □

Proposition 5 (Ridge metric under feature reparameterization)**.** *If $\phi_H = T\phi_I$ for invertible T, unregularized linear prediction classes agree, while isotropic ridge induces the original-basis penalty $\lambda\alpha^\top T^{-1}T^{-\top}\alpha$. Complete polynomial dictionaries are span-invariant under invertible affine state changes; generic nonlinear gauges need not preserve their finite span.*

Proof of Proposition 5. Write $\alpha = T^\top\beta$. Then $\beta^\top T\phi_I = \alpha^\top\phi_I$ and $\beta = T^{-\top}\alpha$, giving $\|\beta\|^2 = \alpha^\top T^{-1}T^{-\top}\alpha$. Closure of all monomials up to degree p under affine substitution yields an invertible feature-basis map when the affine state map is invertible. General nonlinear substitution need not remain in the finite polynomial span. □

Corollary 3 (Topology and persistence stability)**.** *If H is a homeomorphism on latent support Z, then $H_q(Z) \cong H_q(H(Z))$. If it is also κ-bi-Lipschitz, their Vietoris–Rips persistence modules are multiplicatively κ-interleaved; on log filtration scale the shift is at most $\log\kappa$.*

Proof of Corollary 3. Homeomorphisms induce homology isomorphisms. A simplex in $\mathrm{VR}_r(Z)$ maps to one in $\mathrm{VR}_{\kappa r}(H(Z))$; applying the same argument to H^{-1} yields multiplicative interleaving. The log filtration converts it to an additive $\log\kappa$ shift (Chazal et al., 2014). □

Exact conjugacy is not fitted stability. If $\|A^t\|_2 \leq C\alpha^t$ for $\alpha < 1$, $H(0) = 0$, and H is κ-bi-Lipschitz, then the exact conjugate $F_H = H \circ A \circ H^{-1}$ obeys $\|F_H^t(w)\|_2 \leq \kappa^2 C\alpha^t\|w\|_2$. A fitted $\widehat{F}_H$ need not equal this conjugate and inherits no such guarantee. This separates stable-Koopman parameterization from coordinate-map certification.

B Experimental Configuration and Estimation

Study roles and risk estimators are collected in Appendix H. Technical seeds, shared trajectories, independent charts and continuous measured records are distinct units; risks from different interfaces and metrics are not pooled.

B.1 Receiving Learner Configuration and Fairness

For each of the six coordinates and each seed in $\{1943, 2943, 3943\}$, the receiving head is tuned independently on the same six support and three tuning trajectories. Selection minimizes equal-trajectory decoded physical NMSE. At most 1,024 evenly spaced pairs are used per historical trajectory; after selection, the head is refitted on their nine-trajectory union. Fresh trajectories participate only in evaluation, using every valid pair.

Q2 uses the complete degree-two dictionary with an intercept, covariance-aware ridge and relative covariance floor 10^{-8}. The seven ridge candidates are $10^{-12}, 10^{-10}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-4}, 10^{-2}$. RBF concatenates the normalized six-state input with 256 Gaussian radial features. Anchors are sampled without replacement from normalized support states using seed $s + 200003$. The base length scale is the square root of the median squared pairwise anchor distance; the bandwidth multipliers are $.25, .5, 1, 2, 4$. All 35 bandwidth–ridge combinations are considered independently in every coordinate. RBF-512 and RBF-1024 identity references use the same full grid. The feature-standardized RBF regularizer differs from Q2's covariance-aware regularizer. State normalization is fitted within each receiving coordinate.

The residual MLP uses two width-128 SiLU hidden layers, batch size 256, weight decay 10^{-5} and 12,800 updates. Its initial learning rate is chosen from $.0003, .001, .003$; a cosine schedule ends at $.01$ times that rate. Hyperparameter profiles and final refits use identical update budgets.

Within each receiving family, all coordinates receive the same data and search grid. Across families, model sizes, objectives and search costs differ. Historical Q2 and RBF gauge training used 96 outer updates; the neural gauge pilot used six. RBF gauge training did not search the entire final bandwidth grid at every outer update. The final RBF head nevertheless receives the full independent grid in all tested coordinates. The random control matches gauge architecture and certificate cap, not achieved displacement or training cost. A structurally zero neural gauge is aliased to its literal identity candidate before fresh evaluation. These declarations accompany all results, including PCA and the larger identity receivers.

B.2 Original Controlled Protocol: Formulation

The following formulation and method describe the original controlled studies. Later crossed and budget experiments declare their differences in the main text.

Let $X_t \in \mathcal{M} \subset \mathbb{R}^D$ be a physical Markov state, $Y_t = P(X_t)$ the legal measurement, and $O(Y_{\le t})$ a declared observer. Let U_c be the target under condition

$$c = (\text{task}, h, \mathcal{G}, A_B, n, \text{regime}), \tag{15}$$

where $\mathcal{G}$ is the latent hypothesis class and A_B is its training algorithm with budget B. An encoder–decoder (E, D) is admitted before adaptation if it reconstructs independent physical states and retains the information needed for the declared task under a matched flexible probe. These are empirical eligibility tests, not claimed global manifold certificates.

Write $Z = E(O(Y_{\le t}))$. For an invertible gauge H, the decoded finite class is

$$\mathcal{F}_{c,H} = \{DH^{-1}gHE : g \in \mathcal{G}\}. \tag{16}$$

Its infimal risk describes approximation within a fixed class. To include estimation and optimization, define the transformed training sample S_H. For state-transition tasks, $S_H = \{(H(z_i), H(z_i^+))\}_{i=1}^n$. The randomized budgeted algorithm returns $A_{B,\xi}(S_H) = \widehat{g}_{H,S,\xi} \in \mathcal{G}$. We then define

$$\mathcal{R}_{A,n,B}(H) = \mathbb{E}_{S,\xi}\mathbb{E}\,\ell_c\big(DH^{-1}\widehat{g}_{H,S,\xi}(HE(O)), U_c\big)\,. \tag{17}$$

Equation (17) is the object calibrated in the dynamics experiments. The scalar supervised extension instead predicts $\widehat{f}(H(E(s)))$, without decoder compensation, as defined in Appendix L.

Proposition 6 (Exact-orbit invariance)**.** *If H is a bimeasurable bijection on the support of Z, then $(H \circ E, D \circ H^{-1})$ has exactly the same pointwise reconstruction as (E, D) and $\sigma(H(Z)) = \sigma(Z)$. Hence unrestricted Bayes risk is unchanged for every square-integrable target.*

Proof of Proposition 6. For every admitted state x, $(D \circ H^{-1})((H \circ E)(x)) = D(E(x))$ by cancellation. Measurability of H gives $\sigma(H(Z)) \subseteq \sigma(Z)$; writing $Z = H^{-1}(H(Z))$ and using measurability of the inverse

gives the reverse inclusion. For any square-integrable target U, the squared-loss Bayes predictor is the conditional expectation. Thus

$$\mathbb{E}[U \mid \sigma(H(Z))] = \mathbb{E}[U \mid \sigma(Z)] \quad \text{almost surely},$$

and its risk is unchanged. More generally, the measurable decision rules available from the two representations coincide. Reconstruction preservation is relative to $DE(x)$, not an assertion that $DE(x) = x$. □

Theorem 2 (Finite computations can distinguish exact coordinates)**.** *If $\mathcal{F}_{c,H} = \mathcal{F}_{c,I}$, their infimal risks agree for every loss. The sufficient closure condition $H^{-1}\mathcal{G}H = \mathcal{G}$ implies this equality. For squared loss, if the closures of the two decoded classes differ in $L^2(P_O)$, there exists a square-integrable target for which their infimal risks differ. Nonclosure permits, but does not guarantee, improvement.*

Proof of Theorem 2. Class equality makes the two infima identical. If $H^{-1}\mathcal{G}H = \mathcal{G}$, substitution gives decoded-class equality. Conversely, choose a target in one L^2 closure at positive distance from the other. Its approximation risk is zero for the first class and positive for the second. Failure of conjugacy closure without decoded-closure separation is insufficient by itself. □

Exact gauges are computation-equivalent when their complete decoded predictors agree for every admitted training sample and profiled hyperparameter. For the ideal covariance-aware complete-polynomial learner, Proposition 1 places invertible affine gauges in the identity class, whereas a generic nonlinear polynomial pullback leaves a finite degree-p dictionary. Auxiliary class and nonclosure statements are in Appendix A.4.

Let $A_{B,\lambda}$ denote a declared learner indexed by $\lambda \in \Lambda$. Define $R_\lambda(H) = \mathcal{R}_{A_\lambda,n,B}(H)$, $R_\Lambda(H) = \inf_{\lambda\in\Lambda} R_\lambda(H)$, and tuning regret $T_{\lambda_0}(H) = R_{\lambda_0}(H) - R_\Lambda(H)$.

Proposition 7 (Fixed versus profiled coordinate gain)**.** *For $G_{\text{fixed}} = R_{\lambda_0}(I) - R_{\lambda_0}(H)$ and $G_{\text{prof}} = R_\Lambda(I) - R_\Lambda(H)$,*

$$G_{\text{fixed}} = G_{\text{prof}} + T_{\lambda_0}(I) - T_{\lambda_0}(H). \tag{18}$$

Thus a fixed-algorithm gain combines residual profiled coordinate value with a coordinate-dependent change in hyperparameter mismatch. The equality is on absolute risk; subtracting relative percentages is not an exact decomposition.

Proof of Proposition 7. By definition, $R_{\lambda_0}(Q) = R_\Lambda(Q) + T_{\lambda_0}(Q)$ for $Q \in \{I, H\}$. Subtract these two identities:

$$R_{\lambda_0}(I) - R_{\lambda_0}(H) = \big[R_\Lambda(I) - R_\Lambda(H)\big] + T_{\lambda_0}(I) - T_{\lambda_0}(H).$$

Identifying the bracketed term as G_{prof} proves the result. This is an identity for finite absolute risks, not an assumption about how validation approximates their infima. □

For a direct h-step task, write $G_{H,h} = H \circ F^h \circ H^{-1}$ and fix a partition $\{K_j\}_{j=1}^m$ on identity support before transporting it as $H(K_j)$. Let V_p^d be global vector-valued degree-p polynomials and $W_{p,m}^d$ their piecewise extension; $V_p^d \subset W_{p,m}^d \subset L^2(\mu_H; \mathbb{R}^d)$, where μ_H is the transported support measure and all norms below are in this space. Assume the partition is measurable and $G_{H,h} \in L^2(\mu_H; \mathbb{R}^d)$; the dictionary functions must also be square-integrable as required by the stated inclusions. These finite-dimensional subspaces are then closed, so both orthogonal projections are well-defined.

Proposition 8 (Finite-learner local–coherence–fit decomposition)**.** *For orthogonal projections P_V, P_W and any fitted $\widehat{G}_H \in V_p^d$,*

$$\begin{aligned}\|G_{H,h} - \widehat{G}_H\|^2 = \|G_{H,h} - P_W G_{H,h}\|^2 + \|P_W G_{H,h} - P_V G_{H,h}\|^2 \\ + \|P_V G_{H,h} - \widehat{G}_H\|^2.\end{aligned} \tag{19}$$

The terms are local approximation, the burden of sharing one global model across patches, and finite fitting/regularization error, respectively.

Proof of Proposition 8. Because $V_p^d \subset W_{p,m}^d$, nested orthogonal projections satisfy $P_V P_W = P_V$. Hence $G - P_W G$ is orthogonal to W, while $P_W G - P_V G$ is orthogonal to V. Since $P_V G - \widehat{G} \in V$, the three summands

$$G - \widehat{G} = (G - P_W G) + (P_W G - P_V G) + (P_V G - \widehat{G})$$

are pairwise orthogonal. Pythagoras gives Eq. (19). Fixing the identity-support partition and transporting sample membership under H is essential: reclustering separately in every coordinate would change both the partition and the coordinate intervention. □

This coordinate-space identity does not assert that any term must decrease or that decoder-weighted physical risk decomposes identically.

For multiple conditions, allowing one gauge per condition cannot have larger profiled risk than forcing a common gauge. Appendix A.4 formalizes the universal-coordinate regret inequality; the frozen horizon-atlas test is in Archive H3 (Table 37).

B.3 Reproducibility Details

Evaluation artifacts retain protocol, code, checkpoint, generated-data and environment hashes; numerical amendment summaries are in Archive H1 (Table 37). The original experiments used the 'ae14' environment. The controlled six-receiver benchmark records its separate Python 3.12.8 / PyTorch 2.5.1 environment in its evidence component. Training and selector roles may fit or choose models; fresh-holdout roles only accept or reject one frozen candidate; report-only roles never alter a checkpoint or threshold. Failed protocols, including the certified v2 omnibus failure and fixed-budget MLP identity fallback, remain in the release.

Original learned-chart construction. The original L3F chart has the statewise convolutional architecture detailed in Appendix C.2, with six latent coordinates and 32 convolution channels. It was constructed on a seed-42 trajectory of 6,000 states, split contiguously at 4,200 and 5,100 states. Length-64 windows were formed within each partition; the network itself never mixes time points. Training used three successive stages, all with batch size 64, seed 42 and AdamW weight decay 10^{-4}. Reconstruction pretraining used learning rate .001, at most 100 epochs and patience 15. BLAE fine-tuning used learning rate .0002, weight .001, at most 45 epochs and patience 10. The final stage used learning rate .00005, BLAE weight .001, RTD-Lite weight .01, at most 20 epochs and patience 8; its retained checkpoint was epoch 18. Forecasting-loss weights were zero throughout these stages. The BLAE penalty uses finite-sample geometry and sampled decoder Jacobians; it is not a global certificate for the base encoder or decoder. The independent-chart study instead trains every AE from random initialization with reconstruction MSE alone.

Silverbox fitting configuration. The original frozen protocol uses $q = 24$, support $[0, 20000)$, inner $[20500, 28000)$, outer $[28500, 39037)$, Checkpoint $[39037, 52049)$ and Selector $[52049, 65062)$ (zero-based, half-open, no boundary-crossing pairs). Complete degree-two features include every state–control cross term. The ridge grid is $\{10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}\}$, implemented by feature whitening with relative covariance floor 10^{-8}. With centered feature covariance C, let $\epsilon = 10^{-8} \max\{\mathrm{tr}(C)/d_\phi, 10^{-12}\}$ and $C + \epsilon I = U \,\mathrm{diag}(s_i) U^\top$; use $T = U \,\mathrm{diag}(\max\{s_i, \epsilon\}^{-1/2}) U^\top$ and isotropic ridge on transformed nonconstant features, unpenalized intercept. Support pairs are capped at 8,192 deterministic rows. Three retained seeds 1943/2943/3943 train four width-64 couplings for 60 AdamW updates, rate 3×10^{-4}, decay 10^{-6}, clipping one, inner/outer horizon 64. Checkpoint uses sixteen 256-step starts and retains epoch zero. Selector heads refit Development+Checkpoint and score thirty-two 256-step windows. Official heads keep selected ridge and refit the union of Development, Checkpoint and Selector, using at most 8,192 pairs from each role. Multisine $[105712, 127400)$ and arrow $[100, 40575)$ initialize from 50 measured outputs, then free-run without future observed outputs; only multisine is the primary official endpoint. The corresponding frozen development and official protocols are included in the companion archive; their pre-evaluation statements do not revise the later observed rollout failures.

Measured-system evaluation roles. Silverbox's selector source and official evaluator were frozen before official-test access. Wiener–Hammerstein used a separately frozen PCEGA-only protocol after a broader train-derived authorization stopped. The broader stop remains in the evidence ledger; it is not counted as a successful full-system protocol. The tables below state the retained gates. Source amendments, timestamps and one-time evaluation manifests are retained in the accompanying record rather than treated as additional scientific evidence.

Table 6: Silverbox's separately frozen selector and official multisine gates. The 32 selector blocks and eight official segments are correlated diagnostics, not independent statistical units. All Δ values use the same error-change convention: negative improves and positive worsens identity.

Gate	Frozen threshold	Selector	Official multisine
Learned RMSE Δ	$\leq -3\%$	-42.938%	-46.375%
Improving seeds	$\geq 2/3$	$3/3$	$3/3$
Finite / NMAE	all / no regression	yes / 44.707%	yes / 47.755%
Certificate	$\kappa \leq 3$	2.995	2.995
Round-trip MSE	$\leq 10^{-10}$	3.72–3.91×10^{-18}	same gauges
Random RMSE Δ	learned margin ≥ 1 pp	$+34.899\%$	$+25.799\%$
Arrow role	secondary diagnostic	sealed	instability boundary

Table 7: Wiener–Hammerstein PCEGA-only official audit. The eight continuous segments are descriptive, not independent units. Negative Δ improves identity.

Gate	Frozen threshold	Train-derived selector	Official free run
Learned error Δ	$\leq -0.5\%$	NMAE -8.957%	RMSE -10.466%
Improving seeds/cells	$3/3$; selector $\geq 80\%$	$3/3$; $24/24$	$3/3$
Finite / NMAE	all / no regression	yes / yes	yes / -9.376%
Certificate	$\kappa \leq 3$	2.995	2.995
Round-trip MSE	$\leq 10^{-10}$	$\leq 1.83 \times 10^{-16}$	same gauges
Random RMSE margin	learned ≥ 1 pp better	> 1 pp	9.866 pp
Official access	sealed until P-only lock	unopened	78,750 scored samples

The three official PCEGA seed RMSEs are .140649, .140602, and .140021, corresponding to -10.323%, -10.353%, and -10.723% relative changes from identity. The three random-gauge RMSEs are .155149, .161233, and .151314 (-1.078%, $+2.801\%$, and -3.523%). The PCEGA family improves all eight descriptive segments, with segment effects from -11.816% to -8.983%; these segments are correlated and are not used as eight independent replications.

The final shift network contains three tanh-separated linear maps per conditioner. Effective weights are projected by spectral norm on every use; four alternating additive couplings and a scalar conjugated normalizer add no uncounted anisotropic scaling. Effective checkpoint weights are independently re-audited rather than trusting metadata. Every final run exposes epoch zero and 96 outer epochs to its internal checkpoint rule. Three optimization seeds are retained as a family; no best-seed selection is allowed.

L3F and Rössler trajectories each contain 6000 post-burn-in steps. All state, field, coordinate, and feature statistics are fit from the appropriate support role only. Final L3F training uses six gauge-development and three disjoint checkpoint trajectories; ten selector trajectories authorize fresh seeds 17901–17910. The primary unit is a fresh trajectory: gauge seeds are averaged within trajectory before a 10,000-replicate trajectory bootstrap. Optimization used three GPUs at most, respecting the project-wide five-GPU cap; sealed evaluation used one.

C Component Construction and Search

C.1 Original Layer-Budget Adaptation Method

This section specifies LB-PCEGA and its original study protocol. Its numerical grid and validation rule belong to that protocol, not to the later complete-process comparison. Algorithm 1 specifies the GB workflow; Appendix C.5 distinguishes LB, GB, EC-end and EC-best checkpoint policies.

C.1.1 Globally certified exact gauges

The additive coupling and its forward/inverse bound are stated and proved in Appendix A.2, Eqs. (12)–(13).

Each shift is a three-layer tanh MLP. Every matrix is projected to a declared spectral-norm cap on every forward pass, so the product of layer caps bounds L globally. For m alternating couplings, we choose per-layer budgets so that

$$\prod_{j=1}^{m} \rho(L_j) \leq \kappa_{\text{build}} < \kappa_{\text{public}}. \tag{20}$$

The original layer-budget study uses $\kappa_{\text{build}} = 2.995$ and publicly certifies $\kappa \leq 3$. The original L3F checkpoints have maximum weight-derived bound 1.508; this is not a maximum across all systems. The original measured branches in Table 16 report 2.995. An earlier fixed-ridge precursor uses the tighter public bound 1.25. A scalar train-only normalizer is conjugated inside the flow and therefore does not alter its Euclidean bi-Lipschitz constant. The final shift matrices are zero initialized, making epoch zero exactly identity.

An optional orthogonal-mixing refinement is 1-Lipschitz in both directions and spends no certificate budget; its architecture and separately frozen ablation remain in Archive H3 (Table 37).

C.1.2 Cross-fitted adaptation and identity fallback

The base chart is frozen. For each candidate H, disposable learners are fit on support trajectories, λ is selected on an inner disjoint trajectory, and the gauge is scored after inverse transformation and physical decoding on an

Algorithm 2 Original layer-budget search and its study-specific acceptance protocol.

Input: Admitted frozen (E, D); learner A_B; separated $S_{\mathrm{sup}}, S_{\mathrm{in}}, S_{\mathrm{out}}, S_{\mathrm{sel}}$, public κ, and a sealed fresh protocol $\mathcal{P}_{\mathrm{fresh}}$.
Output: Accepted coordinate map or identity under the study-specific rule.
1: Initialize certified H at identity.
2: Fit disposable heads on S_{sup} for every $\lambda \in \Lambda$.
3: Choose $\lambda^*(H)$ on S_{in}; refit on support+inner.
4: Update only H using decoded physical risk on S_{out} while enforcing the certificate.
5: Select a feasible checkpoint using development roles only.
6: Accept it on S_{sel} under the frozen effect rule, else return I.
7: If authorized, lock the evaluator and open $\mathcal{P}_{\mathrm{fresh}}$ once; accept the frozen candidate or return I. No runner-up may be promoted.

outer disjoint trajectory:

$$\begin{aligned} &\lambda^*(H) = \arg\min_{\lambda \in \Lambda} \widehat{L}_{\mathrm{in}}(H, \lambda), \\ &\min_{H \in \mathcal{H}_\kappa} \widehat{L}_{\mathrm{out}}\big(DH^{-1} A_{B,\lambda^*(H)}(S_{\mathrm{sup}\cup\mathrm{in},H}) HE(O), U_c\big)\,. \end{aligned} \tag{21}$$

No reconstruction, PHATE, or persistent-homology term appears in Eq. (21). A nonidentity checkpoint is accepted only under a predeclared effect-size rule on independent statistical units; otherwise the output is the exact identity pair. A one-time fresh-holdout role may accept or reject one frozen checkpoint but cannot promote a runner-up.

For reproducibility, the declared learner is an EDMD-style direct h-step degree-2 polynomial state-transition head: it regresses target state coordinates on a complete source dictionary and recomputes features from predicted states, rather than advancing the full observable vector by one Koopman matrix. Support coordinates are centered and scaled coordinatewise (scale floor 10^{-4}), and ϕ contains all degree-1 and unique degree-2 monomials. If Φ is the resulting support feature matrix, c its support-only mean, and Y the normalized target coordinates, we set

$$C_\Phi = \frac{(\Phi - c)^\top (\Phi - c)}{n} + \epsilon I, \qquad \epsilon = 10^{-8}\ \mathrm{tr}(C_{\Phi,0})/d_\phi, \tag{22}$$

where $C_{\Phi,0}$ denotes the covariance before flooring. With $X = [\Phi - c, \mathbf{1}]$ and $P = \mathrm{diag}(C_\Phi, 0)$, the disposable head is

$$\widehat{W}_\lambda = \Big(X^\top X/n + \lambda P\Big)^{-1} X^\top Y/n, \qquad \lambda \in \{10^{-7}, \ldots, 10^{-1}\}. \tag{23}$$

The intercept is unpenalized. For each fold, inner losses are detached only for the discrete $\arg\min_\lambda$; the selected head is refit on support+inner data, and the outer decoded loss differentiates through feature statistics (except the numerical floor) and the linear solve into H. No selector or fresh trajectory participates in this optimization.

C.1.3 Controlled-system specialization

For a measured single-input system, we keep the known control outside the gauge and form a delay state

$$s_t = (y_t, y_{t-1}, \ldots, y_{t-q+1}), \quad w_t = H(s_t). \tag{24}$$

Let μ_w and the positive diagonal matrix S_w be the fitting-support state mean and scale, and let $x_t = [\bar{w}_t^\top, \bar{u}_{t+1}]^\top$ concatenate $\bar{w}_t = S_w^{-1}(w_t - \mu_w)$ with the support-normalized control. Write c_ϕ for the support mean of the complete degree-2 state–control dictionary. Using the centered-feature coefficients $\widehat{W}_\lambda \in \mathbb{R}^{(d_\phi+1)\times q}$ from Eq. (23), column-vector prediction is

$$\begin{aligned} \psi_t &= \begin{bmatrix} \phi_2(x_t) - c_\phi \\ 1 \end{bmatrix}, \quad \widehat{\bar{w}}_{t+1} = \widehat{W}_\lambda^\top \psi_t, \\ \widehat{w}_{t+1} &= \mu_w + S_w \widehat{\bar{w}}_{t+1}, \qquad \widehat{y}_{t+1} = \big[H^{-1}(\widehat{w}_{t+1})\big]_1\,. \end{aligned} \tag{25}$$

All preprocessing statistics are fitted on the head's fitting support only and frozen for evaluation. If the implementation uses a feature transform T (including whitening) and weights $\widetilde{W}_\lambda$, its design vector is $[T^\top(\phi_2(x_t) - c_\phi); 1]$; equivalently, $\widehat{W}_\lambda = \mathrm{diag}(T, 1)\widetilde{W}_\lambda$ in Eq. (25). The inverse-updated state is rolled out without measured outputs after initialization. The known u_{t+1} drives the $t \to t+1$ transition and is not gauged. Silverbox uses $q = 24$, 64-step nested training, 256-step selection, and official free simulation after 50 measured initialization outputs; full dictionary and ridge-grid details are in Appendix B.3.

C.1.4 DISTORTION CONTROL AND RIDGE MECHANISM

For the general readout interface $Y = r(O^+)$, define $\epsilon_{\text{rec}} = \mathbb{E}\|D(E(O^+)) - r(O^+)\|^2$. Under the Lipschitz-domain and finite-error conditions of Proposition 4, for every $\eta > 0$,

$$\mathbb{E}\|DH^{-1}(\widehat{w}) - r(O^+)\|^2 \le (1+\eta)L_D^2\kappa^2\mathbb{E}\|\widehat{w} - H(E(O^+))\|^2 + (1+1/\eta)\epsilon_{\text{rec}}. \tag{26}$$

Thus the certificate bounds coordinate-induced decoded amplification but not recursive dynamics stability. A feature reparameterization $\phi_H = T\phi_I$ also changes isotropic ridge into the original-basis metric $\lambda T^{-1}T^{-\top}$; profiling and covariance-aware regularization are therefore required before attributing gains to nonlinear coordinates. Reported ridge-bias and effective-dimension diagnostics are in Archive H3 (Table 37).

Proposition 9 (Task-horizon fitted-residual transport)**.** *Let F_t^H and $\widehat{F}_t^H$ be continuously differentiable true and fitted gauged one-step maps on a convex region containing the compared states, under the same input, and define $r_t(w) = F_t^H(w) - \widehat{F}_t^H(w)$ and $e_t = w_t - \widehat{w}_t$. Then*

$$e_{t+1} = A_t e_t + r_t(\widehat{w}_t), \qquad A_t = \int_0^1 JF_t^H(\widehat{w}_t + se_t)\,ds. \tag{27}$$

For $e_0 = 0$, $e_h = \sum_{j=0}^{h-1} A_{h-1}\cdots A_{j+1}r_j(\widehat{w}_j)$, with an empty product equal to identity.

Proof of Proposition 9. Add and subtract $F_t^H(\widehat{w}_t)$ in $F_t^H(w_t) - \widehat{F}_t^H(\widehat{w}_t)$. The integral mean-value identity along the segment from $\widehat{w}_t$ to w_t gives $F_t^H(w_t) - F_t^H(\widehat{w}_t) = A_t e_t$, proving Eq. (27). For $e_0 = 0$, the first two errors are $e_1 = r_0(\widehat{w}_0)$ and $e_2 = A_1 r_0(\widehat{w}_0) + r_1(\widehat{w}_1)$. Inductively, multiplying the expansion of e_t by A_t and adding $r_t(\widehat{w}_t)$ gives the stated ordered sum for e_{t+1}. For two fixed, time-homogeneous linear operators on a common invariant observable space, with all compositions defined, the analogous identity is

$$\mathcal{U}_H^h - \widehat{\mathcal{U}}_H^h = \sum_{j=0}^{h-1} \mathcal{U}_H^{h-1-j}(\mathcal{U}_H - \widehat{\mathcal{U}}_H)\widehat{\mathcal{U}}_H^j.$$

A finite Galerkin matrix must first be represented on that common space, for example by the extension $\widehat{\mathcal{U}}_H = P\mathcal{U}_H P$ with a specified projection P, before forming the difference. For time-varying controlled maps one instead needs ordered operator products, or an autonomous augmentation of the control-generating state; the fixed operator powers above do not apply directly. No commutativity is required. For an ideal Galerkin learner this separates finite-dictionary closure from transport; for the fitted learner, the defect also contains finite-data and regularization error (Nüske et al., 2023; Conradie et al., 2026). □

Thus task risk depends on residual magnitude, direction, visited state, and subsequent Jacobian products, not one-step error alone. The identity is exact for a deterministic admitted Markov state; restarted diagnostics on noisy measured delay states are mechanism evidence, not proof of that assumption.

The ideal affine-equivariance and controlled-state results, with proofs, are collected in Appendix A.3.

The coordinate-local production floor need not preserve this equality for nonorthogonal bases. In the original learned-chart experiment, a post-confirmatory learned-affine control, trained with the same three seeds, budget, profiling, and public certificate, changes selector and opened fresh NMAE by less than 3×10^{-7} in relative units. It therefore tests the finite-precision loophole rather than serving as a new method branch.

Exact invertibility preserves latent homology, while the bi-Lipschitz budget controls Vietoris–Rips persistence distortion; the formal corollary is in Appendix A.4. Topology is a chart-admission issue, not a gauge training loss.

C.2 INDEPENDENT CHART AND GLOBAL-BUDGET STUDY

Three independently reconstruction-trained six-dimensional charts per system use seeds 142/242/342; no chart is filtered by forecasting performance. The physical field, historical normalization and simulator are fixed. Lorenz–63 uses $(10, 28, 8/3)$, step .01 and 1,000 burn-in states; Rössler uses $(.2, .2, 5.7)$, step .02 and 3,000 burn-in states. Each trajectory retains 6,000 states. For standardized simulator state (a, b, c) and $s = j/100$, $j = 0, \ldots, 99$, the 100D field before fixed training normalization is

$$\begin{aligned} v(s) = {} & .26a\sin(2\pi s) + .24b\cos(2\pi s) + .22c\sin(4\pi s) \\ & + (.72 + .18\tanh a)\sin(2\pi s + .75\tanh b) + .38\tanh(.7b - .25a)\cos(4\pi s + .55\tanh c) \\ & + .52\exp\left[-\frac{(s - .5 - .18\tanh(.85a - .35c))^2}{2(.075 + .025\,\text{sigmoid}(b))^2}\right] + .16\sin(6\pi s + .30ac). \end{aligned}$$

Table 8: Complete deployment on independently trained charts. Counts are technical runs across three charts, not six independent representations. The public cap is 3 for all arms; rejected candidates deploy identity. Gains are ratios of mean risks.

System	Search	NMSE	NMAE	NMSE gain (%)	Certified	Accepted
L3F	Identity	0.007337	0.074329	0.00	—	—
	LB	0.006622	0.070902	9.75	6/6	6/6
	GB	0.003756	0.052491	48.82	6/6	6/6
	EC	0.004672	0.058375	36.32	4/6	4/6
Rössler	Identity	0.001931	0.033665	0.00	—	—
	LB	0.001761	0.031868	8.81	6/6	6/6
	GB	0.001060	0.023353	45.11	6/6	6/6
	EC	0.001724	0.031372	10.70	1/6	1/6

The statewise CNN has a circular width-5 stem (1 to 32 channels), three width-3 residual convolution/one-group GroupNorm/GELU blocks, each followed by ceiling average pooling by two, and a 32×13 to 6 projection. The decoder reverses this using linear interpolation to 26/52/100 and a width-5 output convolution; dropout is zero. Each AE uses six training and three checkpoint trajectories, reconstruction MSE only, AdamW 10^{-3}, weight decay 10^{-5}, batch 512, clipping 5, at most 200 epochs, patience 25 and improvement threshold 10^{-9}.

Every frozen chart then receives separate six/three downstream Development/Checkpoint trajectories. LB, GB and uncapped endpoint-certified (EC) paths use four width-64 couplings, 96 updates and seeds 1943/2943: 36 prescribed runs. All initialize the identity function, but LB and EC do not match effective hidden weights; this compares search implementations. The matched numerical configuration below supplies folds, ridge and optimizer. GB scales only output matrices by a common $s \in [0, 1]$ until $\prod_j \rho(sL_j) \leq 2.995$, using 50 bisection steps and a relative 10^{-6} safety margin. EC checks its selected endpoint, without rescuing an earlier certified checkpoint. Ties choose the earliest minimum, including update zero. Final heads independently tune ridge on Development/Checkpoint and refit their union.

Ten new Selector trajectories per chart follow frozen heads. Acceptance requires a nonidentity candidate, finite bound ≤ 3, round-trip MSE $\leq 10^{-10}$, at least 1% mean NMSE improvement and eight wins; otherwise identity is deployed. Twenty new Report trajectories per system follow all decisions and are shared across charts/seeds. Direct $h = 12$ evaluation uses 1,024 fixed pairs per trajectory. For P targets u_{ij} and temporal channel means $\bar{u}_j$, NMSE divides squared error by $\max\{\sum_{ij}(u_{ij} - \bar{u}_j)^2, 100P10^{-8}\}$; NMAE uses absolute error and $\max\{\sum_{ij} |u_{ij} - \bar{u}_j|, 100P10^{-8}\}$. Effects are ratios of mean risks, including all fallbacks.

All 36 runs completed: LB/GB each accept 12/12, EC 5/12; all seven EC rejections exceed the cap. Technical risks are averaged first, then a 20,000-draw crossed bootstrap resamples three charts and twenty shared trajectories. EC–identity and EC–LB are the four primary system-specific contrasts; adjusted one-sided lower bounds use quantile .0125. GB–EC is estimation-only. Three charts limit training-population inference; zero primary lower bounds do not demonstrate strict superiority.

Table 9: Prespecified primary NMSE contrasts. Positive favors EC. Intervals are pointwise; lower bounds use quantile 0.0125 for four tests.

System	Contrast	Difference	95% interval	Adjusted lower
L3F	Identity−EC	0.00266515	$[0, 0.00532894]$	0
	LB−EC	0.00194957	$[-0.00030018, 0.00394127]$	-0.00030779
Rössler	Identity−EC	0.00020659	$[0, 0.00061446]$	0
	LB−EC	0.00003647	$[-0.00015609, 0.00037572]$	-0.00015849

The fixed perturbation audit uses the first Report trajectory, 128 receiver outputs, eight unit directions and norms $.001, .01, .1$. Sampled inverse amplification/certificate ranges are .8042–.8906 (LB), .4486–.5365 (GB), and .4571–approximately 1 (EC with fallbacks); a float32 fallback ratio 1.000239 is a recorded rounding discrepancy. Physical amplification spans 2.12–3.54 and NMSE changes -3.48×10^{-7} to 4.92×10^{-4}. Sample maxima are not global worst-case estimates or rollout guarantees.

Table 10: Complete-training sensitivity at 192 updates. NMSE includes identity fallbacks, three frozen charts, two seeds and twenty Report trajectories per system. Original denotes normal equations with stopped floor; QR/full differentiates the floor within the selected ridge branch.

	L3F NMSE		Rössler NMSE	
Policy	Original	QR/full	Original	QR/full
GB	0.003941957	0.003941887	0.000979633	0.000979621
LB	0.005340695	0.005341478	0.001258204	0.001258391
EC-best	0.005378553	0.005378569	0.001279055	0.001279069
EC-end	0.007898960	0.007898960	0.001913905	0.001913905

C.3 Numerical fitting and complete-training sensitivity

The original implementation solves covariance-aware normal equations and stops the gradient through the feature-dependent floor. Complete Q2 uses an unpenalized intercept, source-scale floor 10^{-4} and relative covariance floor 10^{-8}. With centered features F, $C = F^\top F/n$, $\epsilon = 10^{-8}\max\{\mathrm{tr}(C)/d, 10^{-12}\}$, $X = [F, \mathbf{1}]$, and $LL^\top = C + \epsilon I$, augmented QR solves

$$A = \begin{bmatrix} X/\sqrt{n} \\ \sqrt{\lambda}[L^\top, 0] \end{bmatrix}, \quad B = \begin{bmatrix} Y/\sqrt{n} \\ 0 \end{bmatrix}, \quad A = QR, \quad W = R^{-1}Q^\top B.$$

It preserves the forward ridge objective; full-floor differentiation changes the backward convention, not the discrete ridge-selection rule. Normal equations require an invertible regularized system. Nonfinite fits or failed training return identity before Report; nonfinite Report output remains a failure, not grounds for reselection or omission.

The matched configuration uses six Development trajectories in six rotating 4-support/1-inner/1-outer folds, at most 768 pairs per trajectory, and the seven-point ridge grid in Appendix B.1. AdamW uses 2×10^{-4}, weight decay 10^{-5}, clipping at five and mean fold loss divided by its initial value (floor 10^{-10}). Every eight updates, three Checkpoint folds use Development as support, one Checkpoint trajectory as inner and the next as outer. Final heads tune on Development/Checkpoint, then refit their union; evaluation uses at most 1,024 pairs per trajectory.

A post-confirmatory solver-by-floor-derivative study on the six matched charts adds 60 paths and 144 selected outputs at 96/192 updates, retaining all fallbacks. All 2,880 new Report rows are finite and acceptance counts match the original. The largest mean-NMSE change is 7.825×10^{-7} (7.000×10^{-8} among GB conditions). At 96 updates GB and both EC policies remain tied, ahead of LB. This reused-data sensitivity analysis is neither independent efficacy evidence nor an equivalence test. The separate 72-head frozen QR/SVD check changes relative NMAE by at most 1.08×10^{-5}; small-sample high-degree failures are documented in Archive H1 (Table 37).

C.4 A simpler training recipe with an unchanged component

The recipe chooses one ridge at identity by mean inner-fold risk, holds it fixed during gauge updates, and retains independent full-grid Checkpoint and final-head tuning. Architecture, solver, stopped floor, weight decay, clipping and GB retraction are unchanged. Its 384-update schedule is

$$\eta_t = \begin{cases} 2 \cdot 10^{-4} + 4 \cdot 10^{-4} t/24, & t \le 24, \\ 2 \cdot 10^{-5} + (5.8 \cdot 10^{-4}/2)[1 + \cos(\pi(t-24)/360)], & t > 24. \end{cases}$$

The 192-update candidate is a prefix, not a separate schedule. Comparisons match updates, not wall time or regression solves. The initial chart-zero, seed-1943 pilot retains all ten candidates, then ten new Selector and twenty new Report trajectories per system. All pass selection. Constant-rate fixed ridge nearly matches profiling; without a profiled-plus-schedule arm, that pilot does not separate ridge and schedule effects. Scheduled-192 wins 20/20 paired trajectories per system; adjusted lower NMSE-difference bounds are .000622555/.000165615, conditional on fixed candidates. The 384 results are descriptive extra-budget comparisons.

A subsequent frozen 192-update protocol repeats the full process on one fixed chart per system with three new joint data–initialization seeds 2943/3943/4943. Each repeat has six Development and three Checkpoint trajectories, matched initialization/pairs, and checkpoints every eight updates including identity. Final fitting precedes ten new Selector trajectories per repeat; all decisions precede twenty additional shared Report trajectories per system, with no post-Selector refit. The original certificate/round-trip/utility gate and identity fallback apply. All twelve paths complete and are accepted; every pair improves NMSE and NMAE.

Table 11: Fixed-candidate Report NMSE. Same charts and training seed within each system; every candidate is retained. The scheduled 192-update result uses the first half of the frozen 384-update schedule.

Training rule	Updates	L3F	Rössler
Original profiled, constant rate	192	0.003494057	0.000962327
Fixed ridge, constant rate	192	0.003494112	0.000962300
Fixed ridge, scheduled rate	192	0.002816616	0.000792493
Fixed ridge, constant rate	384	0.002915712	0.000791186
Fixed ridge, scheduled rate	384	0.002494691	0.000742229

Table 12: Complete-process Report NMSE, including the prescribed identity fallback policy. Repetitions use new fitting data and seeds 2943/3943/4943 but one fixed chart per system; their twenty Report trajectories are shared. Accept (O/S) gives original/scheduled acceptance indicators. Mean rows average risks before computing the relative reduction.

System	Repeat	Original NMSE	Scheduled NMSE	Reduction	Accept (O/S)
L3F	1	0.003365	0.002754	18.14%	1/1
L3F	2	0.003394	0.003015	11.19%	1/1
L3F	3	0.003358	0.002943	12.36%	1/1
Rössler	1	0.000986	0.000816	17.23%	1/1
Rössler	2	0.001027	0.000823	19.86%	1/1
Rössler	3	0.000999	0.000806	19.36%	1/1
L3F	Mean	0.003373	0.002904	13.89%	—
Rössler	Mean	0.001004	0.000815	18.83%	—

Mean NMAE changes .049599 to .046544 (L3F), .022602 to .019546 (Rössler). All select update 192, so convergence is not established. Maximum bound and round-trip MSE are 2.994997 and 2.746×10^{-14}. Mean training/final-fit times are 437.8/205.1 seconds (original/scheduled, two concurrent two-thread CPU workers), descriptive rather than isolated timings. These three repeats do not separate initialization from data variance, and the shared Report set does not create sixty independent training repetitions.

Matched factorial follow-up. We cross training-time ridge profiling versus a ridge fixed at identity with constant versus scheduled learning rates (Table 13). Each system uses three frozen charts and one optimization seed (1943), reusing Development/Checkpoint data from Appendix C.5. Within each chart, initialization, 192 updates, eight-update checkpoints, global budget and independent final-head tuning are matched. Ten historical candidates are reused and fourteen missing cells trained. After all heads are frozen, ten new Selector trajectories per chart apply the existing acceptance rule; all decisions precede twenty new Report trajectories per system, shared across charts, without further fitting. All 24 candidates are accepted; the maximum bound is 2.994997 and round-trip MSE is 3.337×10^{-14}. Scheduling improves both ridge strategies on all six charts; system-mean NMSE differs by at most 0.024% between ridge strategies at a matched learning-rate recipe. These are not equivalence tests or multi-seed training-reliability estimates. The schedule remains the first 192 steps of the earlier 384-step recipe: its cumulative learning rate is 2.469 times the constant-rate arm. Thus this comparison separates ridge strategy from the learning-rate recipe, not schedule shape from step-size magnitude.

C.5 Matched initialization and eligible-checkpoint selection

This separate prospective study retains the six reconstruction-only charts but uses new downstream data. All three paths (LB/GB/uncapped) start with identical raw and effective weights, biases and normalizers: bounded weights are materialized with a $1 - 10^{-6}$ interior margin and zero output matrices. Paths use 192 AdamW updates with the configuration above; 96/192 prefixes are fixed budgets. Checkpoints at $0, 8, \ldots, 192$ all remain eligible for the declared selection rule. GB retracts the composed budget, LB uses per-layer caps, and uncapped applies neither restriction.

LB/GB choose the best certified checkpoint. EC-end chooses the best before checking its certificate and returns identity on failure; EC-best chooses the best among certified checkpoints on the same uncapped path. Ties favor earlier updates; eligibility requires bound ≤ 3 and round-trip MSE $\leq 10^{-10}$. Identity is included. Six Development and three Checkpoint trajectories per chart fit and tune final heads independently. All 36 paths and 96 policy–budget outputs are sealed before ten Selector trajectories per chart; the 1%/eight-win gate precedes twenty shared new Report trajectories per system. Every fallback enters the risk.

Table 13: Matched 2×2 follow-up: mean Report NMSE across three charts per system, including the identity-fallback policy. All final heads are independently tuned.

Training ridge	Learning rate	L3F	Rössler
Profiled	Constant	0.003689413	0.001015880
Fixed	Constant	0.003689407	0.001015791
Profiled	Scheduled	0.002609732	0.000784775
Fixed	Scheduled	0.002609793	0.000784590

Table 14: Matched effective initialization and checkpoint policies on new downstream data. Physical NMSE includes identity fallbacks. Accepted counts are out of six chart–seed runs per system, not six independent charts. EC-end and EC-best share uncapped training paths.

System	Policy	NMSE (96)	Accept	NMSE (192)	Accept
L3F	LB	0.007019	6	0.005341	6
L3F	GB	0.005625	6	0.003942	6
L3F	EC-end	0.005625	6	0.007899	0
L3F	EC-best	0.005625	6	0.005379	6
Rössler	LB	0.001721	6	0.001258	6
Rössler	GB	0.001463	6	0.000980	6
Rössler	EC-end	0.001463	6	0.001914	0
Rössler	EC-best	0.001463	6	0.001279	6

The two EC-best versus EC-end contrasts at 192 updates are primary; 20,000 crossed chart/trajectory resamples after seed averaging use .025 one-sided quantiles. All GB and budget contrasts are estimation-only. All paths finish: EC-end selects update 192 with bounds 3.707–8.055 and returns identity in all twelve runs; EC-best selects updates 96–128 and all twelve are accepted, as are LB/GB. The two primary lower bounds are .001344088/.000371561. The table retains all budgets and policies.

GB retracts 62–96 updates; all LB/GB selections are at 192, not evidence of convergence. EC-best retaining an earlier feasible point does not solve global constrained search. Only three charts limit population inference. Search times (three concurrent two-thread CPU workers) are descriptive; per-chart curves, bound ranges and timings are in Archive H2 (Table 37).

D Original Effects and Receiver Alternatives

D.1 Original Controlled and Measured Experiments

This section retains the original study results; it is not a pooled estimate with the later studies.

D.1.1 Evidence protocol

We separate training, checkpoint discovery, selector, one-time fresh-holdout, and report-only roles. Statistics are paired by trajectory rather than overlapping window. Every table states whether it is mechanism evidence, fresh-holdout evidence, or selector-only development. NMAE and NMSE are computed after decoding to physical units unless explicitly labeled a selector objective.

D.1.2 Lifted Lorenz: isolating the coordinate effect

We integrate Lorenz–63 with $\Delta t = .01$ and map its standardized 3D state to a smooth 100D periodic field using injective Fourier components, state-dependent waves, a moving Gaussian bump, and nonlinear phase interaction. The known physical state enables unusually clean causal controls. The admitted d6 chart attains independent reconstruction NMSE 2.04×10^{-5}, kNN@10 .9738, distance Spearman .99955, zero measured collisions, and decoder-condition p95 2.91.

Profiled certified fresh-holdout confirmation. We first discovered that independently tuning scalar or covariance-aware ridge absorbs most of the gain of a gauge trained against fixed ridge. We therefore freeze the profiled objective in Eq. (21): a degree-2 transition head is fit on support trajectories, ridge is selected from $\{10^{-7}, \ldots, 10^{-1}\}$ on an inner trajectory, and decoded physical risk on a disjoint outer trajectory updates the gauge. Rotating trajectory folds prevent one trajectory from serving two roles in the same fold.

Table 15: Matched-study paired NMSE contrasts. Only the two EC-end minus EC-best comparisons at 192 updates are primary; their adjusted one-sided lower bounds equal the displayed lower endpoints. All other intervals are descriptive.

System	Contrast	Difference	Lower	Upper
L3F	EC-end-192 vs EC-best-192 (primary)	0.0025204	0.0013441	0.0046926
L3F	EC-best-192 vs GB-192	0.0014366	0.0010730	0.0017785
L3F	LB-192 vs GB-192	0.0013987	0.0010563	0.0016649
L3F	LB-96 vs LB-192	0.0016782	0.0008707	0.0029999
L3F	GB-96 vs GB-192	0.0016834	0.0010730	0.0021187
L3F	EC-end-96 vs EC-end-192	-0.0022736	-0.0046926	-0.0009698
L3F	EC-best-96 vs EC-best-192	0.0002468	0.0000000	0.0003971
Rössler	EC-end-192 vs EC-best-192 (primary)	0.0006349	0.0003716	0.0010710
Rössler	EC-best-192 vs GB-192	0.0002994	0.0002510	0.0003334
Rössler	LB-192 vs GB-192	0.0002786	0.0002403	0.0003351
Rössler	LB-96 vs LB-192	0.0004632	0.0002712	0.0007861
Rössler	GB-96 vs GB-192	0.0004836	0.0003746	0.0006549
Rössler	EC-end-96 vs EC-end-192	-0.0004507	-0.0006602	-0.0003216
Rössler	EC-best-96 vs EC-best-192	0.0001842	0.0000433	0.0004044

A train/checkpoint-only convergence audit fixes 96 outer epochs, nominal $\kappa = 3$, and three optimization seeds before new selector data are generated. All three seeds are retained. On ten new selector trajectories the frozen family improves NMAE by 13.406% with 10/10 trajectory wins, authorizing a separately hash-locked fresh protocol. Ten fresh trajectories are generated only after the protocol, candidates, evaluator, generator, chart, and core code are locked. Table 16 reports the one-time result.

Table 16: Unified locked evaluation evidence. All Δ values are relative error changes from identity: negative improves and positive worsens. L3F and Rössler effects average three gauges within each trajectory; measured-system rows are means of three seed-level full-record RMSEs, not ensemble predictions.

System (metric)	Identity	PCEGA	Δ	PCA/affine	Random Δ	κ	Fresh unit
L3F (NMAE)	0.097922	**0.084769**	$\mathbf{-13.421\%}$	$0.000\%/\approx 0.000\%$	$+0.212\%$	1.508	10 trajectories
Rössler (NMAE)	0.033879	**0.031762**	$\mathbf{-6.225\%}$	0.000%/ n/a	-0.303%	1.484	10 trajectories
Silverbox (mV RMSE)	9.0847	**4.8716**	$\mathbf{-46.375\%}$	n/a	$+25.799\%$	2.995	1 record
Wiener–Hammerstein (RMSE)	0.156839	**0.140424**	$\mathbf{-10.466\%}$	n/a	-0.600%	2.995	1 record

For L3F the trajectory-bootstrap 95% CI is $[-13.789, -13.051]\%$, all 10 trajectories and all three seeds improve, NMSE falls 24.331%, and round-trip MSE is at most 5.26×10^{-14}. Additional reported diagnostics are in Archive H3 (Table 37); measured gates remain in Appendix B.3.

This comparison profiles ridge independently for every coordinate and uses covariance-aware regularization, under which orthogonal PCA is numerically equivalent to identity. The matched random gauges share architecture, nominal budget, and replicate count. Thus the result cannot be attributed to scalar ridge mismatch, a linear basis change, gauge architecture alone, best-seed selection, or decoder drift.

D.1.3 Cross-system and measured-system evaluation

Second-system fresh confirmation. On a separately admitted d6 chart of the same 100D lift of Rössler, the unchanged $h = 12$ protocol reduces fresh NMAE by 6.225% (95% CI $[5.637, 6.893]\%$), with 10/10 trajectory and 3/3 seed wins. PCA is neutral, random gauges improve only $.303\%$, and $\kappa \leq 1.484$. This confirms another controlled dynamics, not chart transfer (Archive H3, Table 37).

Measured benchmarks. Silverbox is a measured Duffing-like oscillator with known forcing (Wigren & Schoukens, 2013). We freeze a 24-output delay observer, leave input outside the gauge, and lock candidates, gates, evaluator, and 50-output initialization before official access. On the 21,638-sample multisine free run,

mean seed-level RMSE falls from 9.0847 to 4.8716 mV (46.375%); all three seeds improve and random gauges worsen 25.799%. The prespecified arrow shift can diverge: κ certifies the coordinate map, not fitted rollout stability. The gate table remains in Appendix B.3; reported protocol history and the official trace are in Archive H7 (Table 37); dedicated identification models attain lower absolute error (Champneys et al., 2024).

The Wiener–Hammerstein cascade (Schoukens et al., 2009) uses an eight-output delay and the public loader's 100,000/78,800 train/test split, with 50 outputs initializing 78,750 recursively scored samples. After method-specific selector PASS but broader-protocol STOP, a separate PCEGA-only evaluator was locked while official values remained unopened; the earlier STOP is unchanged. Official RMSE falls from $.156839$ to $.140424$ (10.466%), with 3/3 seed wins versus $.600\%$ for random gauges. This post-selector, test-locked study is a second measured single-record intervention, not device replication or a leaderboard claim (Appendix B.3).

For reproducibility, the Wiener–Hammerstein training roles are the zero-based, half-open intervals $[0, 45000)$ (support), $[47000, 58000)$ (inner), $[60000, 70000)$ (outer), $[72000, 80000)$ (checkpoint), and $[82000, 100000)$ (selector). The ridge grid is $\{10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}\}$. Support fits and inner 16-step rollouts select ridge; a support–inner refit supplies the outer objective. Gauge training uses 60 updates, learning rate 3×10^{-4}, four width-32 coupling conditioners, and 32 training rollout starts. Checkpoint evaluation uses 64-step rollouts; the selector record is split into eight contiguous diagnostic blocks, not eight independent experiments. Final official heads refit support, inner, outer, checkpoint and selector intervals after selection, with at most 8192 pairs per interval and independently selected ridge in each coordinate. The receiver predicts all eight future delay entries in transformed coordinates. Inverse decoding scores the first entry, while the full predicted coordinate is propagated without replacing its history entries by exact shifts. This rule is common to all arms and is distinct from a shift-constrained NARX receiver.

D.2 Unified Fixed-Interface Receiver Comparison

All seven alternatives use the same six reconstruction-trained charts, seeds 1943/2943 and source/target pairs. Existing accepted GB gauges are fixed; uncapped uses original EC endpoints without certificate rejection, not a revision of EC decisions. Each configuration fits six Development trajectories, selects physical NMSE on three Checkpoint trajectories (1,024 pairs each), then refits their union. Targets follow each method's native latent regression objective; this is not equal-loss/equal-compute architecture isolation. All 84 models are sealed before twenty new Report trajectories per system at $h = 12$. Technical seeds average first; 20,000 crossed chart/trajectory bootstrap draws use .025 lower quantiles for the two primary identity-minus-GB comparisons. Other contrasts are estimates.

Table 17: Unified fixed-interface comparison. All methods use the same base E, D and source/target pairs. Means include three charts, both technical seeds and twenty new trajectories per system. Uncapped reports raw endpoints without certificate rejection.

	L3F		Rössler	
Receiver	NMSE	NMAE	NMSE	NMAE
Identity Q2	0.007495	0.075037	0.001919	0.033694
GB–Q2	0.003822	0.053056	0.001051	0.023338
Uncapped–Q2	0.003648	0.051789	0.001007	0.022820
Local Q2	0.000109	0.005579	0.000014	0.002938
RBF	0.000109	0.007588	0.000031	0.003663
Residual MLP	0.000456	0.017959	0.000150	0.008540
FineMorphs	0.000546	0.019108	0.000235	0.010633

Global Q2 branches share ridge $\{10^{-12}, 10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}\}$. Local Q2 crosses $K = 4, 16$ KMeans partitions (ten initializations) with ridge $10^{-8}, 10^{-4}$; clusters below two pairs are failures. RBF crosses 256/512 random training anchors, bandwidth .5/1 times their median distance and ridge $10^{-8}, 10^{-4}$; linear features are concatenated and ridge is feature-standardized, not Q2 covariance ridge. Residual MLP crosses width 64/128 and rate .001/.003, two hidden layers, 2,000 Adam updates, batch 256, decay 10^{-5} and cosine ending at .01 times its starting rate.

FineMorphs uses author commit `a33713c4f7cd`, six-dimensional ADA without augmentation, $\lambda = \gamma = 1$, order-three Laplacian kernel, ten integration steps, 128 k-means++ control points and scales .5/1. All fitting targets remain visible. BFGS permits 2,000 iterations with the author stopping rule and no automatic residual-scale cycles; all twelve refits reach that cap without convergence flags. A SciPy line-search import compatibility patch preserves the underlying functions. Native float64 uses float32 AE interfaces. This is an author-source adaptation, not an exact decoder-compensated intervention or an optimal FineMorphs result.

Table 18: Unified NMSE contrasts. Negative differences favor the comparator. Intervals are pointwise two-sided 95% crossed-bootstrap intervals. For the two identity comparisons, the lower endpoint is also the prespecified Bonferroni-adjusted one-sided bound; other rows are estimation-only.

System	Comparator minus GB	Difference	Lower	Upper	Role
L3F	Identity Q2	3.673×10^{-3}	2.643×10^{-3}	5.454×10^{-3}	Primary
L3F	Uncapped–Q2	-1.740×10^{-4}	-5.193×10^{-4}	-6.692×10^{-8}	Estimate
L3F	Local Q2	-3.713×10^{-3}	-4.715×10^{-3}	-3.068×10^{-3}	Estimate
L3F	RBF	-3.712×10^{-3}	-4.607×10^{-3}	-3.018×10^{-3}	Estimate
L3F	Residual MLP	-3.366×10^{-3}	-4.173×10^{-3}	-2.760×10^{-3}	Estimate
L3F	FineMorphs	-3.275×10^{-3}	-4.103×10^{-3}	-2.623×10^{-3}	Estimate
Rössler	Identity Q2	8.688×10^{-4}	6.376×10^{-4}	1.262×10^{-3}	Primary
Rössler	Uncapped–Q2	-4.347×10^{-5}	-6.639×10^{-5}	-3.823×10^{-6}	Estimate
Rössler	Local Q2	-1.037×10^{-3}	-1.081×10^{-3}	-1.002×10^{-3}	Estimate
Rössler	RBF	-1.020×10^{-3}	-1.069×10^{-3}	-9.846×10^{-4}	Estimate
Rössler	Residual MLP	-9.008×10^{-4}	-9.706×10^{-4}	-8.534×10^{-4}	Estimate
Rössler	FineMorphs	-8.159×10^{-4}	-8.644×10^{-4}	-7.860×10^{-4}	Estimate

Table 19: Unified CPU costs, averaging the six chart/seed runs per system. Fit sums configuration-fitting and final-refit wall times. Gauge is the historical coordinate-training time, reported separately. B1/B128 average per-run median full-pipeline latencies at batches one/128. Numeric counts compact prediction state, including a gauge when present; Package includes retained training state. Both exclude common AE parameters (96.78 KiB). Budgets are method-specific, not equal-compute comparisons.

System	Receiver	Fit (s)	Gauge (s)	B1 (ms)	B128 (ms)	Numeric (KiB)	Package (KiB)
L3F	Identity Q2	0.58	0.00	1.48	11.55	1.78	27.49
Rössler	Identity Q2	0.65	0.00	1.46	9.77	1.78	27.49
L3F	GB–Q2	0.69	267.60	1.94	12.57	80.00	114.77
Rössler	GB–Q2	0.75	250.93	1.94	10.61	80.00	114.72
L3F	Uncapped–Q2	0.73	265.37	1.98	12.65	80.00	114.78
Rössler	Uncapped–Q2	0.74	248.58	1.96	10.57	80.00	114.73
L3F	Local Q2	1.05	0.00	1.64	13.15	25.27	376.24
Rössler	Local Q2	1.08	0.00	1.64	12.16	28.88	429.70
L3F	RBF	2.47	0.00	1.27	10.80	44.47	8429.29
Rössler	RBF	2.53	0.00	1.26	10.27	44.47	8429.28
L3F	Residual MLP	13.26	0.00	1.24	9.79	71.07	76.04
Rössler	Residual MLP	12.97	0.00	1.25	9.65	71.07	76.03
L3F	FineMorphs	4900.91	0.00	1.89	13.20	126.76	23078.92
Rössler	FineMorphs	4867.00	0.00	1.81	12.98	126.76	23078.92

Every model contributes; no failed case is omitted. Costs include fitting all configurations plus final refit, with gauge search listed separately. After training, five warmups and twenty timed calls yield per-run medians for full E–receiver–D inference. Numeric state uses diagonal transforms stored as vectors, without changing predictions; serialized packages can include training state and are not minimum deployment memory. Timings use three two-thread CPU workers during training and are implementation-specific. Local Q2/RBF/MLP/FineMorphs all have lower error than GB–Q2; RBF and MLP also have lower batch-one latency and numerical state. Thus no accuracy–cost dominance or necessity of retaining Q2 is claimed. Related conjugate-mapping work is not counted as another reproduced baseline.

E Receiver Structure and Transfer

E.1 Crossed-Interaction Details

Table 5 reports absolute NMAE for these layer-budget coordinates. Q2 and RBF prefer their own trained coordinates in both systems; the same Q2 coordinate increases MLP error. The four Q2/RBF directional comparisons below are the primary tests.

Table 20: L3F localization at fixed degree two. Effects are trajectory-macro percentages after seed-risk averaging. Transfer holds the global gauge fixed; search retrains it. Physical effects and search outcomes are from the frozen evaluation. Pointwise paired-trajectory bootstrap 95% intervals and the decoded-dynamics/base-latent metrics are post-confirmatory diagnostics.

K	Id. NMAE	Physical [95% CI]	Decoded	Latent	Search
1	0.096836	13.085 [12.901, 13.266]	13.141	12.064	13.085
4	0.017241	4.987 [4.599, 5.391]	5.472	4.396	1.244
16	0.005028	0.333 [0.243, 0.419]	2.481	3.645	$\simeq 0$ (identity)

Each entry below is other-coordinate minus own-coordinate NMAE for one receiving learner. Positive favors its own coordinate. The uncertainty is conditional on the frozen three-seed family and configuration; it does not cover the distribution of all historical development choices.

System / receiver	Difference	98.75% trajectory interval
L3F / Q2	0.01138841	[0.01099883,0.01187312]
L3F / RBF	0.00047371	[0.00045718,0.00049077]
Rössler / Q2	0.00198957	[0.00182053,0.00214983]
Rössler / RBF	0.00010374	[0.00009791,0.00010892]

All 2,400 candidate–trajectory metric rows and saved latent prediction traces are retained. The metric replay is not independent retraining or a clean-environment end-to-end replication. The two systems share an observation lift family. RBF's Rössler identity-relative gain is less robust to seed resampling than the directional preference over the Q2 coordinate. Six-step neural adaptation is a limited-budget experiment. The larger identity RBF configurations are specified in Appendix B.1; the full prediction matrix is a separate experiment artifact, not a table reproduced here.

E.2 Structural Contraction: Evidence Roles and Metrics

The original structural study uses six Development, three Checkpoint, ten Selector and ten Fresh trajectories of the L3F task at $h = 12$. Eighteen gauges cover global Q1/Q2/Q3, patchwise Q2 with $K = 4, 16$, and local Q1 with $K = 32$, retaining all three optimization seeds. Partition memberships are defined in identity coordinates and transported by sample identity. The frozen evaluator already computes both cell-specific reoptimized and global-Q2 frozen-transfer scores. Our new presentation does not rerun or relabel the data. The primary frozen interaction concerns global-Q2 versus local-Q2 reoptimized search; fixed transfer is a retained supporting comparison.

All three $K = 16$ runs retain epoch-zero identity before evaluation. The post-confirmatory optimization audit shows nonzero gradients and training-risk reductions of 76.6–78.2%, but nonidentity checkpoints do not improve validation. Thus this is not an inert optimizer; nor is it evidence that the best coordinate in the full admissible family is identity. The local gate has a one-sided positive-value upper bound, not a two-sided equivalence test.

The fixed gauges give physical/decoded-dynamics/base-latent relative values of $13.085/13.141/12.064\%$ globally, $4.987/5.472/4.396\%$ at $K = 4$, and $0.333/2.481/3.645\%$ at $K = 16$. The latter two metric types were examined after confirmation and are diagnostic, not additional locked primary tests. Identity physical NMAE decreases from .096836 to .017241 to .005028, compared with the base reconstruction reference .004777. This reference can mask changes in latent transition quality but is not asserted to lower-bound all decoded predictors.

The structural intervention also changes model size and per-patch sample allocation; it is not a parameter-count-matched isolation of coefficient sharing alone. Q3 and local-Q1 results rule out a general monotone law in degree or patch count. No population-oracle or universal restoration claim follows from these finite scores. An asymptotic local-approximation theorem would additionally need regularity uniform over the gauges, suitable support coverage and estimator consistency; a bi-Lipschitz budget alone does not bound higher derivatives. Such a theorem is not substituted for this empirical result.

E.3 Frozen Cross-Component Evaluation

The development audit completes the missing off-diagonal Q2/RBF donor–receiver factorial on the existing learned encoder–decoder interfaces. Each system retains three gauge seeds and three rotating historical Checkpoint folds. Each arm fits on the six Development trajectories, profiles on an inner Checkpoint trajectory, and

Table 21: Development donor preferences in pp of the common receiver identity risk. Positive favors the own-donor coordinates; this is not an additive mediation decomposition.

System	Head	Input only	Output only	Joint
L3F	Q2	11.049	-0.246	11.000
L3F	RBF	5.554	0.262	6.483
Rössler	Q2	5.539	0.507	6.060
Rössler	RBF	3.475	0.006	3.457

Table 22: Input-only donor preferences on 20 new trajectories per system. Absolute physical NMAE averages all nine retained seed/fold refits. Preference is other minus own risk, normalized by identity; positive favors own. LB is the adjusted one-sided bootstrap lower endpoint on absolute differences, scaled by the observed identity mean (not a ratio confidence interval).

System	Head	Identity	Own	Other	Pref. (pp)	LB (pp)
L3F	Q2	0.097616	0.085142	0.096840	11.98	11.68
L3F	RBF	0.006208	0.005832	0.006149	5.10	4.93
Rössler	Q2	0.033446	0.031606	0.033260	4.95	4.48
Rössler	RBF	0.004756	0.004658	0.004832	3.65	3.49

refits support plus inner before evaluating the other Checkpoint trajectory. Gauges are not retrained. All 288 labels complete; 108 off-diagonal nonidentity configurations are new, while repeated own arms are numerical bridges. All 144 own-arm scores and profile selections match their earlier results exactly. Saved-head prediction replay has zero maximum error; separate NumPy metric replay differs by at most 2.26×10^{-7} relatively. These reused Checkpoint results are descriptive, not confirmation.

Independent test. After development, 978 artifacts bind the complete fitted matrix, generators, training-derived normalizers, evaluator, statistics and environment before generation. The independent role uses 20 new initial-condition seeds per system, with the original dynamics parameters, observation lift and E/D. Each 6000-observation trajectory contributes 1024 deterministically spaced valid $h = 12$ pairs. This differs from the older all-pairs evaluation and is reported as a separate pipeline, not a replacement of the original table. All four component arms, both donor families and every technical seed/fold are retained without fitting or selection after locking.

Risks, not predictions, are averaged over the nine technical refits before forming each trajectory contrast. The four primary comparisons are the two directed input-only donor preferences in each system. A fixed 20000-resample trajectory bootstrap reports 95% percentile intervals and one-sided lower endpoints at quantile $.05/4$ for the four-direction claim. This is approximate bootstrap/Bonferroni inference conditional on the retained models, not a finite-sample certificate or inference over a population of devices. Time points, seeds and fitting folds are not independent scientific units. Joint/output-only, NMSE and per-seed results remain descriptive supporting outputs. All physical predictions, sampled targets and per-trajectory scores are retained for replay.

All four adjusted lower endpoints are positive: the input-only intervention retains both directed preferences in both systems.

E.4 Controlled Six-Receiver Benchmark

Table 1 in the main text reports the validation-gated results.

States and observations. The benchmark uses explicit systems and generated observations, independently of the pretrained encoders and checkpoints in the earlier experiments. Lorenz–63 has $(\sigma, \rho, \beta) = (10, 28, 8/3)$; Rössler has $(a, b, c) = (.2, .2, 5.7)$. Forced Duffing obeys $\dot{q} = v$, $\dot{v} = q - q^3 - .2v + .3\cos\theta$ and $\dot{\theta} = 1.2$. Its interface includes $(q, v, \sin\theta, \cos\theta)$, so the forcing phase is observed. Hénon iterates $(x, y) \mapsto (1 - 1.4x^2 + y, .3x)$. RK4 step sizes are $.01, .02, .05$ for the continuous systems, with direct horizon 12; Hénon uses horizon 2. Burn-in lengths are $3000, 5000, 3000, 1000$, respectively, followed by 2048 states. Each trajectory contributes 192 evenly spaced valid source–target pairs. Initial states and all role seeds are explicit in the accompanying generator; no unstable trajectory is replaced.

Native states are centered and divided by one scalar RMS fitted on Chart trajectories. For normalized native state $u \in \mathbb{R}^d$, append $.8\tanh(Au)$ to obtain six dimensions, with a fixed $\|A\|_2 = 1$. A Chart-fitted translation,

scalar scale and orthogonal rotation give $z = C(u, v)$. The observation map is

$$D(z) = \left[z,\ z^2,\ \sqrt{2}(z_i z_j)_{i<j},\ \sin z,\ \cos z, \sin(\mathbf{1}^T z/\sqrt{6})\right] \in \mathbb{R}^{40}.$$

The encoder reads the first six coordinates. Thus decoder compensation is exact, while no additional observations enter an adapted head. All comparisons use this same six-dimensional interface; a lower-dimensional native-state learner is not substituted for an identity control.

Candidate training and receiving heads. The four-block additive coupling map has conditioners of dimensions $6 \to 32 \to 32 \to 6$, with two tanh activations and spectral cap .82 on every matrix. Alternating masks, identity initialization and a scalar normalizer produce a global bound $[(\sqrt{4 + .82^6} + .82^3)/2]^4 < 3$. All donors receive 36 AdamW updates at learning rate .002, weight decay 10^{-5} and gradient-norm cap 5. Each update averages six rotating Development folds: four support trajectories, one Inner profile trajectory, and one Outer physical-risk trajectory. After Inner profiling, refit on support plus Inner before differentiating Outer physical MSE normalized by Chart variance. Checkpoint chooses among epochs $0, 6, \ldots, 36$; every update runs irrespective of the retained epoch.

Linear and complete quadratic heads use covariance-aware ridge, including the unpenalized intercept. ELM has 64 fixed random-tanh features plus the state; RBF has 64 source-selected anchors plus the state. Both use feature-standardized ridge. The common ridge grid is $\{10^{-12}, 10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}\}$; ELM basis scales and RBF bandwidth factors are $\{.5, 1, 2\}$. SVD profiles the fixed feature map and augmented QR performs differentiable refits. The floor's full derivative is used here, unlike the historical stopped-gradient convention disclosed in Appendix C.3.

The residual MLP has two tanh hidden layers, widths $32/64$, and learning rates $.0003/.001/.003$, with batch size 128, a cosine schedule and Adam weight decay 10^{-5}. An identity-only comparison of 4000 and 8000 updates improved mean Checkpoint NMAE by $31.83\%, 24.24\%, 30.15\%, 54.76\%$ across the four systems. The predeclared rule therefore assigns every MLP coordinate all six configurations at 8000 updates. ExtraTrees profiles $128/512$ trees and leaf sizes $1/4$, with all features. Heads fit six Development trajectories; three Checkpoint trajectories select hyperparameters and retained MLP epochs. There is no subsequent head refit.

Data roles and selection. Each system has $4/6/3/3/10/12$ trajectories in Chart, Development, Checkpoint, Atlas, Selector and Fresh, respectively. These role seeds are all distinct across the six natural/adaptation systems. Seeds $1943, 2943, 3943$ are technical optimization repeats, never selected or treated as independent trajectories. Atlas chooses the best family from identity and four learned coordinates, averaging seed risks first. It also fixes the best other learned family for each donor receiver and one shared coordinate minimizing mean receiver-relative risk. Stable ties prefer the declared library order, with identity first. Strict wins use tolerance $10^{-10} + 10^{-8}$ times the reference risk.

Selector requires at least 1% mean NMAE reduction, at least $8/10$ strict trajectory wins, and at least .5 percentage points more reduction than a matched random gauge using the identity denominator. Every retained seed must satisfy the global budget, round-trip MSE $\leq 10^{-10}$ and displacement matching. A reference below 10^{-8} returns identity without a relative-risk interpretation. Rejection does not promote a runner-up. All fitted candidates, sources, roles and decisions are fixed before their succeeding evaluation role.

Controls and numerical checks. The primary random construction uses the same architecture and spectral caps, with one fixed random draw per seed. Paired complementary active-channel shears allow a common construction to match all 72 donor displacement RMSs within 1% on Development source states. Its amplitude is selected only by geometry, not prediction risk. Centered displacement and residual after a best affine fit are also recorded. Earlier dense/orthonormal-path distributions remain secondary; unattainable matches retain their failure flags.

Ordinary affine refits and a separately transported full-penalty affine control are distinguished. The coordinate-local covariance floor ϵI is not itself affine covariant; if $X_H = XT$, the diagnostic transports the floor as $\epsilon T^T T$. Consequently, class closure is not misreported as a null property of the ordinary finite-floor algorithm. Numerical identity maps are canonicalized only for exact structural zeros; no small-displacement threshold aliases a nonzero map. This prevents floating-point normalization round trips from changing tree splits in an otherwise identical null control.

Two adaptation systems use a stable three-dimensional linear transition, an orthogonal conjugate of $\mathrm{diag}(.995, .990, .985)$, with 512 transient states and horizon 12. One has a linearly padded chart; the other has the nonlinear chart above. The latter's known straightener is $CT^{-1}C^{-1}$ with $T(u, v) = (u, v + .8\tanh(Au))$. It is an analytic approximation control, not a learned candidate promoted after evaluation. Full-budget shuffled donors permute latent targets and their physical truths together; their receiving heads fit the correct transitions.

The final matrix has 2064 heads: 1944 system/seed/receiver/coordinate fits, 72 affine controls and 48 shuffled controls. Closed-form receivers have 22 coordinates and MLP/ExtraTrees have 10 before the extra controls.

Table 23: Mean Fresh physical NMAE, with seed risks averaged first. Column headings name the donor of each learned coordinate; Known is the fixed analytic straightener, not a test-selected candidate. The same physical target and decoder metric apply across each row.

System	Receiver	I	Linear	Q2	ELM	RBF	Known
Lorenz–63	Linear	0.623153	0.220364	0.545924	0.618958	0.556199	0.809217
	Q2	0.048754	0.042709	0.036629	0.053422	0.063589	0.415353
	ELM	0.010599	0.009852	0.011954	0.007967	0.008082	0.001815
	RBF	0.008835	0.007835	0.011714	0.008083	0.005876	0.002615
	MLP	0.009701	0.012418	0.011403	0.009779	0.011912	0.007739
	Trees	0.035221	0.038551	0.034101	0.034213	0.035456	0.033981
Rössler	Linear	0.205610	0.088036	0.200031	0.205622	0.191647	0.201138
	Q2	0.013372	0.012206	0.008979	0.014311	0.013321	0.035176
	ELM	0.001321	0.001288	0.001420	0.000919	0.001173	0.000436
	RBF	0.006311	0.006333	0.006258	0.005892	0.003978	0.004695
	MLP	0.004656	0.004821	0.004701	0.004990	0.005476	0.004261
	Trees	0.022725	0.024673	0.023554	0.024211	0.024601	0.023543
Duffing	Linear	0.188876	0.159815	0.188879	0.189236	0.189890	0.190065
	Q2	0.189522	0.158598	0.169713	0.185870	0.181860	0.190084
	ELM	0.011583	0.011691	0.011780	0.008708	0.010219	0.008113
	RBF	0.009370	0.012286	0.009316	0.008458	0.007773	0.005569
	MLP	0.008850	0.009724	0.009080	0.009239	0.009264	0.008796
	Trees	0.061742	0.062879	0.061886	0.062742	0.064729	0.062832
Hénon	Linear	0.824553	0.353557	0.794268	0.788547	0.803692	1.055544
	Q2	0.026664	0.022000	0.016714	0.022294	0.021159	0.971465
	ELM	0.004078	0.003914	0.003174	0.002372	0.002153	0.000260
	RBF	0.012991	0.014102	0.011023	0.011449	0.003753	0.002446
	MLP	0.013514	0.016849	0.014673	0.014097	0.019573	0.009913
	Trees	0.008716	0.009439	0.008220	0.009109	0.010189	0.007528

All retained Checkpoint predictions pass independent NumPy physical decoding and metric replay; applicable QR/SVD refits agree within the declared tolerance. Four prespecified Lorenz seed-1943 donor replays reproduce all 37 curve rows and retained tensors exactly. This is numerical repeatability in the installed environment, not a claim about every platform or every training run.

Fresh results and inference. The Atlas candidates give a 24-cell macro reduction of 24.960% (95% CI $[24.368, 25.600]\%$); the frozen deployment policy gives 24.840% ($[24.262, 25.476]\%$). All 17 authorized conditions improve on all twelve Fresh trajectories relative to identity. Own coordinates have lower mean risk than Atlas-best-other learned coordinates in $14/16$ anchor cells; $13/16$ pass the fixed BY strict-win family. Duffing Q2 and Hénon ELM prefer the other coordinate, while the Lorenz ELM own-versus-other comparison is uncertain. The full absolute matrix below also includes the predetermined analytic straightener, outside the learned selection library. These control comparisons are not used to revise deployment.

Shuffled-donor control. All 16 true Own coordinates outperform their full-budget shuffled donors in mean risk, but shuffled training itself improves some heads. These descriptive controls are not a new confirmatory family; the full 16-row matrix is in Archive H5 (Table 37).

Fresh uncertainty resamples twelve complete trajectories, after averaging the three technical seed risks. Percentile intervals use 20000 paired resamples. The four-system macro gives equal weight to 24 system/receiver cells and resamples within each fixed system; it does not estimate uncertainty over a population of dynamical systems. The ratio of pooled absolute NMAEs is reported separately. Intervals condition on retained training samples, candidate library and technical seeds; they do not measure retraining variation over new Development sets. Sign tests concern the probability of a strict trajectory win, with ties as nonwins, not the mean effect. The fixed test families have 24 candidate–identity, 24 candidate–random, 16 own–best-other, and 24 selected–universal contrasts. BY at $q = .10$ is the dependence-robust discovery label; BH and Holm values are also retained. Nonsignificant, rejected, identity and opposite-direction outcomes remain in these families.

The results package reconstructs physical and latent risks from saved predictions and reproduces choices and statistical summaries without a GPU. Full project sources, the role generator, per-seed matrices and the retained-model numerical audit accompany the experiment. The present analytic chart benchmark supplements, rather than replaces, the learned-chart crossings and measured-system evidence.

Table 24: Own coordinate versus the best other learned family selected on Atlas. Positive reductions favor Own. Bootstrap intervals summarize mean-effect uncertainty; BY labels concern strict-win probability, a separate hypothesis.

System	Receiver	Other	Reduction (%)	95% CI	Wins	BY
Lorenz–63	Linear	Q2	+59.63	$[+58.20, +60.98]$	12/12	Yes
Lorenz–63	Q2	Linear	+14.24	$[+13.02, +15.46]$	12/12	Yes
Lorenz–63	ELM	RBF	+1.43	$[-2.63, +6.22]$	8/12	No
Lorenz–63	RBF	Linear	+25.00	$[+23.82, +26.48]$	12/12	Yes
Rössler	Linear	RBF	+54.06	$[+52.98, +55.15]$	12/12	Yes
Rössler	Q2	Linear	+26.43	$[+24.50, +28.25]$	12/12	Yes
Rössler	ELM	RBF	+21.65	$[+20.36, +23.07]$	12/12	Yes
Rössler	RBF	ELM	+32.49	$[+30.45, +34.19]$	12/12	Yes
Duffing	Linear	Q2	+15.39	$[+10.99, +19.53]$	12/12	Yes
Duffing	Q2	Linear	-7.01	$[-8.53, -5.55]$	0/12	No
Duffing	ELM	RBF	+14.79	$[+11.73, +19.68]$	12/12	Yes
Duffing	RBF	ELM	+8.11	$[+6.20, +9.93]$	12/12	Yes
Hénon	Linear	ELM	+55.16	$[+54.07, +56.11]$	12/12	Yes
Hénon	Q2	RBF	+21.01	$[+19.49, +22.34]$	12/12	Yes
Hénon	ELM	RBF	-10.13	$[-12.07, -7.83]$	0/12	No
Hénon	RBF	Q2	+65.96	$[+65.48, +66.48]$	12/12	Yes

Table 25: Adaptation Fresh NMAE. All twelve adaptation receiver conditions deploy identity. The known map recovers the calibrated linear dynamics to numerical precision for polynomial heads; the finite learned candidates need not discover this opportunity. A near-zero reference is not converted to a relative benefit.

System	Receiver	I	Known	Own learned
Linear null	Linear	1.71e-12	1.71e-12	1.71e-12
	Q2	1.7e-12	1.7e-12	1.7e-12
	ELM	0.000133	0.000133	0.000133
	RBF	1.96e-10	1.96e-10	1.96e-10
	MLP	0.00796	0.00796	–
	Trees	0.576	0.576	–
Known conjugacy	Linear	0.0121	1.84e-12	0.0121
	Q2	0.104	1.84e-12	0.0962
	ELM	0.0544	6.89e-05	0.044
	RBF	0.0235	4.45e-10	0.0332
	MLP	0.0082	0.00739	–
	Trees	0.554	0.547	–

E.5 Analytic straightener control

An analytic straightener, fixed independently of donor optimization, gives a particularly direct control. On Lorenz it increases Q2 NMAE from 0.048754 to 0.415353, but reduces ELM NMAE from 0.010599 to 0.001815 under the same exact reconstruction interface. It also improves the MLP by 20.22%, although the learned library selected identity. Table 26 shows the complete control comparison. This is a total intervention on both input features and output coordinates. Separating the two on development data reveals input-feature loss alongside known output-structure recovery (Appendix E.5). Its receiver-dependent total effect also shows why a library fallback does not establish that better coordinates are absent.

The full matrix separates coordinate opportunity from its discovery. In particular, an own-donor coordinate need not outperform the best other learned coordinate selected before validation. The direct MLP also offers a different error–capacity choice from adapting a lower-order receiver. Absolute risks, own-versus-other contrasts, adaptation outcomes and all rejected candidates are reported in Appendix E.4; the systems in this benchmark are not pooled with the learned-chart or measured-system results.

Input/output-factorial interpretation. For $C_0 = I, C_1 = H$, the diagnostic fits $f_{ab}(z) = DC_b^{-1}\widehat{g}_{ab}(C_a z)$ to targets $C_b z^+$, independently profiling Q2/ELM in all four arms on six Development and three already opened Checkpoint trajectories per system, retaining three seeds (96 heads). Input-only transformation increases mean error in all eight system–receiver cells; output-only decreases it. For Lorenz the physical NMAEs

Table 26: One fixed analytic straightener per system, evaluated by six receiving learners, transforming both source and target coordinates. Entries are mean test NMAE divided by that learner's identity NMAE; ratios below one indicate lower error. Absolute risks are in Table 23. The transformation is independent of donor training. These are descriptive comparisons for a prespecified control, not additional selections or a new confirmatory testing family.

System	Linear	Q2	ELM	RBF	MLP	Trees
Lorenz–63	1.30	8.52	0.17	0.30	0.80	0.96
Rössler	0.98	2.63	0.33	0.74	0.92	1.04
Duffing	1.006	1.003	0.70	0.59	0.994	1.018
Hénon	1.28	36.43	0.064	0.19	0.73	0.86

$(00, 10, 01, 11)$ are $(.049600, .458162, .034304, .418354)$ for Q2 and $(.010154, .016203, .001689, .001723)$ for ELM. The sampled Q2 design ranks fall $28 \to 10$ (Lorenz/Rössler), $27 \to 14$ (Duffing) and $28 \to 6$ (Hénon), stable at relative SVD tolerances $10^{-8}, 10^{-10}, 10^{-12}$; these are not population ranks.

On the known chart $M = \{\Gamma(u, .8\tanh(Au))\}$, with fixed invertible affine Γ, the straightener agrees with $L(\Gamma(u, v)) = \Gamma(u, 0)$. At a common fixed feature/ridge profile, target-independent linear readout fitting with an unpenalized intercept gives $\widehat{g}_{a1} = L\widehat{g}_{a0}$; decoding therefore applies the known-graph retraction $H^{-1}L$ to the original prediction. Independent profile choices can break equality of selected heads. Thus the joint intervention combines input-feature changes and output structure recovery, not additive contributions to a unique risk mechanism.

This is development evidence, not another confirmation. The full eight-row table and derivation are in Archive H4 (Table 37).

F Physical-Loss Refitting and Interface Details

F.1 Physical-loss refitting

Freeze one chart per system (chart zero, gauge seed 1943), identity and the 192-update GB/scheduled gauges. Each independently tunes its ridge head on three Checkpoint trajectories after six Development trajectories. Only an additive correction to the full Q2 coefficients is then optimized through fixed H^{-1}, D, with observed or base-reconstructed targets. All $n = 6144$ Development pairs use source-fitted centering/scaling. With $\Phi = U\Sigma V^\top$, the correction is ΦTA, $T = V\,\mathrm{diag}(\sqrt{n}/\max(\sigma_j, 10^{-6}\sigma_{\max}))$, and A starts at zero. This invertible preconditioning does not enlarge Q2. Float64 L-BFGS uses 64 outer steps (one iteration each), learning rate one, history 20, strong-Wolfe search, gradient/change tolerances $10^{-12}/10^{-14}$ and no penalty. Checkpoint selects among $0, 1, 4, 16, 64$ using observed error for either objective, without refitting. Both L3F GB paths choose 16; the other ten choose 64, not a global-optimality certificate.

All twelve training paths and eighteen final predictors are retained, then evaluated on ten new trajectories per system (1,024 direct $h = 12$ pairs). No new Selector gate is tested: these are fixed candidates, not complete training repetitions. Trajectory-centered physical normalizers and ratios of mean risks differ from the original trajectory-macro NMAE estimand. Matched-identity gains before/after refit are 48.69/50.00% and 58.07/58.73%

Table 27: Fixed-candidate physical-loss control. "Original" refers to the head fit in this study, not the manuscript's historical efficacy table. Each physical refit improves trajectory NMSE in 10/10 cases.

System	Coordinate	Original NMSE	Refit NMSE	Original NMAE	Refit NMAE
L3F	Identity	0.006672836	0.006321777	0.069803765	0.067254998
	GB	0.003424089	0.003160752	0.049970628	0.047586845
	Schedule	0.002797710	0.002609086	0.044738768	0.042812381
Rössler	Identity	0.001681776	0.001533940	0.031750462	0.030966986
	GB	0.000958306	0.000896663	0.022307493	0.022171465
	Schedule	0.000795698	0.000745960	0.019193197	0.019213711

(L3F GB/schedule), 43.02/41.55% and 52.69/51.37% (Rössler). Post-confirmatory paired bootstraps (20,000 trajectory draws) give pointwise 95% intervals for changes of [1.080,1.550], [.494,.822], [-2.089,-.907] and [-1.613,-1.010] pp. These are not multiplicity-adjusted, equivalence tests or between-training/chart intervals.

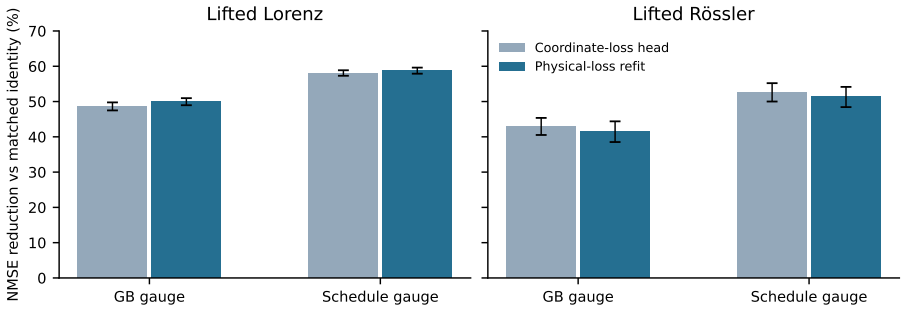

Figure 3: Coordinate value after physical-loss head refitting. Each bar compares a gauge with identity using the same head-fitting rule; gains are ratios of mean physical NMSE. Candidates were frozen before evaluation. The visualization and pointwise paired-trajectory 95% bootstrap intervals are post-confirmatory analyses, conditional on fixed charts and candidates.

Reconstructed-target gains are 49.97/58.68% and 41.50/51.29%. The physical squared error retains the cross term:

$$\|\widehat{x} - x^+\|^2 = \|\widehat{x} - DE(x^+)\|^2 + \|DE(x^+) - x^+\|^2 + 2\langle \widehat{x} - DE(x^+), DE(x^+) - x^+\rangle.$$

Thus reconstruction error is invariant, not a forecasting lower bound. Similar refit outcomes weaken but do not exclude reconstruction compensation; remaining gains can include optimization/estimation effects. Corrections merge into the same 28×6 matrix (180 scalars with centers/scales), with no extra deployed SVD branch; maximum checked discrepancy is 3.73×10^{-14}. All results, including Rössler NMAE regression, remain in the table. Implementation: `metric_alignment_pilot_gpu.py`.

F.2 Fitting and inference

With centered covariances C, B and an unpenalized intercept, covariance-aware regularization gives $W_\lambda = ((1+\lambda)C + \lambda\epsilon I)^{-1}B$. Without flooring and at full rank, $W_\lambda = (1+\lambda)^{-1}W_{\mathrm{OLS}}$: uniform, not direction-dependent shrinkage. Affine invariance requires transporting the metric, including any floor. Independent final head tuning remains part of the comparison even when training-time profiling adds no benefit.

Receiver requirements and inference. Gauge training differentiates through H, H^{-1}, D and the selected head fit; a black-box receiver needs a different optimizer or a fixed coordinate library. Synthetic Q2 directly predicts six future coordinates with a complete quadratic dictionary, not a claimed exact Koopman closure. Measured receivers use known forcing and recursive delay-state updates; we evaluate both the original full-vector update and exact history shifts. Their distinction matters for fixed-gauge transfer (Appendix G). Unlike Algorithm 1, historical measured protocols refit the head with Selector data after fixing the coordinate decision. The selected coordinate is unchanged, but the officially evaluated head is not the head scored on Selector; the utility check does not certify this refitted predictor's risk.

Inference adds one gauge and its inverse per prediction step. The composite predictor is nonlinear, so physical interpretability does not automatically follow from a low-degree receiver. Likewise, the coordinate certificate alone bounds neither the receiver's sensitivity nor recursive rollout error. An identity-coordinate stronger receiver also preserves E, D and has coordinate distortion one; retaining the interface alone does not require adopting this component.

F.2.1 Expanded default GB algorithm

Algorithm 3 expands the fitting and selection steps of Algorithm 1. Here K is the number of prescribed Development splits, λ_r is the ridge coefficient selected for the current map on split r, and λ_Q is selected independently for the final coordinate map Q. Each evaluation of $\widehat{R}_V(H; S, \lambda)$ fits on S_H and scores decoded physical error on V. Ridge ties follow the fixed grid order. The approximate hypergradient $\widetilde{\nabla}$ differentiates the selected refit with the declared covariance-floor convention. RestoreBudget rescales conditioner output matrices; Eligible checks the public cap and inverse round-trip tolerance. SelectCheckpoint uses the fixed Checkpoint folds and favors the earlier update on ties. Accept applies the prespecified mean-error and trajectory-win thresholds. These settings are detailed in Appendices C.5 and C.3. Report data are not inputs to any fitting or selection step.

Algorithm 3 Expanded PCEGA global-budget procedure.

Input: Frozen (E, D); regression fitting algorithm $\mathcal{A}_\lambda$; ridge grid Λ;
disjoint data $S_{\mathrm{dev}}, S_{\mathrm{ckpt}}, S_{\mathrm{sel}}$;
K Development folds, update budget T, checkpoint schedule $\mathcal{T}$, and caps $1 \le \kappa_{\mathrm{int}} \le \kappa$.
Output: Coordinate map $H^\star$ and fitted predictor $\widehat{g}^\star$.

Initialize $H_\theta = I$; $\mathcal{C} \leftarrow \{(0, I)\}$.
for $t = 1, \dots, T$ **do**
  for $r = 1, \dots, K$ **do**
    $(S^r_{\mathrm{sup}}, S^r_{\mathrm{in}}, S^r_{\mathrm{out}}) \leftarrow \textsc{Split}(S_{\mathrm{dev}}, r)$
    $\lambda_r \leftarrow \arg\min_{\lambda \in \Lambda} \widehat{R}_{S^r_{\mathrm{in}}}(H_\theta; S^r_{\mathrm{sup}}, \lambda)$
    $\mathcal{L}_r(\theta) \leftarrow \widehat{R}_{S^r_{\mathrm{out}}}(H_\theta; S^r_{\mathrm{sup}} \cup S^r_{\mathrm{in}}, \lambda_r)$
  end for
  $\theta \leftarrow \textsc{Update}(\theta, \widetilde{\nabla}_\theta K^{-1} \sum_{r=1}^K \mathcal{L}_r(\theta))$
  if $B(\theta) > \kappa_{\mathrm{int}}$ **then**
    $\theta \leftarrow \textsc{RestoreBudget}(\theta, \kappa_{\mathrm{int}})$
  end if
  if $t \in \mathcal{T}$ **and** $\textsc{Eligible}(H_\theta, \kappa)$ **then**
    $\mathcal{C} \leftarrow \mathcal{C} \cup \{(t, H_\theta)\}$
  end if
end for
$(t_c, H_c) \leftarrow \textsc{SelectCheckpoint}(\mathcal{C}, S_{\mathrm{dev}}, S_{\mathrm{ckpt}})$
for all $Q \in \{I, H_c\}$ **do**
  $\lambda_Q \leftarrow \arg\min_{\lambda \in \Lambda} \widehat{R}_{S_{\mathrm{ckpt}}}(Q; S_{\mathrm{dev}}, \lambda)$
  $\widehat{g}_Q \leftarrow \mathcal{A}_{\lambda_Q}((S_{\mathrm{dev}} \cup S_{\mathrm{ckpt}})_Q)$
end for
$(H^\star, \widehat{g}^\star) \leftarrow (I, \widehat{g}_I)$
if $\textsc{Accept}(H_c, \widehat{g}_{H_c}, I, \widehat{g}_I, S_{\mathrm{sel}})$ **then**
  $(H^\star, \widehat{g}^\star) \leftarrow (H_c, \widehat{g}_{H_c})$
end if
Freeze $(H^\star, \widehat{g}^\star)$ before Report evaluation.
return $(H^\star, \widehat{g}^\star)$

F.3 Latent-Coordinate Views and Their Interpretation

The GB/Q2 views in Figure 2 are separate from the LB receiver-crossing study.

Figure 2 uses the fixed L3F chart zero, seed 1943 and 192-update GB/scheduled-GB candidates. Both systems are shown in Figure 4; the trajectory in each is the lowest-ID historical Checkpoint trajectory, with every sixth source frame displayed at horizon 12. The common PCA basis is fitted to per-dimension-standardized Development identity states. The Checkpoint trajectories were used historically in model selection and final fitting, so these plots describe fitted candidates rather than independently estimate gains. Two-dimensional projection does not preserve all distances or transfer the global certificate to the displayed axes.

The artifact also retains the earlier report-only own-PCA, common-frame, local-geometry, displacement-field, and measured seed-level figures generated from frozen checkpoints. No visualization selected a seed, checkpoint, threshold, or claim. Their numerical summary is intentionally concise: learned/random L3F local-distance ratios span .933–1.124 and .987–1.041 (5–95%); on Silverbox median coordinate-RMS displacement is $9.27\%/30.57\%$, although only the learned gauge improves official RMSE. Thus neither visual compactness nor displacement magnitude is a sufficient utility diagnostic. The remaining historical figures and plotting manifests are retained in the artifact.

G Coordinate Sensitivity and Measured Rollout Diagnostics

Sampled inverse amplification reaches 80–89% of the LB bound and 45–54% of the GB bound in a fixed-gauge audit (Appendix C.2). This gives the certificate a measurable coordinate scale, but physical sensitivity also depends on the decoder. On previously opened measured records, Silverbox's frozen gauges fail transfer from full-vector to exact history-shift updates. Receiver-matched retraining instead improves exact-shift RMSE by 24.30% on multisine and 35.83% on arrow, with all three seeds finite. Its full-vector counterpart improves multisine by 45.71% but diverges on arrow in all three seeds. These are post-confirmatory structural diagnostics, not new measured confirmation or evidence that certification prevented divergence (full reported protocols and tables: Archive H6, Table 37). Likewise, lowering the numerical whitening floor makes two of three Silverbox

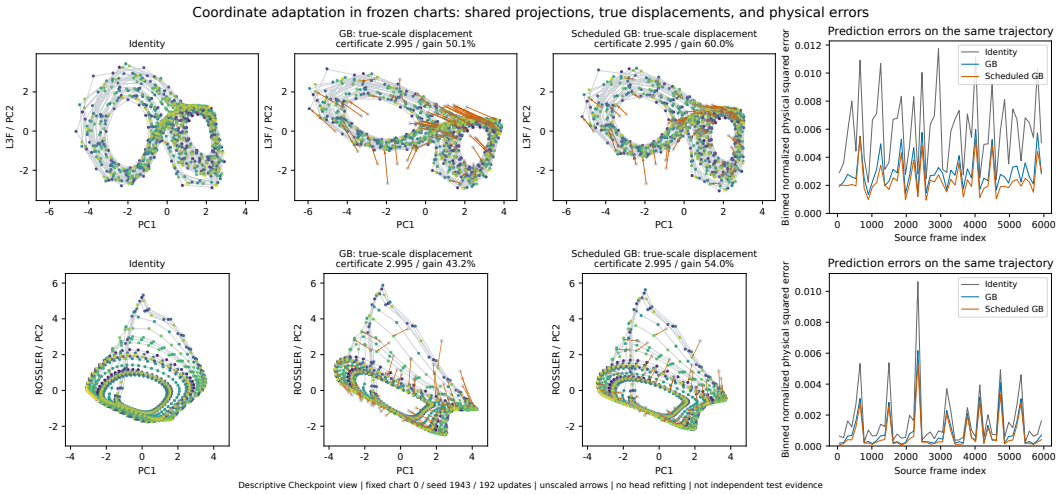

Figure 4: Current frozen-chart coordinate views for both systems. Axes, time colors, sample indices and PCA basis are shared across maps within each system. Arrows retain their actual projected magnitude; right panels average squared errors in 50 fixed time bins, not confidence intervals. These are descriptive historical-Checkpoint views with no new head fitting.

Table 28: Silverbox receiver-matched retraining, physical RMSE in mV. Each learned row retains all three seeds. All six candidates pass Selector, so candidate and selected-pipeline results coincide, including failures. "Fail" means a nonfinite full-record prediction, not a missing seed. Records were previously opened; no new independent measured record is added.

Receiver	Policy	Accepted	Multisine (mV)	Arrow (mV)
full	Identity	–	9.084	14.702
full	Candidate	3/3	4.932	fail (0/3 finite)
full	Selected pipeline	3/3	4.932	fail (0/3 finite)
shift	Identity	–	9.119	14.746
shift	Candidate	3/3	6.903	9.463
shift	Selected pipeline	3/3	6.903	9.463

gauges diverge despite finite Selector rollouts (Archive H1, Table 37). The coordinate guarantee survives these failures; a rollout guarantee was never established.

For physical delay state s and next known input u, the exact-shift receiver applies

$$T_H(s,u) = ([H^{-1} f_H(Hs,u)]_1, s_1, \ldots, s_{q-1}),$$

then re-encodes the simulated state. No future measured output enters its history. Full-vector propagation instead retains all predicted components. The fixed-gauge audit independently refits heads but never retrains gauges; all three Silverbox gauges fail finite exact-shift Checkpoint selection, whereas Wiener–Hammerstein exact-shift RMSE is .1375896 against .1568353 identity (12.27% lower). This changes the receiver class, not information.

Receiver-matched Silverbox retraining pairs seeds 1943/2943/3943, four width-64 couplings and common effective identity initialization across full/shift arms. It uses 60 AdamW updates, rate 3×10^{-4}, decay 10^{-6}, clipping one and GB cap 2.995 (never activated here). The original role intervals and ridge grid above apply; inner/outer use 32 windows of 64 steps, Checkpoint sixteen windows of 256. Final ridge profiles Development-to-Checkpoint, refits these roles for Selector, and additionally Selector for official evaluation. A 1% pooled Selector NMSE gate plus certificate/round-trip tests accepts all six; subsequent official failures cannot trigger identity replacement. No seed is omitted.

The independent fixed-coordinate numerical audit changes whitening floor from 10^{-8} to 10^{-10} or 10^{-6}, and compares normal/QR solves and scalar ridge with separately profiled heads, without retraining gauges. Normal and QR at 10^{-8} give the same recorded risks. At 10^{-10}, two Silverbox runs diverge after 10,457/15,238 finite outputs of 21,638, despite finite Selector predictions. At 10^{-6}, gains are 46.684% (Silverbox) and 10.484% (WH); scalar ridge gives 45.844%/5.644%. These compiled-evaluator audit scores do not overwrite the original table. All records were previously opened; technical repeats and windows are not independent physical replications.

Table 29: Protocol map. Head tuning is independent for each coordinate. Cov denotes covariance-aware Q2 ridge; Diag denotes diagonal ridge. Later Report sets only score frozen decisions; the original fresh-holdout rule could veto a frozen candidate (Appendix H).

Study	Search and final fitting	Evaluation unit
Original efficacy	LB, Cov; measured heads refit after Selector	Trajectories / one measured record
Matched search	LB/GB/uncapped, Cov; final fit before Selector	Three charts; new trajectories
Physical-loss control	Fixed GB maps; Q2 coefficient refit, Checkpoint selection	Fixed candidates; new trajectories
Flow adaptation	GB, Diag; fixed training ridge, final tuning before Selector	Three gauge seeds; one time block

The reported measured-update and whitening-floor protocols and result tables are in Archives H6 and H1 (Table 37); these diagnose already opened records, not new confirmation. The original full-vector and exact-history-shift receivers are different model classes, not interchangeable implementations.

H Study Protocols and Risk Estimators

The studies address incremental utility, search design, fitting objectives and receiver structure under separately specified protocols. The flow extension adds a frozen convolutional interface and prospective temporal holdout (Appendix I.1). The studies use separate frozen protocols; their risks are compared within each study, not pooled across incompatible charts or metrics. Original and structural studies use LB-PCEGA; GB is the default search implementation in the later learned-chart comparisons. The original fresh-holdout rule could reject one frozen candidate without promoting a runner-up. Later Report-only studies decide acceptance on Selector and never revise it on Report. These are distinct protocols. Learned-chart NMSE and NMAE normalize squared and absolute errors by the corresponding per-channel-centered target deviations; measured primary scores are physical RMSE. Original synthetic gains average trajectory-wise relative effects, whereas later search gains are ratios of mean risks. The configuration appendix distinguishes interfaces and independent units; additional trajectories do not increase the number of independently trained charts.

I Coordinate Adaptation for High-Dimensional Flow Observations

Data and evidence role. This development study uses the locally retained Tu–Rowley–Kutz–Shang cylinder PIV record, with 8,000 frames sampled at 20 Hz and two velocity components on a 135×80 grid. The retained data manifest identifies the source as https://zenodo.org/records/20765567. Only the pre-existing train block (frames $[0, 4000)$) and Checkpoint block ($[4096, 5296)$) are accessed. Both are historical development data; the reserved Selector and Fresh blocks were not accessed in that development phase. Their subsequent use is specified in Appendix I.1. This is a new CNN-chart development branch, separate from the earlier POD admission study; it does not revise that study's stopping decision. The development effects below condition on one measured record and technical seed 1943, without an independent-test or statistical-significance claim.

Frozen convolutional charts. Inputs retain all 21,600 observed values per frame. The CNN internally zero-pads height to 144 and crops its output back to 135; the evaluation does not spatially downsample the fields. A per-pixel mean field and channelwise fluctuation scales are fitted on train frames $[0, 2800)$. The encoder has four stride-two convolutional layers with kernel four and channel widths $16, 24, 32, 48$, SiLU activations and a linear latent projection. The decoder uses a linear expansion and four transpose convolutions. Separate 16D and 32D autoencoders receive 1,500 Adam updates at learning rate 10^{-3} and batch size 64. Internal validation on train frames $[2896, 4000)$ selects updates 1,000 and 1,250, respectively. Both encoder–decoders are then frozen. Checkpoint reconstruction MSE normalized by total target-field energy is .009082 and .008141; normalized instead by fluctuations relative to the training mean field, it is .204713 and .183508. These reconstruction errors are references, not lower bounds on forecasting error.

Receiving model and coordinate search. Direct prediction horizons are fixed at 32 and 64 frames (1.6 and 3.2 s). This adaptation uses complete quadratic features, coordinate-specific feature-input standardization and diagonal ridge on nonconstant coefficients, rather than the covariance-regularized fitting convention of the original synthetic studies. The ridge grid is $\{10^{-6}, 10^{-4}, 10^{-2}, 1\}$; the intercept receives a numerical 10^{-10} diagonal term. Source and target states use the same latent centering and scaling within each fitted head. Four

Table 30: Cylinder PIV reserved-time-block evaluation. All three optimization seeds are retained in every setting. Errors are fluctuation-normalized physical MSE; reductions compare the mean of seed-level risks with identity. The three rightmost entries give individual-seed MSE reductions.

d	h	Identity Q2	GB–Q2	Reduction	Seed 1943	Seed 2943	Seed 3943
16	32	0.313452	0.306129	2.336%	2.235%	2.315%	2.460%
16	64	0.457715	0.443818	3.036%	3.605%	2.759%	2.744%
32	32	0.318349	0.309467	2.790%	3.030%	2.707%	2.632%
32	64	0.432661	0.424322	1.927%	2.339%	1.848%	1.595%

additive coupling blocks of width 64 are optimized for 96 updates with AdamW, learning rate 6×10^{-4}, weight decay 10^{-5} and gradient clipping at five. GB retracts the composed bound to 2.995; the uncapped arm has the same initialization and update budget without this retraction. All candidates, including identity at step zero, are retained.

Head support pairs are sampled every two frames within train $[0, 1800)$. Ridge is selected at identity on inner pairs within $[1896, 2800)$ and fixed during coordinate optimization, which differentiates decoded physical error through the fitted head on batches of 32 inner source states. Every 24 updates, internal validation pairs within $[2896, 4000)$, sampled every four frames, select a checkpoint and independently tune ridge. These heads use all eligible pairs within train $[0, 2800)$; the final head is also fitted on this block with coordinate-specific tuning. Every source–target pair stays entirely within its assigned block. No Selector acceptance rule is applied in this development pilot.

I.1 Three-seed evaluation on a reserved time block

Following the development study above, we froze the CNN charts, normalization, four dimension–horizon settings and 96-update GB recipe before opening the remaining data blocks. The protocol retains the existing seed-1943 candidate in each setting and repeats gauge optimization with seeds 2943 and 3943, yielding twelve candidates in total. These are optimization repetitions conditional on the two frozen CNN charts, not independently trained charts or independent flow runs. The Q2 fitting and coordinate training rules are unchanged from this flow development branch.

Selection and temporal holdout. Before any Selector access, all candidates and independently tuned identity heads were frozen and hash-bound. All twelve coordinate maps passed the declared numerical and geometric admission checks: finite outputs, composed bound at most three and sampled inverse round-trip MSE at most 10^{-10} on development inputs. This CNN-branch admission does not assert near-exact AE reconstruction or revise the earlier POD stopping decision. The reserved Selector contains original frames $[5392, 6592)$. Each candidate is accepted only if its full-block physical MSE is at least 1% below its independently fitted identity Q2 reference; otherwise that seed returns identity. All twelve candidates pass. Decisions and model hashes are recorded before the final Report block, frames $[6688, 8000)$, is accessed. No head is refitted after selection. Source–target pairs remain within their respective blocks, with the original 96-frame gaps between roles.

This design evaluates a previously unused time block of one continuous measured record, rather than an independent experimental run. The retained 1,312 frames yield 1,280 and 1,248 direct-prediction pairs for horizons 32 and 64, respectively. Adjacent frames and the three technical seeds are not treated as independent physical replications; no independent-frame confidence interval or significance claim is made.

Results and estimand. Table 30 reports physical MSE normalized by target fluctuations around the fixed training mean field. Procedure risk averages the three seed-level MSEs, including identity fallback where prescribed; it is not the error of a prediction ensemble. Every seed improves Report error in every setting. The procedure-mean reductions are 2.34%, 3.04%, 2.79% and 1.93%, respectively. All candidates are accepted here, so raw candidate and full-procedure means coincide. The largest composed coordinate bound is 2.994996888. On the development admission inputs, the maximum inverse round-trip MSE is 3.33×10^{-13} and the maximum reconstruction drift in channel-normalized field coordinates is 2.60×10^{-14}. The analytic bound certifies coordinate geometry; sampled checks verify numerical consistency of the interface, not recursive flow stability.

J Affine and Physical-Loss Controls on the Flow Representation

These development controls reuse the fixed CNNs and opened training/Checkpoint blocks, not Selector/Report. All four settings and seeds 1943/2943/3943 remain. Affine $H(z) = c + A(z - c)$ uses the first 2,800 frames for c, identity A, and SVD clipping to $[1/2.995, 2.995]$. Both arms share batches, AdamW 6×10^{-4}, decay 10^{-5}, clipping 5, fixed identity-selected training ridge, independent final ridge grid $\{10^{-6}, 10^{-4}, 10^{-2}, 1\}$, and

internal validation every 24 updates including zero. A fixed 384-update extension retains 96/192/384 prefixes, not a result-dependent extension.

Table 31: Complete flow development budget matrix. Reductions are percentages of mean physical MSE relative to identity. Selected steps list all three seeds; prefix selection uses internal validation, not the scored block.

d/h	Budget	Affine	Nonlinear	Affine steps	Nonlinear steps
16/32	96	0.313	1.343	96/96/96	72/72/72
16/32	192	0.596	1.343	192/192/192	72/72/72
16/32	384	1.134	1.343	384/384/384	72/72/72
16/64	96	1.176	0.553	96/96/96	96/96/96
16/64	192	0.650	0.553	144/144/144	96/96/96
16/64	384	0.650	0.553	144/144/144	96/96/96
32/32	96	2.247	2.164	96/96/96	48/48/72
32/32	192	2.283	2.164	120/120/120	48/48/72
32/32	384	2.283	2.164	120/120/120	48/48/72
32/64	96	1.585	0.686	96/96/96	48/48/48
32/64	192	1.521	0.686	120/96/120	48/48/48
32/64	384	1.521	0.686	120/96/120	48/48/48

The 16D/32 nonlinear advantage over affine contracts from 1.033% to .211% at 384 updates; this concerns finite search, not optimal class risks. Training-loop times span 6.41–7.77 seconds for affine and 23.02–26.93 for nonlinear, not equal compute or inference timing.

For physical-loss controls, freeze the 384-budget selected maps. Compile affine into native Q2 coefficients, so affine and identity share the same standardization and function class. All heads fit sources $[0, 2800 - h)$; validation uses $[2896, 4000 - h)$ at stride four. Only the complete Q2 coefficients are updated through fixed decoder/inverse, with correction ΦTB, $\Phi = U\Sigma V^\top$ and $T = V \operatorname{diag}(\sqrt{n}/\max(\sigma_j, 10^{-6}\sigma_{\max}))$. B starts at zero; coefficients/SVD are float64 and CNN interfaces float32. Physical loss uses fixed training-target fluctuation energy and no penalty. Paired Adam batches of 64 use rates $10^{-4}/10^{-3}$, clipping 5 and 192 updates. Validation chooses rate and step $0, 24, 48, 96, 192$, without refit. All 72 paths/36 candidates remain, including 16 choosing zero.

Table 32: Flow development controls after matched physical-loss Q2 fitting. All entries are mean physical-MSE reductions relative to the original unrefitted identity, averaging three optimization seeds on the opened Checkpoint block. These are not the held-out-time-block results.

d	h	Identity + refit	Affine + refit	Nonlinear + refit
16	32	0.000%	1.134%	1.343%
16	64	5.508%	4.796%	5.176%
32	32	1.604%	3.277%	3.302%
32	64	0.140%	1.626%	0.686%

At 32D/32, affine/nonlinear retain 1.700%/1.725% gains versus equally refitted identity; at 16D/64 identity refitting is best. Thus useful adaptation does not establish nonlinear necessity or global optimality. Affine polynomial closure permits a native Q2 deployment without an extra module; the maximum compilation discrepancy on evaluated inputs is 3.60×10^{-14}, not a global floating-point or latency guarantee.

K MLP-SPECIFIC TRAINING-TIME ADAPTATION

This is development evidence on the same fixed six-dimensional charts, not new confirmation. Four of six historical trajectories fit models, two validate, and three already-opened Checkpoint trajectories score the comparison; Selector/Report are not accessed. Direct $h = 12$ uses at most 2,048 fitting and 1,024 validation/evaluation pairs per trajectory. All branches optimize decoded physical MSE with the same normalization. Three seeds profile two-layer residual SiLU MLP widths 64/128 and rates .0003/.001, 8,000 Adam steps, batch 256, no decay, saving every 1,000 steps including zero. The selected base head initializes each extra arm.

Extra head fitting gets 8,000 updates; affine/nonlinear joint adaptation gets 2,000, resetting Adam at one quarter of the selected head rate, coordinate rate .0002, clipping 5 and selection every 250 steps including identity. Affine

singular values are projected to $[1/2.995, 2.995]$. The nonlinear map has two alternating $3 \to 32 \to 3$ tanh shears, zero output initialization and composed-bound retraction below 2.995. Scalar normalization preserves the Euclidean certificate. This is joint supervised training through the fixed decoder, not the closed-form profiled algorithm.

Table 33: MLP-specific development comparison. Positive Δ means lower NMSE than extra head fitting; costs are mean additional training-loop seconds, not full pipeline costs.

System	Branch	NMSE	NMAE	Δ (%)	Seconds
L3F	Original	0.000173	0.011487	-120.76	—
L3F	Extra fitting	0.000079	0.007865	+0.00	120.5
L3F	Affine	0.000114	0.009437	-45.12	40.5
L3F	Nonlinear	0.000106	0.009101	-35.20	46.5
Rössler	Original	0.000072	0.006113	-67.20	—
Rössler	Extra fitting	0.000043	0.004780	+0.00	122.5
Rössler	Affine	0.000056	0.005285	-29.05	41.0
Rössler	Nonlinear	0.000056	0.005295	-29.49	47.0

Joint nonlinear adaptation improves the original head by 38.76%/22.56%, but longer extra fitting is better in both systems. Its fourfold update budget and greater time prevent an equal-cost ranking. Base search is not converged: extending 4,000 to 8,000 steps lowers validation loss by 53.5–58.5% (L3F) and 43.4–55.0% (Rössler).

For the new-head control, freeze each selected map, discard its joint head, and use source seed plus 1,000,000 for a new identity/affine/nonlinear head with paired batches and the same selected width, rate and 8,000-step budget. There are three different maps with one new head each, not three heads per map. Affine conjugacy can be absorbed by unrestricted affine first/last MLP layers and need not expand its hypothesis class.

Table 34: Fresh head initializations, but the same historically opened development data. Gains are relative to the matched newly initialized identity head.

System	Coordinate	NMSE	Gain (%)	Seeds improved
L3F	Identity	0.000161	+0.00	—
L3F	Affine	0.000154	+4.17	3/3
L3F	Nonlinear	0.000152	+5.32	3/3
Rössler	Identity	0.000072	+0.00	—
Rössler	Affine	0.000071	+1.80	1/3
Rössler	Nonlinear	0.000074	-2.31	1/3

Nonlinear transfer improves L3F in all three pairs but worsens Rössler on average; no uniform neural benefit or population-optimality claim follows. All zero-step choices remain included. Additional timings exclude base-grid and new-head fitting costs. Maximum selected bound is 1.121186 and sampled round-trip MSE 2.1×10^{-13}, not a decoded-robustness/rollout guarantee.

L Supervised Coordinate Adaptation for Protein Embeddings

L.1 Task and two-stage protocol

Frozen `esm2_t6_8M_UR50D` residue-mean embeddings retain all 320 dimensions without PCA or fine-tuning (Lin et al., 2023). The FLIP2 amylase archive (Didi et al., 2026) provides 2,574/644/488 train/validation/test examples. Reference-sequence checks identify multi-mutant training and single-mutant testing despite the filename `one_to_many.csv.gz`. Train-only centering, scalar normalization and label standardization are used; pretraining overlap was not audited. The predictor is $\widehat{y} = f_{\widehat{\psi}}(H_{\widehat{\theta}}(E(s)))$: an input-side supervised extension, not a conjugated scalar transition. Invertibility preserves embedding information; the coupling construction certifies coordinate distortion. There is no decoder, reconstruction test or application of the polynomial-dynamics normalizer theorem to these scalar labels.

Two additive couplings with $160 \to 32 \to 160$ tanh conditioners (20,864 map parameters, bound 2.99) train jointly with a linear head under MSE. Two donor weight-decay settings each receive 800 updates. Freeze the map, discard the donor, and independently fit a new $320 \to 64 \to 1$ tanh MLP: rates .0003/.001 crossed with weight decays .0001/.01, 1,200 updates each. This is joint donor training, not nested profiling. Seeds

Table 35: Complete raw-method amylase comparison under the five-repeat protocol. NMSE SD is across technical repeats; it is not a population confidence interval. MAE and Spearman entries are repeat means.

Method	NMSE	MAE	Spearman
Training-mean constant	2.0646	0.06821	—
Identity MLP	0.9800 ± 0.0467	0.04610	0.5238
Longer-trained MLP	0.9497 ± 0.0466	0.04518	0.5380
Wider MLP	0.9767 ± 0.0210	0.04601	0.5354
Certified map + MLP	0.8786 ± 0.0288	0.04124	0.5074
Uncapped shear + MLP	0.8919 ± 0.0567	0.04108	0.4987
Residual adapter + MLP	0.9186 ± 0.1195	0.04308	0.5087
Random certified map + MLP	0.9744 ± 0.0492	0.04564	0.5285
Ridge	1.3195	0.05428	0.4720
RBF kernel ridge	1.1279	0.05061	0.5591

5101/5201/5301/5401/5501 repeat the entire procedure with paired receiving initializations and minibatches. Validation selects configurations and may return the paired identity or training-mean constant; all five main maps are retained. Test labels were previously explored: this is development robustness, not new-test confirmation or five biological replications.

Identity has 20,609 parameters and 4,800 grid updates. Longer training uses 6,400 updates (matching 1,600 donor plus 4,800 head updates, not FLOPs/time). Width 128 has 41,217 parameters against 41,473 for map+MLP. Same-size uncapped shear, residual adapter and random certified map are retained; removing a guarantee does not establish noninvertibility. NMSE uses training-label variance and MAE archived label units. Effects are ratios of mean MSE, not means of percentage changes. Certified gains are 10.34% over identity (five wins, range 3.82–17.24%), 7.48% over longer fitting, 10.04% over wider MLP and 9.83% over random (five wins each). Raw advantages over uncapped/residual are 1.49% (four wins) and 4.36% (three), not a causal benefit of certification alone. Selected NMSE for uncapped/residual/random is .946629/.946953/.968583, with respectively 1/1/2 maps and 4/4/3 identity fallbacks. Spearman falls .5238 to .5074. Recorded CPU times (two single-thread workers, excluding ESM) are 6.50/8.15/8.00/11.68 seconds for identity/longer/wider/certified: descriptive, not a hardware-normalized speed advantage.

L.2 Original-component port with nested profiling

This separate protocol reuses the certified gauge, differentiable ridge solver and rotating folds: four masked $320 \to 64 \to 64 \to 320$ conditioners, 182,016 parameters and construction bound 2.995. Support fitting, inner ridge choice, support+inner refitting and outer supervised-loss gradients retain nested profiling, but replace decoded dynamics loss by scalar loss. MLP transfer is fitted only after the map is frozen.

Three seeds 5101/5201/5301 retrain all arms. Split seed 90217 divides the 644 validation cases into 322 Checkpoint/322 Selector. Four rotating folds train 96 float64 AdamW epochs at 2×10^{-4}, decay 10^{-5}, clipping 5, ridge $\{10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}, 1, 10^{2}, 10^{4}\}$. Three-fold training CV chooses final ridge before Checkpoint scoring; Selector compares candidate, paired identity and constant, without the dynamics trajectory/random gate. Fixed-ridge arms match architecture and epochs; the two-coupling reference is retrained, not architecture-matched. These baselines are not paired with the five-repeat table.

Table 36: Original-component supervised port on amylase: three technical repeats, separate checkpoint/selector protocol. Raw and selected risks are distinguished. These numbers must not be paired with Table 35.

Coordinate training	Receiver	Raw NMSE	Selected NMSE
Identity	Ridge	1.412568	1.412568
Nested profiling	Ridge	1.036196	1.036196
Fixed training ridge	Ridge	1.024171	1.024171
Two-coupling joint training	Ridge	1.027501	1.027501
Identity	MLP	0.966924	0.966924
Nested profiling	MLP	0.893811	0.920582
Fixed training ridge	MLP	0.895307	0.895307
Two-coupling joint training	MLP	0.863981	0.863981

Mean ridge NMSE falls by 26.64%, with three paired wins. MLP raw gain is 7.56% (three wins); one Selector identity fallback reduces pipeline gain to 4.79%. Selected bounds are 2.9809–2.9816. Fixed ridge slightly outperforms profiled ridge and raw MLP transfer differs by only .17%; the simplified joint reference is also better on raw MLP. Fidelity to the original component does not establish superiority of its training recipe.

L.3 Additional settings and evidence boundaries

With 512 independently sampled training cases in each of five repeats, raw/selected amylase gains are 2.88%/2.82% (four raw wins; three maps/two fallbacks). A custom forward split uses 391/98 train/validation variants with at most one mutation and 3,217 multi-mutant tests: raw error worsens 2.71%, while selection improves 1.63% (two maps/three fallbacks). It does not replace the official archive. Rhomax cross-backbone uses 584/116/184 cases; all five pipelines select the constant, with NMSE 1.079851, equal to selected identity. The separate three-repeat original port also selects the constant throughout. Thus utility is conditional on amylase, not general transfer, forward extrapolation or ranking improvement. The existing numerical reports accompany the source package; no new runs or retrospective upgrade of the explored test split is implied.

M Archived Development and Reproducibility

This focused appendix retains the proofs, protocols, principal results and controls supporting the main text. The separate historical archive preserves the unabridged main54 manuscript and its source, including older development branches, numerical replay ledgers, exploratory budgets, Known-conjugacy paths, measured update audits and additional plots. These materials are not required to interpret the main results and are not additional independent confirmation. No result is relabeled as prospective by this editorial separation. Every "Archive H" locator below refers to the supplied standalone file `history/main54_full.pdf`, not to this appendix's page numbers. Its editable source and the same PDF also appear inside `history/main54_source.zip` as `main54_source/main.tex` and `main54_source/main.pdf`. These files accompany the manuscript source; they must also accompany any submission that relies on the archived analyses. The archived paper contains reported analyses and tables, not every raw prediction file or execution log; those belong to the separate experiment artifact.

Table 37: Locators in the accompanying `history/main54_full.pdf`. Page and table numbers refer exclusively to that unchanged archive.

ID	Archived material	Exact locator
H1	Numerical fitting, full-training and measured-floor sensitivity	C.7–C.9, pp.29–33; Table 19, p.33
H2	Earlier search diagnostics and matched checkpoint curves	C.3–C.6, pp.25–29; C.11, pp.35–36
H3	Original effects, ridge/affine controls, atlas and mechanism diagnostics	D.1, pp.36–38; F.2–F.3, pp.56–58
H4	Input/output-factorial table and known-chart retraction	E.3, pp.45–46; Table 32, p.45
H5	Full-budget shuffled-donor comparison	E.6, Table 38, p.50
H6	Receiver-matched retraining and fixed-gauge history-shift audit	E.8–E.9, pp.52–54; Tables 42–44
H7	Measured gates, official trace and stopped-branch history	B.3, pp.19–20; F.3, Figure 12, p.58; F.5, pp.59–60
H8	Budget diagnostics and additional exploratory settings	E.7, p.51; F.1, pp.54–55; F.7–F.8, p.60

Retained scope facts include a stopped certified-v2 omnibus, fixed-budget MLP identity fallback, KSE and C. elegans identity returns, and a previously stopped POD-flow admission branch distinct from the later CNN branch. The Wiener–Hammerstein broader-protocol STOP preceded its separate locked PCEGA-only evaluation. Known favorable coordinates do not establish successful search; decoder compensation does not certify fitted recursive stability. Post-confirmatory near-OLS L3F diagnostics gave 14.20% global versus $.11\%$ 12-patch physical gains and 16.28% lower latent sharing burden; these are different metrics, not additive components of physical error. Input/output-factorial development analyses likewise do not replace the independent input-only crossed test retained above.

Reproducibility statement. The manuscript source is independently compilable but is not the complete training-data/checkpoint artifact. The full study archives retain source, protocol, data-role and checkpoint hashes,

selected and rejected cases, and execution incidents. Historical numerical checks do not establish universal cross-platform reproducibility. No new training or hypothesis testing was performed to shorten the appendix.