GROUP-ORBIT MODERN HOPFIELD NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Modern Hopfield networks retrieve over individual vectors, even when rotational symmetry generates infinitely many equivalent observations of each semantic object. We introduce Group-Orbit Modern Hopfield Networks (G-MHNs), together with a theory and complete T-step update algorithm for aggregating transformation evidence within each quotient orbit and retrieving over semantic orbit classes. The update admits finite-sampled and continuous-Haar evaluation under the same orbit-level retrieval semantics. We prove equivariance of the update, orbit invariance of its energy, and separation-dependent retrieval and capacity certificates for finite banks and smooth positive-dimensional orbits. On official rMD17 splits, the G-MHN continuous posterior mean reduces hidden-atom RMSD by 10.8–22.8% relative to nearest-pose Procrustes retrieval and wins all 40 split-condition comparisons. These results show that orbit memory can improve symmetry-aware retrieval without treating equivalent poses as separate semantic memories.

1 INTRODUCTION

Ramsauer et al. (2021) introduce modern Hopfield networks (MHNs) with exponential memory capacity and one-shot retrieval, proving their update rule is mathematically equivalent to Transformer self-attention. However, they treat each stored vector as an individual memory. This point-level abstraction is natural when different vectors correspond to different semantic objects, but it becomes ambiguous when a known transformation changes the observation without changing the underlying object. A molecule recorded in different laboratory frames, for example, gives infinitely many rotated feature vectors although all of them represent the same molecular structure. The same issue occurs for rotated images and three-dimensional objects. Thus, under symmetry, the fundamental question is not only how to retrieve a memory, but also what should count as one memory.

A standard solution is to remove the nuisance transformation by learning an invariant representation. This makes transformed observations share the same feature, but it also removes the transformation state that may be needed for reconstruction or pose-conditioned retrieval. The opposite solution is to preserve the transformation and explicitly store multiple transformed copies. This retains state information but makes equivalent observations compete as separate memories and causes the memory footprint to depend on the resolution used to discretize the transformation space. For a continuous group such as $SO(3)$, this tension becomes particularly clear: the same semantic object corresponds to a continuum of point memories.

We resolve this tension by changing the memory unit itself. We introduce Group-Orbit Modern Hopfield Networks (G-MHNs), in which a semantic memory is represented by its transformation orbit rather than by a single transformed sample. For a prototype ξ_a and group representation ρ, the memory is the quotient orbit G/H_a, where the stabilizer H_a removes transformations that leave the prototype unchanged. Retrieval therefore operates at two levels: the model first accumulates evidence for the semantic orbit and then infers the transformation-dependent state within that orbit. This construction gives G-MHN a property that neither invariant pooling nor point-level retrieval provides. The posterior over semantic objects is invariant to the input pose, while the retrieved state remains equivariant: changing the query by a group action changes the reconstructed state by the same action without changing its semantic identity. In other words, G-MHN is invariant where memory identity should be invariant and equivariant where the retrieved state should transform.

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.

Our contributions are as follows:

- **Orbit-level memory.** G-MHN aggregates evidence over each quotient orbit before objects compete, yielding a pose-invariant posterior and equivariant state. Prototype storage avoids pose duplication: 32 rMD17 prototypes use 6,912 bytes versus 3.54 GB for an expanded 512,000-pose bank.
- **Retrieval guarantees.** We establish an orbit-invariant energy identity and derive conditional finite-step retrieval, normal-direction contraction, Gaussian-noise failure bounds, and sufficient bank-dependent capacity certificates, accounting for sampling and numerical approximation errors.
- **Sampled and continuous updates.** One T-step rule supports sampled and normalized-Haar retrieval. Fourier/Bessel $\mathrm{SO}(2)$ and fixed-correspondence matrix-Fisher $\mathrm{SO}(3)$ avoid pose grids and their discretization error; two sampled steps improve masked ModelNet40 balanced accuracy over matched TTA (0.2652 versus 0.2045).

2 Related Works

MHNs connect continuous associative retrieval with Transformer self-attention (Hopfield, 1982; Ramsauer et al., 2021), and recent work has substantially broadened their dynamics, architectures, and theoretical interpretations. Universal associative memory, dense associative memory, and Bayesian adaptive variants extend the expressive or probabilistic view of Hopfield retrieval (Millidge et al., 2022; Mimura et al., 2026; Wang et al., 2026a), while recent studies examine parameter-free Hopfield attention, attractor-like dynamics in vision transformers, complex retrieval trajectories, and manifold-constrained Hopfield models (Masumura & Taki, 2026; Xia & Jiang, 2026; Bogakovskaya et al., 2026; Wang et al., 2026b). These developments, however, still primarily regard individual feature vectors as the basic memory units. Murray et al. (Murray et al., 2025) show that classical Hopfield neural networks (HNNs) can implicitly learn group-orbit structure through energy-flow training, but the symmetry is learned from data rather than explicitly represented as a memory primitive.

A complementary line of work builds symmetry directly into feature representations through invariant or equivariant architectures (Cohen & Welling, 2016; Cohen et al., 2019; Fuchs et al., 2020; Liao & Smidt, 2023; Liao et al., 2024). Such methods specify how features should transform under a known group action, but they do not by themselves change the semantic unit stored by an associative memory. This distinction is important for retrieval: making a representation invariant can identify transformed observations with the same object, whereas an orbit-level memory can identify them semantically while still retaining the transformation state needed to reconstruct or retrieve a pose-dependent output. G-MHN therefore addresses symmetry at the level of the associative memory itself, placing semantic identity and transformation state in a single retrieval framework rather than requiring either explicit transformation augmentation or complete removal of the transformation.

3 Sampled and Continuous Group-Orbit Retrieval

3.1 Sampled quotient-orbit update

An MHN with point memories $\Xi = [\xi^1, \ldots, \xi^L] \in \mathbb{R}^{d\times L}$ updates a query x by

$$F_{\mathrm{MHN}}(x) = \Xi \,\mathrm{softmax}(\beta \Xi^\top x). \tag{1}$$

Let a finite or compact group G act through an orthogonal representation $\rho : G \to O(d)$, where every $\rho(g)$ is a d-dimensional matrix and

$$\rho(g)\rho(g^{-1}) = \rho(e) = I_d. \tag{2}$$

Equivalently, $\rho(G)$ is a subgroup of $O(d)$. Object a has prototype ξ^a and stabilizer $H_a = \{g \in G : \rho(g)\xi^a = \xi^a\}$. Intuitively, H_a contains transformations that do not create a distinct pose, such as rotations about the axis of a perfectly axisymmetric object. Distinct poses are therefore indexed by the quotient G/H_a, not by all of G.

For computation, fix Q_a representatives $g_{a,q}$, positive normalized weights $w_{a,q}$, and sampled atoms $z_{a,q} = \rho(g_{a,q})\xi^a$. Given orbit prior π_a, define

$$Z_a(x) = \sum_{q=1}^{Q_a} w_{a,q} \exp(\beta \langle z_{a,q}, x \rangle), \tag{3}$$

$$p_x(q \mid a) = \frac{w_{a,q} \exp(\beta \langle z_{a,q}, x \rangle)}{Z_a(x)}, \qquad \mu_a(x) = \sum_q p_x(q \mid a) z_{a,q}, \tag{4}$$

$$P_x(a) = \frac{\pi_a Z_a(x)}{\sum_b \pi_b Z_b(x)}, \qquad F(x) = \sum_a P_x(a) \mu_a(x). \tag{5}$$

Eq. (5) is Eq. (1) with a hierarchical posterior: pose evidence is normalized within an orbit, then orbit evidence is normalized across semantic objects. The weights prevent oversampled or stabilizer-repeated orbits from receiving an accidental prior advantage. All features and prototypes are L_2-normalized before memory construction.

3.2 Structure and energy

Theorem 1 (Orthogonality and equivariance)**.** *For the Euclidean score used by G-MHN, compatibility with every transformed query and prototype holds if and only if $\rho(G) \subseteq O(d)$. Under this condition, the exact normalized-Haar update is equivariant; a sampled update is equivariant when its weighted bank is closed under the group action. The resulting transformation laws are*

$$F(\rho(k)x) = \rho(k)F(x), \qquad P_{\rho(k)x}(a) = P_x(a), \qquad k \in G. \tag{6}$$

Equivalently, for $\mathcal{O}_x = \{\rho(k)x : k \in G\}$,

$$F(\mathcal{O}_x) = \mathcal{O}_{F(x)}. \tag{7}$$

Thus changing only the pose moves the retrieved state by the same group action, while the semantic object posterior remains unchanged.

Proposition 1 (Energy identity)**.** *For*

$$E_G(x) = \frac{1}{2}\|x\|^2 - \frac{1}{\beta} \log \sum_a \pi_a Z_a(x) + C, \tag{8}$$

we have $\nabla E_G(x) = x - F(x)$. Under the assumptions of Theorem 1, $E_G(\rho(k)x) = E_G(x)$. Consequently, every point on the same orbit $\mathcal{O}_x$ has the same energy; pose changes move along an energy level set rather than creating new semantic basins. Fig. 1 visualizes this orbit-invariant energy landscape on an $\mathrm{SO}(2)$ slice.

Appendix B gives one joint proof of Theorem 1 and Proposition 1 for normalized Haar measures and closed weighted banks. A generic fixed $\mathrm{SO}(3)$ sample is only approximately closed, so its error must be measured rather than assumed to vanish.

3.3 T-step search and computation

Starting from x_0, we apply Eq. (5) for $s = 0, \ldots, T-1$. The final result includes the state x_T, orbit posterior P_T, the highest-posterior orbit, and an optional top-K candidate set. This separates three events: state convergence, target coverage, and selective retrieval. Unrestricted stable-set membership alone is not counted as success because a set containing every object has perfect coverage but no selectivity.

Algorithm 1 is the complete T-step procedure. Figure 2 visualizes one continuous $\mathrm{SO}(3)$ orbit-statistics evaluation inside its second step.

Sampled complexity. For M orbits, Q poses, and dimension d, a materialized scan costs $O(TMQd)$. Expanded point memory stores $O(MQd)$ values; procedural generation stores $O(Md + Qc_G)$, where c_G describes one shared group element. Against Q-view TTA, G-MHN exchanges repeated encoder calls for latent scans. If C_{enc} and C_{scan} are the corresponding costs, the sampled update is favorable when $(T-1)C_{\mathrm{scan}} < (Q-1)C_{\mathrm{enc}}$. This is a workload condition, not a universal speed guarantee.

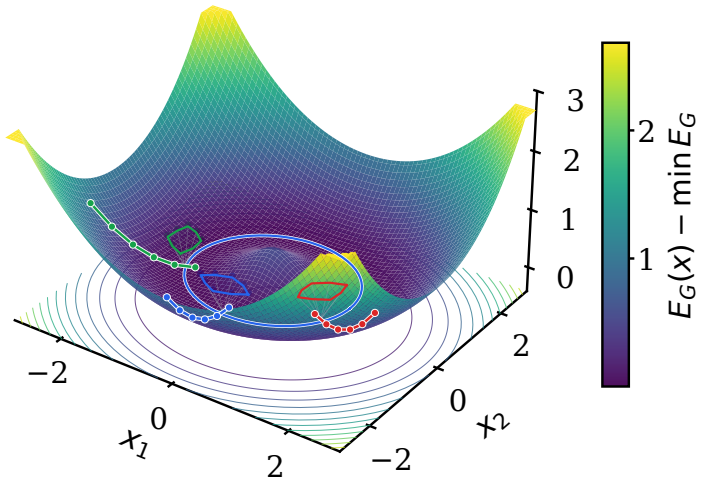


Figure 1: Sampled $\mathrm{SO}(2)$ energy landscape with $M = 50$, $Q = 300$, uniform orbit priors, and $\beta = 7$, shifted by its sampled minimum. The white query orbit illustrates that group-equivalent states share energy up to sampling error. Colored descent trajectories approach orbit-equivalent neighborhoods; the translucent polygons are orientation guides, not estimated basins.

Algorithm 1: Group-orbit Hopfield retrieval

1. **Input:** normalized query x_0, prototypes and priors $\{\xi^a, \pi_a\}$, an orbit-statistics backend, β, T, and candidate budget K.
2. For $s = 0, \ldots, T-1$, use either the sampled sums in Eqs. (3)–(4) or a continuous backend to compute $\log Z_a(x_s)$ and $\mu_a(x_s)$. Form $P_{x_s}(a)$ and set $x_{s+1} = \sum_a P_{x_s}(a)\mu_a(x_s)$.
3. Stop early only when the prescribed state and posterior tolerances both hold; otherwise execute the complete T-step budget.
4. **Output:** final state x_T, orbit posterior P_T, the primary orbit, and $\mathcal{S}_T^{(K)} = \mathcal{S}_T \cap \mathrm{Top}_K(P_T)$.

3.4 Continuous Haar orbit statistics

Let ν_a be the normalized invariant measure on G/H_a. Replacing the finite sums by their mathematical target gives

$$Z_a^{\mathrm{H}}(x) = \int_{G/H_a} e^{\beta\langle\rho(g)\xi^a, x\rangle}\, d\nu_a(gH_a), \tag{9}$$

$$\mu_a^{\mathrm{H}}(x) = \frac{1}{Z_a^{\mathrm{H}}(x)} \int_{G/H_a} \rho(g)\xi^a e^{\beta\langle\rho(g)\xi^a, x\rangle}\, d\nu_a(gH_a). \tag{10}$$

Figure 2 depicts these two statistics for the single-vector $\mathrm{SO}(3)$ quotient orbit. We propose two continuous realizations of the same orbit update: a Fourier/Bessel construction for $\mathrm{SO}(2)$ and a matrix-Fisher construction for fixed-correspondence $\mathrm{SO}(3)$ channels. Both replace only the per-orbit statistics in Eq. (5); the outer posterior, T-step dynamics, and retrieval rule remain unchanged.

Multi-frequency $\mathrm{SO}(2)$. Every finite-dimensional real orthogonal $\mathrm{SO}(2)$ representation decomposes into invariant coordinates and two-dimensional frequency blocks. Its score is a finite trigonometric polynomial

$$\langle\rho(\theta)\xi^a, x\rangle = c_{a,0} + \sum_{k\in\mathcal{K}} r_{a,k}\cos(k\theta - \phi_{a,k}). \tag{11}$$

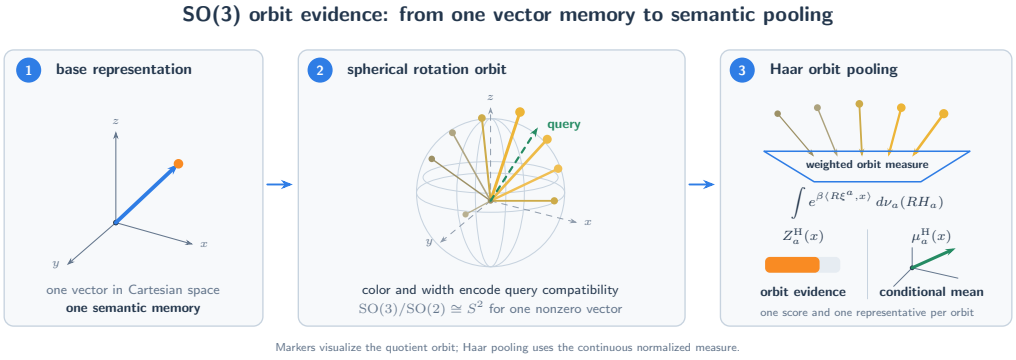


Figure 2: $\mathrm{SO}(3)$ orbit evidence for a nonzero vector memory. (1) A base vector $\xi^a \in \mathbb{R}^3$ is one reference pose. (2) Its orbit is $\mathrm{SO}(3)/\operatorname{Stab}(\xi^a) \cong \mathrm{SO}(3)/\mathrm{SO}(2) \cong S^2$; color and line width encode the compatibility $e^{\beta\langle R\xi^a, x\rangle}$, not separate semantic memories. (3) Normalized-Haar pooling yields the scalar evidence $Z_a^{\mathrm{H}}(x)$ and the conditional mean $\mu_a^{\mathrm{H}}(x)$; the outer normalization in Eq. (5) converts orbit evidences into $P_x(a)$. For a general multi-channel memory, the sphere is replaced by $\mathrm{SO}(3)/H_a$. A sampled backend replaces the integral by weighted quadrature without changing the outer update.

Expanding each exponential into modified-Bessel Fourier coefficients and convolving across frequencies gives coefficients $C_{a,\ell}$, for which

$$Z_a^{\mathrm{H}}(x) = e^{\beta c_{a,0}} C_{a,0}, \qquad \mathbb{E}[e^{i\ell\theta} \mid a, x] = \frac{C_{a,-\ell}}{C_{a,0}}. \tag{12}$$

The required conditional mean follows blockwise from these moments. Symmetric truncation has an explicit Bessel-tail error; the strict implementation uses ball arithmetic, while the ordinary implementation uses scaled coefficients and an adaptive scalar-spectral fallback. Appendix A.2 gives the rotation-group definition, complete Fourier/Bessel derivation, explicit conditional mean, and certified truncation bounds.

Fixed-correspondence $\mathrm{SO}(3)$ vector channels. Write prototype and query channels as $X_a, Y \in \mathbb{R}^{n\times 3}$, acted on by $X_a \mapsto X_a R^\top$. The score becomes $\operatorname{tr}(K_a^\top R)$, where $K_a = \beta Y^\top X_a$. For the proper SVD $K_a = U \operatorname{diag}(s_1, s_2, s_3) V^\top$, define $A(u) = (s_1 - s_2)(1-u)/2$ and $B(u) = (s_1 + s_2)(1+u)/2$. The normalized matrix-Fisher normalizer and mean are

$$c(K_a) = \frac{1}{2}\int_{-1}^{1} I_0(A(u)) I_0(B(u)) e^{s_3 u}\, du, \tag{13}$$

$$\mathbb{E}[R \mid a, Y] = U \operatorname{diag}(\partial_{s_1} \log c, \partial_{s_2} \log c, \partial_{s_3} \log c)\, V^\top, \quad \mu_a^{\mathrm{H}}(Y) = X_a \mathbb{E}[R \mid a, Y]^\top. \tag{14}$$

This removes a three-dimensional pose grid but still evaluates a deterministic one-dimensional integral. It applies to corresponded points or equivariant $\ell = 1$ channels, not arbitrary unordered point sets. Both continuous backends store $O(Md)$ values and have no Q-dependent pose bank; their arithmetic depends instead on effective spectral order, concentration, and requested tolerance. Full derivations and numerical-certification boundaries are in Appendix A.3; the common normalized-Haar objective is stated in Appendix A.1.

4 Retrieval and Capacity Guarantees

Let $\mathcal{M}_a = \mathcal{O}_{\xi^a}$, $d_a(x) = \inf_{y\in\mathcal{M}_a} \|x - y\|$, and assume $\|z\| \le R$ on every orbit. For unit features, define the semantic orbit separation

$$\Delta_{\mathrm{orbit}} = \inf_{a\neq b} \inf_{y\in\mathcal{M}_a,\, z\in\mathcal{M}_b} (1 - \langle y, z\rangle). \tag{15}$$

It ignores pose differences within the same object and measures the closest cross-object collision. Sampled banks replace the two infima by minima.

For either normalized counting measure or normalized quotient Haar measure, write

$$u_a(x) = \log \pi_a + \log Z_a(x),$$
$$\Lambda_a(x) = u_a(x) - \log \sum_{b \neq a} e^{u_b(x)},$$
$$\Lambda_\star(M) = \inf_a \inf_{y \in \mathcal{M}_a} \Lambda_a(y). \tag{16}$$

The last quantity is the worst clean target-versus-competitor free-energy gap. It is defined directly for a continuous orbit and contains no atomic sampling weight.

Theorem 2 (Free-energy retrieval tube)**.** *Let $0 < \varepsilon < 1$. Suppose the logits used at readout satisfy $|\widetilde{u}_a(x) - u_a(x)| \le \zeta$ for every orbit. Define*

$$r^{\mathrm{H}}_{\mathrm{ret}}(M;\zeta) = \frac{\Lambda_\star(M) - 2\zeta - \log((1-\varepsilon)/\varepsilon)}{2\beta R}. \tag{17}$$

If this radius is positive, then

$$d_a(x) < r^{\mathrm{H}}_{\mathrm{ret}}(M;\zeta) \quad \Longrightarrow \quad \widetilde{P}_x(a) > 1 - \varepsilon. \tag{18}$$

For an exact continuous or sampled update, set $\zeta = 0$.

Operationally, $r^{\mathrm{H}}_{\mathrm{ret}}$ is the largest certified perturbation radius for which the target posterior exceeds $1 - \varepsilon$; the term 2ζ explicitly pays for numerical error in the implemented logits.

The proof uses the $2\beta R$-Lipschitz property of the free-energy gap and therefore applies to non-atomic Haar measure. The earlier atomic-weight radius is retained only as a finite-bank corollary in Appendix B.

Let $p_x(dz)$ be the posterior measure on the disjoint union of all memory orbits. Differentiation under the integral gives the same covariance identity for sampled and exact-Haar updates:

$$J_{F_{\mathrm{H}}}(x) = \beta \operatorname{Cov}_{p_x}(z). \tag{19}$$

Proposition 2 (Computable normal-bundle contraction)**.** *Let $\mathcal{M}_a$ be a compact C^3 embedded orbit with positive reach. A finite normal-bundle atlas, posterior-moment enclosures, and analytic off-grid remainders give computable $\widehat{b}_a \ge 0$ and $\widehat{\kappa}^{\mathrm{nor}}_a \ge 0$ such that, throughout a certified radius r_{loc} tube,*

$$d_a(F_{\mathrm{H}}(x)) \le \widehat{b}_a + \widehat{\kappa}^{\mathrm{nor}}_a d_a(x). \tag{20}$$

The derivative tested by the atlas is $DR_a(F_{\mathrm{H}}(x)) J_{F_{\mathrm{H}}}(x) \Pi^{\perp}_a(\pi_a(x))$, not the full ambient Jacobian. Thus tangent pose transport is not counted as retrieval instability.

Operationally, $\widehat{\kappa}^{\mathrm{nor}}_a < 1$ certifies contraction toward the target orbit, while $\widehat{b}_a$ is the residual geometric bias. Motion along the orbit is intentionally excluded because it changes pose, not identity.

Appendix C gives the complete finite-atlas construction. For an implemented continuous step $\widetilde{F}_t$ with uniform state error e_t, Eq. (20) implies

$$d_a(\widetilde{F}_t(x)) \le \widehat{b}_a + e_t + \widehat{\kappa}^{\mathrm{nor}}_a d_a(x). \tag{21}$$

Theorem 3 (Gaussian strict-collapse certificate)**.** *Let $0 < \kappa < 1$ upper-bound the certified normal coefficients and suppose Eq.* (21) *is forward invariant. Set*

$$B^{\mathrm{num}}_{a,T} = \sum_{t=0}^{T-1} \kappa^{T-1-t}(\widehat{b}_a + e_t). \tag{22}$$

If $r^{\mathrm{H}}_{\mathrm{ret}}(M;\zeta_T) > B^{\mathrm{num}}_{a,T}$, define

$$R^{\mathrm{cont}}_{M,a}(T) = \min\left\{ r_{\mathrm{loc}},\, \kappa^{-T}\left(r^{\mathrm{H}}_{\mathrm{ret}}(M;\zeta_T) - B^{\mathrm{num}}_{a,T} \right) \right\}. \tag{23}$$

For $x_0 = y_0 + \eta_0$, y_0 on target orbit a, and $\eta_0 \sim \mathcal{N}(0, \Sigma_0)$ with $\|\Sigma_0\|_2 \le \sigma^2$, let $\widetilde{\tau}^{\mathrm{strict}}_\varepsilon = \inf\{s : \widetilde{P}_s(a) > 1 - \varepsilon\}$. Then

$$\Pr(\widetilde{\tau}^{\mathrm{strict}}_\varepsilon > T) \le \overline{F}_{\chi^2_d}\left(\frac{[R^{\mathrm{cont}}_{M,a}(T)]^2}{\sigma^2} \right), \tag{24}$$

where strict success is evaluated using the implemented logits and ζ_T bounds their final-step error.

This result turns the deterministic contraction certificate into a computable failure-probability bound: dimension, noise level, update depth, geometric bias, and backend error enter only through the certified radius and a chi-square tail.

The chi-square tail is exact for the norm comparison under isotropic covariance. Appendix D derives the contraction-to-probability step, while Appendix B.1 propagates certified backend errors.

Theorem 4 (Continuous-Haar capacity certificate)**.** *Assume the hypotheses above hold uniformly over targets, let $q_d^{\chi}(\delta)$ be the upper chi-square norm quantile, and set*

$$B_T^{\mathrm{num}} = \max_a \sum_{t=0}^{T-1} \kappa^{T-1-t}(\widehat{b}_a + e_t), \qquad \Gamma_T^{\mathrm{num}} = \kappa^T \sigma q_d^{\chi}(\delta) + B_T^{\mathrm{num}}. \tag{25}$$

If

$$r_{\mathrm{loc}} \geq \sigma q_d^{\chi}(\delta), \tag{26}$$

$$\Lambda_\star(M) > \log \frac{1-\varepsilon}{\varepsilon} + 2\beta R \Gamma_T^{\mathrm{num}} + 2\zeta_T, \tag{27}$$

then the implemented T-step continuous-Haar retrieval succeeds with probability at least $1-\delta$, uniformly over targets and target poses.

The certificate is a practical admission test for a prescribed memory bank: local geometry must contain the likely noise, and the worst orbit-level gap must pay for contracted noise, accumulated bias, and numerical error. Eq. (29) below makes the corresponding limit on M explicit.

There is no separation-free universal capacity law: coincident semantic orbits cannot be strictly distinguished. Theorem 4 is therefore a sufficient bank-specific certificate. For unit memories and a uniform orbit prior, define the continuous within-orbit entropy penalty

$$H_M^{\mathrm{H}}(\beta) = \sup_{a, y \in \mathcal{M}_a} -\log \int_{G/H_a} e^{-\frac{\beta}{2}\|z_a(\omega)-y\|^2}\, d\nu_a(\omega). \tag{28}$$

Then $\Lambda_\star(M) \geq \beta\Delta_{\mathrm{orbit}} - H_M^{\mathrm{H}}(\beta) - \log(M-1)$, giving the explicit sufficient condition

$$M < 1 + \frac{\varepsilon}{1-\varepsilon} \exp\left[\beta(\Delta_{\mathrm{orbit}} - 2R\Gamma_T^{\mathrm{num}}) - H_M^{\mathrm{H}}(\beta) - 2\zeta_T\right]. \tag{29}$$

For a smooth q-dimensional quotient orbit with a uniform small-ball lower bound, $H_M^{\mathrm{H}}(\beta) \leq (q/2)\log\beta + O(1)$. Thus intrinsic orbit dimension replaces the artificial $\log Q$ atomic penalty; all geometric constants and Δ_{orbit} may still depend on the prescribed bank. Controlled sweeps confirm these directions but do not claim tightness.

5 Experiments

5.1 Protocol and comparison boundary

Experiments ran on a system with an Intel Core i9-14900HX CPU and an NVIDIA GeForce RTX 4070 Laptop GPU (8 GB); continuous-backend latency was measured on the CPU.

All hyperparameters use validation data and all banks use training features. Learned benchmarks report mean and sample SD over three seeds; rMD17 uses five official splits. Learned features are L_2-normalized, rMD17 coordinates are centered in Angstrom, and stated TTA timings use matched view budgets.

The continuous study uses public rMD17 trajectories and official 1,000/1,000 splits (Christensen & Von Lilienfeld, 2020). Per split, a 50-frame buffer separates 800 design structures from validation candidates; farthest-point traversal selects $M = 32$ training prototypes in atom-pair-distance space. Rotated out-of-gallery queries reveal four or six of nine atoms with 0.05 or 0.15 Angstrom noise, and validation selects β, T by hidden-atom RMSD. Reused trajectories make split means protocol replications. Timing freezes the smallest grid matching the matrix-Fisher $T = 3$ trajectory and posterior within 10^{-3}, then tests five queries; it includes all updates and the terminal posterior but not feature construction. Appendix H.1 retains the 10^{-6} synthetic audit.

Method	Balanced accuracy	ms/query
Sampled G-MHN, $T = 2$	$\mathbf{0.2652 \pm 0.0080}$	$\mathbf{5.03 \pm 0.17}$
Initial orbit evidence	0.1856 ± 0.0069	–
Nearest sampled pose	0.1848 ± 0.0052	–
Prototype mean	0.2023 ± 0.0149	–
3-view TTA	0.2045 ± 0.0079	11.91 ± 0.41

Table 1: Held-out ModelNet40 masked latent completion.

Molecule	observed	noise	nearest pose	continuous mean
ethanol	4	0.05	0.8041 ± 0.0879	$\mathbf{0.6204 \pm 0.0456}$
	4	0.15	0.8870 ± 0.0429	$\mathbf{0.6860 \pm 0.0252}$
	6	0.05	0.5067 ± 0.0305	$\mathbf{0.4348 \pm 0.0589}$
	6	0.15	0.5201 ± 0.0788	$\mathbf{0.4638 \pm 0.0580}$
malonaldehyde	4	0.05	0.9932 ± 0.0655	$\mathbf{0.7897 \pm 0.0345}$
	4	0.15	1.1431 ± 0.0831	$\mathbf{0.8839 \pm 0.0380}$
	6	0.05	0.6991 ± 0.0807	$\mathbf{0.6090 \pm 0.1049}$
	6	0.15	0.7340 ± 0.1055	$\mathbf{0.6235 \pm 0.0712}$

Table 2: Out-of-gallery rMD17 hidden-atom RMSD in Angstrom (mean $\pm$ sample SD over five official splits). Lower is better. Each condition uses 40 test queries per split; split means, not individual frames, are the replication units.

5.2 Controlled and established retrieval results

Image restoration. Two-step retrieval restores corrupted poses of stored Fashion-MNIST and CIFAR-10 exemplars from an augmentation-matched bank; this is an associative-memory visualization, not a supervised benchmark (Appendix G).

ModelNet40 masked completion. A frozen spherical-MLP encodes held-out queries from 11 classes under independent rotation, noise, and 80% masking. Across three banks, two G-MHN steps reach 0.2652 ± 0.0080 balanced accuracy (Table 1), versus 0.2045 ± 0.0079 for matched three-view TTA ($p = 0.0088$), while taking 5.03 ± 0.17 rather than 11.91 ± 0.41 ms/query. Initial orbit evidence and nearest sampled pose reach only 0.1856 and 0.1848, respectively, showing the gain associated with the recurrent update.

Theory audits. At $M = 12, \beta = 24, T = 3$, increasing Δ_{orbit} from 0.12 to 0.48 raises the clean gap from 0.310 to 8.176 and strict success from 0 to 1; increasing Q from 32 to 1024 lowers success from 1 to 0.797, and matched one-, two-, and three-dimensional orbits yield 1, 0.849, and 0. The exact sampled-bank Gaussian audit uses $(\widehat{\kappa}, b_{\max}, R^{\mathrm{bias}}) = (0.48550, 1.62 \times 10^{-7}, 0.18)$: at $\sigma = 0.09$, its 0.26146 chi-square bound compares with 0.00110 empirical failure and no failure inside the certified radius. It checks only the conservative $e_t = \zeta_T = 0$ finite-bank chain; continuous tests audit numerical errors, not a joint rMD17 capacity certificate. The chi-square bound is therefore valid but conservative for this retrieval event.

5.3 Continuous rMD17 completion and matched-accuracy scaling

Hidden-atom completion. The continuous backend integrates all global rotations of each stored conformation through corresponded observed-atom channels. Its conditional mean beats nearest-pose Procrustes retrieval in all 40 cells (Table 2), reducing condition-averaged RMSD by 10.8–22.8% and agreeing with independent soft Procrustes within 0.0022 Angstrom.

Matched-accuracy scaling. With $M = 32, \beta = 25, T = 3$, four observed atoms, and noise 0.05, both molecules first meet 10^{-3} state-RMSD and posterior-TV tolerances at $Q = 80^3 = 512{,}000$ (Figure 3, left). Across five test queries, continuous versus sampled medians are 15.31 versus 2615.97 ms for ethanol and 17.95 versus 2888.23 ms for malonaldehyde. Paired speedups are geometric

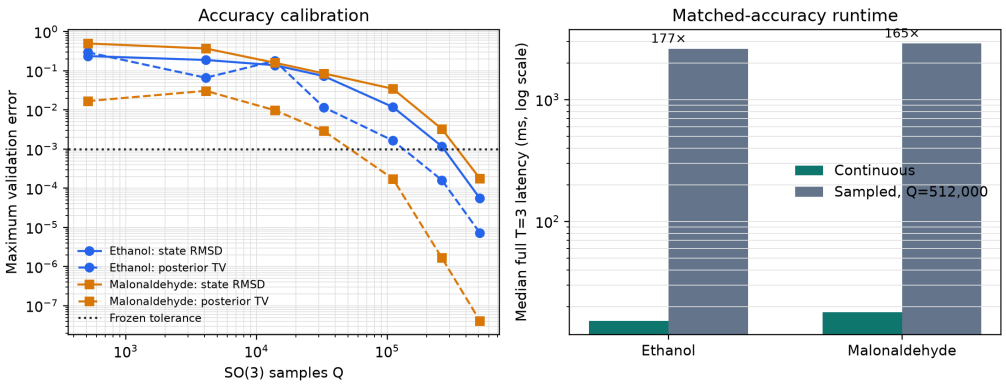


Figure 3: Real-data matched-accuracy scaling. Left: validation error of sampled $\mathrm{SO}(3)$ grids against the continuous three-step trajectory and terminal posterior; both frozen criteria first pass at $Q = 512{,}000$. Right: test latency at that frozen budget on a logarithmic scale. Timings cover the full three-step kernel and terminal posterior, excluding feature construction.

means of $177.23\times$ (95% CI $[154.08, 201.15]$) and $165.17\times$ ($[147.54, 186.12]$); intervals resample held-out queries.

The 32 prototypes occupy 6,912 bytes, versus 40.96 MB for the factorized grid and weights or 3.54 GB for an expanded bank, so continuous evaluation reduces latency and persistent storage in this high-accuracy regime.

What recurrent steps add. The Bayes-optimal conditional mean minimizes squared reconstruction error; recurrence sharpens modes. Across all 40 cells it lowers energy, entropy, and nearest-orbit distance; with four observed atoms, pairwise-distance MAE falls by 0.052–0.095 Angstrom while hidden-coordinate RMSD rises by 0.037–0.093 Angstrom. Thus $T = 0$ favors mean reconstruction and $T > 0$ a stored mode (Appendix J).

5.4 Controlled numerical audits and boundary controls

Both continuous backends match strict interval references near machine precision. Cached grids are faster in regular cases, but the pre-specified concentrated $\mathrm{SO}(3)$ audit needs $Q = 262{,}144$, where matrix-Fisher is $3.62\times$ faster (95% CI $[3.48, 3.79]$) and uses 3 KiB rather than 20.0 MiB (Appendix H.1). Adverse controls show no expected gain when ScanObjectNN memories are insufficiently separated or a QM9 encoder is already invariant (Appendix I), restricting the method to informative group-structured representations and orbit-level retrieval.

6 Conclusion

G-MHN retrieves semantic orbits rather than individual poses. Sampled updates and continuous Fourier/Bessel $\mathrm{SO}(2)$ and matrix-Fisher $\mathrm{SO}(3)$ backends share one orbit-level rule, supported by conditional retrieval and sufficient capacity certificates. Experiments show gains over matched TTA on ModelNet40 and consistent rMD17 completion gains. At rMD17's frozen accuracy budget of 512,000 rotations, continuous evaluation removes grid storage and accelerates full trajectories by over two orders of magnitude.

These gains require a known action, informative equivariant features, and separated semantic orbits. The continuous backends require finite $\mathrm{SO}(2)$ frequencies or fixed-correspondence $\ell = 1$ $\mathrm{SO}(3)$ channels; cached low-Q grids can be faster. rMD17 measures coordinate completion, not forces, on only two nine-atom molecules with known correspondence, CPU kernels, reused trajectories, and five timing queries. ScanObjectNN and QM9 controls show that orbit memory cannot create missing margins or improve already invariant features. Larger systems and GPU implementations remain open.

AI USE STATEMENT

Generative AI assisted with theory and proof drafting, experiments, debugging, code, figures, and English editing. The authors originated the work, verified all claims, proofs, protocols, and code paths, and take responsibility for the content.

REPRODUCIBILITY STATEMENT

Algorithm 1 specifies the shared update. The appendix documents assumptions, proofs, splits, prototype selection, normalization, validation, seeds, results, and implementation boundaries. The anonymous supplement provides scripts and raw outputs for every reported benchmark, figure-generation code, environment notes, and 39 passing unit and regression tests.

ACKNOWLEDGMENTS

Acknowledgments omitted for anonymous review.

REFERENCES

Sofiya Bogakovskaya, Vladimir Fanaskov, and Ivan Oseledets. Dynamics of modern hopfield networks. In *New Frontiers in Associative Memories-Workshop at ICLR 2026*, 2026.

Yiyang Chen, Lunhao Duan, Shanshan Zhao, Changxing Ding, and Dacheng Tao. Local-consistent transformation learning for rotation-invariant point cloud analysis. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5418–5427. IEEE, 2024.

Anders S Christensen and O Anatole Von Lilienfeld. On the role of gradients for machine learning of molecular energies and forces. *Machine Learning: Science and Technology*, 1(4):045018, 2020.

Anders S. Christensen and O. Anatole von Lilienfeld. Revised MD17 dataset (rMD17). figshare, version 4, 2020.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016.

Taco S Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant cnns on homogeneous spaces. *Advances in neural information processing systems*, 32, 2019.

Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in neural information processing systems*, 33: 1970–1981, 2020.

John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554--2558, 1982.

Jaein Kim, Hee Bin Yoo, Dong-Sig Han, Yeon-Ji Song, and Byoung-Tak Zhang. Continuous so (3) equivariant convolution for 3d point cloud analysis. In *European Conference on Computer Vision*, pp. 59–75. Springer, 2024.

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

Taeyoung Lee. Bayesian attitude estimation with the matrix fisher distribution on so (3). *IEEE Transactions on Automatic Control*, 63(10):3377–3392, 2018.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In *International Conference on Learning Representations*, 2023.

Yi-Lun Liao, Brandon Wood, Abhishek Das, and Tess Smidt. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. In *International Conference on Learning Representations*, pp. 39282–39309, 2024.

Tsubasa Masumura and Masato Taki. On the role of hidden states of modern hopfield network in transformer. *Advances in Neural Information Processing Systems*, 38:68607–68626, 2026.

Beren Millidge, Tommaso Salvatori, Yuhang Song, Thomas Lukasiewicz, and Rafal Bogacz. Universal Hopfield networks: A general framework for single-shot associative memory models. In *International Conference on Machine Learning*, pp. 15561–15583. PMLR, 2022.

Kazushi Mimura, Junichi Takeuchi, Yuto Sumikawa, Yoshiyuki Kabashima, and Anthony CC Coolen. Dynamical properties of dense associative memory. In *International Conference on Learning Representations*, volume 2026, pp. 19723–19744, 2026.

Michael Murray, Tenzin Chan, Kedar Karhadker, and Christopher J Hillar. Implicit bias and invariance: How hopfield networks efficiently learn graph orbits. *arXiv preprint arXiv:2512.14338*, 2025.

NIST Digital Library of Mathematical Functions. Modified bessel functions: Generating function and associated series. Section 10.35, `https://dlmf.nist.gov/10.35`, 2026. Accessed 2026-09-14.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017.

Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):140022, 2014.

Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Lukas Gruber, Markus Holzleitner, Thomas Adler, David Kreil, Michael K Kopp, et al. Hopfield networks is all you need. In *International Conference on Learning Representations*, 2021.

Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Thanh Nguyen, and Sai-Kit Yeung. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1588–1597. IEEE, 2019.

Shurong Wang, Yuqi Pan, Zhuoyang Shen, Meng Zhang, Hongwei Wang, and Guoqi Li. Adaptive hopfield network: Rethinking similarities in associative memory. In *International Conference on Learning Representations*, volume 2026, pp. 116942–116974, 2026a.

Tianwei Wang, Xinhui Ma, and Wei Pang. Asymptotically stable quaternion-valued hopfield-structured neural network with periodic projection-based supervised learning rules. *Advances in Neural Information Processing Systems*, 38:87040–87063, 2026b.

Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, 38(5): 1–12, 2019.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1912–1920, 2015.

Runze Xia and Richard Jiang. Energy landscape-aware vision transformers: Layerwise dynamics and adaptive task-specific training via hopfield states. *Advances in Neural Information Processing Systems*, 38:12509–12532, 2026.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

APPENDIX: NOTATION

Table 3 summarizes the symbols used in the main text and proofs. Local indices and backend-specific coefficients are defined where they first appear.

Table 3: Notation and definitions.

Symbol	Definition
M, d, T	Number of semantic memory orbits, state dimension, and recurrent update steps.
G, ρ	Finite or compact group and its orthogonal action $\rho : G \to O(d)$.
$\xi^a, \mathcal{M}_a$	Prototype of object a and its orbit $\{\rho(g)\xi^a : g \in G\}$.
H_a, Ω_a	Stabilizer of ξ^a and quotient G/H_a indexing distinct poses.
ν_a, π_a	Normalized quotient measure and prior weight of orbit a.
$\pi_a(x), \Pi_a^{\perp}$	Nearest point on orbit a and the corresponding normal projector.
$Q_a, g_{a,q}, w_{a,q}$	Sample count, pose representatives, and normalized positive weights.
$z_{a,q}, z_a(\omega)$	Sampled atom and continuous orbit point transformed from ξ^a.
β, x_t, F	Inverse temperature, state at step t, and orbit update map.
Z_a, Z_a^{H}	Sampled orbit evidence and exact normalized-Haar evidence.
$\mathcal{K}, c_{a,0}, r_{a,k}$	Frequency set, invariant score offset, and harmonic amplitudes for $\mathrm{SO}(2)$.
$C_{a,\ell}$	Fourier coefficients used for exact $\mathrm{SO}(2)$ orbit evidence and pose moments.
X_a, Y, K_a	$\mathrm{SO}(3)$ prototype and query channels, with $K_a = \beta Y^\top X_a$.
$c(K_a), \mathbb{E}[R \mid a, Y]$	Matrix-Fisher normalizer and conditional mean rotation.
$p_x(q \mid a), \mu_a(x)$	Conditional pose posterior and within-orbit mean.
$P_x(a), \widetilde{P}_x(a)$	Orbit posterior and its implemented approximation.
$u_a, \Lambda_a, \Lambda_\star$	Orbit logit, target-versus-competitor gap, and worst clean gap.
E_G, J_F	Orbit energy and Jacobian of the update.
$\Delta_{\mathrm{orbit}}, d_a(x)$	Separation between distinct orbits and distance from x to $\mathcal{M}_a$.
$r_{\mathrm{loc}}, r_{\mathrm{ret}}^{\mathrm{H}}$	Certified local tube radius and strict-retrieval radius.
$\widehat{b}_a, \widehat{\kappa}_a^{\mathrm{nor}}$	Residual bias and certified normal contraction coefficient.
$e_t, \zeta_t, B_{a,T}^{\mathrm{num}}$	State and logit errors, and accumulated numerical bias.
$\Gamma_T^{\mathrm{num}}, H_M^{\mathrm{H}}(\beta)$	Noise-plus-bias radius and within-orbit entropy penalty.
q	Intrinsic dimension of a smooth quotient orbit in the capacity estimate.
$\Sigma_0, \sigma, \delta, \varepsilon$	Initial noise covariance, its norm scale, failure tolerance, and posterior threshold.
$q_d^{\chi}(\delta), \overline{F}_{\chi_d^2}$	Chi-square norm quantile and chi-square upper-tail function.

APPENDIX CONTENTS

Appendix A establishes the quotient measures and sampled and continuous updates. Appendix B proves the structural identities. Appendices C–E give contraction, retrieval, capacity, and Gaussian-noise derivations. Appendices F–J provide protocols, visualizations, complete results, boundary controls, and the continuous rMD17 study.

A Measure Conventions and Sampled/Continuous Computation 14

A.1 Exact continuous orbit objective 15

A.2 Multi-frequency $\mathrm{SO}(2)$ backend 16

A.3 Fixed-correspondence $\mathrm{SO}(3)$ backend 17

B Structural Proofs and Strict Retrieval Tube 18

B.1 Complete dynamics and error propagation 20

B.2 Complexity and applicability 20

C Computable Local Contraction Certificates 21

D Retrieval and Capacity Proofs 25

E Gaussian-Noise Orbit Search Derivation 31

F Experimental Protocol Details 32

G Image-Restoration Protocol and Qualitative Visualizations 37

H Extended Result Tables and Plots 38

H.1 Validation-frozen continuous comparison protocol 44

H.2 Complete regular-condition results 46

H.3 High-concentration $\mathrm{SO}(3)$ crossover 46

H.4 Interpretation and remaining boundary 47

I Real-Data Protocols and Complete Results 48

I.1 ScanObjectNN 48

I.2 QM9 48

I.3 Interpretation and non-comparability 49

J rMD17 Continuous Retrieval and Completion 49

J.1 Dataset, splits, and leakage control 49

J.2 Methods and validation selection 50

J.3 Complete hidden-coordinate results 50

J.4 Matched-accuracy continuous versus sampled backend 50

J.5 Mean-to-mode behavior of recurrent retrieval 51

A Measure Conventions and Sampled/Continuous Computation

This appendix fixes the quotient measures, orbit priors, and numerically stable sampled computation used by the group-orbit update and Algorithm 1 of the main paper.

Definition A.1 (Quotient-orbit memory bank)**.** *Let G be a finite or compact group and let $\rho : G \to O(d)$ be an orthogonal representation. For base memory ξ^a, set*

$$H_a = \mathrm{Stab}(\xi^a) = \{h \in G : \rho(h)\xi^a = \xi^a\}, \tag{A.1}$$

$$\Omega_a = G/H_a, \tag{A.2}$$

$$z_a(\omega) = \rho(g_\omega)\xi^a, \tag{A.3}$$

$$\mathcal{O}_{\xi^a} = \{z_a(\omega) : \omega \in \Omega_a\}. \tag{A.4}$$

The measure ν_a is the normalized counting measure for finite quotients and the normalized quotient Haar measure for compact quotients. An orbit prior $\pi_a > 0$, $\sum_a \pi_a = 1$, may be used to encode semantic prior weights; the default is uniform.

Definition A.2 (Quotient-orbit Hopfield update)**.** *For query x, define*

$$s_a(\omega; x) = \beta\langle z_a(\omega), x\rangle, \tag{A.5}$$

$$Z_a(x) = \int_{\Omega_a} e^{s_a(\omega;x)}\, d\nu_a(\omega), \tag{A.6}$$

$$Z(x) = \sum_{b=1}^{M} \pi_b Z_b(x). \tag{A.7}$$

The orbit posterior and conditional orbit representative are

$$P_x(a) = \frac{\pi_a Z_a(x)}{Z(x)}, \tag{A.8}$$

$$p_x(d\omega \mid a) = \frac{e^{s_a(\omega;x)}}{Z_a(x)}\, d\nu_a(\omega), \tag{A.9}$$

$$\mu_a(x) = \int_{\Omega_a} z_a(\omega)\, p_x(d\omega \mid a). \tag{A.10}$$

The update is

$$F(x) = \sum_{a=1}^{M} P_x(a)\mu_a(x). \tag{A.11}$$

The orbit free energy is the negative log-evidence

$$\mathcal{E}_a(x) = -\frac{1}{\beta}\log Z_a(x). \tag{A.12}$$

Including the orbit prior gives the prior-adjusted free energy

$$\widetilde{\mathcal{E}}_a(x) = \mathcal{E}_a(x) - \frac{1}{\beta}\log \pi_a. \tag{A.13}$$

Equivalently,

$$P_x(a) = \frac{\exp(-\beta\widetilde{\mathcal{E}}_a(x))}{\sum_{b=1}^{M}\exp(-\beta\widetilde{\mathcal{E}}_b(x))}. \tag{A.14}$$

Lower $\widetilde{\mathcal{E}}_a$ indicates a stronger target-orbit or metastable candidate after accounting for the orbit prior.

For a finite group, ν_a is normalized counting measure on $\Omega_a = G/H_a$, so all quotient-orbit integrals in the main text are averages over distinct cosets. For a compact group, ν_a is the normalized quotient measure induced by Haar measure. This measure is invariant under left multiplication and is the continuous analogue of a normalized average over distinct poses.

For a sampled quotient orbit $z_{a,q} = \rho(g_{a,q})\xi^a$, $q = 1, \dots, Q$, with optional weights $w_{a,q}$, compute

$$s_{a,q} = \beta\langle z_{a,q}, x\rangle + \log w_{a,q}. \tag{A.15}$$

Let

$$c_a = \max_q s_{a,q}, \tag{A.16}$$

$$\widetilde{e}_{a,q} = \exp(s_{a,q} - c_a). \tag{A.17}$$

Then

$$\log Z_a(x) = c_a + \log \sum_{q=1}^{Q} \widetilde{e}_{a,q}, \tag{A.18}$$

$$\mu_a(x) = \frac{\sum_q \widetilde{e}_{a,q} z_{a,q}}{\sum_q \widetilde{e}_{a,q}}. \tag{A.19}$$

After computing these quantities for $a = 1, \dots, M$, let

$$\mathrm{LSE}_b\, u_b = \log \sum_{b=1}^{M} e^{u_b}. \tag{A.20}$$

Then

$$u_a(x) = \log \pi_a + \log Z_a(x), \tag{A.21}$$

$$\log Z(x) = \mathrm{LSE}_b\ u_b(x), \tag{A.22}$$

$$\ell_a(x) = u_a(x) - \log Z(x), \tag{A.23}$$

$$P_x(a) = \exp\big(\ell_a(x)\big). \tag{A.24}$$

The sampled quantities are substituted into the update in Eq. (A.11). This orbit-wise computation is algebraically equivalent to a single softmax over all sampled orbit representatives, but stores only $\log Z_a$, $P_x(a)$, and μ_a at the orbit level.

A.1 Exact continuous orbit objective

The two continuous backends use the compact rotation groups

$$\mathrm{SO}(2) := \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} : \theta \in [0, 2\pi) \right\}, \tag{A.25}$$

$$\mathrm{SO}(3) := \{R \in \mathbb{R}^{3\times 3} : R^\top R = I_3,\ \det R = 1\}. \tag{A.26}$$

Normalized Haar measure is $d\theta/(2\pi)$ on $\mathrm{SO}(2)$ and the unique probability rotation measure dR on $\mathrm{SO}(3)$. These definitions make the phrase "all rotations" precise without selecting a pose grid.

Let G be compact, let dg be normalized Haar measure, and let $\rho : G \to O(d)$. The orbit map is constant on right cosets of $H_a = \mathrm{Stab}(\xi^a)$, so the pushforward of dg is the normalized invariant measure ν_a on G/H_a. Consequently,

$$\int_{G/H_a} f(\rho(g)\xi^a)\, d\nu_a(gH_a) = \int_G f(\rho(g)\xi^a)\, dg. \tag{A.27}$$

For a query x, define

$$Z_a^{\mathrm{H}}(x) = \int_G e^{\beta\langle \rho(g)\xi^a, x\rangle}\, dg, \tag{A.28}$$

$$\mu_a^{\mathrm{H}}(x) = \frac{1}{Z_a^{\mathrm{H}}(x)} \int_G \rho(g)\xi^a e^{\beta\langle \rho(g)\xi^a, x\rangle}\, dg, \tag{A.29}$$

$$P_x^{\mathrm{H}}(a) = \frac{\pi_a Z_a^{\mathrm{H}}(x)}{\sum_b \pi_b Z_b^{\mathrm{H}}(x)}, \qquad F_{\mathrm{H}}(x) = \sum_a P_x^{\mathrm{H}}(a)\mu_a^{\mathrm{H}}(x). \tag{A.30}$$

These equations define the mathematical target. Here *continuous* means that the production update does not instantiate group poses. It does not imply that every backend is an elementary closed form.

The continuous-Haar update is the invariant-measure case of the joint equivariance and energy theorem in Appendix B. Consequently,

$$Z_a^{\mathrm{H}}(\rho(k)x) = Z_a^{\mathrm{H}}(x), \qquad P^{\mathrm{H}}_{\rho(k)x}(a) = P_x^{\mathrm{H}}(a), \tag{A.31}$$

$$F_{\mathrm{H}}(\rho(k)x) = \rho(k)F_{\mathrm{H}}(x), \qquad \nabla E_{\mathrm{H}}(x) = x - F_{\mathrm{H}}(x), \tag{A.32}$$

where

$$E_{\mathrm{H}}(x) = \frac{1}{2}\|x\|^2 - \frac{1}{\beta}\log\sum_a \pi_a Z_a^{\mathrm{H}}(x) + C. \tag{A.33}$$

A.2 Multi-frequency SO(2) backend

Every finite-dimensional real orthogonal representation of $\mathrm{SO}(2)$ admits an orthogonal decomposition

$$V = V_0 \oplus \bigoplus_{k\in\mathcal{K}} (\mathbb{R}^2)^{m_k}, \tag{A.34}$$

where V_0 is fixed and the k-th block transforms by the planar rotation $R(k\theta)$. For prototype blocks $u_{a,k,j}$ and query blocks $v_{k,j}$, the score coefficients are

$$A_{a,k} = \sum_j u_{a,k,j}^\top v_{k,j}, \tag{A.35}$$

$$B_{a,k} = \sum_j (Ju_{a,k,j})^\top v_{k,j}, \qquad r_{a,k} = (A_{a,k}^2 + B_{a,k}^2)^{1/2}, \tag{A.36}$$

$$\phi_{a,k} = \mathrm{atan2}(B_{a,k}, A_{a,k}), \tag{A.37}$$

where J is a positive quarter turn. Hence

$$\langle \rho(\theta)\xi^a, x\rangle = c_{a,0} + \sum_{k\in\mathcal{K}} r_{a,k}\cos(k\theta - \phi_{a,k}). \tag{A.38}$$

Set $\kappa_{a,k} = \beta r_{a,k}$. The modified-Bessel Jacobi–Anger identity (NIST Digital Library of Mathematical Functions, 2026) gives

$$e^{\kappa_{a,k}\cos(k\theta-\phi_{a,k})} = \sum_{n\in\mathbb{Z}} I_n(\kappa_{a,k}) e^{in(k\theta-\phi_{a,k})}. \tag{A.39}$$

Convolving these finitely many coefficient sequences defines

$$\prod_{k\in\mathcal{K}} e^{\kappa_{a,k}\cos(k\theta-\phi_{a,k})} = \sum_{\ell\in\mathbb{Z}} C_{a,\ell} e^{i\ell\theta}. \tag{A.40}$$

Integration with $d\theta/(2\pi)$ selects the zero coefficient, while multiplication by $e^{i\ell\theta}$ selects coefficient $-\ell$:

$$Z_a^{\mathrm{H}}(x) = e^{\beta c_{a,0}} C_{a,0}, \tag{A.41}$$

$$\mathbb{E}[e^{i\ell\theta} \mid a, x] = \frac{C_{a,-\ell}}{C_{a,0}}. \tag{A.42}$$

Eq. (A.42) supplies exactly the moments needed to rotate each frequency block. To make the update explicit, set

$$m_{a,k}(x) := \mathbb{E}[e^{ik\theta} \mid a, x] = \frac{C_{a,-k}}{C_{a,0}}, \qquad \mathcal{R}(m) := \begin{pmatrix} \Re m & -\Im m \\ \Im m & \Re m \end{pmatrix}. \tag{A.43}$$

Writing the prototype in the decomposition of Eq. (A.34) as $\xi^a = (u_{a,0}, \{u_{a,k,j}\}_{k,j})$, the conditional orbit mean is

$$\mu_a^{\mathrm{H}}(x) = \left(u_{a,0}, \{\mathcal{R}(m_{a,k}(x))u_{a,k,j}\}_{k,j}\right). \tag{A.44}$$

Truncation and numerical enclosure. The implementation uses scaled coefficients

$$b_n(\kappa) = e^{-\kappa} I_n(\kappa), \qquad b_0(\kappa) + 2\sum_{n\geq 1} b_n(\kappa) = 1. \tag{A.45}$$

If factor k is truncated to $|n| \leq N_k$, its omitted mass is

$$\tau_k = 2\sum_{n>N_k} b_n(\kappa_{a,k}), \qquad \varepsilon_C \leq 1 - \prod_{k\in\mathcal{K}} (1-\tau_k). \tag{A.46}$$

For $\widetilde{C}_{a,0} > \varepsilon_C$,

$$\left|\log C_{a,0} - \log \widetilde{C}_{a,0}\right| \leq -\log\left(1 - \frac{\varepsilon_C}{\widetilde{C}_{a,0}}\right). \tag{A.47}$$

For any required moment,

$$\left|\frac{C_{a,-\ell}}{C_{a,0}} - \frac{\widetilde{C}_{a,-\ell}}{\widetilde{C}_{a,0}}\right| \leq \frac{\varepsilon_C(\widetilde{C}_{a,0} + |\widetilde{C}_{a,-\ell}|)}{(\widetilde{C}_{a,0} - \varepsilon_C)\widetilde{C}_{a,0}}. \tag{A.48}$$

The certified path uses Arb/Acb ball arithmetic. If the requested enclosure cannot be resolved within the precision or order budget, it raises rather than returning an uncertified value.

A.3 Fixed-correspondence $\mathrm{SO}(3)$ backend

Let $X_a, Y \in \mathbb{R}^{n\times 3}$ contain prototype and query vector channels with fixed correspondence. The group action is

$$X_a \longmapsto X_a R^\top, \qquad R \in \mathrm{SO}(3). \tag{A.49}$$

The transformations that leave prototype a unchanged form

$$H_a = \mathrm{Stab}(X_a) = \{R \in \mathrm{SO}(3) : X_a R^\top = X_a\}, \tag{A.50}$$

and distinct poses are indexed by $\mathrm{SO}(3)/H_a$. The Frobenius score is

$$\beta\langle X_a R^\top, Y\rangle_F = \mathrm{tr}(K_a^\top R), \qquad K_a = \beta Y^\top X_a. \tag{A.51}$$

Thus the within-orbit posterior is a matrix-Fisher density (Lee, 2018). Take the proper SVD

$$K_a = USV^\top, \qquad U, V \in \mathrm{SO}(3), \qquad S = \mathrm{diag}(s_1, s_2, s_3), \tag{A.52}$$

with $s_1 \geq s_2 \geq |s_3|$. Haar bi-invariance gives the orbit evidence explicitly as

$$Z_a^{\mathrm{H}}(Y) = \int_{\mathrm{SO}(3)} e^{\mathrm{tr}(K_a^\top R)}\, dR = c(K_a) = c(S). \tag{A.53}$$

Define

$$A(u) = \tfrac{1}{2}(s_1 - s_2)(1-u), \qquad B(u) = \tfrac{1}{2}(s_1 + s_2)(1+u). \tag{A.54}$$

The normalized scalar reduction is

$$c(S) = \frac{1}{2}\int_{-1}^{1} I_0(A(u)) I_0(B(u)) e^{s_3 u}\, du. \tag{A.55}$$

Its three derivatives can be evaluated in the same quadrature:

$$\partial_{s_1} c = \frac{1}{4}\int_{-1}^{1} \big[(1-u)I_1(A)I_0(B) + (1+u)I_0(A)I_1(B)\big] e^{s_3 u}\, du, \tag{A.56}$$

$$\partial_{s_2} c = \frac{1}{4}\int_{-1}^{1} \big[-(1-u)I_1(A)I_0(B) + (1+u)I_0(A)I_1(B)\big] e^{s_3 u}\, du, \tag{A.57}$$

$$\partial_{s_3} c = \frac{1}{2}\int_{-1}^{1} u I_0(A) I_0(B) e^{s_3 u}\, du. \tag{A.58}$$

Writing $d_i = \partial_{s_i} \log c(S)$, the conditional rotation and memory mean are

$$\overline{R}_a(Y) := \mathbb{E}[R \mid a, Y] = U\,\mathrm{diag}(d_1, d_2, d_3) V^\top, \tag{A.59}$$

$$\mu_a^{\mathrm{H}}(Y) = X_a \overline{R}_a(Y)^\top. \tag{A.60}$$

Numerical paths. The ordinary path applies exponential scaling and endpoint-resolved adaptive quadrature. It reports a numerical error estimate, not a theorem. The strict path encloses the scalar integrals with Arb/Acb, encloses the floating-point proper-SVD residual, and propagates both through the matrix-Fisher posterior and orbit mixture. The certificate covers fixed-correspondence vector channels. It does not cover unknown point permutations, arbitrary higher $\mathrm{SO}(3)$ irreducible features, encoder error, or training error.

B Structural Proofs and Strict Retrieval Tube

This appendix provides detailed proofs for the structural statements and the strict-collapse certificate. Equivariance and energy invariance are stated for complete finite or compact quotient orbits with invariant reference measures. The strict finite-time retrieval tube is stated only for finite quotient-orbit banks or fixed sampled approximations.

Theorem B.1 (Joint equivariance and energy identity)**.** *For the Euclidean score, the compatibility identity*

$$\langle \rho(g)\xi, \rho(k)x\rangle = \langle \rho(k^{-1}g)\xi, x\rangle \quad \textit{for all } g,k,\xi,x \tag{B.1}$$

holds if and only if $\rho(k)^\top = \rho(k^{-1})$ for every k, or equivalently $\rho(G) \subseteq O(d)$. Assume this condition, a complete quotient-orbit bank, and a reference measure ν_a invariant under left multiplication on $\Omega_a = G/H_a$. Then

$$P_{\rho(k)x}(a) = P_x(a), \tag{B.2}$$

$$F(\rho(k)x) = \rho(k)F(x), \qquad k \in G, \tag{B.3}$$

and

$$E_G(x) = \frac{1}{2}\|x\|^2 - \frac{1}{\beta}\log Z(x) + C, \qquad Z(x) = \sum_a \pi_a Z_a(x), \tag{B.4}$$

$$E_G(\rho(k)x) = E_G(x), \tag{B.5}$$

$$\nabla_x E_G(x) = x - F(x). \tag{B.6}$$

The same conclusions hold for a finite weighted bank when the bank and its weights are closed under the action.

Proof. Taking $g = e$ in Eq. (B.1) and requiring the identity for every ξ, x gives $\rho(k) = \rho(k^{-1})^\top$, which is equivalent to $\rho(k)^\top\rho(k) = I_d$. Conversely, orthogonality and the representation law give Eq. (B.1) directly.

For each orbit a, left multiplication by k^{-1} is measure preserving on Ω_a. Eq. (B.1) therefore gives $Z_a(\rho(k)x) = Z_a(x)$, hence both $P_{\rho(k)x}(a) = P_x(a)$ and $Z(\rho(k)x) = Z(x)$. Moreover,

$$p_{\rho(k)x}(d\omega \mid a) = p_x(d(k^{-1}\omega) \mid a), \tag{B.7}$$

so

$$\mu_a(\rho(k)x) = \int_{\Omega_a} z_a(\omega)p_{\rho(k)x}(d\omega \mid a) = \rho(k)\mu_a(x). \tag{B.8}$$

Summing over orbit posteriors proves Eq. (B.3). For a closed weighted bank, the same argument is a weight-preserving permutation of its atoms.

Orthogonality preserves $\|x\|$, while the preceding argument preserves $Z(x)$; this proves energy invariance. Finally, differentiation under the finite sum or dominated integral yields

$$\nabla_x \log Z(x) = \beta \sum_a P_x(a)\mu_a(x) = \beta F(x), \tag{B.9}$$

and substitution into Eq. (B.4) proves Eq. (B.6). □

For the strict finite-bank certificate, define the semantic orbit margin and inter-orbit diameter jointly as

$$\Delta_{\mathrm{orbit}} = \inf_{a,\omega}\Big[\langle z_a(\omega), z_a(\omega)\rangle - \sup_{\substack{b\neq a\\ \upsilon\in\Omega_b}} \langle z_b(\upsilon), z_a(\omega)\rangle\Big], \tag{B.10}$$

$$D_{\mathrm{orbit}} = \max_{a\neq b} \sup_{\substack{\omega\in\Omega_a\\ \upsilon\in\Omega_b}} \|z_a(\omega) - z_b(\upsilon)\|. \tag{B.11}$$

The exact-pose margin, used only for finite pose grids or a prescribed pose resolution, is

$$\Delta_{\mathrm{pose}} = \inf_{a,\omega}\Big[\langle z_a(\omega), z_a(\omega)\rangle - \sup_{(b,\upsilon)\neq(a,\omega)} \langle z_b(\upsilon), z_a(\omega)\rangle\Big]. \tag{B.12}$$

When all memories are unit norm, orthogonality gives $\|z_a(\omega)\| = 1$, so these score-gap definitions reduce exactly to the cosine-separation formulas in the main paper. The margin is always evaluated on the current memory bank and is therefore recomputed when M changes.

For a target orbit a and $0 < \varepsilon < 1$, define the strict posterior hitting time

$$\tau_\varepsilon^{\mathrm{strict}} := \inf\{s \geq 0 : P_s(a) > 1-\varepsilon\}, \qquad \inf\varnothing := +\infty. \tag{B.13}$$

Theorem B.2 (Sampled-bank strict high-confidence orbit-collapse tube)**.** *Let $M \geq 2$ and $0 < \varepsilon < 1$. Consider a finite quotient-orbit bank or a fixed sampled approximation with normalized within-orbit weights and orbit priors π_a. Let $w_{\min} > 0$ be the smallest positive atomic within-orbit weight. Let*

$$L_{\mathrm{prior}} = \max_a \frac{1-\pi_a}{\pi_a}, \tag{B.14}$$

$$\ell_M = \log\left(\frac{L_{\mathrm{prior}}}{\varepsilon w_{\min}}\right), \tag{B.15}$$

$$\theta_M = \Delta_{\mathrm{orbit}} - \beta^{-1}\ell_M, \tag{B.16}$$

$$r_{\mathrm{ret}}(M) = \frac{\theta_M}{D_{\mathrm{orbit}}}. \tag{B.17}$$

Thus, the positive effective strict-margin condition in the main paper means $\theta_M > 0$, not merely $\Delta_{\mathrm{orbit}} > 0$. This distinction accounts for the inverse temperature, orbit priors, target confidence, and smallest positive within-orbit weight. If $\theta_M > 0$ and $d_a(x_T) < r_{\mathrm{ret}}(M)$, then $P_T(a) > 1-\varepsilon$. Consequently,

$$\mathbb{P}(\tau_\varepsilon^{\mathrm{strict}} > T) \leq \mathbb{P}\big(d_a(x_T) \geq r_{\mathrm{ret}}(M)\big). \tag{B.18}$$

For a uniform orbit prior, $L_{\mathrm{prior}} = M - 1$.

Proof. Choose $y_\star \in \mathcal{O}_{\xi^a}$ with $\|x_T - y_\star\| = d_a(x_T)$. For any $y \in \mathcal{O}_{\xi^b}$, $b \neq a$,

$$\langle y_\star, x_T\rangle - \langle y, x_T\rangle \geq \Delta_{\mathrm{orbit}} - D_{\mathrm{orbit}} d_a(x_T). \tag{B.19}$$

If $d_a(x_T) < r_{\mathrm{ret}}(M)$, then every wrong-orbit score is at least $\beta^{-1}\ell_M$ below the score of $y_\star$. Since each wrong orbit is integrated with normalized total weight one,

$$\pi_b Z_b(x_T) \leq \pi_b e^{\beta\langle y_\star, x_T\rangle - \ell_M}. \tag{B.20}$$

Summing over $b \neq a$ yields

$$\sum_{b\neq a} \pi_b Z_b(x_T) \leq (1-\pi_a) e^{\beta\langle y_\star, x_T\rangle - \ell_M} \leq \varepsilon \pi_a w_{\min} e^{\beta\langle y_\star, x_T\rangle}. \tag{B.21}$$

The correct orbit contributes at least $\pi_a w_{\min} e^{\beta\langle y_\star, x_T\rangle}$, so $P_T(a) \geq (1+\varepsilon)^{-1} > 1-\varepsilon$. Hence

$$\{P_T(a) \leq 1-\varepsilon\} \subseteq \{d_a(x_T) \geq r_{\mathrm{ret}}(M)\}. \tag{B.22}$$

Because $\tau_\varepsilon^{\mathrm{strict}} > T$ implies $P_T(a) \leq 1-\varepsilon$, the probability bound follows. □

The positive atomic mass $w_{\min}$ is essential in this finite-bank corollary. It has no nonzero analogue under continuous Haar measure; exact continuous retrieval instead uses the free-energy radius in Proposition D.2.

B.1 Complete dynamics and error propagation

At every step, either continuous backend returns $(\log Z_a^{\mathrm{H}}(x_t), \mu_a^{\mathrm{H}}(x_t))$ for all a. The outer posterior and state update are unchanged:

$$P_t(a) = \frac{\pi_a Z_a^{\mathrm{H}}(x_t)}{\sum_b \pi_b Z_b^{\mathrm{H}}(x_t)}, \tag{B.23}$$

$$x_{t+1} = \sum_a P_t(a)\mu_a^{\mathrm{H}}(x_t), \qquad t = 0, \dots, T-1. \tag{B.24}$$

Suppose the implemented step has local error e_t and the exact map has Lipschitz constant L_t on a region containing both trajectories. Then

$$\|x_{t+1} - \widetilde{x}_{t+1}\| \le L_t \|x_t - \widetilde{x}_t\| + e_t. \tag{B.25}$$

Iterating gives

$$\|x_T - \widetilde{x}_T\| \le \sum_{t=0}^{T-1} \left(\prod_{j=t+1}^{T-1} L_j \right) e_t. \tag{B.26}$$

As in Eq. (19), the exact-map Jacobian is a posterior covariance, so a local spectral bound can supply L_t. A numerical trajectory bound can grow when no contraction has been established; the implementation does not reinterpret it as a convergence certificate.

Proposition B.1 (Certified backend-to-theory interface)**.** *At step t, suppose the backend returns orbit quantities satisfying*

$$\max_a |\widetilde{u}_{t,a} - u_{t,a}| \le \zeta_t, \qquad \max_a \|\widetilde{\mu}_{t,a} - \mu_{t,a}^{\mathrm{H}}\| \le m_t, \tag{B.27}$$

and $\|\mu_{t,a}^{\mathrm{H}}\| \le R$. Then the implemented outer posterior and state update satisfy

$$\|\widetilde{P}_t - P_t\|_1 \le \delta_P(\zeta_t) := \min\{2, e^{2\zeta_t} - 1\}, \tag{B.28}$$

$$\|\widetilde{F}_t(x) - F_{\mathrm{H}}(x)\| \le e_t := m_t + R\delta_P(\zeta_t). \tag{B.29}$$

Consequently, certified Fourier/Bessel tails or matrix-Fisher interval enclosures provide the (ζ_t, e_t) required by Theorems D.1 and D.4.

Proof. If each logit changes by at most ζ_t, the ratio between corresponding normalized posterior masses lies in $[e^{-2\zeta_t}, e^{2\zeta_t}]$. Summing the resulting pointwise difference gives Eq. (B.28), with the trivial upper bound two. Adding and subtracting $\sum_a \widetilde{P}_t(a)\mu_{t,a}^{\mathrm{H}}$ then yields

$$\begin{aligned} \|\widetilde{F}_t - F_{\mathrm{H}}\| \le & \sum_a \widetilde{P}_t(a) \|\widetilde{\mu}_{t,a} - \mu_{t,a}^{\mathrm{H}}\| \\ & + \sum_a |\widetilde{P}_t(a) - P_t(a)| \, \|\mu_{t,a}^{\mathrm{H}}\|, \end{aligned} \tag{B.30}$$

which is at most $m_t + R\delta_P(\zeta_t)$. □

For the Fourier/Bessel backend, Eq. (A.47) supplies ζ_t, while Eq. (A.48) and the orthogonal block decomposition supply m_t. For the strict matrix-Fisher path, interval enclosures of $c(K_a)$, its three derivatives, and the proper-SVD residual are propagated through Eqs. (A.59)– (A.60) to produce the same pair. An ordinary floating-point error estimate is diagnostic and is not substituted for these certified quantities.

B.2 Complexity and applicability

Backend	Persistent storage	Per-step work	Required structure
Expanded sampling	$O(MQd)$	$O(MQd)$	Any computable finite action
Factorized sampling	$O(Md + Qc_G)$	$O(MQd)$	Shared sampled transformations
SO(2) Fourier/Bessel	$O(Md)$	$O(Md + MC_{\mathrm{FB}})$	Known finite frequency decomposition
SO(3) matrix-Fisher	$O(Md)$	$O(Md + MC_{\mathrm{quad}})$	Corresponded $\ell = 1$ vector channels

Table 4: Asymptotic resource comparison. C_{FB} depends on effective Fourier support and concentration; C_{quad} depends on quadrature tolerance and concentration. Neither is a universal constant.

The $\mathrm{SO}(2)$ path is appropriate for steerable features with finitely many known angular frequencies. The matrix-Fisher path is appropriate when an equivariant encoder emits a small number of semantically corresponded vector channels. Raw unordered point clouds generally do not satisfy the latter condition; permutation handling must occur in the encoder or in a separate group action.

C Computable Local Contraction Certificates

This appendix first turns the finite-bank Jacobian identity into a deterministic uniform certificate over an entire retrieval tube. It then states the corresponding normal-direction result for smooth compact orbits.

Proposition C.1 (Quotient-orbit Jacobian)**.** *For a finite or sampled quotient-orbit memory bank, let ξ^i enumerate the sampled representatives (the $z_{a,q}$ of Appendix A) and let $\varpi_i > 0$ denote the combined orbit-prior and within-orbit weight. With*

$$p_i(x) = \frac{\varpi_i \exp(\beta {\xi^i}^\top x)}{\sum_j \varpi_j \exp(\beta {\xi^j}^\top x)}, \tag{C.1}$$

$$F(x) = \sum_i p_i(x)\xi^i. \tag{C.2}$$

we have

$$J_F(x) = \beta \operatorname{Cov}_{p_x}(\xi). \tag{C.3}$$

Proof. The weights ϖ_i are independent of x, so the softmax derivative is

$$\frac{\partial p_i}{\partial x} = \beta p_i \left(\xi^i - \sum_j p_j \xi^j \right) = \beta p_i (\xi^i - F(x)). \tag{C.4}$$

Therefore

$$\begin{aligned} J_F(x) &= \sum_i \xi^i \left(\frac{\partial p_i}{\partial x} \right)^\top \\ &= \beta \left(\sum_i p_i \xi^i {\xi^i}^\top - F(x)F(x)^\top \right), \end{aligned} \tag{C.5}$$

which is $\beta \operatorname{Cov}_{p_x}(\xi)$. For compact quotient orbits, the same calculation holds with integrals in place of sums. □

For the finite result, write $A_a = \{\xi^{a,q}\}_{q=1}^{Q_a}$ for the sampled target orbit, where $\xi^{a,q} = z_{a,q}$ in Appendix A. Let $d_a(x) = \min_{y \in A_a} \|x - y\|$, and define

$$\mathcal{T}_a(r) = \{x : d_a(x) \le r\}, \tag{C.6}$$

$$R = \max_i \|\xi^i\|. \tag{C.7}$$

For $h > 0$, set $s = 2h/\sqrt{N}$. Define the Cartesian grid, third pairwise moment, and computable Jacobian bound jointly by

$$\mathcal{G}_{a,h}(r) = \{sk : k \in \mathbb{Z}^N,\ d_a(sk) \le r + h\}, \tag{C.8}$$

$$S_3(x) = \sum_{i,j} p_i(x) p_j(x) \|\xi^i - \xi^j\|^3, \tag{C.9}$$

$$\begin{aligned} \widehat{\kappa}_a^{\mathrm{fin}}(h, r) = \max_{g \in \mathcal{G}_{a,h}(r)} \big\{ &\beta \| \operatorname{Cov}_{p_g}(\xi) \|_2 \\ &+ \beta^2 h e^{4\beta R h} S_3(g) \big\}. \end{aligned} \tag{C.10}$$

Every $x \in \mathcal{T}_a(r)$ is within Euclidean distance h of a point in this grid: coordinatewise rounding to $s\mathbb{Z}^N$ incurs distance at most $s\sqrt{N}/2 = h$, and the rounded point has target-set distance at most $r + h$. Thus $\mathcal{G}_{a,h}(r)$ is an explicit finite h-net of the tube, not a random covering sample. Moreover,

$$|\mathcal{G}_{a,h}(r)| \le Q_a \left(2 + \frac{2(r+h)}{s}\right)^N, \tag{C.11}$$

because the retained grid is contained in the union of Q_a axis-aligned cubes of half-width $r + h$. The construction is therefore finite and computable, although exponential in the ambient dimension.

All terms in (C.10) are finite sums evaluated on a finite set.

Theorem C.1 (Finite-cover uniform Jacobian certificate)**.** *For every $x \in \mathcal{T}_a(r)$,*

$$\|J_F(x)\|_2 \le \widehat{\kappa}_a^{\mathrm{fin}}(h, r). \tag{C.12}$$

Let

$$b_a = \max_{y \in A_a} d_a(F(y)). \tag{C.13}$$

Then every $x \in \mathcal{T}_a(r)$ satisfies

$$d_a(F(x)) \le b_a + \widehat{\kappa}_a^{\mathrm{fin}}(h, r) d_a(x). \tag{C.14}$$

Consequently, if $\widehat{\kappa}_a^{\mathrm{fin}}(h, r) < 1$ and

$$b_a + \widehat{\kappa}_a^{\mathrm{fin}}(h, r) r \le r, \tag{C.15}$$

then $\mathcal{T}_a(r)$ is forward invariant and the local-contraction hypothesis used by the Gaussian and capacity theorems holds uniformly on the whole tube.

Proof. Write $\mu_x = F(x)$ and $v_i(x) = \xi^i - \mu_x$. Differentiating the covariance identity in direction u gives

$$DJ_F(x)[u] = \beta^2 \sum_i p_i(x) \langle v_i(x), u \rangle v_i(x) v_i(x)^\top. \tag{C.16}$$

Hence, for $\|u\| = 1$,

$$\|DJ_F(x)[u]\|_2 \le \beta^2 \sum_i p_i(x) \|v_i(x)\|^3. \tag{C.17}$$

Since $v_i(x) = \sum_j p_j(x)(\xi^i - \xi^j)$, convexity of $v \mapsto \|v\|^3$ and Jensen's inequality imply

$$\sum_i p_i(x) \|v_i(x)\|^3 \le \sum_{i,j} p_i(x) p_j(x) \|\xi^i - \xi^j\|^3 = S_3(x). \tag{C.18}$$

Thus, with

$$L_J(x) := \sup_{\|u\|=1} \|DJ_F(x)[u]\|_2, \tag{C.19}$$

we have the directional operator bound $L_J(x) \le \beta^2 S_3(x)$.

If $\|x - g\| \le h$, boundedness of the atoms yields

$$e^{-\beta R h} \le \frac{\exp(\beta {\xi^i}^\top x)}{\exp(\beta {\xi^i}^\top g)} \le e^{\beta R h}. \tag{C.20}$$

The same bounds on the partition functions give

$$\frac{p_i(x)}{p_i(g)} \le e^{2\beta R h}, \tag{C.21}$$

$$S_3(x) \le e^{4\beta R h} S_3(g). \tag{C.22}$$

Integrating (C.16) along the line segment from g to x now gives

$$\|J_F(x) - J_F(g)\|_2 \le \beta^2 h e^{4\beta R h} S_3(g). \tag{C.23}$$

For any $x \in \mathcal{T}_a(r)$, choose its grid point g from the explicit h-net. Combining the last display with $\|J_F(g)\|_2 = \beta \| \mathrm{Cov}_{p_g}(\xi)\|_2$ proves the uniform Jacobian bound.

Finally choose $y \in A_a$ with $\|x - y\| = d_a(x)$. The segment from y to x lies in the ball $B(y, r) \subset \mathcal{T}_a(r)$. The mean-value inequality and the uniform Jacobian bound therefore give

$$\begin{aligned} d_a(F(x)) &\le d_a(F(y)) + \|F(x) - F(y)\| \\ &\le b_a + \widehat{\kappa}_a^{\mathrm{fin}}(h, r) \|x - y\|, \end{aligned} \tag{C.24}$$

which is (C.14). The final two inequalities make the right-hand side at most r throughout the tube. □

Computational scope. The Cartesian certificate above is deterministic but scales exponentially with the ambient dimension. It is directly useful in low dimensions and as a benchmark for reduced coordinates. A positive-dimensional orbit requires a different test: tangent directions transport pose along the same semantic orbit and therefore must not be counted as retrieval instability. We next give a finite normal-bundle certificate that retains the analytic Jacobian remainder while testing only distance-changing directions.

Smooth compact orbits. Let $\mathcal{M}_a = \mathcal{O}_{\xi^a} \subset \mathbb{R}^N$ be a compact C^3 embedded orbit of positive dimension. Assume a certified tubular radius $\tau_a > 0$, so the nearest-point projection π_a is uniquely defined on $U_a(\tau_a) = \{u : d_a(u) < \tau_a\}$. Write

$$R_a(u) = u - \pi_a(u), \tag{C.25}$$

$$\Pi_a^{\perp}(y) = \mathrm{Proj}_{N_y\mathcal{M}_a}. \tag{C.26}$$

On this tube, $\|R_a(u)\| = d_a(u)$. The use of $\Pi_a^{\perp}$ is essential: the full norm $\|J_F\|_2$ can exceed one because an equivariant update may transport tangent perturbations without increasing orbit distance.

Fix an input radius $0 < r < \tau_a$. Choose finitely many normal charts $(K_\ell, \psi_\ell, N_\ell)$, $\ell = 1, \dots, L$, whose images cover $\mathcal{M}_a$. Write $q_a = \dim \mathcal{M}_a$. Here $K_\ell \subset \mathbb{R}^{q_a}$ is compact, $\psi_\ell : K_\ell \to \mathcal{M}_a$, and the columns of $N_\ell(u) \in \mathbb{R}^{N\times(N-q_a)}$ form an orthonormal basis of $N_{\psi_\ell(u)}\mathcal{M}_a$. Let

$$\|\psi_\ell(u) - \psi_\ell(u')\| \le L_{\psi,\ell}\|u - u'\|, \tag{C.27}$$

$$\|N_\ell(u) - N_\ell(u')\|_2 \le L_{N,\ell}\|u - u'\| \tag{C.28}$$

be certified chart and frame bounds. Select a finite $h_{u,\ell}$-net $U_\ell \subset K_\ell$ and a finite $h_{v,\ell}$-net V_ℓ of $\overline{B}_{N-q_a}(0, r)$, allowing its grid centers to lie in $\overline{B}_{N-q_a}(0, r + h_{v,\ell})$, and define

$$\mathcal{C}_{a,\ell} = \{\psi_\ell(u_j) + N_\ell(u_j)v_k : u_j \in U_\ell,\ v_k \in V_\ell\}, \tag{C.29}$$

$$h_\ell = (L_{\psi,\ell} + rL_{N,\ell})h_{u,\ell} + h_{v,\ell}, \tag{C.30}$$

$$\delta_{\Pi,\ell} = 2L_{N,\ell}h_{u,\ell}. \tag{C.31}$$

Every $x \in \mathcal{T}_a(r)$ is within h_ℓ of a center $c \in \mathcal{C}_{a,\ell}$ in a chart containing $\pi_a(x)$, and its normal projector differs from the center-base projector by at most $\delta_{\Pi,\ell}$. The latter follows from $\Pi_a^{\perp}(\psi_\ell(u)) = N_\ell(u)N_\ell(u)^\top$.

The construction applies directly to a continuous bank. Let $p_x(d\xi)$ be the posterior probability measure on the disjoint union of all quotient orbits and assume $\|\xi\| \le R$. Define

$$S_3(x) = \iint \|\xi - \xi'\|^3 p_x(d\xi)p_x(d\xi'). \tag{C.32}$$

For $c \in \mathcal{C}_{a,\ell}$, set

$$L_{J,c} = \beta^2 e^{4\beta R h_\ell} S_3(c), \tag{C.33}$$

$$\overline{J}_a = \max_{\ell, c\in\mathcal{C}_{a,\ell}} \big(\|J_F(c)\|_2 + L_{J,c}h_\ell\big). \tag{C.34}$$

If the support diameter is at most D, the convenient bound

$$S_3(c) \le 2D\ \mathrm{tr}\,\mathrm{Cov}_{p_c}(\xi) \tag{C.35}$$

follows from $\|\xi - \xi'\|^3 \le D\|\xi - \xi'\|^2$ and $\mathbb{E}\|\xi - \xi'\|^2 = 2\,\mathrm{tr}\,\mathrm{Cov}(\xi)$.

Choose an output radius $0 < \rho < \tau_a$. On $\overline{U}_a(\rho)$, let certified geometric bounds satisfy

$$B_R \ge \sup_u \|DR_a(u)\|_2, \tag{C.36}$$

$$\|DR_a(u) - DR_a(v)\|_2 \le L_R\|u - v\|. \tag{C.37}$$

For a center $c = \psi_\ell(u_j) + N_\ell(u_j)v_k$, denote its base point by $y_c = \psi_\ell(u_j)$ and define

$$A_c = DR_a(F(c))J_F(c)\Pi_a^{\perp}(y_c), \tag{C.38}$$

$$\mathrm{Rem}_c = L_R\overline{J}_a^2 h_\ell + B_R L_{J,c}h_\ell + B_R\|J_F(c)\|_2\delta_{\Pi,\ell}, \tag{C.39}$$

$$\widehat{\kappa}_a^{\mathrm{nor}} = \max_{\ell, c\in\mathcal{C}_{a,\ell}} \big(\|A_c\|_2 + \mathrm{Rem}_c\big). \tag{C.40}$$

Finally let $\mathcal{Y}_a$ be a finite h_M-net of $\mathcal{M}_a$ and set

$$\widehat{b}_a = \max_{y_j\in\mathcal{Y}_a} d_a(F(y_j)) + \beta R^2 h_M. \tag{C.41}$$

Theorem C.2 (Computable contraction on a positive-dimensional orbit)**.** *Assume the chart, frame, posterior-moment, and residual-geometry quantities above are certified upper enclosures. Suppose also that the finite output test*

$$\max_{\ell, c\in\mathcal{C}_{a,\ell}} \{d_a(F(c)) + \overline{J}_a h_\ell\} \le \rho \tag{C.42}$$

holds. Then every $x \in \mathcal{T}_a(r)$ satisfies

$$d_a(F(x)) \le \widehat{b}_a + \widehat{\kappa}_a^{\mathrm{nor}} d_a(x). \tag{C.43}$$

Consequently,

$$\widehat{\kappa}_a^{\mathrm{nor}} < 1, \tag{C.44}$$

$$\widehat{b}_a + \widehat{\kappa}_a^{\mathrm{nor}} r \le r \tag{C.45}$$

certify uniform normal contraction and forward invariance on the complete tube of the continuous orbit.

Proof. The posterior derivative calculation used for the finite bank remains valid for the continuous posterior measure by differentiation under the integral; bounded support supplies a uniform dominating function. Thus Eq. (C.3) remains valid, and differentiating it in a unit direction q gives

$$DJ_F(x)[q] = \beta^2 \int \langle \xi - \mu_x, q\rangle (\xi - \mu_x)(\xi - \mu_x)^\top p_x(d\xi). \tag{C.46}$$

The Jensen argument and posterior-ratio comparison from the finite proof therefore give, whenever $\|x - c\| \le h_\ell$,

$$\|J_F(x) - J_F(c)\|_2 \le L_{J,c} h_\ell, \tag{C.47}$$

$$\|J_F(x)\|_2 \le \overline{J}_a. \tag{C.48}$$

Thus Eq. (C.42) implies $F(\mathcal{T}_a(r)) \subset \overline{U}_a(\rho)$, where R_a and its derivative bounds are valid.

For $x \in \mathcal{T}_a(r)$, write $y = \pi_a(x)$, choose a covering center c in a chart containing y, and abbreviate $\Pi_x = \Pi_a^\perp(y)$, $\Pi_c = \Pi_a^\perp(y_c)$. Add and subtract the three factors in $DR_a(F(x))J_F(x)\Pi_x - A_c$. Eqs. (C.36), (C.34), and (C.47) yield

$$\begin{aligned} &\|DR_a(F(x))J_F(x)\Pi_x - A_c\|_2 \\ &\quad \le L_R\|F(x) - F(c)\|\,\|J_F(x)\|_2 \\ &\qquad + B_R\|J_F(x) - J_F(c)\|_2 \\ &\qquad + B_R\|J_F(c)\|_2\|\Pi_x - \Pi_c\|_2 \\ &\quad \le \mathrm{Rem}_c. \end{aligned} \tag{C.49}$$

Hence the normal derivative of $R_a \circ F$ is uniformly bounded by $\widehat{\kappa}_a^{\mathrm{nor}}$ over the entire input tube.

Now write $x = y + n$, where $n \in N_y\mathcal{M}_a$ and $\|n\| = d_a(x)$, and put $x_t = y + tn$. The segment x_t, $0 \le t \le 1$, lies in the tube. The fundamental theorem of calculus, followed by the surface-net bias bound, gives, with $K_t = DR_a(F(x_t))J_F(x_t)\Pi_a^\perp(y)$,

$$\begin{aligned} d_a(F(x)) &= \|R_a(F(x))\| \\ &\le \|R_a(F(y))\| + \int_0^1 \|K_t n\|\, dt \\ &\le \widehat{b}_a + \widehat{\kappa}_a^{\mathrm{nor}}\|n\|. \end{aligned} \tag{C.50}$$

Here the surface-net term is valid because $\|J_F(u)\|_2 \le \beta R^2$ globally and distance to a closed set is one-Lipschitz. This proves Eq. (C.43); Eq. (C.44) then gives the stated conclusions. □

If an implemented backend obeys $\|\widetilde{F}_t(x) - F(x)\| \le e_t$ on the tube, the one-Lipschitz property of distance immediately gives

$$d_a(\widetilde{F}_t(x)) \le \widehat{b}_a + e_t + \widehat{\kappa}_a^{\mathrm{nor}} d_a(x). \tag{C.51}$$

Thus certified numerical error enters as additional accumulated bias rather than being mistaken for a change in semantic orbit geometry.

For subsequent retrieval and capacity results, unify the two certified pairs as

$$(\bar{b}_a, \widehat{\kappa}_a) = \begin{cases} (b_a, \widehat{\kappa}_a^{\mathrm{fin}}), & \text{finite sampled orbit,} \\ (\widehat{b}_a, \widehat{\kappa}_a^{\mathrm{nor}}), & \text{smooth continuous orbit.} \end{cases} \tag{C.52}$$

The abstract theorems set $b_a := \bar{b}_a$, suppress the certification superscripts, and choose $\kappa \geq \sup_a \widehat{\kappa}_a$ as the uniform contraction coefficient.

Symmetry reduction of the normal cover. If the representation is orthogonal, the memory measure is group closed, and F is equivariant, then for every $k \in G$,

$$J_F(\rho(k)x) = \rho(k) J_F(x) \rho(k)^\top, \tag{C.53}$$

$$\Pi_a^\perp(\rho(k)y) = \rho(k) \Pi_a^\perp(y) \rho(k)^\top, \tag{C.54}$$

$$DR_a(\rho(k)u) = \rho(k) DR_a(u) \rho(k)^\top. \tag{C.55}$$

Therefore the normal-map operator norm is constant along group translates. One may cover a quotient of the normal fibers by the stabilizer action rather than repeat equivalent tangent positions. For an $\mathrm{SO}(2)$ circle orbit, a single normal fiber is sufficient.

The theorem is a mathematical certificate only when the finite center values, special functions or orbit integrals, and geometric constants are enclosed with verified numerical error. Ordinary Monte Carlo estimates do not provide such enclosures. The interval-certified Fourier/Bessel and matrix-Fisher paths in Appendix B.1 supply the required backend enclosures within their stated representation classes.

D Retrieval and Capacity Proofs

The main text states a convergence-first, size-controlled retrieval task. Operationally, each step computes $\{Z_{s,a}, P_s(a), \mu_{s,a}\}_{a=1}^M$, tests $(\Delta_x(s), \Delta_P(s))$, and then forms post-convergence candidate sets from $\mathcal{S}_\star(\eta, \gamma)$.

For $s \geq 1$, the state and posterior increments are

$$\Delta_x(s) = \frac{\|x_s - x_{s-1}\|}{\|x_{s-1}\| + \nu}, \tag{D.1}$$

$$\Delta_P(s) = \|P_s - P_{s-1}\|_1. \tag{D.2}$$

With window h, the convergence time is

$$\begin{aligned} \tau_{\mathrm{conv}} = \inf\{ s \geq h : \Delta_x(t) \leq \mathrm{tol}_x,\ \Delta_P(t) \leq \mathrm{tol}_P \\ \text{for all } t = s - h + 1, \ldots, s \}. \end{aligned} \tag{D.3}$$

If $\tau_{\mathrm{conv}} \leq T$, set $x_\star = x_{\tau_{\mathrm{conv}}}$ and $P_\star = P_{\tau_{\mathrm{conv}}}$. The target-orbit distance is

$$d_a(y) = \inf_{\omega \in \Omega_a} \|y - z_a(\omega)\|. \tag{D.4}$$

The accuracy event is

$$\mathcal{A}_T(a) = \{ \tau_{\mathrm{conv}} \leq T,\ d_a(x_\star) \leq r_{\mathrm{acc}} \}. \tag{D.5}$$

Using the prior-adjusted free energy $\widetilde{\mathcal{E}}_b(x) = -\beta^{-1} \log(\pi_b Z_b(x))$, the stable post-convergence orbit set is

$$\mathcal{S}_\star(\eta, \gamma) = \left\{ b : P_\star(b) \geq \eta \text{ or } \widetilde{\mathcal{E}}_b(x_\star) - \min_c \widetilde{\mathcal{E}}_c(x_\star) \leq \gamma \right\}, \tag{D.6}$$

and the target-coverage event is

$$\mathcal{C}_T(a) = \{ a \in \mathcal{S}_\star(\eta, \gamma) \}. \tag{D.7}$$

Coverage alone is deliberately not called retrieval success. For a declared candidate budget $1 \leq K < M$, let $\mathrm{Top}_K(P_\star)$ denote the K largest posterior indices, using a fixed deterministic rule for exact ties, and define

$$\mathcal{S}_\star^{(K)} = \mathcal{S}_\star(\eta, \gamma) \cap \mathrm{Top}_K(P_\star), \tag{D.8}$$

and

$$\mathcal{R}_T^{(K)}(a) = \{a \in \mathcal{S}_\star^{(K)}\}. \tag{D.9}$$

Conditioned on $\mathcal{A}_T(a)$, the target-coverage, selective-retrieval, and set-precision diagnostics are

$$p_{\mathrm{cov}}^{(T)}(a) = \mathbb{P}(\mathcal{C}_T(a) \mid \mathcal{A}_T(a)), \tag{D.10}$$

$$p_{\mathrm{sel}}^{(T)}(a;K) = \mathbb{P}(\mathcal{R}_T^{(K)}(a) \mid \mathcal{A}_T(a)), \tag{D.11}$$

$$\mathrm{Prec}_{\mathrm{set}}^{(T)}(a) = \mathbb{E}\left[\frac{\mathbf{1}\{a \in \mathcal{S}_\star\}}{\max(1, |\mathcal{S}_\star|)} \,\middle|\, \mathcal{A}_T(a)\right]. \tag{D.12}$$

All three are set to zero when $\mathbb{P}(\mathcal{A}_T(a)) = 0$. In particular, $\mathcal{S}_\star = \{1, \dots, M\}$ gives $p_{\mathrm{cov}}^{(T)} = 1$ but set precision $1/M$, so it cannot establish useful retrieval.

Fix a prescribed family of orbit banks $\{\mathcal{B}_M\}_{M\geq 1}$; every probability below is evaluated using $\mathcal{B}_M$, and the supremum of an empty feasible set is defined as zero. Given $\delta_{\mathrm{acc}}, \delta_{\mathrm{sel}} \in (0,1)$, the size-controlled T-step capacity is

$$C_{\mathrm{meta}}^{(T)}(K; \delta_{\mathrm{acc}}, \delta_{\mathrm{sel}}) = \sup\big(\{M > K : \mathcal{Q}_M(K)\} \cup \{0\}\big), \tag{D.13}$$

$$\mathcal{Q}_M(K) = \begin{cases} \inf_a p_{\mathrm{acc}}^{(T)}(a) \geq 1 - \delta_{\mathrm{acc}}, \\ \inf_a p_{\mathrm{sel}}^{(T)}(a;K) \geq 1 - \delta_{\mathrm{sel}}. \end{cases} \tag{D.14}$$

The stricter auxiliary capacity uses the worst-case strict posterior success probability

$$\Phi_M^{(T)}(\varepsilon) := \inf_a \inf_{y_0 \in \mathcal{O}_{\xi^a}} \mathbb{P}(\tau_\varepsilon^{\mathrm{strict}} \leq T \mid x_0 = y_0 + \eta_0), \tag{D.15}$$

and is

$$C_{\mathrm{strict}}^{(T)}(\varepsilon, \delta) = \sup\big(\{M : \Phi_M^{(T)}(\varepsilon) \geq 1 - \delta\} \cup \{0\}\big). \tag{D.16}$$

Proposition D.1 (Two-part selective-retrieval decomposition)**.** *Using the T-step accuracy-convergence event in Eq.* (D.5) *and the size-controlled selection event in Eq.* (D.9)*, define*

$$p_{\mathrm{acc}}^{(T)}(a) = \mathbb{P}(\mathcal{A}_T(a)), \tag{D.17}$$

$$p_{\mathrm{ret}}^{(T)}(a;K) = \mathbb{P}(\mathcal{A}_T(a) \cap \mathcal{R}_T^{(K)}(a)). \tag{D.18}$$

If $\mathbb{P}(\mathcal{A}_T(a)) > 0$, define

$$p_{\mathrm{sel}}^{(T)}(a;K) = \mathbb{P}(\mathcal{R}_T^{(K)}(a) \mid \mathcal{A}_T(a)). \tag{D.19}$$

If $\mathbb{P}(\mathcal{A}_T(a)) = 0$, set $p_{\mathrm{sel}}^{(T)}(a;K) = 0$ by convention. When $\mathbb{P}(\mathcal{A}_T(a)) > 0$,

$$p_{\mathrm{ret}}^{(T)}(a;K) = p_{\mathrm{acc}}^{(T)}(a) p_{\mathrm{sel}}^{(T)}(a;K). \tag{D.20}$$

If $p_{\mathrm{acc}}^{(T)}(a) \geq 1 - \delta_{\mathrm{acc}}$ and $p_{\mathrm{sel}}^{(T)}(a;K) \geq 1 - \delta_{\mathrm{sel}}$, then

$$\mathbb{P}(\mathcal{A}_T(a) \cap \mathcal{R}_T^{(K)}(a)) \geq 1 - \delta_{\mathrm{acc}} - \delta_{\mathrm{sel}}. \tag{D.21}$$

Proof. When $\mathbb{P}(\mathcal{A}_T(a)) > 0$, the product identity is the definition of conditional probability. The last inequality follows from $(1 - \delta_{\mathrm{acc}})(1 - \delta_{\mathrm{sel}}) \geq 1 - \delta_{\mathrm{acc}} - \delta_{\mathrm{sel}}$. □

Proposition D.2 (Continuous free-energy retrieval tube)**.** *Assume $\|z\| \leq R$ on every orbit and let $\Lambda_\star(M)$ be defined by Eq.* (16)*. If the implemented orbit logits obey*

$$\sup_c |\widetilde{u}_c(x) - u_c(x)| \leq \zeta, \tag{D.22}$$

then every target a and state x satisfy

$$\widetilde{\Lambda}_a(x) \geq \Lambda_\star(M) - 2\beta R d_a(x) - 2\zeta. \tag{D.23}$$

Consequently, the positive radius in Eq. (17) *implies Eq.* (18)*. This statement holds for normalized Haar measure and, as a special case, normalized finite counting measure.*

Proof. For any orbit c, any x, x', and any representative z, $\langle z, x - x' \rangle \le R\|x - x'\|$. Integrating the resulting exponential inequality in both directions gives

$$|u_c(x) - u_c(x')| \le \beta R\|x - x'\|. \tag{D.24}$$

Log-sum-exp preserves a common Lipschitz constant, so Λ_a is $2\beta R$-Lipschitz. Choose $y \in \mathcal{M}_a$ with $\|x - y\| = d_a(x)$; compactness ensures the minimum is attained. Then

$$\Lambda_a(x) \ge \Lambda_a(y) - 2\beta R d_a(x) \ge \Lambda_\star(M) - 2\beta R d_a(x). \tag{D.25}$$

If every logit changes by at most ζ, both the target logit and the competitor log-sum-exp change by at most ζ, proving Eq. (D.23). Finally, $\widetilde{P}_x(a) = (1 + e^{-\widetilde{\Lambda}_a(x)})^{-1}$, so the radius condition is exactly sufficient for $\widetilde{P}_x(a) > 1 - \varepsilon$. □

Theorem D.1 (Numerically robust continuous Gaussian certificate)**.** *Suppose that for target orbit a, every implemented step satisfies*

$$d_a(\widetilde{F}_t(x)) \le \widehat{b}_a + e_t + \kappa d_a(x), \qquad d_a(x) \le r_{\mathrm{loc}}, \tag{D.26}$$

where $0 < \kappa < 1$, and suppose

$$\widehat{b}_a + e_t + \kappa r_{\mathrm{loc}} \le r_{\mathrm{loc}} \quad \textit{for } t = 0, \dots, T-1. \tag{D.27}$$

Let

$$B^{\mathrm{num}}_{a,T} = \sum_{t=0}^{T-1} \kappa^{T-1-t}(\widehat{b}_a + e_t), \tag{D.28}$$

and suppose the final logit enclosure is ζ_T. If $r^{\mathrm{H}}_{\mathrm{ret}}(M; \zeta_T) > B^{\mathrm{num}}_{a,T}$, define

$$R^{\mathrm{cont}}_{M,a}(T) = \min\left\{ r_{\mathrm{loc}},\, \kappa^{-T}\left(r^{\mathrm{H}}_{\mathrm{ret}}(M; \zeta_T) - B^{\mathrm{num}}_{a,T} \right) \right\}. \tag{D.29}$$

For $x_0 = y_0 + \eta_0$, $y_0 \in \mathcal{M}_a$, $\eta_0 \sim \mathcal{N}(0, \Sigma_0)$, and $\|\Sigma_0\|_2 \le \sigma^2$,

$$\mathbb{P}(\widetilde{\tau}^{\mathrm{strict}}_\varepsilon > T) \le \overline{F}_{\chi^2_N}\left(\frac{(R^{\mathrm{cont}}_{M,a}(T))^2}{\sigma^2} \right). \tag{D.30}$$

Proof. On $\|\eta_0\| < R^{\mathrm{cont}}_{M,a}(T)$, tube invariance permits iteration of Eq. (D.26), giving

$$d_a(\widetilde{x}_T) \le \kappa^T d_a(x_0) + B^{\mathrm{num}}_{a,T} < r^{\mathrm{H}}_{\mathrm{ret}}(M; \zeta_T). \tag{D.31}$$

Proposition D.2 therefore gives strict implemented retrieval. Failure is contained in $\{\|\eta_0\| \ge R^{\mathrm{cont}}_{M,a}(T)\}$. Diagonalizing Σ_0 and coupling its eigenvalues with the same independent standard Gaussians gives $\|\eta_0\|^2 \le \sigma^2 \chi^2_N$ almost surely, proving the bound. The norm comparison is an equality for $\Sigma_0 = \sigma^2 I_N$. □

Theorem D.2 (Biased local contraction with Gaussian initialization)**.** *Assume the hypotheses of Theorem B.2. Let $\sigma > 0$. Assume further that the target orbit has a biased local contraction tube: there exist $r_{\mathrm{loc}} > 0$, $0 < \kappa < 1$, and $b_a \ge 0$ such that*

$$d_a(F(x)) \le b_a + \kappa d_a(x) \qquad \textit{whenever } d_a(x) \le r_{\mathrm{loc}}, \tag{D.32}$$

and

$$b_a + \kappa r_{\mathrm{loc}} \le r_{\mathrm{loc}}. \tag{D.33}$$

Let

$$x_0 = y_0 + \eta_0, \qquad y_0 \in \mathcal{O}_{\xi^a}, \tag{D.34}$$

$$\eta_0 \sim \mathcal{N}(0, \Sigma_0), \qquad \|\Sigma_0\| \le \sigma^2, \tag{D.35}$$

and define

$$B_{a,T} = \frac{1 - \kappa^T}{1 - \kappa} b_a. \tag{D.36}$$

If $r_{\mathrm{ret}}(M) > B_{a,T}$, set

$$R^{\mathrm{bias}}_{M,a}(T) = \min\left\{ r_{\mathrm{loc}}, \kappa^{-T}\left(r_{\mathrm{ret}}(M) - B_{a,T} \right) \right\}. \tag{D.37}$$

Then

$$\mathbb{P}(\tau_\varepsilon^{\mathrm{strict}} > T) \le \overline{F}_{\chi^2_N}\left(\frac{(R^{\mathrm{bias}}_{M,a}(T))^2}{\sigma^2}\right), \tag{D.38}$$

where $\overline{F}_{\chi^2_N}(u) = \mathbb{P}(\chi^2_N \ge u)$*. The chi-square tail is sharp as a covariance-uniform bound for the Gaussian norm event over* $\{\Sigma_0 : \|\Sigma_0\| \le \sigma^2\}$*: equality in that comparison is attained by* $\Sigma_0 = \sigma^2 I_N$*.*

Proof. Since $y_0 \in \mathcal{O}_{\xi^a}$, $d_a(x_0) \le \|\eta_0\|$. On the event $\|\eta_0\| < R^{\mathrm{bias}}_{M,a}(T)$, we have $d_a(x_0) < r_{\mathrm{loc}}$. The tube-invariance condition gives by induction that $d_a(x_s) \le r_{\mathrm{loc}}$ for all $s \le T$: if $d_a(x_s) \le r_{\mathrm{loc}}$, then

$$d_a(x_{s+1}) \le b_a + \kappa r_{\mathrm{loc}} \le r_{\mathrm{loc}}. \tag{D.39}$$

Thus the local contraction inequality can be applied at every step. A second induction gives

$$d_a(x_T) \le \kappa^T d_a(x_0) + B_{a,T}. \tag{D.40}$$

Using $d_a(x_0) \le \|\eta_0\| < R^{\mathrm{bias}}_{M,a}(T)$, we obtain $d_a(x_T) < r_{\mathrm{ret}}(M)$. Theorem B.2 then implies $\tau_\varepsilon^{\mathrm{strict}} \le T$. Hence failure can occur only when $\|\eta_0\| \ge R^{\mathrm{bias}}_{M,a}(T)$.

For the Gaussian tail, diagonalize $\Sigma_0 = U\,\mathrm{diag}(\lambda_1, \dots, \lambda_N)U^\top$, where $0 \le \lambda_i \le \sigma^2$. With independent $Z_i \sim \mathcal{N}(0,1)$, use the same Z_i's to couple

$$\|\eta_0\|^2 \overset{d}{=} \sum_{i=1}^N \lambda_i Z_i^2 \le \sigma^2 \sum_{i=1}^N Z_i^2 \quad \text{almost surely.} \tag{D.41}$$

Combining this stochastic domination with the preceding failure-event inclusion gives the stated chi-square survival bound. If $\Sigma_0 = \sigma^2 I_N$, the displayed comparison is an equality, which also proves sharpness over the stated covariance class. □

Proposition D.3 (No separation-free strict capacity law)**.** *There is no nontrivial strict-retrieval capacity lower bound depending only on* N, β, G, T *and not on the quotient-orbit geometry. In particular, if two semantic memories have the same quotient orbit and the same prior, then no query can assign posterior mass larger than* $1/2$ *to both memories.*

Proof. Let $\mathcal{O}_{\xi^1} = \mathcal{O}_{\xi^2}$, let $\nu_1 = \nu_2$, and let $\pi_1 = \pi_2$. Then $Z_1(x) = Z_2(x)$ for every x, so

$$P_x(1) = P_x(2). \tag{D.42}$$

Since $P_x(1) + P_x(2) \le 1$, both probabilities are at most $1/2$. Thus, for any $\varepsilon < 1/2$, strict retrieval $P_x(a) > 1 - \varepsilon$ cannot be guaranteed uniformly for both targets. Therefore any meaningful strict capacity certificate must depend on orbit separation, priors, quotient measure, sampling, or other geometric information. □

Theorem D.3 (Exact-map free-energy gap capacity certificate)**.** *Let* $0 < \varepsilon, \delta < 1$ *and* $\sigma > 0$*. Work with the prescribed bank family* $\{\mathcal{B}_M\}_{M \ge 1}$ *fixed above. For normalized counting or quotient Haar measure, set* $u_a(x) = \log \pi_a + \log Z_a(x)$ *and define*

$$\Lambda_a(x) = u_a(x) - \log \sum_{b \ne a} e^{u_b(x)}. \tag{D.43}$$

Assume all orbit representatives satisfy

$$R = \sup_{a,\omega} \|z_a(\omega)\| < \infty. \tag{D.44}$$

Assume uniform biased local contraction: for every target a*,*

$$d_a(F(x)) \le b_a + \kappa d_a(x) \qquad \textit{whenever } d_a(x) \le r_{\mathrm{loc}}, \tag{D.45}$$

with $0 < \kappa < 1$*,* $b_a \le b_{\max}$*, and*

$$b_{\max} + \kappa r_{\mathrm{loc}} \le r_{\mathrm{loc}}. \tag{D.46}$$

Let

$$\overline{F}_{\chi^2_N}(u) = \mathbb{P}(\chi^2_N \geq u), \tag{D.47}$$

$$q^{\chi}_N(\delta) = \sqrt{\inf\{u \geq 0 : \overline{F}_{\chi^2_N}(u) \leq \delta\}}, \tag{D.48}$$

$$B_T = \frac{1-\kappa^T}{1-\kappa} b_{\max}. \tag{D.49}$$

$$\Gamma_T = \kappa^T \sigma q^{\chi}_N(\delta) + B_T, \tag{D.50}$$

$$\Lambda_\star(M) = \inf_a \inf_{y \in \mathcal{O}_{\xi^a}} \Lambda_a(y). \tag{D.51}$$

Equivalently, with the initial high-probability radius $\Gamma_0 := \sigma q^{\chi}_N(\delta)$, Eq. (D.50) *reads $\Gamma_T = \kappa^T \Gamma_0 + B_T$. Only the initial state is randomly perturbed; this theorem does not inject fresh noise at subsequent updates. For initial states $x_0 = y_0 + \eta_0$, $y_0 \in \mathcal{O}_{\xi^a}$, $\eta_0 \sim \mathcal{N}(0, \Sigma_0)$, and $\|\Sigma_0\| \leq \sigma^2$, if*

$$r_{\mathrm{loc}} \geq \sigma q^{\chi}_N(\delta) \tag{D.52}$$

and

$$\Lambda_\star(M) > \log \frac{1-\varepsilon}{\varepsilon} + 2\beta R \Gamma_T, \tag{D.53}$$

then

$$\inf_a \inf_{y_0 \in \mathcal{O}_{\xi^a}} \mathbb{P}\left(\tau^{\mathrm{strict}}_\varepsilon \leq T \mid x_0 = y_0 + \eta_0\right) \geq 1 - \delta. \tag{D.54}$$

Let $\mathsf{CapCond}_M(\varepsilon, \delta)$ denote the conjunction of these two displayed inequalities. Consequently the quantity

$$\underline{C}^{(T)}_{\mathrm{FE}}(\varepsilon, \delta) = \sup\Big(\{M : \mathsf{CapCond}_M(\varepsilon, \delta)\} \cup \{0\}\Big) \tag{D.55}$$

is a sufficient lower certificate for strict T-step single-orbit collapse capacity.

Proof. First, the posterior can be written as

$$P_x(a) = \frac{e^{u_a(x)}}{e^{u_a(x)} + \sum_{b \neq a} e^{u_b(x)}} = \frac{1}{1 + e^{-\Lambda_a(x)}}. \tag{D.56}$$

Therefore

$$P_x(a) > 1 - \varepsilon \quad \Longleftrightarrow \quad \Lambda_a(x) > \log \frac{1-\varepsilon}{\varepsilon}. \tag{D.57}$$

Next we prove the Lipschitz gap bound. For any orbit c and any x, x',

$$\begin{aligned} Z_c(x) &= \int_{\Omega_c} \exp(\beta \langle z_c(\omega), x \rangle)\, d\nu_c(\omega) \\ &\leq \exp(\beta R \|x - x'\|) Z_c(x'). \end{aligned} \tag{D.58}$$

Interchanging x, x' gives

$$|u_c(x) - u_c(x')| \leq \beta R \|x - x'\|. \tag{D.59}$$

Since $\log \sum_{b \neq a} e^{u_b(x)}$ is a log-sum-exp of functions with the same Lipschitz constant, it is also βR-Lipschitz. Hence

$$\Lambda_a(x) \geq \Lambda_a(y) - 2\beta R \|x - y\| \tag{D.60}$$

for every $y \in \mathcal{O}_{\xi^a}$.

It remains to control $\|x_T - y\|$. Write $\eta_0 = \Sigma_0^{1/2} g$, where $g \sim \mathcal{N}(0, I_N)$. The covariance assumption gives $\|\eta_0\| \leq \|\Sigma_0^{1/2}\|\, \|g\| \leq \sigma \|g\|$ pointwise under this coupling. Since $\|g\|^2 \sim \chi^2_N$, the definition of $q^{\chi}_N(\delta)$ implies

$$\mathbb{P}(\|\eta_0\| > \sigma q^{\chi}_N(\delta)) \leq \overline{F}_{\chi^2_N}\left((q^{\chi}_N(\delta))^2\right) \leq \delta. \tag{D.61}$$

Thus the event $\|\eta_0\| \leq \sigma q^{\chi}_N(\delta)$ has probability at least $1 - \delta$. Since $r_{\mathrm{loc}} \geq \sigma q^{\chi}_N(\delta)$, the initial state lies in the local tube around the target orbit. The tube-invariance condition gives $d_a(x_s) \leq r_{\mathrm{loc}}$ for every $s \leq T$, and induction yields

$$d_a(x_T) \leq \kappa^T d_a(x_0) + B_T \leq \Gamma_T. \tag{D.62}$$

Choose $y \in \mathcal{O}_{\xi^a}$ with $\|x_T - y\| \le \Gamma_T$. Then

$$\Lambda_a(x_T) \ge \Lambda_\star(M) - 2\beta R\Gamma_T > \log\frac{1-\varepsilon}{\varepsilon}, \tag{D.63}$$

so $P_T(a) > 1-\varepsilon$ and $\tau_\varepsilon^{\mathrm{strict}} \le T$. The bound is uniform over a and y_0, so taking the infimum preserves the probability statement. □

Theorem D.4 (Numerically robust continuous-Haar capacity certificate)**.** *Let $0 < \varepsilon, \delta < 1$, and assume Eq.* (D.26) *and Eq.* (D.27) *hold uniformly over targets with a common $0 < \kappa < 1$. Define*

$$q_N^\chi(\delta) = \sqrt{\inf\{u \ge 0 : \overline{F}_{\chi^2_N}(u) \le \delta\}}, \tag{D.64}$$

$$B_T^{\mathrm{num}} = \max_a \sum_{t=0}^{T-1} \kappa^{T-1-t}(\widehat{b}_a + e_t), \tag{D.65}$$

$$\Gamma_T^{\mathrm{num}} = \kappa^T \sigma q_N^\chi(\delta) + B_T^{\mathrm{num}}. \tag{D.66}$$

Suppose that final-step logits obey Eq. (D.22) *with $\zeta = \zeta_T$. If*

$$r_{\mathrm{loc}} \ge \sigma q_N^\chi(\delta), \tag{D.67}$$

$$\Lambda_\star(M) > \log\frac{1-\varepsilon}{\varepsilon} + 2\beta R\Gamma_T^{\mathrm{num}} + 2\zeta_T, \tag{D.68}$$

then

$$\inf_a \inf_{y_0 \in \mathcal{M}_a} \mathbb{P}\big(\widetilde{\tau}_\varepsilon^{\mathrm{strict}} \le T \mid x_0 = y_0 + \eta_0\big) \ge 1 - \delta \tag{D.69}$$

for every Gaussian η_0 with $\|\operatorname{Cov}(\eta_0)\|_2 \le \sigma^2$. Consequently,

$$\underline{C}_{\mathrm{H,num}}^{(T)}(\varepsilon, \delta) = \sup(\{M : \textit{Eqs.}\ (\mathrm{D.67})\ \textit{and}\ (\mathrm{D.68})\ \textit{hold}\} \cup \{0\}) \tag{D.70}$$

is a sufficient bank-specific capacity certificate. Setting $e_t = \zeta_T = 0$ recovers the exact-Haar result.

Proof. With probability at least $1-\delta$, Gaussian norm domination gives $d_a(x_0) \le \|\eta_0\| \le \sigma q_N^\chi(\delta)$. The tube condition and forward invariance allow Eq. (D.26) to be iterated, yielding

$$d_a(\widetilde{x}_T) \le \kappa^T \sigma q_N^\chi(\delta) + B_T^{\mathrm{num}} = \Gamma_T^{\mathrm{num}}. \tag{D.71}$$

Proposition D.2 then gives

$$\widetilde{\Lambda}_a(\widetilde{x}_T) \ge \Lambda_\star(M) - 2\beta R\Gamma_T^{\mathrm{num}} - 2\zeta_T > \log\frac{1-\varepsilon}{\varepsilon}. \tag{D.72}$$

The logistic identity therefore implies $\widetilde{P}_T(a) > 1-\varepsilon$. The argument is uniform in the target and clean target pose, which proves the probability and capacity statements. □

Proposition D.4 (Metric-entropy orbit-capacity corollary)**.** *Assume all memories are unit norm and the orbit prior is uniform. Define the worst within-orbit entropy penalty*

$$H_M(\beta) = \sup_{a, y \in \mathcal{O}_{\xi^a}} -\log \int_{\Omega_a} \exp\left[-\frac{\beta}{2}\|z_a(\omega) - y\|^2\right] d\nu_a(\omega). \tag{D.73}$$

Then

$$\Lambda_\star(M) \ge \beta\Delta_{\mathrm{orbit}} - H_M(\beta) - \log(M-1). \tag{D.74}$$

Consequently, under the hypotheses of Theorem D.4, a sufficient closed-form condition for strict implemented T-step collapse is

$$M < 1 + \frac{\varepsilon}{1-\varepsilon}\exp[\beta(\Delta_{\mathrm{orbit}} - 2\Gamma_T^{\mathrm{num}}) - H_M(\beta) - 2\zeta_T]\,. \tag{D.75}$$

Proof. Fix a and $y \in \mathcal{O}_{\xi^a}$. Since $\|z\| = \|y\| = 1$,

$$\langle z, y\rangle = 1 - \frac{1}{2}\|z - y\|^2. \tag{D.76}$$

Set

$$I_a(y) = \int_{\Omega_a} \exp\left[-\frac{\beta}{2}\|z_a(\omega) - y\|^2\right] d\nu_a(\omega). \tag{D.77}$$

Therefore

$$\begin{aligned} u_a(y) &= -\log M + \log \int_{\Omega_a} \exp(\beta\langle z_a(\omega), y\rangle)\, d\nu_a(\omega) \\ &= -\log M + \beta + \log I_a(y) \\ &\geq -\log M + \beta - H_M(\beta). \end{aligned} \tag{D.78}$$

For every wrong orbit $b \neq a$, the margin definition gives $\langle z, y\rangle \leq 1 - \Delta_{\mathrm{orbit}}$. Hence

$$\log \sum_{b\neq a} e^{u_b(y)} \leq -\log M + \log(M-1) + \beta(1 - \Delta_{\mathrm{orbit}}). \tag{D.79}$$

Subtracting the two displays gives

$$\Lambda_a(y) \geq \beta\Delta_{\mathrm{orbit}} - H_M(\beta) - \log(M-1). \tag{D.80}$$

Taking the infimum over a and y proves the first claim. Combining this lower bound with Theorem D.4 and solving for M proves the displayed capacity condition. The earlier exact-map formula is recovered by taking $e_t = \zeta_T = 0$. □

For a finite uniform sampled orbit, if y is restricted to the sampled representatives and each representative has weight $1/Q_a$, then $H_M(\beta) \leq \log Q_{\max}$. For a smooth compact quotient orbit G/H_a of dimension $q_a = \dim G - \dim H_a$, a uniform small-ball lower bound

$$\nu_a\{z : \|z - y\| \leq r\} \geq c\, r^{q_a} \tag{D.81}$$

implies

$$H_M(\beta) \leq \frac{\beta r^2}{2} - \log c - q_a \log r \tag{D.82}$$

for all admissible r. Choosing $r \asymp \beta^{-1/2}$ gives the asymptotic penalty $H_M(\beta) \leq (q_{\max}/2)\log\beta + O(1)$, with the constant depending on the local quotient-volume lower bound.

E Gaussian-Noise Orbit Search Derivation

This appendix gives the derivation behind the qualitative discussion of parameter effects for an orbit-search query with Gaussian initial perturbation. The goal is to separate three ingredients: the deterministic retrieval tube, the biased local orbit-contraction dynamics, and the concentration of the initial noise.

Fix a target orbit $\mathcal{O}_{\xi^a}$ and use the Gaussian initialization and covariance assumptions in Eqs. (D.34)–(D.35). Since y_0 lies on the target orbit, the initial distance to the orbit satisfies

$$d_a(x_0) = \mathrm{dist}(x_0, \mathcal{O}_{\xi^a}) \leq \|x_0 - y_0\| = \|\eta_0\|. \tag{E.1}$$

Assume the biased local orbit-contraction and tube-invariance conditions in Eqs. (D.32) and (D.33) hold for the target orbit, with $0 < \kappa < 1$ and $b_a \geq 0$. The latter condition ensures that once the iterate enters the local tube, every subsequent iterate used in the proof remains in the same tube. Repeated application then gives

$$d_a(x_T) \leq \kappa^T d_a(x_0) + \sum_{s=0}^{T-1} \kappa^s b_a. \tag{E.2}$$

Therefore,

$$d_a(x_T) \leq \kappa^T \|\eta_0\| + B_{a,T}. \tag{E.3}$$

The first term is the contracted initial perturbation, while the second term is the accumulated orbit-preservation bias defined in Eq. (D.36).

Use the deterministic retrieval radius $r_{\mathrm{ret}}(M)$ from Eq. (B.17). It is obtained from the finite or sampled orbit-level margin bound. By the deterministic strict-collapse tube, high-confidence single-orbit collapse is certified at time T whenever

$$d_a(x_T) < r_{\mathrm{ret}}(M). \tag{E.4}$$

Using the previous estimate, a sufficient condition for this is

$$\kappa^T \|\eta_0\| + B_{a,T} < r_{\mathrm{ret}}(M). \tag{E.5}$$

Thus, if

$$r_{\mathrm{ret}}(M) > B_{a,T}, \tag{E.6}$$

it is enough that

$$\|\eta_0\| < \kappa^{-T}\big(r_{\mathrm{ret}}(M) - B_{a,T}\big). \tag{E.7}$$

To ensure that the local contraction condition is valid from the initial step and remains valid throughout the iteration, we also require the initial state to lie inside the invariant local tube. Hence use the bias-corrected radius $R^{\mathrm{bias}}_{M,a}(T)$ from Eq. (D.37). On the event

$$\|\eta_0\| < R^{\mathrm{bias}}_{M,a}(T), \tag{E.8}$$

the initial query lies in the local contraction tube and the T-step iterate lies inside the deterministic strict-collapse tube. Consequently,

$$\tau_\varepsilon^{\mathrm{strict}} \le T. \tag{E.9}$$

Taking complements gives

$$\mathbb{P}(\tau_\varepsilon^{\mathrm{strict}} > T) \le \mathbb{P}\left(\|\eta_0\| \ge R^{\mathrm{bias}}_{M,a}(T)\right). \tag{E.10}$$

It remains to bound the right-hand side. Write

$$\Sigma_0 = U\,\mathrm{diag}(\lambda_1, \dots, \lambda_N)U^\top, \qquad 0 \le \lambda_i \le \sigma^2. \tag{E.11}$$

For independent $Z_i \sim \mathcal{N}(0,1)$, the covariance-uniform coupling is Eq. (D.41). Combining it with the failure-event inclusion in Eq. (E.10) yields the sharp chi-square bound in Eq. (D.38). This is the sharp uniform bound under $\|\Sigma_0\| \le \sigma^2$ for the Gaussian norm event, because $\Sigma_0 = \sigma^2 I_N$ makes the coupling inequality an equality. The preceding deterministic failure-event inclusion may still be conservative.

This derivation shows how the main parameters affect the strict-collapse failure probability. Larger orbit separation and larger inverse temperature increase $r_{\mathrm{ret}}(M)$, while larger prior imbalance, larger inter-orbit diameter, stronger noise, higher dimension, weaker contraction, and larger accumulated bias decrease the strict-collapse effective radius $R^{\mathrm{bias}}_{M,a}(T)$. The argument only uses the Gaussian distribution of the initial perturbation. It does not assume that the nonlinear Hopfield iterates are Gaussian. For the uniform radius in Theorem D.3, let $\Gamma_0 = \sigma q_N^{\chi}(\delta)$ and $\Gamma_\infty = b_{\max}/(1-\kappa)$. Then $\Gamma_T = \Gamma_\infty + \kappa^T(\Gamma_0 - \Gamma_\infty)$, so a larger T tightens the certificate if and only if $\Gamma_0 > \Gamma_\infty$. It leaves the bound unchanged at equality and loosens it when $\Gamma_0 < \Gamma_\infty$. A model with independent per-step noise would require an additional stochastic accumulation term and is outside the scope of Theorem D.3.

F Experimental Protocol Details

This appendix gives the procedures behind the experiments summarized in the main text. The main text reports the result-level evidence; the details below specify how memories, queries, diagnostics, and ablations are constructed. The accompanying experiment archive keeps the SO(2) energy visualization, Exp. 1–5, the free-energy diagnostic, the finite-cover audit (Exp. 6), the high-dimensional held-out diagnostic, the orbit-pooling ablation, and Exp. 7 in separate subdirectories. The image restoration protocol is described separately in Appendix G; it is an in-memory recovery demonstration rather than a source of the quantitative point-cloud claims.

Shared sampled-orbit construction. For synthetic ordered point clouds and the ModelNet10 simulation, each semantic memory is represented by a centered and Frobenius-normalized point matrix $X_a \in \mathbb{R}^{n\times 3}$. We sample rotations $R_1, \dots, R_Q \in \mathrm{SO}(3)$ and form the sampled orbit bank

$$Z_{a,q} = X_a R_q^\top, \qquad q = 1, \dots, Q. \tag{F.1}$$

The group-orbit Hopfield update in Algorithm 1 of the main paper is instantiated with this sampled bank and uniform orbit priors unless stated otherwise. The rotation samples are shared across semantic orbits in the balanced experiments. The ablation below intentionally breaks this balance to test sampling and stabilizer bias.

Normalized-noise retrieval trials. For Exp. 1–5 and the orbit-pooling ablation, a trial samples a target orbit a, an unknown rotation $R_\star \sim \mathrm{Haar}(\mathrm{SO}(3))$, and a normalized Gaussian noise matrix E. The initial query is

$$X_0 = X_a R_\star^\top + \sigma E. \tag{F.2}$$

The run applies the sampled orbit update for at most T iterations and records convergence, target-orbit distance, posterior mass, unrestricted target coverage, size-controlled selection, primary retrieval, and nearest-orbit baselines.

Diagnostics. The reported accuracy, coverage, selective-retrieval, and joint-retrieval quantities are those in Eqs. (D.10)–(D.12) and (D.17)–(D.20), with the zero-probability convention from Proposition D.1. We also report set precision, primary posterior accuracy, stable-set size $|\mathcal{S}_\star|$, the standard point-level Hopfield accuracy, the nearest sampled-orbit baseline, and the sampled equivariance error. For an independent Haar rotation $U \in \mathrm{SO}(3)$, the row-vector point-cloud convention $X \mapsto XU^\top$ gives

$$\mathrm{Err}_{\mathrm{eq}} = \frac{\|F_Q(XU^\top) - F_Q(X)U^\top\|_F}{\|F_Q(X)\|_F + \nu}. \tag{F.3}$$

Synthetic studies. Exp. 1 compares point-level Hopfield retrieval, nearest sampled-orbit readout, and quotient-orbit Hopfield readout under random rotations. Exp. 2 varies the number Q of sampled rotations while keeping the finite-step budget fixed, and then repeats the high-Q cases with a larger T/β budget. Exp. 3 constructs clustered memories with controlled cluster scale α to create metastable orbit families. Exp. 4 fixes $M = 12$, $\beta = 24$, $T = 3$, and per-coordinate Gaussian noise $\sigma = 0.012$, then changes one quantity in the metric-entropy capacity expression at a time. Five fixed seeds provide 240 queries each. Final strict success means $P_T(a) > 0.9$; unlike $\tau_\varepsilon^{\mathrm{strict}} \le T$, this diagnostic requires the threshold to hold at the final step.

For exact control, let $U_a \in \mathbb{R}^{3(M+1)\times 3}$ satisfy

$$U_a^\top U_a = I_3, \tag{F.4}$$

$$U_a^\top U_b = (1 - \Delta_{\mathrm{orbit}}) I_3 \quad (a \neq b), \tag{F.5}$$

and define unit-Frobenius matrix memories $z_a(g) = \mathrm{vec}(U_a A_q(g))$. We use

$$A_1(\theta) = 2^{-1/2}\,\mathrm{diag}(R_\theta, 0), \tag{F.6}$$

$$A_2(v) = e_1 v^\top, \quad v \in \mathbb{S}^2, \tag{F.7}$$

$$A_3(R) = 3^{-1/2} R^\top, \quad R \in \mathrm{SO}(3). \tag{F.8}$$

Their quotient-orbit dimensions are $q = 1, 2, 3$, respectively, while the ambient dimension remains $9(M+1)$. Since $\max_{g,h}\langle A_q(g), A_q(h)\rangle = 1$, this construction gives the prescribed pairwise separation exactly:

$$1 - \max_{g,h}\langle z_a(g), z_b(h)\rangle = \Delta_{\mathrm{orbit}}. \tag{F.9}$$

Thus the separation sweep changes Δ_{orbit} without changing self-orbit geometry. The entropy sweep uses the first $Q \in \{32, 64, 128, 256, 512, 1024\}$ elements of one fixed nested Haar $\mathrm{SO}(3)$ sequence and a common 32-pose query pool. The dimension sweep matches the finite sample count at

$Q = 1728$ for $q = 1, 2, 3$, uses Haar quadrature, and evaluates the analytic partition functions

$$Z_1(\beta) = I_0(\beta), \tag{F.10}$$

$$Z_2(\beta) = \sinh(\beta)/\beta, \tag{F.11}$$

$$Z_3(\beta) = e^{\beta/3}\left[I_0(2\beta/3) - I_1(2\beta/3)\right], \tag{F.12}$$

$$H_q(\beta) = \beta - \log Z_q(\beta). \tag{F.13}$$

Every query runs the complete T-step update. We report the conservative logistic lower bound obtained from $\beta\Delta_{\mathrm{orbit}} - H_M(\beta) - \log(M-1)$, the exact clean posterior computed from the finite bank or analytic Haar integral, and empirical strict T-step success. This isolates the capacity formula's directional predictions; it does not turn its sufficient lower bound into a necessary or tight law.

The Exp. 3 K-sensitivity rerun uses three clusters of 20 memories, 64 sampled $\mathrm{SO}(3)$ poses, $T = 8$, $\beta = 300$, and five fixed seeds with 300 queries per seed. Query poses are drawn from the same sampled rotation grid to isolate semantic ranking ambiguity from pose-discretization error. We report $K \in \{1, 3, 5, 10, 20\}$, convergence, unrestricted coverage, conditional size-controlled selection, joint retrieval, and seed standard deviations.

Free-energy certificate diagnostic. The free-energy diagnostic reported in the main paper uses sampled $\mathrm{SO}(3)$ orbit banks with $M = 48$, $Q = 80$, and $\varepsilon = 0.05$, so $\vartheta_\varepsilon = \log((1-\varepsilon)/\varepsilon) = \log 19$. For a final state x_T, let $y_\star$ be the nearest sampled target-orbit representative. The reported certified lower gap is

$$\underline{\Lambda}_T = \Lambda_a(y_\star) - 2\beta d_a(x_T). \tag{F.14}$$

The certificate passes when $\underline{\Lambda}_T > \vartheta_\varepsilon$. This is a sufficient strict-collapse test, so it is expected to be conservative in close-pair and metastable regimes.

ModelNet10 simulation. The ModelNet10 study is a real-data orbit-retrieval simulation, not a supervised ModelNet benchmark. No classifier is trained. We use the ten ModelNet10 categories (Wu et al., 2015), take three training meshes per category, sample 64 surface points from each mesh, center and Frobenius-normalize each point cloud, and obtain $M = 30$ semantic memory orbits. Queries are randomly rotated and noisy versions of stored CAD shapes, with $\sigma = 0.08$, batch size 64, six batches, and accuracy radius $r_{\mathrm{acc}} = 0.22$. The nearest-orbit baseline performs full sampled-orbit enumeration. Thus the reported numbers should not be compared directly with supervised classification leaderboards; they test the orbit-memory dynamics on real CAD geometry.

ModelNet40 masked latent-completion protocol. Exp. 7 is the real point-cloud memory-layer experiment. We use the ModelNet40 HDF5 point-cloud release used by the PointNet benchmark protocol (Wu et al., 2015; Qi et al., 2017) and select 11 visually similar categories: bench, bottle, bowl, chair, cup, desk, flower pot, sofa, stool, table, and vase. The split is leakage-safe: 48 training examples and 16 validation examples per class are disjoint subsets of the official training split, and 20 examples per class come from the official test split. Test queries are used only for final reporting.

Prototype selection uses training data only. The final reported run sets $K = 48$ prototypes per class, i.e., all available training examples in the selected split; this is the all-prototype endpoint of the farthest-prototype selection code and avoids discarding class modes. Each prototype is expanded with $Q = 32$ Haar-uniform $\mathrm{SO}(3)$ rotations generated from random unit quaternions. The same Q is used for every class and prototype. Stored orbit samples include small coordinate noise $\sigma_{\mathrm{bank}} = 0.003$. Test queries sample 512 points, apply an independent random $\mathrm{SO}(3)$ rotation, add coordinate noise $\sigma_{\mathrm{query}} = 0.012$, and are then centered and normalized.

The encoder is a lightweight spherical-MLP trained only on the training split, with random rotations and Gaussian coordinate noise as augmentation. Its best validation classification accuracy during encoder training is 0.5653. The encoder is frozen before building the memory bank and before validation/test retrieval. Raw-coordinate and tiny-PointNet variants in the experiment archive are treated as smoke tests or negative controls, not as positive evidence in the main text.

The nonlinear encoder is not assumed to realize a known orthogonal feature representation satisfying $\phi(XR^\top) = \rho(R)\phi(X)$. Consequently, Exp. 7 evaluates retrieval over a sampled augmentation orbit in feature space and the practical masked-completion layer; it is not a direct numerical test of the exact equivariance or energy-invariance theorems.

All stored features, query features, and updated states are L2-normalized. For the masked-completion task, 80% of query latent coordinates are masked and the corrupted feature is renormalized. With $\mathsf{M}_{\mathrm{obs}}$ denoting the observed coordinate mask, the application-level completion update is

$$x_{s+1} = \mathrm{norm}\Big(\mathsf{M}_{\mathrm{obs}} \odot x_0 + (1 - \mathsf{M}_{\mathrm{obs}}) \odot \big((1-\alpha)x_s + \alpha F_s(x_s)\big)\Big), \tag{F.15}$$

where F_s is the sampled group-orbit Hopfield update with the validation-selected temperature schedule and $\alpha = 0.4$. The validation grid contains the schedules

$$(20, 40, 80, 160, 320, 320, 320, 320) \tag{F.16}$$

and

$$(80, 160, 320, 640, 800, 800, 800, 800), \tag{F.17}$$

with $T \in \{2, 4, 8\}$. Seeds 11, 17, and 23 govern prototype and orbit-bank construction; each validation run selects the first schedule and $T = 2$. The held-out query transformations use a common fixed seed (20260707) across banks to enable paired comparisons, and the test set is evaluated once after selection.

The baselines are prototype-mean memory, nearest-pose retrieval, initial orbit evidence without a state update, and 3-rotation test-time augmentation. The nearest-pose baseline uses the exact same feature bank $\{z_{c,j,q}\}$ as the Orbit-Hopfield update; only the readout rule differs. For the corrected TTA baseline, all three rotated views receive the same observed-coordinate mask as the corresponding Orbit-Hopfield query. This preserves the 80% latent masking task and prevents independent views from revealing additional coordinates. TTA uses three encoder evaluations; Orbit-Hopfield uses one encoder evaluation and two complete memory updates. Wall-clock time includes these operations and is measured in the same run. The validation-selected schedule is frozen for a mask-severity sweep over $\{0.50, 0.65, 0.80, 0.90\}$. Results are reported as mean and standard deviation over the three bank seeds. Paired inference uses an exact two-sided sign-flip test over 220 base CAD shapes, clustering both query repetitions and all three bank seeds within each shape.

Deterministic finite-cover certificate. Exp. 6 uses nested finite C_4-orbit banks in $\mathbb{R}^3$. Each base memory has a distinct third coordinate, and its four representatives are generated by quarter-turn rotations of the first two coordinates. All atoms are unit normalized and all orbit and within-orbit weights are uniform. The fixed parameters are

$$\beta = 18, \quad r = 0.035, \quad h = 0.0075, \quad \sigma = 0.004, \quad T = 3. \tag{F.18}$$

with $\varepsilon = 0.10$, $\delta = 0.05$, and $M \in \{2, 3, 4, 5, 6, 7\}$. The nested order starts with widely separated orbits and then introduces progressively closer orbit families.

For each target orbit, the implementation enumerates every point of $\mathcal{G}_{a,h}(r)$, computes the covariance spectral norm and S_3, and takes the maximum of (C.10). It also evaluates b_a and $\Lambda_\star(M)$ over every stored representative. A row is reported as passing only if contraction, tube invariance, Gaussian containment, and the exact free-energy-gap inequality all pass. The reported Monte Carlo rate uses 500 Gaussian trials per atom and is diagnostic only; no random sample enters the complete-grid audit. Center values are evaluated in floating point; formal machine certification would additionally require outward-rounded interval enclosures. The full CSV, JSON configuration, figure, and source code are included in `experiments/finite_cover_certificate`.

Direct Gaussian probability-bound audit for the exact sampled-bank corollary. To distinguish a direct audit of the explicit failure bound from the binary Gaussian-radius check in the capacity experiment, we use the first four finite C_4 orbits with heights $(-0.82, 0.82, -0.28, 0.28)$, set

$$\beta = 96, \quad T = 3, \quad \varepsilon = 0.1, \quad r_{\mathrm{loc}} = 0.18, \quad h = 0.004, \tag{F.19}$$

and retain uniform orbit priors and uniform weights over the four atoms in each orbit. The complete Cartesian covers contain 4,237,248 centers in total. Evaluating the covariance Jacobian and the

analytic third-moment off-grid remainder gives

$$\widehat{\kappa} = 0.4854953, \tag{F.20}$$

$$b_{\max} = 1.6216\times10^{-7}, \tag{F.21}$$

$$\text{tube slack} = 0.0926107. \tag{F.22}$$

The finite-bank geometry and strict-tube constants are

$$\Delta_{\text{orbit}} = 0.1568, \tag{F.23}$$

$$D_{\text{orbit}} = 2, \tag{F.24}$$

$$w_{\min} = 0.25, \tag{F.25}$$

$$L_{\text{prior}} = 3, \tag{F.26}$$

$$\ell_M = 4.7874917, \tag{F.27}$$

$$\theta_M = 0.1069303, \tag{F.28}$$

$$r_{\text{ret}}(M) = 0.0534651. \tag{F.29}$$

Consequently,

$$B_T = 2.7912\times10^{-7}, \tag{F.30}$$

$$R^{\text{bias}}_{M,a}(T) = 0.18 \tag{F.31}$$

uniformly over the four targets. Thus the positivity, tube-invariance, and bias-corrected-radius hypotheses of Theorem D.2 all pass.

To make strict failures observable without restricting perturbations to one selected direction, and without changing the bank, priors, or complete T-step dynamics, we use the full-dimensional isotropic Gaussian stress model

$$\eta_0 \sim \mathcal{N}(0, \sigma^2 I_3). \tag{F.32}$$

Its covariance has spectral norm $\|\Sigma_0\| = \sigma^2$, so it satisfies the theorem's covariance assumption while perturbing every ambient direction. For each σ, we draw 25,000 independent perturbations at each of the 16 atoms, giving 400,000 queries and 100,000 queries per semantic orbit. Strict failure means $P_s(a) \le 1 - \varepsilon$ for every $s = 0, \dots, T$, exactly matching $\tau^{\text{strict}}_\varepsilon > T$. Closure under C_4 and rotational invariance of the isotropic law make the four poses within an orbit distributionally equivalent. The reported statistical limit is a one-sided Clopper–Pearson 95% simultaneous upper bound, Bonferroni-corrected over the four target orbits. For this three-dimensional isotropic Gaussian, the sharp sampled-bank bound is evaluated as

$$\overline{F}_{\chi^2_3}\big((R^{\text{bias}}_{M,a}/\sigma)^2\big) = \mathbb{P}\big(\chi^2_3 \ge (R^{\text{bias}}_{M,a}/\sigma)^2\big)\,. \tag{F.33}$$

Because $\Sigma_0 = \sigma^2 I_3$, this bound equals the exact Gaussian noise-escape probability. No Monte Carlo sample is used to select or certify any deterministic constant. Source, CSV, JSON, report, and plotting code are included in `experiments/theorem3_failure_bound`.

Additional high-dimensional held-out diagnostic. The earlier diagnostic uses nested finite sampled banks with

$$\begin{aligned} M \in \{8, 16, 32, 64, 96, 128\}, \quad N = Q = 24, \\ T = 6, \qquad \beta = 45. \end{aligned} \tag{F.34}$$

$\varepsilon = \delta = 0.05$, unit-normalized representatives ($R = 1$), and $\sigma \in \{0.03, 0.09, 0.15\}$. Each query is $x_0 = y_0 + \eta_0$, where y_0 is a sampled target representative and $\eta_0 \sim \mathcal{N}(0, \sigma^2 I_N)$. The sharp radius uses $q^{\chi}_{24}(0.05) = 6.03449$.

For each fixed bank, we compute $b_{\max}$ over all sampled orbit representatives and compute $\Lambda_\star(M)$ exactly over the finite bank. Candidate $(\widehat{\kappa}, r_{\text{loc}})$ values are calibrated from 8,000 random tube perturbations, with a fixed additive safety margin of 0.08, then frozen and checked on 4,000 independently seeded tube perturbations. All reported banks have zero validation violations; the exact one-sided binomial upper bound on an unseen violation rate is 7.49×10^{-4} at 95% confidence. This diagnostic does not prove the contraction inequality uniformly over the entire tube, so its sampled "pass" is not used as the formal certificate reported in Exp. 6.

We then evaluate 20 independent Gaussian trials per target orbit from a held-out test seed. Strict success records whether $P_s(a) > 1 - \varepsilon$ for some $s \le T$, matching $\tau_\varepsilon^{\mathrm{strict}} \le T$. A condition passes only when validation has no tube violations, $r_{\mathrm{loc}} - \sigma q_N^\chi(\delta) \ge 0$, tube invariance holds, and

$$\Lambda_\star(M) - \left[\log \frac{1-\varepsilon}{\varepsilon} + 2\beta R \Gamma_T\right] > 0. \tag{F.35}$$

Mean, 5th-percentile, and worst-orbit success are recorded; the worst-orbit rate is the empirical quantity corresponding to the theorem's uniform target quantifier. Source code, frozen configuration, and all 18 result rows are included with the supplementary files.

Orbit-pooling ablation. The ablation compares four implementations on the same sampled representatives: standard point-level Hopfield retrieval, augmented memories with parent-orbit pooling, the same augmented bank with normalized within-orbit weights, and the orbitwise quotient implementation. Balanced trivial-orbit banks test algebraic equivalence. Unequal Q_a banks test sampling-density bias. Repeated representatives mimic stabilizer multiplicity. A clustered setting tests whether unrestricted candidate coverage conceals ambiguity when primary retrieval is weak.

Hardware. All reported runs were executed on an NVIDIA GeForce RTX 4070 Laptop GPU (8,188 MiB) with an Intel Core i9-14900HX CPU, Windows 11 Home 64-bit (version 10.0.26200), and PyTorch 2.5.1+cu121.

G Image-Restoration Protocol and Qualitative Visualizations

Task and scope. The image experiment is a finite-bank associative-recovery demonstration, not a train/test image-classification or generic image-restoration benchmark. For every selected clean image, the memory bank contains its transformed representatives and the query is a noisy transformed version of that same stored exemplar. Thus exact instance identity is known by construction. The purpose is to visualize how a sampled group-orbit update can denoise while remaining compatible with pose; the main paper makes no state-of-the-art restoration or held-out generalization claim from this experiment.

Image orbit bank and update. Images are flattened after mean centering and ℓ_2 normalization. The label $\mathrm{R20}$ denotes the rotation-only bank, with 20 entries per exemplar obtained at $18°, 36°, \ldots, 360°$. The label $\mathrm{A24}$ denotes the augmentation-matched bank: "A" indicates augmentation matching, and $24 = 20 + 4$ records the 20 sampled rotations plus four additional discrete image-symmetry entries per exemplar. This is a pre-quotient count: coincident transforms can yield fewer distinct orbit points. Because interpolation and the finite augmentation set are used in practice, A24 is an augmentation-matched sampled image bank rather than an exact continuous-group discretization. At each step, the implementation computes the within-orbit log-sum-exp evidence, normalizes orbit posteriors, computes conditional pose means, and applies the full posterior-weighted update in Algorithm 1. The displayed recoveries use two complete Hopfield updates. Writing F_{A24} for the sampled update with this bank, the displayed output is

$$x_2 = F_{\mathrm{A24}}(F_{\mathrm{A24}}(x_0))\,. \tag{G.1}$$

Memory/query construction. Fashion-MNIST uses 80 stored 28×28 grayscale images per class (800 memories total) and 50 noisy pose queries per class (500 queries). The displayed A24 run uses $\beta = 80$ and 20% Gaussian-noisy augmented queries. CIFAR-10 uses 35 stored 32×32 RGB images per class (350 memories total) and 25 noisy pose queries per class (250 queries), with $\beta = 52$ and 10% Gaussian-noisy augmented queries. For each query, one of five independently prepared augmentations of its selected stored image is sampled. Therefore the memory/query pairing is intentional and should not be interpreted as a leakage-safe supervised split.

Diagnostics. The stored-image orbit is the target. We record instance and class top-1 readout, target-orbit distance before and after recovery, posterior mass on the target orbit, and wall-clock time. The A24 two-step update reaches instance/class readout of $0.998/1.000$ on Fashion-MNIST and $0.888/0.912$ on CIFAR-10; the no-orbit point-memory control reaches $0.064/0.204$ and $0.188/0.352$,

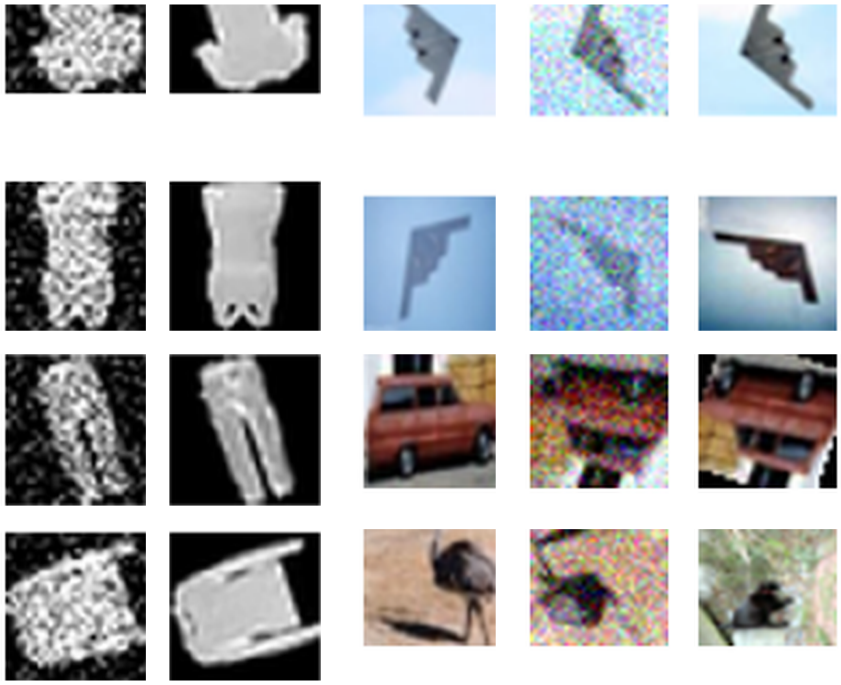


Figure 4: Qualitative image-orbit restoration. In every row, the first two panels are a Fashion-MNIST noisy query and its two-step orbit-Hopfield recovery (Xiao et al., 2017); the final three panels are a CIFAR-10 clean stored memory, rotated/noisy query, and recovery (Krizhevsky et al., 2009). The finite A24 (augmentation-matched, 24-entry) bank comprises 20 rotations and four discrete symmetries.

respectively. These diagnostics are consistent with the visual examples, but their in-memory construction is why they are reported here rather than used as a central quantitative benchmark.

Figure 4 shows representative image-recovery examples selected from the archived image-restoration experiments. The accompanying archive contains the data preparation, `real_dataset_orbit_recovery.py`, per-query records, noise sweeps, and additional recovery grids.

Figure 5 visualizes the real point-cloud setting used in Exp. 7. Each row uses one ModelNet40 class from the 11-class subset. Training prototypes are expanded into sampled $\mathrm{SO}(3)$ orbit poses, while the held-out query is independently rotated and corrupted; prototypes and queries are drawn from disjoint splits.

H Extended Result Tables and Plots

This appendix contains the complete sweeps summarized in Exp. 2–4 and 6 of the main paper. They are supplementary because the main paper states the result-level conclusions and retains the experiments that directly support the central algorithmic and real-data claims.

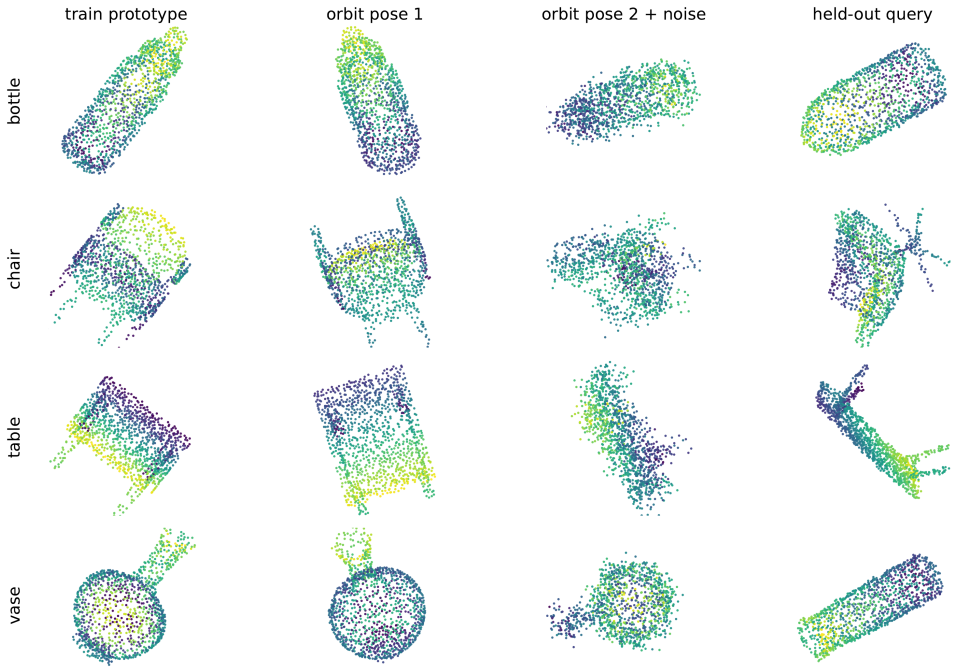


Figure 5: Qualitative ModelNet40 point-cloud examples for Exp. 7. Columns show a training prototype, two sampled orbit poses from the memory bank, and a held-out rotated/noisy query. The second orbit pose includes the small bank noise used in the search database.

Configuration	Q	T	β	$p_{\mathrm{acc}}^{(T)}$	$p_{\mathrm{cov}}^{(T)}$	Point exact	Point class	Eq. err.
Q=120 fixed	120	12	60.00	0.8854	1.0000	0.3203	0.3776	0.4382
Q=320 fixed	320	12	60.00	0.7500	1.0000	0.3203	0.3776	0.2759
Q=320 adapt.	320	18	78.68	0.9323	1.0000	0.3203	0.3776	0.2939

Table 5: Supplementary Exp. 5: ModelNet10 real-data orbit-retrieval simulation with 30 stored CAD shapes. Nearest sampled-orbit exact retrieval is 1.0000 in all rows; $p_{\mathrm{cov}}^{(T)} = 1.0000$ means the target remains in the reported candidate set and is not, by itself, a selectivity claim.

Setting	Point	Naive	Quot.	Set cov.	Bias diagnostic
Balanced orbits	0.275	1.000	1.000	1.000	–
Unequal Q_a	0.244	0.694	0.744	0.744	low-Q: 0.415 $\rightarrow$ 0.511
Stabilizer duplicates	0.233	0.661	0.933	0.933	dup.: 0.670 $\rightarrow$ 0.055
Metastable clusters	0.067	0.389	0.389	0.906	coverage only

Table 6: Supplementary orbit-pooling ablation. Point denotes point-level Hopfield retrieval; Naive pools posterior mass over augmented pose memories. Quotient uses normalized within-orbit weights and removes duplicate representatives. The maximum equivalence error between normalized augmented pooling and the orbitwise quotient update is below 1.3×10^{-13}. Set cov. is unrestricted target coverage and is not interpreted as size-controlled retrieval.

Q	fixed $p_{\mathrm{acc}}^{(T)}$	$p_{\mathrm{cov}}^{(T)}$	$\mathrm{Err}_{\mathrm{eq}}$	adaptive $p_{\mathrm{acc}}^{(T)}$
20	0.957	0.998	0.921	0.957
40	0.941	1.000	0.731	0.936
80	0.940	1.000	0.610	0.922
160	0.891	1.000	0.417	0.967
320	0.740	1.000	0.300	0.965
640	0.461	1.000	0.214	0.928

Table 7: Supplementary Exp. 2: denser orbit sampling improves approximate equivariance but lowers fixed-budget convergence-and-accuracy. The adaptive column uses larger T and β for large Q. At $Q = 640$, the fixed and adaptive entries are the paired runs summarized in the main paper; the independent full fixed-budget sweep gives 0.4707.

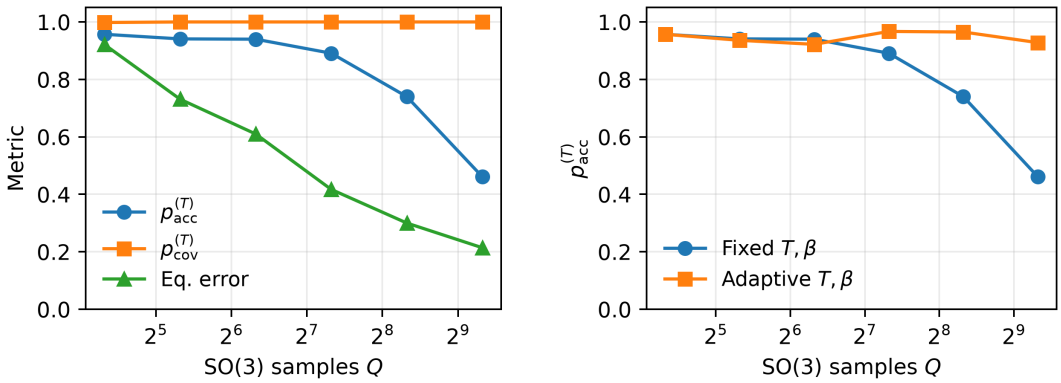


Figure 6: Supplementary Exp. 2. Left: Q improves sampled equivariance but may reduce fixed-budget $p_{\mathrm{acc}}^{(T)}$. Right: a Q-adaptive T/β schedule restores high-Q finite-step accuracy.

α	σ	Acc.	Cov.	Sel.@10
0.003	0.10	0.930	1.000	0.492
0.003	0.25	0.928	1.000	0.499
0.010	0.10	0.924	1.000	0.501
0.050	0.10	0.932	1.000	1.000

α	σ	Primary	Set prec.	Size
0.003	0.10	0.046	0.050	20.0
0.003	0.25	0.051	0.050	20.0
0.010	0.10	0.051	0.050	20.0
0.050	0.10	1.000	0.050	20.0

Table 8: Supplementary Exp. 3: unrestricted coverage is 1.0 but hides a 20-way candidate set. The fixed-budget $K = 10$ metric exposes chance-level selectivity at low α; set precision records the burden of returning the full stable set.

α	$K = 1$	$K = 3$	$K = 5$	$K = 10$	$K = 20$
0.003	0.043	0.150	0.248	0.501	1.000
0.010	0.050	0.154	0.246	0.500	1.000
0.050	1.000	1.000	1.000	1.000	1.000

Table 9: Exp. 3 K-sensitivity: mean $p_{\mathrm{sel}}^{(T)}(K)$ over five seeds. At low separation, selection tracks the candidate budget within a 20-way ambiguous set; increased separation restores single-orbit selection.

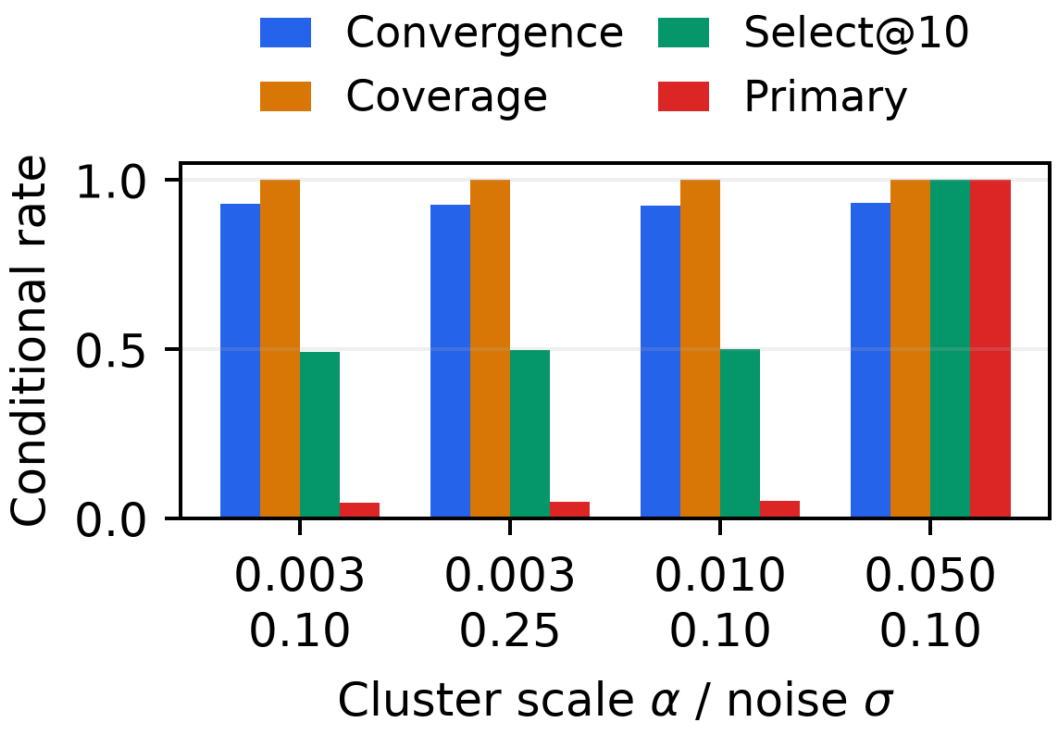


Figure 7: Supplementary Exp. 3: the unrestricted set preserves target coverage at low α, while primary and size-controlled retrieval reveal the intended ambiguity.

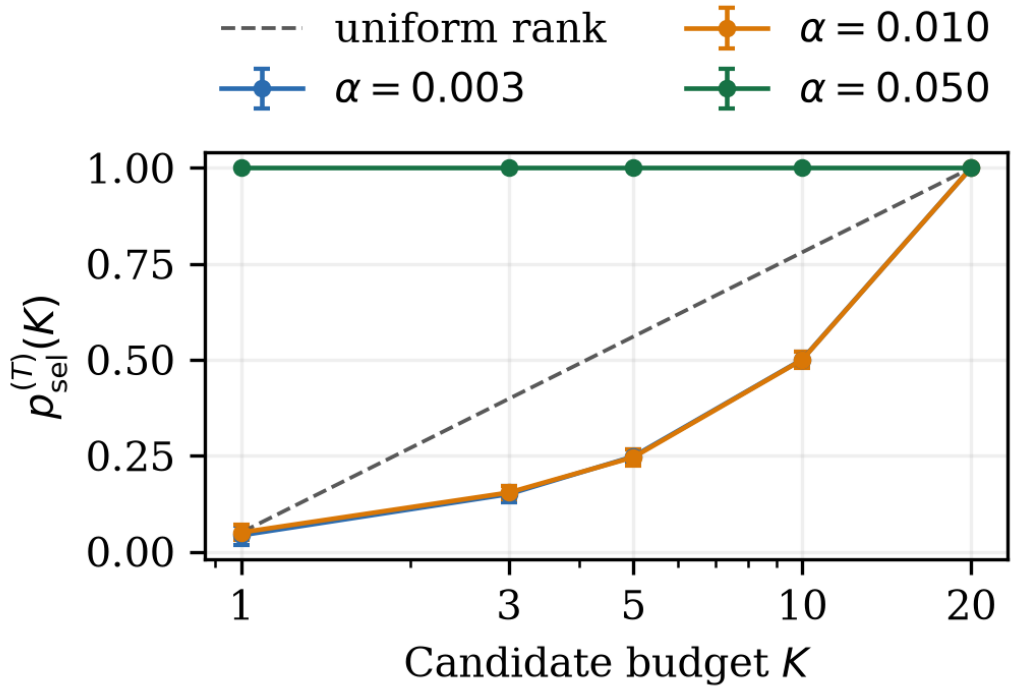


Figure 8: Exp. 3 K-sensitivity with seed standard deviations. Perfect unrestricted coverage at low α does not imply selective retrieval.

Sweep	Level	q	Q	Δ_{orbit}	H	exact $\Lambda_\star$	final strict P_T
Separation	0.12	3	1000	0.120	5.684	0.310	0.000 ± 0.000
	0.16	3	1000	0.160	5.684	1.205	0.000 ± 0.000
	0.20	3	1000	0.200	5.684	2.096	0.000 ± 0.000
	0.24	3	1000	0.240	5.684	2.983	0.023 ± 0.006
	0.28	3	1000	0.280	5.684	3.864	1.000 ± 0.000
	0.32	3	1000	0.320	5.684	4.741	1.000 ± 0.000
	0.40	3	1000	0.400	5.684	6.474	1.000 ± 0.000
	0.44	3	1000	0.440	5.684	7.329	1.000 ± 0.000
	0.48	3	1000	0.480	5.684	8.176	1.000 ± 0.000
Entropy	$Q = 32$	3	32	0.240	3.466	3.250	1.000 ± 0.000
	$Q = 64$	3	64	0.240	4.158	3.184	1.000 ± 0.000
	$Q = 128$	3	128	0.240	4.848	3.107	1.000 ± 0.000
	$Q = 256$	3	256	0.240	5.530	3.027	1.000 ± 0.000
	$Q = 512$	3	512	0.240	6.127	2.929	0.973 ± 0.010
	$Q = 1024$	3	1024	0.240	6.359	2.861	0.797 ± 0.028
Dimension	$q = 1$	1	1728	0.235	2.503	3.106	1.000 ± 0.000
	$q = 2$	2	1728	0.235	3.871	2.974	0.849 ± 0.017
	$q = 3$	3	1728	0.235	5.746	2.832	0.000 ± 0.000

Table 10: Controlled Exp. 4 at fixed $M = 12$. Larger separation improves the exact gap and final strict retrieval; larger finite-orbit entropy and quotient dimension reduce both. Values are means and seed standard deviations over five seeds. The complete CSV also reports initial strict retrieval, top-1 retrieval, mean target posterior, and quadrature error.

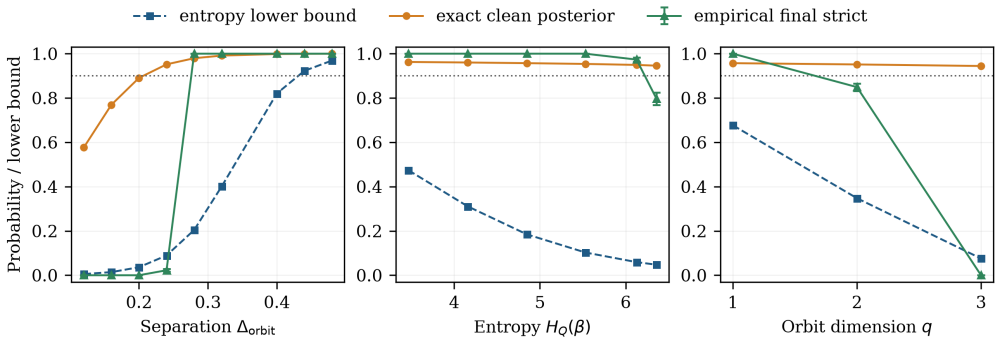


Figure 9: Controlled Exp. 4. The blue curve is the conservative clean posterior lower bound induced by $\beta\Delta_{\mathrm{orbit}} - H_M(\beta) - \log(M-1)$; orange is the exact clean posterior, and green is empirical final strict success. The dotted line is the $1 - \varepsilon = 0.9$ threshold. Exact-gap Spearman correlations with separation, entropy, and dimension are $1, -1, -1$; empirical strict success gives $0.894, -0.845, -1$, respectively. The separation lower bound first crosses the threshold at $\Delta_{\mathrm{orbit}} = 0.44$. Agreement is directional: the lower certificate remains deliberately conservative.

Scenario	Strict succ.	Set cov.	Cert. pass	$\mathbb{E}[\Lambda_T]$	$\mathbb{E}[\underline{\Lambda}_T]$	5% $\underline{\Lambda}_T$	Slack
Separated random	1.000	1.000	1.000	34.56	32.80	25.41	29.85
Close pairs	0.975	0.975	0.906	11.73	8.41	1.21	5.47
Metastable cluster	0.356	0.775	0.281	-6.31	-48.50	-100.93	-51.44

Table 11: Final-state free-energy diagnostic. Strict is posterior-collapse success and Set cov. is target coverage by the unrestricted stable set, not selective retrieval. Here $\underline{\Lambda}_T = \Lambda_a(y_\star) - 2\beta d_a(x_T)$, and Slack is $\mathbb{E}[\underline{\Lambda}_T] - \log 19$. This earlier diagnostic is conservative and does not replace the complete finite-cover certificate in Supplementary Exp. 6.

Mask	T-step	matched TTA	initial	nearest	prototype	p_{cl}	time: Orbit/TTA (ms)
0.50	0.2992 ± 0.0086	0.2606 ± 0.0069	0.2439	0.2439	0.2280	0.0790	4.91 / 12.49
0.65	0.2879 ± 0.0125	0.2424 ± 0.0139	0.2220	0.2189	0.2295	0.0440	4.86 / 12.29
0.80	0.2652 ± 0.0080	0.2045 ± 0.0079	0.1856	0.1848	0.2023	0.0088	5.03 / 11.91
0.90	0.2591 ± 0.0127	0.1462 ± 0.0139	0.1795	0.1765	0.1962	1.26×10^{-7}	4.88 / 12.64

Table 12: Exp. 7 mask-severity and compute audit. TTA uses three rotated encoder evaluations with the same observed-coordinate mask as the query. Orbit time includes one encoder evaluation and two complete Hopfield updates. The validation-selected Orbit-Hopfield schedule is fixed across the sweep; p_{cl} is the exact two-sided shape-cluster sign-flip comparison with matched-mask TTA.

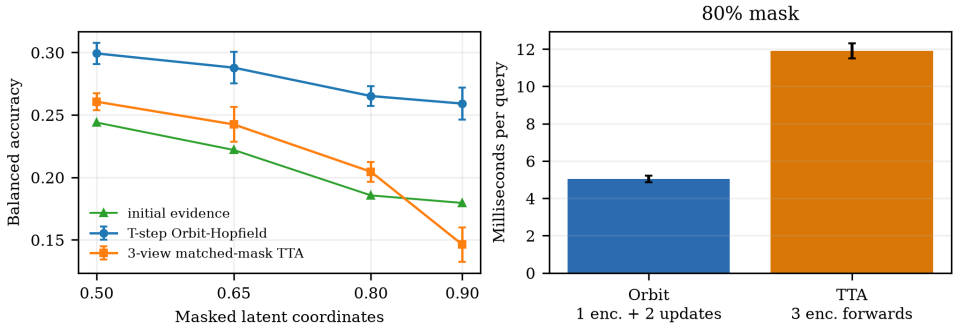


Figure 10: Exp. 7 matched-task comparison. Left: balanced accuracy under increasing latent masking. Right: measured query cost at 80% masking. Error bars are bank-seed standard deviations.

M	$\sum_a \lvert\mathcal{G}_a\rvert$	grid max $\lVert J_F\rVert_2$	$\widehat{\kappa}$	$b_{\max}$	tube slack	$\Lambda_\star$	gap slack	cert.	worst strict
2	3952	0.0553	0.0907	0.0031	0.0287	24.206	21.885	yes	1.000
3	5924	0.2373	0.3267	0.0122	0.0113	3.976	1.134	yes	1.000
4	7896	0.4959	0.6441	0.0225	-0.0101	2.548	-1.427	no	1.000
5	9840	0.7129	0.8560	0.1022	-0.0972	0.026	-11.948	no	0.000
6	11812	0.9721	1.2158	0.1029	-0.1104	-0.020	-16.624	no	0.000
7	13784	0.9723	1.2161	0.0909	-0.0985	-0.064	-15.081	no	0.000

Table 13: Supplementary Exp. 6: deterministic finite-cover numerical audit. The grid column sums the retained lattice points over all target tubes. Tube slack is $r - (b_{\max} + \widehat{\kappa} r)$; gap slack is $\Lambda_\star - [\log((1-\varepsilon)/\varepsilon) + 2\beta R \Gamma_T]$. The pass column evaluates the uniform bound of Theorem C.1 in floating point; the Gaussian trials in the final column are diagnostic only.

σ	strict failures	worst freq.	simultaneous 95% UCB	noise escape emp.	Thm. 3 sharp bound	incl. viol.
0.075	14/400000	7.00×10^{-5}	1.5624×10^{-4}	0.12384	0.12389	0
0.080	43/400000	2.10×10^{-4}	3.3821×10^{-4}	0.16678	0.16728	0
0.085	125/400000	5.60×10^{-4}	7.5238×10^{-4}	0.21342	0.21368	0
0.090	224/400000	0.00110	0.001359	0.26216	0.26146	0
0.095	433/400000	0.00208	0.002427	0.30994	0.30927	0
0.100	665/400000	0.00296	0.003369	0.35587	0.35608	0

Table 14: Direct audit of the explicit Gaussian failure bound for the exact sampled-bank corollary (Theorem D.2), using full-dimensional isotropic Gaussian noise. Strict failure is nonzero and rises with σ, while its worst-orbit frequency remains below the sharp chi-square bound. "Incl. viol." counts strict failures occurring while $\lVert\eta_0\rVert < R^{\mathrm{bias}}_{M,a}(T)$; none are observed.

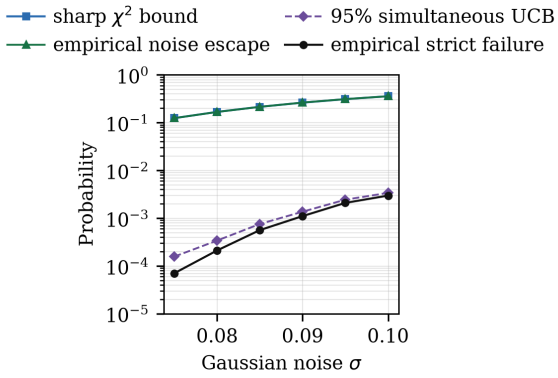


Figure 11: Probability-chain audit for the exact sampled-bank Gaussian corollary (Theorem D.2). The sharp χ^2_3 bound and the empirical isotropic-noise escape frequency nearly coincide. Black circles show the observed worst-orbit strict-failure frequency; diamonds show its simultaneous 95% upper confidence limit. No failure occurs inside the certified radius.

H.1 Validation-frozen continuous comparison protocol

The controlled experiment compares three numerical paths for the same exact orbit objective:

1. weighted deterministic group sampling;
2. ordinary continuous evaluation;
3. interval-certified continuous evaluation used as the reference.

All methods execute the complete five-step dynamics with $M = 8$. The $\mathrm{SO}(2)$ representation has frequencies $\{1, 2, 4, 7\}$, multiplicity four per frequency, and $d = 32$. The $\mathrm{SO}(3)$ representation has 32 fixed-correspondence vectors and $d = 96$. Queries are independently transformed versions of prototype zero with additive Gaussian noise 0.08 and 0.04, respectively, followed by L_2 normalization.

For each $\beta \in \{1, 10, 50\}$, five validation seeds $\{110, \ldots, 114\}$ select the smallest sampled grid whose maximum log-evidence error, orbit-posterior total variation, one-step state error, and five-step

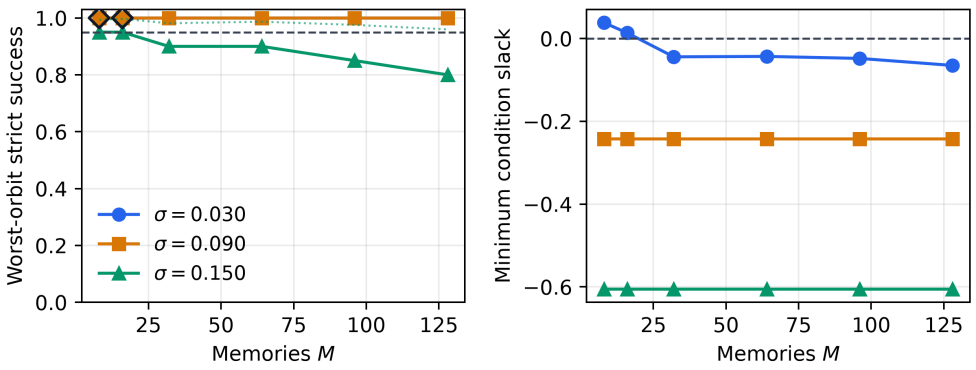


Figure 12: Additional high-dimensional diagnostic. Solid curves are worst-orbit strict success, dotted curves are means, and open diamonds identify rows passing every audited condition. The lower panel shows the minimum normalized condition slack; the sufficient implication is tested only where this value is positive.

M	σ	$\widehat{\kappa}$	$b_{\max}$	$\Lambda_\star$	tube slack	gap slack	pass	mean strict	worst strict
8	0.03	0.313	0.0578	13.94	0.119	3.42	yes	1.000	1.000
8	0.09	0.313	0.0578	13.94	-0.243	3.39	no	1.000	1.000
8	0.15	0.313	0.0578	13.94	-0.605	3.36	no	0.994	0.950
16	0.03	0.356	0.0586	12.35	0.119	1.20	yes	1.000	1.000
16	0.09	0.356	0.0586	12.35	-0.243	1.13	no	1.000	1.000
16	0.15	0.356	0.0586	12.35	-0.605	1.06	no	0.997	0.950
32	0.03	0.338	0.0586	6.92	0.119	-3.99	no	1.000	1.000
32	0.09	0.338	0.0586	6.92	-0.243	-4.04	no	1.000	1.000
32	0.15	0.338	0.0586	6.92	-0.605	-4.09	no	0.981	0.900
64	0.03	0.330	0.0586	6.92	0.119	-3.90	no	1.000	1.000
64	0.09	0.330	0.0586	6.92	-0.243	-3.95	no	1.000	1.000
64	0.15	0.330	0.0586	6.92	-0.605	-3.99	no	0.987	0.900
96	0.03	0.365	0.0586	6.92	0.119	-4.34	no	1.000	1.000
96	0.09	0.365	0.0586	6.92	-0.243	-4.42	no	1.000	1.000
96	0.15	0.365	0.0586	6.92	-0.605	-4.49	no	0.976	0.850
128	0.03	0.461	0.0586	6.92	0.119	-5.86	no	1.000	1.000
128	0.09	0.461	0.0586	6.92	-0.243	-6.18	no	1.000	1.000
128	0.15	0.461	0.0586	6.92	-0.605	-6.49	no	0.960	0.800

Table 15: Additional high-dimensional held-out capacity diagnostic. Tube slack is $r_{\mathrm{loc}} - \sigma q_N^{\chi}(\delta)$; gap slack is $\Lambda_\star - [\log((1-\varepsilon)/\varepsilon) + 2\beta R\Gamma_T]$. Pass requires both slacks, tube invariance, and the independent validation check to succeed. Empirical local constants make this an audit of the theorem conditions on sampled points, not a formal uniform certificate.

final-state error are all at most 10^{-6}. The selected Q is then frozen before evaluating test seeds $\{210, \ldots, 214\}$. No test value enters grid selection.

	SO(2)			SO(3)		
β	1	10	50	1	10	50
Frozen Q	64	128	256	512	4096	64000
Validation pass	yes	yes	yes	yes	yes	yes

Table 16: Validation-selected sampling budgets.

Ordinary CPU timings use one BLAS thread, one warm-up call, and 30 randomized interleaved repetitions for every test seed and method. The reported standard deviations are sample standard deviations across all 150 retained timings. Strict certification uses five repetitions per seed because it is an offline path. The software environment and every timing sample are stored in `continuous/results/formal_three_method_comparison/formal_experiment.json`.

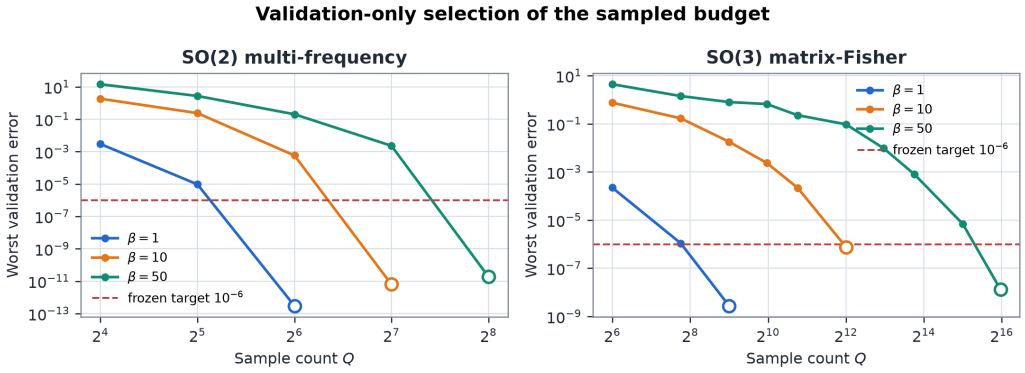


Figure 13: Full validation paths used to freeze Q. The selected point is outlined. The maximum includes log evidence, posterior total variation, one-step state error, and five-step final-state error.

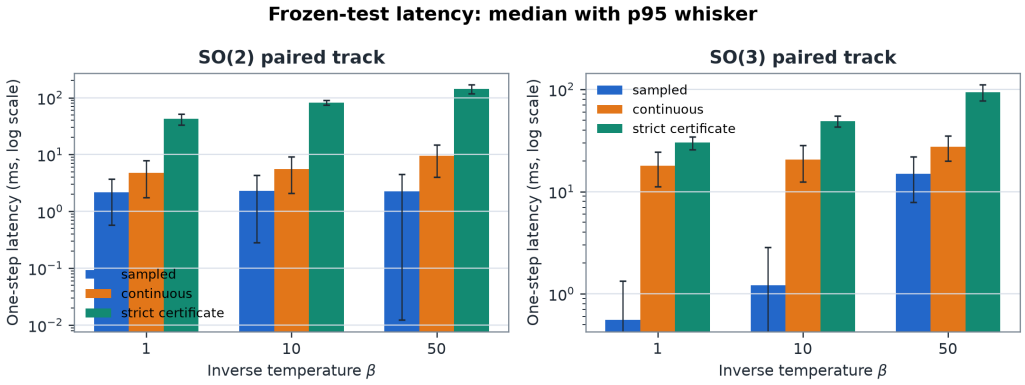


Figure 14: One-step frozen-test latency. Bars show medians and upper whiskers show p95. Ordinary continuous methods are not universally faster than small cached grids; strict interval certification is an offline path.

H.2 Complete regular-condition results

Track	β	Q	sampled T-error	continuous T-error	T-step ms: sampled / continuous
SO(2)	1	64	6.91×10^{-17}	1.40×10^{-14}	12.51±2.43/23.51±3.54
	10	128	1.78×10^{-10}	6.82×10^{-14}	12.28±3.39/29.83±5.74
	50	256	1.28×10^{-10}	1.10×10^{-14}	12.84±3.32/46.81±9.45
SO(3)	1	512	2.14×10^{-15}	1.23×10^{-17}	3.15±1.24/87.57±13.71
	10	4096	2.98×10^{-7}	1.47×10^{-15}	7.39±2.39/98.91±12.29
	50	64000	2.76×10^{-9}	1.07×10^{-15}	76.13±8.00/142.03±13.94

Table 17: Maximum final-state error against the strict continuous reference over five test seeds, and complete five-step latency.

H.3 High-concentration SO(3) crossover

A separate stress protocol fixes $\beta = 200$, $M = 4$, and $T = 1$. Validation seeds $\{410, \ldots, 414\}$ choose the smallest Euler–Legendre product grid satisfying the same 10^{-6} target. The frozen choice is $Q = 262{,}144$. Five disjoint test seeds $\{510, \ldots, 514\}$ are then timed with 30 randomized interleaved repetitions.

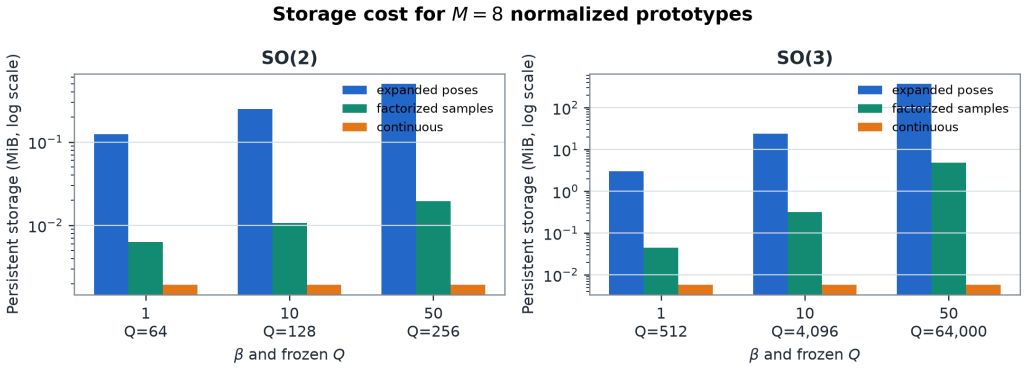


Figure 15: Persistent storage for eight prototypes. Continuous methods store no pose bank, whereas expanded and factorized sampled storage grows with the validation-selected Q.

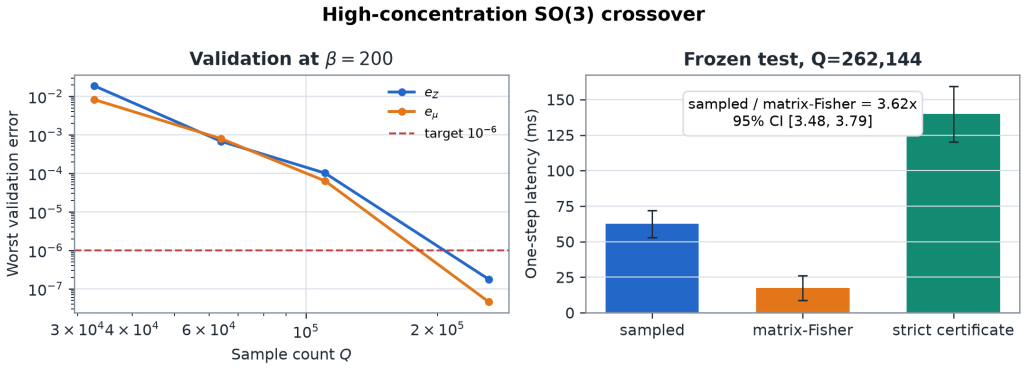


Figure 16: High-concentration crossover. The left panel shows validation grid selection; the right panel reports the frozen test timing.

Method	median ms	p95 ms	persistent storage
Sampled	62.53	72.02	20.0 MiB factorized
Matrix-Fisher ordinary	17.40	26.08	3 KiB
Matrix-Fisher certified	139.77	159.33	3 KiB

Table 18: Frozen high-concentration $\mathrm{SO}(3)$ test. An expanded sampled pose bank would occupy 768 MiB.

The paired geometric-mean speed ratio is

$$\frac{t_{\text{sampled}}}{t_{\text{matrix-Fisher}}} = 3.62, \qquad \mathrm{CI}_{95\%} = [3.48, 3.79]. \tag{H.1}$$

The interval comes from 10,000 paired bootstrap resamples of log timing ratios.

H.4 Interpretation and Remaining Boundary

The controlled results establish numerical agreement and characterize a resource crossover; they do not establish a universal continuous-method speedup or a new real-data accuracy gain. Regular cached grids are faster in all six primary conditions. Continuous evaluation is attractive when pose-bank storage is limiting, exact Haar symmetry is required, or concentration forces large Q. The current matrix-Fisher implementation is deliberately restricted to fixed-correspondence vector channels, and the current Fourier/Bessel implementation requires a finite known $\mathrm{SO}(2)$ frequency decomposition. General higher-order $\mathrm{SO}(3)$ features, end-to-end differentiable training, GPU kernels, and real-data encoders exposing compatible channels remain future work.

I Real-Data Protocols and Complete Results

Scope. This appendix documents the ScanObjectNN (Uy et al., 2019) and QM9 (Ramakrishnan et al., 2014) mechanism comparisons added to the ICLR version. Every G-MHN row uses the finite sampled update in Algorithm 1. No analytic continuous-integration update is used.

I.1 ScanObjectNN

Data and leakage control. We use the official OBJ_BG HDF5 train and test releases. For each seed, 750 objects are sampled from the official training partition for fitting and a disjoint 225-object subset is used for validation. The evaluation set contains 255 objects from the official test partition. Each cloud is centered, scaled to unit radius, and resampled to 256 points. Rotations and additive background noise are generated independently after the split. Prototypes are selected from training features only.

Models and bank. The compact DGCNN control (Wang et al., 2019) uses dynamic $k = 12$ EdgeConv neighborhoods and a 256-dimensional embedding. The PCA-DGCNN control canonicalizes each cloud by principal axes before the same encoder. Sign ambiguities are resolved deterministically. The SO(2)-Fourier control uses 32 scalar channels, 12 Fourier channels, and a 33,915-parameter classifier. These are mechanism-matched controls, not reproductions of the complete LocoTrans or CSEConv systems (Chen et al., 2024; Kim et al., 2024). Training uses six epochs, batch size 16, learning rate 10^{-3}, and weight decay 10^{-4}.

Each G-MHN bank stores two prototypes per class (30 semantic orbits), selected by class-conditional farthest-first traversal of normalized training features. Each orbit has eight uniformly spaced in-plane rotations. We tune $\beta \in \{8, 16, 32\}$ and $T \in \{1, 2, 3\}$ using validation balanced accuracy only. Damping is 0.7. The matched TTA baseline uses the same eight rotations and averages classifier probabilities. Latency includes all encoder calls, orbit construction required at inference, and retrieval, measured on the same CUDA device.

Encoder	Rule	Balanced accuracy	End-to-end ms/query
Compact DGCNN	single view	0.1987 ± 0.0082	1.243 ± 0.376
	8-view TTA	0.2052 ± 0.0120	8.017 ± 1.656
	sampled G-MHN	0.1359 ± 0.0060	2.051 ± 0.238
PCA-DGCNN control	single view	0.2497 ± 0.0082	1.825 ± 0.017
	8-view TTA	0.2497 ± 0.0082	9.643 ± 0.124
	sampled G-MHN	0.2131 ± 0.0294	2.540 ± 0.078
SO(2)-Fourier control	single view	0.1634 ± 0.0460	0.851 ± 0.029
	8-view TTA	0.1634 ± 0.0460	3.438 ± 0.031
	sampled G-MHN	0.0954 ± 0.0185	2.289 ± 0.245

Table 19: Complete ScanObjectNN summary over seeds 11, 17, and 23.

Encoder	seed 11	seed 17	seed 23
Compact DGCNN	$(8, 1)$	$(32, 1)$	$(32, 1)$
PCA-DGCNN	$(32, 1)$	$(32, 1)$	$(8, 2)$
SO(2)-Fourier	$(8, 1)$	$(8, 2)$	$(32, 2)$

Table 20: Validation-selected (β, T) for ScanObjectNN.

I.2 QM9

Data and targets. We use 5,000 real molecules from the PyTorch-Geometric preprocessing of QM9, with fixed disjoint subsets of 1,200 training, 400 validation, and 400 test molecules for each seed. The targets are dipole moment μ, isotropic polarizability α, and HOMO energy. Target normalization is fit on the training subset only. All models train for eight epochs, while the memory adapter trains for three additional epochs with the encoder frozen.

Controls and memory. The distance-message control uses distances and radial basis features. The scalar-vector control propagates scalar and vector channels. The equivariant-attention control adds invariant attention weights over equivariant messages. They isolate increasingly rich mechanisms found in $\mathrm{SE}(3)$-Transformer and Equiformer (Fuchs et al., 2020; Liao & Smidt, 2023) but are not official implementations. For G-MHN, 24 training-only prototypes are selected by quotient k-medoids from 128 candidates. Each has 128 fixed sampled $\mathrm{SO}(3)$ poses. Validation chooses $\beta \in \{4, 8, 16\}$ and $T \in \{1, 2, 3\}$; all seeds choose $(16, 1)$. Rotation discrepancy is the maximum absolute prediction change over the held-out rotation audit.

Method	μ MAE	α MAE	HOMO MAE	mean nMAE	ms/mol.
Distance message	0.9709 ± 0.0398	4.0075 ± 0.2725	0.5256 ± 0.0237	0.58615 ± 0.00540	0.281
Scalar-vector	0.7463 ± 0.0671	4.9515 ± 0.2974	0.4836 ± 0.0147	0.55262 ± 0.02050	0.455
Equivariant attention	0.7281 ± 0.1015	4.6537 ± 0.3940	0.4663 ± 0.0230	$\mathbf{0.52995 \pm 0.00412}$	0.510
Attention + sampled G-MHN	0.7273 ± 0.1189	4.7803 ± 0.4910	0.4651 ± 0.0178	0.53394 ± 0.00856	0.745

Table 21: Complete QM9 target-wise test summary over three seeds. Units follow the standard QM9 target definitions; normalized mean MAE is used only to aggregate targets.

Method	parameters	max rotation discrepancy
Distance message	19,107	1.31×10^{-6}
Scalar-vector	44,788	6.26×10^{-7}
Equivariant attention	48,908	7.15×10^{-7}
Attention + sampled G-MHN	48,908	1.31×10^{-1}

Table 22: QM9 model size and rotation audit. The G-MHN adapter has no learned encoder parameters, but its finite sampled bank is not group closed.

I.3 Interpretation and non-comparability

The controlled studies support three restricted conclusions. First, sampled G-MHN can replace repeated encoder evaluations by cheaper latent scans, but the gain depends on the encoder-to-scan cost ratio. Second, $T > 1$ is useful only when the learned representation produces a trajectory with increasing semantic separation; most ScanObjectNN validation runs select one step. Third, attaching a sampled orbit bank after an already invariant encoder can break numerical invariance unless the finite sample set is group closed or the sampling error is otherwise controlled.

No result in Tables 19–22 is presented as a state-of-the-art comparison. Official implementations use different capacity, augmentation, optimization, and preprocessing. The purpose here is to test compatibility under matched data splits and one measurement harness.

J rMD17 Continuous Retrieval and Completion

This appendix documents the real-data experiment used to evaluate the continuous matrix-Fisher backend. The task is hidden-atom coordinate completion under an unknown global rotation. It is not an energy or force-prediction benchmark, and no encoder is trained.

J.1 Dataset, splits, and leakage control

We use ethanol and malonaldehyde from rMD17 (Christensen & von Lilienfeld, 2020). Each molecule contains 100,000 nine-atom conformations. The published file hashes are checked before use. For each of the five official splits, the 1,000 training indices are divided into 800 design structures and a validation candidate pool. Validation candidates must be at least 50 original trajectory frames from every design structure. Eligible official-test targets must be at least 50 frames from the complete official training side. This buffer reduces direct temporal-neighbor leakage; the five official splits still reuse one molecular trajectory and are not five independent molecular datasets.

From the 800 design structures, we select $M = 32$ prototypes by farthest-point traversal on the vector of all atom-pair distances, starting from the structure nearest the design-set centroid. Test targets never enter the bank. Each query uses the same observed-atom mask for every prototype, is centered

by the noisy observed-atom centroid, and retains physical Angstrom units. Four conditions cross $n_{\mathrm{obs}} \in \{4, 6\}$ with coordinate noise $\sigma \in \{0.05, 0.15\}$ Angstrom. Every split-condition cell contains 12 validation and 40 test queries.

J.2 Methods and validation selection

All methods receive the same bank, mask, centered query, and atom correspondence.

- **Zero fill** leaves hidden centered coordinates at zero.
- **Nearest Procrustes** aligns every prototype to observed atoms and returns the single nearest aligned conformation.
- **Soft Procrustes** applies a softmax to the same best-alignment scores and averages aligned conformations.
- **Continuous mean** uses the matrix-Fisher orbit evidence and conditional rotation mean, then mixes semantic prototypes. This is a direct $T = 0$ readout, not a recurrent update.
- **Clamped G-MHN** initializes from the continuous mean, reapplies the complete update to hidden coordinates, and clamps observed coordinates after every step.
- **Unconstrained G-MHN** starts from zero and updates all coordinates.

For each method, molecule, split, observation count, and noise level, validation hidden-atom RMSD selects $\beta \in \{25, 100, 400\}$ and, where applicable, $T \in \{1, 3, 5, 10\}$. The direct and soft methods select $T = 0$. Across 40 split-condition selections, $\beta = 25$ is chosen 35 times for each direct method; clamped retrieval chooses $T = 1$ in 35 cases. No test result enters selection.

J.3 Complete hidden-coordinate results

Molecule	$(n_{\mathrm{obs}}, \sigma)$	zero	nearest	soft	continuous	clamped	unconstrained
ethanol	$(4, 0.05)$	1.8222±0.0357	0.8041±0.0879	0.6182±0.0467	0.6204±0.0456	0.7030±0.1173	1.2271±0.1033
	$(4, 0.15)$	1.8214±0.0377	0.8870±0.0429	0.6850±0.0248	0.6860±0.0252	0.7938±0.0967	1.2476±0.1113
	$(6, 0.05)$	1.7403±0.0094	0.5067±0.0305	0.4346±0.0593	0.4348±0.0589	0.4740±0.0422	0.7537±0.0755
	$(6, 0.15)$	1.7440±0.0113	0.5201±0.0788	0.4637±0.0587	0.4638±0.0580	0.4905±0.0545	0.7595±0.0734
malonaldehyde	$(4, 0.05)$	2.0142±0.0252	0.9932±0.0655	0.7890±0.0344	0.7897±0.0345	0.8776±0.0837	1.4171±0.0955
	$(4, 0.15)$	2.0175±0.0240	1.1431±0.0831	0.8831±0.0374	0.8839±0.0380	1.0216±0.0756	1.4410±0.1117
	$(6, 0.05)$	1.9395±0.0439	0.6991±0.0807	0.6086±0.1051	0.6090±0.1049	0.6458±0.0907	1.0027±0.0503
	$(6, 0.15)$	1.9427±0.0456	0.7340±0.1055	0.6246±0.0716	0.6235±0.0712	0.6865±0.0909	1.0055±0.0535

Table 23: Hidden-atom RMSD in Angstrom, reported as mean ± sample SD over the five official split means.

The continuous mean beats nearest Procrustes in all five splits of every condition. The eight mean paired improvements range from 0.0563 to 0.2592 Angstrom, or 10.8–22.8% relative to nearest retrieval. Because each condition has only five official splits, the exact two-sided sign-flip value is $p = 0.0625$ even when all five paired differences have the same sign. We therefore emphasize consistency and effect size rather than dichotomous significance. Split-clustered bootstrap intervals for all eight paired mean differences exclude zero; raw values are retained in `results/rmd17_completion/summary.json`.

J.4 Matched-accuracy continuous versus sampled backend

This comparison fixes $M = 32$, $\beta = 25$, $T = 3$, four observed atoms, noise 0.05 Angstrom, and zero initialization. Three validation queries select the smallest Euler–Legendre product grid for which maximum full-trajectory state RMSD and terminal-posterior total variation are both at most 10^{-3}. The first passing level is 80 for both molecules, giving $Q = 80^3 = 512{,}000$. Five distinct test queries are then timed five times per backend with a warm cached grid. The scope is the complete three-step kernel and terminal posterior; feature construction is excluded.

The prototype coordinates occupy 6,912 bytes. The factorized sampled implementation retains a 40.96 MB rotation grid and weight array; explicitly materializing the $M \times Q \times 9 \times 3$ float64 feature bank would require 3.54 GB. These are persistent array sizes rather than process peak memory. The continuous backend also uses transient one-dimensional quadrature workspaces.

Molecule	Q	continuous ms	sampled ms	speedup	95% CI
ethanol	512,000	15.31	2615.97	177.23×	[154.08, 201.15]
malonaldehyde	512,000	17.95	2888.23	165.17×	[147.54, 186.12]

Table 24: Frozen-test matched-accuracy latency. Medians pool the five test queries; paired geometric ratios and query-clustered bootstrap intervals are computed before rounding.

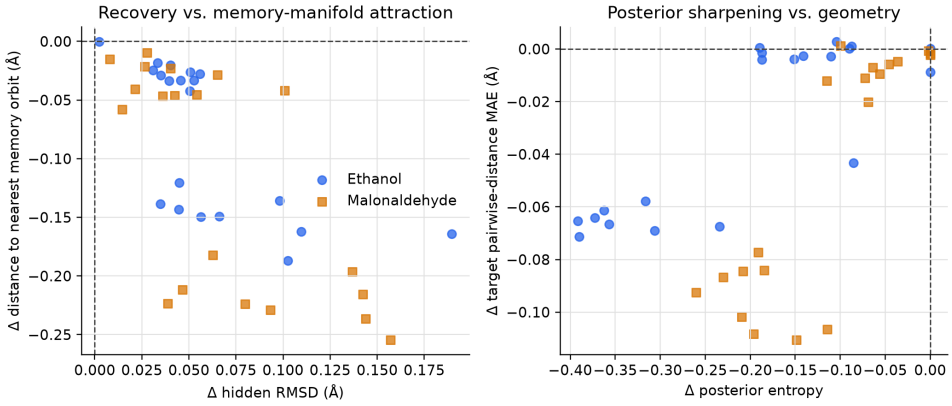


Figure 17: Paired change from the direct conditional mean to the validation-selected clamped recurrent state. Each point is one official split mean in one observation condition. Recurrence consistently attracts the state toward the memory manifold and sharpens the posterior, while hidden-coordinate RMSD generally increases.

J.5 Mean-to-mode behavior of recurrent retrieval

The direct conditional mean and the recurrent clamped state optimize different readouts. Relative to the direct mean, the final clamped state lowers nearest-memory-orbit RMSD in every one of the 40 split-condition cells, with condition means between 0.024 and 0.226 Angstrom. Posterior entropy and Hopfield energy also decrease in all 40 cells; across all 80 recurrent-method cells, the largest recorded energy increase is 1.96×10^{-14}, consistent with floating-point noise. With four observed atoms, clamping reduces target pairwise-distance MAE in every split, with condition means between 0.052 and 0.095 Angstrom. Hidden-atom RMSD increases because the posterior mean, rather than a posterior mode, minimizes expected squared error.

The practical policy follows directly: use the continuous $T = 0$ conditional mean for squared-error completion, and use validation-selected clamped recurrence when returning a sharper, stored-like state is more important than minimum coordinate RMSD. Full paired records are in `results/rmd17_completion/plausibility_analysis.json`; the frozen backend calibration and all timing samples are in `results/rmd17_completion/backend_comparison.json`.