SAME FUNCTION, DIFFERENT GEOMETRY: CHOOSING OPTIMIZATION COORDINATES FOR GAUSSIAN REPRESENTATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Optimization coordinates are part of a learning algorithm, even when they leave its represented functions unchanged. We investigate their practical value in Gaussian representations, where appearance is linear at fixed geometry but joint fitting is nonlinear. Exact appearance absorption and transported Adam/AdamW controls isolate coordinate effects from capacity. We construct energy-normalized coordinates that compensate appearance as shape changes, cancelling continuous single-atom shape–appearance coupling. With the same five-profile iRprop+ calibration allowance for peak and energy coordinates, a prospective study on 64 fresh images yields mean RGB PSNR gains of 1.242 dB over a frozen historical incumbent and 0.939 dB over independently calibrated peak coordinates, averaged across attained 2, 5 and 10-second output budgets. Both medians are positive and mean LPIPS decreases; a complete repaired-arithmetic replay retains these gains. Benefits extend to three of four domains; Aircraft and arithmetic-mean error against calibrated peak remain limitations. Timings measure warm output delivery with monitored host activity. Separately, fixed output measurements yield a sharp, separation-independent angle law and first-order minimax certificate for moving one-dimensional Gaussian clusters. Prospective selection and optimizer-payoff bounds distinguish useful geometry from useful computation. Together, these results make coordinate choice a testable design decision rather than an inherited implementation detail.

1 INTRODUCTION

Representation design determines both the functions a model can express and the coordinates its optimizer updates. Changing the latter can alter learning while preserving every paired prediction. Preconditioning explains this possibility; the practical question is sharper: *does a coordinate design remain useful after fair optimizer calibration and accounting for its computational cost?*

Gaussian representations provide a controlled setting (Kerbl et al., 2023; Zhang et al., 2024; Tai et al., 2025). Rendering is linear in color at fixed geometry, but image fitting jointly learns both. We first absorb a fixed appearance map into optimization-aware (OA) coordinates; optimizer transport identifies the changed dynamics. Our nonlinear energy chart instead compensates color as Gaussian shape changes. This cancels continuous single-atom shape–appearance coupling without whitening every interaction in an image.

Our strongest test gives both charts the same optimizer and five-profile calibration allowance. On 64 fresh images, energy gains 0.939 dB over calibrated peak and 1.242 dB over the frozen incumbent, averaged across output budgets. A complete replay retains these gains after backward-arithmetic repair. Three domains improve; Aircraft remains adverse. A matched-profile control separates the coordinate intervention from rate selection. The benefit survives calibration and the cost of producing the image.

Contributions. Our practical contribution is an explicit nonlinear Gaussian-coordinate recipe evaluated under prospective, symmetric calibration and delivered-image budgets. Separately, we give a sharp fixed-measurement characterization of moving Gaussian clusters. Exact optimizer

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.

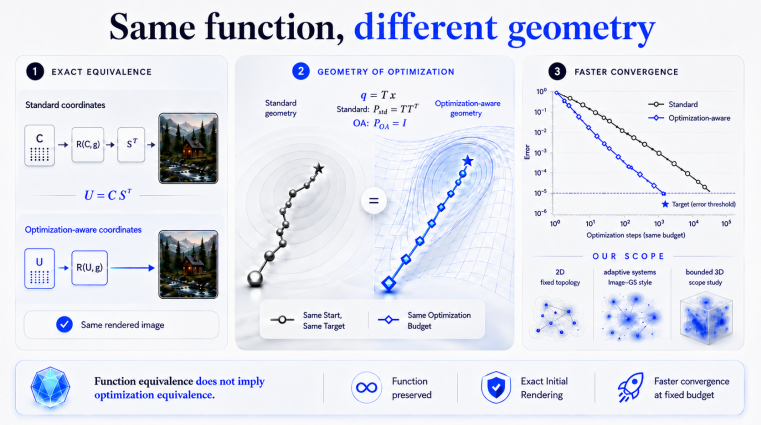


Figure 1: **Same function, different optimization.** Linear appearance absorption preserves predictions while changing Euclidean update geometry. Curves are illustrative, not measured convergence or a native-backward theorem; the 3D icon denotes a scope study. The nonlinear energy study separately tests shape-dependent color compensation under matched initialization, symmetric calibration and actual output budgets.

transport and classical model-decrease bounds specify what these results establish about computation. Shared-Gradient Chart Lookahead (SGCL) tests the information and cost of prospective appearance-policy selection. The two distinct constructions address complementary questions: energy normalization controls an atom's self-coupling, while fixed measurements control interactions in the common-width cluster model.

2 Exact Coordinates and Optimizer Transport

Appearance absorption. A *chart* is an invertible map from optimizer-facing parameters to physical renderer parameters. Let $C \in \mathbb{R}^{N\times d}$ contain Gaussian appearance coefficients, g the geometry, and A fixed image offsets. For a fixed invertible decoder map S, appearance linearity gives

$$\widehat{Y}_{\mathrm{std}} = A + \mathcal{R}(C, g)S^{\top} = A + \mathcal{R}(U, g) = \widehat{Y}_{\mathrm{OA}}, \qquad U = CS^{\top}. \tag{1}$$

The bijection preserves every paired rendering, capacity, attainable loss, and physical initialization. We also map backgrounds, shifts, and constraints. The discovery scale $S = \mathrm{diag}(s_Y, s_{Cb}, s_{Cr})$ contains per-channel ranges of the learned spatial PCA basis in luminance/chrominance (YCbCr) coordinates. It is not the standard deviation or covariance of training-image pixels. Although $\kappa_2(S)$ is only 1.039 to 1.193, appearance is small relative to unchanged geometry: $\kappa_2(T)$ is 24.5 to 30.3 for

$$q = Tx, \qquad T = \mathrm{diag}(S \otimes I_N, I_g), \quad q = (\mathrm{vec}\, U, g), \quad x = (\mathrm{vec}\, C, g). \tag{2}$$

These are map condition numbers, not measured Hessian condition numbers. A global learning rate cannot remove this block ratio while leaving geometry unchanged.

Ordinary gradients and supplied covectors. For an ordinary loss gradient on a smooth active set, a linear chart induces

$$\nabla_x F(Tx) = T^{\top}\nabla_q F(q), \qquad \Delta q_{\mathrm{std}} = -\eta TT^{\top}\nabla_q F, \qquad \Delta q_{\mathrm{OA}} = -\eta \nabla_q F. \tag{3}$$

The original native projector supplies a modified shape covector, not generally the ordinary derivative. Its fixed appearance-chart *transport algebra* still holds for consistently transformed supplied covectors, whereas gradient-descent and Hessian theorems require true derivatives. Energy comparisons share an algebraic conversion to the ordinary pullback. Appendix J.10 distinguishes this conversion from a subsequent arithmetic-precision repair.

Adaptive optimizer equivalence. For $q_i = t_i x_i$, $t_i > 0$, Adam moments transform as $m_i^x = t_i m_i^q$ and $v_i^x = t_i^2 v_i^q$. OA Adam with (α, ϵ) is realized in standard storage by rates α/t_i, stabilizers

$t_i\epsilon$, and mapped moments. AdamW also requires equal rate-decay products (Kingma & Ba, 2015; Loshchilov & Hutter, 2019). In particular, a scalar appearance chart $U = sX$ with chart settings $(\eta, \epsilon, \lambda)$ has equivalent physical settings

$$\alpha_U = \eta s, \qquad \epsilon_U = \epsilon/s, \qquad \lambda_U = \lambda/s. \tag{4}$$

Thus scalar chart selection is block-optimizer-policy selection, not a new source of expressivity. Dense and shape-dependent charts need not reduce to such changes. Exactness is algebraic; float32 rendering, clipping, schedules, topology and unmatched state can separate trajectories. Appendix A supplies proofs and finite-horizon counterexamples.

3 Coordinate Constructions and Their Limits

We distinguish four interventions: fixed appearance absorption, energy normalization, fixed output measurements, and finite-horizon appearance-policy selection. The appendix roadmap summarizes their evidence and scope.

3.1 An energy-normalized shape and appearance chart

Write $L = \left[\begin{smallmatrix} a & 0 \\ b & c \end{smallmatrix}\right]$ and $\Sigma = LL^\top$. For initial factor (a_0, b_0, c_0) with $a_0, c_0 > 0$, absolute units A give

$$a = a_0 e^{u/a_0}, \quad b = b_0 + v, \quad c = c_0 e^{w/c_0}, \qquad C = hQ, \quad h = \sqrt{\frac{a_0 c_0}{ac}}. \tag{5}$$

The relative-unit convention R instead uses $(a, b, c) = (a_0 e^u, b_0 + \sqrt{a_0 c_0} v, c_0 e^w)$ with the same h. Here $C, Q \in \mathbb{R}^3$ are an atom's physical and optimized color. At $(u, v, w) = 0$, $Q = C$ preserves its initial field; the map is invertible throughout the positive-diagonal chart. With Q fixed, changing shape preserves the atom's continuous whole-plane squared energy, since that energy is proportional to $ac\|C\|^2$.

For one untruncated Gaussian, write the self-Gram matrix of output derivatives as $G = \left[\begin{smallmatrix} A & B \\ B^\top & D \end{smallmatrix}\right]$, partitioned into geometry and color. Differentiating its norm gives $D^{-1}B^\top = C(\nabla \log(ac))^\top/2$. With physical geometry as the first block, the appearance-coupling Jacobian $T_h = \left[\begin{smallmatrix} I & 0 \\ -D^{-1}B^\top & hI \end{smallmatrix}\right]$ satisfies

$$T_h^\top G T_h = \begin{bmatrix} A - BD^{-1}B^\top & 0 \\ 0 & h^2 D \end{bmatrix}. \tag{6}$$

The positive-shape map in Eq. (5) adds a block-diagonal Jacobian and preserves the zero cross block. Energy normalization removes this self-coupling, leaving the geometry Schur complement, *not its inverse*. This is classical normalization and block algebra, not full-image whitening. Finite pixels, boundaries, overlap, and residual curvature are outside the identity.

Implemented recipe. Each domain's frozen profile selects A/R units and initial center/shape/color step widths from the same five-profile menu. Starting from shared physical parameters and fresh history, peak uses $C = Q$ and energy uses $C = hQ$. The algebraic projection-convention correction precedes the chart chain rule; both charts use the same loss-aware iRprop+ update and whole-atom numerical guard. Initialization, attempted updates, guards and CPU-image delivery are charged. Appendix R.8 gives the full recurrence, history restoration and deadline rule. This is coordinate-wise resilient backpropagation, not natural Rprop or exact color minimization (Riedmiller & Braun, 1993; Igel & Hüsken, 2003; Igel et al., 2005).

3.2 Fixed output measurements control interactions

For $f(g, c) = \Phi(g)c$, fixed appearance-only whitening changes units inside the appearance tangent space, but not its canonical correlation with geometry. Instead, fix output measurements L and define

$$q = L\Phi(g)c, \qquad f = \Phi(g)[L\Phi(g)]^{-1}q. \tag{7}$$

This is exact wherever $L\Phi(g)$ is invertible. The measurements remain fixed during differentiation: shape changes now induce a compensating appearance motion. Such cardinal coordinates are classical (Maday & Mula, 2013); the following result quantifies their geometry on a full moving-center domain.

Theorem 1 (sharp Gaussian cluster certificate). For $b > 0$, let n distinct one-dimensional Gaussian centers of common fixed width σ lie within $\sqrt{2}b\sigma$ of a fixed anchor, with independent centers and arbitrary nonzero vector appearances. Take $L = H_n^*$, the first n orthonormal Hermite measurements. This is an invertible appearance chart. If $\Delta_n(b)$ is the worst largest-principal-angle sine between the Gaussian span and H_n, then

$$\frac{\sqrt{n}\,b}{\sqrt{1+nb^2}} \le \Delta_n(b) \le \min\{1, \sqrt{n}\,b\}, \qquad \rho \le \Delta_n(b), \qquad I \preceq A^*A \preceq \frac{I}{1-\Delta_n(b)^2}, \quad (8)$$

where $A = \Phi(H_n^*\Phi)^{-1}$, ρ is geometry-appearance canonical correlation, and $\Delta_n(b) < 1$ for finite n, b. The exact worst angle is the limit as distinct centers approach one endpoint. Among all fixed n-dimensional probe spaces, the optimal worst-angle certificate is $\sqrt{n}\,b + O_n(b^3)$ as $b \to 0$.

The bounds have no minimum-separation, amplitude-ratio or channel-count factor. They do not remove geometry rank loss near collisions. A polynomial-remainder decoder and its learned-center derivatives avoid inverse-gap coefficients; the exactly coincident dictionary is excluded. Appendix N supplies the proof and decoder. The useful regime is $\sqrt{n}\,b \ll 1$: for example, $n = 4, b = 0.1$ certifies $\rho \le 0.2$ and $A^*A \preceq 1.042I$, without a separation floor. Geometry can still become ill-conditioned near collisions. Unlike generalized interpolation and RBF-QR/HermiteGF stabilization (Maday & Mula, 2013; Fornberg et al., 2011; Kormann et al., 2019), the result characterizes a worst angle over the full moving-center domain, its endpoint extremizer, and its leading fixed-probe minimax constant. It is not an optimality theorem for arbitrary charts, anisotropic widths, or image-fitting optimizers.

Why a local construction is necessary. For a general linear head, zero coupling requires $\partial_i C = -K_i C$ with $K_i = (\Phi^*\Phi)^{-1}\Phi^*\partial_i\Phi$. Nonzero connection curvature obstructs this on an open color set. Appendix K gives a sharp regional minimax in a full-rank linear-head family, including nonlinear charts, an odd/even head classification, and a two-chart pointwise escape. These are limits of a retained geometry/appearance split, not of unrestricted joint optimization.

3.3 From geometric improvement to optimizer payoff

At an identifiable exact fit, optimal dense tuning within the two tangent blocks leaves condition number $(1+\rho)/(1-\rho)$ and stationary block-gradient local factor ρ (Elsner, 1984; Lin & Sinnamon, 2012). But small coupling is not a finite-step guarantee for Adam or Rprop. For a nonzero loss gradient, after exact scalar calibration of a block-whitened gradient, the joint Gauss–Newton (GN) to block model-decrease ratio satisfies the sharp classical bound

$$1 \le \frac{d_{\text{joint}}}{d_{\text{block}}} \le \frac{1}{1-\rho^2}. \quad (9)$$

Both endpoints occur at the same geometry with different residuals (de Klerk et al., 2017). At a fresh fixed-measurement anchor, exact separate block solves produce the same tangent as Schur GN; with the same finite realization, they produce the same next state. Appendix O proves this identity and actual loss-decrease/cost bounds. These least-squares comparisons delimit the computational opportunity from fixed measurements. The energy study separately evaluates guarded first-order iRprop+ rather than these solves.

4 Prospective Coordinate Selection

Shared-Gradient Chart Lookahead (SGCL) chooses a fixed appearance chart from a small library for a specified horizon. For $U = XB^\top$ and an ordinary physical appearance gradient G_U, the fresh Adam/AdamW candidate update is

$$\Delta U_B = \left[-\eta \frac{G_U B}{|G_U B| + \epsilon} - \eta\lambda U B^{-\top}\right] B^\top, \quad (10)$$

with elementwise division and $\lambda = 0$ for Adam. In deployment, we capture the native backward covector once and transport its appearance block; this calculation is exact for that input, not a certificate of the native shape derivative. Geometry's first update is shared.

Forward-only renders preview each candidate. All candidates share proposed geometry, permitting exact projection/support reuse (Appendix Q). An offline-fitted *ridge regression* predicts its 100-step gain over $4I$ from candidate loss, scale, rendered block balance, gradient balance, optimizer identity, and shape descriptors. A frozen margin rule selects a chart or falls back to $4I$. We install the selected chart at the original state and reuse the captured covector for its first real update, followed by 99 ordinary updates. No candidate trajectory is trained and discarded, but previews cost time.

The eight choices comprise $I, 4I, 8I, 16I$, three render-balance scalar rules, and one scale-8 gradient-eigen rotation. Development uses 48 images from eight domains. The frozen rule accesses neither domain identity nor endpoint outcomes. It chooses scalars in 192/192 initial and 254/256 new-domain runs: primarily scale selection, not the pointwise adaptive escape of Appendix K. Appendix G gives full specifications.

5 Experimental Evidence

Distinct protocols. Coordinate pairs match physical initialization, renderer, loss, and capacity; practical comparisons may independently select optimizer policies. The original six-domain study uses 256×256 targets, 2,000 Gaussians, and 500/60/200 training/validation/held-out images per domain. Its PSNR is *unclamped normalized YCbCr*, not RGB. DIV2K's bicubic ×2 low-resolution track, evaluated at its native image sizes, uses 785 training images for the prior, 12 for validation, 3 for pilots, and 100 held-out official validation images; official test is untouched (Agustsson & Timofte, 2017). Native DIV2K, the separate production-controller study, SGCL, and both energy-chart confirmations report common encoded-RGB metrics. We do not pool the protocols. PSNR is peak signal-to-noise ratio, SSIM structural similarity (Wang et al., 2004), and LPIPS a learned perceptual distance (Zhang et al., 2018). Appendices B, G, I and R give selection, image roles, clocks and numerical paths.

5.1 Nonlinear coordinates after symmetric calibration

An earlier ordinary-Rprop energy study gave $+1.8482$ dB matched and $+0.5909$ dB practical gains, but only 28/64 practical wins and worse LPIPS (Appendix J). Here symmetric iRprop+ calibration uses a separate cohort (Appendix R). We retain its original B0 results before reporting the repair replay.

Both charts select from the same five-profile iRprop+ menu by domain-specific calibration PSNR averaged over three budgets. The 32 calibration images overlap prior training; equal menus do not equalize historical research effort. Confirmation uses 64 unused official-test images, 16/domain, with existing priors, 256×256 targets and 2,000 Gaussians.

Let E denote selected energy, P selected peak, and M peak at E's profile. Incumbent C retains per-domain/budget selections from 111 challengers and is rerun contemporaneously. $E - M$ changes only the amplitude exponent, including resulting dynamics and guard exposure; $E - P$ allows equal calibration. All 512 deployments complete across two repeats and 2/5/10-second CPU-delivery budgets, charging setup, updates and guards. Estimates average repeats/budgets within image and weight domains equally; intervals resample images within domains.

Both practical comparisons improve in three domains and the median image. Excluding GTSRB *descriptively* leaves $+0.5352$ dB over either control. Aircraft loses at every budget; late STL10 nearly ties. Mean PSNR and LPIPS favor energy, while SSIM intervals cross zero. Against P, arithmetic-mean raw fitting MSE and clamped-RGB MSE are 1.15% and 1.57% higher; both difference intervals include zero. PSNR averages log errors. Host-I/O bursts qualify these warm-delivery results, not clean-host or cold-start speedups. The two cohorts do not estimate a paired improvement.

Complete repair replay. All frozen policies, including C, were rerun on the same 64 images under original B0 and repaired B1 backward arithmetic. All 1,024 deployments complete. B1 retains $+1.230$ dB over C (95% interval $[1.028, 1.449]$) and $+0.926$ dB over P ($[0.700, 1.157]$), with lower mean LPIPS. The B1-minus-B0 $E - C$ contrast is $+0.007$ dB ($[-0.014, 0.028]$). This exposed-cohort replication is not new confirmation or an equivalence test. Aircraft remains adverse and arithmetic RGB MSE is 1.23% higher than P (Appendix T).

Table 1: Original B0 confirmation after symmetric calibration. Energy minus control on 64 images fresh at the original evaluation; original B0 arithmetic. The later same-cohort B0/B1 replay is separate (Appendix T). PSNR is in dB; higher PSNR/SSIM and lower LPIPS are favorable. Repeats and budgets stay paired within image. The primary $E - C$ PSNR interval is $[1.0424, 1.4576]$; the descriptive $E - P$ interval is $[0.7073, 1.1760]$. Budget-averaged absolute RGB PSNR is $44.2980/43.3586/43.0559$ dB for $E/P/C$, with LPIPS $0.08582/0.10265/0.09662$. Calibration is symmetric only for the current five-profile E/P menus. Aircraft and the qualified warm-timing boundary remain part of the result.

Control	ΔPSNR	ΔSSIM	ΔLPIPS	Median Δ PSNR	PSNR wins
Frozen incumbent C	$+1.2421$	$+0.00102$	-0.01080	$+0.6185$	47/64
Calibrated peak P	$+0.9394$	$+0.00060$	-0.01683	$+0.3991$	45/64
Matched profile M	$+1.1542$	$+0.00301$	-0.02307	$+0.7625$	53/64

	$E - C$ by output budget			Budget-averaged effects	
Domain	2 s	5 s	10 s	$E - P$	$E - C$ wins
Aircraft	-0.105	-0.191	-0.243	-0.180	6/16
GTSRB	$+3.697$	$+3.544$	$+2.847$	$+2.152$	16/16
STL10	$+0.305$	$+0.131$	$+0.041$	$+0.159$	9/16
SVHN	$+2.273$	$+1.457$	$+1.148$	$+1.626$	16/16

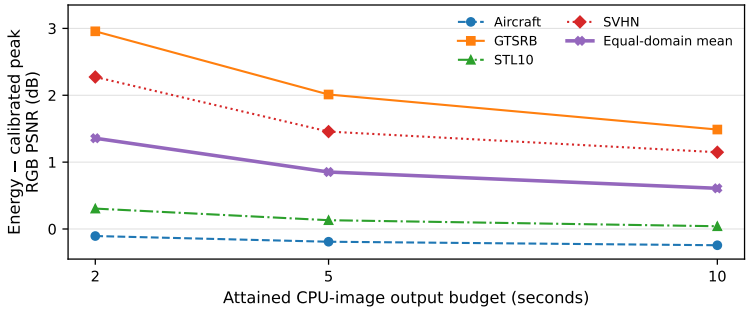


Figure 2: **Energy coordinates after equal current calibration.** Energy minus calibrated peak ($E - P$), with 16 paired images per domain and two executions at every budget. The pooled line weights domains equally. Aircraft remains adverse and late STL10 nearly neutral. These original-precision warm-delivery means complement the primary $E - C$ inference in Table 1. The complete B0/B1 replay is reported in Appendix T.

Fixed normalization is a different intervention. A four-image iRprop+ development check compares identity, inherited scaling, z-scoring and PCA/ZCA whitening using training-only YCbCr statistics and the same decoded loss. Inherited scaling and identity nearly tie after scalar calibration; energy retains a favorable mean within each tested frame. This is not a held-out SGCL/Adam comparison (Appendix P).

5.2 An exact intervention with broad measured effects

Independently selected OA/standard policies improve all 4,800 six-domain image-checkpoint comparisons; matched optimizer/rate pairs retain 2,398/2,400 endpoint wins in normalized YCbCr (Table 2). The 1,200 image means have median gain 11.228 dB and interquartile range $[6.639, 14.474]$ (Appendix C.3). On 100 native DIV2K images, RGB gain contracts from 1.3947 dB at step 100 to 0.3746 at 5,000 under independently selected, not optimizer-matched, policies (Figure 3). Capacity and richer-system results retain their declared comparisons (Figure 4, Appendix C).

Transported Adam/AdamW in standard storage differs from OA by at most 0.0108 dB in pooled endpoint RGB PSNR across four production protocols. Transport also restores the first appear-

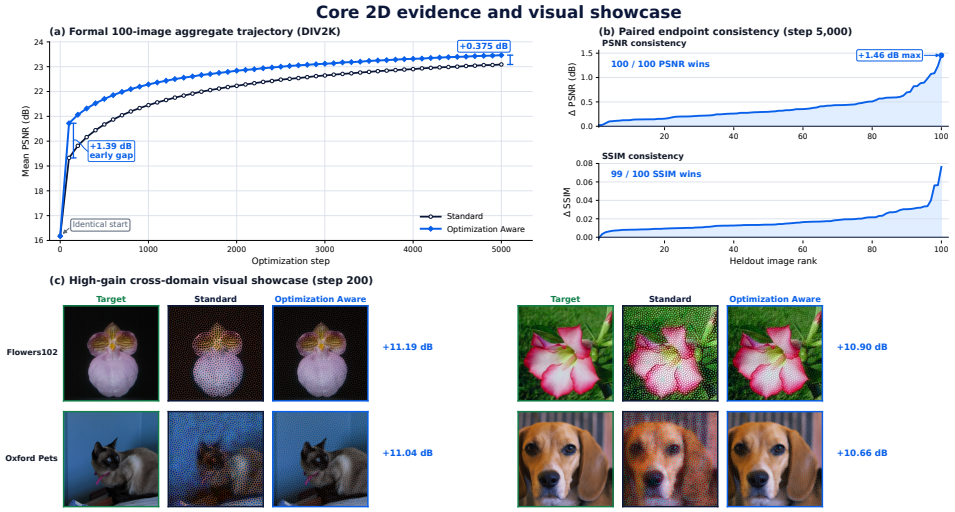


Figure 3: **Native-resolution evidence and visual examples.** **(a,b)** DIV2K: 100 held-out images, 2,000 Gaussians, independently selected policies and identical initial functions. At 5,000 updates, RGB PSNR/SSIM improve on 100/100 and 99/100 images. **(c)** Frozen Flowers102/Pets strong-Adam step-200 examples are shown in RGB; annotations use unclamped normalized-YCbCr PSNR from the separate historical protocol. They show inherited-scale correction, not SGCL or energy versus calibrated controls.

Table 2: Appearance absorption and optimizer controls. All rows use 2,000 Gaussians. Panel B is a separate six-domain mechanism study with 1,200 images per optimizer and 400 updates; its RGB values must not be pooled with Panel A's YCbCr values. DIV2K uses OA Adam versus standard AdamW; both DIV2K rows have ΔSSIM $+0.0171$. Campaign time is descriptive. Isolated time is statistically indistinguishable, with equal whole-device peak memory of 918.8 MiB. W/T/L denotes PSNR wins/ties/losses. Full metrics and controller comparisons appear in Appendix C.4.

A. OA minus standard

Comparison	Images	Updates	ΔPSNR	W/T/L	OA / STD time (s)
Matched optimizer and learning rate; normalized YCbCr PSNR					
Adam	1,200	400	+11.0805	1,199/0/1	-
AdamW	1,200	400	+9.3890	1,199/0/1	-
Independently validation-selected policies; RGB PSNR					
DIV2K heldout	100	5,000	+0.3746	100/0/0	11.406 / 11.630
Isolated hardware	24	5,000	+0.3467	24/0/0	13.917 / 13.896

B. Stronger controls; absolute RGB PSNR (dB)

Optimizer	Matched STD	OA	Transported controller	Tuned LR-only
Adam	20.7373	30.1457	30.1444	**30.3691**
AdamW	21.5439	29.5479	29.5476	**30.5405**

ance/geometry motion ratio from about 1.1%–1.2% of OA's value. A stronger learning-rate-only control can exceed OA (Table 2). Endpoint transport identifies the supplied-covector mechanism, not ordinary-gradient accuracy or optimality.

A separate derivative-only check on six existing validation images preserves the large 100-step OA effect: normalized-YCbCr gains are 7.7342 dB natively and 7.7362 dB with the ordinary shape

Table 3: Prospective selection on unseen images and new domains. SGCL minus appearance-chart controls at 100 updates, with unchanged geometry settings. Panel A has six development domains and 192 Adam/AdamW runs; Panel B has 256 runs and no selector refit. The validation rule is selected on each new domain. Co-primary controls are marked $\star$. Optimizer pairs stay together when resampling images within domains. Higher PSNR/SSIM and lower LPIPS are favorable. Joint-block controls appear in Table 33.

Comparator	ΔPSNR [95% CI]	ΔSSIM	ΔLPIPS
A. Seen domains; 96 fresh images			
Scalar $4I$	+0.0895 [0.0682, 0.1137]	+0.0040	−0.0023
OA (I)	+0.7727 [0.6299, 0.9266]	+0.0307	−0.0162
B. Four new domains; 128 fresh images			
$\star$ Validation-selected rule	**+0.0818 [0.0654, 0.0990]**	+0.0031	−0.0022
$\star$ Prior render-balance rule	+0.0782 [0.0615, 0.0962]	+0.0025	−0.0017
Greedy first-step loss	+0.5791 [0.4588, 0.7047]	+0.0236	−0.0104
OA (I)	+1.1506 [0.9814, 1.3214]	+0.0417	−0.0190

derivative. Seven of 48 attempts are conservatively refused before step 400, leaving that endpoint unaudited, not establishing historical crashes (Appendix J).

5.3 Frozen selection, useful information, and its cost

SGCL tests mainly scalar appearance-policy selection, with inherited geometry. The first study uses 96 new images from six development domains; unchanged-rule confirmation uses 128 images from four new domains and training-only priors. At 100 updates, SGCL gains 0.0818 dB over validation-selected appearance rules, with improved SSIM/LPIPS and positive domain/optimizer means (Table 3). Beating greedy loss supports horizon calibration, not the necessity of candidate-loss features. Overheads are 10.9% initially and $9.9\%/5.6\%$ versus $4I$/render-balance on new domains.

A separate 128-image study refits the predictor with and without candidate-loss features. Removing the eight candidate renders saves 4.17% time and loses 0.0581 dB at equal updates, but gains 0.3317 dB at equal time over the matched full refit. A broader short-budget-calibrated joint-block control beats original SGCL by 3.5563 dB at 100 updates; continuing the same policies to 400 reverses the ranking, with SGCL ahead by 4.7970 dB (Table 33). The latter control is not retuned at 400. At calibrated equal time, the joint control also leads SGCL; the horizon-conditioned predictor does not establish superiority over that control. Appendix I retains all tests and repeat variability.

Conditional color versus joint fitting. Shared-support scoring preserves SGCL decisions and saves 1.73 ms acquisition time on four development images, without a resolved complete-fitting gain. Fixed-geometry SOAP/least-squares references isolate color fitting; periodic corrections lower local losses without improving complete-fitting quality (Appendix Q).

6 Transfer and Scope

OA composes with fixed topology, matched insertion events, Image-GS-derived insertion (Zhang et al., 2025), and a frozen residual-driven allocation rule. Every family improves all nine tested images at 10,000 updates; native adaptive *rules*, not event trajectories, are matched. Unmodified Image-GS remains higher-quality on all nine images at 10,000, so within-system gains are not complete-system superiority (Appendix D).

For SIREN and Fourier-feature MLPs, transforming the full output head preserves the continuous function and trainable body (Sitzmann et al., 2020; Tancik et al., 2020). Dense shape can improve on determinant- or same-scale controls at 100 updates, but the separate scale-shape study does not establish persistent 400-step superiority (Appendix H). A fixed spherical-harmonic chart in official 3D Gaussian Splatting has early gains of 0.2833 to 0.2974 dB that reverse by iteration 7,000

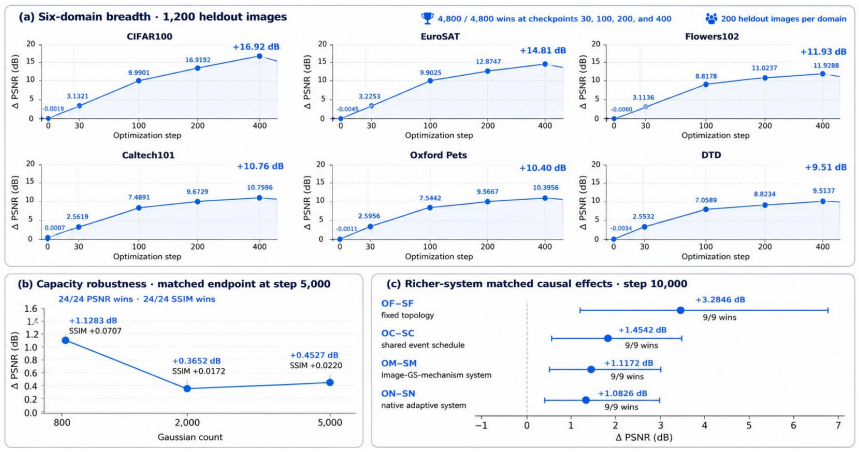


Figure 4: **OA breadth, capacity and richer systems.** **(a)** Six domains, 200 images each, four checkpoints; independently selected policies and normalized YCbCr PSNR. **(b)** Twenty-four native DIV2K images, three capacities, 5,000 updates. **(c)** Nine images, four systems, 10,000 updates. O/S means OA/standard; F/C/M/N means fixed topology/shared events/Image-GS mechanisms/native allocation. Bars show paired-bootstrap intervals. These are OA, not energy or SGCL, comparisons.

(Appendix E). Exact preservation therefore transfers more broadly than a successful fixed policy. Neither the energy construction nor the theoretical adaptive escape has an established general 3D implementation.

Scope of the guarantees. The energy identity covers continuous single atoms; the cluster theorem covers one-dimensional fixed measurements. Matched and calibrated raster gains are empirical, not convergence guarantees. Confirmation adds images, not domains or priors. The adverse 512-pixel test also changes images, priors and units. The complete repair replay validates frozen-policy outcomes, not the entire renderer (Appendix T).

7 Related Work

Natural gradients, whitening, rescaling and implicit preconditioning connect representation to dynamics (Amari, 1998; Martens & Grosse, 2015; Desjardins et al., 2015; Dinh et al., 2017; Arora et al., 2018; Gehring et al., 2021). Adaptive methods and trivializations exploit coordinates (DePavia et al., 2026; Singh, 2026; Vyas et al., 2025; Lezcano-Casado, 2019; 2020). Section 3 distinguishes our Gaussian angle/minimax result from generalized interpolation and stable RBF/Hermite bases (Maday & Mula, 2013; Fornberg et al., 2011; Kormann et al., 2019).

Variable projection, natural Rprop and Schur methods provide controls, not the energy update (Golub & Pereyra, 1973; Igel et al., 2005; Agarwal et al., 2010; Schrader et al., 2024). SGCL selects a frozen policy, unlike learned updates (Andrychowicz et al., 2016) or differentiation through training (Maclaurin et al., 2015); preview cost relates to information seeking (Grand et al., 2026; Rosman et al., 2016). Gaussian work studies adaptive allocation (Zhang et al., 2025), eigenspace representations (Tai et al., 2025), and learned updates (Pearl et al., 2026); we isolate coordinates at fixed capacity (Appendix M).

8 Conclusion

Coordinate design can improve Gaussian fitting without changing the function family. Energy coordinates retain quality–time gains after symmetric calibration and a complete backward-repair replay, with improved mean LPIPS. The cluster characterization and optimizer-payoff bounds delimit the geometric guarantees. Together, these results support evaluating coordinates jointly with the optimizer, calibration access and output budget.

Reproducibility Statement

Appendices A, K, L, N and O provide assumptions, proofs and limitations. Appendices B–R specify the completed study designs, numerical conventions, data roles, optimizer settings, timing boundaries and detailed results; Appendix S gives the secondary predictors' numerical policies. Appendix T reports the complete same-cohort replay under repaired projection-backward arithmetic, including the incumbent comparison, metric checks and adverse outcomes. We intend to publicly release cleaned code and configurations after acceptance.

AI Use Statement

Generative AI tools assisted with language and manuscript editing and with identifying relevant related work. The authors conceived the research, designed and conducted the experiments, performed and verified the analyses, and determined the interpretations and claims. All AI-assisted material, including cited related work, was reviewed and verified by the authors, who take full responsibility for the paper. The first author originated the research idea, set the scientific direction, designed and conducted the experimental program, and made the final decisions on analyses, interpretations, and claims.

Ethics Statement

This study uses existing image benchmarks and synthetic mathematical examples and collects no new human-participant data. Its experiments concern image reconstruction and optimization, not demographic fairness or safety validation. Benchmark-dependent gains, adverse cases and measurement limitations are reported explicitly. Datasets and pretrained evaluation models are used subject to their applicable licenses and redistribution terms.

References

Sameer Agarwal, Noah Snavely, Steven M. Seitz, and Richard Szeliski. Bundle adjustment in the large. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010. URL https://grail.cs.washington.edu/projects/bal/bal.pdf.

Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1122–1131, 2017. doi: 10.1109/CVPRW.2017.150.

Shun-ichi Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998. doi: 10.1162/089976698300017746.

Marcin Andrychowicz, Misha Denil, Sergio Gómez, Matthew W. Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando de Freitas. Learning to learn by gradient descent by gradient descent. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Sanjeev Arora, Nadav Cohen, and Elad Hazan. On the optimization of deep networks: Implicit acceleration by overparameterization. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pp. 244–253, 2018.

Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

Hans-Joachim Baues and Fernando Muro. The symmetric action on secondary homotopy groups. *Bulletin of the Belgian Mathematical Society – Simon Stevin*, 15(4):733–768, 2008. URL https://personal.us.es/fmuro/assets/files/papers/zbMATH05496973.pdf.

Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101: Mining discriminative components with random forests. In *European Conference on Computer Vision*, 2014.

Guangyong Chen, Peng Xue, Min Gan, Jing Chen, Wenzhong Guo, and C. L. Philip Chen. Variable projection algorithms: Theoretical insights and a novel approach for problems with large residual, 2025. URL https://arxiv.org/html/2402.13865v2.

Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3606–3613, 2014.

Adam Coates, Andrew Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15, pp. 215–223, 2011. URL https://proceedings.mlr.press/v15/coates11a.html.

Etienne de Klerk, François Glineur, and Adrien B. Taylor. On the worst-case complexity of the gradient method with exact line search for smooth strongly convex functions. *Optimization Letters*, 11:1185–1199, 2017. doi: 10.1007/s11590-016-1087-4.

Adela DePavia, Jose Cruzado, Jiayou Liang, Vasileios Charisopoulos, and Rebecca Willett. Faster adaptive optimization via expected gradient outer product reparameterization. *arXiv preprint arXiv:2502.01594v2*, 2026.

Guillaume Desjardins, Karen Simonyan, Razvan Pascanu, and Koray Kavukcuoglu. Natural neural networks. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Renjie Ding, Yaonan Wang, Min Liu, Jialin Zhu, Jiazheng Wang, Jiahao Zhao, Wenting Shen, Feixiang He, and Xiang Chen. A step to decouple optimization in 3dgs. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2601.16736.

Laurent Dinh, Razvan Pascanu, Samy Bengio, and Yoshua Bengio. Sharp minima can generalize for deep nets. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pp. 1019–1028, 2017.

John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12:2121–2159, 2011.

Ludwig Elsner. A note on optimal block-scaling of matrices. *Numerische Mathematik*, 44:127–128, 1984. doi: 10.1007/BF01389760.

Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):594–611, 2006. doi: 10.1109/TPAMI.2006.79.

Bengt Fornberg, Elisabeth Larsson, and Natasha Flyer. Stable computations with gaussian radial basis functions. *SIAM Journal on Scientific Computing*, 33(2):869–892, 2011. doi: 10.1137/09076756X.

Clement Gehring, Kenji Kawaguchi, Jiaoyang Huang, and Leslie P. Kaelbling. Understanding end-to-end model-based reinforcement learning methods as implicit parameterization. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Gene H. Golub and Victor Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis*, 10(2):413–432, 1973. doi: 10.1137/0710036.

Gabriel Grand, Valerio Pepe, Joshua B. Tenenbaum, and Jacob Andreas. Shoot first, ask questions later? building rational agents that explore and act like people. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2510.20886.

Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow. Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics*, 37(6), 2018. doi: 10.1145/3272127.3275084.

Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226, 2019. doi: 10.1109/JSTARS.2019.2918242.

Roger Hsiao, Yuchen Fang, Xiangru Huang, Ruilong Li, Hesam Rabeti, Zan Gojcic, Javad Lavaei, James Demmel, and Sophia Shao. 3DGS^2-TR: Scalable second-order trust-region method for 3d gaussian splatting. *arXiv preprint arXiv:2602.00395*, 2026.

Christian Igel and Michael Hüsken. Empirical evaluation of the improved Rprop learning algorithms. *Neurocomputing*, 50:105–123, 2003. doi: 10.1016/S0925-2312(01)00700-7.

Christian Igel, Marc Toussaint, and Wan Weishui. Rprop using the natural gradient. In *Trends and Applications in Constructive Approximation*, pp. 259–272. Birkhäuser, 2005. doi: 10.1007/3-7643-7356-3_19. URL https://christian-igel.github.io/paper/RUtNG.pdf.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, 2023. doi: 10.1145/3592433.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4), 2017. doi: 10.1145/3072959.3073599.

Katharina Kormann, Caroline Lasser, and Anna Yurova. Stable interpolation with isotropic and anisotropic gaussians using hermite generating function, 2019. URL https://arxiv.org/abs/1905.09542.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

Mario Lezcano-Casado. Trivializations for gradient-based optimization on manifolds. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://arxiv.org/abs/1909.09501.

Mario Lezcano-Casado. Adaptive and momentum methods on manifolds through trivializations, 2020. URL https://arxiv.org/abs/2010.04617.

Tiantian Li, Xinjie Zhang, Xingtong Ge, Tongda Xu, Dailan He, Jun Zhang, and Yan Wang. GaussianImage++: Boosted image representation and compression with 2d gaussian splatting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 6442–6449, 2026. doi: 10.1609/aaai.v40i8.37572.

Minghua Lin and Gord Sinnamon. The generalized Wielandt inequality in inner product spaces. *Eurasian Mathematical Journal*, 3:72–85, 2012. URL https://arxiv.org/abs/1201.6294.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Dougal Maclaurin, David Duvenaud, and Ryan Adams. Gradient-based hyperparameter optimization through reversible learning. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pp. 2113–2122, 2015. URL https://proceedings.mlr.press/v37/maclaurin15.html.

Yvon Maday and Olga Mula. A generalized empirical interpolation method: Application of reduced basis techniques to data assimilation. In *Analysis and Numerics of Partial Differential Equations*, volume 4 of *Springer INdAM Series*, pp. 221–235. Springer, 2013. doi: 10.1007/978-88-470-2592-9_13.

Subhransu Maji, Juho Kannala, Esa Rahtu, Matthew Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*, 2013.

James Martens and Roger Grosse. Optimizing neural networks with kronecker-factored approximate curvature. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pp. 2408–2417, 2015.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision*, pp. 405–421, 2020. doi: 10.1007/978-3-030-58452-8_24.

Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y. Ng. Reading digits in natural images with unsupervised feature learning. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*, 2011.

Behnam Neyshabur, Ruslan Salakhutdinov, and Nati Srebro. Path-SGD: Path-normalized optimization in deep neural networks. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *Sixth Indian Conference on Computer Vision, Graphics and Image Processing*, pp. 722–729, 2008. doi: 10.1109/ICVGIP.2008.47.

Zixuan Pan, Kaiyuan Tang, Jun Xia, Yifan Qin, Lin Gu, Chaoli Wang, Jianxu Chen, and Yiyu Shi. SGI: Structured 2d gaussians for efficient and compact large image representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.

Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman, and C. V. Jawahar. Cats and dogs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3498–3505, 2012. doi: 10.1109/CVPR.2012.6248092.

Naama Pearl, Stefano Esposito, Haofei Xu, Amit Peleg, Patricia Gschossmann, Lorenzo Porzi, Peter Kontschieder, Gerard Pons-Moll, and Andreas Geiger. Learn2Splat: Extending the horizon of learned 3dgs optimization. *arXiv preprint arXiv:2605.15760*, 2026.

Boris T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964. doi: 10.1016/0041-5553(64)90137-5.

Martin Riedmiller and Heinrich Braun. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *IEEE International Conference on Neural Networks*, pp. 586–591, 1993. doi: 10.1109/ICNN.1993.298623.

Guy Rosman, Daniela Rus, and John W. Fisher, III. Information-driven adaptive structured-light scanners. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 874–883, 2016.

Tim Salimans and Diederik P. Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Shibani Santurkar, Dimitris Tsipras, Andrew Ilyas, and Aleksander Madry. How does batch normalization help optimization? In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Simon Elias Schrader, Thomas Bondo Pedersen, and Simen Kvaal. Multidimensional quantum dynamics with explicitly correlated gaussian wave packets using rothe's method, 2024. URL `https://arxiv.org/html/2411.05459v1`.

Devender Singh. The loss does not see the basis, but adam does. *arXiv preprint arXiv:2608.05136*, 2026.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Johannes Stallkamp, Marc Schlipsing, Jan Salmen, and Christian Igel. The German Traffic Sign Recognition Benchmark: A multi-class classification competition. In *International Joint Conference on Neural Networks*, pp. 1453–1460, 2011. doi: 10.1109/IJCNN.2011.6033395.

Lo-Wei Tai, Ching-En Li, Cheng-Lin Chen, Chih-Jung Tsai, Hwann-Tzong Chen, and Tyng-Luh Liu. EigenGS representation: From eigenspace to gaussian image space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13487–13496, 2025. doi: 10.1109/CVPR52734.2025.01259.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Nikhil Vyas, Depen Morwani, Rosie Zhao, Itai Shapira, David Brandfonbrener, Lucas Janson, and Sham Kakade. SOAP: Improving and stabilizing shampoo using adam for language modeling. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/e988664070e9591f93fdcf605f7dc623-Abstract-Conference.html.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. doi: 10.1109/TIP.2003.819861.

Zhaojie Zeng, Yuesong Wang, Tao Guan, Chao Yang, and Lili Ju. Instant GaussianImage: A generalizable and self-adaptive image representation via 2d gaussian splatting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 27896–27905, 2025. doi: 10.1109/ICCV51701.2025.02589.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.

Xinjie Zhang, Xingtong Ge, Tongda Xu, Dailan He, Yan Wang, Hongwei Qin, Guo Lu, Jing Geng, and Jun Zhang. GaussianImage: 1000 FPS image representation and compression by 2d gaussian splatting. In *European Conference on Computer Vision*, pp. 327–345, 2024. doi: 10.1007/978-3-031-72673-6_18.

Yunxiang Zhang, Bingxuan Li, Alexandr Kuznetsov, Akshay Jindal, Stavros Diolatzis, Kenneth Chen, Anton Sochenov, Anton Kaplanyan, and Qi Sun. Image-GS: Content-adaptive image representation via 2d gaussians. In *ACM SIGGRAPH Conference Papers*, pp. 1–11, 2025. doi: 10.1145/3721238.3730596.

Hexu Zhao, Haoyang Weng, Daohan Lu, Ang Li, Jinyang Li, Aurojit Panda, and Saining Xie. On scaling up 3d gaussian splatting training. In *International Conference on Learning Representations*, 2025.

Lingting Zhu, Guying Lin, Jinnan Chen, Xinjie Zhang, Zhenchao Jin, Zhao Wang, and Lequan Yu. Large images are gaussians: High-quality large image representation with levels of 2d gaussian splatting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 10977–10985, 2025. doi: 10.1609/aaai.v39i10.33193.

Appendix

Study guide. The calibrated energy recipe and the fixed-measurement theorem are separate contributions. The following index distinguishes their evidence, the exact-transport controls, and the secondary selection and scope studies.

Question	Evidence or specification	Appendix
Energy after symmetric calibration	64 fresh images; five profiles per chart; exact update and clock	R, R.8
Earlier energy study	Separate 64-image cohort; failures, ordinary pullback and guard	J
Sharp Gaussian cluster certificate	Moving-center angle, extremizer, minimax and decoder proofs	N
Coordinate and optimizer controls	Linear transport, 2D effects, metrics and resource contracts	A–F
Prospective appearance selection	Original SGCL, joint-block controls and numerical policies	G, I, S
Transfer and computational limits	3D/non-Gaussian scope; normalization and conditional color	E, H, P–Q
Supporting theory and prior work	Chart obstruction, memory, classical payoff bounds and context	K–O

The targeted backward-precision study is in Appendix J.10. Appendix T adds a complete original/repaired replay of the symmetric-calibration cohort, including the incumbent; historical confirmation records remain separate.

Terminology. STD denotes the inherited standard coordinates; OA, optimization-aware absorbed coordinates; and SGCL, Shared-Gradient Chart Lookahead. RGB denotes red-green-blue color, and YCbCr denotes luminance and two chrominance channels. PCA is principal component analysis, MSE mean squared error, PSNR peak signal-to-noise ratio, SSIM structural similarity, LPIPS learned perceptual image patch similarity, MLP multilayer perceptron, SIREN sinusoidal representation network, and SH spherical harmonics. CI denotes a confidence interval; W/T/L denotes paired wins/ties/losses. A *chart* is an invertible parameter-coordinate map, a *physical state* is the decoded renderer state, and a *fixed rule* is a frozen construction policy that may still use target-dependent statistics. Unless stated otherwise, optimizer-specific outcomes on one image are paired measurements, not independent images.

Construction map.

Construction	Established consequence	Evidence and scope
Fixed absorption	Supplied-covector optimizer transport	Paired OA fitting; no optimality claim
Energy chart	Continuous single-atom self-coupling cancels	Matched and calibrated fitting; no full-image whitening
Fixed measurements	Sharp moving-cluster angle certificate	Common-width 1D theory; not the energy implementation
SGCL	Frozen finite-horizon appearance selection	New-domain library comparison; not a joint-optimizer default

A Extended Theory

Derivative scope. Differential, Hessian and gradient-descent results in this appendix assume ordinary derivatives on a differentiable active set. Fixed-chart optimizer conjugacy also applies to consistently transformed supplied covectors and states. The native shape-covector convention is separately audited in Appendix J; conjugacy is not validation of that derivative implementation.

This appendix supplies the assumptions, proofs, counterexamples, optimizer-state boundaries, and prospective audit supporting Section 2. The presentation is self-contained at the level needed for the paper. We distinguish exact linear-chart statements, local quadratic statements, and adaptive-optimizer transport statements throughout.

A.1 Assumptions, notation, and exact equivalence

Let $C \in \mathbb{R}^{N\times d}$ contain the standard appearance coefficients, let $U = CS^\top$ contain the absorbed physical appearance coefficients, and let g collect every nonappearance parameter. The fixed map $S \in \mathbb{R}^{d\times d}$ is invertible. With column-wise vectorization,

$$c = \mathrm{vec}(C), \qquad u = \mathrm{vec}(U) = Dc, \qquad D = S \otimes I_N. \tag{11}$$

Define the standard and physical coordinates

$$x = \begin{bmatrix} c \\ g \end{bmatrix}, \qquad q = \begin{bmatrix} u \\ g \end{bmatrix} = Tx, \qquad T = \mathrm{diag}(D, I_g), \qquad P = TT^\top \succ 0. \tag{12}$$

For fixed g, assume that the complete predecoder appearance path is linear in its appearance argument, that every predecoder appearance row affected by S is mapped, and that both charts apply the same fixed offsets and post-render operations. If constraints are present, their feasible sets are paired by $q = Tx$.

Proposition A.1 (Global function and capacity equivalence). Under the assumptions above, the paired standard and absorbed coordinates represent exactly the same output for every state. Because T is invertible, they represent the same function family, have the same trainable dimension, and have paired optimum sets.

Proof. For fixed g, write the appearance renderer as $\mathcal{R}(A, g) = W(g)A$. Then

$$\mathcal{R}(C, g)S^\top = W(g)CS^\top = W(g)U = \mathcal{R}(U, g). \tag{13}$$

The common offsets and post-render operations preserve equality. Invertibility gives the inverse $C = US^{-\top}$, so the map is bijective and preserves the represented family and its dimension. Mapping the feasible set by the same bijection preserves the attainable objective values and optimum sets. □

Let $F(q)$ denote the objective in physical coordinates and let

$$f(x) = F(Tx) \tag{14}$$

be the same objective in the standard chart. Derivative claims below apply at differentiable points; piecewise differentiable renderers are covered within each differentiable region.

A.2 Derivative congruence and physical preconditioning

Proposition A.2 (Derivative congruence). Suppose F is twice differentiable at $q = Tx$. Then

$$\nabla_x f(x) = T^\top \nabla_q F(q), \qquad \nabla_x^2 f(x) = T^\top \nabla_q^2 F(q) T. \tag{15}$$

If $r(q)$ is a differentiable residual and $W \succeq 0$ is a fixed weighting matrix, then the Gauss-Newton matrices satisfy

$$G_x = T^\top G_q T, \qquad G_q = J_q^\top W J_q. \tag{16}$$

Proof. The differential of $q = Tx$ is $dq = T\,dx$. Hence

$$df = \langle \nabla_q F, T\,dx\rangle = \langle T^\top \nabla_q F, dx\rangle, \tag{17}$$

which proves the gradient identity. Differentiating again and using that T is constant gives the Hessian identity. The residual Jacobian obeys $J_x = J_q T$, hence $G_x = J_x^\top W J_x = T^\top G_q T$. □

Corollary A.1 (Block curvature). Partition $q = (u, g)$ and write $T = \mathrm{diag}(D, I_g)$. If

$$H_q = \begin{bmatrix} H_{uu} & H_{ug} \\ H_{gu} & H_{gg} \end{bmatrix}, \tag{18}$$

then

$$H_x = \begin{bmatrix} D^\top H_{uu} D & D^\top H_{ug} \\ H_{gu} D & H_{gg} \end{bmatrix}. \tag{19}$$

Thus a chart that rescales only appearance also changes the numerical appearance-geometry cross blocks.

Theorem A.1 (Euclidean descent induces a physical preconditioner). A standard-chart Euclidean step

$$x^+ = x - \eta \nabla_x f(x) \tag{20}$$

maps to

$$q^+ = q - \eta P \nabla_q F(q), \qquad P = TT^\top. \tag{21}$$

Direct descent in the absorbed chart uses $P = I$. The physical updates agree for every gradient if and only if $TT^\top = I$.

Proof. Multiplying the standard update by T and applying Equation 15 yields

$$T(x^+ - x) = -\eta TT^\top \nabla_q F(q). \tag{22}$$

Equality with $-\eta \nabla_q F(q)$ for every gradient is equivalent to $TT^\top = I$. □

Proposition A.3 (Exact transported second-order invariance). Assume the relevant Hessian, Gauss-Newton matrix, or metric is nonsingular and that the complete linear solve is transported consistently. Then a Newton, exact Gauss-Newton, or exact natural-gradient step is invariant to the chart. For example,

$$T\left[-(T^\top HT)^{-1} T^\top g\right] = -H^{-1} g. \tag{23}$$

Proof. Since T is invertible,

$$(T^\top HT)^{-1} = T^{-1} H^{-1} T^{-\top}. \tag{24}$$

Substitution gives Equation 23. The metric argument is identical. □

Independent identity damping, diagonal approximations, or independently selected pseudoinverses are extra chart-dependent choices and need not preserve this invariance.

A.3 Coupled quadratic conditioning

Consider the symmetric positive definite two-block Hessian

$$H_1 = \begin{bmatrix} a & c \\ c & b \end{bmatrix}, \qquad a > 0, \quad b > 0, \quad ab - c^2 > 0. \tag{25}$$

Rescale the appearance coordinate by $s > 0$. The chart Hessian becomes

$$H_s = \begin{bmatrix} as^2 & cs \\ cs & b \end{bmatrix}. \tag{26}$$

Let $\rho = c/\sqrt{ab}$ denote normalized appearance-geometry coupling.

Theorem A.2 (Condition-balancing scale). The spectral condition number of H_s is uniquely minimized at

$$s_\star = \sqrt{\frac{b}{a}}, \tag{27}$$

and the minimum is

$$\kappa_{\min} = \frac{1 + |\rho|}{1 - |\rho|}. \tag{28}$$

Consequently, the absorbed physical chart $s = 1$ balances the two local diagonal curvatures when $a = b$.

Proof. The trace and determinant of H_s are

$$\tau = as^2 + b, \qquad \delta = s^2(ab - c^2). \tag{29}$$

For a positive definite 2×2 matrix, the eigenvalue ratio is a strictly increasing function of $\tau/\sqrt{\delta}$. Up to the positive constant $\sqrt{ab - c^2}$,

$$\frac{\tau}{\sqrt{\delta}} = \frac{as + b/s}{\sqrt{ab - c^2}}. \tag{30}$$

The numerator is strictly convex for $s > 0$ and its derivative is $a - b/s^2$, which vanishes uniquely at Equation 27. At this scale,

$$H_{s_\star} = b \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \tag{31}$$

whose ordered eigenvalues are $b(1 - |\rho|)$ and $b(1 + |\rho|)$. □

Proposition A.4 (Intrinsic coupling is not removed). The geometry Schur complement and normalized coupling are invariant to s:

$$b - (cs)(as^2)^{-1}(cs) = b - \frac{c^2}{a}, \qquad \frac{cs}{\sqrt{(as^2)b}} = \rho. \tag{32}$$

The chart can balance relative block scales, but it cannot eliminate intrinsic coupling.

A.4 When geometry begins to diverge

Theorem A.3 (Second-update geometry divergence). Let $q = (u, g)$ and

$$F(q) = \tfrac{1}{2} q^\top H q + r^\top q, \qquad H = H^\top, \qquad P = \mathrm{diag}(P_u, I_g). \tag{33}$$

Initialize direct absorbed-chart descent and mapped standard-chart descent at the same physical state q_0 with the same step size η. Then the first geometry update is identical, while

$$u_1^{\mathrm{std}} - u_1^{\mathrm{OA}} = -\eta(P_u - I)\nabla_u F(q_0), \tag{34}$$

$$g_2^{\mathrm{std}} - g_2^{\mathrm{OA}} = \eta^2 H_{gu}(P_u - I)\nabla_u F(q_0). \tag{35}$$

For a three-times continuously differentiable objective whose Hessian is locally Lipschitz, the second identity holds to leading order with remainder $O(\eta^3)$ in a bounded neighborhood of q_0.

Proof. The first updates are

$$q_1^{\mathrm{std}} = q_0 - \eta P \nabla F(q_0), \qquad q_1^{\mathrm{OA}} = q_0 - \eta \nabla F(q_0). \tag{36}$$

The geometry block of P is identity, so the first geometry updates agree and the appearance difference is Equation 34. For the quadratic,

$$\nabla_g F(q_1^{\mathrm{std}}) - \nabla_g F(q_1^{\mathrm{OA}}) = H_{gu}(u_1^{\mathrm{std}} - u_1^{\mathrm{OA}}). \tag{37}$$

Multiplying by the second geometry step $-\eta$ gives Equation 35. A Taylor expansion about q_0, together with the locally Lipschitz Hessian, contributes an $O(\eta^2)$ gradient remainder and therefore an $O(\eta^3)$ update remainder. □

The leading geometry separation vanishes when $P_u = I$, $H_{gu} = 0$, or $\nabla_u F(q_0) = 0$. These are exact null regimes for this local mechanism.

A.5 Finite-horizon convergence and gap shape

Proposition A.5 (Exact modal loss under a fixed chart). Let

$$F(q) = \tfrac{1}{2}(q - q^\star)^\top H (q - q^\star), \qquad H \succ 0, \qquad P \succ 0. \tag{38}$$

Define

$$e_k = q_k - q^\star, \qquad z_k = P^{-1/2} e_k, \qquad K_P = P^{1/2} H P^{1/2}. \tag{39}$$

Write $z_0 = \sum_i a_i v_i$ in an orthonormal eigensystem $K_P v_i = \lambda_i v_i$. If $|1 - \eta\lambda_i| < 1$ for every active mode, then

$$F(q_k) - F(q^\star) = \tfrac{1}{2} \sum_i \lambda_i a_i^2 (1 - \eta\lambda_i)^{2k}. \tag{40}$$

Proof. The preconditioned update gives

$$z_{k+1} = (I - \eta K_P) z_k. \tag{41}$$

Each modal coefficient is multiplied by $1 - \eta\lambda_i$ per update. Substituting the resulting expansion into

$$F(q_k) - F(q^\star) = \tfrac{1}{2} z_k^\top K_P z_k \tag{42}$$

yields Equation 40. □

Equation 40 shows why condition number alone is insufficient for a finite-budget prediction. The actual step size, individual contraction factors, initial modal weights, and target alignment all enter the exact error.

Proposition A.6 (A gap can widen and later contract). Let $0 < r < s < 1$ and $A > 0$. For real $k \geq 0$, define

$$G(k) = A\left(s^{2k} - r^{2k}\right). \tag{43}$$

Then $G(0) = 0$, $G(k) > 0$ for $k > 0$, $G(k) \to 0$ as $k \to \infty$, and G has a unique maximum at

$$k^\star = \frac{\log(\log r / \log s)}{2\log(s/r)}. \tag{44}$$

Proof. Differentiating gives

$$G'(k) = 2A\left[s^{2k}\log s - r^{2k}\log r\right]. \tag{45}$$

The equation $G'(k) = 0$ is equivalent to

$$(s/r)^{2k} = \log r / \log s, \tag{46}$$

which yields Equation 44. The ratio of the derivative terms is strictly monotone, so the critical point is unique and changes the derivative from positive to negative. □

Thus a shrinking late gap does not negate meaningful finite-compute acceleration, and a widening early gap does not imply different attainable optima. Multimode and nonlinear trajectories may contain several extrema, so the one-mode theorem is not used to claim that every empirical curve is strictly unimodal.

Proposition A.7 (Sufficient condition for reversal after a regime change). Suppose OA has lower modal error at a switch time k_0:

$$E_{\mathrm{OA}}(k_0) < E_{\mathrm{std}}(k_0). \tag{47}$$

After the switch, suppose the active mode contracts by r_{OA} and r_{std}, where

$$0 < r_{\mathrm{std}} < r_{\mathrm{OA}} < 1. \tag{48}$$

Then the standard chart overtakes OA after any integer n satisfying

$$n > \frac{\log(E_{\mathrm{OA}}(k_0)/E_{\mathrm{std}}(k_0))}{2\log(r_{\mathrm{std}}/r_{\mathrm{OA}})}. \tag{49}$$

Proof. After n additional updates, standard is better exactly when

$$E_{\mathrm{std}}(k_0)r_{\mathrm{std}}^{2n} < E_{\mathrm{OA}}(k_0)r_{\mathrm{OA}}^{2n}. \tag{50}$$

Rearranging and dividing by the negative quantity $2\log(r_{\mathrm{std}}/r_{\mathrm{OA}})$ gives Equation 49. □

This proposition formalizes how a chart can help under one active spectrum and alignment, then lose that advantage after the basis, objective, residual alignment, topology, or optimizer state changes.

A.6 No universal fixed-chart dominance

Theorem A.4 (No nonidentity fixed preconditioner dominates every objective). Let $P \succ 0$ and $P \neq I$. Under a shared step size $\eta = 1$, there are positive definite quadratic objectives and nonzero initial errors for which the P-preconditioned update is strictly better, and others for which direct physical descent is strictly better.

Proof. For $H = P^{-1}$, the P-preconditioned update matrix is $I - PH = 0$, so it reaches the minimizer in one step. Choose the initial error along an eigendirection of P whose eigenvalue differs from one. Direct physical descent has update matrix $I - P^{-1}$ and leaves a nonzero residual. Conversely, for $H = I$, direct physical descent reaches the minimizer in one step, whereas the P update matrix $I - P$ leaves a nonzero residual along the same eigendirection. □

If $P = cI$, a scalar learning-rate retuning can neutralize the distinction for fixed-step gradient descent. A nonscalar P changes anisotropy and cannot generally be removed by one scalar learning rate. The theorem therefore supports a conditional representation principle, not universal superiority.

A.7 Adaptive optimizers and exact controls

For a positive diagonal map $q_i = t_i x_i$, paired physical gradients obey $g_i^x = t_i g_i^q$. Adam uses

$$m_{k,i} = \beta_1 m_{k-1,i} + (1-\beta_1) g_{k,i}, \tag{51}$$

$$v_{k,i} = \beta_2 v_{k-1,i} + (1-\beta_2) g_{k,i}^2, \tag{52}$$

followed by the usual bias corrections $\widehat{m}_{k,i}$ and $\widehat{v}_{k,i}$.

Proposition A.8 (Adam state scaling). Assume paired physical states, paired gradient sequences, positive diagonal T, and zero or correctly mapped initial moments. Then

$$m_{k,i}^x = t_i m_{k,i}^q, \qquad v_{k,i}^x = t_i^2 v_{k,i}^q, \tag{53}$$

and the same relations hold after bias correction.

Proof. The first-moment recurrence is linear in the gradient and the second-moment recurrence is quadratic. Induction gives Equation 53. Bias correction divides by iteration-dependent scalars and preserves both relations. □

Theorem A.5 (Mapped Adam step and exact diagonal conjugacy). Under Proposition A.8, one standard-chart Adam step with learning rate α and stabilizer ϵ maps to

$$\Delta q_i = -\alpha t_i \frac{\widehat{m}_{k,i}^q}{\sqrt{\widehat{v}_{k,i}^q} + \epsilon/t_i}. \tag{54}$$

A physical-coordinate controller with $\alpha_i = \alpha t_i$ and $\epsilon_i = \epsilon/t_i$ reproduces this update exactly. Conversely, global OA Adam with $(\alpha_{\mathrm{OA}}, \epsilon_{\mathrm{OA}})$ is reproduced in the standard chart by $\alpha_i^x = \alpha_{\mathrm{OA}}/t_i$ and $\epsilon_i^x = t_i \epsilon_{\mathrm{OA}}$. If the initial physical states and moments are paired and no additional chart-dependent operation is introduced, the correspondence holds for the complete trajectory by induction.

Proof. The standard-coordinate update is

$$\Delta x_i = -\alpha \frac{t_i \widehat{m}_{k,i}^q}{t_i \sqrt{\widehat{v}_{k,i}^q} + \epsilon}. \tag{55}$$

Multiplying by t_i yields Equation 54. The stated coordinate-dependent hyperparameters match numerator and denominator exactly. Therefore the physical updates agree, the next states remain paired, and Proposition A.8 applies again; induction completes the trajectory claim. The converse follows by solving the same identities for the standard-chart hyperparameters. □

One global pair (α, ϵ) is not generally conjugate across coordinates when the t_i differ. Dense coordinate mixing cannot in general be represented by elementwise Adam hyperparameters, while clipping, schedules, topology changes, and incompletely transported state require separate analysis.

AdamW boundary. AdamW applies the same adaptive update as Adam and a decoupled shrinkage. Under coordinate-dependent controller settings, exact physical conjugacy additionally requires the decay products to match, $\alpha_i^x \lambda_i^x = \alpha_{\mathrm{OA}} \lambda_{\mathrm{OA}}$. With one matched scalar product within a parameter group, shrinkage commutes with a fixed linear map:

$$T\big[(1-\alpha\lambda)x\big] = (1-\alpha\lambda)q. \tag{56}$$

Chart-misaligned parameter groups, coordinate-dependent decay, or coupled ℓ_2 regularization do not transport automatically. In the coupled case, $\frac{\lambda}{2}\|x\|^2$ becomes $\frac{\lambda}{2} q^\top T^{-\top} T^{-1} q$, rather than the physical penalty $\frac{\lambda}{2}\|q\|^2$.

A.8 Controller validation and numerical scope

The exact positive-diagonal controller was implemented independently in both mapping directions for Adam and AdamW. Twenty-nine deterministic float64 tests cover fresh and mapped nonzero moments, 500-step trajectories, mixed charted and identity parameters in one group, background

mapping, frozen parameters, a scheduled learning-rate transition, and checkpoint save/load. Every tested physical parameter, update, first moment, and second moment deviation is at most 10^{-12}. A bounded CPU execution of the project renderer, including the PCA decoder, background map, pixelwise MSE, geometry optimization, and restart at step 50, also satisfies its predeclared numerical tolerances.

Table 4: Optimizer-transport validation. Exact-controller validation under deterministic arithmetic and the bounded project-renderer test. Deviations compare the direct source-coordinate reference with the transported controller.

Validation layer	Coverage	Largest observed deviations	Outcome
Deterministic float64	29 Adam/AdamW tests	parameters, updates, and moments $\leq 10^{-12}$	Pass
Project-renderer CPU	100 steps plus restart	state 3.58×10^{-7}; RGB 5.18×10^{-7}; PSNR 1.06×10^{-6} dB	Pass
Production endpoints	Four protocols (2,524 optimizer rows)	largest mean PSNR gap 0.0108 dB; all LPIPS intervals include zero	Practical match

The theorem is exact in real arithmetic. Production validation instead asks whether the transported controller reproduces OA's finite-budget image quality when a float32 rasterizer contains discrete contribution tests. Across the four pooled Adam and AdamW protocols, the largest absolute mean PSNR difference is 0.0108 dB; SSIM differences are negligible and every OA-controller LPIPS interval includes zero. Thus the controller is a practical endpoint match even though roundoff can separate the detailed trajectories. We make no claim of bitwise-identical production states. This numerical scope does not alter the OA-versus-standard comparisons, which evaluate both representations as actually executed.

Counterexample 1 (Dense transforms are not Adam invariant). At the first Adam step with zero moments and $\epsilon = 0$, let $g = (1, 2)^\top$ and

$$Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}. \tag{57}$$

Direct Adam moves along $-\operatorname{sign}(g) = -(1, 1)^\top$. In the rotated chart,

$$Q^\top g = (3, 1)^\top/\sqrt{2}, \tag{58}$$

so the mapped step is

$$-Q \operatorname{sign}(Q^\top g) = (0, -\sqrt{2})^\top, \tag{59}$$

which is not parallel to the direct step.

Theorem A.6 (Exact SGD controller). Suppose the standard-coordinate gradient is multiplied by a matrix M before an SGD step. The mapped physical update equals direct OA SGD for every gradient if and only if

$$M = T^{-1}T^{-\top} = (T^\top T)^{-1}. \tag{60}$$

Proof. The controlled mapped update is $-\eta T M T^\top \nabla_q F$. Equality with $-\eta \nabla_q F$ for every gradient is equivalent to $TMT^\top = I$, whose unique solution is Equation 60. □

Heuristic multiplication by s, $1/s$, or another scalar is therefore not generally the exact controller. For Adam, fixed scalar gradient multiplication can also cancel inside $m/\sqrt{v}$ when ϵ is negligible, even though the parameter chart changes the mapped physical step. The empirical manual-scaling experiment rules out only its tested heuristic and does not substitute for Theorem A.5.

A.9 PATH DEPENDENCE UNDER CHART HANDOFF AND OPTIMIZER REBOOT

Proposition A.9 (Identity handoff is not a rewind). Suppose an OA trajectory reaches physical state q_τ^{OA} at time τ. Re-expressing that state in the identity chart preserves the represented function at τ. Subsequent identity-chart optimization starts from q_τ^{OA} and the retained optimizer state; it is not the counterfactual trajectory obtained by running the standard system from q_0 for τ updates.

Proof. The handoff is a function-preserving re-expression of the current physical state. Unless both the physical and optimizer states coincide with the standard trajectory at τ, deterministic future updates begin from different states and need not coincide. □

Proposition A.10 (Optimizer reboot boundary). Resetting Adam moments, variance estimates, and local step counters removes optimizer-state memory, but it does not remove differences in physical parameters, active basis, topology, residuals, schedule phase, or future gradients. A reboot therefore provides no general guarantee that an earlier OA advantage will persist, vanish, or reverse.

Proof. After reboot, both methods use the same moment initialization, but their physical states can differ. Their next gradients and all subsequent optimizer states are evaluated at different points, so the ordering depends on the post-reboot geometry and states, not only on the discarded moments. □

A.10 APPLICABILITY TAXONOMY

The analysis distinguishes four qualitatively distinct outcomes:

- **Help:** target-aligned modal error contracts faster under the induced physical metric, reducing finite-budget error.
- **Contraction:** both charts approach the same attainable optimum after the accelerated modes are exhausted, so the gap narrows without negating earlier acceleration.
- **Null:** the chart is identity or the complete update is transported exactly. Orthogonal maps are null for Euclidean descent because $TT^\top = I$, but not generally for coordinate-wise Adam; scalar retuning is null only under the optimizer and ϵ conditions being matched.
- **Reversal:** the active spectrum, basis, topology, residual alignment, objective, or optimizer state changes, making a formerly favorable chart less favorable in the new regime.

The exact claims concern invertible linear charts, mapped feasible sets, and the stated optimization assumptions. They do not establish global nonconvex convergence, universal chart superiority, or practical improvement in every Gaussian system. Section 5 tests sustained finite-budget help in stable two-dimensional fitting, while Appendix E evaluates the anticipated nonstationary boundary in a changing three-dimensional system.

A.11 PROSPECTIVE PREDICTION AUDIT

Before the richer-system and official 3D experiments, we froze twelve result-blind predictions together with explicit support, falsification, and scope criteria. The ledger and test contract were immutable: later evidence was appended without rewriting the original predictions. The initial audit contained six supported, one partially supported, no falsified, and five inconclusive outcomes. Subsequent coherent 2D and 3D experiments sharpened several entries but did not retroactively convert invalid or unavailable tests into successes.

Predictions P1-P6.

Locked prediction	Terminal status	Evidence and boundary
P1: exact paired initialization	Supported	Paired initialization passed throughout the EigenGS-style pipeline and the official 3D fixed-state map. The statement remains conditional on mapping the complete renderer contract.
P2: identity-case null	Supported	The isolated $S = I$ intervention is algebraically inert and passed exact SGD and Adam controls. Hybrid-system results are separate interventions.
P3: curvature-balance predictor	Inconclusive; revised	Condition number alone missed its frozen operational threshold. A useful finite-horizon predictor must also include the actual step, spectral radius, modal projection, target alignment, and curvature drift.
P4: appearance-geometry coupling	Supported locally	The first geometry update matched, and the second followed the predicted sign and η^2 law in all three controlled seeds; halving η changed the effect by a factor of 4.0011. This local test does not identify the full 3D reversal cause.
P5: finite-horizon gap shape	Exact modal result supported	The one-mode curve peaked analytically at 55.4041 updates and discretely at 55, then contracted toward zero; all five frozen targets were reached earlier. The nonlinear 2D and 3D curves are theory-aligned, not certified single-mode trajectories.
P6: coordinate-invariant optimizer control	Supported where defined	Exact transported Newton, Gauss-Newton, and natural-metric steps agreed to at most 1.001×10^{-16}. Independent damping, diagonal approximations, and singular solve choices remain outside the theorem.

Predictions P7-P12.

Locked prediction	Terminal status	Evidence and boundary
P7: OA-initialized official Image-GS	Not validly tested	Direct cross-renderer state transfer was structurally incoherent under the frozen contract. The later coherent hybrid is a different system and is not counted as this test.
P8: fixed-topology Image-GS transfer control	Not validly tested	No direct official Image-GS transfer satisfying the original state-mapping contract was completed. Same-renderer fixed-topology pairs provide separate causal evidence.
P9: adaptive insertion-event amplification	Not validly tested	The direct transfer program closed before a coherent adaptive comparison. No event-amplification claim is made.
P10: 3D all-view equivalence	Supported	The nontrivial spherical-harmonic appearance chart preserved all audited initial views under official rendering within the calibrated numerical envelope. This tests equivalence, not trajectory superiority.
P11: reversal regime	Partially supported; materially instantiated	Exact adverse quadratics passed and official 3D fitting showed broad early help followed by late reversal. The condition-only operational diagnostic remained imperfect, so no complete reversal predictor is claimed.
P12: algorithm-axis versus system-axis hierarchy	Supported	Original OA pairs, coherent hybrid systems, official Image-GS, and official 3D scope evidence remained separately attributed. No complete-system result was relabeled as the original coordinate intervention.

The audit is stronger because it preserves boundaries rather than selecting only favorable outcomes. Exact local predictions received prospective tests; the unsuccessful condition-only predictor was narrowed; and incoherent cross-renderer transfers were marked untested rather than forced into a success or failure label. The resulting claim is analytical and scoped: exact function-preserving coordinates can alter finite-horizon optimization, while the sign and duration depend on step size,

modal alignment, objective geometry, optimizer state, and system stationarity. The frozen condition-only predictor remains inconclusive. The subsequently calibrated SGCL rule (Appendix G) supplies separate prospective selection evidence for 100-step Gaussian fitting, not a validation of the earlier predictor or a guarantee for a new system.

Why the horizon matters. For a local quadratic with physical Hessian $H \succ 0$ and fixed $P \succ 0$, let $K_P = P^{1/2} H P^{1/2}$ have eigenpairs (λ_i, v_i), and $a_i = v_i^\top P^{-1/2}(q_0 - q^\star)$. Then

$$F(q_k) - F(q^\star) = \tfrac{1}{2} \sum_i \lambda_i a_i^2 (1 - \eta\lambda_i)^{2k}. \tag{61}$$

Convergence requires $|1 - \eta\lambda_i| < 1$ on active modes. A chart changes both contraction and target-aligned weights, so condition number alone does not determine finite-budget quality. This motivates finite-horizon calibration, without proving that a first-step-loss feature is necessary or sufficient for selection. An early advantage can contract or reverse; immediate loss reduction need not predict a later endpoint. No fixed nonidentity P dominates all objectives: at step size one, $H = P^{-1}$ favors P, whereas $H = I$ favors identity. These are statements about fixed-metric quadratic descent, not an Adam convergence theorem.

Mapped native-storage step. Under the notation of Section 2, the standard-chart Adam step maps to

$$\Delta q_i = -\alpha \frac{t_i \widehat{m}_i^q}{\sqrt{\widehat{v}_i^q + \epsilon/t_i}}. \tag{62}$$

The covector and optimizer moments must be transformed consistently. This identity does not certify that a renderer-supplied shape covector is the ordinary derivative.

B Full Experimental Protocol

Appendices B-I specify the historical OA and SGCL protocols. The energy-coordinate Rprop study uses the separate renderer conventions, controls, image roles, and delivered-output clocks in Appendices J and R.

Metric and numerical paths. The original six-domain sweep reports unclamped normalized-YCbCr PSNR. Native-resolution DIV2K, the separate production-controller campaign, SGCL and coupled-coordinate confirmation report their specified common encoded-RGB metrics. References to RGB in this appendix apply only to the explicitly RGB-evaluated protocols, not retroactively to the original sweep. Legacy runs use the native supplied shape covector; the later coupled-coordinate comparison uses the shared corrected projection branch and guarded forward described in Appendix J.

This appendix expands the OA protocols in Section 5; the distinct SGCL and non-Gaussian protocols appear in Appendices G and H. It separates the six-domain breadth study, the native-resolution DIV2K study, the isolated hardware study, and the richer matched-system study because they answer different questions and use different experimental units.

B.1 Common intervention, empirical map, and equivalence gates

Every coordinate comparison begins from one physical Gaussian state and two exactly paired coordinate descriptions. In the natural-image experiments, $d = 3$ and $S = \mathrm{diag}(s_Y, s_{Cb}, s_{Cr})$, where each positive entry is the frozen max-minus-min range stored in the corresponding per-channel PCA normalization metadata. The triplet is broadcast over all Gaussian appearance rows and is fixed during online fitting. It can vary only when the offline dataset, resolution, native shape, or basis artifact changes. Each experimental unit records the exact triplet, its singular values, and $\kappa_2(S) = \max_j s_j / \min_j s_j$.

For a fixed-policy causal pair, target, renderer, pixelwise MSE objective, Gaussian primitives, Gaussian budget, physical initialization, represented initial image, optimizer family and hyperparameters, and update budget are matched. Only the chart changes. The practical headline comparisons instead allow each chart to select its own policy using validation data; they are labeled validation-selected

and are not used alone as chart-only attribution. Each run receives a fresh optimizer state unless an experiment explicitly studies optimizer-state transport or reboot.

Before optimization, we require the maximum absolute rendered-image discrepancy between paired charts to remain below 10^{-4}. This numerical gate checks the complete implemented renderer contract, including the matched background scale and fixed offsets; the exact algebraic equivalence is proved in Section 2 and Appendix A. Heldout outcomes are never used to choose a learning rate, optimizer, checkpoint, image identity, capacity, or stopping rule.

Table 5: Appearance-map scales and condition numbers. Exact appearance-map diagnostics for the formal six-domain breadth study. Since the geometry block is identity and every listed scale is below one, $\kappa_2(T) = 1/\min_j s_j$ and $\kappa_2(P_{\mathrm{std}}) = \kappa_2(T)^2$. These are coordinate-map scale ratios, not Hessian condition numbers.

Domain	s_Y	s_{Cb}	s_{Cr}	$\kappa_2(S)$	$\kappa_2(T)$	$\kappa_2(P_{\mathrm{std}})$
Caltech101	0.041831	0.038054	0.039336	1.099	26.278	690.557
CIFAR-100	0.033037	0.035688	0.037244	1.127	30.269	916.222
DTD	0.037004	0.042225	0.044164	1.193	27.024	730.299
EuroSAT	0.036537	0.042588	0.038993	1.166	27.369	749.072
Flowers102	0.038635	0.041267	0.037384	1.104	26.749	715.532
Oxford-IIIT Pets	0.042379	0.040781	0.041172	1.039	24.521	601.293

The 24 native DIV2K shapes used in the formal resource study have $s_j \in [0.01062, 0.01688]$, $\kappa_2(S) \in [1.017, 1.167]$, $\kappa_2(T) \in [62.3, 94.2]$, and $\kappa_2(P_{\mathrm{std}}) \in [3.88{\times}10^3, 8.87{\times}10^3]$. Their smaller long-horizon gains despite larger map-scale ratios reinforce that the map alone is not an outcome predictor; step size, curvature, modal alignment, and horizon remain essential.

B.2 Loss, optimizer, and schedule settings

Online 2D fitting uses pixelwise mean-squared error in the internal fitting domain. There are no perceptual, structural, or learned terms in the optimized loss. Adam and AdamW (Kingma & Ba, 2015; Loshchilov & Hutter, 2019) use $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. AdamW uses decoupled weight decay 0.01; Adam uses no weight decay. The StepLR schedule has step size 20,000 and factor 0.5, so all reported trajectories through 10,000 updates use a constant learning rate. The offline Gaussian-basis construction is a separate preparation stage and is not part of the online optimizer comparison.

The six-domain headline chooses OA Adam at 0.016 and standard AdamW at 0.008. The native DIV2K validation grid includes Adam and AdamW at 0.002, 0.004, and 0.008; it selects OA Adam at 0.008 and standard AdamW at 0.008. Same-optimizer controls keep the optimizer family fixed and select only within the predeclared learning-rate grid for that family.

B.3 Transported-controller and perceptual-control protocol

A separate fresh mechanism campaign compares four arms from the same physical initialization: OA with ordinary global Adam or AdamW; standard storage with the same global optimizer settings; standard storage with the complete transported OA controller; and a chart-informed learning-rate-only approximation. For appearance coordinate t_i, the transported controller uses $\alpha_i^x = \alpha_{\mathrm{OA}}/t_i$ and $\epsilon_i^x = t_i \epsilon_{\mathrm{OA}}$, transports first and second moments, schedules, and restart state, and matches AdamW's decoupled decay product. Geometry remains in the identity block. The learning-rate-only arm uses appearance multipliers γ/t_i while retaining ordinary stored-coordinate epsilon, moments, and decay semantics. Its single scalar γ is selected at step 30 on the 12 DIV2K validation images from $\{0.5, 1, 2\}$, yielding 2 for Adam and 0.5 for AdamW, and is frozen before every heldout protocol.

The campaign evaluates 100 native-resolution DIV2K images with Adam for 5,000 updates, the frozen 24-image subset with AdamW for 5,000 updates, and all 1,200 six-domain images with both optimizers for 400 updates. All arms are freshly executed and must pass the same initial-render gate. In addition to PSNR and SSIM, every endpoint is evaluated with VGG LPIPS v0.1 (Zhang et al.,

2018) on encoded RGB in $[0, 1]$, after one reconstruction clamp and without resizing, cropping, padding, rotation, gamma linearization, or color conversion. The campaign is a separate mechanism control and does not replace the locked primary trajectory in Figure 3.

B.4 SIX-DOMAIN BREADTH STUDY

Datasets and roles. The breadth study uses Flowers102, Oxford-IIIT Pets, DTD, Caltech101, CIFAR-100, and EuroSAT. Each domain contributes exactly 500 training images, 60 validation images, and 200 heldout images, giving 1,200 heldout images in total. Role selection is performed before resizing or color conversion. Stable source identifiers, encoded-file hashes, and decoded RGB hashes are checked for pairwise role overlap.

Role construction.

Domain	Training role	Validation role	Heldout role
Flowers102	Official train	Official validation	Official test
Oxford-IIIT Pets	Deterministic subset of official trainval	Disjoint official trainval subset	Official test
DTD	Partition-1 train	Partition-1 validation	Partition-1 test
CIFAR-100	Deterministic subset of official train	Disjoint official-train subset	Official test
Caltech101	Deterministic all-pool split	Disjoint all-pool split	Disjoint all-pool split
EuroSAT	Deterministic all-pool split	Disjoint all-pool split	Disjoint all-pool split

"Deterministic split" denotes a frozen, pairwise disjoint partition created before any outcome was observed.

Targets are represented at 256×256. Validation selects the optimizer and learning rate independently for each method, and the choices are then frozen before heldout evaluation. This validation-selected practical comparison uses Adam 0.016 for OA and AdamW 0.008 for the standard chart in all six domains. Separate Adam and AdamW controls keep optimizer family fixed. We evaluate updates 0, 30, 100, 200, and 400; the four optimized checkpoints over six domains and 200 heldout images per domain give $6 \times 200 \times 4 = 4{,}800$ image-checkpoint comparisons. The update-zero rows come from independently regenerated roots and are not used as the exact equivalence test; the within-root forward gate above provides that test.

B.5 NATIVE-RESOLUTION DIV2K PROGRAM

Frozen data roles. The program uses the bicubic $\times 2$ low-resolution track of DIV2K. The role split is:

- 785 official training images for shared offline appearance-basis preparation only;
- 12 disjoint official training images for policy selection only: 0135, 0230, 0623, 0260, 0145, 0734, 0326, 0053, 0742, 0664, 0653, and 0640;
- 3 disjoint official training images for bounded pilot checks only: 0417, 0506, and 0079;
- all 100 official validation images, 0801x2.png through 0900x2.png, as the formal heldout set;
- the official test set untouched.

The 12 validation identities are the only images used to select policies or quality targets. The 3 pilot identities are not merged into the final heldout statistics.

Common image and metric domain. Targets retain their encoded RGB values, native dimensions, and orientation. Evaluation applies no resize, crop, padding, rotation, gamma linearization, or color-profile transformation. OA and standard coordinates invert their internal YCbCr representation before common metrics. Every system is then evaluated by float64 RGB PSNR and one frozen Gaussian-window SSIM implementation.

Systems and budgets. The OA-versus-standard trajectory is a paired representation comparison under independently validation-selected policies. Same-optimizer experiments provide the fixed-family causal controls. These paired 2D arms use the same PCA-derived EigenGS-style appearance construction. Complete-system context additionally includes the official GaussianImage Cholesky parameterization, GaussianImage rotation-scale parameterization, and official Image-GS, which optimize direct physical appearance. These complete-system comparisons are not interpreted as matched coordinate effects.

The primary heldout endpoint uses 2,000 active Gaussians, seed 1, and 5,000 online updates. The same 24-image subset was frozen before sensitivity analysis and is reused for 800, 2,000, and 5,000 Gaussian capacities, additional seeds, and longer-horizon checks. For the count-800 official-system implementation, renderer-only zero-opacity padding preserves 800 active and saved Gaussians; the padding contributes no rendered signal or trainable capacity, while its renderer time remains included in recorded timing.

B.6 Richer matched systems

The richer-system study reuses the same nine frozen heldout identities across all four families at 10,000 updates. Each row in Table 16 compares two charts within one surrounding system and matched optimization schedule:

- **Fixed topology:** both runs keep the same fixed 2,000-Gaussian state and receive no allocation events.
- **Shared event schedule:** an exogenous frozen ledger gives both charts the same event times and the same inserted physical states.
- **Image-GS-derived mechanisms:** both charts use the same portable residual-driven insertion mechanism and the same surrounding implementation.
- **Native adaptive allocation:** both charts use one frozen residual-driven rule, evaluated independently on each chart's own trajectory. The rule is matched, but its event times and inserted locations may diverge after the residuals diverge.

The first three comparisons isolate the chart under increasingly rich but explicitly coupled surroundings. The final comparison additionally tests the feedback loop from coordinates to residuals, from residuals to allocation events, and from those events to later optimization. Official Image-GS is an unchanged complete-system reference and is analyzed separately.

The progressive systems begin with 1,000 active Gaussians and allocate four blocks of 250 complete physical states. Inactive storage has zero contribution before activation. Fixed-count invariance and shared-ledger equality are checked where required. The nine-image study is a controlled mechanism-transfer test, not the source of the paper's breadth claim; breadth comes from the separate 1,200-image and 100-image protocols.

B.7 Formal isolated timing and resource protocol

Formal timing and memory claims use five systems, 24 frozen DIV2K identities, six native image shapes, and three fresh-process repetitions per valid unit. The six shapes are 1020×678, 1020×834, 678×1020, 684×1020, 768×1020, and 936×1020. With 24 images, 5 systems, and 3 repetitions, the design contains 360 formal units and 1.8 million optimizer updates.

Each unit runs in a fresh process under exclusive GPU isolation. Phase boundaries are synchronized before timing. The instrumentation records online optimization time, evaluation time, serialization time, end-to-end time, process-level CUDA allocation and reservation, whole-device memory, renderer and online peaks, and host resident memory. First-use compilation is retained in the declared first-step accounting. Offline basis preparation is excluded from online timing and disclosed separately; historical offline timing is descriptive unless measured under the same repeated isolation contract.

Formal OA-versus-standard hardware inference uses paired image-level summaries over the fresh-process repetitions. Whole-device telemetry is definition-matched across systems. Quality targets

are selected from validation trajectories only. If a run does not reach a target, it remains right-censored; no target crossing is extrapolated beyond the observed trajectory. Target-step and target-time summaries therefore use censoring-aware analyses rather than assigning an invented crossing time.

B.8 Statistics and reporting rules

All primary comparisons are paired by heldout identity. We report the paired mean difference and wins, ties, and losses. Paired bootstrap intervals use the frozen resampling contract for the relevant study. Exact paired sign or binomial tests are used for directional confirmation, with Holm correction when a family of related hypotheses is tested. Capacity results use the same frozen 24 identities at every Gaussian count. The richer-system intervals use the same nine identities within each matched family.

The following estimands remain separate throughout the paper:

1. validation-selected practical policy comparisons;
2. fixed-policy OA-versus-standard coordinate effects;
3. complete-system quality-time comparisons against official methods;
4. formal repeated-hardware timing and memory;
5. descriptive campaign-recorded timing;
6. retained-byte accounting and offline preparation cost.

No unavailable field is imputed, no censored target is treated as reached, and no unfavorable heldout outcome is selectively rerun or removed.

B.9 Prospective freezing and reproducibility

Dataset roles, selection policies, capacities, checkpoints, metrics, statistical rules, and prediction criteria are frozen before the corresponding heldout outcomes. Machine-readable manifests record source identities and hashes. The prospective theory ledger and its preoutcome contract are reported in Appendix A.11. Appendix E gives the separate official 3D protocol and endpoint controls. Appendix F records software, environment, and implementation-level reproducibility details.

C Extended 2D Results

This appendix expands the two-dimensional evidence summarized in Sections 5 and 6. It keeps validation-selected comparisons, same-optimizer controls, native-resolution sensitivities, and complete-system context separate. Deltas are OA minus the named comparator; where no alternative is named, the reference is the corresponding standard-coordinate system.

C.1 Six-domain heldout breadth

Table 6 gives every optimized checkpoint from the frozen six-domain study. Optimizer and learning rate were selected independently for each chart using validation data, then fixed before evaluation. The selected policy was Adam at learning rate 0.016 for OA and AdamW at learning rate 0.008 for the standard chart in every domain. OA improves all 200 heldout images in every domain at every optimized checkpoint, for $4{,}800/4{,}800$ positive image-checkpoint comparisons. These independently regenerated roots are not used to certify exact update-zero equivalence; that property is checked by the within-root forward gate in Appendix B.

Table 6: Six-domain heldout breadth in unclamped normalized YCbCr PSNR. Each domain contains 200 heldout images. "Wins" aggregates the four optimized checkpoints, so each row contains 800 paired outcomes.

Domain	n	ΔPSNR_{30}	ΔPSNR_{100}	ΔPSNR_{200}	ΔPSNR_{400}	Wins
Caltech101	200	+2.5619	+7.4891	+9.6729	+10.7596	800/800
CIFAR-100	200	+3.1321	+9.9901	+13.7312	+16.9193	800/800
DTD	200	+2.5532	+7.0589	+8.8234	+9.5137	800/800
EuroSAT	200	+3.2253	+9.9025	+12.8747	+14.8118	800/800
Flowers102	200	+3.1136	+8.8178	+11.0237	+11.9288	800/800
Oxford-IIIT Pets	200	+2.5956	+7.5442	+9.5667	+10.3956	800/800

C.2 Six-domain fixed-policy controls

The headline breadth result lets each chart use its independently validation-selected policy. The same completed campaign also contains a stricter chart-only control. For each optimizer family, one learning rate shared by standard and OA is selected from the joint validation mean, frozen, and applied to all 1,200 heldout images. Combining Adam and AdamW gives 2,400 paired image-optimizer outcomes per checkpoint. Table 7 shows that OA wins every pair through step 200 and 2,398/2,400 pairs at step 400. At step 400, the mean gain is $+11.0805$ dB under Adam and $+9.3890$ dB under AdamW. The two losses are the same EuroSAT identity under Adam (-0.3082 dB) and AdamW (-0.1001 dB); that identity is positive through step 200. These chart comparisons use the recorded native derivative convention. A derivative-only six-image validation diagnostic retains the large gap at its complete 100-step prefix; seven conservatively refused arms leave the full 400-step corrected-derivative population unresolved (Appendix J).

Table 7: Matched-optimizer breadth controls. Six-domain normalized-YCbCr-PSNR fixed-optimizer, shared-learning-rate controls. A single validation-selected learning rate is shared by both charts within each optimizer family. Counts aggregate 1,200 heldout images under Adam and AdamW.

Step	Mean ΔPSNR	Wins	Losses
30	+2.9879	2,400	0
100	+7.7802	2,400	0
200	+9.7050	2,400	0
400	+10.2347	2,398	2

C.3 Effect sizes for the matched endpoint comparison

The 2,398/2,400 win count describes optimizer-image pairs at 400 updates, not effect magnitude. Table 8 reconstructs the distribution from the same fixed-optimizer, shared-learning-rate records. Averaging Adam and AdamW within each image produces 1,200 image means, with 1,199 positive means. The metric is unclamped normalized-YCbCr PSNR, not decoded RGB PSNR. These descriptive distributions neither introduce a new experiment nor establish the effect of using an ordinary rather than the historical supplied shape derivative.

Table 8: Distribution of the matched OA endpoint effect. OA minus standard, in normalized-YCbCr PSNR (dB), at 400 updates. The second row averages both optimizer contexts within each image. Repeats and optimizer contexts are not treated as independent images.

Population	n	Mean	Median	Interquartile range	Wins
Optimizer-image pairs	2,400	10.2347	11.0744	[6.5871, 14.2512]	2,398
Image means	1,200	10.2347	11.2276	[6.6394, 14.4742]	1,199

C.4 TRANSPORTED-CONTROLLER AND PERCEPTUAL CLOSURE

The four production protocols contain 2,524 optimizer-specific rows: 100 native DIV2K/Adam, 24 native DIV2K/AdamW, and 1,200 six-domain rows for each optimizer. These are not 2,524 independent images; the breadth optimizer outcomes are paired. The fresh mechanism campaign in Appendix B.3 tests whether the analytically transported standard-storage controller recovers OA under production Adam and AdamW. Table 9 reports absolute endpoint PSNR. The controller remains within 0.0108 dB of OA in every pooled protocol, while both substantially outperform the ordinary matched standard arm. These common-RGB results use the native supplied-covector implementation; endpoint agreement does not validate its ordinary shape derivative. They are separate from, and do not replace, the locked Figure 3 trajectory.

Table 9: Absolute RGB PSNR in the transport study. Absolute endpoint PSNR for the fresh production mechanism campaign. "Matched STD" uses the same global optimizer settings as OA. The LR-only arm uses the validation-selected chart-informed approximation.

Protocol	Optimizer	n	OA	Controller	Matched STD	LR-only	OA-STD
DIV2K, step 5,000	Adam	100	23.4612	23.4720	22.8353	23.4535	+0.6259
DIV2K, step 5,000	AdamW	24	22.9260	22.9333	22.6591	22.5797	+0.2669
Six domains, step 400	Adam	1,200	30.1457	30.1444	20.7373	30.3691	+9.4083
Six domains, step 400	AdamW	1,200	29.5479	29.5476	21.5439	30.5405	+8.0040

These are fixed-optimizer mechanism comparisons. The locked Figure 3 endpoint instead compares OA with the independently validation-selected standard AdamW policy, so its $+0.3746$ dB practical-policy gap remains unchanged.

Table 10 shows paired endpoint discrepancies and complete perceptual results. The DIV2K/Adam PSNR interval excludes zero despite the small mean difference; these measurements do not constitute a predeclared empirical equivalence test. Intervals containing zero in other metrics do not prove equivalence either. Deltas are OA minus the listed reference, so lower LPIPS is favorable. Every pooled OA-controller LPIPS interval includes zero. The corresponding OA-controller mean SSIM gaps range only from -1.03×10^{-4} to $+8.5\times10^{-6}$, with every paired-bootstrap interval containing zero. In contrast, OA reduces LPIPS relative to ordinary matched standard storage by 0.0314 on the 100-image DIV2K Adam protocol, by 0.0077 on the 24-image AdamW protocol, and by 0.2063 and 0.1752 across the six domains under Adam and AdamW.

Table 10: Paired controller agreement and perceptual improvement. Brackets give paired-bootstrap 95% intervals. LPIPS deltas below zero favor OA; the final column also gives OA perceptual wins against ordinary matched standard storage.

Protocol	OA-controller ΔPSNR	OA-controller ΔLPIPS	OA-matched STD ΔLPIPS (wins)
DIV2K Adam	-0.01077 $[-0.02128, -0.00120]$	$+0.000201$ $[-0.000155, +0.000565]$	-0.03142 (97/100)
DIV2K AdamW	-0.00732 $[-0.04058, +0.01795]$	$+0.000149$ $[-0.000272, +0.000599]$	-0.00770 (19/24)
Six domains Adam	$+0.00130$ $[-0.00040, +0.00348]$	$+0.000020$ $[-0.000001, +0.000040]$	-0.20626 (1,195/1,200)
Six domains AdamW	$+0.00032$ $[-0.00110, +0.00183]$	-0.000019 $[-0.000040, +0.000002]$	-0.17523 (1,195/1,200)

The chart-informed learning-rate-only arm is deliberately stronger than a global learning-rate retuning: it applies the known $1/t_i$ appearance multipliers, leaves geometry at its base rate, and selects only the scalar γ on validation data. It is close to OA on the 100-image Adam protocol, trails OA

by 0.3463 dB on the 5,000-step AdamW protocol, and exceeds OA by 0.2235 and 0.9926 dB in the 400-step six-domain Adam and AdamW protocols. This regime dependence supports the finite-horizon analysis and rules out universal superiority of one fixed chart or controller. OA's practical distinction is that the same geometry is exposed through the representation while retaining an ordinary optimizer interface.

A validation-only block diagnostic directly measures the mechanism. At the first update, ordinary standard storage produces only 1.10% and 1.21% of OA's physical appearance-update RMS under Adam and AdamW, while its geometry-update RMS remains at the OA scale. The transported controller restores both the physical appearance update and the appearance-to-geometry ratio to essentially 100% of OA. Table 11 shows that this agreement persists through step 30.

Table 11: Physical appearance-to-geometry motion. Validation-set physical appearance-to-geometry update ratio relative to OA. Values near one reproduce OA's block balance.

Optimizer	Ordinary STD, step 1	Controller, step 1	Ordinary STD, step 30	Controller, step 30
Adam	0.0110	1.0000	0.0393	1.0000
AdamW	0.0121	1.0000	0.0461	1.0000

C.5 Independent 512-pixel optimizer controls

Four frozen 512-pixel settings provide an independent scale confirmation under fixed Adam and fixed AdamW. Each aggregate contains 96 image-setting rows per method and checkpoint. Table 12 shows exact initialization to numerical precision, a positive mean gain at every optimized checkpoint, and a 100% paired win rate. A separate heuristic manual gradient-scaling control does not reproduce the OA gain. This establishes robustness across Adam and AdamW. The exact diagonal Adam controller is derived in Appendix A; the manual-scaling result evaluates only the tested heuristic and should not be interpreted as an exact conjugate control.

Table 12: Same-optimizer 512-pixel controls. The equivalence column is the maximum absolute paired rendering discrepancy before optimization.

Optimizer	Initial max error	ΔPSNR$_{30}$	ΔPSNR$_{100}$	ΔPSNR$_{200}$	ΔPSNR$_{400}$	Win rate	Manual scaling
Adam	7.2×10^{-7}	+0.1626	+0.5528	+1.2957	+4.5050	100%	Heuristic fails
AdamW	9.5×10^{-7}	+0.1624	+0.5548	+1.2229	+3.7905	100%	Heuristic fails

C.6 Scale, category, and hard-domain checks

Table 13 reports three independently frozen checks beyond the six-domain headline. The Flowers102 experiment uses near-native 512-pixel targets and selects Adam at learning rate 0.016 for both charts. The DTD experiment is deliberately labeled supplemental: it uses a category-balanced, resolution-ranked subset of the official first partition with 141 training, 47 validation, and 47 held-out images, one heldout image per category. Food101 (Bossard et al., 2014) was selected before observing its outcomes as the hard-domain check; it uses 202 heldout images, two per class, and selects OA Adam at 0.016 versus standard AdamW at 0.008. All optimized images improve in all three experiments, and every DTD category and Food101 class has a positive mean gain.

Table 13: Additional scale and domain checks. Values are mean paired PSNR gains in dB.

Study	Heldout n	Δ_{30}	Δ_{100}	Δ_{200}	Δ_{400}	Image checkpoint wins	Group consistency
Flowers102, 512 px	50	+0.1901	+1.0174	+5.9830	+9.7575	200/200	50/50 at every checkpoint
DTD supplemental, 512 px	47	+0.1101	+0.9229	+5.4257	+8.3781	188/188	47/47 categories positive
Food101, 256 px	202	+3.2199	+8.6203	+10.5473	+11.1572	808/808	101/101 class means positive

At step 400, the minimum per-image gain is $+0.9465$ dB for the Flowers102 512-pixel check and $+2.9238$ dB for Food101. The DTD category-mean gains range from $+0.7704$ to $+12.3139$ dB. These results strengthen scale and domain robustness while preserving their protocol-specific roles; the repaired DTD subset is not presented as an unconditioned full-dataset benchmark.

C.7 Native-resolution DIV2K trajectory and endpoint distribution

Table 14 gives the complete set of principal checkpoints for the formal 100-image native-resolution trajectory. The methods start from the same represented image. The mean PSNR gap peaks early and contracts but remains positive through 5,000 updates, matching the finite-horizon interpretation rather than implying different attainable optima.

Table 14: Native-resolution DIV2K convergence. Formal 100-image native-resolution DIV2K trajectory with 2,000 Gaussians.

Step	OA PSNR	STD PSNR	ΔPSNR	OA SSIM	STD SSIM	ΔSSIM	PSNR wins
0	16.1756	16.1756	0.0000	0.3622	0.3622	0.0000	-
100	20.7205	19.3259	+1.3947	0.4643	0.4321	+0.0322	100/100
500	21.7010	20.6737	+1.0273	0.5211	0.4797	+0.0414	100/100
1,000	22.2869	21.4512	+0.8357	0.5571	0.5154	+0.0417	100/100
2,000	22.8404	22.2298	+0.6106	0.5872	0.5569	+0.0303	100/100
3,000	23.1192	22.6461	+0.4731	0.6014	0.5790	+0.0223	100/100
4,000	23.3175	22.9017	+0.4158	0.6108	0.5917	+0.0191	100/100
5,000	23.4642	23.0896	+0.3746	0.6177	0.6006	+0.0171	100/100

At step 5,000, the paired PSNR-gain distribution has minimum, first quartile, median, third quartile, and maximum values of 0.0211, 0.2002, 0.2998, 0.4533, and 1.4560 dB. Its paired-bootstrap 95% interval is $[0.3256, 0.4282]$ dB. SSIM improves on 99 of 100 images; the single loss is -0.000923, while that image still gains $+0.1242$ dB in PSNR. The corresponding SSIM-gain quartiles are 0.0101, 0.0137, and 0.0198, with a maximum of 0.0763.

The 24-image long-horizon sensitivity remains positive at every optimized mean checkpoint and gives $+0.2252$ dB at 10,000 updates. Extra seeds 2 and 3 reproduce 24/24 PSNR wins; across seeds 1, 2, and 3, the standard deviation of the mean endpoint gap is 0.0052 dB. These checks support persistence and seed stability, while the shrinking long-horizon gap remains part of the claimed finite-budget behavior.

C.8 Capacity sensitivity

Table 15 reports absolute metrics as well as deltas on the same frozen 24 identities. Every image improves in both PSNR and SSIM at 800, 2,000, and 5,000 active Gaussians. Campaign-recorded

times are included for completeness but are not formal hardware comparisons; Appendix D reports the isolated protocol.

Table 15: Capacity sensitivity on DIV2K. Capacity sensitivity at step 5,000 on 24 native-resolution DIV2K images.

Gaussians	OA PSNR	STD PSNR	ΔPSNR	OA SSIM	STD SSIM	ΔSSIM	PSNR W/T/L	SSIM W/T/L
800	20.5582	19.4299	+1.1283	0.5123	0.4416	+0.0707	24/0/0	24/0/0
2,000	23.0173	22.6521	+0.3652	0.6136	0.5965	+0.0172	24/0/0	24/0/0
5,000	25.0159	24.5632	+0.4527	0.7061	0.6841	+0.0220	24/0/0	24/0/0

D System and Efficiency Results

This appendix separates three questions that are easy to conflate: the matched coordinate effect, complete-system quality-time context, and formal isolated resource measurement. Only OA versus its matched standard chart identifies the coordinate intervention causally.

Figure 4 in the main text summarizes the breadth, capacity, and system comparisons; the tables below retain the full numerical and resource context.

Table 16: Matched richer-system generality. Matched richer-system generality. Each row compares optimization-aware and standard coordinates within the same system family and optimization schedule. Fixed topology uses no allocation events; the shared event schedule fixes identical event times and inserted physical states; Image-GS-derived mechanisms use the same portable residual-driven insertion policy; and native adaptive allocation evaluates the same frozen residual-driven rule separately on each chart's own trajectory, so the resulting events may diverge. Brackets report the frozen paired-bootstrap interval for the PSNR difference.

System family	Steps	ΔPSNR [95% CI] ↑	PSNR W/T/L	ΔSSIM ↑	SSIM W/T/L
Fixed topology	10,000	+3.2846 [0.9011, 6.4285]	9/0/0	+0.1108	9/0/0
Shared events	10,000	+1.4542 [0.6461, 2.4203]	9/0/0	+0.0526	9/0/0
Image-GS-derived	10,000	+1.1172 [0.6570, 1.5656]	9/0/0	+0.0274	9/0/0
Native allocation	10,000	+1.0826 [0.5461, 1.6903]	9/0/0	+0.0252	9/0/0

D.1 Complete-system endpoint context

Table 17 compares independently selected complete systems on the 100-image DIV2K heldout set at 2,000 active Gaussians and 5,000 updates. GaussianImage Cholesky (GI-C), GaussianImage rotation-scale (GI-RS), and Image-GS (IGS) retain their official system mechanisms and optimize direct physical appearance. For GI-C and GI-RS, the labels describe spatial covariance parameterizations, not PCA appearance. Their rows therefore answer a practical system-level question, not the matched chart question. OA has the best quality among the three roughly 11-12 second systems, whereas IGS reaches substantially higher quality at greater recorded online time.

Table 17: Complete-system comparison on DIV2K. Complete-system context on 100 native-resolution DIV2K images at 5,000 updates. Times are campaign records, not the formal isolated measurements in Table 19.

System	Mean PSNR	Mean SSIM	Online sec.	Active Gaussians
GI-C	23.0077	0.6013	11.5926	2,000
GI-RS	23.0688	0.6041	12.1238	2,000
IGS	25.8652	0.7338	30.7556	2,000
OA	23.4642	0.6177	11.4060	2,000
STD	23.0896	0.6006	11.6301	2,000

The official longer schedules sharpen the frontier. On the frozen 24-image subset, GI-C and GI-RS at 50,000 updates reach 24.3626 and 24.3961 dB, with SSIM 0.6831 and 0.6837, in 121.521 and 126.245 recorded seconds. Official IGS stops according to its native rule between 4,400 and 7,600 updates, averaging 5,608, and reaches 25.8124 dB and 0.7342 SSIM in 36.479 seconds. Its native endpoint is not detectably different from its selected 5,000-update endpoint. These endpoints define the complete-system frontier: OA leads the similarly timed systems, while IGS attains the highest quality at greater online cost.

D.2 Richer-system quality-time boundary

The matched 10,000-update effects in Table 16 establish that OA composes with fixed topology, shared events, transferred Image-GS mechanisms, and trajectory-dependent allocation. Comparison with unchanged official IGS answers a different question. Table 18 shows the key boundary: the OA mechanism-transfer systems can occupy an earlier quality-time frontier, but official IGS attains the stronger late endpoint.

Table 18: Complete-system comparison with Image-GS. Mechanism-transfer systems versus unchanged official IGS on nine images. Time and quality differences are candidate minus IGS. These are complete-system comparisons, not matched coordinate effects.

Candidate	Step	ΔPSNR vs IGS	PSNR W/T/L	Δonline sec.	Interpretation
OM	2,000	+2.2220	7/0/2	-20.884	Early quality-time advantage
ON	2,000	+1.0233	7/0/2	-19.905	Early quality-time advantage
OM	10,000	-1.6154	0/0/9	-103.192	Faster, lower late endpoint
ON	10,000	-1.9338	0/0/9	-98.139	Faster, lower late endpoint

OM denotes OA inside the Image-GS-derived mechanism system; ON denotes OA inside native adaptive allocation. Together, they establish composability and an early Pareto frontier relative to official IGS.

D.3 Formal isolated timing and memory

The formal study contains 360 complete units: five systems, 24 frozen images spanning six native shapes, and three fresh-process repetitions per system-image pair. Each unit runs under exclusive GPU isolation. CUDA phases are synchronized, first-use compilation is retained in the declared accounting, and whole-device telemetry is checked for foreign processes. The primary time estimand is online end-to-end time, including the declared online phases rather than only the optimizer loop.

At the locked 5,000-update quality point, OA improves its matched standard chart by $+0.3467$ dB PSNR and $+0.0171$ SSIM, with 24/24 wins in both metrics. Mean online end-to-end time is 13.9168 seconds for OA and 13.8962 seconds for STD; paired timing and throughput are statistically indistinguishable. Their measured GPU-memory values are exactly equal under every reported GPU definition. Thus OA provides better quality per update and earlier attainment of validation-defined targets without a detectable online-time or GPU-memory penalty.

Table 19: Formal isolated online cost and resources. GPU and host quantities are means in MiB.

System	Online end-to-end sec.	CUDA allocated	CUDA reserved	Whole-device peak	Model state	Host RSS
OA	13.9168	92.062	109.333	918.833	16.943	2762.872
STD	13.8962	92.062	109.333	918.833	16.943	2762.899
GI-C	14.4048	60.577	69.750	869.250	0.313	1362.682
GI-RS	14.9844	60.589	69.750	869.250	0.313	1398.536
IGS	35.4591	116.644	143.000	954.500	0.206	1701.742

Additional phase-matched quantities reinforce the causal OA/STD equality: both have 898.917 MiB online whole-device peak, 893.750 MiB renderer whole-device peak, and 83.768 MiB temporary allocator peak. OA uses more GPU and host memory than GI-C and GI-RS. It uses less GPU memory than IGS in this fixed-count protocol but has a much larger deployed model state and higher host RSS. These measurements expose the system-specific memory tradeoffs alongside the matched OA/STD equality.

D.4 Target attainment, censoring, and offline cost

Quality targets are selected from validation trajectories only. Unreached targets remain right-censored, and crossing times are interpolated only within observed intervals. Under this contract, OA reaches frozen standard-chart and GaussianImage targets earlier in the censoring-aware analysis. No crossing is extrapolated and no unreached target is assigned an artificial time.

OA and STD share an offline appearance prior. Across the 24 native shapes used by the formal study, the retained artifacts total 52,212,713,622 bytes and use 785 source images per shape. Recorded historical wall time is 1,314.182 seconds for PCA fitting, 11,365.899 seconds for Gaussian-basis construction, and 14,184.058 seconds in total. The retained-byte count is exact, but these times were not rerun under the fresh-process isolation protocol and are therefore descriptive. Offline preparation is consequently reported as a separate cost rather than folded into the online efficiency estimand.

E Prospective 3D Stress Test

Purpose and protocol. The main paper studies two-dimensional Gaussian image fitting. This bounded 3D extension tests whether a nontrivial exact appearance chart transfers the early finite-horizon mechanism into a system whose spherical-harmonic basis, geometry, optimizer state, and topology evolve during training. It uses Mip-NeRF 360 *bonsai* (Barron et al., 2022), Tanks and Temples *truck* (Knapitsch et al., 2017), and Deep Blending *playroom* (Hedman et al., 2018), one scene per benchmark family, seed 1, and three topology protocols: fixed, shared-event, and trajectory-dependent adaptive. The primary static-chart package therefore contains nine scene-topology units, each run to 7,000 iterations with checkpoints at 0, 500, 1,000, 1,500, 2,000, 3,000, 4,000, 5,000, 6,000, and 7,000. All comparisons preserve every audited initial rendered view.

The exact non-PCA map. The official spherical-harmonic basis through degree 3 factors as $\phi(v) = Kp(v)$, with fixed signed, nonzero diagonal normalization constants K and view direction v. For per-Gaussian RGB coefficients A_{SH}, define $\psi(v) = K^{-1}\phi(v) = p(v)$ and $U_{\mathrm{SH}} = A_{\mathrm{SH}}K$. Then

$$A_{\mathrm{SH}}\phi(v) = U_{\mathrm{SH}}\psi(v), \qquad A_{\mathrm{SH}} = U_{\mathrm{SH}}K^{-1}. \tag{63}$$

The map uses the official degree-wise source constants, is compatible with active-degree prefixes and the DC/rest optimizer split, and adds no appearance capacity. Complete pre-clamp colors, clamp flags, compositing, background, exposure, and final clamp are paired under the fixed state. Densification, pruning, opacity reset, and geometry are not absorbed into this appearance identity. Because K is signed, the positive-diagonal Adam formula in the main text is not applied to it without the corresponding sign-aware gradient/state map.

Table 20: Endpoint summary for the prospective 3D stress test. Deltas are optimization-aware minus the matched standard chart in source-balanced PSNR. Static rows report the exact source-balanced means shown in the paper. PACO's early values are summarized approximately because its controller transitions at iteration 1,000. Exact PSNR, SSIM, and LPIPS (Zhang et al., 2018) deltas for the bounded controls appear in Table 21. The text below specifies the controller events, measured overhead and scope of these bounded tests.

Control	Topology protocol	ΔPSNR at 500	ΔPSNR at 7,000	Interpretation
Static exact chart	Fixed topology	+0.297	-0.185	Early gain, late reversal
Static exact chart	Shared event schedule	+0.284	-0.195	Early gain, late reversal
Static exact chart	Trajectory-dependent adaptive	+0.283	-0.309	Early gain, late reversal
PACO controller	Fixed topology	$\approx +0.20$	-0.035	Identity handoff, near parity
PACO controller	Shared event schedule	$\approx +0.20$	-0.061	Identity handoff, near parity
PACO controller	Trajectory-dependent adaptive	$\approx +0.20$	+0.007	Identity handoff, near parity
OABR optimizer reboot	Fixed topology	+0.205	-0.073	Reboot alone is insufficient

Table 21: Perceptual results in the bounded 3D study. Exact source-balanced perceptual boundary for the bounded 3D controllers. Deltas are candidate minus the matched standard control; higher PSNR and SSIM are better, while lower LPIPS is better. All four controls improve all three metrics at iteration 500. At iteration 7,000, LPIPS is unfavorable in every row, including the native PACO case whose PSNR is within 0.008 dB of parity.

Control	Topology	Iteration	ΔPSNR	ΔSSIM	ΔLPIPS
PACO	Fixed topology	500	+0.203579	+0.003205	-0.002638
PACO	Fixed topology	7,000	-0.034609	-0.001480	+0.001932
PACO	Shared event schedule	500	+0.202317	+0.003296	-0.002656
PACO	Shared event schedule	7,000	-0.061361	-0.001723	+0.001877
PACO	Trajectory-dependent adaptive	500	+0.200980	+0.003265	-0.002744
PACO	Trajectory-dependent adaptive	7,000	+0.007142	-0.000603	+0.001752
OABR	Fixed topology	500	+0.204623	+0.003297	-0.002734
OABR	Fixed topology	7,000	-0.073035	-0.001584	+0.002367

Static-chart boundary. The exact static chart improves source-balanced PSNR at iteration 500 in every topology protocol, but its ordering reverses by iteration 7,000. This is the direct nonstationary-boundary test: a chart that is favorable in the early low-order appearance regime need not remain favorable after the active basis, geometry, optimizer state, and topology have evolved. Because the bounded study has one scene per benchmark family and one seed, it is mechanism and limitation evidence rather than a statistically powered 3D benchmark claim.

PACO identity handoff. Pareto-Aligned Chart Optimization (PACO) is a training-only diagnostic controller with two components: a degree-wise exact Adam realization of the chart update and a controller that evaluates source-independent training-camera signals. It can return each spherical-harmonic degree block to identity without changing the current physical parameters. Every controller instance selected identity at iteration 1,000 through the frozen identity-only verified fallback; all 45 subsequent recorded decisions retained identity. Thus these outcomes do not demonstrate that a flexible search learned a successful nonidentity chart. At iteration 7,000, its PSNR differences are -0.035, -0.061, and $+0.007$ dB under fixed-topology, shared-event, and trajectory-dependent adaptive protocols, respectively. Thus, all three protocols lie within 0.061 dB of parity. PACO therefore brings the late PSNR endpoint near parity, while the complete perceptual, time-to-target,

runtime, and memory criteria remain neutral or standard-favoring. Aggregate online overhead is 3.199220%, and the maximum paired peak-reserved-memory increase is 618,659,840 bytes. Only 3/9 units are nonworse in interpolated steps to target and 2/9 in online time.

Optimizer reboot. Optimization-Aware Burn-In Reboot (OABR) tests whether the early transformed trajectory produces a better physical state whose later disadvantage is carried only by Adam history. At the reboot point, the physical tensors and rendered function are preserved while Adam moments and local step counters are reset under a matched contract. The early advantage remains, but the iteration-7,000 mean is -0.073 dB; all three source endpoints also have lower SSIM and higher LPIPS than the matched standard reboot. Adam history is therefore not the sole cause of the late reversal.

Interpretation. The static, identity-handoff, and optimizer-reboot results support one coherent conclusion. The exact chart improves the early low-order appearance regime, but a fixed coordinate choice can become mismatched as the active basis, geometry, optimizer state, and topology evolve. The bounded controls expose a limitation consistent with nonstationarity, but do not identify a unique reversal cause or establish stationarity as a necessary or sufficient condition. They motivate future state- and duration-aware chart design without establishing a practical 3D improvement.

F Reproducibility Details

This appendix records the original OA implementation and evaluation contracts; later extensions are specified in Appendices G, J, P and R. The appendices provide the analytical and experimental specifications.

F.1 Minimal coordinate implementation

The OA natural-image implementation uses $S = \mathrm{diag}(s_Y, s_{Cb}, s_{Cr})$ with $s_j = m_j^{\max} - m_j^{\min}$ from frozen per-channel PCA normalization metadata. For every appearance row affected by S, it:

1. constructs one physical Gaussian initialization (C_0, g_0) using the standard EigenGS-style initializer;
2. retains the standard arm with colors C_0, fixed decoder shift, fixed decoder scale S, and unit background factor;
3. forms $U_0 = C_0 S^\top$, folds the fixed decoder shift into the image mean, replaces the online shift and scale by zero and identity, and applies the same S to the matched background contribution;
4. leaves every geometry, opacity, covariance, renderer, target, and capacity tensor unchanged;
5. initializes fresh optimizer state for each arm and optimizes the same pixelwise MSE for the same update budget.

Every affected appearance channel must be mapped. If appearance is constrained, the feasible set must be transported by the same bijection. Before optimization, the arms are evaluated in one forward pass and must satisfy the 10^{-4} maximum rendering-discrepancy gate from Appendix B. The reported implementation tests cover identity, paired forward equivalence, background and offset equivalence, update-count equality, and state-shape equality.

For every offline basis, the experiment records contain the exact S triplet, singular values, $\kappa_2(S)$, dataset and resolution scope, PCA metadata source, and sharing scope across targets. They also record the online objective, optimizer, β values, ϵ, weight decay, scheduler, parameter groups, and validation grid.

F.2 Exact optimizer-controller relation

For the empirical positive diagonal map, the representation change has an exact optimizer realization. To reproduce a standard-chart global Adam step in physical coordinates, use $\alpha_i = \alpha t_i$ and $\epsilon_i = \epsilon / t_i$. To reproduce global OA Adam in the standard chart, use $\alpha_i^x = \alpha_{\mathrm{OA}} / t_i$ and $\epsilon_i^x = t_i \epsilon_{\mathrm{OA}}$.

With paired moments and no additional chart-dependent operation, the trajectories remain conjugate in real arithmetic; AdamW additionally requires matching the decay products. The executable controller transports coordinate-wise learning rates, epsilon, moments, schedules, decoupled decay, and restart state. It passes both mapping directions in deterministic float64 tests, a bounded CPU renderer test, and the practical production endpoint campaign in Appendix C.4. The learning-rate-only arm instead changes only the appearance learning-rate numerator through γ/t_i multipliers and retains ordinary epsilon, moment, and decay semantics; γ is selected on validation data as specified in Appendix B.3. Heuristic manual gradient scaling remains a separate nonconjugate control.

F.3 Data, image, and metric contract

Appendix B specifies every role construction and heldout identity policy. Role assignment occurs before image conversion, and source identifiers plus encoded and decoded-content checks prevent overlap. The native DIV2K program preserves encoded RGB values, orientation, and native dimensions; it applies no crop, resize, padding, rotation, gamma linearization, or color-profile transformation. Intermediate YCbCr outputs are inverted before common evaluation.

The common evaluator computes PSNR in float64 encoded RGB. The original normalized-YCbCr sweep is separately labeled in Appendix C.1; its scores are not RGB scores. For common-RGB 2D evaluation, SSIM is computed in float64 separately on the three encoded-RGB channels with data range one. The normalized Gaussian window has size 11×11, standard deviation 1.5, and truncation parameter 3.5. Filtering uses reflection at the boundary; a five-pixel border is then excluded from the spatial mean. Weighted population covariance is used without sample-covariance correction, with $K_1 = 0.01$ and $K_2 = 0.03$, so $C_1 = 10^{-4}$ and $C_2 = 9 \times 10^{-4}$. The three channel means are averaged equally. Targets are not clamped; reconstructions are clamped once to $[0, 1]$ before PSNR/SSIM. This specifies the shared evaluator used by the energy confirmations, not the historical normalized-YCbCr PSNR or the separate 3D evaluator. The fresh 2D mechanism campaign evaluates every endpoint with the official VGG LPIPS v0.1 implementation (Zhang et al., 2018) on native encoded RGB in $[0, 1]$, using one reconstruction clamp and no spatial transformation. For the energy studies, the authenticated Graphdeco VGG v0.1 path casts the clamped RGB input to float32 and applies $2x - 1$ before network evaluation (Appendix T); this provenance check does not re-audit every historical evaluator. The 3D scope study reports LPIPS under one shared evaluator together with PSNR and SSIM. Metric code, peak-value conventions, channel order, and aggregation rules are fixed before heldout analysis. Figures are generated from machine-readable aggregate and paired files rather than transcribed manually.

F.4 Policies, pairing, and statistical reconstruction

Validation data alone selects optimizer family, learning rate, quality targets, and reported operating points. After selection, each policy is frozen and applied once to the heldout role. Fixed-policy pairs share target identity, physical initialization, Gaussian budget, renderer, loss, optimizer contract, and update budget. Richer-system pairs additionally obey the event-matching contract in Appendix B. Each record includes system, policy, image identity, seed, checkpoint, active Gaussian count, exact S, quality metrics, and declared timing fields.

The experimental analyses reconstruct paired means, wins, ties, losses, bootstrap intervals, exact sign or binomial tests, and Holm correction from per-image records. They also reconstruct controller endpoint agreement, absolute PSNR, SSIM, and LPIPS, the validation-only block diagnostic, and the frozen learning-rate-only selection. Target-attainment scripts preserve right censoring and forbid extrapolation beyond the final checkpoint. Validation-selected policy, fixed-policy coordinate, controller, complete-system, campaign-timing, formal-hardware, offline-cost, and retained-byte estimands remain distinct.

F.5 Software and hardware environments

The core EigenGS-style experiments use a pinned renderer environment with `gsplat` 0.1.3. Formal cross-system experiments use Python 3.12.13, PyTorch 2.11.0 with CUDA 12.8, official vendored `gsplat` 1.4.0, and a pinned fused-SSIM dependency. The environments remain separate to preserve the implementation contract of each official system.

Formal 2D timing and memory are measured on an NVIDIA GeForce RTX 5090 Laptop GPU in fresh, exclusive processes with synchronized phase boundaries, retained first-use compilation, allocator statistics, whole-device monitoring, and host-RSS sampling. These environment and measurement definitions apply to the formal resource study; other studies retain their own explicitly reported timing boundaries.

G Shared-Gradient Chart Lookahead: Frozen Protocol and Complete Results

G.1 Exact construction and full-budget execution

The physical appearance has rows U and the optimizer-facing chart has rows $X = UB^{-\top}$. At the common initialization, $G_X = G_U B$. For fresh moments,

$$m_1 = (1-\beta_1)G_X, \quad v_1 = (1-\beta_2)G_X^{\odot 2}, \quad \widehat{m}_1 = G_X, \quad \widehat{v}_1 = G_X^{\odot 2}. \tag{64}$$

Consequently Equation 10 is the exact real-arithmetic first Adam/AdamW update, including stabilizer and decoupled decay. The implementation explicitly forms and bias-corrects the moments; it does not approximate the adaptive step by its zero-ϵ limit. Geometry is in the identity chart and its first update is common. Equality at initialization therefore supplies all candidates from one physical backward pass.

Deployment procedure. Given an initialized physical model and target, SGCL (1) evaluates the initial loss and physical gradient; (2) constructs candidate charts and the common geometry update; (3) evaluates candidate first-step losses using temporary physical states and forward passes only; (4) selects the chart using the frozen score below; and (5) installs that chart at the original physical state, applies the shared gradient through an ordinary first optimizer step, and performs the remaining 99 updates. The temporary candidate states never mutate a live optimizer. The first installed update must agree with its analytical prediction, and every arm must have exactly 100 updates with zero discarded updates. No JVP is used. Baseline counterfactual trajectories are run for research evaluation, not for the deployment decision.

A scalar *appearance* chart $B = sI_3$ is not a scalar rescaling of the complete model: its full map is $\mathrm{diag}(sI_{3N}, I_g)$. It changes appearance relative to geometry. Equation 4 gives its exact positive-scalar Adam/AdamW realization under the stated state and decay conditions; the operational question is choosing the block geometry rather than asserting that representation storage is the only implementation.

G.2 Candidate library and rendered block balance

Let Y_0 be the common rendering, Y_g the rendering after the common geometry-only first update, and Y_a the rendering after the identity-chart adaptive appearance-only update. For AdamW, the appearance decay contribution is removed from the latter and audited separately. Define

$$r = \frac{\|Y_g - Y_0\|_F}{\max(\|Y_a - Y_0\|_F, 10^{-18})}, \quad \bar{r} = \min(16, \max(0.25, r)), \quad s_\rho = \exp\big((1-\rho)\log 4 + \rho \log \bar{r}\big). \tag{65}$$

The three regularized rules use $B = s_\rho I_3$ for $\rho \in \{0.25, 0.50, 0.75\}$. Clipping occurs *before* log-space shrinkage. The eight selectable rules are these three, I_3, $4I_3$, $8I_3$, $16I_3$, and a scale-8 gradient-eigen rotation. The last uses an orthonormal eigenbasis of the three-channel physical gradient second moment with zero eigenvalue whitening power. Inherited standard coordinates are an additional research reference, not a selectable rule. A fixed rule may be target-dependent through r; "fixed" means its construction rule is selected once, not that its realized matrix is identical across targets.

The development lattice also included scalar 2, determinant-normalized covariance shapes, quarter-power eigen charts, and low-rank gradient shapes at matched outer scales. Only the outcome-blind operational library was carried into confirmation. These exploratory dense alternatives did not establish a Gaussian shape advantage. The first prospective policy selects only scalar charts; the new-domain policy selects the eigen rotation on 2/256 optimizer units and scalar charts on the remaining

254. The result is primarily adaptive appearance-scale selection, not demonstrated discovery of a generally superior dense Gaussian chart.

G.3 Frozen score, coefficients, and development roles

The predictor is $u(B) = b + \sum_{j=1}^{10} w_j \phi_j(B)$, trained to predict 100-step PSNR gain relative to $4I$. Let L_0 be initial training loss, L_B the hypothetical first-step loss, $s = |\det B|^{1/3}$, $\ell = \log(s/4)$, and $Q = B/s$. Let

$$h = \log \frac{\max(\|G_U\|_F, 10^{-18})}{\max(\|G_g\|_2, 10^{-18})}, \qquad z = \mathbf{1}\{\text{AdamW}\},$$
$$o = \frac{\|Q - \operatorname{diag}(Q)\|_F}{\|Q\|_F}, \qquad e = \|Q^\top Q - I\|_F. \tag{66}$$

Here G_g concatenates the position and Cholesky-geometry gradients. The ten unstandardized features are specified with their fitted coefficients in Table 22, using the stored raw-feature coefficients. The features use $\log\max(r, 10^{-18})$ and $\log\max(\kappa_2(B), 1)$ where appropriate.

Table 22: Frozen SGCL features and coefficients. Frozen SGCL feature map and stored raw-feature ridge coefficients. The displayed decimal representations preserve the supplied numerical parameters.

j	Feature $\phi_j(B)$	w_j
1	$(L_{4I} - L_B)/\max(\lvert L_0\rvert, 10^{-18})$	+0.16385070964963422
2	ℓ	+0.0041453634186663325
3	ℓ^2	−0.12982475370981858
4	$\ell \log\max(r, 10^{-18})$	+0.048724223979023505
5	ℓh	−0.045929912617550042
6	ℓz	−0.024838277942260286
7	$\log\max(\kappa_2(B), 1)$	−0.027745023947292887
8	o	−0.013372521420979777
9	e	−0.01607678473466749
10	oz	+0.0049550089922128487

The intercept is $b = -0.020754156668509174$ and the ridge regularization parameter is 100. Candidates are ordered by decreasing utility. Exact ties use the fixed priority I, scale-8 eigen rotation, $16I$, $4I$, $8I$, RB-0.25, RB-0.50, RB-0.75, which is the lexical order of the original identifiers. If the best candidate differs from $4I$ but exceeds its score by less than $\delta = 0.01446052745764661$ dB, the policy falls back to $4I$. This margin is 0.1 times the development RMSE (0.1446052745764661 dB); it is a calibrated abstention heuristic, not a confidence bound or an optimality guarantee.

Development used 48 images from eight domains: the six breadth domains, DIV2K, and Food101. Adam and AdamW form paired observations, not separate independent images. Grouped leave-one-domain-out evaluation kept these pairs together; endpoint utility and model selection were domain-balanced. Horizon and library reduction were development decisions: the confirmed policy targets 100 updates. Results at 30 and 400 updates examined during development are not new heldout confirmations. The policy, candidate rules, margin, roles, and comparisons were frozen before either prospective endpoint campaign. No domain label is an input feature.

G.4 Initial prospective study

The first prospective set contains 96 canonical images (16 per breadth domain), disjoint from the 48 development identities, and 192 optimizer-specific units. These are same-domain, outcome-sequestered identities: the source archives can contain unrelated prior outcomes, but those outcome records were not consulted for development, identity selection, power analysis, or the frozen contract. Adam and AdamW share learning rate 0.016, $(\beta_1, \beta_2) = (0.9, 0.999)$, and $\epsilon = 10^{-8}$; decay is zero for Adam and 0.01 for AdamW. Each candidate and the policy receive 100 updates. The frozen estimand averages Adam/AdamW within each image and domains equally. Intervals use 50,000

domain-stratified percentile-bootstrap replicates, seed 20260904. This preserves image-level dependence and does not turn multiple optimizers into additional images.

Table 23: Initial prospective SGCL comparisons. Initial prospective SGCL study: 96 images, six domains, 192 optimizer units. Differences are selector minus comparator at 100 updates.

Comparator	ΔPSNR [95% CI]	ΔSSIM [95% CI]	ΔLPIPS [95% CI]
OA (I)	+0.7727 [+0.6299, +0.9266]	+0.03074 [+0.02613, +0.03552]	-0.01616 [-0.01875, -0.01376]
Scalar $4I$	+0.0895 [+0.0682, +0.1137]	+0.00400 [+0.00313, +0.00494]	-0.00233 [-0.00298, -0.00173]
Scalar $8I$	+0.1339 [+0.0717, +0.2091]	+0.00355 [+0.00257, +0.00462]	-0.00368 [-0.00552, -0.00217]

Against scalar $4I$, the image-level win/tie/loss count is 69/15/12 with 0.001 dB tolerance. Every domain and optimizer mean is positive. Selected optimizer-unit counts are: scalar 8, 59; RB-0.75, 47; scalar 4, 37; RB-0.25, 28; scalar 16, 13; and RB-0.50, 8. Seventeen decisions abstain to scalar 4. Mean candidate-oracle opportunity over scalar 4 is 0.125841 dB and residual regret is 0.036377 dB, yielding 71.1% recovery. This oracle uses observed endpoints only for diagnosis and is never available to the selector.

G.5 New-domain confirmation and stronger comparators

The second campaign freezes the same policy and tests FGVC Aircraft, STL-10, SVHN, and GTSRB. Per domain, 500 training images build the offline basis, 16 validation images select fixed rules, and 32 heldout images supply confirmation: 128 images and 256 optimizer units. Aircraft uses official train/validation/test partitions. For the other domains, training and validation roles are disjoint within the official training partition, with heldout targets from official test. Domain availability and roles are settled without heldout outcomes. Images are converted to RGB and fitted to 256×256 using the frozen bicubic `ImageOps.fit` preprocessing before basis construction and fitting. This is controlled image reconstruction, not a native-resolution or classification benchmark.

The two predeclared co-primary baselines are RB-0.50, chosen in prior development, and a rule selected on each new domain's own validation outcomes, separately for Adam and AdamW. Both are chosen from the tested library, not from the confirmation results. Aircraft selects RB-0.75 for both optimizers; GTSRB selects RB-0.25 for both; STL-10 and SVHN select RB-0.50 for Adam and RB-0.25 for AdamW. The greedy baseline selects the lowest hypothetical first-step training loss. The predeclared requirements were superiority to RB-0.50 and noninferiority to the domain-selected rule. The observed interval against the latter is also strictly positive.

Against RB-0.50, image-level wins/ties/losses are 87/30/11; against the domain-selected rule they are 94/20/14. The corresponding domain-mean gains are $0.0682/0.0370$ dB (Aircraft), $0.0814/0.1021$ (GTSRB), $0.0523/0.0670$ (STL-10), and $0.1110/0.1210$ (SVHN), with the first value against RB-0.50. Adam improves by $0.0841/0.0818$ dB and AdamW by $0.0724/0.0817$ dB. Oracle opportunity over RB-0.50 is 0.136127 dB, residual regret 0.057904 dB, and recovery 57.46%. For this evaluated library, population, horizon, and objective, the remaining candidate-oracle opportunity is only 0.057904 dB beyond SGCL. It is an empirical ceiling on selecting among these fixed candidates, not on alternative policies or coordinate methods. Selected counts are scalar 8, 73; RB-0.75, 69; RB-0.25, 49; scalar 16, 36; scalar 4, 25; rotation at 8, 2; and RB-0.50, 2. The between-image bootstrap does not quantify uncertainty over an unlimited population of domains, seeds, or implementation reruns.

G.6 Numerical equivalence, restart, and cost

The external campaign records zero maximum initial-render discrepancy, maximum chart roundtrip error 9.5367×10^{-7}, and maximum analytical/live first-step loss discrepancy 4.7684×10^{-7}. All 256 units have complete budgets and zero discarded updates. Restart checks preserve selector choice and ranking, but CUDA endpoint trajectories are not byte-identical: maximum PSNR drift is 0.0017084 dB for the selector, 0.0071463 dB for RB-0.50, and 0.0272325 dB across candidate controls. An

Table 24: New-domain SGCL paired comparisons. New-domain confirmation: 128 images, four domains, 256 optimizer units. Full paired endpoint intervals; lower LPIPS is favorable.

Comparator	ΔPSNR [95% CI]	ΔSSIM [95% CI]	ΔLPIPS [95% CI]
New-domain validation rule	+0.0818 [+0.0654, +0.0990]	+0.00312 [+0.00251, +0.00376]	-0.00218 [-0.00284, -0.00158]
Prior RB-0.50	+0.0782 [+0.0615, +0.0962]	+0.00248 [+0.00198, +0.00299]	-0.00171 [-0.00220, -0.00124]
Greedy first-step loss	+0.5791 [+0.4588, +0.7047]	+0.02362 [+0.01950, +0.02799]	-0.01040 [-0.01249, -0.00830]
Scalar $4I$	+0.1838 [+0.1537, +0.2150]	+0.00720 [+0.00618, +0.00823]	-0.00329 [-0.00403, -0.00257]
Scalar $8I$	+0.2298 [+0.1212, +0.3680]	+0.00496 [+0.00349, +0.00670]	-0.00436 [-0.00677, -0.00247]
OA (I)	+1.1506 [+0.9814, +1.3214]	+0.04166 [+0.03692, +0.04647]	-0.01896 [-0.02158, -0.01632]
Inherited standard	+7.7077 [+7.4639, +7.9455]	+0.25223 [+0.23639, +0.26797]	-0.12881 [-0.13562, -0.12211]

Table 25: Absolute new-domain SGCL metrics. Absolute 100-step new-domain metrics, equally averaging domains and averaging optimizer pairs within images. These are the same runs as Table 3, not an additional experiment.

Method	PSNR (dB)	SSIM	LPIPS
SGCL	25.833842	0.497047	0.511408
New-domain validation rule	25.752064	0.493931	0.513588
Prior RB-0.50	25.755619	0.494567	0.513116
Greedy first-step loss	25.254751	0.473429	0.521809
Scalar $4I$	25.650086	0.489845	0.514695
Scalar $8I$	25.604030	0.492082	0.515772
OA (I)	24.683209	0.455382	0.530369
Inherited standard	18.126180	0.244820	0.640222

initial byte-equality validation requirement failed; the validation-only repair distinguishes exact decision replay from bounded numerical trajectory agreement. It does not change the frozen policy, returned outcomes, bootstrap estimates, or scientific configuration. Mathematical chart equivalence must not be conflated with bitwise renderer determinism.

Operational timing charges the common forward/backward, balance renders, chart construction, candidate forward scoring, decision, and the full selected trajectory. Unselected full trajectories, research-arm initialization, and endpoint metric evaluation are reported separately. In the initial study, warm overhead excluding the first campaign unit is 10.88% (95th percentile 21.02%); first-unit cold overhead is 90.49%. A separate balanced-order, timing-only development audit gives mean paired overhead 6.41%, but order effects are substantial; it is not a replacement for the prospective measurement. New-domain warm overhead is 9.94% against scalar 4 (95th percentile 19.35%) and 5.57% against RB-0.50. These are comparator-relative costs, not a formal matched-wall-time quality result. Allocator peaks recorded while research arms coexist are not deployment-memory superiority evidence. Offline selector calibration and domain-specific basis preparation are not free and are outside the per-target online timing.

Separate joint-block and deployment study. The two campaigns above retain their original appearance-chart library, geometry settings, and timing definitions. Appendix I reports a later, disjoint-image study with refitted candidate-loss ablations, broader joint-block controls, and charged equal-time evaluation. Its stronger comparator is not the fixed-rule comparator in Tables 3 or the original campaign tables; the studies must not be pooled as one test.

G.7 Original feature normalization and invariant-feature audit

The original regression was fitted on 16 development candidates, not only the reduced eight-rule confirmation library. The covariance, fractional-eigen and low-rank shapes in that fitting population make log condition number and orthogonality defect nonconstant. Their nonzero coefficients in Table 22 therefore do not imply that a constant feature was trained on the eight-rule library.

For 1,536 development rows (48 images, two optimizers, 16 candidates), feature means and population standard deviations are unweighted across rows. Ridge fitting then gives domains equal total weight, with sample weights normalized to mean one and regularization 100. If $\widetilde{w}$ and $\widetilde{b}$ are fitted on standardized features, deployment uses $w_j = \widetilde{w}_j/\sigma_j$ and $b = \widetilde{b} - \sum_j \widetilde{w}_j\mu_j/\sigma_j$ directly on raw features. There is no online standardization using the reduced candidate library. The reconstruction of this fit agrees with stored raw coefficients within 1.06×10^{-15} and the intercept within 4.2×10^{-17}.

Table 26: Original SGCL training standardization. The displayed training statistics explain the offline standardization. Deployment uses the raw-feature coefficients in Table 22, not these rounded statistics. The original policy is preserved, not refitted for confirmation.

Feature index in Table 22	Mean μ_j	Standard deviation σ_j
1	−0.276224402200	0.622864498602
2	0.289898431663	0.654443531678
3	0.512337436836	0.600783480536
4	0.932054727310	2.129175562714
5	−0.864780137778	1.995259203643
6	0.144949334296	0.484931420589
7	0.154906067136	0.212542386655
8	0.240925818978	0.373931563671
9	0.242860903047	0.327425159519
10	0.120462909506	0.290557601433

For the eight deployed rules, features 7 and 9 are zero in exact arithmetic. The earlier 96-unit development replay changed no decision when those two features were set to zero. We additionally recover all 448 original prospective feature sets and selection records: 192 from the six-domain study and 256 from the new-domain confirmation. Applying the unchanged raw coefficients, intercept, candidate list, lexical tie rule and strict fallback margin reproduces every stored ranking and decision

exactly. Setting only the two invariants to zero changes 0/448 decisions. Their maximum absolute score contributions are 3.5992×10^{-9} and 3.4589×10^{-9} dB in the two campaigns. This analysis covers all 3,584 recorded candidate rows. This closes the historical feature-replay limitation without refitting, restandardizing, rendering or substituting the separate Appendix I policies. It certifies those saved decisions, not all future numerical inputs or SGCL's practical superiority.

H Non-Gaussian Transfer and the Scale-Shape Boundary

H.1 Exact full-head chart and common protocol

For a neural field with a linear RGB head, write $y(z) = Wh_\theta(z) + b$. For any fixed invertible $B \in \mathbb{R}^{3\times3}$, define

$$\widetilde{W} = B^{-1}W, \qquad \widetilde{b} = B^{-1}b, \qquad y_B(z) = B\big(\widetilde{W}h_\theta(z) + \widetilde{b}\big) = y(z). \tag{67}$$

Both weight and bias are mapped. The body remains trainable and is identical at the paired initial state; it is not frozen, nor required to follow the same later trajectory. The complete continuous function, capacity, parameter count, initialization, objective, and coordinate batches are paired. This is a three-channel output chart, *not* whitening of hidden features or a change to the network architecture.

Architecture and base learning rate are selected on development data before confirmation. The Fourier-feature MLP uses three hidden layers of width 96, 64 Fourier features with scale 4, 31,299 parameters, and learning rate 0.003. SIREN uses four hidden layers of width 128, first-layer frequency 30 and hidden frequency 1, 50,307 parameters, and learning rate 0.001. Adam and AdamW use the same betas and stabilizer as the Gaussian studies; AdamW decay is 0.01, applied to the complete head including bias under the same policy as other parameters. Targets are 256×256; the objective is unclamped RGB MSE with a linear output head, not an additional perceptual loss. A deterministic image/architecture-specific schedule draws 8,192 pixel coordinates uniformly with replacement per update, identically across paired chart arms. The physical bias is initialized to the target mean in every arm. Reconstructions are clamped once for the common endpoint metrics.

In the first campaign, 13 CPU tests pass and one CUDA-only test is skipped in that process; the separate CUDA smoke passes all four architecture-optimizer cells. Maximum CUDA function error is 5.9605×10^{-7} and maximum analytical/live first-head update error is 8.9407×10^{-8}. The subsequent scale-shape implementation passes 22 CPU tests (four CUDA cases excluded from that CPU run) and all four separate CUDA cases, including bias, head gradients, optimizer state, restart, parameter counts, and same-scale factorization. All selected trajectories receive their full update budget.

H.2 First study: shape matters at a fixed scalar component

This study contains 64 heldout images, 16 from each of Aircraft, STL-10, SVHN, and GTSRB, evaluated under two architectures and two optimizers: 256 optimizer-specific outcomes, not 256 independent images. Validation selects a gradient-eigen rotation at outer scalar 4 for the Fourier MLP and a determinant-normalized target-covariance inverse-square-root chart at scalar 4 for SIREN, for both optimizers. The final policy record was frozen before these heldout outcomes. An independent read-only reconstruction of the returned validation and endpoint tables verifies the choices and reports Table 27. This is a secondary analysis of the prior experiment, not a new prospective replication; those outcomes subsequently count as development information for the separate scale-shape study.

The reconstruction averages domains and architectures equally and keeps Adam/AdamW paired in canonical-image resampling, with 20,000 replicates and seed 2026090404. Architecture means are $+0.122681$ dB for Fourier features and $+0.377206$ dB for SIREN. A single target-covariance chart is not uniformly useful: its pooled gain is $+0.104723$ dB, but the architecture means have opposite signs ($+0.377206$ for SIREN and -0.167761 for Fourier features). A separate family-calibrated selector is -0.005980 dB below its strongest validation-selected comparator (reconstructed interval $[-0.007487, -0.004618]$), despite a $+1.302933$ dB gain over conventional physical coordinates. Thus conventional-head improvement alone does not establish successful operational selection.

Table 27: Dense versus determinant-matched scalar coordinates. Prior validation-selected architecture-conditioned dense policy versus determinant-matched scalar $4I$ at 100 steps. The paired intervals are independently reconstructed from all returned rows.

Metric	Dense policy	Scalar $4I$	Difference [95% CI]
PSNR (dB)	29.784632	29.534689	+0.249944 [0.223557, 0.276834]
SSIM	0.768371	0.758453	+0.009918 [0.009164, 0.010629]
LPIPS	0.483962	0.491681	−0.007719 [−0.008794, −0.006626]

H.3 Separate prospective scale-shape test

Factor every chart as

$$B = sQ, \qquad s = |\det B|^{1/3} > 0, \qquad |\det Q| = 1. \tag{68}$$

Comparing sQ to sI matches determinant magnitude and isolates shape at that scalar component. It does *not* establish superiority to a scalar chart with a different, better-tuned scale. Likewise, neither comparison is the same as testing a selector against the strongest validation-selected fixed rule.

The new prospective set contains 128 images (32 per domain), two architectures, and two optimizers: 512 units at 100 updates. A preselected nested subset of 32 images (8 per domain) provides 128 units at 400 updates. Development reduces the 45-chart lattice to 27 complete candidates; validation selects baselines and one of four already-frozen selector forms separately at each horizon. These forms are joint one-step selection, hierarchical scale-then-shape selection, setting-shortlist lookahead, and setting-conditioned ridge utility. The chosen form at both horizons is the setting-shortlist rule; its selection does not imply positive validation performance. No validation refitting or heldout-based removal is permitted. The frozen estimand is equal-domain/equal-architecture canonical-image-clustered, pairing Adam/AdamW; 20,000 bootstrap replicates use base seed 2026090406 with fixed horizon-, comparison-, and metric-specific offsets recorded in the analysis tables.

At 100 steps, the selected dense charts are gradient-covariance inverse-square-root at scale $0.5/0.25$ for Fourier Adam/AdamW, and target-covariance inverse-square-root at scale 16 for both SIREN settings. At 400, Fourier uses a positive low-rank chart at scale 0.5 and SIREN a gradient-eigen chart at scale 16. Strongest separately validation-selected scalar scales are $0.25/0.5$ for Fourier Adam/AdamW and 32 for SIREN at 100; at 400 they are 0.25 for Fourier and 16 for SIREN. Hence the 400-step aggregate is neither a full-set 400-step test nor the continuation of one identical 100-step policy. The persistence comparison uses the frozen nested population and horizon-specific rules.

Table 28: Separate prospective scale-shape study. Dense means the validation-selected sQ, compared with its exact same-scale sI. Selector means the independently frozen scale-shape decision rule, compared with the strongest validation-selected fixed comparator.

Comparison	Metric	100 updates (512 units)	400 updates (128 units)
Dense − same-scale scalar	PSNR	+0.1194 [0.0835, 0.1528]	−0.0038 [−0.1576, 0.1509]
	SSIM	−0.00193 [−0.00345, −0.00049]	+0.00042 [−0.00039, 0.00164]
	LPIPS	−0.00903 [−0.01053, −0.00752]	+0.00101 [−0.00044, 0.00267]
Selector − strongest fixed	PSNR	+0.0175 [−0.0323, 0.0727]	−0.2549 [−0.4777, −0.0309]

At 100 steps, dense charts also improve on the validation-selected output-head learning-rate baseline by $+0.118651$ dB ($[0.082014, 0.152964]$). Yet SSIM is lower than the same-scale scalar control, and the 400-step PSNR advantage is not established. The selector improves 100-step SSIM by

$+0.010125$ and LPIPS by -0.003508 relative to its primary comparator, but its PSNR interval spans zero; at 400 it loses 0.254852 dB and worsens LPIPS by 0.003182. Dense selections occur on 78.52% and 64.84% of units, respectively; frequent selection is not evidence of successful selection.

Cost and supported conclusion. Balanced timing uses one cold and five warm repetitions per architecture-optimizer setting. The full 100-update selector has mean warm overhead 20.678% versus the primary fixed comparator, 22.498% versus the strongest scalar chart, and 22.448% versus the strongest output-head learning-rate control. These measurements are not quality-matched speedups. Taken together, the studies establish exact non-Gaussian mechanism transfer and a short-horizon shape effect at matched scale, but not persistent dense superiority, uniformly nondegrading perceptual quality, or a selector that improves on its strongest fixed comparator.

H.4 Direct comparison with the best validation-selected scalar

The saved scale-shape rows also permit a direct comparison of the frozen dense rule with the strongest separately validation-selected scalar at each architecture, optimizer and horizon. Table 29 uses those existing choices and endpoints, without selecting a new policy. Means average the four paired architecture–optimizer units within image and domains equally. Descriptive intervals resample canonical images within each domain with 20,000 draws, keeping all paired units together. This is a retrospective contrast of an existing prospective study, not a new prospective confirmation.

Table 29: Dense minus the strongest validation-selected scalar chart. Lower LPIPS is favorable; higher PSNR and SSIM are favorable. Each entry gives a mean paired difference and descriptive 95% interval. The 400-update cohort is nested and its policies were selected separately.

Metric	100 updates, 128 images	400 updates, 32 nested images
PSNR (dB)	-0.0027 $[-0.0614, +0.0502]$	-0.0508 $[-0.2321, +0.1014]$
SSIM	-0.00980 $[-0.01206, -0.00774]$	$+0.00088$ $[+0.00004, +0.00222]$
LPIPS	$+0.00343$ $[+0.00187, +0.00503]$	$+0.00007$ $[-0.00102, +0.00120]$

The dense rule does not establish a PSNR advantage over the best allowed scalar at either horizon. Its positive same-scale effect at 100 updates is therefore a shape-isolation finding, not a best-scalar performance claim. At 100 updates the direct comparison also worsens mean SSIM and LPIPS. These contrasts complement, rather than replace, the original same-scale and selector comparisons.

I Joint-Block Controls and the Cost of Candidate Information

Separate from the two unchanged SGCL campaigns (Appendix G), this study tests the cost of candidate-loss features and competitiveness under independent geometry/appearance tuning.

I.1 Roles, implementations, and controlled ablation

The confirmation contains 128 new source/content-disjoint images, 32 each from FGVC Aircraft, GTSRB, STL-10, and SVHN: fresh images, not new domains. Adam/AdamW pairs and two extra repetitions of the first eight images per domain yield 384 optimizer/repetition observations per arm, 1,920 across five arms. Repeats are averaged within image without increasing sample size. The unchanged 256×256 preprocessing and 2,000-Gaussian initialization do not make resized low-resolution sources native high-resolution evidence.

The full and no-candidate-loss refits use the original 48-image development ledger, identical eight-candidate access, whole-domain folds, domain weighting, 100-step utility, and equal hyperparameter-search budgets. The reduced predictor removes the hypothetical-loss feature and its eight candidate renders, but retains the physical gradient and two isolated block-response renders. It is not a zero-forward method. Since every library member is scaled orthogonal in exact arithmetic, both refits set the log condition number and normalized orthogonality defect to zero rather than amplify floating-point residuals during feature standardization. Their nominal ten/nine

feature masks thus have eight/seven nonconstant values. The original SGCL coefficients remain untouched. Refitted ridge penalties are 100 (full) and 10 (reduced); fallback margins are 0.04794345 and 0.01988081 dB, respectively. Appendix S supplies their coefficients and decision conventions.

The broader simple family includes independent geometry/appearance learning rates, scalar refinements, a rendered-balance rule, and one-switch cooldown schedules. Settings are selected separately by domain and optimizer on the existing 16-image-per-domain validation roles; the fresh cohort is not used for selection. A shared online budget is frozen before timing the top three validation settings near 100 updates. The final simple comparator is the primary-time winner of that shortlist, not a claim of global optimality or necessarily the 100-update-only winner. It retains access to all nine candidates available to the joint-block predictor.

For a fixed scalar chart $U = X(sI)^\top$, fresh-state native AdamW can realize chart-space settings $(\eta, \epsilon, \lambda)$ through

$$\alpha_U = \eta s, \qquad \epsilon_U = \epsilon/s, \qquad \lambda_U = \lambda/s.$$

Block multipliers and schedules are applied consistently. The native learning-rate-only control instead keeps physical ϵ and decay: it is a distinct, deliberately strong optimizer policy, not exact transport. Thus all arms retain the reconstruction objective, but their effective AdamW decay need not match. Ordinary Adam has zero decay.

Table 30: Frozen simple controls. All use scalar 8, geometry multiplier 4 and appearance multiplier 2 relative to the base policy. "Transported" uses the scalar identities above; "LR-only" retains native epsilon/decay. The base rate is 0.016. Budgets are set on validation, not fresh outcomes.

Domain	Optimizer	Implementation / schedule	Budget (ms)
FGVC Aircraft	Adam	Transported	128.730
FGVC Aircraft	AdamW	LR-only	128.347
GTSRB	Adam	Transported	106.665
GTSRB	AdamW	LR-only	109.306
STL-10	Adam	Transported; appearance $\times 0.5$ at step 50	103.364
STL-10	AdamW	LR-only	110.731
SVHN	Adam	LR-only	107.401
SVHN	AdamW	LR-only	108.382

The joint-block predictor is a separate bounded extension, not a coupling intervention. Its scalar-8 candidate rate pairs are $(0.5, 0.5)$, $(1, 1)$, $(2, 1)$, $(2, 2)$, $(4, 2)$, $(4, 4)$, $(8, 4)$, $(8, 8)$ in transported coordinates, plus the $(4, 2)$ LR-only arm. It uses one physical gradient and two block-response renders, rather than nine candidate-loss probes. A ridge model combines log rates, log requested horizon, optimizer identity, normalized geometry/appearance response magnitudes, and the gradient-norm ratio. It is fit on development horizons 30, 100, and 400; 200 is reserved. A fitted margin selects a constant fallback, not a confidence certificate. The requested horizon can change the initial decision; a 200-step request is not the continuation of a 100-step decision. No new cross-block chart, neural representation, or native 3D experiment is part of this study.

I.2 Frozen inference and primary contrasts

Means average optimizer/repetition observations within each image, then weight the four domain means equally. The bootstrap resamples image clusters within domain with 10,000 draws (seed 20260909). Three prespecified primary PSNR contrasts use Bonferroni-adjusted 98.33% intervals. The 0.02 dB preservation margin is an engineering tolerance, below one quarter of the earlier approximately 0.08 dB gain; it is not a perceptual threshold. Secondary 95% intervals are descriptive, not additional adjusted primary successes.

The reduced predictor's upper endpoint is below -0.02: preservation is not merely unresolved. The joint predictor's interval crosses zero, establishing neither superiority nor universal inferiority. First-repetition-only means (-0.05785, -4.36734, -0.21738 dB) change no decision. Table 33 gives absolute PSNR and equal-time perceptual metrics; additional endpoints/resources follow.

Removing candidate-loss renders saves 5.518 ms at 100 updates (95% CI $[4.332, 6.754]$ ms). At the common online budget, reduced minus full refit is $+0.331665$ dB (95% CI $[0.254659, 0.409639]$),

Table 31: All prespecified primary contrasts. Positive favors the left arm. The first requires a lower endpoint above -0.02 dB; the other two require it above zero. None meets its criterion. There are 128 image clusters and no unavailable primary states.

Contrast	Mean PSNR	98.33% interval	Wins/losses
No-loss − full, 100 updates	-0.05814	$[-0.09761, -0.03719]$	24/104
No-loss − simple, equal time	-4.36718	$[-4.84884, -3.82246]$	7/121
Joint predictor − simple, equal time	-0.23300	$[-0.51690, +0.08474]$	52/76

Table 32: Perceptual endpoints and resources. SSIM/LPIPS are at 100 updates. Setup is part of online time, not added to it again. Allocated/reserved peaks are process-allocator measurements; these are not quality-matched memory-efficiency claims.

Policy	SSIM	LPIPS	Setup (ms)	Alloc. (MiB)	Res. (MiB)
Original SGCL	0.52037	0.53640	33.291	19.645	32.000
Full refit	0.52063	0.53620	33.447	19.645	32.000
No-candidate-loss refit	0.51880	0.53713	27.993	18.830	30.000
Joint-block predictor	0.68183	0.45579	25.329	18.646	30.000
Simple control	0.69709	0.44481	23.877	18.627	29.948

Table 33: Joint-block controls across update and time budgets. 128 fresh images in four studied domains. Full refit is matched; no-loss refit removes candidate-loss features. Joint rows use broader blockwise choices. The 400-update column continues the short-budget policies without retuning. Equal-time budgets (103.4–128.7 ms) and controls are calibrated per domain/optimizer before fresh outcomes. Time includes initialization and probes, excluding startup, warmup, prior preload and final metrics. PSNR is in dB; see Appendix I.

	100 updates		400 updates	Equal online time		
Policy	PSNR	Time (ms)	PSNR	PSNR	SSIM	LPIPS
SGCL	25.947	132.29	31.941	24.690	0.4739	0.5605
Full refit	25.953	132.17	31.904	24.705	0.4745	0.5602
No-loss refit	25.895	126.65	32.047	25.037	0.4867	0.5542
Joint predictor	29.330	114.28	29.730	29.171	0.6759	0.4593
Joint control	**29.503**	**113.02**	27.144	**29.404**	**0.6935**	**0.4475**

$+0.012176$ SSIM, and -0.005953 LPIPS. This supports a measured cost-quality option relative to the full selector, not superiority over the broader simple family. The original/full/reduced policies all use the narrower appearance-chart pool.

I.3 Clock, numerical limits, and duration

The clock includes target reading/hashing/decoding/preprocessing, model initialization/transfer, gradient/probe acquisition, chart installation, updates, and synchronized state recording. It excludes shared immutable prior preload/offline fitting, process/CUDA startup, common synthetic warmup, and final metrics/export. These subsecond budgets are not end-to-end service latency or interchangeable with older timings. Over-budget states receive no credit. Offline LPIPS evaluates 29,521 saved float-RGB states, not quantized images; cold-process LPIPS was not measured.

Scalar first-step and dedicated render/gradient checks pass, and reported initial PSNR is identical across all five arms in every paired block. An optional scalar fast implementation was not used for selector confirmation because its long-trajectory tolerance was not met, including some unchanged-implementation pairs. The main comparisons therefore include native-optimizer versus hidden-chart-wrapper cost; they do not attribute all overhead to probes alone. The 32-image repeat subset has median/maximum within-policy PSNR ranges of $0.000151/0.621364$ dB at 100 updates and

$0.313344/20.597152$ dB at 400. All runs are retained. Late variability is not uniquely attributed to CUDA or interpreted as a new mechanism.

Against the same near-100-step-selected simple control, original SGCL changes from -3.556302 dB at 100 (descriptive 95% CI $[-3.957989, -3.106981]$, 9/128 image wins) to $+4.797011$ dB at 400 ($[4.098421, 5.546579]$, 125/128 wins). The joint predictor changes from -0.173051 to $+2.585899$ dB at 100/400. None of these continuations establishes superiority to a 400-step-selected control. On prior validation, substituting a 400-step-selected control reverses the joint predictor's apparent 400-step gain from $+3.273576$ to -2.956010 dB. On 32 nested fresh images, the separately requested 200-step predictor is -0.130905 dB (95% CI $[-0.417385, 0.175846]$) against a 200-step-selected simple control.

A nested 16-image, two-domain check at 800 Gaussians gives -0.273748 dB (95% CI $[-1.286126, 0.939736]$) for the unchanged joint predictor against the capacity-selected simple control, and -0.425579 dB against the unchanged primary control. Neither establishes superiority. Capacity also changes the common learned basis; cross-capacity absolute differences are not isolated chart effects. These nested tests add no independent confirmation images.

Time-to-target uses validation-fixed thresholds and only observed completed states, retaining right censoring without interpolation. In the primary window, neither full nor reduced refit attains its target in any of their 384 paired optimizer/repeat observations. Over the longer reporting window, both attain in 297, only reduced in 21, only full in 7, and neither in 59. Equal censoring is not equal efficiency. The evidence establishes a bounded within-family tradeoff and a stronger simple comparator, not a generally superior learned controller.

J Energy Coordinates: Mechanism, Confirmation, and Numerical Attribution

Study scope. This appendix preserves the first fixed-energy ordinary-Rprop confirmation, its original cohort and all calibrated-control tradeoffs. The later symmetric iRprop+ calibration uses a separate 64-image cohort and is reported in Appendix R; it does not replace these rows. Appendix J.10 tests the repair on a targeted panel; Appendix T then replays the entire symmetric-calibration cohort under original and repaired arithmetic. Neither replaces the earlier ordinary-Rprop cohort here.

J.1 Self-Gram identity and exact chart domain

Let $z = (r_x, r_y, u, v, w)$ denote the peak chart's five geometry coordinates, and let $C \in \mathbb{R}^3$ be one atom's physical color. For a whole-plane untruncated Gaussian with fixed opacity and fixed nonzero channel scales, write its self-Gram matrix as

$$G = \begin{bmatrix} A & B \\ B^\top & D \end{bmatrix}, \qquad D = kac\,\mathrm{diag}(s_1^2, s_2^2, s_3^2), \quad k > 0.$$

Here k includes the fixed opacity and common loss normalization; G is not the full overlapping-image Gram matrix or the residual-weighted Hessian. With $\ell = \nabla_z \log(ac) = (0, 0, 1/a_0, 0, 1/c_0)^\top$, differentiation of the atom's squared norm gives

$$D^{-1}B^\top = \tfrac{1}{2}C\ell^\top.$$

Set $h(z) = \sqrt{a_0 c_0/(ac)}$ and $C = hQ$. Its differential is

$$T = \begin{bmatrix} I & 0 \\ -C\ell^\top/2 & hI \end{bmatrix}, \qquad T^\top G T = \begin{bmatrix} A - BD^{-1}B^\top & 0 \\ 0 & h^2 D \end{bmatrix}.$$

The geometry derivative in these coordinates is $g_z - BD^{-1}g_C$, and the color derivative is hg_C. The cancellation is exact under these self-atom assumptions. It does not invert the remaining geometry block, remove interactions between atoms, or account for finite pixels, clipping, support boundaries or covariance singularities. Vanishing colors can make geometry directions degenerate. Here $a = a_0 e^{u/a_0}$ and $c = c_0 e^{w/c_0}$, so $\partial_u \log(ac) = 1/a_0$ and $\partial_w \log(ac) = 1/c_0$. The initial shear b_0 is unrestricted; only the two diagonal anchors must be positive. Chart bijectivity applies on that positive-diagonal chart, not as an identification of every possible signed-factor parameterization.

J.2 Relation to metric methods and variable projection

Natural Rprop already combines inverse-metric directions and resilient updates, including trace loading in its published construction (Igel et al., 2005). Our frozen candidate instead passes its ordinary chart gradient to ordinary Rprop. Later per-atom natural-metric development controls are not candidate components. Nor is the chart variable projection: for an invertible color map $C = D(z)Q$, the set of attainable linear-color fits at fixed geometry is unchanged. Exact minimization over color gives the same reduced objective under fixed-rank differentiability assumptions. At an unconstrained color optimum $g_C = 0$, the extra shape-color chain term vanishes. Finite Rprop updates do not perform that solve. Rank changes, coordinate-dependent regularization and coordinate-dependent minimum-norm choices need separate treatment; see classical variable-projection analysis and Gaussian-wavepacket applications (Chen et al., 2025; Schrader et al., 2024).

J.3 Matched implementation

Peak and energy use the same positive anchored shape coordinates, initial physical reconstruction and fresh optimizer state. The center and shape initial widths are 0.032×4, and color width is 0.128×4; Rprop multipliers are $(0.5, 1.2)$ with bounds $(10^{-6}, 50)$. The only chart difference is the amplitude exponent, zero versus $1/2$. The renderer wrapper uses independent at-most-255-atom chunks plus a zero-opacity sentinel, with the same background-compatibility restriction and algebraic projection-convention correction in both arms. The correction is applied before, not after, combining the shape and amplitude chains. These original measurements retain float32 internal projection-backward products; the convention conversion itself is evaluated in float64. The shared numeric safeguard rejects a whole atom's inadmissible proposed state, restores its previous raw coordinates, clears its previous-gradient history, contracts its preceding widths and retains attempted-step counters. It performs no retry. This is a declared guarded optimizer policy, not unmodified fail-stop Rprop at the numeric boundary.

J.4 Frozen confirmation and costs

The 64 reserved source identities comprise 16 images each from FGVC Aircraft, GTSRB, STL10 and SVHN. Existing domain priors use 500 training images, 128 channelwise PCA components and 2,000 Gaussians. Targets use the frozen RGB bicubic center-fit to 256×256, not a native-pixel protocol. All 1,024 attempts and 3,072 requested-budget records are retained: 3,066 records have delivered status and six retain failed-arm status, including any available pre-failure measurements. The primary estimand averages paired repeats and the 2/5/10-second budgets within each image, then images within each fixed domain, then the domains. The bootstrap uses 10,000 image-cluster draws and frozen seed 62017. Its interval conditions on calibration, domains, priors and observed executions; it does not include fresh-calibration or domain-population uncertainty. Secondary intervals are descriptive, not multiplicity-adjusted primary discoveries.

The primary per-domain PSNR gains are -0.3511, $+3.3100$, -0.3108 and -0.2844 dB, in the domain order above. The median image effect is -0.1039 dB. Primary gains at 2/5/10 seconds are respectively $+0.7615$, $+0.7185$ and $+0.2928$ dB. Positive mean PSNR and adverse arithmetic mean MSE are compatible because PSNR averages log errors. The geometric clamped-MSE ratio is 0.8728, whereas the ratio of arithmetic mean clamped errors is 1.0397; neither is an image win rate.

The 111 calibration challengers comprise 108 classical peak systems and three native anchors. All 48 frozen policy choices are iRprop+; four reach the appearance-factor-16 search boundary, including primary GTSRB at 10 seconds. This finite menu is not every possible optimizer. Calibration used 32 prior-training identities with withheld calibration outcomes; the candidate had already been developed in all four domains. Other-domain control calibration does not establish unseen-domain transfer of the candidate.

The successful calibration root charges 71,668.81 online seconds for 7,168 attempts, including four failures; confirmation charges 10,235.25 seconds for 1,024 attempts, including two failures. One trajectory serves all three budgets, so these costs must not be tripled. Historical shared-prior fitting costs 929.05 seconds; 12 inference assets total 391,581,942 bytes. Earlier research and repairs are separate, not free or included in these two online totals. Warm charged work includes preparation, acquisition, updates, rejected work and synchronized float32 YCbCr CPU-output delivery. Process

startup, resident-prior preload, specified housekeeping, archival file I/O and offline metrics lie outside that boundary. GPU isolation passed, but host timing was qualified; no clean-host speed or memory-saving claim follows.

J.5 Transfer and comparator boundaries

The primary comparison has worse mean LPIPS ($+0.00821$), despite positive mean PSNR. The matched comparison improves LPIPS on 56/64 image means and SSIM on 61/64, not on every image. A separate, already-observed 512-pixel development configuration loses about 0.75 dB against matched peak under both tested ordinary-pullback realizations, on all four images. Because images, priors and coordinate units also differ, this is not a resolution-only causal result. No new candidate, rate grid or router is selected from confirmation. The $+8.738$ dB difference against original SGCL also changes optimizer and numerical path; it is not an isolated coupled-chart benefit. Native 3D transfer and a practical lifetime predictor remain unestablished.

J.6 Historical numerical attribution

For a nonsingular triangular factor (a, b, c) and the native returned projection covector (n_a, n_b, n_c), the ordinary branch correction is

$$o_b = \tfrac{1}{2}(n_b + (b/c)n_c), \qquad o_a = n_a - (b/a)o_b + (b^2/(ac))n_c, \qquad o_c = n_c.$$

This corrects the doubled packed off-diagonal covariance contribution. It is not a scalar learning-rate change. Ordinary-gradient descent and Hessian/GGN interpretations require an ordinary derivative on a smooth active set. By contrast, fixed-chart optimizer conjugacy requires only that the supplied covectors and optimizer states are consistently transported; this algebra does not certify loss-gradient accuracy.

A bounded diagnostic preserved the original six-domain shared-Adam configuration, cached priors and preprocessing on one hash-selected existing validation image per domain. It crossed STD/OA, native/corrected projection derivatives and two repeats, without changing the forward renderer. Four synthetic finite-difference checks and 12 initial field/captured-covector checks passed. At 100 updates, all six images are available: the mean OA advantage is 7.7342 versus 7.7362 dB in the original YCbCr metric, with interaction $+0.00204$ dB. Common RGB gives 6.2675 versus 6.2713 dB. These are diagnostic prefix results, not new heldout estimates or a replacement endpoint.

Only 41/48 arms reach 400 updates under the predeclared conservative pre-render refusal. Seven STD arms in Caltech101, DTD and EuroSAT approach float32 covariance cancellation. They are unavailable, not losses imputed to STD; no repairs or retries were allowed. The six-domain 400-step factorial therefore remains unresolved. A conservative refusal does not prove the historical CUDA execution crashed. Maximum observed tile occupancy is 48 across 17,899 trajectory projection checks, below the conditional 256-atom first-batch defect threshold. This rules out that defect only on inspected valid states, not every historical trajectory. The historical E2A record lacks an original binary hash, so this is not certified bitwise reproduction of the earlier experiment.

J.7 Complete principal contrasts

The candidate wins on 28/64 primary image means, versus 58/64 under the matched-coordinate comparison. The two native joint-policy failures make the corresponding complete contrast unavailable; their earlier outputs are retained but cannot replace the failed full attempts in a headline complete comparison. The shared rejection policy's absence of an unsafe CUDA call is a safety property, not proof that an unguarded historical implementation would have crashed.

No retrospective method selection. Neither the energy exponent, optimizer family, guard, nor starting widths were changed using confirmation outcomes. The reference control is chosen separately for each declared calibration regime and output budget. The favorable matched comparison isolates the nonlinear amplitude coordinate within its shared numerical implementation. The primary independently calibrated comparison instead answers a practical finite-menu question, with the selection uncertainty and adverse perceptual/domain effects stated above. These two questions must not be interchanged.

Table 34: First energy confirmation: matched effect and calibrated-control scope. Energy minus control on 64 fresh images. Repeats and 2/5/10-second budgets are averaged within image; domains are equally weighted. PSNR is in dB; higher PSNR/SSIM and lower LPIPS are favorable. Conditional PSNR 95% intervals are $[0.3268, 0.8746]$ (primary) and $[1.6255, 2.0699]$ (secondary matched contrast). These are attained warm CPU-output budgets with qualified host timing. The matched effect does not replace the calibrated primary test.

Control	ΔPSNR	ΔSSIM	ΔLPIPS	Median PSNR	PSNR wins
Calibrated (primary)	$+0.5909$	-0.00048	$+0.00821$	-0.1039	28/64
Matched peak	$+1.8482$	$+0.00742$	-0.05135	$+1.1002$	58/64
	Calibrated-control ΔPSNR by output budget			Budget-mean ΔPSNR	
Domain	2 s	5 s	10 s	Calibrated	Matched
Aircraft	-0.321	-0.380	-0.352	-0.351	$+0.241$
GTSRB	$+3.850$	$+3.896$	$+2.184$	$+3.310$	$+2.840$
STL10	-0.223	-0.339	-0.371	-0.311	$+0.393$
SVHN	-0.260	-0.303	-0.290	-0.284	$+3.918$

Table 35: Energy-coordinate paired intervals. Energy minus comparator, averaged first within each image across repeats and attained budgets. Positive PSNR/SSIM and negative LPIPS are favorable. The primary interval conditions on the frozen calibration choices and four domains; the matched comparison is secondary.

Comparison	ΔPSNR (dB)	ΔSSIM	ΔLPIPS
Calibrated controls, primary	$+0.5909$	-0.00048	$+0.00821$
Conditional 95% interval	$[0.3268, 0.8746]$	$[-0.00181, 0.00092]$	$[0.00149, 0.01606]$
Matched peak coordinates	$+1.8482$	$+0.00742$	-0.05135
Descriptive 95% interval	$[1.6255, 2.0699]$	$[0.00570, 0.00940]$	$[-0.06099, -0.04188]$

J.8 Complete calibration menu and selected policies

Let (c_f, s_f, a_f) multiply the inherited initial Rprop widths for centers, shape and appearance, respectively: $(0.032c_f, 0.032s_f, 0.128a_f)$. Each of the absolute and relative peak-coordinate families contains the same 26 triples:

$$\mathcal{T}_0 = \{(k, k, a_f) : k \in \{0.25, 1, 4, 16, 64\},\ a_f \in \{0.25, 1, 4, 16\}\},$$
$$\mathcal{T}_1 = \{(0.25, 1, 1), (0.25, 1, 4), (0.25, 4, 1), (0.25, 4, 4), (4, 1, 1), (4, 1, 4)\}.$$

The full set is $\mathcal{T}_0 \cup \mathcal{T}_1$. The absolute shape chart is $(a, b, c) = (a_0 e^{u/a_0}, b_0 + v, c_0 e^{w/c_0})$; the relative chart is $(a_0 e^u, b_0 + \sqrt{a_0 c_0}\, v, c_0 e^w)$. Each optimizer additionally has two signed physical-coordinate controls with triples $(4, 4, 4)$ and $(16, 16, 1)$. Their equivalent positive-factor view preserves the represented covariance and differentiates back to the signed coordinates. Thus ordinary Rprop and loss-aware iRprop+ each contribute $26 + 26 + 2 = 54$ controls, or 108 in total. Original SGCL, its no-candidate-loss refit, and the frozen joint-block Adam policy provide the remaining three anchors. The energy candidate is separate from this 111-challenger pool. All Rprop-family settings use multipliers $(0.5, 1.2)$, width limits $[10^{-6}, 50]$, and zero decay.

For each domain and output budget, four calibration choices are frozen: maximize mean RGB PSNR or minimize mean raw decoded YCbCr fitting loss, using either that domain's eight calibration images or the other domains' 24 images. Both repeats enter each image's mean equally. A setting is eligible only when every required deployment is complete and its requested output is delivered; failures are not replaced by valid earlier prefixes. Exact ties use the frozen identifier order. This yields $4 \times 3 \times 2 \times 2 = 48$ selections, all iRprop+. Table 37 enumerates them. Reconstruction from all 21,504 calibration rows reproduces all 48 settings and eligibility counts; means agree within 10^{-10}. Four choices reach appearance factor 16. The finite-menu conclusion does not claim optimization over untested rates.

Table 36: Energy-coordinate effects by domain and budget. The positive overall primary PSNR result is heterogeneous across the predeclared domains and budgets. Domain rows average all budgets; budget rows average all domains. They are not independent experiments.

Primary contrast stratum	Mean ΔPSNR (dB)	Interpretation
FGVC Aircraft	-0.3511	16 image means
GTSRB	$+3.3100$	16 image means
STL10	-0.3108	16 image means
SVHN	-0.2844	16 image means
2-second output	$+0.7615$	64 image means
5-second output	$+0.7185$	64 image means
10-second output	$+0.2928$	64 image means

Table 37: All 48 energy-comparator calibration choices. A and R denote absolute and relative peak coordinates; each triple is (center, shape, appearance) initial-width multipliers. All cells use iRprop+. The primary comparison uses the same-domain PSNR column. Selection objectives are not chosen using confirmation outcomes.

		Same-domain calibration		Other-domain calibration	
Domain	Seconds	PSNR	Fitting loss	PSNR	Fitting loss
Aircraft	2	R(4,4,4)	R(4,4,4)	R(4,4,4)	R(4,4,4)
	5	R(4,4,4)	R(4,4,4)	R(4,4,4)	R(4,4,4)
	10	R(4,4,4)	R(4,4,4)	R(4,4,4)	R(4,4,4)
GTSRB	2	A(16,16,4)	R(4,4,16)	R(4,4,4)	R(4,4,4)
	5	A(16,16,4)	R(4,4,16)	R(4,4,4)	R(4,4,4)
	10	R(4,4,16)	R(4,4,16)	R(4,4,4)	R(4,4,4)
STL10	2	R(4,4,4)	R(4,4,4)	A(16,16,4)	R(4,4,4)
	5	R(4,4,4)	R(4,4,4)	R(4,4,4)	R(4,4,4)
	10	R(4,4,4)	R(4,4,4)	R(4,4,4)	R(4,4,4)
SVHN	2	A(16,16,1)	A(16,16,1)	R(4,4,4)	R(4,4,4)
	5	A(16,16,1)	A(16,16,1)	R(4,4,4)	R(4,4,4)
	10	A(16,16,1)	A(16,16,1)	R(4,4,4)	R(4,4,4)

The secondary other-domain-PSNR comparison gives $+1.1247$ dB (conditional interval $[0.9312, 1.3164]$), while same-domain and other-domain raw-loss selection give $+0.4917$ and $+0.7042$ dB. These are the original frozen secondary comparisons, not alternative primary findings. Other-domain calibration changes the challenger's information, not the candidate's prior exposure to the four developed domains.

J.9 Image-role ledger and numerical safeguards

Each of the four priors uses 500 training identities. The 32-image development cohort and a separate 32-image calibration cohort each contain eight official training images per domain and overlap the corresponding prior-training set. The calibration identities were not previously evaluated in this campaign; their fitting outcomes were withheld until the candidate, menu and analysis were frozen. Thus calibration is separate from candidate development outcomes, but not independent of prior fitting. The 64 confirmation identities are 16 official-test images per domain, with no declared prior-training or previous-evaluation overlap after identity and decoded-content checks. Role checks use canonical identities and source/prepared-content hashes. Confirmation uses new images from the previously studied domains.

For each proposed positive triangular factor (a, b, c), the range check uses float32 operations in the projector's order:

$$x = a^2, \qquad y = ab, \qquad z = b^2 + c^2, \qquad d = xz - y^2,$$
$$t = (x + z)/2, \qquad r = \left\lceil 3\sqrt{t + \sqrt{\max(t^2 - d, 0.1)}} \right\rceil .$$

Raw parameters, decoded colors, the covariance entries and determinant must be finite, with $a, c > 0$. For $d \neq 0$, the reciprocal, conic $(z/d, -y/d, x/d)$, trace arithmetic and radius must also be finite, with $0 < r \leq 2147483520$, the largest float32 value below 2^{31}. Exactly zero determinant is the renderer's inactive/culling case; its reciprocal and radius do not define admissibility. Signed-factor controls use the equivalent positive-factor audit. Source states and gradients must be admissible and finite before a step.

An inadmissible proposal restores that atom's entire preceding raw center, shape and color state, clears its stored previous gradients, and replaces its widths by half the *preceding* widths clipped to $[10^{-6}, 50]$. Attempted-step counters are retained; no retry occurs. The iRprop+ realization also clears that atom's stored previous update. Native Adam-based anchors instead use the declared pre-render fail-stop contract; a failed deployment stays unavailable. These policies must not be pooled as one identical optimizer-boundary intervention.

Table 38: Confirmation deployments, guard exposure and delivered updates. Exposed trials have at least one rejected atom during the full 10-second deployment. Rejection totals count atom-events, not distinct atoms or rejected iterations. Width contractions are not extra completed updates. Native anchors have no atom-rejection policy, hence n/a. Joint-block Adam's update means use its 126 complete deployments; its two failures remain unavailable in complete-population quality contrasts. Failures remain in the denominators and are not converted to complete quality observations.

System	Complete	Exposed trials	Atom rejections	Mean delivered updates		
				2 s	5 s	10 s
Energy / Rprop A(4,4,4)	128/128	6	16	288.0	720.2	1441.8
Peak / iRprop+ A(16,16,1)	32/32	2	3	291.0	727.5	1453.8
Peak / iRprop+ A(16,16,4)	64/64	2	1006	287.6	723.3	1449.2
Peak / iRprop+ R(4,4,16)	32/32	6	3839	290.5	725.9	1451.8
Peak / iRprop+ R(4,4,4)	128/128	17	2625	286.2	720.8	1446.5
Joint-block Adam	126/128	n/a	n/a	825.1	2075.9	4157.7
No-candidate-loss refit	128/128	n/a	n/a	809.2	2038.4	4084.5
Original SGCL	128/128	n/a	n/a	799.5	2022.7	4062.5
Peak / Rprop A(4,4,4)	128/128	13	1040	290.6	729.3	1459.8
Peak / Rprop R(4,4,4)	128/128	14	5075	290.2	730.3	1463.4

Descriptive no-rejection subgroup. As a post-hoc check, require both repeats of the candidate and every relevant comparator to have zero atom rejections throughout their complete deployments. This retains 55/64 images for the matched contrast and 49/64 for the primary contrast. Equal-domain means within these restricted groups are respectively $+1.8601$ and $+0.6061$ dB; all four matched domain means remain positive. The subgroup is selected by treatment-dependent outcomes, not randomized. It neither estimates a guard-free counterfactual nor attributes the total benefit to decorrelation alone. The full 64-image comparisons remain primary; all excluded images remain in their original tables. These are retrospective record summaries, with no new trajectories or claim of independent replication.

J.10 Isolated backward-precision sensitivity

Intervention and numerical contract. The algebraic convention correction above and the precision of the calculation it converts are separate issues. A targeted audit identified cancellation error in the native projection-backward products, including ordinary-gradient sign disagreements at captured inputs. B0 retains the original internal float32 arithmetic. B1 replaces only that backward

component with double-precision products and returns the same float32 interface. Both use the existing float64 algebraic conversion exactly once, before the chart chain rule. Forward projection, rasterization, support, opacity, affine decoding, optimizer settings and guards are unchanged. B0 identifies the retained arithmetic in this assay, not bitwise replay of every earlier experiment.

Specifically, for the returned symmetric conic K and symmetric covector H (whose off-diagonal entry follows the native half-packed convention), the covariance pullback is $V = -KHK$. For factor (a, b, c), its exact symmetric form gives the native return

$$n = (2aV_{00} + 4bV_{01},\ 4aV_{01} + 2bV_{11},\ 2cV_{11}).$$

B1 computes KH, $-(KH)K$ and the factor products in float64, retaining $V_{01} + V_{10}$ for the packed off-diagonal calculation before the float32 return. The build disables fused multiply-add contraction. It introduces no determinant floor, regularization or new support rule. Subsequent conversion to (o_a, o_b, o_c) is unchanged; increasing that conversion's precision alone would not recover already lost internal accuracy.

Component checks. An independently expanded polynomial reference uses 100-digit arithmetic on the captured binary inputs for 120 selected three-component pullbacks. The all-atom screen contains 100 replays of 2,001 atoms (including the sentinel), or 600,300 scalar positions *per output convention*. Its predeclared component envelope uses $\gamma_{64} = 64\,2^{-53}/(1 - 64\,2^{-53})$ times absolute monomial sums, adds float32 return rounding, and propagates these allowances through the absolute coefficients of the ordinary conversion. Float32 unit roundoff is 2^{-24}, with absolute subnormal allowance 2^{-149}; thresholds are not fitted to outcomes. B1 has zero recorded native- or ordinary-envelope violations. The selected ordinary-sign disagreements fall from six pullbacks/twelve scalar components in B0 to zero in B1. These are finite component checks, not a theorem of whole-renderer accuracy. B0/B1 forward, projection and support outputs match exactly at the five frozen states; original forward rounding and support limitations remain.

Fixed-policy fitting and population. The panel contains the five already-exposed official-training images in Table 39, not new confirmation images. It includes two Aircraft cases selected for the numerical investigation and overlaps prior training/calibration. The E, M and P profiles, 256×256 preprocessing, 2,000-Gaussian priors and guarded iRprop+ recurrence are those in Appendix R, without retuning. Each image/policy has two executions per backend from matched saved physical initializations and fresh optimizer history. $P = M$ on GTSRB and SVHN, yielding 52 distinct fixed-update trajectories and 208 checkpoints at 0, 30, 100 and 400 attempted updates. Repeats are averaged within image, then the five images equally. No population interval or equivalence margin is asserted.

At 400 updates, none of the 20 execution/comparator cells changes its native-MSE or RGB-PSNR contrast sign; aliases reduce these to 16 distinct execution/peer cells, not 20 independent images. E beats M on all five image means in native MSE but only four in RGB PSNR. Both Aircraft images remain worse than P in RGB PSNR. Arithmetic-mean native-MSE benefit, defined as peer minus E, is -7.461607×10^{-6} for B0 and -7.627841×10^{-6} for B1 against P: favorable mean log RGB quality does not imply favorable arithmetic fitting error. Individual fixed-400 B1-minus-B0 PSNR changes span $[-0.089245, +0.070778]$ dB. Sign preservation therefore does not mean identical trajectories or statistical equivalence.

Delivered time and metric accounting. A separate 52-trajectory assay delivers all 156 requested 2/5/10-second CPU outputs. The warm clock charges target loading, initialization, chart/optimizer setup, attempted work, guards, synchronization and delivery; resident-prior preload, synthetic warmup and offline metrics are excluded. At ten seconds, three native-MSE comparator cells reverse sign, while none reverses in RGB PSNR; earlier deadlines can have RGB reversals. Attained update counts vary, so these observations establish neither zero repair overhead nor a systematic speedup. Native sensitivity uses float64 MSE recomputed from retained unclamped YCbCr images, not a captured live optimizer scalar. RGB metrics use the existing single-clamp decoder. The 2,416 offline metric entries include the 208 fixed-update, 156 timed and 2,000 pulse checkpoints below, plus 52 timed initial images; these counts do not increase the image sample size.

One-step intervention diagnostic. One saved state per image is tested: energy at 400 updates for Aircraft 1251094, and matched peak at 30 updates for the other four. Each backend has ten

Table 39: Backward-precision sensitivity on five exposed training images. B0 retains the original internal projection-backward arithmetic; B1 uses double-precision internal products and the same float32 interface. Each image first averages two executions. Both Aircraft images are retained, so these descriptive equal-image means are not equal-domain estimates or independent confirmation. Aircraft labels are official image identifiers; other labels are zero-based official-training sample indices. $P = M$ for GTSRB and SVHN; shared trajectories do not create additional independent observations. The full-cohort Table 1 remains unchanged.

	$E - M$ RGB PSNR (dB)		$E - P$ RGB PSNR (dB)	
Image or output budget	B0	B1	B0	B1
Fixed 400 attempted updates: two executions per image/policy				
Aircraft 0875344	-0.044186	-0.018410	-0.595684	-0.607099
Aircraft 1251094	$+0.466945$	$+0.462438$	-0.365925	-0.361860
GTSRB 24271	$+2.554283$	$+2.553219$	$+2.554283$	$+2.553219$
STL10 898	$+0.543345$	$+0.558723$	$+0.671020$	$+0.627469$
SVHN 32181	$+1.794356$	$+1.921897$	$+1.794356$	$+1.921897$
Equal-image mean	$+1.062949$	$+1.095574$	$+0.811610$	$+0.826725$
Image wins	4/5	4/5	3/5	3/5
Attained output budgets: equal-image means				
2 seconds	$+1.358807$	$+1.465463$	$+1.029556$	$+1.118124$
5 seconds	$+0.744206$	$+0.739810$	$+0.528610$	$+0.500047$
10 seconds	$+0.514356$	$+0.522179$	$+0.305028$	$+0.261780$

executions per state. One fresh backward evaluation is shared by four branches: the unchanged update; a gradient-chain pulse toggling only the shape–color chain term for one step; an amplitude pulse toggling that step's amplitude realization and re-encoding its physical color into the original chart while retaining the control's geometry and optimizer history, after checking equal rejection masks; and a same-color re-encoding sham. All branches resume the donor's original chart and optimizer for 100 further updates, with observations at 0, 1, 5, 20 and 100 suffix updates. All 400 branches are available. This fresh-gradient assay is distinct from replaying an archived gradient.

Effects remain mixed. Using control-minus-branch native MSE for *both* donor types, the gradient-chain pulse has positive mean effects at suffix 100 on both Aircraft states and GTSRB, but negative effects on STL10 and SVHN. The amplitude pulse's Aircraft 1251094 mean changes from -2.0885×10^{-8} to $+8.9460 \times 10^{-8}$; its shift of 1.1034×10^{-7} is nearly the sham's 1.1065×10^{-7} shift. This does not isolate a newly positive mechanism. Across amplitude, gradient and sham means, there is one primary strict reversal, two sham strict reversals and one sham transition to zero, not four primary-effect reversals.

Supported conclusion. The localized arithmetic repair passes its declared component checks and retains the tested fixed-update E/M and E/P contrast signs on this targeted panel. It does not explain the full finite-horizon benefit, certify the entire renderer, or establish broad practical equivalence. This five-image study did not retest either confirmation cohort or C. The subsequent complete symmetric-cohort replay, including C, is reported in Appendix T. Original headline values and figures remain labeled B0 measurements; the repaired replay is separate.

K Sharp Fixed-Chart Limits and Constructive Escapes

This appendix proves the complementary fixed-chart results summarized in Section 3, their stationary-optimizer consequences, and the constructive escape. All output inner products are real Euclidean inner products, including when complex notation abbreviates a real map. The appearance coordinate is allowed to depend on geometry, but the geometry coordinate itself is retained. The theorem is therefore not an obstruction to unrestricted joint coordinates. We first give the classical pointwise block calculation, then the regional chart argument, keeping their contributions distinct.

Theorem K.1 (regional minimax). For $f_p(z,c) = (\cos r, \mathrm{sinc}(r)(z_1 + iz_2)) \otimes c$, $r = \|z\| \le R < \pi/2$, $c \in \mathbb{C}^p \setminus \{0\}$, write $C_z(q) = C(z,q)$ and let $\mathcal{C}$ contain smooth real appearance charts $c = C(z,q)$ that retain z, are global fiber diffeomorphisms of $\mathbb{R}^{2p}$, and extend beyond the closed disk. Then

$$\inf_{C\in\mathcal{C}} \sup_{\|z\|\le R, c\neq 0} \rho_C(z, C_z^{-1}(c)) = \begin{cases} \sin R, & p \text{ odd}, \\ \min\{\sin R, 1/\sqrt{2}\}, & p \text{ even}. \end{cases} \tag{69}$$

Identity attains the odd-head limit; the real quaternion construction below attains the even-head cap. Complex-linear charts retain $\sin R$ at every width. The full-space and punctured-space cases are distinguished in the proof. The result concerns one fixed chart before the adverse state and exact-fit target are chosen.

K.1 Canonical correlation and optimal within-block scaling

Let a full-column-rank output Jacobian be partitioned as $J = [J_g, J_c]$, and write

$$G = J^\top J = \begin{bmatrix} A & B \\ B^\top & D \end{bmatrix} \succ 0, \qquad \rho = \|A^{-1/2} B D^{-1/2}\|_2 < 1.$$

This ρ is the largest canonical correlation between the two tangent subspaces. For every invertible block-diagonal $T = \mathrm{diag}(T_g, T_c)$,

$$\inf_T \mathrm{cond}(T^\top G T) = \frac{1+\rho}{1-\rho}. \tag{70}$$

This is an optimal block-scaling identity, not a new general matrix-conditioning principle (Elsner, 1984; Lin & Sinnamon, 2012). A proof is included because unrestricted dense transforms *inside* each block are important to our theorem.

Block whitening produces $\left[\begin{smallmatrix} I & R \\ R^\top & I \end{smallmatrix}\right]$, with $R = A^{-1/2} B D^{-1/2}$ and extreme eigenvalues $1 \pm \rho$, proving attainability. For any other block congruence, canonical correlation is unchanged. Choose an attaining pair of block-supported vectors, normalize them in Euclidean norm, and compress the congruent Gram to their span. The compression is $\left[\begin{smallmatrix} a & b \\ b & d \end{smallmatrix}\right]$, with $|b| = \rho\sqrt{ad}$. Its relative spectral spread satisfies

$$s^2 = \frac{(a-d)^2 + 4b^2}{(a+d)^2} = \rho^2 + (1-\rho^2)\frac{(a-d)^2}{(a+d)^2} \ge \rho^2.$$

Its condition number is $(1+s)/(1-s)$. Rayleigh-quotient bounds place the full matrix's extreme eigenvalues outside those of the compression, proving the lower bound in (70).

For a fixed block-SPD preconditioner P, PG is similar to $P^{1/2} G P^{1/2}$. Optimizing a positive constant rate gives the stationary error-map spectral radius $(\kappa - 1)/(\kappa + 1)$. Therefore

$$\inf_{P \text{ block SPD},\, \eta > 0} \mathrm{spr}(I - \eta P G) = \rho. \tag{71}$$

For the optimally block-whitened quadratic, the worst one-step objective ratio is ρ^2. This last statement is specific to that quadratic update, not a claim about every nonlinear loss or momentum method.

K.2 The smooth full-rank model and its exact angle

For $z = (z_1, z_2)$, $r = \|z\|$, and $0 < R < \pi/2$, define

$$s(z) = \big(\cos r, \mathrm{sinc}(r)(z_1 + iz_2)\big), \qquad f_p(z,c) = s(z) \otimes c, \quad c \in \mathbb{C}^p.$$

Here $\mathrm{sinc}(r) = \sin(r)/r$ has value one at zero. The functions $\cos r$ and $\mathrm{sinc}(r)$ are smooth functions of r^2 at the origin. With $z = r(\cos\phi, \sin\phi)$, the appearance feature frame has Gram I_{2p}, connection $K = i\omega I_p$, and horizontal geometry metric

$$\omega = \sin^2 r \, d\phi, \qquad h = dr^2 + \sin^2 r \cos^2 r \, d\phi^2, \qquad d\omega = 2 \, d\,\mathrm{vol}_h \,. \tag{72}$$

In Cartesian coordinates, the radial and tangential eigenvalues of h are 1 and $[\sin(2r)/(2r)]^2$, respectively. Both are positive for $r \le R < \pi/2$.

For $c = C(z, q) \neq 0$ and geometry velocity v, decompose the geometry output derivative orthogonally as

$$J_g v = N_v + \Phi e_v, \qquad e_v = C_v + i\omega(v)c, \qquad \|N_v\|^2 = \|c\|^2 h(v, v).$$

The appearance derivatives span $\mathrm{col}(\Phi)$ because C_q is invertible. Consequently

$$d_C = \sup_{v \neq 0} \frac{\|e_v\|}{\|c\|\sqrt{h(v, v)}}, \qquad \rho_C = \frac{d_C}{\sqrt{1 + d_C^2}}. \tag{73}$$

This is invariant to arbitrary invertible units within either tangent block. The horizontal metric also proves that the full real joint Jacobian has rank $2 + 2p$ at every $c \neq 0$. For $p = 1$ its determinant magnitude in the identity chart is $\|c\|^2 \sin(2r)/(2r)$, with limit $\|c\|^2$ at zero. The zero-color rank deficiency is excluded rather than used to create a conditioning obstruction.

K.3 One head: a sharp lower for any common nonzero slice

Fix one appearance coordinate q_0 such that $C(z, q_0)$ is nonzero throughout the closed disk. With $p = 1$, its complex value has a smooth global expression $c = ae^{i\chi}$, $a > 0$, because the disk is simply connected. Equation (73) gives

$$\frac{\|e_v\|^2}{|c|^2} = (d \log a(v))^2 + (d\chi(v) + \omega(v))^2.$$

The exact phase has zero circulation around the boundary. Since that circle has h-length $2\pi \sin R \cos R$,

$$2\pi \sin^2 R = \oint_{\partial D_R} (d\chi + \omega) \leq 2\pi \sin R \cos R \sup_{D_R} d_C.$$

Thus $\sup d_C \geq \tan R$ and $\sup \rho_C \geq \sin R$. Identity has $d_C = \tan r$, so equality is attained. The conclusion permits arbitrary nonlinear appearance dependence and arbitrary magnitude a, not only orthogonal phase coordinates. It requires one common nonzero slice; it does not cover unrelated small appearance patches chosen separately at each geometry state.

K.4 Every width: the small-radius lower and quaternion upper

On a circle of radius r, parallel appearance transport is $H_r = e^{-2\pi i \sin^2 r} I_p$. Put $r_* = \min(R, \pi/4)$. Every unit color is separated from its transported endpoint by spherical angle $\beta = 2\pi \sin^2 r_* \leq \pi$. Undoing parallel transport along a common nonzero slice and normalizing its length gives a sphere path whose speed is at most $d_C \sqrt{h(\dot{z}, \dot{z})}$. Its endpoint separation therefore gives

$$\sup d_C \geq \frac{\beta}{2\pi \sin r_* \cos r_*} = \tan r_*, \qquad \sup \rho_C \geq \min(\sin R, 1/\sqrt{2}).$$

For full-space diffeomorphisms a fixed coordinate q_0 outside the bounded set of zero preimages supplies such a slice; punctured-space diffeomorphisms supply it directly. This argument gives the sharp lower at small radii for all widths and the cap lower for even widths.

For two complex heads, identify (q_1, q_2) with the quaternion $q_1 + q_2 j$. Write

$$a = \cos r, \qquad b = \mathrm{sinc}(r)(z_1 - iz_2), \qquad S(z)(q_1, q_2) = (aq_1 - b\overline{q_2},\ aq_2 + b\overline{q_1}).$$

This is left multiplication by the unit quaternion $a + bj$. In real blocks, if $B = \left[\begin{smallmatrix} \Re b & \Im b \\ \Im b & -\Re b \end{smallmatrix}\right]$, then $S = \left[\begin{smallmatrix} aI & -B \\ B & aI \end{smallmatrix}\right]$ and $S^\top S = I$ because $B^2 = |b|^2 I$. Its two complex coefficient functions are $\overline{s}$, so

$$|d(a + bj) + i\omega(a + bj)|^2 = h.$$

Quaternion norm multiplicativity gives, for *every* appearance vector,

$$\|(dS + i\omega S)q\|^2 = \|q\|^2 h = \|Sq\|^2 h.$$

Hence $d_C = 1$ and $\rho_C = 1/\sqrt{2}$ everywhere. Applying this construction blockwise to pairs of heads supplies the upper for every even p. Choosing identity instead when $R \leq \pi/4$ attains the smaller $\sin R$ bound. This chart is real-linear but not complex-linear; that distinction is essential.

K.5 Full-space nonlinear charts reduce to a linear lower

Let $V(C) = \sup_{z,c\neq 0} \rho_C(z, C_z^{-1}(c))$. For a smooth global fiber diffeomorphism of $\mathbb{R}^{2p}$, let $q_0(z) = C_z^{-1}(0)$.

If q_0 is not constant, some derivative $\partial_v q_0$ is nonzero. Differentiating $C(z, q_0(z)) = 0$ gives $C_v = -C_q \partial_v q_0 \neq 0$ at that state. As physical color tends to zero, the normal derivative shrinks proportionally to $\|c\|$, but $C_v + i\omega(v)c$ tends to this nonzero vector. Thus $V(C) = 1$.

Otherwise translate the constant q_0 to zero and set $S(z) = C_q(z, 0)$. Taylor expansion yields

$$C(z, \epsilon u) = \epsilon S(z)u + O(\epsilon^2), \qquad C_v(z, \epsilon u) = \epsilon S_v(z)u + o(\epsilon).$$

The lower singular value of S and the horizontal metric are bounded away from zero on compact unit spheres. Substitution into (73) gives $\rho_C(z, \epsilon u) \to \rho_S(z, u)$, so $V(C) \geq V(S)$. Linear charts are admissible; therefore the uniform nonlinear minimax equals the linear minimax. This reduction uses arbitrarily small nonzero appearances and the full color domain. It is not a claim that nonlinear coordinates cannot help on a bounded application-specific color range.

For $C = S(z)q$, put $J_0 = \mathrm{diag}(J_2, \ldots, J_2)$, $J_2 = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$, and

$$A_v = S_v S^{-1} + \omega(v) J_0, \qquad \Lambda(S) = \sup_{z,v\neq 0} \frac{\|A_v\|_2}{\sqrt{h(v,v)}}.$$

The uniform correlation equals $\Lambda/\sqrt{1+\Lambda^2}$.

K.6 Odd-head lower from a matrix loop

Suppose p is odd, $R > \pi/4$, and $\Lambda(S) < \tan R$. Let $\tau = \sin^2 R$, $\ell = \sin R \cos R$, and $S_\phi = S(R\cos\phi, R\sin\phi)$. The matrix path

$$U_\phi = e^{\tau\phi J_0} S_\phi S_0^{-1}, \qquad 0 \leq \phi \leq 2\pi,$$

starts at I, ends at $e^{2\pi\tau J_0}$, and obeys $\|U_\phi' U_\phi^{-1}\|_2 \leq \Lambda\ell$. If $y' = By$, the normalized vector has derivative $(I - \widehat{y}\widehat{y}^\top)B\widehat{y}$, of norm at most $\|B\|_2$. Close the path by the positive rotation from angle $2\pi\tau$ to 2π, giving a based matrix loop T whose every normalized vector orbit has length

$$< 2\pi \tan R\, \ell + 2\pi(1-\tau) = 2\pi.$$

A closed sphere path of length less than 2π cannot pass through the antipode of its start. Thus no $T(t)$ has a negative real eigenvalue. The interpolation $(1-s)T(t) + sI$ remains invertible and contracts the loop inside $GL^+(2p, \mathbb{R})$.

However, the boundary loop $S_\phi S_0^{-1}$ is based-nullhomotopic since S extends over the disk. Explicitly, evaluate S on $(1-s)R(\cos\phi, \sin\phi) + s(R, 0)$. Substituting this contraction into U shows that T is homotopic to the simultaneous 2π rotation in p real coordinate planes. Polar decomposition retracts $GL^+(2p, \mathbb{R})$ onto $SO(2p)$. For $2p \geq 3$, a single-plane rotation generates $\pi_1 SO(2p) = \mathbb{Z}/2$; the simultaneous loop has class $p \bmod 2$. For $p = 1$, the corresponding group is $\pi_1 SO(2) = \mathbb{Z}$. These classical rotation facts imply a contradiction for odd p (Baues & Muro, 2008). Thus $\Lambda \geq \tan R$. Combined with the nonlinear reduction, identity upper, and even-head quaternion construction, this proves Theorem K.1.

K.7 Punctured-space extension and actual adverse targets

The derivative-at-zero proof does not apply to maps defined only on punctured appearance space. The same minimax nevertheless holds for global diffeomorphisms of that space, and its lower is attained or exceeded at an actual nonzero state, as follows.

For a full-space chart choose a coordinate sphere of radius A containing all zero preimages; for a punctured chart use $A = 1$. The normalized map $F_z(u) = C(z, Au)/\|C(z, Au)\|$ on S^{2p-1} has degree $+1$ or -1. It extends over the geometry disk, so its boundary loop is based-nullhomotopic. The degree is the orientation of the enclosed unique zero in the first case; in the second it follows from the diffeomorphism and radial deformation retraction. For odd $p \geq 3$ and $R > \pi/4$, suppose

the attained correlation maximum on this compact circle-times-sphere were below $\sin R$. Select $d < \tan R$ bounding the associated quantity in (73). The paths

$$A_\phi(u) = e^{i\tau\phi} F_{z_\phi}(u)$$

have speeds at most $d\ell$. Close each by the complementary positive rotation. All resulting sphere loops have length $< 2\pi$ and avoid $-F_{z_0}(u)$. The normalized straight interpolation to F_{z_0} is therefore a continuous contraction of the entire loop of sphere maps.

The original based disk contraction instead homotopes it to $u \mapsto e^{i\phi} F_{z_0}(u)$. For $n = 2p - 1 \geq 3$, the degree-one component of sphere maps has $\pi_1 \operatorname{Map}_1(S^n, S^n) \simeq \pi_{n+1}(S^n) \simeq \mathbb{Z}/2$. Evaluation at a base point gives the first isomorphism because $\pi_1(S^n) = \pi_2(S^n) = 0$; adjunction identifies the based group. A full rotation in one real plane maps to the nonzero degree-one unstable J class; simultaneous rotation in p planes adds this class p times (Baues & Muro, 2008, Remark 6.2 and Eq. (6.3)). Right composition by a degree ± 1 map is a homotopy equivalence and preserves non-triviality. For odd p the supposed contraction is impossible. The one-head phase proof and the compact small-radius angle proof cover the remaining cases.

Thus the compact maximum is at least the stated threshold. Its maximizing nonzero state supplies an actual exact-fit target. The *larger* full-domain supremum of an arbitrary nonlinear chart may still be unattained. These are different assertions; only the compact adverse-target existence is used in the local optimization interpretation.

K.8 Complex-linear restriction and prescribed-slice comparison

If $C = S(z)q$, $S \in GL(p, \mathbb{C})$, the uniform optimum is $\sin R$ for *every* p. Indeed,

$$\operatorname{Im} \oint \operatorname{tr}(dS\, S^{-1} + i\omega I_p) = 2\pi p \sin^2 R,$$

because $\det S$ is nonzero over the disk and its phase has zero boundary winding. The inequality $|\operatorname{tr} A| \leq p\|A\|_2$ gives $\Lambda \geq \tan R$. The quaternion upper lies outside this complex-linear class.

If one instead evaluates a single prescribed nonzero coordinate slice, its physical colors change with the chart. For $p \geq 2$ the sharp value is $\min(\sin R, 1/\sqrt{2})$, without a parity distinction. The small-radius lower still applies. The complex-unitary matrix

$$\begin{bmatrix} \cos r & -\sin r e^{i\phi} \\ \sin r e^{-i\phi} & \cos r \end{bmatrix}$$

applied to the slice $q = e_1$ gives $c = \overline{s}$ and $d_C = 1$; padding handles larger p. It does not furnish the uniform upper over all colors. At $q = e_2$ its angular term retains an additional connection contribution. This distinction prevents a favorable slice from being mistaken for a common-target-family comparison.

K.9 Stationary gradient and scalar heavy-ball consequences

For any fixed chart, choose a nonzero state x_* and set $y = f_C(x_*)$. The loss $L_y(x) = \frac{1}{2}\|f_C(x) - y\|^2$ has Hessian $G_C = J_C^\top J_C \succ 0$ at x_*. For a smooth stationary rule $x^+ = x - \eta(x)P(x)\nabla L_y(x)$, the derivatives of ηP multiply the zero gradient and disappear from its linearization. The local error map is $I - \eta_* P_* G_C$. Equations (70)-(71) apply even when each block's dense SPD scaling is optimally tuned at that state. Fixing the chart before choosing the adverse state and exact-fit target gives the sharp regional factors in Theorem K.1. This is a minimax local asymptotic assertion, not a fixed target's global complexity.

For scalar heavy-ball with smooth stationary coefficients, the augmented linearization is

$$\mathcal{A} = \begin{bmatrix} (1+\beta)I - \eta P G_C & -\beta I \\ I & 0 \end{bmatrix}, \qquad 0 \leq \beta < 1.$$

After diagonalizing $P^{1/2} G_C P^{1/2}$, each eigenvalue $\lambda \in [m, L]$ gives roots of $z^2 - (1 + \beta - \eta\lambda)z + \beta = 0$. If all root magnitudes are bounded by $q < 1$, then $\beta \leq q^2$ and the real trace obeys $|1 + \beta - \eta\lambda| \leq q + \beta/q$. Evaluating at m, L, and using $\beta \leq q^2$, implies

$$\frac{L}{m} \leq \left(\frac{1+q}{1-q}\right)^2, \qquad q \geq \frac{\sqrt{L} - \sqrt{m}}{\sqrt{L} + \sqrt{m}}.$$

The classical choices $\eta = 4/(\sqrt{L}+\sqrt{m})^2$ and $\beta = q^2$ attain this bound for the actual finite spectrum (Polyak, 1964). Optimizing the within-block scaling therefore gives

$$Q(\rho_C) = \frac{\rho_C}{1 + \sqrt{1 - \rho_C^2}}. \tag{74}$$

The one-head regional factor is $Q(\sin R) = \tan(R/2)$; the general full-domain factors follow by substituting Theorem K.1. For identity, a single smooth model-specific attaining rule is $P = \mathrm{diag}(A^{-1}, I)$, $\eta = \sec^2(r/2)$, and $\beta = \tan^2(r/2)$, where A is the geometry Gram block. Smoothness at $r = 0$ follows from their power series in r^2.

Necessary exclusions. At optimal heavy-ball parameters, repeated endpoint roots can have Jordan factors tq^{t-1}; q is not a guaranteed one-step loss ratio. For $G = \mathrm{diag}(1, 9)$, $\eta = \beta = 1/4$, and equal current/previous error $(0, 1)$, the next error is $(0, -1.25)$ and the objective increases by a factor 1.5625. On the block-whitened one-head quadratic the distinct eigenvalues are $1 - \rho, 1, 1 + \rho$. Three scheduled positive rates reciprocal to these values annihilate the frozen quadratic. Thus no unrestricted scheduled-method or finite-horizon lower follows. A full structured Schur solve also bypasses the retained block restriction. Since geometry dimension is two, that solve and the quaternion chart both have $O(p)$ arithmetic in this example; we claim no same-cost separation. Adam, Rprop, general momentum rules, Krylov methods, and line searches are not covered by (74).

K.10 Nonzero residuals: invariant robustness and an exact boundary lift

At an actual stationary least-squares state, let $r = f - y$, $J^\top r = 0$, $G = J^\top J \succ 0$, and

$$H = G + E, \qquad E = \sum_\alpha r_\alpha \nabla^2 f_\alpha, \qquad \epsilon = \|G^{-1/2} E G^{-1/2}\|_2 < 1.$$

Let ρ_G, ρ_H denote the normalized cross-block norms in these two SPD inner products. The generalized angle inequality (Lin & Sinnamon, 2012) gives

$$\max\left\{0, \frac{\rho_G - \epsilon}{1 - \epsilon\rho_G}\right\} \le \rho_H \le \frac{\rho_G + \epsilon}{1 + \epsilon\rho_G}. \tag{75}$$

To see the constants, $(1 - \epsilon)G \preceq H \preceq (1 + \epsilon)G$, so the identity map between the two inner products has norm ratio at most $\sqrt{(1 + \epsilon)/(1 - \epsilon)}$. The half-angle distortion bound, converted by $\cos\theta = (1 - \tan^2(\theta/2))/(1 + \tan^2(\theta/2))$, sends a nonnegative cosine a to bounds $(a \mp \epsilon)/(1 \mp \epsilon a)$. Taking the extremal pair across the two subspaces proves (75), with zero truncation when the lower numerator is negative.

At a stationary state, a smooth coordinate change gives congruences for both H and G, so ϵ and the block angles are invariant to invertible within-block units. Away from stationarity, the Hessian has an additional coordinate-second-derivative term and this assertion fails. Also, ϵ is controlled by residual size times a normalized second fundamental form, not by appearance-connection curvature alone. In the original four-output one-head model the joint Jacobian is square invertible, so stationarity forces zero residual; one cannot simply assign it an arbitrary normal residual.

A separate six-output construction supplies a nonzero-residual boundary theorem. Augment the one-head frame by $\widetilde{\Phi}(z) = [\Phi(z)^\top, \psi(z) I_2]^\top$, where $\psi(z) = (\|z\|^2 - R^2)^2$. On $\|z\| = R$, $\psi = d\psi = 0$, so the original Gram and connection are unchanged. For a selected boundary state and $\delta \ge 0$, choose target so the last two residual components are $\delta c/\|c\|$ and the first four are zero. This state is stationary. Because $\nabla^2\psi = 8zz^\top$ on the boundary, the residual Hessian adds only the positive radial geometry form $8\delta\|c\| zz^\top$. It leaves the tangential geometry-appearance correlation used in the circulation lower unchanged. Identity has only tangential cross-coupling and still attains $\sin R$. Hence the exact one-head boundary minimax is $\sin R$ for this declared nonnegative-residual target family, and its heavy-ball factor is $\tan(R/2)$. This is not an interior minimax theorem for arbitrary residuals or targets.

K.11 A quantitative loop bound for a general linear head

For $f(g, c) = \Phi(g)c$, write $D = \Phi^\top\Phi$, $\Pi = \Phi D^{-1}\Phi^\top$, and $K = D^{-1}\Phi^\top d\Phi$. For a closed geometry path, let $P' = -K(\dot{g})P$, $P(0) = I$, and $H = P(T)$. Metric compatibility gives

$P^\top D(g)P = D(g(0)) =: D_0$. Define the horizontal operator and its path length by

$$B_v = (I - \Pi)(\partial_v \Phi) D^{-1/2}, \qquad L_s = \int_0^T \|B_{\dot g}\|_2 \, dt.$$

Assume a common nonzero chart slice, full joint rank along the loop, and $L_s > 0$. Undo transport of the chart color and normalize it in D_0. Its sphere speed is bounded by $d_C \|B_{\dot g}\|_2$, using $\|N_v\| \le \|B_v\|_2 \|c\|_D$. If β is the spherical angle between the initial color and its holonomy-transported endpoint, then

$$\sup_\gamma \rho_C \ge \frac{\beta / L_s}{\sqrt{1 + (\beta/L_s)^2}}. \tag{76}$$

In a rank-two oriented real head with a nonzero slice extending across a spanning disk, the normalized color has a global phase lift. Its unwrapped connection circulation can replace the shortest endpoint angle β in this argument. In higher appearance dimension the path can leave a single phase circle; this unwrapped replacement is not generally valid.

K.12 A NONFLAT FAMILY OF OVERLAPPING GAUSSIANS

Fix $\sigma > 0$ and $a > 0$. Define two individually unit-norm features on the real line,

$$g_\mu(x) = (\pi\sigma^2)^{-1/4} e^{-(x-\mu)^2/(2\sigma^2)}, \qquad \Phi(m, a) = [g_{m-a}, g_{m+a}].$$

Their overlap is $s = e^{-a^2/\sigma^2}$. The orthonormal sum/difference frame is

$$Q_+ = \frac{g_{m-a} + g_{m+a}}{\sqrt{2(1+s)}}, \qquad Q_- = \frac{g_{m+a} - g_{m-a}}{\sqrt{2(1-s)}}.$$

Gaussian integration yields $K_m = k(a) J_2$, $K_a = 0$, with

$$k(a) = \frac{a}{\sigma^2 \sqrt{e^{2a^2/\sigma^2} - 1}}, \qquad k'(a) = k(a)\left[\frac{1}{a} - \frac{2a}{\sigma^2(1 - e^{-2a^2/\sigma^2})}\right] < 0.$$

The final inequality follows from $1 - e^{-2t} < 2t$ for $t > 0$. Thus $F_{ma} = -k'(a) J_2 \neq 0$: normalization of each individual feature does not remove the connection of the two-feature family. The two shifted Gaussians and their two center derivatives are linearly independent. Dividing a putative relation by a common Gaussian gives two distinct exponentials times degree-one polynomials, forcing all four coefficients to vanish. Consequently the joint Jacobian is full rank when both physical amplitudes are nonzero; a zero amplitude excludes an identifiable center and is not covered.

For completeness, the squared operator norms of the normal feature derivatives are, with $t = a^2/\sigma^2$,

$$b_m = \frac{1}{\sigma^2}\left(\frac{1}{2} + \frac{t}{2\sinh t}\right), \qquad b_a = \frac{1}{\sigma^2}\left(\frac{1}{2}\tanh(t/2) + \frac{t e^{-t}}{(1+e^{-t})^2}\right).$$

These follow by differentiating the sum/difference frame, subtracting its appearance projection, and integrating products of Gaussians and degree-two polynomials. On a rectangle of translation width W and separation $a \in [a_0, a_1]$, the flux is $W|k(a_0) - k(a_1)|$ and

$$L_s = W\big(\sqrt{b_m(a_0)} + \sqrt{b_m(a_1)}\big) + 2\int_{a_0}^{a_1} \sqrt{b_a(a)}\, da.$$

For $(\sigma, a_0, a_1, W) = (1, 0.35, 1.8, 8)$ these are 4.74973024 and 16.12693766. The rank-two unwrapped bound gives $\rho \ge 0.28252288$ somewhere on each admissible common slice. The shortest-angle version gives 0.09465960. This is a positive continuum lower, not a sharp Gaussian minimax, a hard-support-renderer theorem, or a guarantee that the confirmed image trajectories visit such states. Fixed column spaces can have overlapping features and zero curvature, so overlap alone is not a sufficient condition.

K.13 TWO SMOOTH GLOBAL PHASE CHARTS ESCAPE THE ONE-CHART LIMIT

For the one-head disk, take charts $c = e^{i\chi(z)}q$. Then $d_\chi = \|d\chi + \omega\|_{h^{-1}}$ and $\rho_\chi = d_\chi/\sqrt{1+d_\chi^2}$, independently of the nonzero appearance. We prove

$$\inf_{\chi_1,\chi_2} \sup_{\|z\|\le R} \min_{j=1,2} \rho_{\chi_j} = 0, \tag{77}$$

where both phases are smooth over a neighborhood of the entire disk. This contrasts with the one-chart value $\sin R$ and uses identical pointwise block-scaling access.

Let $\psi(x) = e^{-1/[x(1-x)]}$ for $0 < x < 1$ and zero otherwise, $Z = \int_0^1 \psi$, and $F(x) = Z^{-1}\int_{-\infty}^x \psi$. The elementary bound $Z \ge \frac{1}{2}e^{-16/3}$ implies $0 \le F' < 8$. With $\ell = \pi/2$, define on $[0, 2\pi]$ and extend periodically

$$p_1(s) = \frac{2\pi}{\ell Z}\psi\left(\frac{s-\pi/4}{\ell}\right), \qquad b_1(s) = s - 2\pi F\left(\frac{s-\pi/4}{\ell}\right).$$

Endpoint derivatives agree, $b_1' = 1 - p_1$, and $|b_1| \le B = 2\pi$. Let $p_2(s) = p_1(s-\pi)$, $b_2(s) = b_1(s-\pi)$. Their bump supports are disjoint, so at each angle at least one p_j is zero.

Given $\tau > 0$, choose $t = \min(R/4, \tau/16, 1/64)$, $N = \lceil 2B/\tau \rceil$, $\eta(r) = F((r-t)/t)$, and

$$\chi_j(r,\phi) = -\frac{\eta(r)\sin^2 r}{N} b_j(N\phi).$$

The integer frequency makes the phase single-valued; its identity core $r \le t$ removes the polar singularity. Select j with $p_j(N\phi) = 0$. Its connection residual is

$$d\chi_j + \omega = -\frac{b_j(N\phi)}{N}[\eta' \sin^2 r + \eta \sin 2r]dr + (1-\eta)\sin^2 r\, d\phi.$$

Since η' is supported on $[t, 2t]$, the radial dual component is at most $(B/N)(1+32t) \le 3\tau/4$. The angular dual component is $(1-\eta)\tan r \le \tan(2t) \le 4t \le \tau/4$. Therefore $d_{\chi_j} \le \sqrt{10}\tau/4 < \tau$, uniformly in state and color. Letting $\tau \downarrow 0$ proves (77).

For any fixed phase, the zero set of $d\chi + \omega$ is closed and has empty interior, since $d(d\chi + \omega) = 2d\,\mathrm{vol}_h \ne 0$. A finite union of these closed sets cannot cover an open disk. Thus no finite phase family attains zero everywhere. This nonattainment statement is not automatically the same claim for nonlinear charts evaluated at varying coordinate slices.

Installation between charts at unchanged physical state is $q_j = e^{i(\chi_i - \chi_j)}q_i$. The construction's angular second derivatives grow proportionally to N at a fixed radius outside the core, and a direct safe-chart rule has up to order N decisions per circular traversal. These are properties of this construction, not a universal lower bound on interfaces, evaluation time, or training iterations. Its bump/CDF evaluations and the within-block solves are not free. No trained policy, practical runtime, or adaptive-optimizer memory theorem is claimed from (77).

K.14 A REGULAR LOCAL ATLAS AND ITS RESTRICTED COVERAGE BOUNDS

The metric h is the round-sphere metric of radius $1/2$ in geodesic polar coordinates. Its distance is $d_h(a,z) = \arccos|s(a)^* s(z)|$ and its ball area is $\pi\sin^2\delta$ for $\delta < \pi/2$. The phase anchored by $\overline{s(a)^* s(z)}/|s(a)^* s(z)|$ is smooth wherever that overlap is nonzero. In polar coordinates about a it has the same connection form, so its correlation is $\sin d_h(a,z)$.

A cover of the radius-R disk by *full* geodesic balls of radius δ requires at least $\sin^2 R/\sin^2\delta$ balls by area. A maximal δ-separated set gives a radius-δ cover, while its disjoint radius-$\delta/2$ balls lie inside radius $R + \delta/2$; when $R + \delta/2 < \pi/2$ this gives the upper count $\sin^2(R+\delta/2)/\sin^2(\delta/2)$. Thus regular full-ball covers have order δ^{-2}. Choose a finer net so a newly installed anchor starts within $\delta/2$ and retain it until distance δ; after the first installation every refresh consumes at least $\delta/2$ path length. Hence refresh count is at most $1 + 2\,\mathrm{length}_h(\gamma)/\delta$. This bound conditions on an actual finite path length, not on a number of optimization iterations.

The full-ball count lower does not apply to arbitrary patches: on a thin radial sector the phase $\chi = -\sin^2 r(\phi - \phi_0)$ leaves a radial residual bounded by the sector width, allowing an order-δ^{-1} sector cover. The two global charts above further demonstrate why chart count without regularity or domain restrictions is not a universal complexity measure.

K.15 Verification scope

Eight separate numerical suites check the feature derivatives and full rank, classical scaling and scalar momentum, quaternion identities, overlapping-Gaussian integrals, stationary residual perturbations, the boundary lift, anchored atlases, and the two-chart construction. Their independent reruns pass. These finite checks corroborate the implementations and counterexamples; the minimax, topology, and nonattainment claims rest on the proofs above. Neither the explicit model nor its phase atlas is asserted to describe the entire Gaussian image renderer, and no empirical gain is inferred from these local spectral results.

L Quantitative Coupling and Optimizer-Memory Boundaries

The following complementary results make the retained-geometry restriction precise for a general smooth linear appearance head. They concern declared coordinate and memory contracts, not a new optimizer or a sign prediction for the measured image gains.

L.1 Integrability for arbitrary nonlinear appearance charts

Let $f(g, c) = \Phi(g)c$, $\Phi \in C^2$, $D = \Phi^\top \Phi \succ 0$, $\Pi = \Phi D^{-1} \Phi^\top$, and

$$K_i = D^{-1}\Phi^\top \Phi_i, \qquad F_{ij} = \partial_i K_j - \partial_j K_i + [K_i, K_j].$$

For $c = C(g, q)$ with invertible C_q, the appearance and geometry Jacobians are $J_q = \Phi C_q$ and $J_i = \Phi_i C + \Phi C_i$. Their cross Gram is

$$b_i = J_q^\top J_i = C_q^\top D(C_i + K_i C). \tag{78}$$

Thus cross-coupling vanishes exactly when $C_i = -K_i C$. Commuting mixed derivatives gives $F_{ij}C = 0$. A local diffeomorphism has an open appearance image; a linear matrix annihilating all its colors must be zero. Flatness $F = 0$ is therefore necessary for cancellation on an open geometry-appearance neighborhood.

Conversely, on a sufficiently small simply connected neighborhood, the compatible system $\partial_i S = -K_i S$, $S(g_0) = I$, has a path-independent invertible solution. For example, the horizontal vector fields $\partial_i - (K_i c) \cdot \partial_c$ commute precisely when $F = 0$; their local flows give S. ODE uniqueness preserves invertibility. Then $C = S(g)h(q)$, with any local diffeomorphism h, cancels the cross block. This is a local integrability statement. A multiply connected base can be flat with nontrivial global holonomy; cancellation on one special color does not imply matrix flatness.

Differentiating K and canceling mixed derivatives gives

$$F_{ij} = D^{-1}\left[\Phi_i^\top (I - \Pi)\Phi_j - \Phi_j^\top (I - \Pi)\Phi_i\right]. \tag{79}$$

In particular, $\partial_i D = K_i^\top D + D K_i$ and $F_{ij}^\top D + D F_{ij} = 0$. Under an invertible appearance basis $\widetilde{\Phi} = \Phi B$, $\widetilde{K}_i = B^{-1} K_i B + B^{-1} B_i$, $\widetilde{F}_{ij} = B^{-1} F_{ij} B$, and $\widetilde{D} = B^\top D B$. Zero curvature and its metric operator norm are appearance-basis invariant; the magnitude of a geometry component still depends on geometry units unless a metric or physical loop is specified. For a scalar head, $K = \frac{1}{2} d \log D$ is flat. A fixed feature span $\Phi = \Phi_0 B(g)$ is also flat, regardless of feature overlap.

L.2 Finite-loop lower with appearance-normalized units

Let $M = C_q^\top D C_q$ and $b(v) = \sum_i v_i b_i$. The appearance-normalized cross magnitude is

$$\kappa_C(g, q; v)^2 = b(v)^\top M^{-1} b(v) = \|C_v + K(v)C\|_D^2.$$

It measures the appearance-span part of the output motion at fixed q, without an arbitrary appearance-unit scale. Consider a closed geometry path γ, transport $P' = -K(\dot\gamma)P$, $P(0) = I$, and

$H = P(T)$. Metric compatibility yields $P^\top D(\gamma)P = D_0$ and $H^\top D_0 H = D_0$. For a single-valued chart on a common fixed slice q_0, set $c(t) = C(\gamma(t), q_0)$ and $w = P^{-1}c$. Then

$$\dot{w} = P^{-1}(\dot{c} + K(\dot{\gamma})c), \qquad w(T) - w(0) = (H^{-1} - I)c_0.$$

Taking the base-metric norm and integrating proves

$$\int_0^T \kappa_C(\gamma(t), q_0; \dot{\gamma}(t))dt \geq \|(I - H)c_0\|_{D_0}. \tag{80}$$

For a declared geometry arc length L, the supremum of unit-speed κ_C is at least the right side divided by L. Holonomy-fixed colors give a zero lower. This is accumulated projected output motion, not a dimensionless convergence rate; the stronger angle formulation appears in Appendix K.

L.3 An exactly solvable connection and sharp gauge constants

For $|x| < 1$, let $a = \sqrt{(1+x)/2}$, $b = \sqrt{(1-x)/2}$, and

$$\Phi(x, y) = \begin{bmatrix} a & 0 \\ 0 & a \\ b\cos y & -b\sin y \\ b\sin y & b\cos y \end{bmatrix}.$$

Then $D = I$, $K_x = 0$, $K_y = (1-x)J_2/2$, and $F_{xy} = -J_2/2$. The horizontal Gram has diagonal entries $\|c\|^2/[4(1-x^2)]$ and $\|c\|^2(1-x^2)/4$, giving joint determinant magnitude $\|c\|^2/4 > 0$ for $c \neq 0$. A positively oriented loop of Euclidean area $\mathcal{A}$ has exact holonomy $H = e^{\mathcal{A}J_2/2}$.

Use the matrix-valued one-form norm $\|K^S\| = \sup_{\|v\|_2=1} \|K^S(v)\|_2$. On a disk of radius $R < 1$, among smooth orthonormal linear frames $C = S(x, y)q$ the exact bounds are

$$\inf_S \sup \|K^S\| = R/4, \qquad \inf_S \int \|K^S\|^2 dxdy = \pi R^4/32.$$

Up to a constant reflection, write $S = e^{\chi J_2}$. Stokes gives $|\oint((1-x)dy/2 + d\chi)| = \pi R^2/2$, proving the first lower. The phase $\chi = -y/2 + xy/4$ gives the residual one-form $(y\,dx - x\,dy)/4$, attaining it. This field is divergence-free and tangent to the boundary, so its L^2 inner product with every exact differential is zero. Adding $d\eta$ can only increase its squared integral; polar integration yields the second equality.

For general nonlinear charts, the finite-loop lower instead gives $\sup \kappa_C \geq [\sin(\pi R^2/4)/(\pi R)]\|c_0\| = (R/4)\|c_0\| + O(R^5)$. Exact finite-radius equality above is claimed only for the orthonormal linear class. Deliberate horizontal transport around the loop changes the represented output by $2|\sin(\pi R^2/4)|\|c_0\|$; it is not the unchanged-output chart installation considered next.

L.4 A specified moving-frame memory-compression law

Normalize the head at a base state to $U = \Phi_0 D_0^{-1/2}$, so $U^\top U = I$. A closed frame loop supplies an orthogonal matrix H. Installing its endpoint frame without changing the output uses $z = H^\top q$ and $UHz = Uq$. Let $A \succ 0$ be an already stored diagonal update operator. Preserving that operator exactly requires $A_H = H^\top AH$ in the new frame. Compressing it to a diagonal B is a different memory policy. These statements freeze the operator before the next gradient is selected; they do not assert an Adam or Rprop recurrence.

For a loop of side length order ϵ and signed area $\mathcal{A}$, C^3 feature regularity gives $H = I - \mathcal{A}\Omega + O(\epsilon^3)$, where $\Omega = D_0^{1/2} F_{ij} D_0^{-1/2}$ is skew-symmetric. Consequently

$$A_H = A + \mathcal{A}[\Omega, A] + O(\epsilon^3), \qquad \inf_{B \text{ diagonal}} \|A_H - B\|_F = |\mathcal{A}|\|[\Omega, A]\|_F + O(\epsilon^3).$$

The commutator is symmetric with zero diagonal, and deleting off-diagonal entries is the orthogonal projection in Frobenius norm, proving the equality. Its entries are $\Omega_{ij}(A_{jj} - A_{ii})$, so both frame rotation and unequal eigenvalues are needed for the leading error. Isotropic A is an exact null for all orthogonal H.

In two dimensions, $A = \operatorname{diag}(a, b)$ and $H = R_\alpha$ give

$$\inf_{B \text{ diagonal}} \|H^\top A H - B\|_2 = |b - a|\,|\sin\alpha\cos\alpha|.$$

The off-diagonal entry is fixed for every choice of B, giving the lower; matching the diagonal leaves a symmetric matrix with eigenvalues equal to plus/minus that entry, attaining it. In the isometric output frame the worst-unit-gradient next-update error is exactly the rate times this quantity. Spectral optimality of diagonal deletion is not asserted in arbitrary dimension.

Even exact operator preservation need not lower the next loss. At fixed normalized head, the squared-loss Hessian is I and the exact difference is

$$L_{\text{diag}} - L_{\text{exact}} = \eta g^\top (A_H - B) g + \frac{\eta^2}{2} g^\top (B^2 - A_H^2) g.$$

Taking $g = (1, 1)/\sqrt{2}$ or $(1, -1)/\sqrt{2}$ can change its sign while both methods decrease loss. An operator-discrepancy result is therefore not an optimization-benefit theorem.

L.5 Which historical diagonal moments can be reconstructed?

For invertible H, a function of $\operatorname{diag} M$ that recovers $\operatorname{diag}(H^\top M H)$ for *every* PSD matrix M exists if and only if H is monomial (one nonzero entry in each row and column). For necessity, choose $M_\pm = I \pm t(e_a e_b^\top + e_b e_a^\top)$, $a \neq b$, $0 < t < 1$. They have the same diagonal but transformed diagonals differ by $4tH_{aj}H_{bj}$. Every column must therefore have at most one nonzero entry; invertibility forces a monomial matrix. Permuting and squaring the coordinate scales proves sufficiency. For orthogonal H, the condition is a signed permutation.

Adam's first moment does not restore the missing information. Starting at zero with histories $g_1 = u_\pm = (1, \pm 1)$ and $g_2 = -\beta_1 u_\pm$ gives

$$m_2 = 0, \quad M_2 = \lambda u_\pm u_\pm^\top, \quad v_2 = \lambda(1, 1), \quad \lambda = (1 - \beta_2)(\beta_2 + \beta_1^2).$$

The stored (m_2, v_2, t) agrees, but literal replay after a dense frame rotation produces distinct new second-moment diagonals. Bias correction does not change this indistinguishability. This theorem concerns counterfactual historical moment reconstruction; preserving a current preconditioner with an explicit frame is a different contract. Even a full second-moment replay does not make coordinatewise square roots and divisions rotationally equivariant.

For Rprop, positive widths δ, δ' independent of the future gradient satisfy

$$H \operatorname{diag}(\delta') \operatorname{sign}(H^\top g) = \operatorname{diag}(\delta) \operatorname{sign}(g) \quad \text{for every } g$$

away from zero-sign boundaries if and only if H is monomial. The right side jumps only across coordinate hyperplanes. Each column h_j creates a nonzero jump $2\delta_j' h_j$ on $h_j^\top g = 0$ on the left. Distinct independent columns cannot cancel that jump at a generic point, so each hyperplane must be a coordinate hyperplane. Monomial transport supplies the converse. This is a one-step obstruction even for isotropic widths; it does not deny useful chart changes that intentionally alter the optimizer. No universal quadratic-memory complexity lower follows, because structured, low-rank, isotropic, or explicit-frame policies can be cheaper.

L.6 Why decoupling does not predict the sign of a gain

Metric compatibility implies horizontal transport preserves $\|f\|$. If a prescribed loop changes f_0 to f_1, write $\Delta f = f_1 - f_0$. Squared loss obeys $L_y(f_1) - L_y(f_0) = -y^\top \Delta f$. The targets $(f_0 + f_1)/2 \pm \Delta f/4$ have equal norm and give opposite changes $\mp\|\Delta f\|^2/4$. The midpoint gives zero change. This counterexample is about a specified path, not a path selected by an objective-aware optimizer.

There is also no uniform fresh-GD advantage for local orthogonalization. Let $T = [\Phi_i c]$, $Z = \Pi T$, and $N = (I - \Pi)T$. Native and horizontal Jacobians are $[T, \Phi]$ and $[N, \Phi]$. Their output-kernel difference is $\Delta = -(NZ^\top + ZN^\top + ZZ^\top)$. When the joint Jacobian is full rank and $Z \neq 0$, N is injective and $NZ^\top \neq 0$. In the normal/appearance decomposition, Δ has a zero normal diagonal block and a nonzero off-diagonal block, hence is indefinite. For small positive common rate, $L_{\text{horizontal}} - L_{\text{native}} = -\eta r^\top \Delta r + O(\eta^2)$ therefore has either sign for suitable unit residuals. Nonzero curvature can coexist with zero local connection and identical first steps; flat normalization can likewise help or harm.

L.7 Projection, residual dependence, and exact-color elimination

For a finite geometry change $\Phi_0 \to \Phi_1$, the least-output-change color is $c_{\mathrm{preserve}} = D_1^{-1}\Phi_1^\top \Phi_0 c_0$. Its infinitesimal derivative is $-K_i c$. The target-optimal color is instead

$$c_1^* = c_{\mathrm{preserve}} - D_1^{-1}\Phi_1^\top r_0, \qquad r_0 = \Phi_0 c_0 - y.$$

This is classical projection, not a novel optimizer. At fixed full-rank Ψ, $G = \Psi^\top\Psi$, $a_* = G^{-1}\Psi^\top y$, and $g_a = \Psi^\top(\Psi a - y)$, the normal equations give

$$L(\Psi a) - L(\Psi a_*) = \tfrac{1}{2} g_a^\top G^{-1} g_a \le \frac{\|g_a\|^2}{2\lambda_{\min}(G)}.$$

Differentiating $\Phi^\top r = 0$ at the optimum yields

$$\partial_i c_* = -K_i c_* - D^{-1}\Phi_i^\top r.$$

Thus connection-only compensation omits a residual-dependent term, even for a flat connection. For example, $\phi(t) = (\cos t, \sin t)$ has $K = F = 0$, but at $t = 0, c = 1, y = (1, 1.3)$ the optimal-color derivative is 1.3.

An invertible appearance change leaves the attainable colors at fixed geometry unchanged, so exact color minimization eliminates that chart choice from the reduced objective under the usual fixed-rank and differentiability assumptions (Golub & Pereyra, 1973; Chen et al., 2025). Coordinate-dependent regularization or minimum-norm conventions need not preserve it. Dense Gram assembly, factorization, and solves cost $O(mk^2)$, $O(k^3)$, and $O(k^2)$, respectively; matrix-free methods trade factorization for repeated normal-operator applications. No compute advantage over these established methods is inferred from the identities. Finite coordinatewise Rprop updates do not carry out this elimination, which is why the normalization intervention can still change their measured trajectories.

M Additional context and protocol details

Representations and optimization geometry. Neural fields change architectural or input coordinates (Sitzmann et al., 2020; Tancik et al., 2020; Mildenhall et al., 2020); Gaussian systems improve initialization, allocation, compactness, or multiscale structure (Zhu et al., 2025; Zeng et al., 2025; Zhang et al., 2025; Tai et al., 2025; Li et al., 2026; Pan et al., 2026). Our intervention fixes the renderer and function family. Natural gradient, K-FAC, adaptive updates, whitening, and normalization already exploit parameter geometry (Amari, 1998; Martens & Grosse, 2015; Duchi et al., 2011; Desjardins et al., 2015; Neyshabur et al., 2015; Salimans & Kingma, 2016; Santurkar et al., 2018). Function-preserving rescalings and implicit parameterizations alter geometry or induce preconditioning (Dinh et al., 2017; Arora et al., 2018; Gehring et al., 2021). We do not claim these principles as new; these classical principles motivate the exactly paired intervention, coupled-coordinate construction, chart-family limitation, and prospective selection studied here.

Adaptive bases and learned decisions. Expected-gradient reparameterization, basis-dependent Adam, and SOAP study favorable adaptive-optimizer coordinates (DePavia et al., 2026; Singh, 2026; Vyas et al., 2025). Learned optimizers and differentiable hyperparameter optimization learn update rules or optimize through training (Andrychowicz et al., 2016; Maclaurin et al., 2015). SGCL instead uses an offline-calibrated score to select *one fixed chart* from a small library using a common initial gradient and forward-only candidate evaluations. The Gaussian result is primarily block-scale selection; the non-Gaussian controls separately test shape. Recent 3D work modifies batching, regularization, second-order steps, or learned optimizers (Zhao et al., 2025; Ding et al., 2026; Hsiao et al., 2026; Pearl et al., 2026); our bounded 3D test diagnoses a fixed-chart boundary rather than claiming to replace those methods.

M.1 Original selection and control protocols

OA breadth and native resolution. The breadth study uses 256×256 targets and 2,000 Gaussians in Flowers102, Oxford-IIIT Pets, DTD, Caltech101, CIFAR-100, and EuroSAT (Nilsback &

Zisserman, 2008; Parkhi et al., 2012; Cimpoi et al., 2014; Fei-Fei et al., 2006; Krizhevsky, 2009; Helber et al., 2019), with 500 training, 60 validation, and 200 held-out images per domain. Independent validation selects OA Adam at 0.016 and standard AdamW at 0.008; matched-policy controls instead fix optimizer and learning rate within each coordinate pair. Native-resolution DIV2K uses the bicubic $\times 2$ track (Agustsson & Timofte, 2017): 785 official training images for the basis, 12 for validation, 3 for pilots, and all 100 official validation images held out; official test is untouched. Its primary test uses 2,000 Gaussians, 5,000 updates, seed 1, and unmodified encoded-RGB targets.

Selection and stronger controls. SGCL is frozen after 48-image, eight-domain development, then tested on 96 new images from six development domains. An unchanged-policy confirmation uses 128 images in FGVC Aircraft, STL-10, SVHN, and GTSRB (Maji et al., 2013; Coates et al., 2011; Netzer et al., 2011; Stallkamp et al., 2011), absent from selector development. Each supplies 500 training images for its basis, 16 validation images for fixed-rule selection, and 32 held-out targets. A subsequent control study uses another 128 source/content-disjoint images in these four *already studied* domains, paired under Adam/AdamW; a nested 32-image timing-repeat subset adds no independent images. It evaluates matched full/no-candidate-loss refits, broader joint-block controls, and a horizon-conditioned joint-block predictor. Policies, budgets, and comparisons are frozen before fresh outcomes. All selector/control targets are 256×256 with 2,000 Gaussians, not native-resolution evidence.

Matching and measurement. Coordinate pairs preserve physical initialization, renderer, objective, and capacity; independently selected optimizer policies are labeled separately. Fits minimize internal-domain mean squared error (MSE); base Adam/AdamW use $(\beta_1, \beta_2) = (0.9, 0.999)$, $\epsilon = 10^{-8}$, and AdamW decay 0.01. SGCL's base learning rate is 0.016. In the common-RGB protocols, peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and learned perceptual image patch similarity (LPIPS; VGG v0.1) (Zhang et al., 2018) use common encoded-RGB reconstructions. Image-cluster bootstrap keeps optimizer pairs and repetitions together and weights domains equally. The joint-block study charges target initialization, acquisition, installation, updates, and synchronized state recording; startup, warmup, immutable prior residency, and final metrics are separate. Appendices B, G, and I specify the distinct protocols and clocks.

N Fixed-Measurement Coordinates for Interacting Gaussian Clusters

This appendix proves Theorem 1. The model is a continuum, common-width, one-dimensional Gaussian cluster. All centers move independently, and every appearance channel remains trainable. The result is not a theorem for the finite-pixel anisotropic renderer. We distinguish a stable field representation, the conditioning of its appearance block, and the identifiability of its geometry.

What is characterized beyond the existing constructions. Generalized empirical interpolation supplies coordinates defined by fixed measurements (Maday & Mula, 2013); RBF-QR and HermiteGF stabilize Gaussian field representations, including flat-kernel regimes (Fornberg et al., 2011; Kormann et al., 2019). We use these established ideas but characterize a different quantity: the worst Gaussian-span/probe-space angle over all distinct centers in a complete radius-bounded domain. The endpoint-coalescence extremizer, separation-independent uniform certificate and leading minimax constant concern that angle, not a new interpolation principle or unrestricted floating-point stability.

The explicit upper bound is informative when $\sqrt{n}\, b < 1$. For $n = 4, b = 0.1$, it gives $\rho \le 0.2$ and $A^*A \preceq (1 - 0.2^2)^{-1} I \approx 1.042 I$. At an identifiable exact fit this also bounds the optimally block-tuned stationary-gradient factor by 0.2, under the assumptions of Section 3.3. It does not bound Rprop's finite-step image error. Keeping the certificate below a fixed constant as n grows requires $b = O(n^{-1/2})$; at fixed $b > 0$ the lower bound tends to one. Distinct centers may approach one another, but geometry conditioning and numerical recovery of native coefficients can still deteriorate. No measured general-image decoder-cost advantage is inferred from this example.

N.1 Model, exact coordinates, and the angle certificate

Fix $\sigma > 0$ and an anchor m_0. Write

$$u_\mu(x) = (\pi\sigma^2)^{-1/4} e^{-(x-\mu)^2/(2\sigma^2)}, \qquad f(x) = \sum_{i=1}^{n} u_{\mu_i}(x) c_i, \qquad z_i = \frac{\mu_i - m_0}{\sqrt{2}\sigma}, \quad |z_i| \le b.$$

Centers are distinct and $c_i \in \mathbb{R}^p \setminus \{0\}$. Let H_n be the isometry whose columns are the first n orthonormal Hermite functions at this anchor and width. Its adjoint takes output inner products, not target-dependent statistics. With $\Phi = [u_{\mu_1}, \dots, u_{\mu_n}]$, define

$$W = H_n^* \Phi, \qquad q = (W \otimes I_p) c, \qquad A = \Phi W^{-1}, \qquad f = (A \otimes I_p) q.$$

The tensor notation only records independent color channels; all statements can be read columnwise. The anchor and measurements remain fixed during differentiation.

Put $y = (x - m_0)/\sigma$ and

$$h_k(x) = (\pi\sigma^2)^{-1/4} e^{-y^2/2} \frac{\mathcal{H}_k(y)}{\sqrt{2^k k!}},$$

where $\mathcal{H}_k$ is the physicists' Hermite polynomial. The generating function gives

$$u_{m_0 + \sqrt{2}\sigma z} = e^{-z^2/2} \sum_{k=0}^{\infty} \frac{z^k}{\sqrt{k!}} h_k, \qquad W_{ji} = e^{-z_i^2/2} \frac{z_i^j}{\sqrt{j!}}, \quad 0 \le j < n. \tag{81}$$

Thus W is a Vandermonde matrix with invertible row and column scalings. Distinct centers make the appearance chart bijective. Its field can be evaluated without explicitly recovering an ill-conditioned native coefficient vector, as shown below. This is a cardinal-coordinate construction, related to classical generalized interpolation and stable Gaussian basis changes (Maday & Mula, 2013; Fornberg et al., 2011; Kormann et al., 2019).

Let $S = \operatorname{range} \Phi$ and $P_H = H_n H_n^*$. The identity $H_n^* A = I$ implies

$$A = H_n + H_n^{\perp} T, \qquad A^* A = I + T^* T, \qquad \delta(S, H_n) = \frac{\|T\|}{\sqrt{1 + \|T\|^2}},$$

where $H_n^{\perp} T$ denotes the orthogonal tail, without requiring a finite complement basis. Holding q fixed gives $H_n^* \partial_g f = 0$. Every geometry tangent therefore lies in $H_n^{\perp} \otimes \mathbb{R}^p$, while appearance tangents span $S \otimes \mathbb{R}^p$. Their canonical correlation is bounded by

$$\rho \le \|P_S (I - P_H)\| = \delta(S, H_n), \qquad I \preceq A^* A \preceq [1 - \delta(S, H_n)^2]^{-1} I. \tag{82}$$

Neither inequality requires a lower amplitude ratio or a lower separation. The full differential nevertheless loses rank at a zero atom amplitude or an exact collision. For distinct centers and nonzero vector amplitudes, the functions u_{μ_i} and their center derivatives are linearly independent: after dividing by a common Gaussian, a putative relation is a sum of distinct exponentials times degree-one polynomials. All coefficients vanish. Applying this channelwise establishes the stated identifiable geometry domain.

N.2 Polynomial remainders give the exact graph

Let $P(t) = \prod_{i=1}^{n} (t - z_i)$, and write the polynomial remainder

$$R_k(t) = \operatorname{rem}_P(t^k) = \sum_{j=0}^{n-1} R_{kj} t^j.$$

Any sequence $m_k = \sum_i d_i z_i^k$ obeys $m_k = \sum_{j<n} R_{kj} m_j$. From Eq. (81), the field coefficients in probe coordinates are therefore

$$\widehat{f}_k = \sum_{j=0}^{n-1} V_{kj} q_j, \qquad V_{kj} = \sqrt{\frac{j!}{k!}} R_{kj}. \tag{83}$$

The first n rows form the identity and the remaining rows form T. This equality is exact for every distinct configuration, not an approximation of the represented function.

For $k \geq n$, Vandermonde interpolation gives the hook-Schur identity

$$R_{kj} = (-1)^{n-1-j} s_{(k-n+1,1^{n-1-j})}(z_1, \ldots, z_n). \tag{84}$$

Here a Schur polynomial is the sum of monomials of semistandard tableaux: rows are weakly increasing, columns strictly increasing, and entries belong to $\{1, \ldots, n\}$. One derivation uses the generating relation

$$\sum_{k \geq 0} R_{kj} v^k = \frac{v^j \sum_{\ell=0}^{n-1-j} (-1)^\ell e_\ell(z) v^\ell}{\prod_i (1 - z_i v)}.$$

Expanding the denominator in complete homogeneous polynomials and matching coefficients yields the hook Jacobi-Trudi expression. Equivalently, Cramer's rule for interpolation gives the alternating determinant quotient for the same Schur polynomial. These are classical symmetric-polynomial identities; the role here is to bound the entire Gaussian graph, including arbitrarily close centers.

Every tableau coefficient is nonnegative. Homogeneity and $|z_i| \leq b$ imply

$$|R_{kj}(z)| \leq C_{nkj} b^{k-j}, \qquad C_{nkj} = s_{(k-n+1,1^{n-1-j})}(1^n) = \frac{k!}{j!(k-j)(k-n)!(n-1-j)!}. \tag{85}$$

The displayed constant follows by counting hook tableaux, or the hook-content formula. At $z_1 = \cdots = z_n = b$, equality holds entrywise, and column j has sign $(-1)^{n-1-j}$ for every tail row. Hence

$$|T(z)x| \leq |T(b\mathbf{1})|\,|x|, \qquad \|T(z)\| \leq \|T(b\mathbf{1})\|.$$

For fixed n, j, C_{nkj} has polynomial growth in k, while $\sqrt{k!}$ dominates any fixed exponential. The squared majorants are summable. Distinct nodes tending to b converge to the endpoint graph in Hilbert-Schmidt norm by dominated convergence. This proves that the endpoint is the exact worst angle as a *supremum* over distinct dictionaries.

The confluent limit consists of a Gaussian centered at $m_0 + \sqrt{2} b\sigma$ times polynomials of degree below n. It is the translated Hermite space $D(b)H_n$. We use this rank-n limiting space only to characterize conditioning. The original dictionary at an actual collision has lower rank and is not admitted as an unchanged-capacity chart.

N.3 Sharp radius law and fixed-probe optimality

Let P_n project onto H_n. The unitary translation is $D(b) = \exp[b(a^* - a)]$, in oscillator notation. The commutator $[a^* - a, P_n]$ acts only between h_{n-1} and h_n, and has norm $\sqrt{n}$. Thus

$$D(b) P_n D(b)^* - P_n = \int_0^b D(t)[a^* - a, P_n] D(t)^*\, dt,$$

so the largest-principal-angle sine is at most $\sqrt{n}\, b$. It is at most one by definition. At the confluent endpoint, the first tail entry obeys $|T_{n,n-1}| = \sqrt{n}\, b$. The graph-angle formula then gives

$$\frac{\sqrt{n}\, b}{\sqrt{1 + nb^2}} \leq \Delta_n(b) \leq \min\{1, \sqrt{n}\, b\}.$$

The projection of the translated Hermite frame onto H_n has determinant $e^{-nb^2/2}$. To see this, normal ordering gives the finite product $H_n^* D(b) H_n = e^{-b^2/2} \exp(bL_n) \exp(-bL_n^*)$, where L_n is the truncated raising matrix. The two triangular exponentials have determinant one. Projection is invertible for every finite n, b, so $\Delta_n(b) < 1$. In particular $\Delta_n(b) = \sqrt{n}\, b + O_n(b^3)$ as $b \to 0$. At fixed $b > 0$ the lower bound tends to one as $n \to \infty$, identifying a genuine cardinality/radius tradeoff for these probes.

For any fixed real n-dimensional probe space H, let

$$\mathfrak{R}_n(b) = \inf_{\dim H = n} \sup_{S \in \mathcal{F}_n(b)} \delta(S, H),$$

where $\mathcal{F}_n(b)$ contains all distinct-center spans in the full anchor interval. Both $D(b)H_n$ and $D(-b)H_n$ are limits in this family. Their mutual largest-angle sine equals $\Delta_n(2b)$. The largest principal angle $\vartheta = \arcsin\delta$ satisfies the triangle inequality. Indeed, project any unit vector successively onto the two intermediate subspaces, normalize each nonzero projection, and apply the spherical-angle triangle inequality; when the sum of angles is at least $\pi/2$, the bound is immediate. Consequently

$$\sin\big(\tfrac{1}{2}\arcsin\Delta_n(2b)\big) \le \mathfrak{R}_n(b) \le \Delta_n(b). \tag{86}$$

Combining this with the radius law proves $\mathfrak{R}_n(b) = \sqrt{n}\,b + O_n(b^3)$. It is first-order minimax optimality of the *angle certificate*. We do not infer exact finite-radius minimax optimality of Hermite probes, or optimal actual coupling among all nonlinear charts.

N.4 Stable derivatives and explicit tails

Write $P(t) = t^n + \sum_{j<n} p_j t^j$. The remainder coefficients satisfy

$$R_{k+1,0} = -p_0 R_{k,n-1}, \qquad R_{k+1,j} = R_{k,j-1} - p_j R_{k,n-1} \quad (1 \le j < n).$$

Start with $R_{kj} = \delta_{kj}$ for $k < n$. Differentiating this recurrence, using $\partial_{z_i} P(t) = -\prod_{\ell\neq i}(t - z_\ell)$, gives every learned-center derivative without inverse node differences. On $|z_i| \le b$, $b > 0$, positivity and homogeneity in Eq. (84) also give

$$\sum_i |\partial_{z_i} R_{kj}| \le (k-j) C_{nkj} b^{k-j-1}.$$

For a truncation at $K \ge n$, orthonormality bounds the omitted field norm by $\|q\|_F E_K$, where

$$E_K^2 = \sum_{k\ge K}\sum_{j<n} \frac{j!}{k!} C_{nkj}^2 b^{2(k-j)}.$$

A derivative-tail bound replaces each factor $C_{nkj}b^{k-j}$ by $(k-j)C_{nkj}b^{k-j-1}$; converting z_i derivatives to physical centers multiplies by $1/(\sqrt{2}\sigma)$. Both tails converge. Higher fixed-order derivatives follow from the differentiated polynomial recurrence with summable majorants. These bounds establish continuity and inverse-gap-free field/derivative evaluation, not a universal floating-point error bound for every implementation.

Geometry conditioning is a different question. Even when the appearance Gram is uniformly bounded, arbitrarily weak atoms or almost coincident centers can make the full differential nearly singular. Local exact-fit convergence neighborhoods can shrink accordingly. Allowing collision-limit derivative Gaussians as extra independent parameters would change the represented family and is not part of the theorem.

N.5 Fixed normalization and a concrete pair

A fixed invertible appearance-only map multiplies J_c on the right and leaves J_g unchanged at the paired physical state. Both tangent spans, and therefore their canonical correlation, remain the same. This does not imply that Adam or a sign optimizer has the same trajectory: the metric inside the appearance block and the coordinate orientation seen by such optimizers can change. Differentiating a *geometry-dependent* decoder changes J_g as well and is a different intervention.

For a centered fixed-width pair at $\mu = \pm a\sigma$ and fixed probes h_0, h_1, the normalized sum and difference features have squared projection magnitudes

$$\gamma_+^2 = \operatorname{sech}(a^2/2), \qquad \gamma_-^2 = \frac{a^2/2}{\sinh(a^2/2)}.$$

Parity makes the two projections orthogonal. Since $t\cosh t \ge \sinh t$ for $t \ge 0$, the largest-angle sine is

$$\delta = \sqrt{1 - \operatorname{sech}(a^2/2)} = \frac{a^2}{2\sqrt{2}} + O(a^6).$$

The appearance Gram eigenvalues are $\cosh(a^2/2)$ and $\sinh(a^2/2)/(a^2/2)$, hence its defect from identity is $O(a^4)$. Both center derivatives are retained at this symmetric state; the anchor does not follow their midpoint. This is a pointwise illustration within the stated domain, not a two-dimensional image optimizer or a collision-uniform optimization theorem.

O RESIDUAL, FINITE-STEP, AND COST CONDITIONS FOR OPTIMIZER PAYOFF

The geometric statements do not characterize arbitrary optimizer histories. Here we give the exact comparison underlying Eq. (9), its matching classical control, and an actual finite-step bound. The generic matrix inequality is classical; the purpose is to specify precisely what it implies for a coordinate intervention (de Klerk et al., 2017; Golub & Pereyra, 1973; Agarwal et al., 2010).

O.1 SHARP LOCAL GAIN AFTER SCALAR CALIBRATION

Let $F(\theta, a) = \Phi(\theta)a$ have full-column-rank Jacobian $J = [B, A]$ at a state, and residual $r = F - y$ in one declared output Hilbert space. Within-block whitening gives

$$S = \operatorname{diag}(B^*B, A^*A)^{-1/2}, \quad G = SJ^*JS = \begin{bmatrix} I & C \\ C^* & I \end{bmatrix}, \quad b = SJ^*r, \quad \rho = \|C\| < 1.$$

A normalized tangent u denotes physical tangent Su, with quadratic model decrease $d(u) = -b^*u - u^*Gu/2$. For $b \neq 0$, compare

$$u_N = -G^{-1}b, \qquad u_E = -\alpha b, \qquad \alpha = \frac{\|b\|^2}{b^*Gb}.$$

The second action performs exact within-block solves and optimal scalar calibration; it is not the production Rprop baseline. Direct substitution gives

$$d(u_N) = \tfrac{1}{2}b^*G^{-1}b, \quad d(u_E) = \frac{\|b\|^4}{2b^*Gb}, \quad \frac{d(u_N)}{d(u_E)} = \frac{(b^*G^{-1}b)(b^*Gb)}{\|b\|^4}.$$

Cauchy-Schwarz yields the lower bound one. For completeness, let $m = 1 - \rho$, $M = 1 + \rho$. Spectral calculus gives $G + mMG^{-1} \preceq (M + m)I$. If $x = b^*Gb/\|b\|^2$ and $y = b^*G^{-1}b/\|b\|^2$, then $x + mMy \leq M + m$. The product xy is at most $(M + m)^2/(4mM) = 1/(1 - \rho^2)$. This is the Kantorovich argument.

At fixed G, an eigenvector attains the lower bound. Equal normalized weights on eigenvectors at m and M attain the upper bound when $\rho > 0$; for $\rho = 0$ every nonzero direction gives one. Since J has full rank, each such b can be realized by an output residual. Thus high coupling alone does not determine useful improvement. At $b = 0$, both decreases are zero and their ratio is undefined.

For any tentative normalized action u, define its linear-system residual $s = b + Gu$. Completing the square gives

$$d(u_N) - d(u) = \tfrac{1}{2}s^*G^{-1}s, \qquad \frac{\|s\|^2}{2(1+\rho)} \leq d(u_N) - d(u) \leq \frac{\|s\|^2}{2(1-\rho)}. \tag{87}$$

For the following comparison specialize Eq. (87) to $u = u_E$, so $s = b + Gu_E$. Suppose the complete cost of the joint action is at most $\gamma > 1$ times that of u_E. The lower bound exceeding $(\gamma - 1)d(u_E)$ is sufficient for a quadratic-model decrease/work advantage. A necessary condition based on an actual cost ratio, or its *lower* bound γ, is $\rho^2 > 1 - 1/\gamma$. An upper cost bound alone does not imply that necessary condition. Computing the model and the screen is charged.

For a separable cluster model, write baseline and joint decreases as E_c, N_c, baseline total $E = \sum_c E_c$, and total baseline work C_0. If selecting clusters $\mathcal{S}$ costs an additional $h_0 + \sum_{c\in\mathcal{S}} h_c$, no selected joint action beats baseline decrease/work when

$$\sum_{c\in\mathcal{S}} (N_c - E_c) \leq \frac{E}{C_0}\left(h_0 + \sum_{c\in\mathcal{S}} h_c\right).$$

This follows by cross-multiplication of positive costs; strict reversal is sufficient if the joint minima and stated costs are attained. The ceiling $N_c - E_c \leq E_c\rho_c^2/(1 - \rho_c^2)$ shows why residual leverage matters in addition to correlation. Actual overlapping image clusters are not separable. A bound $(1 - \varepsilon)G_0 \preceq J^*J \preceq (1 + \varepsilon)G_0$ would require the extra quadratic allowance $\varepsilon \sum_c \|J_c d_c\|^2/2$ for each compared action. No global raster bound of this kind was measured.

O.2 Fresh anchored block solves equal Schur GN

Fix measurements L whose adjoint range equals $\mathrm{range}(A)$ at installation, with LA invertible. In coordinates $q = L\Phi(\theta)a$, let $W = LA$ and $K = W^{-1}LB$. Then $AK = P_A B$, since L extracts a full basis of the same appearance range. The geometry tangent at fixed q is

$$Z = B - AK = (I - P_A)B,$$

and the appearance tangent is AW^{-1}. These are orthogonal. Exact block least-squares minimization therefore solves

$$Z^*(r + Z\,d\theta) = 0, \qquad (AW^{-1})^*(r + AW^{-1}dq) = 0.$$

Pushing the chart tangent to native coordinates gives $da = W^{-1}dq - Kd\theta$. The complete output tangent is $B\,d\theta + A\,da = Z\,d\theta + AW^{-1}dq$, the full least-squares minimizer. Hence

$$p_{\mathrm{chart}} = -(J^*J)^{-1}J^*r = p_{\mathrm{Schur}}. \tag{88}$$

This is a full-rank tangent identity for every residual, not only at exact fit. Damping preserves it only when both methods use the same transported quadratic penalty. Separate identity damping in different charts defines different actions.

Chart Euler and native Euler can differ at second order. Give Schur the same finite realization,

$$\theta^+ = \theta + t\,d\theta, \qquad q^+ = q + t\,dq, \qquad a^+ = [L\Phi(\theta^+)]^{-1}q^+,$$

and its next state is identical in exact arithmetic on the valid chart domain. Matching repeated installations and realizations gives identical deterministic GN trajectories. A stable basis, retained factors and output statistics may be shared with the classical comparator. This does not prove equal implementation constants; it rules out attributing a distinct optimizer action to the chart in this comparison. Away from installation, block solves use a reusable triangular approximation, which can equally be represented by a native preconditioner.

Equation (88) concerns exact joint least squares. A damped continuum-local or same-pair approximation is not an exact global raster GN solve.

O.3 Actual finite-step loss and cost intervals

Let x^+ be the fully specified accepted endpoint of a method, including its state, chart realization and rejection, and set $d = x^+ - x$ in common physical coordinates. Assume the whole segment is admissible and DF is L_F-Lipschitz on it. Integration of the derivative remainder gives

$$F(x + d) - F(x) - Jd = e, \qquad \|e\| \le \eta := \tfrac{1}{2}L_F\|d\|^2.$$

With $R = \|r + Jd\|$, the triangle and reverse-triangle inequalities imply

$$\tfrac{1}{2}\|r\|^2 - \tfrac{1}{2}(R + \eta)^2 \le D := \ell(x) - \ell(x^+) \le \tfrac{1}{2}\|r\|^2 - \tfrac{1}{2}\max\{0, R - \eta\}^2. \tag{89}$$

These are actual loss-decrease bounds, including possible increases. For exact positive costs, candidate lower decrease/cost above comparator upper decrease/cost certifies a one-action advantage. For a signed decrease interval $[D_-, D_+]$ and positive cost interval $[c_-, c_+]$, take the minimum and maximum of the four endpoint quotients. Choosing a cost endpoint without checking the sign of the numerator is invalid. Overlapping ratio intervals are unresolved. A rejection has $d = 0$ but retains spent work.

A continuum or sparse surrogate needs a separate discrepancy allowance. If its Hessian error is bounded by ξ in a declared positive reference metric and gradient error by ζ, add $\xi\|d\|_{\mathrm{ref}}^2/2 + \zeta\|d\|_{\mathrm{ref}}$ to model uncertainty, and bound the true linearized residual norm. No smooth-segment or surrogate-error certificate for all native raster steps is inferred from synthetic checks. A one-action decrease/work certificate is also not a proof of better images at a fixed delivery deadline.

O.4 The sign-optimizer boundary

The production implementations use Rprop or its loss-conditioned reversal variant iRprop+, not exact block least-squares solves (Riedmiller & Braun, 1993; Igel & Hüsken, 2003). A fresh sign step of positive width w on $\ell(x) = x^2/2$ sends $0 < x \ll w$ to $x - w$. The loss ratio $(1 - w/x)^2$ is

arbitrarily large although the Hessian is identity and there is no cross-block coupling. With a positive width floor h, a state $x = h/2$ can enter the cycle $h/2, -h/2, h/2$: reversal detection inserts zero-update attempts, but equal losses do not trigger a beneficial loss-conditioned reversal. The floor is reachable by repeatedly shrinking a wider initial width. These are mathematical counterexamples, not descriptions of measured image failures.

Dense orthogonal coordinate changes can also alter componentwise signs even when they preserve the Hessian spectrum and step norm. Universal affine transport of a diagonal sign-width state is possible only for a monomial map (permutation times nonsingular diagonal scaling): the image of every axis-aligned update box must remain axis aligned. Generic mixed-atom charts violate that condition. Thus PCA and ZCA can share a covariance metric yet differ for iRprop+, and no exact dense sign-state transport or finite-step convergence follows from the coupling certificate. A safe-guarded step-shortening rule can enforce descent under additional bounds, but changes optimizer semantics and must be offered to the comparators at its complete cost.

P FIXED CHANNEL NORMALIZATION

P.1 WHAT IS NORMALIZED, AND WHAT REMAINS UNCHANGED

The inherited scale S is the range of each learned spatial PCA basis channel, not the standard deviation or covariance of training-image pixels. Absorbing it into physical coefficients and applying conventional fixed channel normalization are therefore different operations. The energy map additionally depends on each Gaussian's shape; the fixed-measurement construction mixes atom coefficients. These maps act on different axes and can be composed.

We estimated the mean m and population covariance Σ from the existing 500 official-training images in each of FGVC Aircraft, GTSRB, STL10 and SVHN. Every processed image has 256×256 pixels. Float64 accumulation is applied to the exact float32 limited-range YCbCr values. Variance and eigenvalue floors are $10^{-8} \operatorname{tr} \Sigma$. With $\Sigma = V \Lambda V^\top$, the unscaled coefficient decoders are

$$B_I = I, \quad B_z = \operatorname{diag}(\sigma), \quad B_{\mathrm{PCA}} = V \Lambda^{1/2}, \quad B_{\mathrm{ZCA}} = V \Lambda^{1/2} V^\top, \quad B_{\mathrm{rot}} = V.$$

These are the inverse maps of the corresponding pixel encoders, not inverse-covariance-shaped decoders. Eigenvalues are ordered and eigenvector signs fixed. We also test S. Each decoder except the separately labeled raw-S reference is divided by its RMS singular value $\|B\|_F/\sqrt{3}$ to separate scalar size from channel shape.

For a fixed frame B, physical colors satisfy $U = XB^\top$, and initialization is $X_0 = U_0 B^{-\top}$. In the energy chart this is composed with its shape-dependent factor. All original image-space means, shifts and backgrounds remain unchanged. The training mean is recorded but is not subtracted from every Gaussian coefficient: doing so would add a spatially varying term proportional to the sum of Gaussian weights. Neither the target nor the decoded objective is whitened. Thus this experiment tests matched coefficient-coordinate versions of standard normalization, not altered-target preprocessing or LayerNorm.

PCA and ZCA have the same physical covariance metric $BB^\top$ before common RMS normalization. They can still yield different sign updates because their orthogonal coordinate frames differ. The same-metric comparison must not collapse them into one optimizer policy.

P.2 COMPLETE MATCHED DEVELOPMENT COMPARISON

There is one already-known development image per domain, not a new held-out cohort. Every fit retains 2,000 Gaussians, all position, shape and appearance parameters, the original physical initialization, decoded loss and numerical path. No prior is retrained. The geometry multiplier is fixed at four from the prior development screen. For each RMS-normalized frame and each peak/energy chart, the same appearance-multiplier menu $\{1, 4, 16\}$ is available. A single multiplier is selected by mean raw decoded YCbCr MSE at 400 updates over these same four images. The raw-S reference uses multiplier four only. These are in-sample calibration summaries; selecting by PSNR afterward would define a different comparison.

The optimizer is iRprop+, with the common numerical safeguard and loss-conditioned reversal (Igel & Hüsken, 2003). It is not the deployed SGCL/Adam protocol. Initialization and differentiated decoder checks pass. Table 40 reports the entire frame set, including rotation-only and raw scale. RGB metrics are computed using the common evaluator; the selection loss remains raw YCbCr MSE.

Table 40: Fixed normalization and energy coupling. Four-image, 400-update development means after raw-loss selection within each fixed frame and shape chart. All normalized frames receive the same three color-rate choices. Raw scale is an unnormalized, fixed-rate reference. The calibration images are also the displayed images; no confirmation interval or equal-time normalization claim is made. Lower LPIPS and raw MSE are favorable.

Fixed color frame	Color factor	PSNR (dB)	SSIM	LPIPS	Raw MSE $\times 10^5$
Peak coordinates					
Identity	4	42.600	0.97294	0.14676	5.0479
Inherited scale, RMS-matched	4	42.589	0.97292	0.14576	5.0029
Z-score decoder	4	42.891	0.97542	0.13859	5.1135
PCA-whitening decoder	1	42.432	0.97106	0.14460	5.3337
ZCA-whitening decoder	4	42.252	0.97143	0.16092	5.8069
Rotation only	4	42.847	0.97473	0.14267	4.7778
Raw inherited scale	4	41.538	0.96183	0.16688	5.6975
Energy coordinates					
Identity	4	44.881	0.98081	0.11180	4.5741
Inherited scale, RMS-matched	4	44.869	0.98056	0.11354	4.4776
Z-score decoder	4	45.141	0.98113	0.10849	4.6280
PCA-whitening decoder	16	44.516	0.98066	0.11295	4.3823
ZCA-whitening decoder	16	44.028	0.97874	0.12414	4.5828
Rotation only	1	44.862	0.98045	0.11886	4.1476
Raw inherited scale	4	43.957	0.97789	0.12475	5.0240

Inherited scale and identity nearly tie after RMS matching and scalar calibration: peak PSNR is 42.5894 versus 42.6000 dB. Energy remains favorable relative to peak within each tested fixed frame. For z-scoring the difference is $45.1415 - 42.8911 = 2.2504$ dB; with identity it is 2.2808 dB. The rotation-only energy setting has the smallest mean raw selection loss, not the highest mean PSNR. These findings separate fixed normalization from geometry-dependent coupling in this development setting. They do not establish held-out whitening superiority, a result for raw-RGB statistics, equal-time normalization gains, or a direct SGCL/Adam comparison.

Q Conditional Color Structure and Joint-Fitting Cost

Q.1 Exact scores at the shared proposed geometry

Let g^+ be the first geometry update shared by all SGCL candidates. With geometry, opacity and affine decoding fixed, write the fitting image as $A_+c + b$, where c stacks physical color coefficients and A_+ includes the decoder. For $r_+ = A_+c + b - y$, M fitting scalars, and the actual rounded candidate increment d_j, the exact-arithmetic mean-squared loss is

$$L_j = L_0 + \frac{2r_+^\top A_+ d_j + \|A_+ d_j\|^2}{M}, \qquad L_0 = \frac{\|r_+\|^2}{M}. \tag{90}$$

The initial shared gradient constructs candidate updates, but alone does not determine this loss: the residual and directional curvature must be evaluated at g^+, not the original geometry. The color Hessian contains inter-atom overlap, not merely a 3×3 channel covariance. Rank deficiency requires a supported inverse or damping. As in variable projection, an invertible color chart leaves exact unregularized color least-squares predictions unchanged (Golub & Pereyra, 1973).

We implement the narrower exact optimization of reusing projection, sorting and active support at g^+ while retaining native RGB kernels and the original decoder and loss reduction. The frozen eight-candidate ridge policy, utility margin, fallback, epsilon and decay are unchanged. We do not substitute greedy one-step selection for the trained 100-step objective. Across eight initial states (four images and Adam/AdamW), same-covector checks preserve candidate images, features, decisions and the first real physical update and optimizer state. Separate backwards retain native atomic variability. Cache validity is restricted to this one shared geometry; no stale-geometry reuse is claimed.

A batched prototype and increment-quadratic implementation were diagnostic alternatives, not the timed deployed scorer. Measured numerical discrepancies are not universal error certificates.

Q.2 Measured acquisition and complete fitting

One already-observed image from each of Caltech101, CIFAR100, DTD and DIV2K low-resolution $\times 2$ is evaluated with two optimizer contexts and two timing repeats. The ten systems in Table 41 each have 16 trajectories; all 160 complete and all 800 retained outputs validate. Contexts and repeats are averaged within each of four images before computing means. This is a development diagnostic, not an independent confirmation or an estimate of a domain population. LPIPS was not evaluated.

The warm resident-prior clock includes target loading, initialization, selection, optimizer construction, accepted and rejected work, and synchronized CPU RGB output after every update. Prior preload, synthetic warmup and final archival are separate. Only actual outputs delivered within a budget receive credit. The pilot and two later control completions use the same protocol, but run in separate locked sessions. No foreign GPU workload is recorded; host storage quiescence was not enforced between trajectories, so these are qualified workstation timings.

Table 41: Shared previews and conditional-color controls. All quality columns are decoded RGB PSNR (dB); acquisition means cover the entire selector, not candidate kernels alone. No-loss refits omit candidate-loss features. Earlier and final controls retain their historical 100-update calibration; none is a newly selected 10-second optimum. Energy uses a different optimizer and algebraic covector conversion, so that row is a complete-system comparison, not a chart-only intervention. The CG and reset controls clear appearance moments and counters while preserving geometry optimizer state at each intervention. No curve, interval or inference treats the 160 trajectories as 160 independent images.

	100 updates	CPU-output budget			Acquisition (ms)
System	PSNR	2 s	5 s	10 s	
SGCL	23.036	28.528	28.694	28.817	10.996
SGCL, shared support	23.036	28.425	28.579	28.789	9.267
Appearance references and broader controls					
Scalar $8I$	22.989	29.071	29.358	29.228	1.176
No-loss refit (earlier)	22.957	29.430	29.641	29.935	9.824
No-loss refit (final)	22.965	29.524	29.508	29.772	10.006
Joint rates (earlier)	24.850	25.818	25.950	27.031	1.160
Joint rates (final)	26.632	25.065	25.205	27.118	1.177
Energy iRprop+	33.343	34.839	35.760	36.223	-
Periodic color interventions, every ten updates					
Shared SGCL + reset	20.044	18.738	19.080	19.109	8.874
Shared SGCL + CG	22.346	20.097	18.710	19.016	9.030

Mean whole acquisition falls from 10.996 to 9.267 ms, a saving of 1.729 ms (15.73% of acquisition). This equals only 0.0173% of a ten-second budget. At 100 updates, shared minus original SGCL is -0.00000824 dB. At ten seconds it is -0.0276 dB, with one positive image mean out of four; the descriptive four-image bootstrap interval is $[-0.2943, 0.2240]$ dB. Mean time to the 100-update CPU output is 0.4508 s originally and 0.4598 s with sharing. Thus exact cheaper acquisition is established on the tested states, but complete-fitting acceleration or degradation is not resolved. These four-image measurements do not establish a complete-fitting speedup.

Q.3 Fixed-geometry color references and the scope of a local solve

At two fixed-geometry Adam states with $2{,}000 \times 3$ color tensors, native-image and adjoint checks precede a damped least-squares correction. Unpreconditioned conjugate gradients (CG), with a 128-iteration cap and requested true relative residual 10^{-4}, converges in nine iterations in both cases (Table 42). Damping is 10^{-4} times a positive-floored deterministic probe estimate of mean curvature. The original reconstruction loss is evaluated separately from this damped correction objective.

Table 42: Fixed-geometry color references. One already-open image per domain. Adam and SOAP columns use 30 actual parameter updates. SOAP's first gradient-only call is charged, so 31 gradient calls produce 30 updates. These are untuned fixed-state references, not complete-fitting or best-calibrated SOAP comparisons. CG times are 15.45 and 7.79 ms, excluding separately recorded geometry setup. The first Adam timing includes lazy initialization and is not used for a speed ratio.

	Fitting-channel MSE					CG true relative residual
Image	Initial	CG	Adam	Scalar-8 Adam	SOAP	
Caltech101	.041398	.039230	.039269	.039311	.041317	7.50×10^{-5}
DTD	.061381	.060022	.060045	.060092	.061318	9.51×10^{-5}

The unchanged official SOAP implementation (Vyas et al., 2025) uses the full atom and channel axes, maximum preconditioned dimension 2048, no dimension merging, refresh period ten, learning rate 0.003, betas $(0.95, 0.95)$, epsilon 10^{-8} and zero decay. State projection, refresh, first-call behavior and dimension skipping are checked. Its accumulated gradient statistics are not the exact overlap Gram. One untuned setting cannot establish superiority over properly calibrated SOAP.

For the separate joint experiment, color CG runs every ten accepted updates with a cap of 32 iterations. All 3,795 installed corrections are finite and nonincreasing in the local native MSE; only 1,271 meet the requested true-residual tolerance, so these are not collectively exact least-squares solutions. CG exceeds its reset-only control by 2.302 dB at 100 updates, but remains 0.690 dB below ordinary SGCL. At ten seconds it is 9.801 dB below SGCL and 0.092 dB below reset-only. These results limit this schedule and memory treatment, not conditional optimization in general. Geometry state is unchanged at an intervention, but later geometry paths can differ. A local decrease is therefore not a certificate of better future joint fitting.

Reporting and numerical scope. Targets of $+1$, $+3$ and $+6$ dB from the shared initial RGB reconstruction are checked at retained CPU outputs. Although delivery is charged every update, only initial and 2/5/10-second images define the reported sampled attainment. These observations are not exact first-passage times; a missed transient crossing cannot be excluded and no interpolation is used. The shared scorer deliberately preserves native tile-cap behavior, verified on synthetic overfull tiles; initial real states are below that cap. This is not a guarantee for every later support or extreme geometry. No general 3D, cold-start, best-optimizer or asymptotic claim follows from the conditional-color experiment.

R Symmetric Calibration and a Separate Fresh-Image Confirmation

R.1 Question, scope, and frozen policy family

The first energy confirmation in Appendix J keeps energy at an ordinary-Rprop setting while its practical challengers receive a much larger calibration menu. The new study asks a different question: does the existing energy coordinate design remain useful when peak and energy receive the same restricted iRprop+ calibration opportunity? It introduces neither a new optimizer nor a new exponent, and does not test whether the fixed-measurement cluster theorem explains the measured benefit. The old candidate, policies, outcomes and adverse cases are preserved. All results in this

appendix retain original B0 projection-backward arithmetic. Appendix J.10 gives a five-image sensitivity study; Appendix T gives the subsequent complete B0/B1 replay of this cohort, including C, without changing the historical record.

For each atom let $a_0, c_0 > 0$ and b_0 be its initial triangular shape factor. The two existing geometry-unit conventions are

$$\begin{aligned} \mathrm{A}: &\quad (a,b,c) = (a_0 e^{u/a_0}, b_0 + v, c_0 e^{w/c_0}), \\ \mathrm{R}: &\quad (a,b,c) = (a_0 e^{u}, b_0 + \sqrt{a_0 c_0}\, v, c_0 e^{w}). \end{aligned}$$

The appearance map is $C = (a_0 c_0/ac)^p Q$, with $p = 0$ for peak and $p = 1/2$ for energy. Both are exact on the positive-diagonal chart and preserve the same anchored physical image. Five profiles are used for *each* chart, in the following deterministic tie-breaking order:

$$\mathrm{A444}, \quad \mathrm{A16161}, \quad \mathrm{A16164}, \quad \mathrm{R444}, \quad \mathrm{R4416}.$$

Their center/shape/appearance factors are respectively $(4,4,4)$, $(16,16,1)$, $(16,16,4)$, $(4,4,4)$ and $(4,4,16)$, multiplying base initial widths $(0.032, 0.032, 0.128)$. A/R identify the two maps, not additional optimized parameters. iRprop+ uses multiplicative width factors $(0.5, 1.2)$, bounds $[10^{-6}, 50]$, zero decay, and the published loss-conditioned reversal idea (Igel & Hüsken, 2003). Its project-specific whole-atom restoration and history clearing are shared by both charts and specified in Appendix J.9. No search grid, guard threshold, method or candidate exponent is changed after calibration.

R.2 Calibration and independently reserved targets

The 32 calibration images contain eight official-training identities per domain and overlap the frozen priors' training sets. They are the disclosed existing calibration cohort, not a new prior-independent validation set. Each of the ten chart/profile systems runs twice and delivers outputs at 2, 5 and 10 seconds: 640 unique calibration deployments and 1,920 endpoints. A profile is eligible only if every required deployment and output is complete. All 40 domain/chart/profile scores are eligible. For each domain and chart, the selected profile maximizes mean encoded-RGB PSNR over its eight images, two executions and all three budgets. This yields one profile per chart/domain, not a separate profile for each requested horizon.

Table 43: Frozen symmetric-calibration choices and historical incumbents. All displayed profiles use iRprop+. C is peak and retains the previous same-domain PSNR selections from the 111-challenger menu. E/P have the same current five-profile allowance, not equal cumulative research expenditure. The selection criterion and deterministic tie order are specified above.

Domain	Energy E	Peak P	C at 2 s	C at 5 s	C at 10 s
Aircraft	R4416	R444	R444	R444	R444
GTSRB	R444	R444	A16164	A16164	R4416
STL10	A444	R444	R444	R444	R444
SVHN	A16161	A16161	A16161	A16161	A16161

Confirmation contains 64 new official-test identities, 16 each from FGVC Aircraft, GTSRB, STL10 and SVHN. Before evaluation, a fixed SHA256 ordering selects identities not in the recorded exposure registry. The identity and content checks cover 2,795 known identities and 9,342 hashes from prior training, development, calibration, earlier diagnostics and confirmations. The prepared cohort has no recorded identity or source/preprocessed-content overlap with that registry. This is a statement about the recorded exposure history, not about undocumented data use. All four domains and their priors were used earlier; this is *fresh-image*, not fresh-domain, confirmation. Each prior contains 2,000 Gaussians for 256×256 fitting and is reused without retraining. The preprocessing remains RGB bicubic center fitting; small source images are upsampled, so this is not native-resolution evidence.

Five comparison roles are frozen before confirmation. E and P use the selected profiles; M is peak with exactly E's profile and iRprop+ settings; E_0 is the original energy A444 policy with ordinary Rprop; and C uses the historical per-domain/per-budget assignments in Table 43. Historical profiles are rerun on the new images, not paired with measurements from the old cohort. Only exactly

equivalent semantic policies share a trajectory. The resulting 512 unique confirmation deployments yield all 1,536 requested outputs with no failed or missing endpoint. Aliased roles add neither independent trials nor independent images.

R.3 Output clock, estimand, and main comparisons

Both executions use the same per-image initial physical parameters and seed across methods. Within-image order is balanced by cyclic rotation; the second fresh-process execution reverses order. The clock includes target preparation, initialization, gradients, optimizer updates, rejected proposals, rendering and synchronized delivery of the float32 YCbCr image to CPU. A budget receives only an output whose recorded delivery completed before its deadline, never a later render or interpolated quality. Resident-prior loading, synthetic warmup, process startup, final archival and offline evaluation are separately recorded. Outputs are buffered without persistent file writes during fitting.

RGB PSNR/SSIM use the frozen decoded evaluator. LPIPS uses the authenticated Graphdeco VGG v0.1 implementation and cached weights: decoded RGB is clamped once to $[0, 1]$, cast to float32, then mapped by $2x - 1$. Evaluation is in inference mode, with no additional resizing or quantization, after fitting. LPIPS is recorded for all 1,920 calibration and 1,536 confirmation endpoints; calibration selection uses PSNR only. Raw fitting MSE is in normalized YCbCr, not encoded RGB. A mean log-quality gain need not improve an arithmetic mean error.

The predeclared primary contrast is $E-C$ RGB PSNR. We average the two repeats and three budgets within each image, then image means within each domain, then the four domains equally. The 20,000-resample percentile bootstrap uses seed 2026091608 and resamples images independently within domains while retaining all their paired observations. Intervals condition on these domains, priors, calibration selections and observed executions; they do not include variation from drawing new domains or rerunning calibration. Secondary intervals remain descriptive. No delivery failure is replaced by a favorable prefix or survivor mean; all confirmation comparisons here are complete.

Table 44: Symmetric-calibration confirmation: all paired metric intervals. The primary inference is $E - C$ RGB PSNR. Other intervals are descriptive and not adjusted for multiple comparisons. Higher PSNR/SSIM and lower LPIPS are favorable. SSIM intervals cross zero for both practical comparisons. $E - E_0$ changes optimizer and profile as well as calibration; it is not an exponent-only effect.

Contrast	ΔPSNR (dB)	ΔSSIM	ΔLPIPS
$E - C$ (primary)	$+1.2421$	$+0.00102$	-0.01080
95% image interval	$[+1.04236, +1.45757]$	$[-0.00012, +0.00218]$	$[-0.01489, -0.00707]$
$E - P$	$+0.9394$	$+0.00060$	-0.01683
95% image interval	$[+0.70729, +1.17601]$	$[-0.00047, +0.00169]$	$[-0.02330, -0.01047]$
$E - M$	$+1.1542$	$+0.00301$	-0.02307
95% image interval	$[+0.91820, +1.40116]$	$[+0.00143, +0.00452]$	$[-0.02991, -0.01638]$
$E - E_0$	$+0.7389$	$+0.00127$	-0.01506
95% image interval	$[+0.58658, +0.88709]$	$[+0.00023, +0.00227]$	$[-0.02190, -0.00865]$

For $E - C$, the median image effect is $+0.6185$ dB with 47/64 wins. For $E - P$, it is $+0.3991$ dB with 45/64 wins. $E - M$ yields a median of $+0.7625$ dB with 53/64 wins, isolating the exponent at the profile selected for energy, not claiming that profile is optimal for peak. $E - E_0$ gains 0.7389 dB with 53/64 wins, but changes optimizer/profile and cannot be attributed solely to the coordinate map.

Mean raw fitting MSE is 7.53549×10^{-5} for E, 7.60256×10^{-5} for C, and 7.44965×10^{-5} for P. The ratios of arithmetic means are therefore 0.99118 and 1.01152: the latter is unfavorable despite favorable RGB PSNR and LPIPS. The corresponding raw-loss-difference intervals cross zero, as do SSIM intervals for both practical comparisons. Arithmetic clamped-RGB MSE is also 1.57% higher for E than P, consistent with the distinction between arithmetic and logarithmic aggregation. A descriptive reconstruction from the same paired records gives mean clamped-RGB MSE 2.9139648×10^{-4} for E and 2.8690138×10^{-4} for P. Their difference is 4.4951×10^{-6},

with image-cluster interval $[-6.8710 \times 10^{-6}, 1.8009 \times 10^{-5}]$ under the same 20,000-resample rule; it includes zero. These limits remain part of the result; "all metrics improve" is not supported.

R.4 Domain and budget boundaries

Table 45: All domain-budget cells for the two practical comparisons. Each cell first averages the two executions within each image. The same 16 images enter all three budgets in a domain. PSNR is in dB and lower LPIPS is favorable. The three budgets and the two contrasts are paired observations, not separate sample populations.

		$E-C$			$E-P$		
Domain	Seconds	ΔPSNR	Wins	ΔLPIPS	ΔPSNR	Wins	ΔLPIPS
Aircraft	2	-0.1053	7/16	$+0.00777$	-0.1053	7/16	$+0.00777$
	5	-0.1905	5/16	$+0.01237$	-0.1905	5/16	$+0.01237$
	10	-0.2431	4/16	$+0.01336$	-0.2431	4/16	$+0.01336$
GTSRB	2	$+3.6968$	16/16	-0.01778	$+2.9566$	15/16	-0.07580
	5	$+3.5438$	16/16	-0.01331	$+2.0114$	14/16	-0.05476
	10	$+2.8472$	15/16	-0.06812	$+1.4874$	13/16	-0.04102
STL10	2	$+0.3051$	13/16	-0.01057	$+0.3051$	13/16	-0.01057
	5	$+0.1314$	8/16	-0.00345	$+0.1314$	8/16	-0.00345
	10	$+0.0410$	8/16	-0.00108	$+0.0410$	8/16	-0.00108
SVHN	2	$+2.2735$	16/16	-0.03410	$+2.2735$	16/16	-0.03410
	5	$+1.4568$	16/16	-0.01018	$+1.4568$	16/16	-0.01018
	10	$+1.1481$	16/16	-0.00446	$+1.1481$	16/16	-0.00446

Budget-averaged $E-C$ domain effects are -0.1797, $+3.3626$, $+0.1592$ and $+1.6261$ dB, with 6/16, 16/16, 9/16 and 16/16 wins, respectively. Aircraft loses at every budget in both PSNR and LPIPS; STL10 is modest and nearly neutral at ten seconds. Against P, the GTSRB mean is $+2.1518$ dB; the other three domain effects are identical because $P = C$ there. Excluding GTSRB as a *descriptive* check gives $+0.5352$ dB over each practical comparator across the other three domains, with 31/48 image wins. This is not a new selected primary estimand.

Pooled $E - C$ gains at 2/5/10 seconds are $+1.5425$, $+1.2354$ and $+0.9483$ dB; $E - P$ gains are $+1.3575$, $+0.8522$ and $+0.6084$ dB. LPIPS mean differences are favorable at all three pooled budgets in both comparisons. At those deadlines, E delivers on average 272.9, 685.9 and 1,370.3 updates, versus 277.5, 699.1 and 1,405.8 for C, and 277.1, 698.6 and 1,403.1 for P. Fewer energy updates per deadline are charged; the result is not an equal-update advantage presented as a timing win. The two execution-level $E-C$ means are $+1.2462$ and $+1.2379$ dB; $E-P$ gives $+0.9485$ and $+0.9302$ dB. This descriptive repeat agreement does not create 128 independent images or remove timing uncertainty.

R.5 Numerical events, host activity, and technical recovery

All 16 calibration/confirmation batches pass the recorded GPU-isolation and host-preflight gates. No foreign CUDA workload or fatal monitor violation is recorded. Whole-host disk counters nevertheless show 50 confirmation sampling intervals above the inherited 30 MiB/s *preflight* threshold, lasting a total of 16.749 seconds and overlapping 47 of the 512 trajectory windows; the maximum is 480.792 MiB/s. The threshold is not a frozen rule for deleting post-launch outcomes. These counters identify neither the process nor its causal effect on fitting. Every output and both repeats are retained. The claim is observed delivered-budget quality under monitored warm execution, not perfectly isolated host latency, cold-start performance or a universal hardware speedup.

A technical recovery preserved the first completed 80-trajectory calibration batch and all frozen scientific settings; only the seven unstarted batches resumed. No trajectory or fresh image was rerun, and no calibration quality was inspected to choose the repair. This recovery created no second scientific selection opportunity.

Table 46: Symmetric confirmation: unique deployments and numerical-guard exposure. Counts refer to complete ten-second deployments; updates are attempted optimizer updates, not a matched-update comparison. Rejections count atom-events, not distinct atoms or rejected iterations. Pooled output-delivered updates at each budget are reported in Appendix R. The maximum allocated memory per system is a recorded resource statistic, not a memory-superiority claim. Exact semantic aliases share execution and do not multiply the sample count.

Domain	Policy	Complete	Mean updates	Atom rejections	Peak allocated MiB
Aircraft	iRprop+ energy R4416	32/32	1362.9	0	26.06
	iRprop+ peak R4416	32/32	1399.5	469	25.83
	iRprop+ peak R444	32/32	1382.8	5793	25.86
	Rprop energy A444	32/32	1383.2	0	25.96
GTSRB	iRprop+ energy R444	32/32	1369.5	0	25.92
	iRprop+ peak A16164	32/32	1398.0	512	25.63
	iRprop+ peak R4416	32/32	1411.2	0	25.57
	iRprop+ peak R444	32/32	1400.6	6	25.59
	Rprop energy A444	32/32	1412.6	697	25.70
STL10	iRprop+ energy A444	32/32	1402.6	0	25.86
	iRprop+ peak A444	32/32	1437.4	455	25.63
	iRprop+ peak R444	32/32	1428.3	12	25.67
	Rprop energy A444	32/32	1432.4	0	25.80
SVHN	iRprop+ energy A16161	32/32	1350.3	535	26.20
	iRprop+ peak A16161	32/32	1404.7	230	25.58
	Rprop energy A444	32/32	1411.6	399	25.65

R.6 Original-record closure and interpretation

The original prior role manifests, historical 64-image confirmation manifest, scale-shape validation choices and preoutcome/freeze records were recovered without retraining. The recorded historical confirmation identities are disjoint from the known training/exposure records, while all 32 historical calibration identities overlap prior training. This scopes the disjointness assertion to confirmation, not every study role. Source identities and hashes are auditable; these record checks do not re-decode every original dataset.

The historical no-rejection subgroup now exports separate equal-domain and equal-image aggregates. Primary means are respectively $+0.606135$ and $+0.587896$ dB on 49 images; matched means are $+1.860147$ and $+2.023121$ dB on 55 images. The original manuscript uses the equal-domain estimand. These treatment-dependent subgroups are not guard-free counterfactuals. Appendix G.7 additionally closes the original 448-unit prospective SGCL replay with zero changed decisions under exact invariant zeroing; no policy is refitted or replaced.

The new study supports broader practical usefulness under symmetric current calibration. It does not establish a universal default, unseen-domain energy transfer, or a causal explanation of finite-raster gains by the cluster theorem. The earlier $+1.8482$ dB matched and $+0.5909$ dB practical results remain separate cohort/protocol estimates. Subtracting either from a new mean would not estimate a paired improvement from calibration or optimizer changes. Anonymous end-to-end reproduction remains distinct from successful CPU table, provenance and policy replay.

R.7 Calibration scores and exact image identities

Table 47 reports all current calibration scores. Each score averages eight calibration images, two executions and all three output budgets. Selection uses the unrounded recorded score; the displayed values are rounded only for presentation. These are existing calibration outcomes, not additional confirmation evidence.

Table 48 fixes the 32 calibration and 64 confirmation identities. Aircraft entries are official seven-digit image identifiers (without the image extension). GTSRB entries are zero-based source sample

Table 47: Complete current calibration menu: mean RGB PSNR (dB). Bold selects the maximum within each row. All 40 scores are eligible. The peak and energy menus contain the same five profiles.

Domain	Chart	A444	A16161	A16164	R444	R4416
Aircraft	Peak	32.356282	30.402728	30.062153	**32.744137**	32.099016
Aircraft	Energy	32.384554	31.231808	30.807309	32.628445	**32.643067**
GTSRB	Peak	44.098772	44.686286	45.745092	**45.750495**	45.519532
GTSRB	Energy	48.347310	47.500857	46.430478	**48.894384**	47.618404
STL10	Peak	38.833960	35.031903	34.701150	**39.220995**	38.350886
STL10	Energy	**39.428578**	36.233090	35.191313	39.362901	38.842527
SVHN	Peak	51.512007	**55.080643**	54.672010	52.817731	51.807704
SVHN	Energy	54.949644	**56.218591**	55.547800	55.401687	54.492390

indices in the stated official split; STL10 and SVHN entries are zero-based sample indices in their official split arrays. Calibration uses the training split and overlaps prior training; confirmation uses the test split. Row order is within the frozen cohort, not a quality ranking. The existing 256×256 RGB bicubic center-fit preprocessing is unchanged.

Table 48: Symmetric-study calibration and confirmation image identities. The table lists calibration and confirmation source identifiers. It does not change the frozen split; data-role and overlap qualifications are given in Appendix R.

Row	Aircraft	GTSRB	STL10	SVHN
Calibration: official training splits				
1	0875344	62	506	71166
2	0958085	20446	996	6897
3	0114420	17846	2596	59069
4	1251094	8992	2507	23028
5	1718899	11984	2233	32181
6	1286492	12871	4220	439
7	0884477	4175	898	44864
8	1723789	24271	3466	24941
Confirmation: official test splits				
1	1891293	2416	5733	10520
2	1590457	1761	7442	21566
3	0525555	12314	5303	23007
4	1222395	7078	5649	10190
5	1852007	8310	789	6732
6	0977340	269	3699	22503
7	1092039	10265	2502	2356
8	0315569	1546	2939	9647
9	2156795	5274	3721	11846
10	1282206	9682	2819	10227
11	1201877	7681	284	24967
12	1831096	8085	7648	949
13	1646930	11484	6429	17002
14	1292176	6544	7641	23025
15	0109458	3472	2102	3919
16	1823272	4587	4599	19911

R.8 Complete deployed coordinate and optimizer recurrence

This specification consolidates the implemented peak/energy comparison, not a new optimizer. Its inputs are a common initialized Gaussian state, fixed opacity and affine image decoder, the frozen domain-specific profile in Table 43, and budgets $\{2, 5, 10\}$ seconds. The chart exponent is $p = 0$ for peak and $p = 1/2$ for energy. Every comparison uses fresh optimizer history.

Coordinates and pullback. For each atom, the raw state is $x = (r, u, v, w, Q)$, with $r \in \mathbb{R}^2$ and $Q \in \mathbb{R}^3$. The normalized center is $\tanh r$; its inherited projection to pixels is unchanged. With positive initial diagonals a_0, c_0, define

$$(d_a, d_b, d_c) = \begin{cases} (a_0, 1, c_0), & \text{A units,} \\ (1, \sqrt{a_0 c_0}, 1), & \text{R units,} \end{cases} \qquad a = a_0 e^{u/d_a}, \quad b = b_0 + d_b v, \quad c = c_0 e^{w/d_c},$$

$$h_p = e^{-p(u/d_a + w/d_c)}, \qquad C = h_p Q.$$

Initially $(u, v, w) = 0$ and $Q = C_0$; r is copied from the common initial raw center. The fitting loss is float32 pixelwise MSE in the original decoded, unclamped normalized-YCbCr domain. Fixed image mean, shifts and channel scales are applied exactly as in the shared renderer.

The shared sum-renderer wrapper projects the real atoms with a fixed zero-opacity sentinel and rasterizes at most 255 real atoms plus that sentinel per call on 16×16 tiles. Partial images are accumulated in float64 and returned in the native float32 image dtype; the common affine decoder is applied once. The sentinel has zero center, zero color and opacity, and triangular diagonals equal to the larger image dimension; it contributes no trainable parameter or rendered signal. Its coverage of every geometric tile is checked before rasterization. The kernel background is zero, and the external affine term is applied only once. Native support tests are unchanged between peak and energy.

Let (n_a, n_b, n_c) be the projection branch's returned shape covector. Before adding any amplitude-chain term, form the ordinary branch pullback

$$o_b = \tfrac{1}{2}\{n_b + (b/c)n_c\}, \qquad o_a = n_a - (b/a)o_b + \{b^2/(ac)\}n_c, \qquad o_c = n_c.$$

This algebraic conversion is evaluated in float64 and cast back to the incoming covector dtype; the preceding native projection-backward products remain float32 in the reported confirmation. B1 raises those internal products to float64 without changing this conversion, first in the targeted sensitivity study and then in the complete cohort replay (Appendices J.10 and T). If g_C is the ordinary physical-color derivative of the same loss, the raw chart derivatives are

$$g_u = \frac{a\, o_a - p\langle C, g_C\rangle}{d_a}, \qquad g_v = d_b o_b, \qquad g_w = \frac{c\, o_c - p\langle C, g_C\rangle}{d_c}, \qquad g_Q = h_p g_C.$$

The center derivative retains the shared projection and $\tanh$ chain. These are the actual branch-composition conventions; correcting the already combined shape/amplitude gradient would be a different update.

Algorithm 1: guarded iRprop+ in the frozen chart. Each raw coordinate has a positive width δ, preceding masked gradient z, and preceding intended update υ. Each parameter group also has an attempted-step counter. Let ℓ_{prev} denote the preceding pre-update loss. Initialize $z = \upsilon = 0$, counters to zero and ℓ_{prev} as undefined. The initial center/shape/color widths are $(0.032c_f, 0.032s_f, 0.128a_f)$ from the selected profile. All groups use $\eta_- = 0.5$, $\eta_+ = 1.2$, width bounds $[10^{-6}, 50]$ and zero decay. Repeat the following for each attempted update:

1. Verify that the current raw state, decoded colors and projector-range predicate of Appendix J are admissible. Evaluate the current loss ℓ and the branch-corrected raw gradient g. Refuse a nonfinite loss or gradient. Retain the current state x and widths δ before changing either.
2. Compute $s = g \odot z$ elementwise and set

$$\delta^+ = \mathrm{clip}_{[10^{-6}, 50]}\left(\delta \odot \begin{cases} 1.2, & s > 0, \\ 0.5, & s < 0, \\ 1, & s = 0. \end{cases}\right), \qquad \widetilde{g} = \begin{cases} 0, & s < 0, \\ g, & s \geq 0. \end{cases}$$

The provisional intended update is $\upsilon^* = -\operatorname{sign}(\widetilde{g}) \odot \delta^+$, with $\operatorname{sign}(0) = 0$. At entries with $s < 0$, replace υ^* by $-\upsilon$ only when ℓ_{prev} exists and $\ell > \ell_{\mathrm{prev}}$ strictly. Otherwise those entries stay zero. Propose $x^* = x + \upsilon^*$ and advance attempted-step counters. The stored υ is the intended coordinate delta, not a difference recomputed from rounded parameter tensors.

3. Decode the proposal and apply the same float32 projector-range predicate as in Appendix J. For every inadmissible atom, restore *all* of its raw center, shape and color coordinates to x; set its next $z = \upsilon = 0$ and its widths to $\mathrm{clip}_{[10^{-6},50]}(0.5\delta)$ using the widths from *before* this attempted step. For each admissible atom, retain x^* and set the next $z = \widetilde{g}$, $\upsilon = \upsilon^*$ and widths to δ^+. Keep all attempted-step counters and set the global $\ell_{\mathrm{prev}} = \ell$, including after atom rejections. There is no retry, line search, extra objective evaluation or best-state selection.
4. Account for elapsed time after the attempted update. If requested deadlines remain open, render the admitted state and synchronously copy an independent float32 YCbCr image to CPU. Each observation stores its completed-update age and *completed delivery time*. The initial image at age zero is delivered and charged in the same way.
5. For each deadline, retain only the most recent CPU image whose delivery completed at or before that deadline. Close elapsed deadlines before replacing the last image with a later delivery; never interpolate or credit an over-budget image. Stop once all deadlines are closed. A failed deployment keeps its available pre-failure images but marks every requested-budget record as failed-arm, not complete. Offline RGB conversion, PSNR/SSIM/LPIPS and archival are outside this output clock.

The projector-range predicate permits exactly zero determinant only as the declared inactive/culling case; nonzero determinants must also pass the finite reciprocal, conic and radius checks. Whole-atom rollback is a numerical admissibility rule, not a loss-decrease test. Different chart trajectories can encounter different rejections even though this rule is shared. Thus the exponent-only comparison identifies the coordinate policy's total effect, not a guard-free or decorrelation-only causal mechanism. Process startup, resident-prior preload and common synthetic warmup precede the clock; target preparation, initialization, attempted work and CPU-image delivery are included.

S Numerical Policies for the Secondary Predictors

This appendix specifies the two refitted appearance selectors and the horizon-conditioned joint-block predictor from Appendix I. They are separate from the unchanged original SGCL policy in Appendix G; none is fitted on confirmation outcomes. All logarithms below are natural. These numerical specifications do not add a new experimental comparison.

S.1 Full and no-candidate-loss appearance refits

The full refit uses the ten raw features of Table 22. The reduced refit deletes feature 1. Both set features 7 and 9 to zero exactly. Scores are the intercept plus the dot product with the raw features; no deployment-time standardization is performed. Candidate construction, tie priority and the $4I$ fallback are as in Appendix G. A different winner is replaced by $4I$ when its score margin is strictly below the stated threshold.

Table 49: Stored numerical parameters of the appearance-selector refits. Indices refer to Table 22. A missing loss feature is not a zero-forward method: the two block-response renders are retained.

Feature index	Full refit	No-candidate-loss refit
1	0.16034853788619383	—
2	0.024661572042685261	−0.45761738805347796
3	−0.11835647920867359	−0.20871938082734273
4	0.0442005491827814	0.14129486070000366
5	−0.047919561207869651	−0.076132186076390593
6	−0.02056096992763317	−0.040327522603806558
7	0	0
8	−0.018680487254259692	−0.015984250652384995
9	0	0
10	−0.00046146976230654265	0.018471103740614402
Intercept	−0.045608482328803143	−0.0090768818140156138
Fallback margin (dB)	0.047943445758444996	0.019880805664034988
Ridge penalty	100	10

S.2 Horizon-conditioned joint-block predictor

Let g_m, a_m be the candidate geometry and appearance multipliers, $H > 0$ the requested update horizon, $z = \mathbf{1}\{\mathrm{AdamW}\}$, and $v = \mathbf{1}\{\text{LR-only implementation}\}$. Put

$$x = \log g_m, \qquad y = \log a_m, \qquad t = \log(H/100).$$

Let M be the number of fitting scalars and L_0 the initial MSE. With $\tau = 10^{-12}$, set $d_0 = \sqrt{\max(ML_0, \tau)}$. Using the same geometry-only and appearance-only initial render changes as the block-response acquisition, define

$$\begin{aligned} G_r &= \log\max(\|Y_g - Y_0\|_F/d_0, \tau), \\ A_r &= \log\max(\|Y_a - Y_0\|_F/d_0, \tau), \\ R_g &= \log\max\left(\frac{\|G_U\|_F}{\max(\sqrt{\|G_{\mathrm{position}}\|_F^2 + \|G_{\mathrm{shape}}\|_F^2}, \tau)}, \tau\right). \end{aligned}$$

The appearance render change excludes the separately accounted AdamW decay. The frozen score is

$$u = b + \sum_{j=1}^{22} w_j \frac{\psi_j - \mu_j}{\sigma_j}, \qquad b = -1.7874283108145297.$$

Table 50 supplies the feature order and stored standardization. The ridge penalty is 0.1; fitting uses equal total domain weights, feature-scale floor 10^{-6} and the development horizons $30, 100, 400$. The requested 200-update horizon is reserved rather than a fitting horizon.

Table 50: Stored horizon-conditioned predictor parameters. These coefficients multiply standardized features, unlike the raw-feature SGCL coefficients. Decimal representations preserve the supplied parameters.

ψ_j	μ_j	σ_j	w_j
x	1.0012125941421595	0.8713405719720404	1.5839370675989626
y	0.69314718055994917	0.80037742256862265	−2.5958993820295304
x^2	1.7616610510334263	1.5361144730796592	−0.91451103762485986
y^2	1.1210570324758184	1.3206390163540636	−0.62378059868231595
xy	1.3345917053283429	1.3735276528381595	−1.68375938457343
x^2t	0.10706292845589405	2.4754591795619123	−1.7694126281075508
y^2t	0.06813095447193239	1.8351247970718159	0.8670230569253472
v	0.1111111111111111	0.31426968052735682	0.12792939009017645
vt	0.0067526502516279292	0.35329829048043837	0.037462547854394214
vz	0.055555555555555552	0.22906142364542073	−0.049361543429893691
xt	0.060847546281903392	1.4057134434700735	2.0935986837965759
xt^2	1.1251502212439251	1.4617522075444833	−3.3382393621919646
xz	0.50060629707107052	0.79386639987438212	0.006861579326335015
xG_r	0.68015936710806923	0.72660232102100075	0.81111432875944223
xA_r	−2.6615524224888065	2.3847550581574208	−3.2882053157993885
xR_g	−3.0166063473200539	2.8141403821364461	2.6773234490098323
yt	0.042125224349010212	1.121630675056158	−1.2185432560913203
yt^2	0.7789501531688694	1.248431051871441	0.26586154419110086
yz	0.34657359027997187	0.6636379003329852	−0.039264548609627969
yG_r	0.47087956184404894	0.63924463261932485	−2.3892837206112034
yA_r	−1.842613215569167	2.1752527086494031	0.66391468105171514
yR_g	−2.088419778913877	2.5434162908271261	−5.6350030388336272

All candidates use scalar 8. Eight use transported (g_m, a_m) pairs $(0.5, 0.5)$, $(1, 1)$, $(2, 1)$, $(2, 2)$, $(4, 2)$, $(4, 4)$, $(8, 4)$ and $(8, 8)$; the ninth is LR-only $(4, 2)$. Descending utility is primary. Exact ties use that listed pair order with LR-only $(4, 2)$ immediately after transported $(4, 2)$. The fallback horizon minimizes $|\log(h/H)|$ over $h \in \{30, 100, 400\}$, breaking a tie toward smaller h. At 30 the fallback is transported $(1, 1)$ for both optimizers; at 100 it is transported $(4, 2)$ for Adam and LR-only $(4, 2)$ for AdamW; at 400 it is transported $(2, 1)$ for both. A winning score less than 0.1 dB above this fallback selects the fallback. Thus a 200-update request uses the 100-update fallback when logarithmic distances tie, but still evaluates its features at $H = 200$. No margin is a statistical confidence guarantee.

T COMPLETE SAME-COHORT REPLAY UNDER REPAIRED ARITHMETIC

Question and intervention. Does the calibrated energy comparison survive the projection-backward arithmetic repair when the incumbent and every frozen comparison role are included? We repeat the complete symmetric-study cohort from Appendix R, using both original B0 and repaired B1 arithmetic from Appendix J.10. B1 computes internal projection-backward matrix and factor products in float64 and returns float32; both backends retain the existing algebraic covector conversion. Forward rendering, support, initialization, data, priors, coordinate maps, optimizer policies and output budgets are unchanged. This is replication on 64 now-exposed images, not another independent confirmation cohort or a new B1 calibration. The historical B0 table and figure remain separate records.

Complete population and execution. The frozen 64 images comprise 16 per domain, with two fresh-process repeats and 2, 5 and 10 second budgets. Exactly equivalent semantic policies share a trajectory as before; E_0 is ordinary-Rprop energy, not an alias for iRprop+ energy. Each backend executes 512 deployments and supplies all 1,536 requested outputs: 1,024 complete deployments and 3,072 attained-budget records overall. No scientific deployment fails. The 256 image/backend/repeat processes reverse backend and arm order between repeats. All 128 image/repeat pairs retain bit-identical initial physical images across backends and methods. Policies, aliases and thresholds are frozen before fitting; no outcomes select a replacement policy.

The original warm CPU-delivery boundary charges target preparation, initialization, chart/optimizer setup, attempted updates, numerical guards, rendering, synchronization and delivery. Resident-prior preload, startup, common synthetic warmup, archival, RGB conversion and metrics remain excluded. Only the latest image fully delivered before each deadline receives credit. Two setup/import failures before scientific measurements were corrected by restoring authenticated historical dependencies; the complete frozen scientific manifest was then executed. No quality outcome was used to select a restart.

Table 51: Complete frozen-policy replay under original and repaired arithmetic. Original B0 and repaired B1 arithmetic. Each backend contains 512 deployments and 1,536 attained-budget outputs on the same 64 images, with two fresh-process repeats and 2, 5, and 10 second warm CPU-delivery budgets. Contrasts average repeats and budgets within image and weight the four domains equally. Intervals use 20,000 image-within-domain bootstrap draws and are descriptive. Positive PSNR and negative LPIPS favor energy. The MSE column is the ratio of arithmetic mean clamped RGB errors; a value above one favors the control. C is the frozen incumbent, P the equally calibrated peak policy, M the matched-profile peak intervention, and E_0 the ordinary-Rprop energy reference. Policies were not recalibrated for B1; this is exposed-cohort repair replication.

Backend	Contrast	ΔPSNR (dB), 95% CI	ΔLPIPS	RGB MSE ratio
B0	$E - C$	$+1.223\ [1.021, 1.438]$	-0.0105	0.9925
B0	$E - P$	$+0.907\ [0.680, 1.141]$	-0.0162	1.0129
B0	$E - M$	$+1.120\ [0.887, 1.363]$	-0.0221	0.9534
B0	$E - E_0$	$+0.723\ [0.574, 0.867]$	-0.0147	0.9824
B1	$E - C$	$+1.230\ [1.028, 1.449]$	-0.0106	0.9921
B1	$E - P$	$+0.926\ [0.700, 1.157]$	-0.0164	1.0123
B1	$E - M$	$+1.135\ [0.903, 1.375]$	-0.0224	0.9509
B1	$E - E_0$	$+0.726\ [0.579, 0.869]$	-0.0146	0.9809

Estimand and uncertainty. For each contrast, average the two repeats and three budgets within image, then average image effects within domain and the four domains equally. The 20,000-draw percentile bootstrap uses the original seed 2026091608 and resamples images within domain, retaining all paired methods, repeats, budgets and backends. Intervals condition on the observed domains, priors and frozen selections; they are descriptive, not simultaneous, and do not include recalibration uncertainty. Neither the number of outputs nor repeated executions increases the sample beyond 64 images.

Results and adverse strata. The repaired $E - C$ gain is 1.22995 dB, with 49/64 favorable image means; $E - P$ gives 0.92605 dB and 48/64. Median gains are 0.60688 and 0.45228 dB, respectively. The contemporaneous B0 values are 1.22273 and 0.90735 dB; the historical B0 values remain 1.24206 and 0.93935 dB. Their numerical differences are not a claim of bitwise trajectory reproduction. The B1-minus-B0 $E - C$ interaction is $+0.00722$ dB with interval $[-0.01444, 0.02812]$; the corresponding $E - P$ and $E - M$ interactions are $+0.01869$ ($[-0.00143, 0.03885]$) and $+0.01437$ ($[-0.00469, 0.03365]$). Intervals spanning zero do not establish equivalence.

Table 52: Repaired B1 PSNR contrasts by domain. Mean PSNR differences (dB), averaged over both repeats and all three budgets. Each domain contains 16 images. Aircraft remains adverse for the incumbent and equally calibrated peak comparisons.

Contrast	Aircraft	GTSRB	STL10	SVHN
$E - C$	-0.147	$+3.315$	$+0.164$	$+1.587$
$E - P$	-0.147	$+2.100$	$+0.164$	$+1.587$
$E - M$	$+0.180$	$+2.100$	$+0.672$	$+1.587$
$E - E_0$	$+0.070$	$+0.683$	$+0.281$	$+1.872$

Aircraft remains adverse for $E - C$ and $E - P$, including a mean LPIPS increase of 0.01033. Across domains, B1 mean LPIPS decreases by 0.01058 against C (interval $[-0.01474, -0.00680]$), 0.01641 against P ($[-0.02285, -0.01020]$), and 0.02243 against M ($[-0.02916, -0.01586]$). B1 SSIM changes against C and P are $+0.00097$ ($[-0.00015, 0.00210]$) and $+0.00053$ ($[-0.00050, 0.00158]$); both intervals include zero. Against M, SSIM changes by $+0.00294$ ($[0.00136, 0.00445]$).

Higher mean PSNR does not imply lower arithmetic error. Against P, B1 arithmetic mean clamped RGB MSE is 1.23% higher, raw RGB MSE 1.12% higher, and fitting-space MSE 0.80% higher. The geometric clamped-MSE ratio is instead 0.80797, consistent with the positive mean logarithmic-quality gain. All adverse images and domains remain in the estimates.

Metric and numerical provenance. The authenticated Graphdeco VGG v0.1 LPIPS path applies one clamp to decoded RGB, casts to float32 and maps inputs by $2x - 1$, with fixed cached weights and inference mode. Four original GPU metric checks reproduce archived values exactly. CPU evaluations differ by at most 1.9431×10^{-5} and do not replace the GPU table. Saved-output fitting MSE is recomputed in float64 from unclamped normalized-YCbCr arrays; it is distinct from the live float32 optimizer scalar. Every returned PSNR value agrees with its clamped RGB MSE, and all paired means and bootstrap intervals are independently reconstructed from the output records. Eight preselected lossless arrays additionally reproduce RGB MSE and PSNR directly; this compact-array check is not a replay of every archived tensor.

Formal isolation records no foreign CUDA workload or fatal monitor violation over 36,878 samples, with a maximum observed GPU temperature of $65°$C. Whole-campaign disk activity includes archival, so these receipts do not establish a universal clean-host latency bound. The shared guards reject 9,513 atom proposals under B0 and 19,430 under B1; all attempted work remains charged and every deployment completes. These counts are atom-events, not missing images, failures or additional observations.

Supported scope. The complete frozen-policy replay closes the missing population-level repair comparison, including C, while retaining the original calibrated-control conclusion and its metric/domain limitations. It does not establish a universal optimizer, whole-renderer derivative accuracy, equivalence between arithmetic paths, an explanation of every finite-raster gain by continuous self-coupling, or transfer to new domains. The earlier targeted fixed-update and intervention diagnostics remain separately reported in Appendix J.10.