TOWARD SPECTRAL KINEMATICS: TRANSFORMATION LAWS AND THEIR LIMITS

Anonymous authors
Paper under double-blind review

ABSTRACT

Pretrained encoders map samples to feature vectors that can be stored and reused across tasks. Pooling, compression, and reweighting change their spectral diversity, the effective number of directions carrying feature energy. We study what information determines pooled diversity and what partial observations can guarantee. Building on classical transformation identities, we construct sources with identical spectra, energies, and principal angles but different pooled effective ranks. These summaries miss which energetic directions overlap. For second-order effective rank, which emphasizes high-energy directions, source spectra and energies plus one energy-weighted overlap total determine the answer exactly. Bounding unmeasured contributions to this total yields computable relative-error guarantees. Compatible completions, which preserve all observations while varying unmeasured overlaps, distinguish conservative bounds from information that permits opposite diversity-gain decisions. In a four-encoder vision cohort, raw features require a median of 1% of direction-pair measurements to guarantee at most 1% relative error; centering raises the median to full coupling on the original query ladder. Fewer measurements do not always reduce runtime: gains depend on cached source factors and bound-tightening costs. These results connect the information needed for pooled spectral diversity to reliable prediction and spectral-diversity batch admission, while delimiting their computational benefits.

1 INTRODUCTION

A pretrained encoder maps samples to reusable feature vectors. Pooling expands collections, compression reduces storage, and reweighting adjusts the relative energy carried by feature directions. Different collections may concentrate energy along the same directions, so adding samples need not add diversity. This leads us to ask how much spectral diversity each operation preserves or adds.

Singular value decomposition (SVD) separates a feature matrix into orthogonal directions (modes) and strengths (singular values). A mode's *energy* is its squared singular value, equal to the sum of squared sample projections onto that direction. The spectrum records strengths but omits directions. *Effective rank* summarizes a spectrum as an effective number of active directions, discounting weak modes. For uncentered features, energy includes both variation around the mean and the mean itself. A large shared mean can concentrate energy along a common direction even if it does not distinguish labels. Effective rank describes this energy distribution; task relevance requires separate evaluation.

A concrete use is batch admission based on spectral diversity: a feature bank under a fixed encoder assesses whether adding a candidate batch meets a specified diversity gain. With cached source spectra and directions, cross-source overlaps can be evaluated progressively until the decision is certified or exact computation is used. This use requires a reliable gain interval, not just a point estimate.

Roy and Vetterli's effective rank (Roy & Vetterli, 2007), used by RankMe (Garrido et al., 2023), exponentiates the Shannon entropy of normalized singular values. Vendi Score (Friedman & Dieng, 2023) applies this construction to normalized positive semidefinite similarity eigenvalues. AlphaReQ (Agrawal et al., 2022) measures covariance spectral decay; its exponent differs from our entropy order. We normalize squared singular values as energy shares. Their Rényi effective ranks (Rényi, 1961; Hill, 1973) form a family in which the order controls sensitivity to weak versus dominant modes.

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.

Recent work studies single-sample second-order novelty updates (Pasarkar & Dieng, 2026) and faster spectral data selection (Bilmes et al., 2026). Our *spectral kinematics* addresses sufficient information and partial-observation guarantees for general pooling. Power reweighting (Bercher, 2012) adjusts mode energies, restriction retains modes, and pooling combines sources. STAR (Lee et al., 2025) truncates and rescales singular components of parameter updates before combining models; here we stack sample feature rows and analyze their spectral diversity.

Coupling means overlap between feature directions from different sources. Overlap between strong modes affects diversity differently from overlap between weak modes. Principal angles summarize subspace geometry; coupling records which directions overlap and how strongly, preserving their association with energy when combined with source spectra.

From partial observations, we compute an upper bound on relative prediction error. If it does not exceed the chosen tolerance, prediction accuracy is guaranteed; we call this computable guarantee an *error certificate*. The bounds also support diversity-gain decisions. An inconclusive bound may reflect conservatism or missing information. *Feasible completions* assign possible values to unobserved couplings while preserving measured entries and subspace geometry constraints. If two completions yield opposite diversity-gain decisions, the observations cannot determine the answer. If tightening the bound resolves the decision without new measurements, the earlier uncertainty came from conservatism.

The classical identities in Section 2 provide the foundation. Our contribution identifies information gaps in general pooling and uses partial-observation bounds and compatible completions to analyze prediction and decision reliability:

1. **Information requirements for general pooling (Section 3).** Counterexamples show that identical spectra, energies, and principal angles can yield different pooled ranks. We identify the missing energy–direction association and a sufficient weighted-overlap scalar at order two.
2. **Guarantees from partial observations (Section 4).** We bound prediction error from partial coupling. Tighter bounds and opposite-decision completions distinguish bound conservatism from information insufficiency.
3. **Empirical scope and computational limits (Section 5).** Controlled interventions test the information gap; raw and centered features reveal when partial observation helps. Direct-product baselines separate measurement savings from runtime gains.

2 Spectral Preliminaries and Classical Identities

Samples share encoder $f_\theta : \mathcal{X} \to \mathbb{R}^d$, checkpoint, layer, and preprocessing. Pooling stacks $H_M = [H_A; H_B]$ without joint centering. Raw features are the baseline; source centering and row normalization test sensitivity. Equal dimensions do not align different encoders. For nonzero $H \in \mathbb{R}^{N\times d}$, compact SVD gives $H = U\Sigma V^\top$, where $k = \operatorname{rank}(H)$, U, V have orthonormal columns, and Σ contains all positive singular values. Thus

$$H = \sum_{i=1}^{k} \sigma_i u_i v_i^\top, \qquad Hv_i = \sigma_i u_i, \qquad \|Hv_i\|_2^2 = \sigma_i^2. \tag{1}$$

The entries of Hv_i describe how the samples vary along direction v_i. A mode can combine several feature coordinates, so restricting modes selects patterns of variation rather than individual columns. Set $E = \|H\|_F^2 = \sum_i \sigma_i^2$ and $p_i = \sigma_i^2/E$. Define

$$S_\alpha(p) = \sum_i p_i^\alpha, \qquad r_\alpha(p) = S_\alpha(p)^{1/(1-\alpha)} \quad (\alpha > 0,\ \alpha \neq 1), \tag{2}$$

with $r_1(p) = \exp(-\sum_i p_i \log p_i)$ and $0 \log 0 = 0$. Explicitly, $r_\alpha(p) = \exp(H_\alpha(p))$: effective rank is the exponential of Rényi entropy $H_\alpha(p) = \log S_\alpha(p)/(1-\alpha)$ for $\alpha \neq 1$. Low orders emphasize weak modes; high orders emphasize dominant modes. Its limits are $r_0 = |\operatorname{supp}(p)|$ and $r_\infty = 1/\max_i p_i$. The order changes the observable, not the matrix.

Theorem 1 (Classical transformation identities)**.** *Let p have positive masses on its support and $\alpha > 0$, $\alpha \neq 1$.*

(1) Power reweighting. *For a reweighting exponent $\beta > 0$, let $q_i = p_i^\beta / S_\beta(p)$. Then*

$$r_\alpha(q) = \left[\frac{S_{\alpha\beta}(p)}{S_\beta(p)^\alpha}\right]^{1/(1-\alpha)}. \tag{3}$$

(2) Subset restriction. *For nonempty $I \subseteq \mathrm{supp}(p)$, let $W_I = \sum_{i\in I} p_i$ and $q_i = p_i/W_I$ on I. Then*

$$r_\alpha(q) = \left[\sum_{i\in I}\left(\frac{p_i}{W_I}\right)^\alpha\right]^{1/(1-\alpha)}. \tag{4}$$

(3) Orthogonal merge. *Let $E_A = \|H_A\|_F^2$ and $E_B = \|H_B\|_F^2$ be the source energies, and $\gamma = E_B/(E_A + E_B)$ the energy fraction of source B. If $V_A^\top V_B = 0$, then*

$$r_\alpha(p_M) = \left[(1-\gamma)^\alpha r_\alpha(p_A)^{1-\alpha} + \gamma^\alpha r_\alpha(p_B)^{1-\alpha}\right]^{1/(1-\alpha)}. \tag{5}$$

Shannon limits and derivations are in Appendix B.

Second-order rank r_2 counts equivalent equally energetic directions: two sharing all energy give 2; one gives 1. It emphasizes high-energy modes more than r_1; many nonzero weak modes can still yield small r_2, unlike algebraic rank.

Power reweighting preserves directions: $0 < \beta < 1$ flattens energy shares, $\beta = 1$ leaves them unchanged, and $\beta > 1$ concentrates them. Restriction measures diversity within selected modes; an unrescaled projection retains the fraction W_I of source energy, giving EW_I, whereas a fixed-energy power transform preserves E. Identical normalized spectra need not imply identical pooling weights. Compositions must track energies and retained directions; Law 3 applies only when the retained source directions are orthogonal. Matrix realizations, examples, and fixed-subset composition are in Appendix A.

3 What General Merging Requires

Predicting pooled diversity requires knowing whether sources contribute distinct directions or repeat one another. We first identify sufficient information for exact prediction, then ask whether simpler overlap summaries can replace it. Their failure motivates a target-specific reduction: second-order rank needs only one energy-weighted overlap total. Section 4 then asks when partial measurements suffice to control its error.

3.1 From Source Spectra to Mode-Resolved Coupling

To combine source strengths with cross-source directional overlap, define

$$C = V_A^\top V_B, \qquad M = \Sigma_A C \Sigma_B, \qquad K = \begin{bmatrix}\Sigma_A^2 & M\\ M^\top & \Sigma_B^2\end{bmatrix}. \tag{6}$$

$C_{ij} = (v_i^A)^\top v_j^B$ is the cosine between unit source directions: zero means orthogonal; absolute value one means the same axis. The coupling C records overlap, while $M_{ij} = \sigma_i^A \sigma_j^B C_{ij}$ weights it by source strengths.

Proposition 2 (Sufficient state via a classical Gram identity)**.** *The positive eigenvalues of K are the squared positive singular values of $[H_A; H_B]$, and $\mathrm{tr}\, K = E_A + E_B$. Thus (Σ_A, Σ_B, C) determines every merged r_α.*

Proof. For $W = [V_A\Sigma_A \mid V_B\Sigma_B]$, $H_M^\top H_M = WW^\top$ and $K = W^\top W$ share their nonzero eigenvalues. □

Thus source strengths and their directional correspondence suffice to recover the entire pooled spectrum. This gives an exact reference for evaluating approximations; it does not establish that every

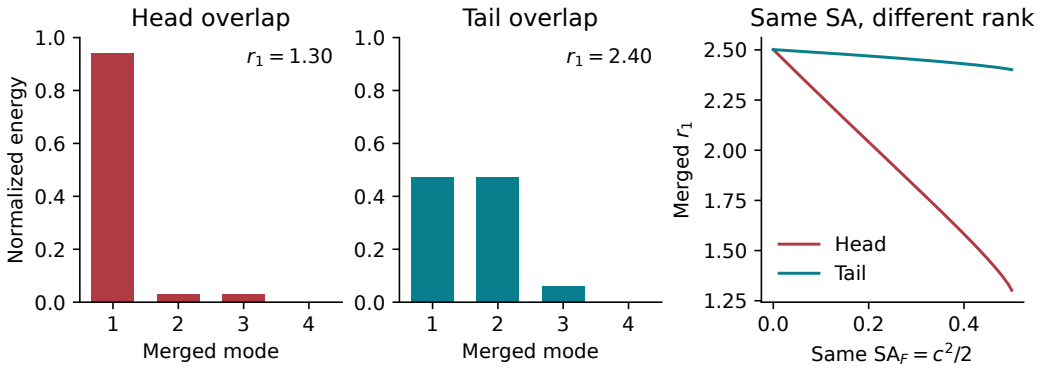

Figure 1: Merged diversity in two synthetic configurations with source singular values $(4, 1)$ and identical principal angles. Overlap joins either the leading or trailing modes. At $c = 1$, the normalized merged energies are $(32, 1, 1, 0)/34$ and $(16, 16, 2, 0)/34$, respectively.

target requires the full coupling. We call the calculation *full Gram*, following classical small-core SVD identities (Zha & Simon, 1999; Brand, 2006). Can a simpler geometric summary retain enough information?

The scalar alignment score is

$$\mathrm{SA}_F = \frac{\|C\|_F^2}{\min(k_A, k_B)} \in [0, 1]. \tag{7}$$

The singular values of C are principal-angle cosines (Björck & Golub, 1973). They describe subspace overlap, but not whether it occurs between strong or weak source modes. The next result shows why that missing association matters.

Proposition 3 (Even principal angles do not identify the merged spectrum)**.** *Sources can have identical singular values, energies, and principal angles, but different merged spectra and Shannon effective ranks.*

The construction and proof appear in Appendix C.2. Figure 1 compares two sources, each with singular values $(4, 1)$ and hence mode energies $(16, 1)$. Here c is the cosine between the overlapping directions; the other direction pair is orthogonal. At $c = 1$, overlapping strong directions combine their energies into 32, leaving $(32, 1, 1)$ (left); overlapping weak directions instead gives $(16, 16, 2)$ (middle). Bars show energy divided by the common total 34; the fourth mode has zero energy. Both matrices have rank three, but energy concentration gives Shannon ranks 1.30 and 2.40. The right panel varies overlap from $c = 0$ to $c = 1$: at each shared horizontal coordinate $\mathrm{SA}_F = c^2/2$, strong-mode overlap reduces diversity more. Thus equal subspace overlap does not identify pooled diversity. For isotropic sources, principal angles suffice (Appendix C.3).

This nonuniqueness has a quantitative consequence: a predictor receiving these summaries cannot distinguish the two cases, so it cannot be accurate on both.

Corollary 4 (An unavoidable error for indistinguishable summaries)**.** *Any deterministic predictor of merged Shannon rank that uses only source spectra, energies, and principal angles has worst-case relative error at least* 29.66% *on the two* $c = 1$ *configurations in Proposition 3.*

The construction and proof appear in Appendix C.2. Identical retained inputs force a predictor to give the same answer to two different targets. The error therefore reflects lost information, not a poorly fitted predictor. The bound also covers SAF-only and source-only predictors, but is a worst-case statement for this construction, not a universal error rate on real data.

3.2 The Orthogonal Reference and Spectral Redundancy Analysis

The preceding results rule out generally exact prediction from geometric summaries alone. We now ask what can still be inferred, and whether a particular effective-rank order needs less information. Let $K_0 = \mathrm{diag}(\Sigma_A^2, \Sigma_B^2)$ and $p_\perp$ be its normalized eigenvalues. This orthogonal reference treats

Table 1: Sufficiency of retained information for a general merge. Source state includes spectra and energies. "No" means insufficient in general; isotropic sources are a special case.

Retained state	Exact r_2	Entire r_α family
Source state only	No	No
Source state and SA_F	No	No
Source state and all principal angles	No	No
Source state and $\|\Sigma_A C \Sigma_B\|_F^2$	Yes	No, in general
Source state and full C	Yes	Yes

source directions as independent. Ignoring their overlap can overstate diversity; the next result bounds that overstatement and corrects it exactly at order two.

Proposition 5 (Upper envelope and exact order-two correction)**.** *For every feasible C and finite $\alpha > 0$,*

$$r_\alpha(p_M) \le r_\alpha(p_\perp). \tag{8}$$

With $E = E_A + E_B$ and $Q = \operatorname{tr}\Sigma_A^4 + \operatorname{tr}\Sigma_B^4$,

$$r_2(p_M) = \frac{E^2}{Q + 2\|M\|_F^2}, \qquad \frac{r_2(p_\perp) - r_2(p_M)}{r_2(p_M)} = \frac{2\|M\|_F^2}{Q}. \tag{9}$$

No small-overlap or positive merged-eigenvalue-floor assumption is required.

Proof outline. The eigenvalues of K majorize its diagonal. Schur concavity of effective rank gives the upper envelope. Expanding $\operatorname{tr} K^2 = Q + 2\operatorname{tr}(MM^\top)$ gives the correction (Appendix C). □

The first conclusion makes the orthogonal calculation an upper bound on diversity. The second changes the information requirement: r_2 depends on total energy and its concentration, so source spectra plus the weighted overlap total suffice, without reconstructing the entire merged spectrum. Order two means $\alpha = 2$, not a Taylor approximation. The identity is equivalently $\|M\|_F^2 = \operatorname{tr}(G_A G_B)$ for $G_X = H_X^\top H_X$. To locate which mode pairs contribute to the total, define

$$D_{ij} = \frac{2(\sigma_i^A)^2(\sigma_j^B)^2 C_{ij}^2}{Q}, \qquad \sum_{ij} D_{ij} = \frac{r_2(p_\perp) - r_2(p_M)}{r_2(p_M)}. \tag{10}$$

Squared overlap in Eq. (6) treats opposite orientations along the same axis equally. We call this *spectral redundancy analysis*. Each D_{ij} attributes the relative order-two deficit to one energy-weighted mode pair; row and column sums attribute it to source modes. Equal squared overlap contributes 256 times more between energy-16 modes than between energy-1 modes. Table 1, row four, retains only the total weighted overlap: $\|M\|_F^2 = (Q/2)\sum_{ij} D_{ij}$. Equations (9)–(10) explain why it suffices for r_2; individual terms additionally locate the overlap. These are second-moment attributions, not semantic redundancy or deletion criteria. Within repeated-eigenvalue blocks, individual attributions depend on the basis, but block aggregates and the total remain invariant.

One weighted scalar therefore suffices for r_2, but not the whole curve. For the same $(4, 1)$ sources, $C = \operatorname{diag}(.05, 0)$ and $C' = \operatorname{diag}(0, .8)$ both give $\|M\|_F^2 = .64$, yet spectra $\{16.8, 15.2, 1, 1\}$ and $\{16, 16, 1.8, .2\}$. Table 1 summarizes the distinction.

The same correction composes with the first two laws: powers change energy weights, subsets select mode pairs, and pooling adds their surviving interaction. Dense evaluation from cached C costs $O(|I_A||I_B|)$ for retained subsets. Subsequent merges need output directions, recoverable through W (Appendix C). We next ask when a terminal order-two query can avoid reading the complete coupling.

4 Error Guarantees from Partial Coupling

We study restricted observations: estimators use only source spectra, energies, and queried coupling. In the cached-factor implementation, each query computes a directional inner product. Information

insufficiency concerns these observations; additional queries or direct computation remain available. Write $a_i = (\sigma_i^A)^2$, $b_j = (\sigma_j^B)^2$, and $L = \sum_{ij} a_i b_j C_{ij}^2$. The queried region is a rectangle $I \times J$ of C. Let L_{IJ} be its weighted sum and $R(x, y) = \sum_\ell x_\ell^\downarrow y_\ell^\downarrow$, padding the shorter vector with zeros. Since C is a contraction, $u_i = 1 - \sum_{j \in J} C_{ij}^2$ and $v_j = 1 - \sum_{i \in I} C_{ij}^2$ bound the squared overlap still available in each queried row or column. These capacities and mode energies limit omitted contributions.

Proposition 6 (A sufficient partial-coupling certificate)**.** *With complements taken within source supports and empty maxima set to zero, define*

$$U_r = \max(b_{J^c}) \sum_{i \in I} a_i u_i + R(a_{I^c}, b), \qquad U_c = \max(a_{I^c}) \sum_{j \in J} b_j v_j + R(a, b_{J^c}),$$

$$U = \min\{U_r, U_c, R(a, b) - L_{IJ}\}, \qquad \widetilde{r}_2 = \frac{E^2}{Q + 2L_{IJ}}. \tag{11}$$

Then $0 \le L - L_{IJ} \le U$ *and*

$$0 \le \frac{\widetilde{r}_2}{r_2} - 1 = \frac{2(L - L_{IJ})}{Q + 2L_{IJ}} \le \frac{2U}{Q + 2L_{IJ}}. \tag{12}$$

Using only measured overlap underestimates energy concentration and therefore overestimates rank. Proposition 6 bounds this overestimate using observable quantities alone. Its row, column, and global bounds constrain the same omitted contribution; taking their minimum retains the tightest guarantee. Querying nested leading rectangles can stop once the relative-error bound is at most ε. Failure to meet the tolerance does not imply inaccurate prediction, and the resulting budget is sufficient for this schedule, not a minimum requirement. Appendix D gives the proof and a Shannon certificate.

When the bound remains large, is more information necessary, or is the guarantee merely loose? To separate these possibilities, fix the observed block $D = C_{I,J}$ and let $S(D)$ be the largest omitted weighted interaction over compatible contractions. When $d \ge k_A + k_B$, these are realizable by source subspaces; zero-filling attains zero omitted interaction. Constructed completions give a lower bound $W(D)$ on $S(D)$; a transport relaxation retaining residual row and column capacities gives an upper bound $U_{\mathrm{tr}}(D)$. With $U' = \min(U, U_{\mathrm{tr}})$,

$$W(D) \le S(D) \le U' \le U, \qquad U - U' \le U - S(D) \le U - W(D). \tag{13}$$

The first interval brackets remaining ambiguity; the second brackets bound conservatism. Tightening can settle a diversity-gain decision without new measurements. Conversely, compatible completions with opposite decisions prove that the observations cannot support a uniform decision guarantee. Some states remain unresolved by either test. Neither the actual pair's error nor a sampled maximum identifies $S(D)$; Appendix D.2 gives constructions and dual checks.

5 Numerical Validation and Computational Evaluation

We first examine how coupling affects pooled spectral diversity and when partial measurements suffice to guarantee prediction accuracy. When the resulting bounds cannot certify the required accuracy or determine a diversity-gain decision, we investigate whether tighter bounds can resolve the uncertainty or compatible completions can establish information insufficiency. We then assess whether measurement savings translate into runtime gains after accounting for source preparation and bound-tightening costs. Average errors do not estimate the theory's worst-case lower bounds.

5.1 Why Coupling Matters for Pooled Diversity

We first compare exact reconstruction with predictions that omit coupling. Direct SVD supplies targets for 672 *within-encoder* pairs from 192 matrices, covering 24 encoders and eight datasets. Sources have 800 rows and compact ranks 192–800, including 14 unequal-rank pairs (Appendix E). Float64 implementation checks give maximum relative error 1.51×10^{-14} for power and subset identities; general-Gram and order-two identities also agree at numerical precision on unequal-rank synthetic pairs (Appendix E.2).

Table 2: **Effect of directional overlap on merged effective-rank accuracy.** Across 672 pairs, direct SVD supplies exact targets. Orthogonal reference ignores overlap; source sum adds source effective ranks; full Gram includes overlap. α: rank order; ρ: Spearman correlation (Gram: 1). Relative errors: mean $\pm$ SD in percent for orthogonal reference; unscaled mean and maximum for Gram.

	Orthogonal reference		Source sum	Full Gram rel. error	
α	ρ	Rel. error (%)	ρ	Mean	Maximum
0.5	0.976	37.5 ± 17.3	0.977	4.02×10^{-9}	9.25×10^{-8}
1	0.990	25.5 ± 14.8	0.966	1.09×10^{-14}	2.98×10^{-13}
2	0.975	24.7 ± 23.8	0.912	1.95×10^{-15}	1.84×10^{-14}

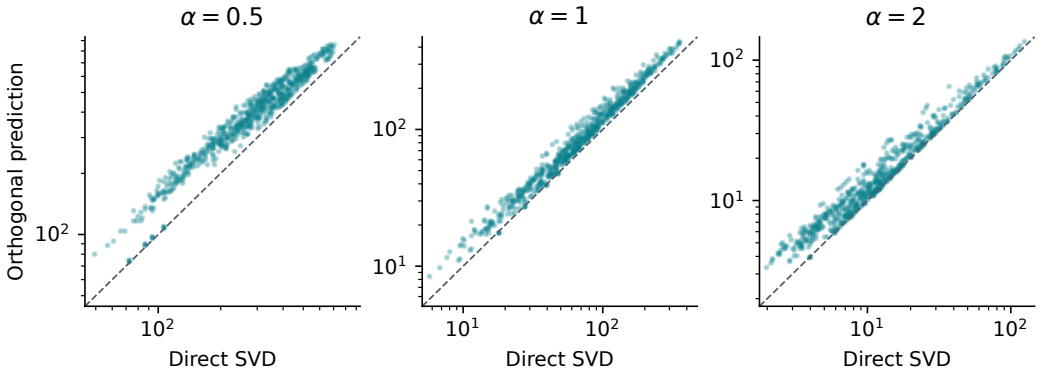

Figure 2: Predicted and directly measured diversity for archived full-rank merges, shown on logarithmic axes. The diagonal marks exact agreement. The orthogonal reference and full Gram calculation use the same source factors.

Table 2 and Figure 2 show that the orthogonal reference ranks pairs well but systematically overestimates diversity. Shared directions concentrate energy that this reference assigns to separate modes, as predicted by Proposition 5. Full Gram restores agreement; its largest numerical error occurs at the lowest tested order, which emphasizes tiny eigenvalues. A separate, traceable extraction on two encoders and four datasets reproduces the contrast: 41.4% mean Shannon-rank error for the orthogonal reference and near-machine-precision Gram agreement (Appendix E.3).

The preceding comparison shows that omitting coupling affects prediction accuracy. To identify which coupling information matters, we construct new matrices by permuting the directions associated with fixed singular values in 168 measured pairs from six encoders. These permutations preserve source spectra, energies, and principal angles while changing the association between energy and direction. For each pair, we measure the range of Shannon effective rank across the tested assignments relative to the original rank; the median across pairs is 9.73%. This demonstrates the ambiguity in Proposition 3 using measured source factors. Full coupling resolves this ambiguity by retaining which source modes overlap and, together with their spectra, how much energy those modes carry. We next examine whether partial coupling measurements suffice for reliable prediction of second-order effective rank.

5.2 When Partial Measurements Suffice

A pair comprises two sample-block feature matrices from the same encoder. Each encoder/partition supplies two 200-row blocks per dataset across four datasets: eight sources and 28 pairs. Four encoders and five partitions yield 560 pairs, with 1,680 raw, source-centered, or unit-row evaluations sharing sources.

We fix the order-two relative-error tolerance at 1% before querying. Given source SVDs, we order directions by energy and query leading-mode inner products, expanding rectangles along a preset ladder while reusing entries. Queries concern direction pairs, not new samples or labels. Measured

Table 3: Entry budgets for 1% relative error in second-order rank: median percentages across each row's pairs at first actual-error or error-bound stopping on the refined ladder. Actual-error stopping uses exact targets retrospectively. Conditions share sources.

Group, preprocessing	Pairs	Actual-error stopping budget (%)	Error-bound stopping budget (%)
Four encoders, raw	560	0.062	1.000
Four encoders, centered	560	16.161	64.645
Supervised, raw	30	0.062	1.000
Supervised, centered	30	16.161	64.645
DINO, raw	30	0.062	4.000
DINO, centered	30	16.161	98.998

overlap gives the rank estimate; bounding unmeasured contributions supplies the guarantee. We stop when the bound is at most 1%, without consulting exact targets. Entry budget multiplies the retained-mode fractions. On the original ladder, raw and unit-row median budgets are 1%; the centered median reaches full coupling. Every certified stop meets the tolerance.

The same policy reproduces the raw/centered contrast externally with supervised and DINO ResNet-50 on 2,000 Flowers-102 and 2,000 Oxford-IIIT Pets images each: five partitions give 30 pairs per condition, 180 total (Appendix E.4, Table 9). Query location also matters: at a fixed 1% entry budget on raw cohort features, leading-mode median error is 0.0833%, versus 62.204% for random and 63.076% for trailing queries.

Mean controls probe preprocessing. On 560 raw pairs, the median mean-related overlap share is 99.31%; mean attenuation increases certification budgets but changes spectra, directions, and normalization together. Both external models exceed 99.7% median mean share yet have different budgets, so mean share alone is insufficient. Controlled tests find flat spectra and balanced energies harder to certify; all 98,496 checks across 3,456 cases hold within tolerance (Appendix E.4).

Sufficient budgets need not be necessary. Table 3 compares actual-error and error-bound stopping budgets on a *refined* query ladder, avoiding the original ladder's coarse jump to full coupling. For centered DINO, the median budgets are 16.161% and 98.998%, respectively, expressed as percentages of queried entries. Actual-error stopping uses the exact target for retrospective evaluation. Actual error remains below its certified bound at every tested observation state, yet it can meet the tolerance well before the bound certifies it. Thus, the tested certificate can require substantially more measurements than suffice for accuracy on the actual pair. This gap alone does not establish bound conservatism: unobserved couplings consistent with the same measurements may yield larger errors. We next examine these two sources of uncertainty.

5.3 Separating Bound Conservatism from Missing Information

We fix spectra and queried entries, varying only compatible unobserved coupling. On 120 raw/centered external pairs, querying $.4$ of modes per side measures approximately 16% of entries. Centered DINO medians are 0.434% actual error, 7.043% completion-witness error, and 7.213% transport upper bound. Accuracy on the actual pair thus need not guarantee accuracy on all compatible pairs. The median minimax lower bound is 3.402%; each bound concerns deterministic estimation at fixed observations, not query complexity across adaptive policies (Appendix D.2, Table 5).

Spectral-diversity batch admission assesses candidates B separately against a fixed feature bank A. Gain intervals entirely at or above the threshold certify admission; those entirely below certify rejection, while straddling intervals require further analysis. A new label-free partition, distinct from the error experiment, gives 240 pairs. We test $g = r_2([H_A; H_B])/r_2(H_A) - 1 \geq .05$. At per-side fraction $.4$, Table 4 partitions 80 centered decisions: 29 originally certified, 16 additionally resolved by tightening, 26 with opposite-decision completions, and nine unresolved. Raw and unit-row pairs fall below the threshold (Appendix F.2). This tests spectral admission, not downstream training value.

Table 4: Fixed-observation decisions at 5% spectral gain (80 pairs/row). Per-side fraction $.4$ queries about 16% of entries. Counts: original certification, additional transport resolution, opposite-decision completions, unresolved. Ambiguity does not denote actual errors.

Preprocessing, fraction	Original	Tightened	Ambiguous	Unresolved
Raw, 0.1	80	0	0	0
Centered, 0.1	3	3	64	10
Centered, 0.4	29	16	26	9
Centered, 0.8	60	13	3	4
Unit-row, 0.1	80	0	0	0

At fixed observations, tighter bounds can resolve additional admission decisions without new measurements, whereas compatible completions yielding opposite decisions establish that the available information cannot support a uniform decision guarantee. These results distinguish cases where improving the bound suffices from those requiring additional information, while leaving some cases unresolved.

5.4 Computational Benefits and Limits of Guarantees

Exact order-two rank uses $L = \|H_A H_B^\top\|_F^2$ and analogous products for Q, avoiding source SVD (Appendix F.1). Cached raw or unit-row factors make certified policies faster than this baseline, but charging preparation reverses the aggregate comparison. On centered cached inputs, median times per four-candidate bank are 7.30 ms for the original certificate, 4004.55 ms with online transport, and 2.30 ms for direct products. Partial queries may thus suit batch assessments reusing factors; one-off merges or expensive tightening may favor direct computation.

Decisions certified by gain intervals agree with exact decisions; 48-probe Hutch++ (Meyer et al., 2021) has seven disagreements in 864 recorded, dependent decisions (Appendix F.2).

In negative controls, at equal exact-call budgets, the 2% source-gap rule has higher Shannon-rank regret than random allocation (Appendix G.1, Table 11). Train-only pruning by retained $r_{.5}$ keeps more dimensions without improving mean ridge-probe accuracy over the energy baseline (Appendix G.2). Spectral guarantees alone establish neither allocation benefits nor label relevance.

6 Discussion and Limitations

Sufficient information depends on the target: order-two rank needs weighted overlap, whereas broader targets may need full coupling. Weighted overlaps also expose repeated high-energy directions for redundancy diagnosis. Candidate screening could allocate queries through gain intervals, but cost benefits remain untested. Stream monitoring additionally requires updating merged spectra and directions, beyond our fixed-base experiments.

Evidence concerns fixed vision encoders with shared sources. Archive metadata remain incomplete; timings cover tested shapes and hardware, and pretraining comparisons vary multiple factors. Spectral diversity describes geometry, not label relevance. Task-aware extensions need additional links, such as Neural Collapse (Papyan et al., 2020) under its specific training conditions.

7 Conclusion

We study the information needed to predict pooled spectral diversity. Source spectra and principal angles can miss which energetic directions overlap. Given source spectra and energies, one weighted overlap total determines second-order effective rank. Bounding unmeasured overlap guarantees prediction accuracy, while compatible completions distinguish conservative bounds from information insufficiency. Vision-feature experiments validate these guarantees and show that tightening resolves additional diversity-gain decisions without new measurements. The results connect information requirements to reliable prediction and batch admission based on spectral diversity.

AI USE STATEMENT

Generative AI assisted with conceptual framing, mathematical derivations and proofs, interpretation of experimental results, preparation and checking of experiment and analysis code, figures, literature lookup, and manuscript drafting and editing. Real-feature measurements originate from the identified archive and completed extraction and analysis runs. Manuscript builders recompute summaries and identified synthetic checks, not raw-feature extraction. Counterfactual matrices are not represented as natural observations. Authors remain responsible for claims, provenance, and the final submission.

REPRODUCIBILITY STATEMENT

The anonymous supplementary package provides one entry point with three levels. Records mode rebuilds manuscript evidence from saved measurements. Cached mode recomputes the pilot, budget, coupling, mean-control, stochastic-baseline, external, stress, and fixed-observation completion experiments from supplied features or new synthetic constructions. Raw mode downloads official data and weights, extracts the three traceable collections, and runs their analyses; cohort and external extraction require CUDA. Inputs are fingerprinted, outputs use new directories, and each stage reports execution and verification separately. Separate clean-directory runs validated the earlier cached suite and its completion extension; a small public-data extraction was also tested, not another full GPU extraction. Images and weights are not redistributed. The inherited extraction remains unreconstructable from available metadata. Appendices E and H document coverage. The fixed-base decision extension uses the same three levels through `-suite admission`. Its full server run and independent verification are complete; packaging checks and any subsequent cached rerun are recorded separately.

REFERENCES

Kumar Krishna Agrawal, Arnab Kumar Mondal, Arna Ghosh, and Blake A. Richards. α-ReQ: Assessing representation quality in self-supervised learning by measuring eigenspectrum decay. In *Advances in Neural Information Processing Systems*, volume 35, pp. 17626–17638, 2022. doi: 10.52202/068431-1281. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/70596d70542c51c8d9b4e423f4bf2736-Abstract-Conference.html.

Koenraad M. R. Audenaert. A sharp continuity estimate for the von Neumann entropy. *Journal of Physics A: Mathematical and Theoretical*, 40(28):8127–8136, 2007. doi: 10.1088/1751-8113/40/28/S18. URL https://doi.org/10.1088/1751-8113/40/28/S18.

Jean-François Bercher. A simple probabilistic construction yielding generalized entropies and divergences, escort distributions and q-Gaussians. *Physica A: Statistical Mechanics and its Applications*, 391(19):4460–4469, 2012. doi: 10.1016/j.physa.2012.04.024. URL https://arxiv.org/abs/1206.0561.

Jeff A. Bilmes, Gantavya Bhatt, and Arnav M. Das. How much is a dataset worth? scaling laws, the Vendi score, and matrix spectral functions. *arXiv preprint arXiv:2605.29448*, 2026. URL https://arxiv.org/abs/2605.29448.

Åke Björck and Gene H. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–594, 1973. doi: 10.1090/S0025-5718-1973-0348991-3.

Matthew Brand. Fast low-rank modifications of the thin singular value decomposition. *Linear Algebra and its Applications*, 415(1):20–30, 2006. doi: 10.1016/j.laa.2005.07.021. URL https://www.merl.com/publications/TR2006-059.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9650–9660, 2021. URL https://openaccess.thecvf.com/content/ICCV2021/html/Caron_Emerging_Properties_in_Self-Supervised_Vision_Transformers_ICCV_2021_paper.html.

Dan Friedman and Adji Bousso Dieng. The Vendi Score: A diversity evaluation metric for machine learning. *Transactions on Machine Learning Research*, 2023. URL https://openreview.net/forum?id=g97OHbQyk1.

Antonio Andrea Gargiulo, Donato Crisostomi, Maria Sofia Bucarelli, Simone Scardapane, Fabrizio Silvestri, and Emanuele Rodolà. Task singular vectors: Reducing task interference in model merging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18695–18705, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Gargiulo_Task_Singular_Vectors_Reducing_Task_Interference_in_Model_Merging_CVPR_2025_paper.html.

Quentin Garrido, Randall Balestriero, Laurent Najman, and Yann LeCun. RankMe: Assessing the downstream performance of pretrained self-supervised representations by their rank. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 10929–10974, 2023. URL https://proceedings.mlr.press/v202/garrido23a.html.

M. O. Hill. Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54(2): 427–432, 1973. doi: 10.2307/1934352.

Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2nd edition, 2013. ISBN 9780521839402. doi: 10.1017/CBO9781139020411.

Yu-Ang Lee, Ching-Yun Ko, Tejaswini Pedapati, I-Hsin Chung, Mi-Yen Yeh, and Pin-Yu Chen. STAR: Spectral truncation and rescale for model merging. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pp. 496–505, 2025. doi: 10.18653/v1/2025.naacl-short.42. URL https://aclanthology.org/2025.naacl-short.42/.

Daniel Marczak, Simone Magistri, Sebastian Cygert, Bartłomiej Twardowski, Andrew D. Bagdanov, and Joost van de Weijer. No task left behind: Isotropic model merging with common and task-specific subspaces. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 43177–43199, 2025. URL https://proceedings.mlr.press/v267/marczak25a.html.

Raphael A. Meyer, Cameron Musco, Christopher Musco, and David P. Woodruff. Hutch++: Optimal stochastic trace estimation. In *Symposium on Simplicity in Algorithms (SOSA)*, pp. 142–155, 2021. doi: 10.1137/1.9781611976496.16. URL https://arxiv.org/abs/2010.09649.

Quan Nguyen and Adji Bousso Dieng. Quality-weighted Vendi scores and their application to diverse experimental design. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 37667–37682, 2024. URL https://proceedings.mlr.press/v235/nguyen24d.html.

Vardan Papyan, X.Y. Han, and David L. Donoho. Prevalence of neural collapse during the terminal phase of deep learning training. *Proceedings of the National Academy of Sciences*, 117(40): 24652–24663, 2020. doi: 10.1073/pnas.2015509117.

Amey P. Pasarkar and Adji Bousso Dieng. Vendi novelty scores for out-of-distribution detection. *arXiv preprint arXiv:2602.10062*, 2026. URL https://arxiv.org/abs/2602.10062.

Alfréd Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pp. 547–561. University of California Press, 1961.

P. L. Robinson. Julia operators and Halmos dilations. *arXiv preprint arXiv:1803.09329*, 2018. URL https://arxiv.org/abs/1803.09329.

Olivier Roy and Martin Vetterli. The effective rank: A measure of effective dimensionality. In *Proceedings of the 15th European Signal Processing Conference*, pp. 606–610, 2007. doi: 10.5281/zenodo.40328. URL https://doi.org/10.5281/zenodo.40328.

Hongyuan Zha and Horst D. Simon. On updating problems in latent semantic indexing. *SIAM Journal on Scientific Computing*, 21(2):782–791, 1999. doi: 10.1137/S1064827597329266.

Xiaohua Zhai, Joan Puigcerver, Alexander Kolesnikov, Pierre Ruyssen, Carlos Riquelme, Mario Lucic, Josip Djolonga, André Susano Pinto, Maxim Neumann, Alexey Dosovitskiy, Lucas Beyer, Olivier Bachem, Michael Tschannen, Marcin Michalski, Olivier Bousquet, Sylvain Gelly, and Neil Houlsby. A large-scale study of representation learning with the visual task adaptation benchmark. *arXiv preprint arXiv:1910.04867*, 2019. URL https://arxiv.org/abs/1910.04867.

A Feature Modes and Operational Examples

Quality-weighted Vendi scores (Nguyen & Dieng, 2024) combine sample quality with similarity-based diversity for experimental design. Their quality weights encode application utility; our mode weights describe feature energy. In model merging, TSV-Merge (Gargiulo et al., 2025) compresses task matrices and decorrelates their singular vectors, while isotropic merging (Marczak et al., 2025) flattens singular values and combines common and task-specific subspaces. These methods motivate studying spectral operations, but operate on parameter updates. Pooling stacks feature rows, so its second-moment matrix is $H_A^\top H_A + H_B^\top H_B$. In contrast, adding parameter-update matrices produces cross terms in the second moment. Thus related spectral operations in STAR do not make parameter averaging equivalent to the pooling studied here.

Let $Z_\beta = S_\beta(p) = \sum_i p_i^\beta$. One fixed-energy realization of Law 1 is

$$H_\beta = U \operatorname{diag}\big(\sqrt{Eq_1}, \ldots, \sqrt{Eq_k}\big) V^\top, \qquad q_i = \frac{p_i^\beta}{Z_\beta}, \qquad \|H_\beta\|_F^2 = E. \tag{14}$$

The new singular values are proportional to σ_i^β. As $\beta \to 0^+$, positive modes approach equal energy without increasing algebraic rank. Weak modes may contain signal or noise; their amplification does not guarantee task benefit. Unlike the observable order α, the exponent β changes the representation.

For a nonempty subset I, let V_I contain the retained right singular vectors. Then

$$H_I = H V_I V_I^\top = U_I \Sigma_I V_I^\top, \qquad Z_I = H V_I = U_I \Sigma_I, \qquad \|H_I\|_F^2 = E W_I. \tag{15}$$

H_I keeps ambient coordinates; Z_I stores the same positive singular values in $|I|$ coordinates. Conditioning normalizes probabilities, not the physical matrix energy. In $p = (.6, .3, .1)$, reweighting with $\beta = 2$ gives $(36, 9, 1)/46$. Retaining $I = \{1, 2\}$ instead gives $(2/3, 1/3)$ and 90% energy; $I = \{2, 3\}$ gives $(3/4, 1/4)$ and 40% energy. These illustrative spectra show why retained mass and retained diversity answer different questions. Independently compressed local coordinates cannot be stacked as though their axes coincided: retain ambient projections or their relative basis alignment. A subset within a repeated-singular-value block also requires a specified basis.

Power transforms act on spectral energy, not prediction logits. Subsets may contain leading, trailing, or noncontiguous modes. For example, an unrescaled projection retaining 80% of energy leaves total energy $0.8E$. Power reweighting and restriction commute after normalization for fixed indices; data-dependent selection may change those indices. Source interaction $C = V_A^\top V_B$ becomes $C_{I,J}$ after restriction, and the orthogonal merge formula applies only when this submatrix vanishes.

For a composed example, apply a fixed-energy power transform with exponent β_X to source $X \in \{A, B\}$, then project onto I_X without rescaling. Retained energies and masses are

$$e_i^X = \frac{E_X (p_i^X)^{\beta_X}}{S_{\beta_X}(p^X)}, \qquad \widetilde{E}_X = \sum_{i \in I_X} e_i^X, \qquad \widetilde{p}_i^X = \frac{e_i^X}{\widetilde{E}_X} \quad (i \in I_X). \tag{16}$$

The merge weight is $\widetilde{\gamma} = \widetilde{E}_B / (\widetilde{E}_A + \widetilde{E}_B)$. Rescaling a compressed source back to its original energy changes this weight without changing its normalized spectrum. For general coupling, define

$$\widetilde{Q} = \sum_{i \in I_A} (e_i^A)^2 + \sum_{j \in I_B} (e_j^B)^2, \qquad \widetilde{L} = \sum_{\substack{i \in I_A \\ j \in I_B}} e_i^A e_j^B C_{ij}^2, \qquad r_2(\widetilde{p}_M) = \frac{(\widetilde{E}_A + \widetilde{E}_B)^2}{\widetilde{Q} + 2\widetilde{L}}. \tag{17}$$

Substitution into the trace identity shows that powers change energy weights, restriction changes retained mode pairs, and pooling adds their surviving interaction. Removing a tail with little weighted overlap leaves most of the correction intact; downweighting strongly overlapping leading modes can reduce it substantially. This distinguishes retained diversity from retained energy and explains why later operations need both.

Cached source factors permit repeated powers and fixed subsets without reconstructing feature matrices. Spectral sums still have to be computed before the constant-size scalar formulas apply. A terminal diversity query uses eigenvalues alone, whereas a subsequent merge also needs output directions.

B Exact Laws and Composition

All probability vectors below are restricted to positive support before logarithms are evaluated. Zeros may be reinserted without changing any r_α with $\alpha > 0$. Define $\Phi_\nu(p) = \log \sum_i p_i^\nu$ for $\nu > 0$. For $\nu \neq 1$,

$$\Phi_\nu(p) = (1-\nu)\log r_\nu(p). \tag{18}$$

At $\nu = 1$, $\Phi_1 = 0$. Working with Φ avoids ambiguous products such as $0 \log r_1$ in parameter limits.

B.1 Power Reweighting, Including the Shannon Limit

For $q_i = p_i^\beta / Z_\beta$ and any $\alpha > 0$,

$$\sum_i q_i^\alpha = Z_\beta^{-\alpha} \sum_i p_i^{\alpha\beta}, \tag{19}$$

$$\Phi_\alpha(q) = \Phi_{\alpha\beta}(p) - \alpha\Phi_\beta(p). \tag{20}$$

Division by $1-\alpha$ proves Law 1 for $\alpha \neq 1$. At $\alpha = 1$, direct substitution gives

$$\log r_1(q) = \log Z_\beta - \beta \sum_i q_i \log p_i. \tag{21}$$

This also follows by differentiating $\Phi_{\alpha\beta} - \alpha\Phi_\beta$ with respect to α and taking the limit at one. In particular $\beta = 1$ returns p, and $\tau_{\beta_2}\tau_{\beta_1} = \tau_{\beta_1\beta_2}$. For positive β on a fixed positive support, the inverse is $\tau_{1/\beta}$. Invertibility of this family does not classify every possible support-preserving spectral map.

One matrix realization with deliberately fixed energy is

$$H_\beta = \sqrt{E}\, U \operatorname{diag}(\sqrt{q_i}) V^\top. \tag{22}$$

Another is $U \operatorname{diag}(\sigma_i^\beta) V^\top$, whose energy generally changes. Both have the same normalized masses, but their later merges differ unless energy rescaling is accounted for. A formula for normalized shape must not silently specify absolute scale.

B.2 Arbitrary-Subset Conditioning

For a nonempty $I \subseteq \operatorname{supp}(p)$, $W_I > 0$, and

$$\sum_{i\in I} (p_i/W_I)^\alpha = W_I^{-\alpha} \sum_{i\in I} p_i^\alpha. \tag{23}$$

This proves Law 2 without an ordering assumption on I. At $\alpha = 1$,

$$\log r_1(P_I p) = \log W_I - \frac{1}{W_I} \sum_{i\in I} p_i \log p_i. \tag{24}$$

For a singleton the rank is one, and for the full support the original rank is recovered. Leading-k_0 PCA, a retained tail, and arbitrary indexed spectral masks are all special cases. A mask that is empty or has zero retained mass has no conditional distribution and is outside the statement. Selection by coordinate features rather than singular modes does not satisfy this law in general.

For a fixed set of original mode indices, restriction and power reweighting commute:

$$(P_I \tau_\beta p)_i = \frac{p_i^\beta}{\sum_{j\in I} p_j^\beta} = (\tau_\beta P_I p)_i \qquad (i \in I). \tag{25}$$

For $\alpha > 0$, $\alpha \neq 1$, the composed log-rank is

$$\log r_\alpha(P_I \tau_\beta p) = \frac{\log \sum_{i\in I} p_i^{\alpha\beta} - \alpha \log \sum_{i\in I} p_i^\beta}{1-\alpha}. \tag{26}$$

At $\alpha = 1$, take the continuous limit, giving $\log Z_{\beta,I} - \beta \sum_{i\in I} q_i \log p_i$, where $Z_{\beta,I} = \sum_{i\in I} p_i^\beta$ and $q_i = p_i^\beta / Z_{\beta,I}$. This does not imply that two *data-dependent selection rules* commute: an energy threshold may select a different index set after reweighting.

B.3 Orthogonal Union and Energy Weights

When $V_A^\top V_B = 0$, the two positive eigenspaces of $H_A^\top H_A$ and $H_B^\top H_B$ are orthogonal. Their sum therefore has the concatenated positive eigenvalues, normalized as $(1-\gamma)p_A$ and γp_B. Taking the power sum proves Law 3. The Shannon limit is

$$\log r_1(p_M) = h(\gamma) + (1-\gamma)\log r_1(p_A) + \gamma \log r_1(p_B), \tag{27}$$

$$h(\gamma) = -(1-\gamma)\log(1-\gamma) - \gamma \log \gamma. \tag{28}$$

At equal energies this becomes $r_1(p_M) = 2\sqrt{r_1(p_A) r_1(p_B)}$. Equal row counts alone do not imply equal energies, although a particular feature-normalization protocol can make them approximately equal. For repeated weighted unions, the physical energies must determine branch weights consistently.

C General Merge: Identification and Corrections

C.1 Full Reconstruction and Feasibility

The proof of Proposition 2 uses the thin rectangular matrix $W = [V_A\Sigma_A \mid V_B\Sigma_B]$. Eigenvectors can also be recovered: if $Kz = \lambda z$ with $\lambda > 0$ and $\|z\|_2 = 1$, then $Wz/\sqrt{\lambda}$ is a unit right singular direction of the merged matrix. This may be needed for a subsequent nonorthogonal merge; caching only the new eigenvalues is not sufficient to propagate all future directional interactions.

Any contraction $C \in \mathbb{R}^{k_A \times k_B}$ is realizable in ambient dimension $k_A + k_B$. Let

$$V_A = \begin{bmatrix} I_{k_A} \\ 0 \end{bmatrix}, \qquad V_B = \begin{bmatrix} C \\ (I_{k_B} - C^\top C)^{1/2} \end{bmatrix}. \tag{29}$$

Then $V_A^\top V_A = I$, $V_B^\top V_B = I$, and $V_A^\top V_B = C$. Choosing $H_X = \Sigma_X V_X^\top$ realizes the desired positive singular values. This verifies feasibility of both constructions in Proposition 3, including $c = 1$ where the merged rank decreases.

C.2 Counterexample and Prediction Lower Bound

Proofs of Proposition 3 and Corollary 4.

Proof. Take $\Sigma_A = \Sigma_B = \mathrm{diag}(4, 1)$ and

$$C_h = \mathrm{diag}(c, 0), \qquad C_t = \mathrm{diag}(0, c), \qquad 0 < c \le 1. \tag{30}$$

Both are feasible in $\mathbb{R}^4$ (Appendix C), with identical singular values $(c, 0)$ and $\mathrm{SA}_F = c^2/2$. Their Gram spectra are

$$\lambda_h = \{16(1+c), 16(1-c), 1, 1\}, \quad \lambda_t = \{16, 16, 1+c, 1-c\}. \tag{31}$$

At $c = 1$, common normalization by 34 gives Shannon ranks 1.30 and 2.40. □

Proof. Its retained inputs are identical, so it must return a common value z. Let the two exact ranks be $a = 1.302768\ldots$ and $b = 2.401496\ldots$, with $a < b$. Then

$$\max\left\{\frac{|z-a|}{a}, \frac{|z-b|}{b}\right\} \ge \frac{b-a}{b+a} = 0.296611\ldots. \tag{32}$$

Indeed, if both errors are at most ε, the triangle inequality gives $b - a \le \varepsilon(a+b)$. Equality is attained at $z = 2ab/(a+b)$. □

The equal-angle example yields distinct Shannon ranks for every $0 < c \le 1$, not only at its endpoint. If $b(c) = (1+c)\log(1+c) + (1-c)\log(1-c)$, with the continuous zero convention at one, then

$$\log r_1(p_t) - \log r_1(p_h) = \frac{15}{34} b(c) > 0. \tag{33}$$

Indeed, splitting a block of two equal masses $a/34$ into $a(1 \pm c)/34$ reduces entropy by $(a/34)b(c)$. Here $a = 16$ in the head case and $a = 1$ in the tail case. Since $b(0) = 0$ and $b'(c) = \log((1+c)/(1-c)) > 0$ for $0 < c < 1$, the difference is strict. The principal angles cannot distinguish these cases because they discard which axis with which energy participates in the overlap.

C.3 When Principal Angles Suffice

Suppose $\Sigma_A = aI_{k_A}$ and $\Sigma_B = bI_{k_B}$. An SVD of C block-diagonalizes K into 2×2 blocks with eigenvalues

$$\lambda_{j,\pm} = \frac{a^2 + b^2 \pm \sqrt{(a^2 - b^2)^2 + 4a^2b^2 s_j(C)^2}}{2}. \tag{34}$$

If source ranks differ, the remaining uncoupled eigenvalues are a^2 or b^2. Thus principal angles and source scales suffice in this isotropic case. In the anisotropic case, the singular-vector changes that diagonalize C generally rotate the diagonal source-energy matrices into nondiagonal ones. Dropping those rotations drops information. The counterexample holds anisotropic source spectra fixed precisely to expose this dependence.

C.4 Proof of the Upper Envelope

Diagonalize $K = O\,\mathrm{diag}(\lambda)O^\top$ with O orthogonal. Its diagonal entries obey $\mu_i = \sum_j O_{ij}^2 \lambda_j$, where μ is the diagonal of K_0. The matrix with entries O_{ij}^2 is doubly stochastic, so the normalized diagonal is a mixing of $p_M = \lambda/E$. This is the Schur–Horn relation (Horn & Johnson, 2013). For $0 < \alpha < 1$, concavity of x^α gives $S_\alpha(p_\perp) \geq S_\alpha(p_M)$; for $\alpha > 1$, convexity reverses this inequality but $1/(1-\alpha) < 0$ reverses it again upon converting to effective rank. Concavity of $-x \log x$ treats $\alpha = 1$. This proves (8) even if K has zero eigenvalues. It is an upper bound on diversity at a fixed order, not a comparison of ranking accuracy between different orders.

C.5 Proof and Meaning of the Order-Two Correction

Block multiplication gives

$$\operatorname{tr} K^2 = \operatorname{tr} \Sigma_A^4 + \operatorname{tr} \Sigma_B^4 + 2 \operatorname{tr} MM^\top, \tag{35}$$

$$\|M\|_F^2 = \sum_{i=1}^{k_A} \sum_{j=1}^{k_B} (\sigma_i^A)^2 (\sigma_j^B)^2 C_{ij}^2. \tag{36}$$

Normalize by E^2 and use $r_2 = 1/S_2$. Since $r_2(p_\perp) = E^2/Q$ and $r_2(p_M) = E^2/(Q + 2\|M\|_F^2)$, their relative difference is $2\|M\|_F^2/Q$, as claimed. This is a direct consequence of the Gram identity, not an estimator fitted to the feature archive. At general integer orders, traces of higher powers of K contain more complicated weighted interactions. A single Frobenius norm cannot generally substitute for them.

For unit-row features and a single added sample, this trace expansion specializes to the exact order-two update in Vendi Novelty Scores (Pasarkar & Dieng, 2026, Eq. 5); that work uses a leading-mode approximation for efficient detection. Here we treat arbitrary source energies and multiple added samples, retaining energy-weighted mode interactions and bounding unobserved coupling. Separately, Bilmes et al. (2026, Section 3.3) accelerate spectral objectives for data selection through secular-equation eigenvalue updates. Such updates exploit candidate coordinates in the current eigenbasis; our partial-observation bounds address uncertainty when coupling information is incomplete.

D Partial-Coupling Certificates and Preprocessing Identities

D.1 Proof of the Order-Two Certificate

For orthonormal source factors, $\|C\|_2 \leq 1$. Consequently $X_{ij} = C_{ij}^2$ is nonnegative with every row and column sum at most one. A rectangular doubly substochastic matrix satisfies, for nonnegative vectors x, y,

$$\sum_{ij} x_i y_j X_{ij} \leq R(x, y). \tag{37}$$

One direct argument pads to a square matrix and fills its row/column deficits to obtain a doubly stochastic matrix. Its decomposition into permutation matrices and the rearrangement inequality bound the weighted sum by the similarly sorted product; nonnegative padding cannot decrease the

original sum. This uses classical matrix/majorization tools (Horn & Johnson, 2013). Submatrices remain doubly substochastic.

Split the unmeasured contribution into $I \times J^c$ and $I^c \times \mathrm{supp}(b)$. The first strip is bounded rowwise by $\max(b_{J^c}) \sum_{i \in I} a_i u_i$, because the unknown squared overlap in row i is at most its residual capacity u_i. The second is at most $R(a_{I^c}, b)$ by the preceding inequality. This proves $L - L_{IJ} \le U_r$. Splitting instead into $I^c \times J$ and $\mathrm{supp}(a) \times J^c$ gives U_c. Applying the global rearrangement bound gives $L - L_{IJ} \le R(a,b) - L_{IJ}$. Every contribution is nonnegative, so their minimum is also an upper bound, including empty-query and full-query cases. Substitution into the exact order-two denominator proves Proposition 6.

The implementation clips tiny negative residual capacities caused by rounding and adds 64 machine epsilons times $R(a,b)$ whenever entries remain unknown. The theorem concerns exact arithmetic; numerical checks allow a relative tolerance of 10^{-8}. Each new rectangle queries only new entries and retains previous measurements. Known spectra, selected indices, and the queried submatrix are the only certificate inputs; full targets are reserved for verification. Selecting leading modes is a tested policy, not a requirement of the bound. Within repeated-eigenvalue blocks, coordinate-level budgets can depend on the chosen basis even though the complete interaction is invariant. No query-complexity lower bound is asserted.

D.2 Fixed-Observation Completions and Bound Loss

We now distinguish the largest compatible interaction from a computable upper bound on it. Fix the source energies, the two spectra, and the signed block $D = C_{I,J}$. Principal angles of the full coupling are not fixed in this experiment; this differs from the earlier energy-assignment counterexample. Relabeling indices makes D the leading $m \times n$ block. We assume $d \ge k_A + k_B$, as holds for every measured pair tested here. Any contraction C can then be realized by $V_A = [I; 0]$ and $V_B = [C; (I - C^\top C)^{1/2}]$, with zero padding when needed. The ambient-dimension assumption is essential: if the summed ranks exceed d, not every contraction is a feasible completion in that space.

Let $\mathcal{C}(D) = \{C : \|C\|_2 \le 1,\ C_{I,J} = D\}$ and $S(D) = \max_{C \in \mathcal{C}(D)} \sum_{(i,j) \notin I \times J} a_i b_j C_{ij}^2$. This compact set contains the zero-filled completion, whose omitted interaction is zero. A segment from that completion to a maximizer also lies in $\mathcal{C}(D)$; its omitted interaction varies as the square of the segment parameter. Hence all values in $[0, S(D)]$ are attainable. For $G = Q + 2L_{IJ}$, the compatible order-two ranks form

$$\left[\frac{E^2}{G + 2S(D)}, \frac{E^2}{G}\right], \qquad \inf_z \sup_{C \in \mathcal{C}(D)} \frac{|z - r_2(C)|}{r_2(C)} = \frac{S(D)}{G + S(D)} \ge \frac{W(D)}{G + W(D)}. \tag{38}$$

The minimax expression equalizes the relative errors at the interval endpoints. It concerns this fixed observation state and all deterministic scalar estimates, not a lower bound over all adaptive query policies. The plug-in estimator E^2/G has worst error $2S(D)/G$, which is generally larger.

To construct witnesses without using hidden entries, set $p = k_A - m$, $q = k_B - n$, $D_L = (I_m - DD^\top)^{1/2}$, and $D_R = (I_n - D^\top D)^{1/2}$. Classical Julia dilation (Robinson, 2018) gives an orthogonal matrix, and contraction by partial isometries $A \in \mathbb{R}^{p \times n}$, $B \in \mathbb{R}^{m \times q}$ gives

$$J_D = \begin{bmatrix} D & D_L \\ D_R & -D^\top \end{bmatrix}, \qquad C_{A,B} = \begin{bmatrix} D & D_L B \\ AD_R & -AD^\top B \end{bmatrix} = \mathrm{diag}(I_m, A) J_D \,\mathrm{diag}(I_n, B). \tag{39}$$

The intertwining identity $D_L D = D D_R$ follows from an SVD of D and verifies orthogonality directly. Thus every displayed completion preserves D and has operator norm at most one. We use two coordinate-ordered and three seeded-permutation choices for A, B, alongside zero filling and two tail-only partial isometries, giving eight witnesses per observation. Their maximum defines $W(D)$; this finite family is not claimed to exhaust $\mathcal{C}(D)$.

For an upper bound, let $\Omega = (I \times J)^c$, $\rho_i = 1 - \sum_{j:(i,j) \notin \Omega} D_{ij}^2$, and $\kappa_j = 1 - \sum_{i:(i,j) \notin \Omega} D_{ij}^2$, where empty sums vanish. Squared entries of every feasible completion satisfy the transportation relaxation

$$U_{\mathrm{tr}}^\star = \max_{x \ge 0} \sum_{(i,j) \in \Omega} a_i b_j x_{ij} \quad \text{subject to} \quad \sum_{j:(i,j) \in \Omega} x_{ij} \le \rho_i, \quad \sum_{i:(i,j) \in \Omega} x_{ij} \le \kappa_j, \tag{40}$$

$$U_{\mathrm{tr}} = \sum_i \rho_i y_i + \sum_j \kappa_j z_j, \qquad y_i, z_j \ge 0, \quad y_i + z_j \ge a_i b_j \quad ((i,j) \in \Omega). \tag{41}$$

Table 5: Fixed-observation completion diagnostics. Entry fractions and error/bound medians are percentages; each row has 30 source pairs. Actual is the plug-in relative error on the measured pair. Witness is the largest plug-in error among the constructed compatible completions, a lower bound on the worst-case error at fixed observations; it need not be below Actual. Transport and Original are the upper bounds $2U'/(Q+2L_{IJ})$ and $2U/(Q+2L_{IJ})$, respectively. Removed is the median guaranteed removable fraction $(U-U')/U$ in percent, not the difference of column medians.

Group, preprocessing, fraction	Entries	Actual	Witness	Transport	Original	Removed
DINO, raw, 0.1	1.000	0.061	0.767	0.814	1.192	32.13
DINO, raw, 0.4	16.000	0.013	0.203	0.209	0.307	32.50
DINO, raw, 0.8	64.000	0.002	0.062	0.066	0.104	37.70
DINO, centered, 0.1	1.010	1.821	22.493	24.269	27.581	13.03
DINO, centered, 0.4	16.161	0.434	7.043	7.213	8.249	13.36
DINO, centered, 0.8	64.645	0.066	2.028	2.167	2.829	23.96
Supervised, raw, 0.1	1.000	0.067	0.359	0.376	0.599	38.18
Supervised, raw, 0.4	16.000	0.008	0.042	0.046	0.085	47.31
Supervised, raw, 0.8	64.000	0.000	0.006	0.007	0.017	55.73
Supervised, centered, 0.1	1.010	4.246	21.709	22.989	27.228	15.47
Supervised, centered, 0.4	16.161	0.627	3.199	3.475	4.495	22.54
Supervised, centered, 0.8	64.645	0.040	0.456	0.574	0.839	31.48

Any feasible dual potentials give $S(D) \le U_{\mathrm{tr}}^{\star} \le U_{\mathrm{tr}}$. We solve the linear program, clip negative dual roundoff, repair every constraint rowwise, and check feasibility independently. This removes the separate-strip maximizations in U. It still relaxes the full contraction constraint to row and column capacities, so equality with $S(D)$ is not assumed. Equation (13) consequently brackets the unknown residual ambiguity and the certificate loss separately. Solver time is additional diagnostic work, not an asserted speedup.

A two-mode family makes the separation exact. Take sorted energies $a = (A, 1)$, $b = (B, 1)$ with $A, B \ge 1$, and observe only $C_{11} = c \in [0, 1]$. Then

$$C_0 = \begin{bmatrix} c & 0 \\ 0 & 0 \end{bmatrix}, \qquad C_* = \begin{bmatrix} c & \sqrt{1-c^2} \\ \sqrt{1-c^2} & -c \end{bmatrix}, \qquad S(D) = U_{\mathrm{tr}}^{\star} = (A+B)(1-c^2)+c^2. \tag{42}$$

To see optimality, write unknown squared entries as x, y, z. Capacities imply $x, y \le 1 - c^2$ and $z \le 1 - \max(x, y)$. The objective $Ax + By + z$ is nondecreasing in each of x, y when $A, B \ge 1$, so its maximum is attained at $x = y = 1 - c^2$, $z = c^2$. The orthogonal completion C_* realizes these values; C_0 realizes zero omitted interaction. For $A = B = 16$, $c = .8$, the original bound is $U = 21.76$ but the exact maximum is 12.16. The plug-in error is therefore at most 2.889%, rather than its 5.171% certificate, and 44.12% of the latter's width is removable without any new observation. This is an exactly solved family, not a claim about general completion optima.

The experiment checks 60 such cases with $A \in \{1, 4, 16, 64\}$, $B \in \{1, 4, 16\}$ and $c \in \{0, .25, .5, .8, 1\}$. For measured-source constructions we use all 120 raw/centered external pairs and fixed per-side fractions $.1, .4, .8$. All 360 observations retain the original source energies and signed queried block. A verifier regenerates and checks every witness, all dual constraints, and separate Gram-eigenvalue targets on 60 cases. The sampled matrices are counterfactual completions, not additional natural representations; conditions share sources. Table 5 reports every declared fraction.

D.3 A Shannon Certificate Without a Positive Spectral Floor

Let $\widetilde{K}$ retain the joint $I \cup J$ principal block of K and leave every other coordinate as a singleton block. This is a pinching of K, hence positive semidefinite with the same trace E and no smaller entropy. For $m = k_A + k_B \ge 2$, set $\rho = K/E$ and $\widetilde{\rho} = \widetilde{K}/E$. The omitted off-diagonal entries give

$$\|\rho - \widetilde{\rho}\|_F^2 = \frac{2(L - L_{IJ})}{E^2}, \qquad \frac{1}{2}\|\rho - \widetilde{\rho}\|_1 \le \sqrt{\frac{mU}{2E^2}}. \tag{43}$$

Here $\|\cdot\|_1$ is the trace norm, bounded by $\sqrt{m}$ times the Frobenius norm. Entropy continuity (Audenaert, 2007) and nonnegative entropy change imply

$$t = \min\left\{\sqrt{\frac{mU}{2E^2}}, 1 - \frac{1}{m}\right\}, \qquad 0 \le \frac{\widetilde{r}_1}{r_1} - 1 \le \exp[t\log(m-1) + h(t)] - 1. \tag{44}$$

Here $h(t) = -t\log t - (1-t)\log(1-t)$ is binary entropy with natural logarithms and $0\log 0 = 0$. The entropy-continuity function increases up to $1 - 1/m$, where it equals the maximal entropy gap $\log m$; clipping there remains valid even if the distance upper bound is larger. No positive eigenvalue floor is required. The price is conservatism: dimension, norm conversion, and capacity bounds can all loosen the certificate. Increasing nested rectangles reverses pinching and decreases the upper rank estimates toward the exact value. A failure to certify early is not proof that unqueried entries matter that much.

D.4 Mean-Related Coupling

For raw sources, write $H_A = X_A + \mathbf{1}\mu_A^\top$ and $H_B = X_B + \mathbf{1}\mu_B^\top$, where X_A, X_B have zero column means. The four terms in their cross-product are pairwise Frobenius orthogonal because centered columns sum to zero. Therefore

$$L = \underbrace{\|X_A X_B^\top\|_F^2}_{L_{cc}} + \underbrace{n_B\|X_A\mu_B\|_2^2}_{L_{cm}} + \underbrace{n_A\|X_B\mu_A\|_2^2}_{L_{mc}} + \underbrace{n_A n_B(\mu_A^\top\mu_B)^2}_{L_{mm}}, \tag{45}$$

$$L(t) = L_{cc} + t^2(L_{cm} + L_{mc}) + t^4 L_{mm}. \tag{46}$$

The second identity applies to simultaneous attenuation $H_X(t) = X_X + t\mathbf{1}\mu_X^\top$. All terms are nonnegative; the mean-related share is $(L_{cm} + L_{mc} + L_{mm})/L$ for $L > 0$. This isolates a preprocessing intervention, not a training mechanism. It changes E, Q, and the mode bases as well as L.

E Core Experimental Protocols and Numerical Checks

The primary collection comprises the complete 24×8 encoder–dataset grid in Table 6. Files contain up to 1,000 frozen feature rows with labels; merge analysis uses the first 800. Merging is performed only within a fixed encoder, where both sources share the ambient coordinate system. The loader casts arrays to float64 and does not add mean centering or row normalization. Existing preprocessing of the cached features is inherited, not independently regenerated by the manuscript builder. Accordingly, the current results concern the supplied feature matrices rather than a newly standardized raw-image benchmark.

Table 6: Exact encoder identifiers in the 672-pair archive, grouped by architecture for held-out evaluation and cluster resampling. Family membership is a grouping variable, not a causal explanation of spectral differences.

Family	Encoders
convnext	`convnext_base`, `convnext_small`, `convnext_tiny`
deit	`deit_small_patch16_224`, `deit_tiny_patch16_224`
densenet	`densenet121`, `densenet169`
efficientnet	`efficientnet_b0`, `efficientnet_b2`, `efficientnet_b3`
mobilenet	`mobilenetv3_large_100`, `mobilenetv3_small_100`
regnet	`regnetx_032`, `regnety_032`
resnet	`resnet101`, `resnet18`, `resnet34`, `resnet50`, `wide_resnet50_2`
swin	`swin_small_patch4_window7_224`, `swin_tiny_patch4_window7_224`
vit	`vit_base_patch16_224`, `vit_small_patch16_224`, `vit_tiny_patch16_224`

The eight dataset identifiers are `cifar10`, `cifar100`, `dtd`, `fashion_mnist`, `flowers102`, `pets`, `stl10`, and `svhn`. The 193-file pruning archive additionally contains `resnet152__cifar10.npz`, which has no partners and is excluded from the complete

merge grid. The distinct counts 192 and 193 therefore refer to different inclusion rules, not interchangeable sample sizes.

The legacy archive uses VTAB in cache and record names; these are internal identifiers, not a claim of compliance with the standard 19-task VTAB benchmark (Zhai et al., 2019). Five datasets overlap its task list: CIFAR-100, DTD, Flowers-102, Oxford-IIIT Pets, and SVHN. CIFAR-10, Fashion-MNIST, and STL-10 are not in that list, so our collection is not simply an eight-task subset of VTAB. Our evaluation concerns pooled frozen-feature spectra, not the standard downstream-task evaluation. The reported aggregate errors, correlations, and selection rates therefore characterize this collection and sampling protocol, not full VTAB performance. Legacy filenames are retained to preserve traceability to the original records.

The original direct-SVD implementation retains source singular values above 10^{-10}. The Gram implementation clips negative eigenvalues caused by roundoff and retains positive eigenvalues above 10^{-12} before computing effective ranks. The numerical cutoffs are not mathematically identical after squaring singular values. This matters especially at low order. The manuscript reports actual observed differences rather than describing every real-data comparison as machine precision. Fourteen pairs have unequal numerical source ranks; no smaller-rank alignment truncation is used in the primary records.

E.1 Error Metrics

Spearman correlations use average ranks for ties. For a prediction $\hat{r}_i$ and direct-SVD target $r_i > 0$, mean relative error is $n^{-1} \sum_i |\hat{r}_i - r_i|/r_i$. SD in Table 2 describes the distribution of pair errors, not sampling uncertainty for its mean. Gram mean and maximum errors are separate statistics. All 672 primary pairs are unique in the encoder and unordered dataset-pair key.

E.2 Numerical Identity Checks

The evidence builder uses NumPy's generator with seed 20260913 and float64 arithmetic. Each of 512 trials samples a Dirichlet spectrum with rank 2–32, order in $\{.5, 1, 2, 4\}$, and power $\beta \in \{.25, .5, 1, 2, 4\}$. Subsets include a leading block, its complementary tail, and an independently sampled index set. The same three choices give 1,536 power–subset compositions. Predictions from the formulas are checked against directly transformed probability vectors. Another orthogonal-union test uses independently sampled source masses and an energy fraction in $[.05, .95]$.

General-merging checks use independently generated orthonormal bases in $\mathbb{R}^{12}$ with unequal ranks three and five, and lognormal singular values. A direct SVD of the explicitly stacked matrices is compared with Gram reconstruction and the order-two trace formula. Table 7 gives the realized maxima for this seed and implementation. A finite test set cannot replace the proofs, but catches normalization and composition mistakes.

Table 7: Fresh numerical identity checks, separate from archived real-data measurements. Errors are relative to the directly evaluated result.

Check	Evaluations	Maximum rel. error
Power reweighting	512	5.77×10^{-15}
Subset	1536	1.78×10^{-15}
Power–subset composition	1536	1.51×10^{-14}
Orthogonal merge	512	1.11×10^{-15}
Gram	512	5.11×10^{-15}
Order two	512	1.11×10^{-15}

E.3 Traceable Extraction and Controlled Interventions

A separate extraction uses torchvision IMAGENET1K_V1 ResNet-18 and ViT-B/16 on 1,000 label-independently selected training images per dataset from CIFAR-10, CIFAR-100, MNIST, and Fashion-MNIST (seed 20260916). Classifier heads are removed; grayscale images become RGB. Both use

Table 8: Certified coupling-entry budgets at 1% order-two tolerance, with 140 pairs per row. IQR describes the observed pair distribution. Full is the percentage reaching the final query step. All stopping decisions meet the tolerance.

Encoder, preprocessing	Median (%)	IQR (%)	Mean (%)	Full (%)
ResNet-18, raw	0.25	[0.06, 0.25]	0.20	0.00
ResNet-18, centered	74.75	[49.49, 100.00]	73.56	50.00
ResNet-18, unit rows	0.25	[0.06, 0.25]	0.20	0.00
DenseNet-121, raw	1.00	[0.25, 1.00]	0.64	0.00
DenseNet-121, centered	100.00	[100.00, 100.00]	87.99	78.57
DenseNet-121, unit rows	1.00	[0.25, 1.00]	0.65	0.00
ConvNeXt-Tiny, raw	4.00	[1.00, 4.00]	6.16	0.00
ConvNeXt-Tiny, centered	100.00	[49.49, 100.00]	82.21	67.14
ConvNeXt-Tiny, unit rows	4.00	[1.00, 4.00]	6.38	0.00
ViT-B/16, raw	32.50	[16.00, 49.00]	40.24	21.43
ViT-B/16, centered	100.00	[100.00, 100.00]	87.99	78.57
ViT-B/16, unit rows	32.50	[16.00, 49.00]	40.24	21.43

resize 256, center crop 224, bilinear interpolation, and ImageNet normalization. Evaluation-mode features are float32 and spectral analysis float64, without feature centering or row normalization. Weight/state hashes, original indices, image-content hashes, and feature hashes are recorded. All six dataset pairs per encoder use 800-row subsamples at seeds 20260917–20260919: 36 dependent evaluations of 12 distinct pairs. Mean orthogonal-reference Shannon error is 41.4%; Gram agrees near numerical precision. These measurements are separate from the inherited archive.

For the controlled intervention, all 28 dataset pairs from each of ResNet-18, ResNet-50, DenseNet-121, ViT-Base/16, Swin-Tiny, and ConvNeXt-Base give 168 pairs. Keeping singular values fixed, columns of the source direction matrices are permuted. Each pair uses the identity, two orderings by squared-overlap sums, and five seeded permutations. This preserves source subspaces and principal angles while changing their association with energy; it constructs matrices, not new natural observations. The range over these eight assignments is normalized by the identity-assignment rank, not claimed to be a global extremum. Its median at order one is 9.73%.

On the same cohort, matched merged-energy cutoffs $\tau\lambda_{\max}^{\mathrm{direct}}$ with $\tau \in \{10^{-14}, 10^{-12}, 10^{-10}, 10^{-8}\}$ give identical retained ranks in all 672 pair–cutoff cases; maximum effective-rank error is 1.17×10^{-12}. Nested sample sizes 200/400/800 at three seeds give 1,512 dependent evaluations, recomputing source factors and merged targets. The preferred ranking order changes with sample size, while the exact order-two identity has maximum relative error 4.33×10^{-15}. Repeated rows, source pairs, and interventions are not independent datasets.

E.4 Partial-Coupling Protocol

The four-encoder cohort uses torchvision IMAGENET1K_V1 ResNet-18, DenseNet-121, ConvNeXt-Tiny, and ViT-B/16, with official weight-specific preprocessing and classifier heads removed. Each encodes the same 4,000 label-independently selected test images per dataset in CIFAR-10, CIFAR-100, MNIST, and Fashion-MNIST. The resulting 16 matrices contain 64,000 encodings of 16,000 images. Five partitions per dataset each supply two disjoint 400-row blocks; coupling experiments use their 200-row prefixes. Models share image indices, and pairs share sources. Historical archive metadata cannot establish disjointness from every earlier cache.

Each encoder/partition supplies eight sources and 28 pairs: 560 pairs in total. Raw, independently source-centered, and unit-row conditions give 1,680 evaluations, without joint post-merge centering. Float64 analysis retains singular values above 10^{-10}. Leading rectangles use per-side fractions $0, .025, .05, .1, .2, .4, .7, 1$, with counts rounded upward; entry budget is the product of retained counts divided by $k_A k_B$. Both orders and tolerances $.01, .02, .05, .1$ are retained; 1% order-two error is primary. The refined ladder adds $.8, .9, .95, .99, .995$. Actual-error stopping uses exact targets and is diagnostic; certification uses only queried information.

Table 9: External certification at 1% relative error in r_2, original leading-mode ladder. Budgets are percentages of coupling entries. Full counts pairs querying all entries; Fail counts tolerance violations, out of 30 pairs per row.

ResNet-50, preprocessing	Median (%)	Mean (%)	Full	Fail
Supervised, raw	1.00	0.69	0/30	0/30
Supervised, centered	100.00	83.16	20/30	0/30
Supervised, unit rows	1.00	0.69	0/30	0/30
DINO, raw	4.00	2.71	0/30	0/30
DINO, centered	100.00	100.00	30/30	0/30
DINO, unit rows	4.00	2.70	0/30	0/30

The raw median entry budget is 1%, but the mean is 11.81% and 5.36% of pairs query everything; centered values are 100%, 82.94%, and 68.57%. Thus the small median does not describe every pair. At fixed 1% entries, raw head-query median error is 0.0833%, versus 62.204% for random and 63.076% for tail queries. Mean attenuation at $t = 0, .25, .5, .75, 1$ gives median budgets $100, 49, 16, 4, 1\%$ over 2,800 evaluations; spectra and directions also change. Mean share alone is not a universal predictor of cost.

The stress suite varies ranks $(8, 8), (16, 8), (32, 32), (64, 32)$, eight spectral families, energy ratios $.01, 1, 100$, overlap scales $0, .5, 1$, and overlap locations, with three seeds: 3,456 feasible cases and 98,496 checks. Direct stacked SVD supplies targets. Maximum numerical excess over a bound is 9.33×10^{-15}, below the 10^{-8} tolerance. Flat spectra and balanced energies are harder to certify; Shannon guarantees are especially conservative.

E.5 External Validation

The external configuration is frozen before extraction. We use torchvision IMAGENET1K_V1 ResNet-50 and the official DINO ResNet-50 checkpoint (Caron et al., 2021), both with the same resize/crop and ImageNet normalization. We select 2,000 label-independent indices from the official Flowers-102 test split (6,149 images) and 2,000 from Oxford-IIIT Pets test (3,669). Thus 4,000 unique selected images produce 8,000 encodings in four 2000×2048 matrices. The model comparison holds architecture and evaluation inputs fixed, not their entire pretraining recipes. Its interpretation is therefore external replication, not causal identification of self-supervision.

Five disjoint partitions per dataset each contain two 200-row blocks. Four sources yield six pairs per partition: 30 per model/preprocessing, comprising 10 same-dataset and 20 cross-dataset pairs. The 180 evaluations retain the original ladder, its refined diagnostic, all four tolerances, both targets, and five stochastic seeds per probe budget. Because conditions share images and pairs share sources, we report descriptive summaries of these fixed evaluations. Table 9 includes every primary condition.

Raw certified median budgets are 1% (supervised) and 4% (DINO), with maximum actual stopping errors 0.2113% and 0.3724%. Their mean-related coupling shares are similar (99.8275% and 99.7645% medians), so this share alone does not explain the difference in query cost. Centered refined median certificates still require 64.645% and 98.998%, versus 16.161% observed-error budgets for both. The larger certificate–observation gap for DINO identifies a case where a tighter bound could be particularly useful. Every real-feature Shannon certificate reaches full coupling at 1% tolerance, including the refined ladder; this records the stopping behavior of the tested bound, not a necessity result.

F Spectral Decisions and Computational Baselines

F.1 Exact and Stochastic Baselines

The exact baseline computes $E_X = \|H_X\|_F^2$, $Q_X = \|H_X H_X^\top\|_F^2$, and $L = \|H_A H_B^\top\|_F^2$ directly. It needs neither source SVD nor the complete ambient second moment. Its usefulness depends on matrix shape, so avoiding SVD alone is not an efficiency claim. We also compare full spectral coupling, adaptive partial coupling, Hutchinson, and Hutch++ (Meyer et al., 2021).

For stochastic methods, set $F = H_A H_B^\top$ and $B = FF^\top \succeq 0$. Then $L = \operatorname{tr} B$ and BZ is evaluated implicitly from the feature matrices. Hutchinson averages Rademacher quadratic forms. For a Hutch++ budget of $3s$, draw independent sign matrices S, G with s columns, form $Q_B = \operatorname{orth}(BS)$ and $G_\perp = (I - Q_B Q_B^\top)G$, and estimate

$$\widehat{L} = \|F^\top Q_B\|_F^2 + \frac{1}{s}\|F^\top G_\perp\|_F^2, \qquad \widehat{r}_2 = \frac{E^2}{Q + 2\widehat{L}}. \tag{47}$$

The $3s$ convention counts sketch, projected-trace, and residual directions; exploiting $B = FF^\top$ reduces arithmetic for quadratic forms. It is not a coupling-entry count. Even an unbiased trace estimate does not imply an unbiased reciprocal rank estimate. We test budgets 12/24/48 with five seeds, giving 50,400 cohort and 5,400 external estimates across both stochastic methods. No confidence parameter is calibrated to produce a per-pair guarantee at the selected budgets.

F.2 Fixed-Base Decisions and Timing Protocol

For a fixed base A, let $r_A = r_2(H_A)$ and define the relative diversity gain and its certified interval:

$$g(L) = \frac{E^2}{r_A(Q + 2L)} - 1, \qquad \mathcal{I}_U(D) = [g(L_{IJ} + U), g(L_{IJ})]. \tag{48}$$

The rule is ADMIT when $g \geq \eta$, SKIP otherwise. Since g decreases with L, the lower endpoint at least η certifies ADMIT, and the upper endpoint strictly below η certifies SKIP. A crossing interval remains undecided, however small its rank error. The absolute gain threshold is ηr_A, not a deficit relative to orthogonality; equality belongs to ADMIT.

Four states partition the fixed observations. *Original* means $\mathcal{I}_U$ certifies a decision. *Tightened* means only $\mathcal{I}_{U'}$ certifies it, without new queries. *Ambiguous* requires two feasible completions preserving the signed observed block and source spectra, with gains on opposite sides. No deterministic rule using that observation can be correct on both. *Unresolved* means neither bounds nor finite witnesses settle the question. A crossing upper bound alone is not an impossibility result. Witnesses range over the contraction model with $d \geq k_A + k_B$, not a learned distribution of natural features.

We reuse the four external 2000×2048 feature matrices from supervised/DINO ResNet-50 on Flowers-102 and Pets. Each dataset has five disjoint random partitions, each with a fixed 128-row base and four 64-row candidate shards. Models share indices; 1,920 rows are used per matrix. There is no sequential base update or label access. Raw, independently source-centered, and unit-row conditions yield 60 banks and 240 pairs. Fractions $.1, .4, .8$ per side give 720 observations; thresholds $\eta = 0, .01, .05, .10$ give 2,880 states. The primary threshold $.05$ was fixed before this run. This is a frozen follow-up analysis of previously inspected features, not new unseen data or public preregistration.

Timing covers partitions 0 and 1 in every model/dataset/preprocessing condition: 24 banks, five methods, three preparation scenarios, and three randomized-order repeats, giving 1,080 records. Adaptive policies query fractions $0, .025, .05, .1, .2, .4, .7, .8, .9, .95, .99, .995, 1$, stopping only when the gain interval resolves the threshold. Transport is solved online only when the original interval is undecided and the block nonempty and incomplete. Complete policies may reach full coupling; fixed-observation classification never adds queries. Intervals are expanded by $10^{-10}\max(1, |g_{\mathrm{lo}}|, |g_{\mathrm{hi}}|)$; full-query point comparison handles a remaining numerical boundary. Verification uses 10^{-8} tolerance, with boundary flags recorded separately.

Table 10 includes solver and preparation costs. Cached raw and unit-row policies are faster than direct products here, whereas centered transport solves take seconds per bank. Direct products win every preparation-charged aggregate comparison at these shapes. Each certified policy agrees with all 864 recorded exact decisions; repeats are not independent examples. Hutch++ disagrees seven times, all on centered inputs outside the numerical-boundary tolerance. Agreement checks execution of the spectral rule, not downstream accuracy.

The run uses four CPU numerical threads, Python 3.12.3, NumPy 2.4.4 and SciPy 1.17.1 on Linux/OpenBLAS. Short workloads use calibrated loops; timings exclude image encoding, disk access and separately recorded preprocessing. The scientific run took 669.7 seconds. A separate verifier regenerated source factors and all 5,760 witnesses, checked dual constraints, 60 Gram-eigenvalue

Table 10: Median milliseconds per four-candidate bank, 24 timing records per cell. Original/transport are adaptive policies; full computes all spectral coupling; direct uses no-SVD feature products; H++ uses 48 probes. Reuse charges preparation once per bank; per-pair charges it anew.

Preprocessing, scenario	Original	Transport	Full	Direct	H++
Raw, cached	1.03	0.99	2.27	2.00	4.82
Raw, reuse	100.27	99.45	101.58	4.86	7.65
Raw, per-pair	233.11	228.93	235.62	7.96	11.05
Centered, cached	7.30	4004.55	2.73	2.30	4.94
Centered, reuse	112.62	3944.48	108.67	5.49	8.35
Centered, per-pair	244.75	3967.72	243.18	8.56	11.31
Unit-row, cached	0.96	0.99	2.41	2.24	5.02
Unit-row, reuse	96.63	98.28	98.98	4.90	7.31
Unit-row, per-pair	223.37	226.02	225.41	7.73	10.29

Table 11: Budget-matched spectral partner selection on 160 queries with seven candidates and 200 rows per source. Regret is relative Shannon-rank loss; calls are out of 1,120. Random allocation matches the screening budget in every batch and reports analytic expectations. Queries share candidate pools.

Method	Top-1 (%)	Regret (%)	Gram calls
Source-rank sum	89.375	0.10265	0
Gap screening (2%)	95.000	0.06955	546
Budget-matched random (expected)	95.625	0.04552	546
Full Gram	100.000	0.00000	1120

targets, decision histories and summaries. It did not independently repeat timings. Three counterexample tests and five additional boundary/integration tests passed. New run identifiers keep these outputs separate from historical results.

G Negative Controls for Selection and Task Utility

G.1 Budget-Matched Partner Selection

The cohort in Appendix E.4 also supplies 160 seven-candidate queries with 200 rows per source. The fixed rule escalates a query to exact evaluation when its top-two source-sum gap is at most 2%; otherwise it uses the source-sum maximizer. Regret is relative loss against the largest directly measured merged Shannon rank. The control randomly escalates exactly the same number of queries within every encoder–partition batch, with source-only fallback. Reported control metrics are analytic expectations: each query is escalated with that batch's fixed sampling fraction. Targets are used only for evaluation, not allocation. Ties within 10^{-10} receive uniform expected choices. Queries share candidates; these are descriptive spectral outcomes, not task-accuracy or independent-population estimates.

At the same 546/1,120 exact-call budget, screening achieves 152/160 correct choices versus the random expectation 153/160, with higher regret (0.06955% versus 0.04552%). Its gain over source-only selection therefore does not establish useful allocation. All 216 sample-size/pool/policy cells remain in the records: the 400-row sensitivity is favorable, but 100-row and cross-dataset-only comparisons are not. The favorable sensitivity does not replace the fixed primary test.

G.2 Train-Only Pruning

The 193-file archive uses a seed-zero 80/20 split of up to 1,000 rows. Uncentered SVD and dimension selection use training rows only; test rows use the same projection. Energy thresholds select the smallest leading dimension retaining the prescribed energy share. The diversity rule retains conditional effective rank at least .9 times its full training value. Nearest-class means use training class means;

the ridge probe uses strength one without task-specific tuning. Table 12 compares rules at their own selected dimensions, not matched dimensions.

Table 12: Pruning rules on the same 193 files. Values are mean ± SD across files; SD reflects heterogeneous tasks and encoders, not CI. These own-dimension comparisons are not dimension matched.

Rule	Retained dimension	NCM accuracy	Ridge accuracy
Train full	684.6 ± 197.4	0.7189 ± 0.2153	0.7292 ± 0.1996
Energy 95%	227.5 ± 110.0	0.7163 ± 0.2175	0.7361 ± 0.2006
Energy 99%	409.5 ± 146.1	0.7181 ± 0.2158	0.7437 ± 0.1948
Law 2, $\alpha = .5$	503.3 ± 145.0	0.7186 ± 0.2152	0.7425 ± 0.1936
Law 2, $\alpha = 1$	365.9 ± 128.1	0.7179 ± 0.2159	0.7441 ± 0.1950

The order-.5 rule retains more dimensions than the 99% energy rule with similar ridge accuracy. At a matched dimension the projections are identical; this audit does not show an improved accuracy–dimension trade-off.

H Reproduction Scope

The anonymous package separates `records`, `cached`, and `raw` modes. Seven record builders reconstruct manuscript summaries from saved measurements; cached mode recomputes analyses from supplied features; raw mode downloads official data and weights before extraction. Cohort and external raw extraction require CUDA. The `completion` and `admission` suites regenerate witnesses and transport solutions, with admission also measuring timings. Inputs are fingerprinted, outputs use new directories, and images/weights are not redistributed.

Clean-directory runs validated the earlier cached suite and the completion extension; the latter reproduced all 629 declared summary fields. A small public-data extraction was tested, but not another full CUDA extraction. The full fixed-base server run passed independent verification of 5,760 witnesses, dual constraints, selected Gram targets, and decision summaries; independent verification did not repeat timing. Packaging/record checks, cached analysis, and raw extraction are reported separately. Hardware-dependent timings and timed stochastic draws are excluded from summary equality checks.

The inherited archive lacks enough metadata to reconstruct its original extraction. No executable entry point, numerical agreement, or successful package build fills that gap or establishes downstream utility. Source-sharing and nested conditions are retained in interpretation; full records include secondary sensitivities and earlier pilot runs beyond the essential evidence printed here.