RIGOROUS STATE EVOLUTION FOR GENERAL-CORE ASYMMETRIC MULTILINEAR AMP

Anonymous authors
Paper under double-blind review

ABSTRACT

State evolution (SE) provides an asymptotically exact finite-dimensional characterization of approximate message passing (AMP) dynamics in many high-dimensional inference problems. For asymmetric multilinear AMP, however, reusing the same Gaussian tensor across mode contractions and iterations creates dependencies that complicate both the Onsager correction and rigorous SE analysis. Existing tensor-AMP results do not directly cover general-core rectangular models with empirical covariance inputs and Onsager reaction matrices. We derive core-induced matrix-valued maps that govern the signal contribution and Gaussian covariance and whose derivatives yield an exact finite-system factorization of the Onsager reaction. For fully observed Gaussian models of order at least three with a fixed known core and fixed, mode-dependent channel dimensions, we prove fixed-horizon SE for the joint empirical law of planted rows, factor-iterate histories, and effective-field histories within each mode. The theorem allows initialization correlated with the planted factors but independent of the Gaussian disorder. The closed SE recursion tracks signal and iterate overlaps separately and does not require Bayes matching. The proof combines a symmetric multi-species embedding and scalar-query serialization with Gaussian conditioning supported by quantitative bounds on the stacked-response covariance in an auxiliary regularized model. The joint law yields limits for representation statistics and simultaneous convergence of order-two pseudo-Lipschitz losses over fixed finite families of admissible recursions and independently fitted, frozen terminal readouts. Numerical experiments assess SE predictions for non-diagonal cores and covariance-dependent updates with unequal channel dimensions, and illustrate estimation gains from SE-calibrated terminal history readouts.

1 INTRODUCTION

Many high-dimensional inference problems seek low-dimensional latent structure from noisy observations, and low-rank tensor models capture such structure in multiway data. In many applications, tensor modes index different types of entities or measurements and therefore differ in ambient dimension and statistical role (Kolda & Bader, 2009). This heterogeneity naturally leads to asymmetric, rectangular models: modes have distinct factor matrices and may use different update maps, while their ambient dimensions may differ. This also motivates mode-specific latent dimensions. A general core further allows interactions among different latent channels across modes, instead of restricting every interaction to a shared channel index.

Approximate message passing (AMP) is a class of iterative algorithms for high-dimensional inference, combining denoising updates with Onsager corrections that compensate for correlations created by iterative reuse of the same random operator (Donoho et al., 2009; Feng et al., 2022). In many Gaussian-matrix settings, the high-dimensional AMP trajectory admits an asymptotically exact finite-dimensional characterization through state evolution (SE), which describes its asymptotic empirical law and predicts overlaps, estimation errors, and other fixed-dimensional observables across iterations (Bayati & Montanari, 2011; Javanmard & Montanari, 2013; Feng et al., 2022). In the tensor models considered here, the AMP field is multilinear: the same Gaussian disorder tensor enters multiple mode contractions and is reused across iterations. These dependencies make both the correct Onsager reaction matrix and a rigorous state-evolution argument more delicate.

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.

Rigorous SE results for AMP in tensor models cover symmetric scalar-coordinate recursions (El Alaoui et al., 2021; Huang & Sellke, 2024), whereas the closest asymmetric analysis treats a Bayes-matched diagonal-core CANDECOMP/PARAFAC (CP) model without a rigorous theorem for the empirical trajectory (Kadmon & Ganguli, 2018). Structured matrix-AMP theorems likewise do not directly cover the dependent mode contractions generated by repeated use of one Gaussian tensor (Berthier et al., 2020; Gerbelot & Berthier, 2023). See Section 2 and Table 1 for details.

Key idea. A single core-induced map determines the signal contribution and Gaussian covariance, while its derivatives generate the Onsager reaction. For an order-N tensor, $n, m \in \{1, \dots, N\}$ index modes, s, t are AMP iteration indices, and T denotes the fixed time horizon. A fixed multilinear core $\boldsymbol{\mathcal{C}}$ induces, for each output mode n, a matrix-valued covariance map $\mathcal{K}_n$. If $\mathbf{Q}_m^{s,t}$ is the matrix-valued cross-time overlap in mode m, this map contracts the tuple $(\mathbf{Q}_m^{s,t})_{m \neq n}$ into the covariance $\mathcal{K}_n(\mathbf{Q}^{s,t})$ of the mode-n Gaussian field in state evolution. For $1 \leq t \leq T$, the Onsager reaction matrix is

$$\mathbf{\Omega}_n^t = \sum_{m \neq n} D_m \mathcal{K}_n(\mathbf{Q}^{t-1,t})[\overline{\mathbf{J}}_m^t], \tag{1}$$

where D_m differentiates with respect to the mode-m overlap argument, and $\overline{\mathbf{J}}_m^t$ is the average denoiser Jacobian. Its finite-system counterpart is the Onsager factorization in (12).

Contributions.

- **Finite-system calculus for a general core.** We derive core-induced matrix-valued covariance maps and an exact finite-system factorization of the Onsager reaction through their derivatives. The resulting factorization computes the reaction from modewise cross-time iterate overlaps and average denoiser Jacobians.
- **Rigorous joint row-law state evolution.** We prove fixed-horizon SE in the original mode-specific vector coordinates for asymmetric multilinear AMP with mode-dependent ambient and channel dimensions, a general core, empirical covariance inputs and Onsager reaction matrices, and initialization correlated with the planted rows but independent of the Gaussian disorder. The theorem jointly characterizes, within each mode, the planted row together with the factor-iterate and effective-field histories, and allows non-Bayesian row maps.
- **Consequences and numerical validation.** The joint row law yields fixed-dimensional representation statistics and simultaneous loss convergence over fixed finite candidate families, including consistent selection when the SE population minimizer is unique. Experiments test SE predictions and illustrate estimation gains from terminal history readouts.

Proof overview. We handle adaptive reuse of the Gaussian tensor by embedding mode blocks as species (coordinate groups) and serializing vector updates into scalar queries. The key step is quantitative control of the stacked-response covariance in a direct Gaussian-conditioning argument. Fixed-horizon coupling then transfers the resulting population-coefficient recursion to empirical AMP. Section 4.3 gives the proof architecture.

The theorem covers fixed-horizon, fully observed Gaussian multilinear models with fixed channel dimensions and a known core shared by the model and AMP recursion. Section 7 discusses the remaining limitations.

2 Related Work

Structured matrix AMP. Rigorous SE results cover fixed-dimensional vector-valued Gaussian-matrix recursions and extensions with non-separable maps, graph-indexed recursions, rotationally invariant ensembles, spectral initialization, multilayer structure, and signal-correlated side information (Bayati & Montanari, 2011; Javanmard & Montanari, 2013; Berthier et al., 2020; Gerbelot & Berthier, 2023; Fan, 2022; Zhong et al., 2024; Daniels et al., 2022; Liu et al., 2023). The graph-indexed formulation of Gerbelot & Berthier (2023) uses independent Gaussian edge matrices, apart from transpose pairing on reverse edges. Here the mode unfoldings are dependent rearrangements of the same Gaussian tensor, so they do not have this edge-wise independence structure. Rossetti & Reeves (2023) establish SE for non-separable AMP in a matrix tensor product model combining multiple types of pairwise observations.

Tensor AMP and state evolution. Kadmon & Ganguli (2018) derive Bayesian AMP and dynamic mean-field equations for asymmetric diagonal-core CP tensors with unequal ambient dimensions. Their analysis does not establish a pseudo-Lipschitz limit theorem for the empirical trajectory. Rigorous SE for symmetric mixed-spin tensor recursions is established by El Alaoui et al. (2021), with a multi-species extension in Huang & Sellke (2024). Montanari & Zhou (2026, Proposition D.3) extend the single-species symmetric tensor-AMP result to allow different nonlinearities for different tensor orders, as part of their analysis of multi-index empirical-risk minimization. Table 1 compares the original asymmetric CP and multi-species formulations with the recursion studied here.

Table 1: Comparison of the original formulations and SE guarantees of two closely related tensor-AMP analyses. A check mark indicates that the property is supported directly by the stated recursion or theorem, without an auxiliary embedding.

Directly supported property	Huang & Sellke (2024)	Kadmon & Ganguli (2018)	This work
Rectangular factorization in original mode coordinates	×	✓	✓
Mode-dependent latent-channel dimensions	×	×	✓
General multilinear core	×	×	✓
Rigorous fixed-horizon joint row law	✓	×	✓

An algebraic CP representation does not by itself specify an equivalent AMP recursion. Preserving arbitrary original-coordinate histories losslessly requires an injective state embedding and compatible lifted denoisers. Appendix G.2 makes these conditions precise.

Relation to prior rigorous state evolution. Our analysis builds on the Gaussian-conditioning approach of Huang & Sellke (2024, Appendix A). We extend it to the pure-degree multi-species state evolution required for the rectangular vector recursion, including correlated planted–initial covariates, and then transfer the resulting source limit to empirical covariance inputs and Onsager reaction matrices.

3 Asymmetric Multilinear AMP

3.1 Gaussian multilinear model

Consider a tensor of order $N \geq 3$, with ambient dimension I_n and latent-channel dimension P_n in mode n. The order and latent-channel dimensions remain fixed as the ambient dimensions grow. Define the reference ambient size by $I := \left(\prod_{n=1}^{N} I_n\right)^{1/N}$. We assume proportional growth:

$$\frac{I_n}{I} \longrightarrow c_n \in (0, \infty) \quad \Longrightarrow \quad \prod_{n=1}^{N} c_n = 1. \tag{2}$$

For each mode n, the planted factor is $\mathbf{A}_n \in \mathbb{R}^{I_n \times P_n}$, and $\boldsymbol{\mathcal{C}} \in \mathbb{R}^{P_1 \times \cdots \times P_N}$ is a fixed deterministic core. With $[k] := \{1, \ldots, k\}$, the ambient multi-index $\boldsymbol{i} = (i_1, \ldots, i_N) \in \prod_{n=1}^{N} [I_n]$ selects a tensor entry, while the channel multi-index $\boldsymbol{p} = (p_1, \ldots, p_N) \in \prod_{n=1}^{N} [P_n]$ selects a core entry. For a fixed noise level $\nu > 0$, the latent tensor and its noisy observation can be expressed as

$$[\boldsymbol{\mathcal{Z}}]_{\boldsymbol{i}} = \sum_{\boldsymbol{p}} \boldsymbol{\mathcal{C}}_{\boldsymbol{p}} \prod_{n=1}^{N} [\mathbf{A}_n]_{i_n p_n}, \tag{3}$$

$$[\boldsymbol{\mathcal{Y}}]_{\boldsymbol{i}} = [\boldsymbol{\mathcal{Z}}]_{\boldsymbol{i}} + \sqrt{\nu_I}\, [\boldsymbol{\mathcal{G}}]_{\boldsymbol{i}}, \qquad \nu_I = \nu I^{N-1}, \tag{4}$$

where $\boldsymbol{\mathcal{G}}$ has independent and identically distributed (i.i.d.) standard Gaussian entries. After division by $I^{(N-1)/2}$, (4) has the conventional dense tensor-inference normalization (Montanari & Richard, 2014; Lesieur et al., 2017; Kadmon & Ganguli, 2018). Equation (3) is a fixed-core Tucker signal (Tucker, 1966; Kolda & Bader, 2009).

3.2 FINITE-SYSTEM AMP RECURSION

Given the observation $\boldsymbol{\mathcal{Y}}$, an initialization $\mathbf{X}^0 = (\mathbf{X}_n^0)_{n=1}^N$, and the same core $\boldsymbol{\mathcal{C}}$ used in the data-generating model (3), AMP alternates an effective-field update with rowwise denoising. In mode n, the algorithm maintains a factor iterate $\mathbf{X}_n^t \in \mathbb{R}^{I_n \times P_n}$, representing its iteration-t estimate of $\mathbf{A}_n$, and an effective field $\mathbf{B}_n^t \in \mathbb{R}^{I_n \times P_n}$. The row denoiser $f_{n,t} : \mathbb{R}^{P_n} \times \mathbb{S}_+^{P_n} \to \mathbb{R}^{P_n}$ maps an effective-field row and a mode-shared covariance parameter to the corresponding row of the next factor iterate $\mathbf{X}_n^{t+1}$, where $\mathbb{S}_+^{P_n}$ is the cone of real symmetric positive-semidefinite $P_n \times P_n$ matrices.

For a mode-indexed tuple, the subscript $\backslash n$ deletes its nth component, so $\boldsymbol{i}_{\backslash n} = (i_m)_{m \neq n}$ and $\boldsymbol{p}_{\backslash n} = (p_m)_{m \neq n}$. Set $\mathbf{X}_n^{-1} = 0$ and $\boldsymbol{\Omega}_{n,I}^0 = 0$. For $t \geq 0$, the mode-n update is

$$[\mathbf{B}_n^t]_{i_n,\cdot} = \frac{1}{\nu_I} \sum_{\boldsymbol{i}_{\backslash n}} [\boldsymbol{\mathcal{Y}}]_{\boldsymbol{i}} (\boldsymbol{\phi}_{n,\boldsymbol{i}_{\backslash n}}^t)^\top - [\mathbf{X}_n^{t-1}]_{i_n,\cdot} \boldsymbol{\Omega}_{n,I}^t, \tag{5}$$

$$[\mathbf{X}_n^{t+1}]_{i_n,\cdot} = f_{n,t}([\mathbf{B}_n^t]_{i_n,\cdot}, \mathbf{K}_{n,I}^{t,t}), \tag{6}$$

where $\boldsymbol{\phi}_{n,\boldsymbol{i}_{\backslash n}}^t$ is the partial core contraction entering the field update, $\mathbf{K}_{n,I}^{t,t}$ is the covariance input to the denoiser, and $\boldsymbol{\Omega}_{n,I}^t$ is the Onsager reaction matrix. The subtracted product $[\mathbf{X}_n^{t-1}]_{i_n,\cdot} \boldsymbol{\Omega}_{n,I}^t$ is the Onsager correction.

For each complementary ambient multi-index $\boldsymbol{i}_{\backslash n}$, the entries of $\boldsymbol{\phi}_{n,\boldsymbol{i}_{\backslash n}}^t \in \mathbb{R}^{P_n}$ are

$$[\boldsymbol{\phi}_{n,\boldsymbol{i}_{\backslash n}}^t]_{p_n} = \sum_{\boldsymbol{p}_{\backslash n}} \boldsymbol{\mathcal{C}}_{p_n, \boldsymbol{p}_{\backslash n}} \prod_{m \neq n} [\mathbf{X}_m^t]_{i_m p_m}. \tag{7}$$

The covariance input is computed as the noise-scaled Gram matrix of the contracted features,

$$\mathbf{K}_{n,I}^{t,t} := \frac{1}{\nu_I} \sum_{\boldsymbol{i}_{\backslash n}} \boldsymbol{\phi}_{n,\boldsymbol{i}_{\backslash n}}^t (\boldsymbol{\phi}_{n,\boldsymbol{i}_{\backslash n}}^t)^\top \in \mathbb{S}_+^{P_n}.$$

This Gram matrix admits the exact finite-system overlap representation

$$\mathbf{K}_{n,I}^{t,t} = \mathcal{K}_{n,I}(\mathbf{Q}_I^{t,t}), \tag{8}$$

where $\mathbf{Q}_I^{s,t} = (\mathbf{Q}_{m,I}^{s,t})_{m=1}^N$ collects across modes the iterate overlaps

$$\mathbf{Q}_{n,I}^{s,t} = \frac{1}{I_n} (\mathbf{X}_n^s)^\top \mathbf{X}_n^t, \qquad 0 \leq s, t. \tag{9}$$

The finite-system core-induced covariance map $\mathcal{K}_{n,I}$ acts on any matrix tuple $\mathbf{E} = (\mathbf{E}_m)_{m=1}^N$ with $\mathbf{E}_m \in \mathbb{R}^{P_m \times P_m}$. To specify its entries, use $\boldsymbol{r} = (r_1, \ldots, r_N) \in \prod_{m=1}^N [P_m]$ as a second channel multi-index alongside $\boldsymbol{p}$. p_n and r_n index the row and column of the resulting mode-n matrix:

$$[\mathcal{K}_{n,I}(\mathbf{E})]_{p_n r_n} = \alpha_{n,I} \sum_{\boldsymbol{p}_{\backslash n}, \boldsymbol{r}_{\backslash n}} \boldsymbol{\mathcal{C}}_{p_n, \boldsymbol{p}_{\backslash n}} \boldsymbol{\mathcal{C}}_{r_n, \boldsymbol{r}_{\backslash n}} \prod_{m \neq n} \mathbf{E}_m[p_m, r_m], \tag{10}$$

where $I_{\backslash n} = \prod_{m \neq n} I_m$ and $\alpha_{n,I} = I_{\backslash n} / \nu_I$.

For $t \geq 1$, the preceding denoising step supplies the average denoiser Jacobian. Under the convention $\delta f = \delta \boldsymbol{b}\, \mathbf{J}$, it is

$$\overline{\mathbf{J}}_{n,I}^t = \frac{1}{I_n} \sum_{i_n=1}^{I_n} D_{\boldsymbol{b}} f_{n,t-1}\big([\mathbf{B}_n^{t-1}]_{i_n,\cdot}, \mathbf{K}_{n,I}^{t-1,t-1}\big), \tag{11}$$

where $D_{\boldsymbol{b}} f_{n,t-1}$ denotes the denoiser Jacobian with respect to its field argument, with the covariance input held fixed.

Since $\mathcal{K}_{n,I}$ is linear in each complementary overlap, $D_m \mathcal{K}_{n,I}(\mathbf{Q})[\mathbf{H}]$ is obtained by replacing the mth input $\mathbf{Q}_m$ with $\mathbf{H} \in \mathbb{R}^{P_m \times P_m}$ in (10), for $m \neq n$. The Onsager reaction matrix is

$$\boldsymbol{\Omega}_{n,I}^t = \sum_{m \neq n} D_m \mathcal{K}_{n,I}(\mathbf{Q}_I^{t-1,t})[\overline{\mathbf{J}}_{m,I}^t], \tag{12}$$

where $D_m \mathcal{K}_{n,I}$ denotes the Fréchet derivative with respect to the mode-m overlap argument.

With the stated initialization, equations (5)–(12) recursively determine the finite-system trajectory.

4 State Evolution and the Onsager Principle

The same core-induced map governs the finite-system signal contribution and Gaussian covariance, while its derivatives generate the chain-rule Onsager reaction. Appendix G.7 relates this derivative–response form to bilinear and linear AMP.

4.1 Finite-system calculus and the population map

Section 3.2 already expresses the covariance input and reaction matrix through iterate overlaps and average denoiser Jacobians. Characterizing the signal contribution requires one additional quantity unavailable to the algorithm, the signal overlap and its mode tuple:

$$\mathbf{M}^t_{n,I} = \frac{1}{I_n}\mathbf{A}_n^\top \mathbf{X}^t_n, \qquad \mathbf{M}^t_I = (\mathbf{M}^t_{m,I})_{m=1}^N. \tag{13}$$

Thus $\mathbf{M}^t_{n,I}$ measures alignment with the planted factor, whereas $\mathbf{Q}^{s,t}_{n,I}$ in (9) is computed from the AMP trajectory alone.

Proposition 4.1 (Finite-system core calculus)**.** *For any deterministic collection of mode iterates, the signal contribution to (5) is $[\mathbf{A}_n]_{i_n,\cdot}\mathcal{K}_{n,I}(\mathbf{M}^t_I)$. For any s,t and fixed output row, the time-s and time-t Gaussian contractions generated by these arrays have cross-time covariance $\mathcal{K}_{n,I}(\mathbf{Q}^{s,t}_I)$. The reaction in (12) is exactly the chain-rule Onsager reaction matrix induced by this covariance kernel.*

Appendix C proves these frozen-input identities. Theorem 4.3 handles the dependencies created by adaptive AMP iterates.

The proportional-growth condition (2) gives $\alpha_{n,I} \to \alpha_n = (\nu c_n)^{-1}$. Replacing $\alpha_{n,I}$ by this limit in (10) defines the population core-induced covariance map $\mathcal{K}_n$ and its derivatives. The finite-system identities in Proposition 4.1 then lead to the representative-row recursion below, in which finite-system overlaps are replaced by population moments.

4.2 State evolution

We impose the following regularity conditions.

Assumption 4.2. *The core, channel dimensions, denoisers, and horizon are deterministic and fixed, and updates are synchronous. Planted rows are i.i.d. within each mode, independent across modes and of $\mathcal{G}$, and have a finite moment of some order greater than four. The initialization may depend on all planted arrays, but the full planted–initial collection is independent of $\mathcal{G}$. The modewise empirical row laws of the planted–initial pairs converge in Wasserstein-2, in probability, to deterministic laws represented by $(\boldsymbol{a}_n, \boldsymbol{x}^0_n)$. The row maps are continuously differentiable in their field argument. Their field and covariance dependence satisfies the Lipschitz, Jacobian, and linear-growth bounds in Appendix B.*

For any time-indexed object, a superscript $s:t$ denotes the ordered history from time s through t, with row extraction applied componentwise. Starting from the limiting pairs $(\boldsymbol{a}_n, \boldsymbol{x}^0_n)$, construct the synchronous recursion for $t = 0, \dots, T$. Appendix C shows that the block matrix $(\mathbf{K}^{s,\ell}_n)_{0\le s,\ell\le t}$ is positive semidefinite, so the centered jointly Gaussian residual history $\boldsymbol{u}^{0:t}_n$ below is well defined. For $0 \le s, \ell \le t$, one SE step is

$$\mathbf{M}^t_n = \mathbb{E}[\boldsymbol{a}_n^\top \boldsymbol{x}^t_n], \qquad \mathbf{Q}^{s,\ell}_n = \mathbb{E}[(\boldsymbol{x}^s_n)^\top \boldsymbol{x}^\ell_n], \tag{14}$$

$$\mathbf{\Theta}^t_n = \mathcal{K}_n(\mathbf{M}^t_1, \dots, \mathbf{M}^t_N), \qquad \mathbf{K}^{s,\ell}_n = \mathcal{K}_n(\mathbf{Q}^{s,\ell}_1, \dots, \mathbf{Q}^{s,\ell}_N), \tag{15}$$

$$\boldsymbol{u}^{0:t}_n \sim \mathcal{N}\Big(0, (\mathbf{K}^{s,\ell}_n)_{0\le s,\ell\le t}\Big), \qquad \boldsymbol{u}^{0:t}_n \text{ is independent of } (\boldsymbol{a}_n, \boldsymbol{x}^0_n), \tag{16}$$

$$\boldsymbol{b}^t_n = \boldsymbol{a}_n \mathbf{\Theta}^t_n + \boldsymbol{u}^t_n, \qquad \boldsymbol{x}^{t+1}_n = f_{n,t}(\boldsymbol{b}^t_n, \mathbf{K}^{t,t}_n), \tag{17}$$

where $\mathbf{M}^t = (\mathbf{M}^t_m)_m$ and $\mathbf{Q}^{s,\ell} = (\mathbf{Q}^{s,\ell}_m)_m$ are the mode tuples. For each mode n, the matrices $\mathbf{M}^t_n$, $\mathbf{Q}^{s,\ell}_n$, and $\mathbf{K}^{s,\ell}_n$ have size $P_n \times P_n$, with $\mathbf{Q}^{\ell,s}_n = (\mathbf{Q}^{s,\ell}_n)^\top$ and $\mathbf{K}^{\ell,s}_n = (\mathbf{K}^{s,\ell}_n)^\top$.

The same SE recursion also determines the average denoiser Jacobians and Onsager reaction matrices in the limit. For $1 \le t \le T$,

$$\overline{\mathbf{J}}_n^t = \mathbb{E}[D_{\boldsymbol{b}} f_{n,t-1}(\boldsymbol{b}_n^{t-1}, \mathbf{K}_n^{t-1,t-1})], \qquad \boldsymbol{\Omega}_n^t = \sum_{m \neq n} D_m \mathcal{K}_n(\mathbf{Q}^{t-1,t})[\overline{\mathbf{J}}_m^t]. \tag{18}$$

Pseudo-Lipschitz tests of order two are defined in Appendix B, with time histories concatenated as their arguments.

Theorem 4.3 (Fixed-horizon state evolution)**.** *Under Assumption 4.2 and the aspect-ratio regime (2), consider the model (3)–(4) and the AMP recursion (5)–(12), using the same core $\boldsymbol{\mathcal{C}}$ as the data-generating model. For every fixed T, mode n, and pseudo-Lipschitz order-2 function ψ,*

$$\frac{1}{I_n} \sum_{i_n=1}^{I_n} \psi\Big([\mathbf{A}_n]_{i_n,\cdot}, [\mathbf{X}_n^{0:T+1}]_{i_n,\cdot}, [\mathbf{B}_n^{0:T}]_{i_n,\cdot} \Big) \xrightarrow{\mathbb{P}} \mathbb{E}\psi(\boldsymbol{a}_n, \boldsymbol{x}_n^{0:T+1}, \boldsymbol{b}_n^{0:T}). \tag{19}$$

The limiting law is the recursion (14)–(17). Moreover, the empirical signal and iterate overlaps and the covariance inputs used by the denoisers converge jointly over the fixed modes and times:

$$\max_{\substack{n \in [N] \\ 0 \le t \le T}} \left\| \mathbf{M}_{n,I}^t - \mathbf{M}_n^t \right\| + \max_{\substack{n \in [N] \\ 0 \le s,t \le T}} \left\| \mathbf{Q}_{n,I}^{s,t} - \mathbf{Q}_n^{s,t} \right\| + \max_{\substack{n \in [N] \\ 0 \le t \le T}} \left\| \mathbf{K}_{n,I}^{t,t} - \mathbf{K}_n^{t,t} \right\| \xrightarrow{\mathbb{P}} 0. \tag{20}$$

The empirical average denoiser Jacobians and Onsager reaction matrices converge to their SE counterparts in (18):

$$\max_{\substack{n \in [N] \\ 1 \le t \le T}} \left\{ \left\| \overline{\mathbf{J}}_{n,I}^t - \overline{\mathbf{J}}_n^t \right\| + \left\| \boldsymbol{\Omega}_{n,I}^t - \boldsymbol{\Omega}_n^t \right\| \right\} \xrightarrow{\mathbb{P}} 0. \tag{21}$$

In particular, squared-error tests give the SE limits of factor mean-squared error (MSE), and the exact factorization in Appendix C then yields the corresponding tensor-reconstruction MSE limit. The result applies to both non-Bayesian denoisers and Bayesian denoisers with misspecified priors or noise parameters, provided they satisfy Assumption 4.2.

Connection to diagonal-core tensor AMP. For the equality-indicator diagonal CP core, when all channel dimensions P_n are equal, the finite-system and population maps reduce to modewise Hadamard products:

$$\mathcal{K}_{n,I}(\mathbf{E}) = \alpha_{n,I} \bigodot_{m \neq n} \mathbf{E}_m, \qquad \mathcal{K}_n(\mathbf{E}) = \alpha_n \bigodot_{m \neq n} \mathbf{E}_m, \tag{22}$$

where $\bigodot$ denotes the entrywise (Hadamard) product over the indicated modes. Appendix G.1 derives the corresponding Hadamard-product Onsager reaction. The Bayes-matched corollary in Appendix G.3 gives the precise conditions under which $\mathbf{M}_n^t = \mathbf{Q}_n^{t,t}$ propagates. Under the additional factorization and diagonality conditions stated there, its single-time equations have the same Gaussian-channel form as the dynamic mean-field equations of Kadmon & Ganguli (2018), and the aspect-ratio normalization is specified by (22). Theorem 4.3 proves convergence for this specialization.

Beyond signal-overlap closure. Outside the Bayes-matched (Nishimori) manifold, signal overlaps alone need not close the recursion: Appendix G.4 gives admissible initial laws with identical signal overlaps but different field and iterate laws. Mixed-time overlaps also cannot generally be replaced by current-time ones. In the scalar three-mode example of Appendix G.6, $\alpha_n = 1$, $Q_n^{0,1} = \overline{J}_n^1 = 1/\sqrt{2}$, and $Q_n^{1,1} = 1$, so the exact reaction is $\Omega_n^1 = 1$ but the current–current substitute is $\sqrt{2}$.

4.3 Proof Architecture

Appendix D embeds mode blocks as symmetric-tensor species and serializes vector channels into scalar queries. Conditioning on adaptive queries from the same tensor is the central difficulty. For a fixed query stack F, $\mathcal{A}_F$ maps Gaussian coefficients to stacked gradient responses. For $N \ge 3$, normalized query-Gram bounds and auxiliary source species-interaction amplitudes bounded above and away from zero imply

$$c\,\mathrm{Id} \preceq \mathcal{A}_F \mathcal{A}_F^* \preceq C\,\mathrm{Id}, \qquad 0 < c \le C < \infty,$$

where $\mathcal{A}_F^*$ is the Euclidean/Frobenius adjoint and Id is the stacked-response identity. For fixed order and bounds, c, C are dimension-independent (Lemma D.3). This controls the conditioning inverse. Integration by parts identifies the Onsager term from the finite-rank covariance correction, which gives the source row law (Appendix D.4).

Same-degree Gaussian fill and independent query dither remove sparse-interaction and singular-history degeneracies by fixed-horizon continuity, while conditioning on planted–initial arrays handles their empirical Wasserstein–2 limits. Theorem D.8 and Appendix D.6 give the source limit and its identification with the original-coordinate population-coefficient recursion. Appendix E restores the finite-system Gaussian normalization and couples the resulting population-coefficient recursion to the empirical-coefficient AMP, proving Theorem 4.3.

5 Consequences of the Joint Row Law

The results below follow from the joint row law in Theorem 4.3 through continuous moment functionals, conditioning, and finite-family arguments.

5.1 Representation statistics

The joint row law determines statistics used in canonical correlation analysis (CCA) and affine probing (Alain & Bengio, 2017; Raghu et al., 2017; Kornblith et al., 2019). Empirical means, covariances, and cross-covariances between planted rows and AMP iterates converge. With covariance nondegeneracy, this gives consistent estimates of canonical correlations and population affine-probe risks for latent-linear labels. Appendix F gives the full statement and formulas, and Appendix H.4 evaluates the predictions.

5.2 Finite-family selection and terminal readouts

For a fixed finite family of admissible AMP recursions, the joint row law gives simultaneous convergence of row-history losses that are pseudo-Lipschitz of order two.

Corollary 5.1 (Finite-family loss convergence and selection). *Fix $N_{\mathrm{cand}} < \infty$ candidate AMP recursions, each with deterministic row maps satisfying Assumption 4.2, and evaluate them on the same admissible model realization and initialization. Then*

$$\max_{1\le j\le N_{\mathrm{cand}}} \left|\widehat{L}_{I,j} - L_j\right| \xrightarrow{\mathbb{P}} 0, \tag{23}$$

where $\widehat{L}_{I,j}$ denotes the empirical pseudo-Lipschitz row-history loss and L_j its SE population counterpart. If $j^\star = \arg\min_j L_j$ is unique, any measurably tie-broken empirical minimizer $\widehat{j}_I$ satisfies $\mathbb{P}(\widehat{j}_I = j^\star) \to 1$.

Appendix F.3 defines these losses and proves the corollary. Conditioning extends the conclusions to admissible row maps and finite panels selected independently of the model and initialization randomness used for evaluation and held fixed across ambient sizes. Appendix F gives the precise conditions for these conditional statements and the additional score-consistency requirement for selection by Monte Carlo SE scores.

We can also hold the AMP recursion fixed and compare terminal readouts of its initialization and effective-field history. With column-concatenated histories and fixed $\mathbf{W}_n^{\mathrm{out}} \in \mathbb{R}^{(T+2)P_n \times P_n}$, define

$$\mathbf{H}_{n,I} = [\mathbf{X}_n^0, \mathbf{B}_n^{0:T}], \qquad \widehat{\mathbf{A}}_n = \mathbf{H}_{n,I}\mathbf{W}_n^{\mathrm{out}}. \tag{24}$$

For each fixed readout, squared estimation error is pseudo-Lipschitz of order two. Corollary F.5 extends the loss and selection conclusions to independently fitted readouts frozen across sizes.

For the zero-mean setting of the experiments, the SE row $\boldsymbol{h}_n = (\boldsymbol{x}_n^0, \boldsymbol{b}_n^{0:T})$ determines the population ridge fit

$$\mathbf{W}_{n,\lambda}^{\mathrm{out}} = \left(\mathbb{E}[\boldsymbol{h}_n^\top \boldsymbol{h}_n] + \lambda \mathbf{I}_{(T+2)P_n}\right)^{-1} \mathbb{E}[\boldsymbol{h}_n^\top \boldsymbol{a}_n], \qquad \lambda > 0, \tag{25}$$

where $\mathbf{I}_{(T+2)P_n}$ is the identity. Positive ridge permits singular history Grams. In the experiments, independent SE samples estimate the moments, separate samples score candidates, and the fitted matrices are frozen before finite-system evaluation.

6 NUMERICAL VALIDATION

We compare fixed-horizon SE with finite-system trajectories and use Monte Carlo SE (MC-SE) to fit and select terminal readouts.

All in-model tests use (4), synchronous AMP, and planted-correlated initialization independent of $\mathcal{G}$. Appendix H gives full settings and supplementary results.

6.1 STATE-EVOLUTION VALIDATION

The first study uses $N = 3$, $P_1 = P_2 = P_3 = 2$, and $T = 5$. Our non-diagonal family sets $\mathcal{C}_{111} = \mathcal{C}_{222} = 1$ and $\mathcal{C}_{122} = \mathcal{C}_{212} = \mathcal{C}_{221} = \gamma \in [0, 1]$, with all other entries zero. Since $\|\mathcal{C}\|_F^2 = 2 + 3\gamma^2$, varying γ changes both cross-channel interactions and core magnitude. At $\gamma = 0.7$, $\nu = 0.6$, and dimensions $(120, 150, 180)$, the left panel of Figure 1 compares the mode-1 off-diagonal entries of the empirical covariance input and Onsager reaction matrix with SE values over 30 seeds.

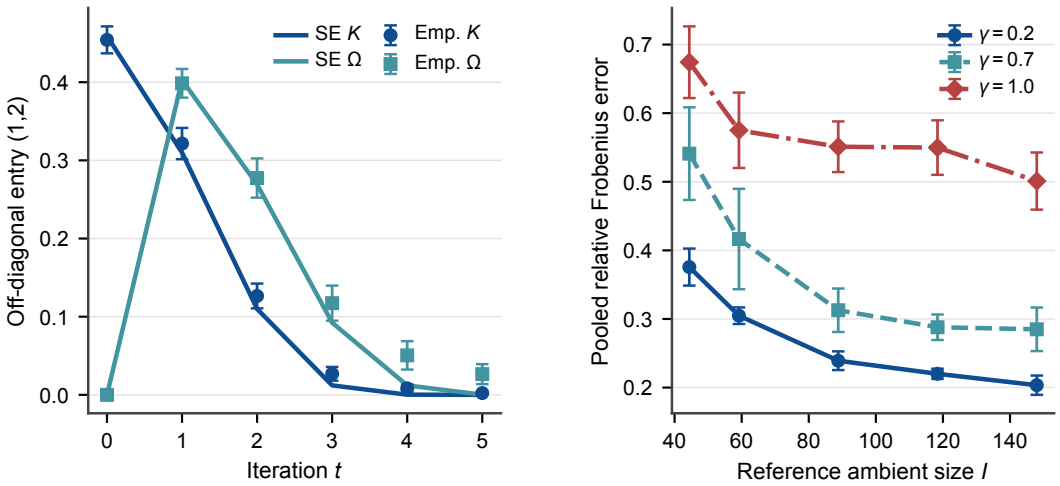

Figure 1: General-core SE validation ($N = 3$, $P_n = 2$, $T = 5$). Left: mode-1 off-diagonal empirical and SE coefficients at $\gamma = 0.7$. Right: per-realization pooled relative Frobenius errors of history overlaps over all modes and time pairs, averaged across realizations, versus size. MC-SE predictions use 120,000 representative rows per mode. Bars show one standard error over 30 realizations.

The pooled relative Frobenius error decreases overall to approximately 0.203, 0.285, and 0.501 for $\gamma = 0.2, 0.7, 1$, respectively, with larger discrepancies for the strongest core. Appendix H defines the error and provides extended-size and order-four results.

Unequal channels and covariance-dependent updates. We also test $(P_1, P_2, P_3) = (2, 3, 4)$ with a fixed $2 \times 3 \times 4$ core and the row map $f_n(\boldsymbol{b}, \mathbf{S}) = \boldsymbol{b}(\mathbf{I}_{P_n} + \mathbf{S})^{-1}$, at $T = 3$ and $\nu = 0.6$. Table 2 compares full-history statistics with a deterministic Gaussian-moment SE reference. The mean errors decrease across sizes, although covariance discrepancies remain appreciable. Appendix H.2 specifies the core and initialization.

Table 2: Fixed-horizon calibration with unequal channel dimensions and covariance-dependent Gaussian row maps. Entries are mean pooled relative Frobenius errors against deterministic SE, with one standard error across 30 realizations. Appendix H.2 gives the pooling ranges.

Ambient dimensions	$\mathbf{M}$	$\mathbf{Q}$	$\mathbf{K}$	$\bar{\mathbf{J}}$	$\mathbf{\Omega}$
$(36, 45, 54)$	0.483 ± 0.022	0.478 ± 0.034	0.590 ± 0.076	0.103 ± 0.010	0.360 ± 0.038
$(72, 90, 108)$	0.397 ± 0.022	0.434 ± 0.034	0.585 ± 0.078	0.089 ± 0.008	0.275 ± 0.029
$(120, 150, 180)$	0.287 ± 0.018	0.304 ± 0.029	0.390 ± 0.061	0.063 ± 0.007	0.190 ± 0.023

Factorized reaction computation. Appendix H.3 compares the factorized and row-explicit reactions. Their maximum relative discrepancy is 1.65×10^{-15}. At $(120, 150, 180)$, the mean reaction-only speedup is $14.10\times$. The timings exclude field contraction and denoising, isolating the computational benefit of modewise overlap and Jacobian summaries.

6.2 SE-CALIBRATED TERMINAL HISTORY READOUTS

We fit the terminal readouts defined in Section 5.2 while keeping the AMP recursion unchanged. The feature sets are X^0, $[X^0, B^0]$, and $[X^0, B^{0:3}]$. At $\gamma = 0.7$ and $T = 3$, the fixed, untrained $\tanh$ row map in Appendix H.5 and the initialization are shared across $\nu \in \{0.05, 0.1, 0.2, 0.5\}$. Four ridge strengths per feature set give twelve candidates.

Separate SE samples from the specified generative model are used for fitting, selection by factor MSE, and held-out loss prediction. The fitted readout matrices are held fixed across three ambient sizes, with 30 paired finite-system realizations per setting. The principal baseline is the analytic Gaussian estimator based on X^0 alone (calibrated initialization).

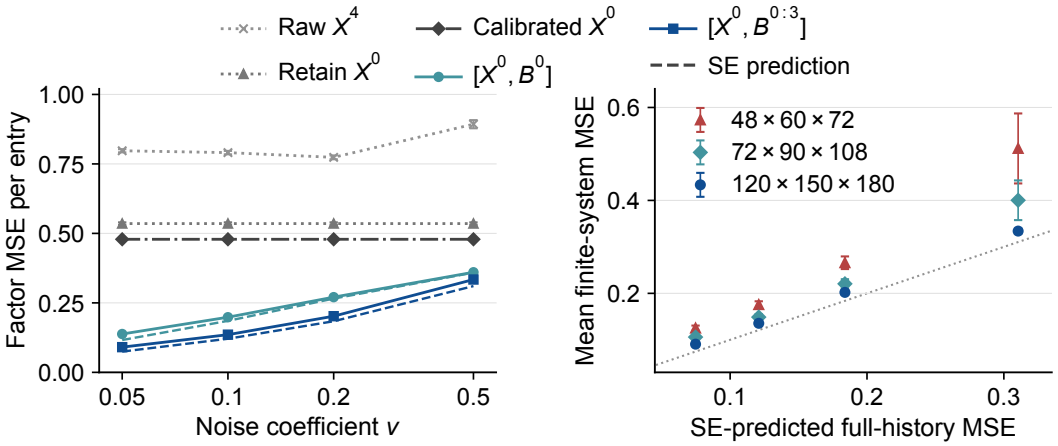

Figure 2: SE-calibrated terminal readouts ($T = 3$). Independent MC-SE scores select each family's ridge and full history overall at all four noise levels. Left: factor MSE at $(120, 150, 180)$, with dashed held-out SE predictions. Raw X^4 denotes the final uncalibrated iterate. Right: empirical full-history MSE versus SE predictions at three sizes. The dotted line is the identity. Bars show one standard error over 30 realizations.

Figure 2 shows that, at $(120, 150, 180)$, full history improves on calibrated initialization at all four noise levels. At $\nu = 0.05$, factor MSE falls from 0.4788 to 0.0907, below the first-field readout's 0.1380. Separately evaluated tensor normalized mean-square error (NMSE) falls from 0.8669 to 0.2310. The SE choice also minimizes empirical mean loss at both larger sizes for $\nu = 0.05, 0.1, 0.2$. At the smallest size and $\nu = 0.5$, however, SE predicts 0.310 for full history, versus observed MSEs 0.512 (full history) and 0.367 (first field). Appendix H.5 gives the full results. Gains arise from terminal readout, while raw final iterates remain worse than initialization.

7 CONCLUSION AND LIMITATIONS

We establish fixed-horizon joint row-law SE for general-core asymmetric multilinear AMP with empirical coefficients. The resulting factorization identifies the mixed-time Onsager reaction required by this limit, and the recursion closes without a Bayes-matched overlap identity. The joint history law determines moment-based statistics and frozen-readout losses, supporting SE-calibrated estimation from a fixed AMP trajectory.

The fixed-horizon formulation leaves long-time stability and monotone reconstruction improvement open, as well as quantitative finite-sample prediction and selection bounds. Extensions to missing observations, learned cores, and growing channel dimensions are also outside the present scope.

REPRODUCIBILITY STATEMENT

The assumptions and complete proofs of the new results are provided in the appendices. Appendix H records the experimental settings, seeds, and diagnostics. Code for anonymous review is available at https://anonymous.4open.science/r/Tensor_AMP_SE-5E05.

AI USE STATEMENT

Generative AI tools assisted in proof auditing, experimental design and interpretation, code implementation and debugging, and manuscript polishing. Numerical results were generated using simulation scripts. The authors are solely responsible for all mathematical statements, interpretations, source code, experimental outcomes, and the final manuscript.

REFERENCES

Guillaume Alain and Yoshua Bengio. Understanding intermediate layers using linear classifier probes. In *International Conference on Learning Representations Workshop Track*, 2017. URL https://arxiv.org/abs/1610.01644.

Brett W. Bader and Tamara G. Kolda. Algorithm 862: MATLAB tensor classes for fast algorithm prototyping. *ACM Transactions on Mathematical Software*, 32(4):635–653, 2006. doi: 10.1145/1186785.1186794.

Antoine Baker, Florent Krzakala, Benjamin Aubin, and Lenka Zdeborová. Tree-AMP: Compositional inference with tree approximate message passing. *Journal of Machine Learning Research*, 24(57):1–89, 2023. URL https://www.jmlr.org/papers/v24/20-695.html.

Jean Barbier, Nicolas Macris, and Léo Miolane. The layered structure of tensor estimation and its mutual information. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing*, pp. 1056–1063, 2017. doi: 10.1109/ALLERTON.2017.8262854.

Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011. doi: 10.1109/TIT.2010.2094817.

Raphaël Berthier, Andrea Montanari, and Phan-Minh Nguyen. State evolution for approximate message passing with non-separable functions. *Information and Inference: A Journal of the IMA*, 9(1):33–79, 2020. doi: 10.1093/imaiai/iay021.

Mark Borgerding, Philip Schniter, and Sundeep Rangan. AMP-inspired deep networks for sparse linear inverse problems. *IEEE Transactions on Signal Processing*, 65(16):4293–4308, 2017. doi: 10.1109/TSP.2017.2708040.

Rasmus Bro. PARAFAC: Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38(2):149–171, 1997. doi: 10.1016/S0169-7439(97)00032-4.

Xinshi Chen, Haoran Sun, and Le Song. Provable learning-based algorithm for sparse recovery. In *International Conference on Learning Representations*. OpenReview.net, 2022. URL https://openreview.net/forum?id=BwPaPxwgyQb.

Bingyang Cheng, Zhongtao Chen, Yichen Jin, Hao Zhang, Chen Zhang, Edmund Y. Lam, and Yik-Chung Wu. Large-scale Bayesian tensor reconstruction via approximate message passing. arXiv preprint arXiv:2505.16305, 2025. URL https://arxiv.org/abs/2505.16305v3.

Mara Daniels, Cédric Gerbelot, Florent Krzakala, and Lenka Zdeborová. Multi-layer state evolution under random convolutional design. In *Advances in Neural Information Processing Systems*, volume 35, pp. 7089–7102, 2022. doi: 10.52202/068431-0514. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/2eb74636e69ca26ab6cba8b724b0776e-Abstract-Conference.html.

Leonardo Defilippis, Yizhou Xu, Julius Girardin, Emanuele Troiani, Vittorio Erba, Lenka Zdeborová, Bruno Loureiro, and Florent Krzakala. Scaling laws and spectra of shallow neural networks in the feature learning regime. In *International Conference on Learning Representations*, volume 2026, pp. 108893–108934, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/b117cd5a0af6f5ecfdac7b47b893b96f-Abstract-Conference.html.

David L. Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009. doi: 10.1073/pnas.0909892106.

Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *The Annals of Probability*, 49(6):2922–2960, 2021. doi: 10.1214/21-AOP1519.

Zhou Fan. Approximate message passing algorithms for rotationally invariant matrices. *The Annals of Statistics*, 50(1):197–224, 2022. doi: 10.1214/21-AOS2101.

Oliver Y. Feng, Ramji Venkataramanan, Cynthia G. Rush, and Richard J. Samworth. A unifying tutorial on approximate message passing. *Foundations and Trends in Machine Learning*, 15(4): 335–536, 2022. doi: 10.1561/2200000092.

Tomoharu Furudoi, Takumi Takahashi, and Hideki Ochiai. State-Evolution-based score matching for generalized approximate message passing. arXiv preprint arXiv:2606.29224, 2026. URL https://arxiv.org/abs/2606.29224.

Cédric Gerbelot and Raphaël Berthier. Graph-based approximate message passing iterations. *Information and Inference: A Journal of the IMA*, 12(4):2562–2628, 2023. doi: 10.1093/imaiai/iaad020.

Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th International Conference on Machine Learning*, pp. 399–406, 2010. URL https://icml.cc/Conferences/2010/papers/449.pdf.

Brice Huang and Mark Sellke. Optimization algorithms for multi-species spherical spin glasses. *Journal of Statistical Physics*, 191(2):29, 2024. doi: 10.1007/s10955-024-03242-7.

Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference: A Journal of the IMA*, 2(2):115–144, 2013. doi: 10.1093/imaiai/iat004.

Jonathan Kadmon and Surya Ganguli. Statistical mechanics of low-rank tensor decomposition. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/hash/b3848d61bbbc6207c6668a8a9e2730ed-Abstract.html.

Ulugbek S. Kamilov, Sundeep Rangan, Alyson K. Fletcher, and Michael Unser. Approximate message passing with consistent parameter estimation and applications to sparse learning. *IEEE Transactions on Information Theory*, 60(5):2969–2985, 2014. doi: 10.1109/TIT.2014.2309005.

Aayush Karan, Kulin Shah, Sitan Chen, and Yonina C. Eldar. Unrolled denoising networks provably learn to perform optimal Bayesian inference. In *Advances in Neural Information Processing Systems*, volume 37, pp. 135264–135298. Curran Associates, Inc., 2024. doi: 10.52202/079017-4299. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/f40b522ef32bbc050d37a2a1ab7ffc2b-Abstract-Conference.html.

Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3): 455–500, 2009. doi: 10.1137/07070111X.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3519–3529. PMLR, 2019. URL https://proceedings.mlr.press/v97/kornblith19a.html.

Hugo Lebeau, Florent Chatelain, and Romain Couillet. A random matrix approach to low-multilinear-rank tensor approximation. *Journal of Machine Learning Research*, 26(7):1–64, 2025. URL https://www.jmlr.org/papers/v26/24-0193.html.

Thibault Lesieur, Léo Miolane, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. In *2017 IEEE International Symposium on Information Theory*, pp. 511–515, 2017. doi: 10.1109/ISIT.2017.8006580.

Ping Li and Phan-Minh Nguyen. On random deep weight-tied autoencoders: Exact asymptotic analysis, phase transitions, and implications to training. In *International Conference on Learning Representations*. OpenReview.net, 2019. URL https://openreview.net/forum?id=HJx54i05tX.

Yinchuan Li, Guangchen Lan, and Xiaodong Wang. Tensor generalized approximate message passing. arXiv preprint arXiv:2504.00008, 2025.

Hangjin Liu, Cynthia Rush, and Dror Baron. Rigorous state evolution analysis for approximate message passing with side information. *IEEE Transactions on Information Theory*, 69(6):3989–4013, 2023. doi: 10.1109/TIT.2022.3220046.

Ryosuke Matsushita and Toshiyuki Tanaka. Low-rank matrix reconstruction and clustering via approximate message passing. In *Advances in Neural Information Processing Systems*, volume 26, pp. 917–925, 2013.

Christopher A. Metzler, Ali Mousavi, and Richard G. Baraniuk. Learned D-AMP: Principled neural network based compressive image recovery. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/8597a6cfa74defcbde3047c891d78f90-Abstract.html.

Marco Mondelli and Ramji Venkataramanan. Approximate message passing with spectral initialization for generalized linear models. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 397–405. PMLR, 2021. URL https://proceedings.mlr.press/v130/mondelli21a.html.

Andrea Montanari and Emile Richard. A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems*, volume 27, pp. 2897–2905, 2014.

Andrea Montanari and Ramji Venkataramanan. Estimation of low-rank matrices via approximate message passing. *The Annals of Statistics*, 49(1):321–345, 2021. doi: 10.1214/20-AOS1958.

Andrea Montanari and Kangjie Zhou. Replica symmetry breaking and algorithmic thresholds in empirical risk minimization under multi-index model. arXiv preprint arXiv:2606.28573, 2026. URL https://arxiv.org/abs/2606.28573.

Jason T. Parker, Philip Schniter, and Volkan Cevher. Bilinear generalized approximate message passing—part I: Derivation. *IEEE Transactions on Signal Processing*, 62(22):5839–5853, 2014. doi: 10.1109/TSP.2014.2357776.

Maithra Raghu, Justin Gilmer, Jason Yosinski, and Jascha Sohl-Dickstein. SVCCA: Singular vector canonical correlation analysis for deep learning dynamics and interpretability. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/dc6a7e655d7e5840e66733e9ee67cc69-Abstract.html.

Sundeep Rangan. Generalized approximate message passing for estimation with random linear mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pp. 2168–2172, 2011. doi: 10.1109/ISIT.2011.6033942.

Riccardo Rossetti and Galen Reeves. Approximate message passing for the matrix tensor product model. arXiv preprint arXiv:2306.15580, 2023.

Riccardo Rossetti, Bobak Nazer, and Galen Reeves. Linear operator approximate message passing (OpAMP). *Information and Inference: A Journal of the IMA*, 14(4):iaaf032, 2025. doi: 10.1093/imaiai/iaaf032.

Cynthia Rush and Ramji Venkataramanan. Finite sample analysis of approximate message passing algorithms. *IEEE Transactions on Information Theory*, 64(11):7264–7286, 2018. doi: 10.1109/TIT.2018.2816681.

Nelvin Tan and Ramji Venkataramanan. Mixed regression via approximate message passing. *Journal of Machine Learning Research*, 24(317):1–44, 2023. URL https://www.jmlr.org/papers/v24/23-0473.html.

Ledyard R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966. doi: 10.1007/BF02289464.

Ramji Venkataramanan, Kevin Kögler, and Marco Mondelli. Estimation in rotationally invariant generalized linear models via approximate message passing. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 22120–22144. PMLR, 2022. URL https://proceedings.mlr.press/v162/venkataramanan22a.html.

Tianhao Wang, Xinyi Zhong, and Zhou Fan. Universality of approximate message passing algorithms and tensor networks. *The Annals of Applied Probability*, 34(4):3943–3994, 2024. doi: 10.1214/24-AAP2056.

Anru Zhang and Dong Xia. Tensor SVD: Statistical and computational limits. *IEEE Transactions on Information Theory*, 64(11):7311–7338, 2018. doi: 10.1109/TIT.2018.2841377.

Xinyi Zhong, Tianhao Wang, and Zhou Fan. Approximate message passing for orthogonally invariant ensembles: Multivariate non-linearities and spectral initialization. *Information and Inference: A Journal of the IMA*, 13(3):iaae024, 2024. doi: 10.1093/imaiai/iaae024.

Qiuyun Zou, Haochuan Zhang, and Hongwen Yang. Multi-layer bilinear generalized approximate message passing. *IEEE Transactions on Signal Processing*, 69:4529–4543, 2021. doi: 10.1109/TSP.2021.3100305.

APPENDIX CONTENTS

A Detailed Comparison with Prior Work **17**

B Notation and Quantitative Assumptions **17**

C Finite-System General-Core Algebra **18**

C.1 Signal and covariance factorization 18

C.2 Reaction factorization and orientation 19

C.3 Reconstruction-error consequences 19

C.4 Stability tools used below 20

D Gaussian Embedding, Serialization, and Source State Evolution **21**

D.1 Exact multi-species embedding 21

D.2 Predetermined channel-index queries 22

D.3 Corrected answers and the block identity 23

D.4 Pure-degree source state evolution 24

D.5 Same-degree regularization and degeneracy removal 30

D.6 Identification of the fixed-interaction oracle 32

E Oracle AMP and Empirical-Coefficient Transfer **33**

E.1 Fixed-interaction and finite oracles 33

E.2 Moment localization 33

E.3 Interaction-scale transfer 34

E.4 Coupling recurrence 34

F Representation Consequences and Frozen-Map Selection **35**

F.1 Moment-based statistics 35

F.2 Independently selected frozen row maps 36

F.3 Finite-family loss convergence and selection 37

F.4 Terminal history readouts 38

G Specializations and Separating Examples **39**

G.1 Diagonal CP core 39

G.2 Original coordinates versus a diagonal-core lift 39

G.3 Bayes-matched reduction and Gaussian calculation 41

G.4 Off-Nishimori non-closure 41

G.5 One-step behavior near zero signal overlap 42

G.6 An admissible mixed-time discrepancy 42

G.7 Covariance–response form across interaction orders 42

H Experimental Settings and Supplementary Results **44**

H.1 Trajectory-level settings and diagnostics 45

H.2 Unequal-channel covariance-dependent updates 48

H.3 Exact factorized implementation of the Onsager reaction 48

H.4 Representation calibration 50

H.5 SE-calibrated terminal history readouts 51

H.6 Frozen-denoiser surrogate-ranking settings 53

H.7 One-step state evolution near zero overlap 54

H.8 Order-four tensor validation 55

H.9 Out-of-model AMP–ALS illustration on amino-acid fluorescence data 55

A Detailed Comparison with Prior Work

This appendix supplements Section 2 with broader matrix-AMP, tensor-inference, and fixed-depth computation context.

Matrix AMP foundations and structured extensions. Beyond the Gaussian-matrix results discussed in Section 2, algorithmic relatives include generalized AMP (GAMP), bilinear generalized AMP (BiG-AMP), low-rank matrix AMP, and multi-layer bilinear AMP (Rangan, 2011; Parker et al., 2014; Matsushita & Tanaka, 2013; Montanari & Venkataramanan, 2021; Zou et al., 2021). Further extensions include compositional factor graphs (Baker et al., 2023), iteration-dependent linear operators (Rossetti et al., 2025), invariant matrix ensembles, vector nonlinearities, spectral initialization, and mixed regression (Fan, 2022; Venkataramanan et al., 2022; Zhong et al., 2024; Tan & Venkataramanan, 2023), and universality results based on matrix-AMP moment expansions encoded by tensor networks (Wang et al., 2024). Data-dependent spectral initialization has rigorous guarantees in specific matrix and generalized-linear models (Montanari & Venkataramanan, 2021; Mondelli & Venkataramanan, 2021), and classical Gaussian-matrix AMP admits nonasymptotic concentration bounds (Rush & Venkataramanan, 2018). Our planted-correlated initialization instead remains independent of the Gaussian disorder tensor.

Tensor inference and core structure. Tensor inference includes threshold analyses for spiked models (Montanari & Richard, 2014; Lesieur et al., 2017; Barbier et al., 2017), statistical and computational guarantees for higher-order orthogonal iteration (HOOI) (Zhang & Xia, 2018), and random-matrix characterizations of truncated multilinear singular value decomposition and HOOI convergence (Lebeau et al., 2025). Algorithmic work includes tensor GAMP for tensor-ring inference with CP and Tucker specializations (Li et al., 2025), as well as CP-GAMP with expectation-maximization-based rank and noise-power learning for partially observed tensors (Cheng et al., 2025). General-core tensor models are classical in Tucker decomposition (Tucker, 1966) and are reviewed by Kolda & Bader (2009). This literature supplies the statistical, algorithmic, and modeling background. The remaining gap is a rigorous fixed-horizon row law for the fully observed general-core vector-valued recursion studied here.

Fixed-depth computation and representation context. Algorithm unrolling and learned AMP motivate viewing finite trajectories as representation maps (Gregor & LeCun, 2010; Borgerding et al., 2017; Metzler et al., 2017). Related learning guarantees concern sparse-recovery unrolling (Chen et al., 2022) and AMP-based denoising networks (Karan et al., 2024). Asymptotic analyses cover random weight-tied autoencoders (Li & Nguyen, 2019), multilayer AMP with convolutional designs (Daniels et al., 2022), and shallow feature learning (Defilippis et al., 2026). SE-generated samples can train Bayesian denoiser surrogates (Furudoi et al., 2026), while adaptive GAMP estimates prior and channel parameters within the recursion (Kamilov et al., 2014). Our results instead concern losses and selection for independently frozen maps and readouts, not learning within the evaluated AMP trajectory.

B Notation and Quantitative Assumptions

This appendix records the normalized row norm and the quantitative row-map bounds used by the finite-system stability estimates, source reduction, and oracle-to-empirical coupling in Appendices C–E.

Unqualified vector and matrix norms are Euclidean and Frobenius norms, respectively. A function $\psi : \mathbb{R}^d \to \mathbb{R}$ is pseudo-Lipschitz of order two, or PL(2), if some $L < \infty$ satisfies

$$|\psi(\boldsymbol{v}) - \psi(\boldsymbol{w})| \le L(1 + \|\boldsymbol{v}\| + \|\boldsymbol{w}\|)\|\boldsymbol{v} - \boldsymbol{w}\| \quad \text{for all } \boldsymbol{v}, \boldsymbol{w}. \tag{26}$$

Time histories are concatenated when passed to such a test.

For a mode-n row array $\mathbf{V}_n \in \mathbb{R}^{I_n \times d}$, write

$$\|\mathbf{V}_n\|_n := \left(\frac{1}{I_n}\|\mathbf{V}_n\|_{\mathrm{F}}^2\right)^{1/2}. \tag{27}$$

The row-map condition used in the transfer is the following quantitative version of Assumption 4.2. For $\mathcal{B}^+_{n,K} = \{\mathbf{S} \in \mathbb{S}^{P_n}_+ : \|\mathbf{S}\|_{\mathrm{op}} \le K\}$ (the subscript op denotes the spectral (operator) norm), there is $L_{T,K} < \infty$ such that, uniformly over the fixed modes and times, for all $\boldsymbol{b}, \boldsymbol{b}' \in \mathbb{R}^{P_n}$ and $\mathbf{S}, \mathbf{S}' \in \mathcal{B}^+_{n,K}$,

$$\left\|f_{n,t}(\boldsymbol{b}, \mathbf{S}) - f_{n,t}(\boldsymbol{b}', \mathbf{S}')\right\| \le L_{T,K}\left\|\boldsymbol{b} - \boldsymbol{b}'\right\| + L_{T,K}(1 + \left\|\boldsymbol{b}'\right\|)\left\|\mathbf{S} - \mathbf{S}'\right\|, \tag{28}$$

$$\left\|D_{\boldsymbol{b}} f_{n,t}(\boldsymbol{b}, \mathbf{S}) - D_{\boldsymbol{b}} f_{n,t}(\boldsymbol{b}', \mathbf{S}')\right\| \le L_{T,K}\{\left\|\boldsymbol{b} - \boldsymbol{b}'\right\| + \left\|\mathbf{S} - \mathbf{S}'\right\|\}, \tag{29}$$

$$\sup_{\mathbf{S} \in \mathcal{B}^+_{n,K}} \left\|f_{n,t}(0, \mathbf{S})\right\| \le L_{T,K}, \qquad \sup_{\substack{\boldsymbol{b} \in \mathbb{R}^{P_n} \\ \mathbf{S} \in \mathcal{B}^+_{n,K}}} \left\|D_{\boldsymbol{b}} f_{n,t}(\boldsymbol{b}, \mathbf{S})\right\| \le L_{T,K}. \tag{30}$$

The core entries and all P_n are fixed, so their norms are absorbed into constants. No uniform positive-definiteness condition is imposed: the single-time SE covariances and Gaussian history covariance blocks may be singular.

C Finite-System General-Core Algebra

Recall that the finite-system core-induced covariance map $\mathcal{K}_{n,I}$ is defined in (10) with $\alpha_{n,I} = I_{\setminus n}/\nu_I$. Although that definition excludes $\mathbf{E}_n$ from the product, it retains the full tuple $\mathbf{E} = (\mathbf{E}_m)_{m=1}^N$ so that the same notation applies across output modes and cross-time overlaps. The map is evaluated on the signal overlaps (13) for the mean calculation and on the iterate overlaps (9) for the covariance calculation. By (2) and $\nu_I = \nu I^{N-1}$, $\alpha_{n,I} \to \alpha_n$.

C.1 Signal and covariance factorization

Expanding the signal part of (5), its r_nth coordinate is

$$\begin{aligned}
&\frac{1}{\nu_I} \sum_{\boldsymbol{i}_{\setminus n}} [\boldsymbol{\mathcal{Z}}]_{\boldsymbol{i}} [\boldsymbol{\phi}^t_{n,\boldsymbol{i}_{\setminus n}}]_{r_n} \\
&= \frac{1}{\nu_I} \sum_{\boldsymbol{p}, \boldsymbol{r}_{\setminus n}} \boldsymbol{\mathcal{C}}_{\boldsymbol{p}} \boldsymbol{\mathcal{C}}_{r_n, \boldsymbol{r}_{\setminus n}} [\mathbf{A}_n]_{i_n p_n} \prod_{m \ne n} \left\{ \sum_{i_m=1}^{I_m} [\mathbf{A}_m]_{i_m p_m} [\mathbf{X}^t_m]_{i_m r_m} \right\} \\
&= \sum_{p_n=1}^{P_n} [\mathbf{A}_n]_{i_n p_n} [\mathcal{K}_{n,I}(\mathbf{M}^t_I)]_{p_n r_n}.
\end{aligned} \tag{31}$$

The second line exchanges fixed channel sums with the sum over ambient multi-indices, and the last line uses (13).

For deterministic arrays $\mathbf{X}^s, \mathbf{X}^t$, define

$$[\mathcal{H}_{n,I}(\mathbf{X}^t)]_{i_n, r_n} = \frac{1}{\sqrt{\nu_I}} \sum_{\boldsymbol{i}_{\setminus n}} [\boldsymbol{\mathcal{G}}]_{\boldsymbol{i}} [\boldsymbol{\phi}^t_{n,\boldsymbol{i}_{\setminus n}}]_{r_n}. \tag{32}$$

For a fixed output row, independence of the Gaussian entries and the same modewise factorization of the sum over ambient multi-indices give

$$\begin{aligned}
&\mathbb{E}\Big[[\mathcal{H}_{n,I}(\mathbf{X}^s)]_{i_n, p_n} [\mathcal{H}_{n,I}(\mathbf{X}^t)]_{i_n, r_n}\Big] \\
&= \alpha_{n,I} \sum_{\boldsymbol{p}_{\setminus n}, \boldsymbol{r}_{\setminus n}} \boldsymbol{\mathcal{C}}_{p_n, \boldsymbol{p}_{\setminus n}} \boldsymbol{\mathcal{C}}_{r_n, \boldsymbol{r}_{\setminus n}} \prod_{m \ne n} \mathbf{Q}^{s,t}_{m,I}[p_m, r_m] = [\mathcal{K}_{n,I}(\mathbf{Q}^{s,t}_I)]_{p_n r_n}.
\end{aligned} \tag{33}$$

Rows with different i_n use disjoint Gaussian entries for deterministic inputs. Equation (33) is a covariance-kernel identity for the Gaussian operator. Replacing the deterministic factor rows by independent copies of the limiting row histories and $\alpha_{n,I}$ by α_n gives the population Gram representation used in Section 4. Hence every finite time-indexed history block $(\mathbf{K}^{s,t}_n)_{s,t}$ is positive semidefinite.

C.2 Reaction factorization and orientation

Although (7) displays the partial core contraction as a column, all differentials use the row map $\boldsymbol{\phi}_n^{\mathrm{row}} := \boldsymbol{\phi}_n^\top$. Write $D_{\boldsymbol{x}_m}\boldsymbol{\phi}_n^{\mathrm{row}} \in \mathbb{R}^{P_m \times P_n}$ so that $\delta\boldsymbol{\phi}_n^{\mathrm{row}} = \delta\boldsymbol{x}_m D_{\boldsymbol{x}_m}\boldsymbol{\phi}_n^{\mathrm{row}}$. For $t \geq 1$, put

$$\mathbf{J}_{m,i_m}^t = D_{\boldsymbol{b}} f_{m,t-1}\big([\mathbf{B}_m^{t-1}]_{i_m,\cdot}, \mathbf{K}_{m,I}^{t-1,t-1}\big).$$

This derivative holds the covariance argument fixed, consistently with (11). The notation $\mathbf{X}_{\setminus m}^t$ denotes the tuple of mode factors at time t with mode m omitted. Before factorizing the sums over ambient multi-indices, the finite-system chain-rule reaction is

$$\boldsymbol{\Omega}_{n,I}^t = \frac{1}{\nu_I} \sum_{\boldsymbol{i}_{\setminus n}} \sum_{m \neq n} D_{\boldsymbol{x}_n}\boldsymbol{\phi}_m^{\mathrm{row}}(\mathbf{X}_{\setminus m}^{t-1})\, \mathbf{J}_{m,i_m}^t\, D_{\boldsymbol{x}_m}\boldsymbol{\phi}_n^{\mathrm{row}}(\mathbf{X}_{\setminus n}^t). \tag{34}$$

The three factors have dimensions $P_n \times P_m$, $P_m \times P_m$, and $P_m \times P_n$. Thus (34) acts on a previous mode-n row from the right, exactly as in (5).

With the row convention $\delta f = \delta\boldsymbol{b}\,\mathbf{J}$, the (p_n, r_n) entry of the mode-m summand in (34) is

$$\alpha_{n,I} \sum_{\boldsymbol{p}_{\setminus n}, \boldsymbol{r}_{\setminus n}} \boldsymbol{\mathcal{C}}_{p_n, \boldsymbol{p}_{\setminus n}} \boldsymbol{\mathcal{C}}_{r_n, \boldsymbol{r}_{\setminus n}} \overline{\mathbf{J}}_{m,I}^t[p_m, r_m] \prod_{k \neq n,m} \mathbf{Q}_{k,I}^{t-1,t}[p_k, r_k]. \tag{35}$$

The first channel index belongs to the previous query, which multiplies the reaction on the left as $\mathbf{X}_n^{t-1}\boldsymbol{\Omega}_n^t$, and the second belongs to the current field. Comparing (35) with the derivative of (10) proves the reaction claim in Proposition 4.1. The two copies of $\boldsymbol{\mathcal{C}}$ arise from the current contraction and the earlier contraction differentiated through the denoiser. Appendix G.1 specializes this identity to the diagonal CP core.

C.3 Reconstruction-error consequences

Taking a squared-error test in Theorem 4.3 and summing over the fixed modes gives, for every fixed $0 \leq t \leq T+1$,

$$\frac{1}{\sum_{n=1}^N P_n} \sum_{n=1}^N \frac{1}{I_n} \left\| \mathbf{X}_n^t - \mathbf{A}_n \right\|_{\mathrm{F}}^2 \xrightarrow{\mathbb{P}} \frac{1}{\sum_{n=1}^N P_n} \sum_{n=1}^N \mathbb{E}\left\| \boldsymbol{x}_n^t - \boldsymbol{a}_n \right\|_2^2.$$

Thus SE predicts the asymptotic planted-coordinate factor MSE at every fixed iteration.

The same Cartesian factorization gives a tensor-level consequence. Reconstruct the latent tensor $\boldsymbol{\mathcal{Z}}$ from the iteration-t factors by

$$[\widehat{\boldsymbol{\mathcal{Z}}}^t]_{\boldsymbol{i}} = \sum_{\boldsymbol{p}} \boldsymbol{\mathcal{C}}_{\boldsymbol{p}} \prod_{n=1}^N [\mathbf{X}_n^t]_{i_n p_n}, \qquad \mathbf{S}_{aa,n,I} = \frac{1}{I_n}\mathbf{A}_n^\top \mathbf{A}_n, \quad \mathbf{S}_{aa,n} = \mathbb{E}[\boldsymbol{a}_n^\top \boldsymbol{a}_n].$$

Expanding the squared error and summing independently over the ambient modes gives the exact finite-system identity

$$\begin{aligned} &\frac{1}{\prod_{n=1}^N I_n} \left\| \widehat{\boldsymbol{\mathcal{Z}}}^t - \boldsymbol{\mathcal{Z}} \right\|_{\mathrm{F}}^2 \\ &\quad = \sum_{\boldsymbol{p},\boldsymbol{r}} \boldsymbol{\mathcal{C}}_{\boldsymbol{p}} \boldsymbol{\mathcal{C}}_{\boldsymbol{r}} \left\{ \prod_{n=1}^N \mathbf{S}_{aa,n,I}[p_n, r_n] + \prod_{n=1}^N \mathbf{Q}_{n,I}^{t,t}[p_n, r_n] - 2 \prod_{n=1}^N \mathbf{M}_{n,I}^t[p_n, r_n] \right\}. \end{aligned} \tag{36}$$

The planted-row assumptions give $\mathbf{S}_{aa,n,I} \xrightarrow{\mathbb{P}} \mathbf{S}_{aa,n}$. At $t = T+1$, extend the population notation by the same moment definitions $\mathbf{M}_n^{T+1} = \mathbb{E}[\boldsymbol{a}_n^\top \boldsymbol{x}_n^{T+1}]$ and $\mathbf{Q}_n^{T+1,T+1} = \mathbb{E}[(\boldsymbol{x}_n^{T+1})^\top \boldsymbol{x}_n^{T+1}]$. The corresponding empirical moments converge to these quantities by the row-law conclusion. Hence, for every fixed $0 \leq t \leq T+1$, the right-hand side converges in probability to

$$\sum_{\boldsymbol{p},\boldsymbol{r}} \boldsymbol{\mathcal{C}}_{\boldsymbol{p}} \boldsymbol{\mathcal{C}}_{\boldsymbol{r}} \left\{ \prod_{n=1}^N \mathbf{S}_{aa,n}[p_n, r_n] + \prod_{n=1}^N \mathbf{Q}_n^{t,t}[p_n, r_n] - 2 \prod_{n=1}^N \mathbf{M}_n^t[p_n, r_n] \right\}.$$

This is the SE prediction for reconstruction error relative to the latent tensor $\boldsymbol{\mathcal{Z}}$.

C.4 STABILITY TOOLS USED BELOW

The source state-evolution and oracle-transfer arguments use two estimates that depend only on the finite-system core calculus and the Gaussian multilinear operator. We establish them here before either reduction is invoked.

Lemma C.1 (Bounded-set map stability)**.** *Define the population Onsager reaction map* $\mathcal{O}_n(\mathbf{Q},\mathbf{J}) = \sum_{m\neq n} D_m\mathcal{K}_n(\mathbf{Q})[\mathbf{J}_m]$*. For every* $K<\infty$*, there is a constant* C_K *depending only on* K*, the fixed core and dimensions, noise level* ν*, and the aspect ratios such that, whenever all arguments have Frobenius norm at most* K*,*

$$\left\|\mathcal{K}_n(\mathbf{E})-\mathcal{K}_n(\widetilde{\mathbf{E}})\right\| \le C_K\sum_{m\neq n}\left\|\mathbf{E}_m-\widetilde{\mathbf{E}}_m\right\|, \tag{37}$$

$$\left\|\mathcal{O}_n(\mathbf{Q},\mathbf{J})-\mathcal{O}_n(\widetilde{\mathbf{Q}},\widetilde{\mathbf{J}})\right\| \le C_K\left\{\sum_m\left\|\mathbf{Q}_m-\widetilde{\mathbf{Q}}_m\right\|+\sum_m\left\|\mathbf{J}_m-\widetilde{\mathbf{J}}_m\right\|\right\}, \tag{38}$$

The bounds hold uniformly for the finite-system maps $\mathcal{K}_{n,I}$ *once* $\sup_I \alpha_{n,I}<\infty$*.*

Proof. Each entry is a finite sum of products of fixed core entries and one matrix entry from every mode except n. Telescope each product, bound the remaining factors by K, and sum over the fixed channel sets. For (38), one factor is a Jacobian entry and the remaining factors are overlap entries. The same telescoping argument applies. □

Lemma C.2 (Uniform Gaussian multilinear field stability)**.** *There exists a constant* $c>0$ *such that, for every deterministic* $K<\infty$*, there is a constant* $C_K<\infty$ *for which, with probability at least* $1-2e^{-cI}$*,*

$$\max_n\left\|\mathcal{H}_{n,I}(\mathbf{X})-\mathcal{H}_{n,I}(\widetilde{\mathbf{X}})\right\|_n \le C_K\max_n\sum_{m\neq n}\left\|\mathbf{X}_m-\widetilde{\mathbf{X}}_m\right\|_m \tag{39}$$

simultaneously over all pairs of row-array collections satisfying $\max_m\{\|\mathbf{X}_m\|_m, \left\|\widetilde{\mathbf{X}}_m\right\|_m\}\le K$*.*

Proof. We use $\otimes$ for the usual tensor product. For vectors $\boldsymbol{z}_m\in\mathbb{R}^{I_m}$, the tensor $\boldsymbol{z}_1\otimes\cdots\otimes\boldsymbol{z}_N$ is the rank-one outer product whose entry at $\boldsymbol{i}$ is $\prod_{m=1}^N[\boldsymbol{z}_m]_{i_m}$. Define the rectangular injective norm

$$\|\boldsymbol{\mathcal{G}}\|_{\mathrm{inj}} = \sup_{\|\boldsymbol{z}_m\|\le 1}\left|\langle\boldsymbol{\mathcal{G}},\boldsymbol{z}_1\otimes\cdots\otimes\boldsymbol{z}_N\rangle\right|. \tag{40}$$

By multilinearity, the supremum may be restricted to the unit spheres. For each mode m, choose a finite $1/(2N)$-net $\mathcal{N}_m$ of the Euclidean unit sphere, so every unit vector lies within Euclidean distance $1/(2N)$ of some point in $\mathcal{N}_m$. Such a net may be obtained as a maximal $1/(2N)$-separated subset. The disjoint Euclidean balls of radius $1/(4N)$ around its points lie in the ball of radius $1+1/(4N)$, so

$$|\mathcal{N}_m|\le(1+4N)^{I_m}, \qquad \left|\prod_{m=1}^N\mathcal{N}_m\right| \le \exp\left(C_N\sum_{m=1}^N I_m\right).$$

At each product-net point the multilinear form in (40) is standard Gaussian because the corresponding rank-one tensor has Frobenius norm one. For any tuple of unit vectors, choose a net point in each mode. Telescoping one argument at a time and using (40) gives

$$\|\boldsymbol{\mathcal{G}}\|_{\mathrm{inj}} \le \max_{\widehat{\boldsymbol{z}}_m\in\mathcal{N}_m}\left|\langle\boldsymbol{\mathcal{G}},\widehat{\boldsymbol{z}}_1\otimes\cdots\otimes\widehat{\boldsymbol{z}}_N\rangle\right| + \frac{1}{2}\|\boldsymbol{\mathcal{G}}\|_{\mathrm{inj}}.$$

Hence the injective norm is at most twice the product-net maximum. For every $u>0$, the standard Gaussian tail bound and a union bound therefore give

$$\mathbb{P}\{\|\boldsymbol{\mathcal{G}}\|_{\mathrm{inj}}>2u\} \le 2\exp\left(C_N\sum_{m=1}^N I_m-\frac{u^2}{2}\right).$$

Since $\sum_m I_m = \Theta(I)$, choosing u as a sufficiently large fixed multiple of $\sqrt{I}$ gives constants $C, c > 0$ such that

$$\mathbb{P}\{\|\boldsymbol{\mathcal{G}}\|_{\mathrm{inj}} > C\sqrt{I}\} \le 2e^{-cI}. \tag{41}$$

Fix an output mode and channel, expand the core sum, and telescope each order-$(N-1)$ product between $\mathbf{X}$ and $\widetilde{\mathbf{X}}$. Test the output column against a unit vector in mode n and apply (40). Summing over the fixed channel multi-indices and using $\left\|[\mathbf{X}_m]_{\cdot,r_m}\right\| \le \sqrt{I_m}\|\mathbf{X}_m\|_m$ gives the deterministic bound

$$\left\|\mathcal{H}_{n,I}(\mathbf{X}) - \mathcal{H}_{n,I}(\widetilde{\mathbf{X}})\right\|_n \le C_0 K^{N-2}\|\boldsymbol{\mathcal{G}}\|_{\mathrm{inj}} \frac{\sqrt{I_{\setminus n}}}{\sqrt{I_n \nu_I}} \sum_{m \ne n} \left\|\mathbf{X}_m - \widetilde{\mathbf{X}}_m\right\|_m, \tag{42}$$

where C_0 depends only on the fixed core and channel dimensions. The size factor is

$$\frac{\sqrt{I_{\setminus n}}}{\sqrt{I_n \nu_I}} = \frac{\sqrt{I}}{\sqrt{\nu} I_n} = \frac{1}{\sqrt{\nu}(I_n/I)\sqrt{I}}. \tag{43}$$

The aspect-ratio assumption bounds I_n/I away from zero. Combining (41)–(43) proves (39). The deterministic inequality holds simultaneously for all bounded arrays on the single injective-norm event, so no independence between the arrays and the tensor is required. In particular, it applies to adaptive trajectories driven by the same tensor. □

D Gaussian Embedding, Serialization, and Source State Evolution

In the proof appendices, *native* refers to the original mode-specific vector coordinates of Section 3, before the scalar channel-index serialization introduced below. An oracle recursion replaces empirical coefficients by their deterministic SE counterparts. Section E.1 states the two auxiliary recursions explicitly. This appendix first constructs the exact serialization of the fixed-interaction oracle, proves the required pure-degree source row law, and removes its history and interaction degeneracies.

Proof structure. The argument separates the algebraic reduction, the source-SE proof, and the two limiting transfers.

1. The first three subsections embed the mode blocks as species and replace each vector-valued update by a predetermined finite list of scalar channel-index queries. Lemma D.1 then shows that their core-weighted corrected answers reproduce the original-coordinate Gaussian field, covariance, and chain-rule Onsager reaction while retaining the same Gaussian disorder.
2. For the single active degree $N \ge 3$, the exact covariance operator of the stacked Gaussian responses is coercive whenever the query-history Gram is nondegenerate. Adaptive Gaussian conditioning and integration by parts then prove the source row law at strictly positive interactions and nondegenerate histories.
3. Same-degree Gaussian fill makes the sparse species interaction strictly positive, while independent query dither makes the population query-history Grams positive definite. Fixed-horizon L^2 continuity removes both regularizations. The proof treats deterministic row covariates directly through their empirical Wasserstein–2 limits and handles random planted–initial arrays by conditioning.
4. Appendix D.6 identifies the zero-regularization endpoint with the fixed-interaction oracle. Appendix E then restores the finite-system Gaussian scale and replaces population covariance and reaction coefficients by the empirical coefficients of the AMP recursion.

The regularizations are fixed before the ambient limit and removed only after the source row law has been established at each positive regularization.

D.1 Exact multi-species embedding

Set $L_I = \sum_n I_n$ and partition $[L_I]$ into mode species $\mathcal{V}_{1,I}, \dots, \mathcal{V}_{N,I}$. The tensor $\mathcal{G} \in (\mathbb{R}^{L_I})^{\otimes N}$ has i.i.d. standard Gaussian entries. For any positive integer k, S_k denotes the permutation group on k

symbols. For an ambient multi-index $\boldsymbol{i}$, viewed through the species embedding, put

$$\overline{G}_{\boldsymbol{i}} = \frac{1}{\sqrt{N!}} \sum_{\pi \in S_N} \mathcal{G}_{i_{\pi(1)},\dots,i_{\pi(N)}}. \tag{44}$$

Different ambient multi-indices have disjoint permutation orbits, so the $\overline{G}_{\boldsymbol{i}}$ are i.i.d. $\mathcal{N}(0,1)$. The finite-system interaction scale is

$$g_I = \left(\frac{L_I^{N-1}}{N! \nu_I} \right)^{1/2} \tag{45}$$

Assign this coefficient to species tuples that are permutations of $(1,\dots,N)$ and assign zero elsewhere. With the symmetric contraction normalized by $1/(N-1)!$, this scale produces exactly the species-n field

$$\frac{1}{\sqrt{\nu_I}} \sum_{\boldsymbol{i}_{\setminus n}} \overline{G}_{\boldsymbol{i}} \prod_{m \neq n} x_{m,i_m}. \tag{46}$$

The $N!$ in g_I normalizes the orbit, while $1/(N-1)!$ removes the admissible orderings of the remaining species. Since $L_I/I \to \sum_n c_n$, the scale converges to the fixed positive coefficient g defined by

$$g^2 = \frac{(\sum_m c_m)^{N-1}}{N! \nu}. \tag{47}$$

The source analysis below is carried out at this fixed coefficient. On the same symmetric Gaussian tensor, define its mode-n operator by

$$\mathcal{H}^{\circ}_{n,I} = \frac{g}{g_I} \mathcal{H}_{n,I}. \tag{48}$$

Thus $\mathcal{H}^{\circ}_{n,I}$ differs from the finite-system Gaussian operator only by a deterministic factor tending to one. The source normalization follows from the degree-N source polynomial and its relevant derivatives:

$$\begin{aligned} \xi_N(\boldsymbol{q}) &= N! g^2 \prod_m (\lambda_m q_m), \\ \lambda_n^{-1} \partial_{q_n} \xi_N(\boldsymbol{q}) &= \alpha_n \prod_{m \neq n} q_m, \qquad \partial_{q_m} \Big(\lambda_n^{-1} \partial_{q_n} \xi_N(\boldsymbol{q}) \Big) = \alpha_n \prod_{k \neq n,m} q_k, \end{aligned} \tag{49}$$

where $\lambda_n = c_n/(\sum_m c_m)$ and $\boldsymbol{q} = (q_1,\dots,q_N)$ collects the species-overlap arguments. At the finite scale g_I, the same calculation gives

$$N! g_I^2 \prod_{m \neq n} \frac{I_m}{L_I} = \frac{I_{\setminus n}}{\nu_I} = \alpha_{n,I}. \tag{50}$$

D.2 Predetermined channel-index queries

Define

$$\mathcal{D} = [P_1] \times \cdots \times [P_N], \qquad L_{\mathrm{tuple}} = |\mathcal{D}| = \prod_{n=1}^{N} P_n. \tag{51}$$

At macro time t, the primitive query indexed by $\boldsymbol{r} = (r_1,\dots,r_N) \in \mathcal{D}$ restricts on species m to the scalar column $[\mathbf{X}_m^t]_{\cdot,r_m}$. Enumerate these L_{tuple} queries in any deterministic order. The fixed-interaction primitive answer in output species n is

$$h^{\circ,t}_{n,\boldsymbol{r}}(i_n) = \frac{g}{g_I \sqrt{\nu_I}} \sum_{\boldsymbol{i}_{\setminus n}} \overline{G}_{\boldsymbol{i}} \prod_{m \neq n} [\mathbf{X}_m^t]_{i_m,r_m}. \tag{52}$$

Although this answer does not depend on r_n, retaining the full channel multi-index gives a common schedule for all output species. Repeated or linearly dependent answers are allowed.

More explicitly, fix $\mathrm{ord} : \mathcal{D} \to \{0,\dots,L_{\mathrm{tuple}}-1\}$ and assign

$$u(t,\boldsymbol{r}) = t L_{\mathrm{tuple}} + \mathrm{ord}(\boldsymbol{r}) \tag{53}$$

as the global source micro-time. At $u(t, \boldsymbol{r})$, the coordinate map on a mode-m species returns $[\mathbf{X}_m^t]_{\cdot,r_m}$, and the prehistory before the first block is fixed at zero. This makes the macro-to-micro correspondence deterministic.

The r_nth core field is the deterministic linear combination

$$[\mathcal{H}_{n,I}^{\circ}(\mathbf{X}^t)]_{\cdot,r_n} = \sum_{\boldsymbol{r}_{\setminus n}} \boldsymbol{\mathcal{C}}_{r_n,\boldsymbol{r}_{\setminus n}}\, h^{\circ,t}_{n,(r_n,\boldsymbol{r}_{\setminus n})}. \tag{54}$$

Every primitive query in macro block t is a function only of answers from blocks $0,\dots,t-1$ and the row side information. In a sequential micro-time representation, its coordinate map ignores all answers revealed earlier in the same block. The source recursion therefore has a fixed list of $L_{\mathrm{tuple}}(T+1)$ admissible scalar queries.

D.3 Corrected answers and the block identity

The primitive quantity in (52) is an uncorrected tensor answer. To connect the serialized source AMP to the fixed-interaction oracle, define its corrected answer by $w^{\circ,0}_{n,\boldsymbol{r}} = h^{\circ,0}_{n,\boldsymbol{r}}$ and, for $t \ge 1$,

$$w^{\circ,t}_{n,\boldsymbol{r}} = h^{\circ,t}_{n,\boldsymbol{r}} - \sum_{\boldsymbol{p}\in\mathcal{D}} [\mathbf{X}_n^{t-1}]_{\cdot,p_n}\, d^t_{n,\boldsymbol{r},\boldsymbol{p}}, \tag{55}$$

$$d^t_{n,\boldsymbol{r},\boldsymbol{p}} = \alpha_n \sum_{m\neq n} \boldsymbol{\mathcal{C}}_{\boldsymbol{p}}\, \overline{\mathbf{J}}^t_m[p_m, r_m] \prod_{k\neq n,m} \mathbf{Q}_k^{t-1,t}[p_k, r_k]. \tag{56}$$

Example: channel-index serialization for $N = 3$. Take $P_1 = P_2 = P_3 = 2$. Then $\mathcal{D} = \{1,2\}^3$ and $L_{\mathrm{tuple}} = 8$. Fix an enumeration $\boldsymbol{r}^{(1)},\dots,\boldsymbol{r}^{(8)}$ and use micro-times $\ell = 1,\dots,8$ inside macro block t. For output species 1,

$$h^{\circ,t}_{1,(r_1,r_2,r_3)}(i_1) = \frac{g}{g_I\sqrt{\nu_I}} \sum_{i_2,i_3} \overline{G}_{i_1 i_2 i_3} [\mathbf{X}_2^t]_{i_2,r_2} [\mathbf{X}_3^t]_{i_3,r_3}.$$

All columns $\mathbf{X}_m^t$ are fixed before this block is exposed. Hence a query at micro-time ℓ has zero derivative with respect to every same-block answer at a micro-time $\ell' < \ell$. Its only relevant derivatives are through the preceding block, and for $m \in \{2,3\}$ they are

$$\frac{\partial [\mathbf{X}_m^t]_{i_m,r_m}}{\partial w^{\circ,t-1}_{m,\boldsymbol{p}}(i_m)} = \boldsymbol{\mathcal{C}}_{\boldsymbol{p}}\, \mathbf{J}^t_{m,i_m}[p_m, r_m]. \tag{57}$$

Although the primitive answer is repeated across r_1, fixing the output channel in the core-weighted field visits each pair (r_2, r_3) exactly once. For fixed $(\boldsymbol{r}, \boldsymbol{p})$, the coefficient (56) sums the two distinct derivative paths $m = 2$ and $m = 3$ once each. Fixing the current output channel r_1 and the previous output channel p_1 likewise visits every current–previous channel multi-index pair once. Thus the serialization introduces no multiplicity beyond the two chain-rule paths.

These are precisely the nonzero source-AMP coefficients: combining (49) with (57) gives (56), and a query in block t depends explicitly only on the corrected answers in block $t-1$. The vector field supplied to the next row denoiser is

$$[\mathbf{B}_n^t]_{\cdot,r_n} = [\mathbf{A}_n \boldsymbol{\Theta}_n^t]_{\cdot,r_n} + \sum_{\boldsymbol{r}_{\setminus n}} \boldsymbol{\mathcal{C}}_{r_n,\boldsymbol{r}_{\setminus n}}\, w^{\circ,t}_{n,(r_n,\boldsymbol{r}_{\setminus n})}. \tag{58}$$

Thus every scalar query map is a fixed Lipschitz function of the preceding corrected block and the row side information.

Lemma D.1 (Same-disorder block serialization)**.** *The core-weighted corrected answers obey*

$$\sum_{\boldsymbol{r}_{\setminus n}} \boldsymbol{\mathcal{C}}_{r_n,\boldsymbol{r}_{\setminus n}}\, w^{\circ,t}_{n,(r_n,\boldsymbol{r}_{\setminus n})} = [\mathcal{H}_{n,I}^{\circ}(\mathbf{X}^t)]_{\cdot,r_n} - [\mathbf{X}_n^{t-1} \boldsymbol{\Omega}_n^t]_{\cdot,r_n}, \tag{59}$$

where $\boldsymbol{\Omega}_n^t = \sum_{m\neq n} D_m \mathcal{K}_n(\mathbf{Q}^{t-1,t})[\overline{\mathbf{J}}^t_m]$.

Proof. For the derivative coefficient, a current query column r_m depends on the previous vector field through $D_{\boldsymbol{b}} f_{m,t-1}[\cdot, r_m]$. More explicitly,

$$\frac{\partial x^t_{m,r_m}}{\partial w^{\circ,t-1}_{m,\boldsymbol{p}}} = \boldsymbol{\mathcal{C}}_{\boldsymbol{p}}\, \mathbf{J}^t_{m,i_m}[p_m, r_m]. \tag{60}$$

The corresponding source-SE coefficient uses the population history-Jacobian expectation $\overline{\mathbf{J}}^t_m$. The coefficient against the previous query $[\mathbf{X}^{t-1}_n]_{\cdot,p_n}$ is therefore (56). After weighting the current answers by the core, the coefficient from previous channel p_n to current channel r_n is

$$\sum_{\boldsymbol{r}_{\setminus n}, \boldsymbol{p}_{\setminus n}} \boldsymbol{\mathcal{C}}_{r_n, \boldsymbol{r}_{\setminus n}} d^t_{n,(r_n,\boldsymbol{r}_{\setminus n}),(p_n,\boldsymbol{p}_{\setminus n})}. \tag{61}$$

Substituting (56) and collecting the previous–current indices in their displayed order gives

$$\begin{aligned}
&\sum_{\boldsymbol{r}_{\setminus n}, \boldsymbol{p}_{\setminus n}} \boldsymbol{\mathcal{C}}_{r_n, \boldsymbol{r}_{\setminus n}} d^t_{n,(r_n,\boldsymbol{r}_{\setminus n}),(p_n,\boldsymbol{p}_{\setminus n})} \\
&\quad = \alpha_n \sum_{m \neq n} \sum_{\boldsymbol{r}_{\setminus n}, \boldsymbol{p}_{\setminus n}} \boldsymbol{\mathcal{C}}_{r_n, \boldsymbol{r}_{\setminus n}} \boldsymbol{\mathcal{C}}_{p_n, \boldsymbol{p}_{\setminus n}} \overline{\mathbf{J}}^t_m[p_m, r_m] \prod_{k \neq n, m} \mathbf{Q}^{t-1,t}_k[p_k, r_k] \\
&\quad = \left[\sum_{m \neq n} D_m \mathcal{K}_n(\mathbf{Q}^{t-1,t})[\overline{\mathbf{J}}^t_m] \right]_{p_n, r_n}.
\end{aligned} \tag{62}$$

The first channel multi-index in the derivative of $\mathcal{K}_n$ is therefore $\boldsymbol{p}$ and the second is $\boldsymbol{r}$, which is why a previous channel indexes the row and a current channel the column of the reaction. Thus (61) is $\boldsymbol{\Omega}^t_n[p_n, r_n]$, which proves (59). □

D.4 Pure-degree source state evolution

We prove the pure-degree source result needed by the serialized recursion. The source degree $N \geq 3$, species set, and query horizon are fixed. The proof follows the adaptive Gaussian-conditioning architecture of Huang & Sellke (2024, Appendix A), while deriving the exact covariance operator and its coercivity directly for this pure-degree setting.

Write d for the source ambient dimension. A fixed finite species set $\mathcal{S}$, whose cardinality need not equal N, partitions the coordinates into blocks $V_{s,d}$ of sizes d_s and proportions

$$\lambda_{s,d} = d_s/d \longrightarrow \lambda_s > 0, \qquad \sum_s \lambda_s = 1.$$

All exact finite-dimensional identities below use $\lambda_{s,d}$, not its limit. Put

$$R_{s,d}(x,y) = d_s^{-1} x_s^\top y_s, \qquad \|Z\|_d = d^{-1/2} \|Z\|_{\mathrm{F}}$$

for a vector or a fixed-column matrix. Let $\gamma_{s_1 \ldots s_N} \geq 0$ be fixed, permutation-symmetric tensor amplitudes. For an ordered i.i.d. standard Gaussian tensor G define

$$H_d(x) = d^{-(N-1)/2} \sum_{i_1, \ldots, i_N = 1}^{d} \gamma_{s(i_1)\ldots s(i_N)} G_{i_1 \ldots i_N} x_{i_1} \cdots x_{i_N}, \qquad h_d = \nabla H_d. \tag{63}$$

In entrywise products, Γ denotes the lifted ambient multiplier $\Gamma_{i_1 \ldots i_N} = \gamma_{s(i_1)\ldots s(i_N)}$. Thus symmetrization is implicit in the gradient. Equivalently, the symmetric tensor in source conventions is $N!\, d^{-(N-1)/2}(\Gamma \odot \operatorname{Sym} G)$ and its contraction is divided by $(N-1)!$. Here Sym denotes the average over all tensor-coordinate permutations, so it is the orthogonal projection onto the symmetric tensor subspace for the Frobenius inner product.

Define the finite-system covariance polynomial by

$$\xi_d(q) = \sum_{s_1, \ldots, s_N} \gamma^2_{s_1 \ldots s_N} \prod_{a=1}^{N} \lambda_{s_a,d} q_{s_a}. \tag{64}$$

The population polynomial ξ is obtained by replacing $\lambda_{s,d}$ with λ_s. Its normalized derivatives and their finite-system counterparts are

$$\xi_{s,d}(q) = \lambda^{-1}_{s,d} \partial_s \xi_d(q), \qquad \xi_s(q) = \lambda_s^{-1} \partial_s \xi(q).$$

For deterministic x, y, $\mathbb{E} H_d(x) H_d(y) = d \xi_d((R_{s,d}(x,y))_s)$.

Rows and maps. For $i \in V_{s,d}$, let $e_i \in \mathbb{R}^{q_s}$ be deterministic row covariates such that, within every species,

$$\frac{1}{d_s} \sum_{i \in V_{s,d}} \delta_{e_i} \longrightarrow \mu_s \quad \text{in } \mathsf{W}_2. \tag{65}$$

The species-dependent dimensions q_s are fixed. At source time t, a query uses a fixed map

$$\phi_{s,t} : \mathbb{R}^t \times \mathbb{R}^{q_s} \to \mathbb{R}.$$

For the fixed horizon under consideration, these maps are globally Lipschitz jointly in the history and the covariate, are C^1 in the history, and have bounded history gradients that are globally Lipschitz jointly in the history and the covariate. Constants may depend on the fixed horizon. An independent finite-dimensional external vector, including the query dither introduced later, can be appended to e_i. Lemma D.7 justifies this extension. The notation $w^{1:0}$ means the empty history.

D.4.1 Deterministic coercivity

Lemma D.2 (Symmetrized-query norm identity for $N \geq 3$)**.** *Let $v_1, \dots, v_u$ be orthonormal in $\mathbb{R}^d$ and let $z_1, \dots, z_u \in \mathbb{R}^d$. Extend these v_r to an orthonormal basis and write $z_{jr} = \langle v_j, z_r \rangle$. For $N \geq 3$,*

$$\left\| \sum_{r=1}^{u} \mathrm{Sym}(z_r \otimes v_r^{\otimes(N-1)}) \right\|_F^2 = \sum_{r=1}^{u} z_{rr}^2 + \frac{1}{N} \sum_{r=1}^{u} \sum_{\substack{1 \leq j \leq d \\ j \neq r}} z_{jr}^2. \tag{66}$$

Consequently this squared norm lies between $N^{-1} \sum_r \|z_r\|^2$ and $\sum_r \|z_r\|^2$.

Proof. For $j = r$, the tensor is $v_r^{\otimes N}$ and has norm one. For $j \neq r$, the symmetrized tensor has N orthogonal ordered tensors, each with coefficient $1/N$, and thus squared norm $1/N$. Distinct pairs (j, r) correspond to distinct multisets of basis indices when $N \geq 3$: the index repeated $N - 1$ times determines r. All these tensor components are therefore orthogonal, proving (66). At $N = 2$, the pairs (j, r) and (r, j) yield the same multiset, producing the skew-symmetric null directions excluded by the assumption $N \geq 3$. □

Lemma D.3 (Coercivity of the stacked-response covariance operator)**.** *Let $F = [f^0 | \cdots | f^{u-1}] \in \mathbb{R}^{d \times u}$, with*

$$a I_u \preceq d^{-1} F^\top F \preceq b I_u, \qquad 0 < a \leq b < \infty.$$

Suppose $0 < \gamma_- \leq \gamma_{s_1 \dots s_N} \leq \gamma_+ < \infty$. Freeze F and let $\mathcal{A}_F$ send the standard ordered Gaussian tensor to the stacked gradient observations $(h_d(f^r))_{r<u}$. Let

$$\mathcal{M}_F = \mathcal{A}_F \mathcal{A}_F^*$$

be their covariance operator, in the ordinary Euclidean/Frobenius metric. Set

$$\rho_- = \min(1, \sqrt{a}), \qquad \rho_+ = \max(1, \sqrt{b}).$$

Then

$$\frac{N \gamma_-^2 \rho_-^{2N}}{\rho_+^2} \, \mathrm{Id} \preceq \mathcal{M}_F \preceq \frac{N^2 \gamma_+^2 \rho_+^{2N}}{\rho_-^2} \, \mathrm{Id}\,. \tag{67}$$

The bounds are independent of d.

Proof. Write $q_r = f^r / \sqrt{d}$. For $Z = [z_0 | \cdots | z_{u-1}]$, differentiation of (63) gives the exact adjoint identity

$$\mathcal{A}_F^* Z = N \, \Gamma \odot \sum_{r<u} \mathrm{Sym}(z_r \otimes q_r^{\otimes(N-1)}). \tag{68}$$

In particular,

$$\langle Z, \mathcal{M}_F Z \rangle = \|\mathcal{A}_F^* Z\|_F^2\,.$$

The normalized query matrix $\widetilde{F} = F/\sqrt{d} = [q_0 | \cdots | q_{u-1}]$ has polar decomposition $\widetilde{F} = V(\widetilde{F}^\top \widetilde{F})^{1/2}$, where V has orthonormal columns, and write v_r for the rth column of V. There

is a positive-definite operator B on $\mathbb{R}^d$ that acts as $V(\widetilde{F}^\top \widetilde{F})^{1/2} V^\top$ on $\mathrm{range}(V)$ and as the identity on its orthogonal complement. Hence $q_r = Bv_r$, and $\rho_- \,\mathrm{Id} \preceq B \preceq \rho_+ \,\mathrm{Id}$. With $y_r = B^{-1} z_r$,

$$\sum_r \mathrm{Sym}(z_r \otimes q_r^{\otimes(N-1)}) = B^{\otimes N} \sum_r \mathrm{Sym}(y_r \otimes v_r^{\otimes(N-1)}).$$

Lemma D.2 and the bounds on B give

$$\frac{\rho_-^{2N}}{N\rho_+^2}\|Z\|_{\mathrm{F}}^2 \le \left\| \sum_r \mathrm{Sym}(z_r \otimes q_r^{\otimes(N-1)}) \right\|_F^2 \le \frac{\rho_+^{2N}}{\rho_-^2}\|Z\|_{\mathrm{F}}^2 .$$

Entrywise multiplication by Γ multiplies squared norms by factors between γ_-^2 and γ_+^2. Applying the bounds on $B^{\otimes N}$ and the species-amplitude multiplier successively, then using (68), proves (67). □

Remark D.4. The Gaussian fill introduced below is pure degree N, so the regularized models remain within the $N \ge 3$ regime of Lemma D.2.

D.4.2 Exact covariance and Gaussian conditioning

For frozen queries define $Q_d^{\ell r} = (R_{s,d}(f^\ell, f^r))_s$. Differentiating the exact Hamiltonian covariance twice gives

$$\mathrm{Cov}(h_d(f^\ell)_i, h_d(f^r)_j) = \mathbf{1}_{\{i=j\}}\, \xi_{s,d}(Q_d^{\ell r}) + \frac{f_i^r f_j^\ell}{d_v}\, \partial_v \xi_{s,d}(Q_d^{\ell r}), \quad s = s(i),\ v = s(j). \tag{69}$$

Covariance symmetry follows from

$$\frac{\partial_v \xi_{s,d}}{d_v} = \frac{\partial_{sv} \xi_d}{d\lambda_{s,d}\lambda_{v,d}} = \frac{\partial_s \xi_{v,d}}{d_s}.$$

For a stack of u queries, define operators on $\mathbb{R}^{d\times u}$ by

$$(\mathcal{D}_u Z)_s^\ell = \sum_{r<u} D_{s,d}[\ell, r] Z_s^r, \tag{70}$$

$$(\mathcal{B}_u Z)_s^\ell = \sum_{r<u} \sum_v f_s^r\, \partial_v \xi_{s,d}(Q_d^{\ell r}) R_{v,d}(f^\ell, Z^r), \tag{71}$$

where $D_{s,d}[\ell, r] = \xi_{s,d}(Q_d^{\ell r})$. Then the stacked response covariance is exactly

$$\mathcal{M}_u = \mathcal{D}_u + \mathcal{B}_u. \tag{72}$$

The finite-rank operator $\mathcal{B}_u$ may be indefinite, whereas $\mathcal{M}_u$ is positive semidefinite by construction. The range of $\mathcal{B}_u$ lies in the space in which each history column, in each species, belongs to $\mathrm{span}\{f_s^r : r < u\}$. Its rank is at most $|\mathcal{S}|u^2$.

Lemma D.5 (Response-Gram and preconditioned-operator bounds)**.** *Under Lemma D.3,*

$$N\gamma_-^2 a^{N-1}\,\mathrm{Id} \preceq \mathcal{D}_u \preceq N\gamma_+^2 b^{N-1}\,\mathrm{Id}\,. \tag{73}$$

Thus the preconditioned normal operator

$$\mathcal{L}_u := \mathcal{M}_u \mathcal{D}_u^{-1} = \mathrm{Id} + \mathcal{B}_u \mathcal{D}_u^{-1} \tag{74}$$

has its smallest and largest singular values bounded uniformly away from zero and infinity. In particular, if $\kappa\,\mathrm{Id} \preceq \mathcal{M}_u \preceq K\,\mathrm{Id}$ and $d_-\,\mathrm{Id} \preceq \mathcal{D}_u \preceq d_+\,\mathrm{Id}$, then

$$\sigma_{\min}(\mathcal{L}_u) \ge \kappa/d_+, \qquad \|\mathcal{L}_u\|_{\mathrm{op}} \le K/d_-.$$

Proof. Define $\mathbf{P}_{s,d} = (R_{s,d}(f^\ell, f^r))_{\ell, r<u}$ and $\mathbf{G}_d = \sum_s \lambda_{s,d}\mathbf{P}_{s,d} = d^{-1}F^\top F$. Expanding $\xi_{s,d}$ as a weighted sum of Hadamard products gives

$$N\gamma_-^2 \mathbf{G}_d^{\odot(N-1)} \preceq D_{s,d} \preceq N\gamma_+^2 \mathbf{G}_d^{\odot(N-1)}.$$

Each difference is a nonnegative weighted sum of Schur products of positive-semidefinite matrices. From $\mathbf{G}_d \succeq aI$, expand $\mathbf{G}_d = aI + \mathbf{R}$ and use the Schur product theorem to obtain $\mathbf{G}_d^{\odot(N-1)} \succeq a^{N-1}I$. The upper bound follows by regarding the Hadamard product as a principal compression of the tensor product, so its operator norm is at most b^{N-1}. This proves (73). The singular-value bounds follow from (74). Although $\mathcal{L}_u$ may be non-self-adjoint, it is similar to the positive-definite operator $\mathcal{D}_u^{-1/2}\mathcal{M}_u\mathcal{D}_u^{-1/2}$. □

Exact adaptive Gaussian conditioning. Suppose the query f^r is a measurable function of the covariates and the corrected responses before its query time. The corrections used below are themselves functions of these earlier responses and fixed coefficients. Consequently, conditional on the complete past sigma-field $\mathcal{F}_u$, all past queries are fixed and all observed raw responses impose linear constraints on the Gaussian tensor. If $Y_u = (h_d(f^r))_{r<u}$ and $\mathcal{M}_u$ is invertible, the conditional law of the ordered standard Gaussian tensor is

$$G \mid \mathcal{F}_u \overset{d}{=} \mathcal{A}_F^* \mathcal{M}_u^{-1} Y_u + (\mathrm{Id} - \mathcal{A}_F^* \mathcal{M}_u^{-1} \mathcal{A}_F)\widetilde{G}, \tag{75}$$

where $\widetilde{G}$ is an independent standard tensor. This follows first for a fixed linear map by orthogonal Gaussian decomposition. For causal queries, condition successively on each next response: the next query is already fixed, and the accumulated constraints are precisely the stacked linear constraints in (75).

D.4.3 Direct source state evolution

Recursion and population law. Define the finite recursion by

$$f_i^t = \phi_{s(i),t}(w_i^{1:t}, e_i), \tag{76}$$

$$w_i^{t+1} = h_d(f^t)_i - \sum_{r<t} b_{tr,s(i)} f_i^r. \tag{77}$$

The coefficients below are deterministic population coefficients.

For each species let $E_s \sim \mu_s$. Construct centered Gaussian $W_s^{1:T+1}$ independent of E_s, together with

$$F_s^t = \phi_{s,t}(W_s^{1:t}, E_s), \qquad Q_s^{tr} = \mathbb{E}[F_s^t F_s^r], \tag{78}$$

$$\mathbb{E}[W_s^{t+1} W_s^{r+1}] = \xi_s(Q^{tr}), \qquad Q^{tr} = (Q_v^{tr})_v, \tag{79}$$

$$j_{tr,v} = \mathbb{E}[\partial_{w^{r+1}} \phi_{v,t}(W_v^{1:t}, E_v)], \qquad b_{tr,s} = \sum_v \partial_v \xi_s(Q^{tr}) j_{tr,v}, \quad r < t. \tag{80}$$

For $r \geq t$, set $j_{tr,v} = 0$. These equations are causal: F^t uses only $W^{1:t}$, and its overlaps determine the enlarged covariance. That covariance is positive semidefinite because ξ_s is a nonnegative sum of Schur products of query Gram matrices. Linear growth ensures all displayed second moments and coefficients are finite at every fixed time. This constructs a unique consistent sequence of finite-horizon laws, including singular covariances.

Proposition D.6 (Positive-amplitude source SE with nondegenerate query histories)**.** *Assume $\gamma_- > 0$ and the map and covariate conditions stated above. For the fixed horizon, suppose that the limiting query-history Gram matrix $(Q_s^{tr})_{0 \leq t,r \leq T}$ is positive definite for each species. Then, for every species,*

$$\frac{1}{d_s} \sum_{i \in V_{s,d}} \delta_{(e_i, w_i^{1:T+1})} \xrightarrow{\mathbb{P}} \mathrm{Law}(E_s, W_s^{1:T+1}) \quad \text{in } \mathsf{W}_2. \tag{81}$$

The conclusion also includes any fixed collection of queries, by Lipschitz mapping. In particular it holds for all pseudo-Lipschitz order-two tests.

Proof. We give the conditioning induction directly. Let $\mathcal{F}_t$ be the sigma-field generated by the static covariates and $w^1, \ldots, w^t$. All independently sampled external coordinates are included when present. All estimates are at a fixed $t \leq T$. On events whose probabilities tend to one, all finite history Grams and their inverses used below have fixed bounds. Auxiliary inverses can be defined arbitrarily on the complement. The compact recursion (77) itself uses no such inverses.

Step 1: Query-Gram and inverse bounds. Assume the joint row-law conclusion through $w^{1:t}$. It gives convergence of the empirical overlaps of $f^0, \ldots, f^t$, since each is a fixed Lipschitz function of this history and the covariate. Thus the empirical query Grams converge to their positive-definite limits. In particular, the full query matrix $d^{-1} F^\top F$ has a positive lower bound before observing w^{t+1}. Lemmas D.3 and D.5 therefore give the bounds on $\mathcal{M}_t^{-1}$ and $\mathcal{D}_t^{-1}$ required by (75). Upper bounds hold for the enlarged stack $f^0, \ldots, f^t$ as well.

Step 2: Integration-by-parts identification. Write $W_t = [w^1|\cdots|w^t]$ and

$$O_t^\ell = \sum_{r<\ell} b_{\ell r} \diamond f^r, \qquad Y_t = W_t + O_t, \qquad 0 \le \ell < t,$$

where $(b_{\ell r} \diamond f^r)_i = b_{\ell r, s(i)} f_i^r$. Thus a species scalar acts on every row of that species. In species v,

$$Z_v^r = \sum_{a<t} (D_{v,d}^{-1})_{ra} w_v^{a+1},$$

where $Z = \mathcal{D}_t^{-1} W_t$. Consequently,

$$R_{v,d}(f^\ell, Z^r) = \sum_{a<t} (D_{v,d}^{-1})_{ra} \frac{1}{d_v} \sum_{i \in V_{v,d}} \phi_{v,\ell}(w_i^{1:\ell}, e_i) w_i^{a+1}.$$

Each summand is a pseudo-Lipschitz function of order two of the row history and covariate. The induction row law therefore gives convergence of the empirical averages, while $D_{v,d}^{-1} \xrightarrow{\mathbb{P}} D_v^{-1}$ follows from query-Gram convergence and its positive-definite limit. Gaussian integration by parts conditional on E_v then gives the second equality below:

$$\begin{aligned} R_{v,d}(f^\ell, Z^r) &\xrightarrow{\mathbb{P}} \sum_{a<t} (D_v^{-1})_{ra} \mathbb{E}[F_v^\ell W_v^{a+1}] \\ &= \sum_{a<t} (D_v^{-1})_{ra} \sum_{b<t} D_v[a,b] \mathbb{E}[\partial_{w^{b+1}} \phi_{v,\ell}(W_v^{1:\ell}, E_v)] \\ &= j_{\ell r, v}, \qquad 0 \le \ell \le t, \quad r < t, \end{aligned} \tag{82}$$

where $D_v[a,b] = \xi_v(Q^{ab})$. The derivative is zero for $b \ge \ell$. The bounded history gradient and finite second moments justify integration by parts, including nonzero query means.

Inserting (82) into (71), for $\ell < t$, gives

$$\left\| \mathcal{B}_t \mathcal{D}_t^{-1} W_t - O_t \right\|_d \xrightarrow{\mathbb{P}} 0. \tag{83}$$

There are only finitely many terms, the scalar coefficient errors tend to zero, and all normalized query norms are tight. Since $\mathcal{M}_t = \mathcal{D}_t + \mathcal{B}_t$, its uniformly bounded inverse now implies

$$\left\| \mathcal{M}_t^{-1} Y_t - \mathcal{D}_t^{-1} W_t \right\|_d \xrightarrow{\mathbb{P}} 0. \tag{84}$$

This is the normal-equation identity required by the induction.

Step 3: Conditional mean and Onsager cancellation. For the enlarged frozen query list, decompose the cross-covariance from the past responses to the new raw response as

$$\mathcal{M}_{t,<} = \mathcal{D}_{t,<} + \mathcal{B}_{t,<}.$$

The diagonal part acts specieswise by the vector

$$g_{s,d}[r] = \xi_{s,d}(Q_d^{tr}), \qquad r < t.$$

The finite-rank part is

$$(\mathcal{B}_{t,<} Z)_s = \sum_{r<t} \sum_v f_s^r \, \partial_v \xi_{s,d}(Q_d^{tr}) R_{v,d}(f^t, Z^r). \tag{85}$$

The enlarged covariance has bounded operator norm, so (84) can be multiplied by $\mathcal{M}_{t,<}$. Applying (82) with $\ell = t$ gives

$$\mathbb{E}[h_d(f^t)_s \mid \mathcal{F}_t] = \sum_{r<t} \beta_{sr} w_s^{r+1} + \sum_{r<t} b_{tr,s} f_s^r + o_{\|\cdot\|_d}(1), \tag{86}$$

where $\beta_s = D_s^{-1} g_s$ and $g_s[r] = \xi_s(Q^{tr})$. The notation $o_{\|\cdot\|_d}(1)$ means a vector whose normalized norm tends to zero in probability. Equation (77) cancels the second sum exactly, leaving Gaussian-history regression as the conditional mean.

Step 4: Conditional-covariance comparison. The exact covariance of the new raw innovation is

$$\Sigma_{t,d} = \mathcal{M}_{tt} - \mathcal{M}_{t,<}\mathcal{M}_{t}^{-1}\mathcal{M}_{<,t}. \tag{87}$$

Define the diagonal, species-constant operator $V_{t,d}$ by the Schur complements

$$v_{s,d} = \xi_{s,d}(Q_d^{tt}) - g_{s,d}^{\top} D_{s,d}^{-1} g_{s,d} \geq 0.$$

The inequality holds because the corresponding enlarged response Gram is positive semidefinite. The enlarged query subspace is

$$\mathcal{V}_{t,d} = \bigoplus_s \operatorname{span}\{f_s^0, \dots, f_s^t\} \subseteq \mathbb{R}^d, \qquad \dim \mathcal{V}_{t,d} \leq |\mathcal{S}|(t+1).$$

Introduce the stacked old-query subspace

$$\mathcal{U}_{t,d} = \left\{ Z \in \mathbb{R}^{d\times t} : Z_s^r \in \operatorname{span}\{f_s^0, \dots, f_s^{t-1}\} \text{ for every } s \text{ and } r < t \right\}.$$

The operator $\mathcal{D}_t$ preserves $\mathcal{U}_{t,d}$ and its orthogonal complement, whereas $\mathcal{B}_t$ vanishes on $\mathcal{U}_{t,d}^{\perp}$. Hence

$$\mathcal{M}_t|_{\mathcal{U}_{t,d}^{\perp}} = \mathcal{D}_t|_{\mathcal{U}_{t,d}^{\perp}}, \qquad \mathcal{M}_t^{-1}\Big|_{\mathcal{U}_{t,d}^{\perp}} = \mathcal{D}_t^{-1}\Big|_{\mathcal{U}_{t,d}^{\perp}}.$$

For $z \perp \mathcal{V}_{t,d}$, the finite-rank parts in (69) vanish: $\mathcal{B}_{tt}z = 0$ and $\mathcal{B}_{<,t}z = 0$. Moreover, $\mathcal{D}_{<,t}z \in \mathcal{U}_{t,d}^{\perp}$, so the displayed inverse identity applies, and $\mathcal{B}_{t,<}\mathcal{D}_t^{-1}\mathcal{D}_{<,t}z = 0$ because $z_s \perp f_s^t$. Therefore

$$\Sigma_{t,d}\big|_{\mathcal{V}_{t,d}^{\perp}} = V_{t,d}\big|_{\mathcal{V}_{t,d}^{\perp}}.$$

Both operators are self-adjoint and preserve this orthogonal decomposition. They are positive semidefinite and have bounded operator norm. Their principal square roots coincide on $\mathcal{V}_{t,d}^{\perp}$, so, on the localization event,

$$\left\| \Sigma_{t,d}^{1/2} - V_{t,d}^{1/2} \right\|_F^2 \leq C \dim \mathcal{V}_{t,d} \leq C|\mathcal{S}|(t+1). \tag{88}$$

Use a new standard Gaussian vector g^t independent of the history in the conditional representation. Equation (88), divided by d, proves that the innovation is

$$(v_{s,d}^{1/2} g_s^t)_s + o_{\|\cdot\|_d}(1).$$

The finite Gram convergence replaces $v_{s,d}$ by

$$v_s = \xi_s(Q^{tt}) - g_s^{\top} D_s^{-1} g_s.$$

Combining with (86) yields

$$w_s^{t+1} = \sum_{r<t} \beta_{sr} w_s^{r+1} + \sqrt{v_s}\, g_s^t + o_{\|\cdot\|_d}(1). \tag{89}$$

Step 5: Empirical convergence and induction closure. Conditional on the history, the new g_i^t are independent. For bounded Lipschitz tests, the variance of their species row average is $O(d_s^{-1})$. Integrating over the new scalar Gaussian converts the test into a bounded Lipschitz function of the old history and covariate, to which the induction hypothesis applies. Second moments converge separately: expand the squared norm in (89), use the Gaussian law of large numbers, and note that the variance of the normalized cross term with the old history is $O(d_s^{-1})$ on bounded normalized histories. The error term changes the second moment by $o_{\mathrm{P}}(1)$ by Cauchy–Schwarz. Hence joint weak convergence and convergence of second moments hold, which is equivalent to Wasserstein–2 convergence in fixed dimension. This also justifies all PL(2) moments in the next step.

At $t = 0$, the mean is zero, the past stack is empty, and the same finite-rank covariance argument gives $w_s^1 = \sqrt{\xi_s(Q^{00})}\, g_s^0 + o_{\|\cdot\|_d}(1)$. Together with (65), this starts the induction. The regression coefficients and innovation variance in (89) are precisely those of (79). Since each fresh Gaussian is independent of the covariates, the limiting Gaussian history is independent of E_s. This proves (81). □

D.5 SAME-DEGREE REGULARIZATION AND DEGENERACY REMOVAL

Lemma D.7 (Wasserstein–2 convergence under independent covariate extension)**.** *Suppose (65) holds and η_i are i.i.d. with a fixed finite-second-moment law ρ, independent of everything else. Then the empirical law of (e_i, η_i) converges in probability in W_2 to $\mu_s \otimes \rho$ in each species.*

Proof. For a bounded Lipschitz test, the conditional variance of its average is $O(d_s^{-1})$ and its conditional mean converges by weak convergence of the deterministic covariate law. The empirical second moment of (e_i, η_i) splits into those of e_i and η_i. The former converges by the assumed Wasserstein–2 convergence, and the latter by the law of large numbers. The fixed-dimensional weak-plus-second-moment criterion gives the result. □

D.5.1 SAME-DEGREE FILL AND QUERY DITHER

Let Γ_0 be the possibly sparse target amplitude tensor. Fix $\tau > 0$ and introduce a fresh independent ordered standard Gaussian tensor $\widetilde{G}$ of the *same degree* N:

$$H_d^\epsilon(x) = H_d^0(x) + \sqrt{\epsilon\tau}\, d^{-(N-1)/2} \left\langle \widetilde{G}, x^{\otimes N} \right\rangle. \tag{90}$$

The covariance polynomial is

$$\xi_d^\epsilon(q) = \xi_d^0(q) + \epsilon\tau \left(\sum_s \lambda_{s,d} q_s \right)^N. \tag{91}$$

At every fixed $\epsilon > 0$, this is equal in law to a model with amplitudes

$$\gamma_{\boldsymbol{s}}^\epsilon = \sqrt{(\gamma_{\boldsymbol{s}}^0)^2 + \epsilon\tau}, \qquad \gamma_-^\epsilon \geq \sqrt{\epsilon\tau} > 0.$$

The filled model therefore remains pure degree N.

Independently adjoin one standard Gaussian coordinate η_i^t per query time and set

$$\phi_{s,t}^\delta(w^{1:t}, e, \eta) = \phi_{s,t}(w^{1:t}, e) + \sqrt{\delta}\, \eta^t, \qquad \delta > 0. \tag{92}$$

For each (ϵ, δ), the source SE moments and deterministic coefficients (80) are recomputed for the corresponding regularized model.

D.5.2 FIXED-REGULARIZATION ANALYSIS

By construction, the regularized population Gaussian field history is independent of the enlarged covariate (E_s, η_s). The undithered summands are functions of (W_s, E_s) alone. Hence their query Gram obeys the exact population identity

$$\mathbf{P}_s^{\epsilon,\delta} = \overline{\mathbf{P}}_s^{\epsilon,\delta} + \delta I \succeq \delta I, \tag{93}$$

where $\overline{\mathbf{P}}_s^{\epsilon,\delta}$ denotes the population Gram of the original maps evaluated along the same regularized Gaussian history. Lemma D.7 supplies the empirical law for the enlarged covariates. For random enlarged covariates, take an almost-sure Wasserstein–2 convergent further subsequence and apply Proposition D.6 conditionally on those arrays. Bounded convergence of the conditional failure probabilities then returns convergence in probability. At every fixed $\epsilon, \delta > 0$, the time-t row law yields the empirical query-Gram limit before the inverse required to analyze time $t + 1$ is formed.

D.5.3 POPULATION CONTINUITY AND UNIFORM MOMENT LOCALIZATION

For the fixed horizon, the population second moments of the regularized queries and fields are bounded uniformly for $0 \leq \epsilon, \delta \leq 1$. Uniform linear growth and the polynomial dependence of field variances on preceding query moments give these moment bounds by induction. The bounded history derivatives then yield uniform bounds on the population Onsager coefficients.

As $(\epsilon, \delta) \to (0, 0)$, the corresponding finite collection of population moments and Onsager coefficients converges to its target. To see this, couple the original and regularized covariates using the same

(E_s, η_s), and at each horizon couple the *whole stacked* Gaussian field vectors by their principal covariance square roots applied to a common standard Gaussian vector. Covariance square roots are continuous on the positive-semidefinite cone, including at singular matrices. The Lipschitz map bounds give L^2 convergence of the next query, and the bounded Lipschitz history derivatives give convergence of the average derivatives. Polynomial covariance and Onsager formulas close the induction.

For the finite paths, a bounded-amplitude ordered Gaussian tensor has injective norm $O_{\mathsf{P}}(\sqrt{d})$: a product of fixed-resolution sphere nets has size $\exp(O(d))$, and the variance of the contraction at each net point is at most the squared maximum amplitude. A union bound and multilinear telescoping give the claim, just as in Lemma C.2. The event can be chosen uniformly for $0 \le \epsilon \le 1$. Indeed, under the common coupling,

$$\sup_{0\le\epsilon\le1} \left\| \Gamma_0 \odot G + \sqrt{\epsilon\tau}\, \widetilde{G} \right\|_{\mathrm{inj}} \le \|\Gamma_0 \odot G\|_{\mathrm{inj}} + \sqrt{\tau} \left\| \widetilde{G} \right\|_{\mathrm{inj}} = O_{\mathsf{P}}(\sqrt{d}).$$

Consequently, for every fixed R,

$$\left\| h_d^0(x) - h_d^0(y) \right\|_d \le C_R \|x - y\|_d\,, \qquad \left\| h_d^{\mathrm{fill}}(x) \right\|_d \le C_R, \tag{94}$$

simultaneously on $\|x\|_d, \|y\|_d \le R$, with probability tending to one. These events apply to adaptive query vectors. The original covariate norms, dither norms, and the operator bound, together with the compact recursion and the already bounded deterministic SE coefficients, yield tight finite-path norms, uniformly for small regularizations, by fixed-horizon induction.

D.5.4 Finite-path coupling and the limiting source theorem

Couple all finite recursions with the same $G, \widetilde{G}, e, \eta$. On a common localization event, (90), (92), (94), and the bounded deterministic coefficients give

$$\Delta_{t+1} \le C_{T,R} \left\{ \Delta_t + \sqrt{\epsilon} + \sqrt{\delta} + \rho_T(\epsilon, \delta) \right\}, \tag{95}$$

where Δ_t is the maximum normalized discrepancy of the corrected histories through time t and of all queries already formed, and $\rho_T(\epsilon, \delta) \to 0$ bounds the differences of the finitely many deterministic Onsager coefficients. The initial query discrepancy is bounded by $\sqrt{\delta}$ times the normalized dither norm. Fixed-horizon induction, followed by removal of the localization event, gives for every $\zeta > 0$

$$\lim_{\epsilon\downarrow0} \lim_{\delta\downarrow0} \limsup_{d\to\infty} \mathbb{P}(\Delta_{T+1} > \zeta) = 0. \tag{96}$$

Because the constants in this comparison depend only on the compact AMP recursion, they remain uniform as ϵ or δ vanishes.

Write $\widehat{\mu}_{s,d}^{\epsilon,\delta}$ and $\mu_s^{\epsilon,\delta}$ for the empirical and population laws of the original covariate and the corrected field history, and project out the auxiliary dither coordinates. The triangle inequality gives

$$\mathsf{W}_2(\widehat{\mu}_{s,d}^{0,0}, \mu_s^{0,0}) \le \mathsf{W}_2(\widehat{\mu}_{s,d}^{0,0}, \widehat{\mu}_{s,d}^{\epsilon,\delta}) + \mathsf{W}_2(\widehat{\mu}_{s,d}^{\epsilon,\delta}, \mu_s^{\epsilon,\delta}) + \mathsf{W}_2(\mu_s^{\epsilon,\delta}, \mu_s^{0,0}).$$

Pairing identical row indices bounds the first term by a constant times the coupled normalized history discrepancy. The second tends to zero at each fixed regularization by the proposition, and the third tends to zero by the population L^2 coupling. Equation (96) therefore transfers Wasserstein–2 convergence in the stated nested order. For a PL(2) test, Cauchy–Schwarz bounds the difference of expectations by the Wasserstein–2 coupling discrepancy times a bounded second-moment factor. The PL(2) conclusion follows as well.

Theorem D.8 (Pure-degree source state evolution). *For fixed $N \ge 3$, a fixed finite species set with deterministic proportions $\lambda_{s,d} \to \lambda_s > 0$, nonnegative fixed amplitudes Γ_0, and the maps and deterministic row covariates specified above, the compact recursion (76)–(77) has the joint specieswise row law (78)–(80) at every fixed horizon, in the sense of (81). No nondegeneracy condition is imposed on the target query or field covariance histories. The same conclusion holds for random row-covariate arrays independent of the Gaussian disorder if their empirical Wasserstein–2 convergence to deterministic laws holds in probability.*

Proof. The preceding coupling argument establishes the deterministic conclusion in the nested limit order of (96). For random covariates, condition on their complete arrays while leaving the original Gaussian disorder, Gaussian fill, and query dither unconditioned. Every subsequence has a further subsequence on which the empirical Wasserstein–2 convergence holds almost surely. For almost every realized sequence, the deterministic theorem applies. The conditional failure probabilities tend to zero and are bounded by one, so bounded convergence integrates them. The subsequence criterion gives convergence in probability for the original sequence. □

D.6 Identification of the fixed-interaction oracle

Apply Theorem D.8 to the embedding and channel schedule in Sections D.1–D.3. The active degree is N, and the base species are the native modes. With $d = L_I = \sum_n I_n$, the target source amplitudes are g on permutations of the mode tuple $(1, \dots, N)$ and zero elsewhere. In the limit,

$$\xi^{\mathrm{emb}}(q) = N! g^2 \prod_m \lambda_m q_m, \qquad \xi_n(q) = \alpha_n \prod_{m \neq n} q_m, \qquad \alpha_n = (\nu c_n)^{-1}.$$

These identities follow from (49)–(50).

Set the static covariate in mode n to $e_{n,i_n} = ([\mathbf{A}_n]_{i_n,\cdot}, [\mathbf{X}_n^0]_{i_n,\cdot})$ and first condition on these arrays. Assumption 4.2 supplies the deterministic empirical Wasserstein–2 hypotheses along almost-sure subsequences, while conditioning leaves the law of the Gaussian disorder tensor unchanged. The finite list of scalar query maps returns the channel columns of $\mathbf{X}_n^t$. The deterministic population controls $\mathbf{\Theta}_n^t$ and $\mathbf{K}_n^{t,t}$ from the candidate SE are held fixed inside these maps. The row-map assumptions give the joint history/covariate Lipschitz conditions and bounded Lipschitz history derivatives required above.

Throughout the auxiliary fill/dither construction, these original deterministic oracle controls remain fixed, while the source SE and source Onsager coefficients are recomputed for the perturbed scalar-query system. The intermediate source paths serve as auxiliary regularizations, and the zero-regularization endpoint is identified with the native vector oracle.

At the zero-regularization endpoint, applying Theorem D.8 to the finite query list gives, for the core-weighted Gaussian limits,

$$\begin{aligned} \mathbb{E}[u^s_{n,p_n} u^t_{n,r_n}] &= \alpha_n \sum_{\boldsymbol{p}_{\setminus n}, \boldsymbol{r}_{\setminus n}} \boldsymbol{\mathcal{C}}_{p_n, \boldsymbol{p}_{\setminus n}} \boldsymbol{\mathcal{C}}_{r_n, \boldsymbol{r}_{\setminus n}} \prod_{m \neq n} \mathbf{Q}_m^{s,t}[p_m, r_m] \\ &= [\mathcal{K}_n(\mathbf{Q}^{s,t})]_{p_n, r_n}. \end{aligned} \tag{97}$$

At the same endpoint, scalar derivatives and the source Onsager formula give, for current macro channel $\boldsymbol{r}$ and previous macro channel $\boldsymbol{p}$,

$$b_{u(t,\boldsymbol{r}),u(t-1,\boldsymbol{p}),n} = \alpha_n \sum_{m \neq n} \boldsymbol{\mathcal{C}}_{\boldsymbol{p}} \overline{\mathbf{J}}^t_m[p_m, r_m] \prod_{k \neq n, m} \mathbf{Q}_k^{t-1,t}[p_k, r_k]. \tag{98}$$

Same-block and earlier-than-previous-block derivatives vanish because of the fixed source schedule.

Weighting the current source Onsager coefficients by $\boldsymbol{\mathcal{C}}_{r_n, \boldsymbol{r}_{\setminus n}}$ recovers the native reaction matrix

$$\mathbf{\Omega}_n^t = \sum_{m \neq n} D_m \mathcal{K}_n(\mathbf{Q}^{t-1,t})[\overline{\mathbf{J}}^t_m].$$

This is the same finite channel-sum calculation as Lemma D.1. The resulting effective native field is $\boldsymbol{a}_n \mathbf{\Theta}_n^t + \boldsymbol{u}_n^t$. By induction in macro time, its deterministic source law agrees with the candidate SE: the initial pairs agree, the covariance agrees by the core contraction, and the row denoiser with its fixed covariance input produces the same next pair of moments.

Theorem D.8, followed by the fixed linear channel reconstruction and Lipschitz denoising maps, therefore proves the joint row law for the fixed-interaction oracle of Appendix E. Conditioning then restores random planted–initial arrays as in the last part of Theorem D.8.

E ORACLE AMP AND EMPIRICAL-COEFFICIENT TRANSFER

E.1 FIXED-INTERACTION AND FINITE ORACLES

The source analysis first considers the fixed interaction in (47). Replace every empirical coefficient by its deterministic state-evolution value and use the coupled operator (48):

$$\mathbf{B}_n^{\circ,t} = \mathbf{A}_n \mathbf{\Theta}_n^t + \mathcal{H}_{n,I}^{\circ}(\mathbf{X}^{\circ,t}) - \mathbf{X}_n^{\circ,t-1}\mathbf{\Omega}_n^t, \tag{99}$$

$$\mathbf{X}_n^{\circ,t+1} = f_{n,t}(\mathbf{B}_n^{\circ,t}, \mathbf{K}_n^{t,t}). \tag{100}$$

Theorem D.8 and the identification in Appendix D.6 give the empirical row law (14)–(17) for this recursion. The limiting side information is $(\boldsymbol{a}_n, \boldsymbol{x}_n^0)$. Appendix D handles its empirical Wasserstein–2 limit directly.

The finite oracle retains these deterministic coefficients but restores the Gaussian operator used by the actual algorithm:

$$\overline{\mathbf{B}}_n^t = \mathbf{A}_n \mathbf{\Theta}_n^t + \mathcal{H}_{n,I}(\overline{\mathbf{X}}^t) - \overline{\mathbf{X}}_n^{t-1}\mathbf{\Omega}_n^t, \tag{101}$$

$$\overline{\mathbf{X}}_n^{t+1} = f_{n,t}(\overline{\mathbf{B}}_n^t, \mathbf{K}_n^{t,t}). \tag{102}$$

Both oracle paths use the initialization of the actual recursion: $\mathbf{X}^{\circ,0} = \overline{\mathbf{X}}^0 = \mathbf{X}^0$ and $\mathbf{X}^{\circ,-1} = \overline{\mathbf{X}}^{-1} = 0$.

E.2 MOMENT LOCALIZATION

The actual AMP coefficients are trajectory-dependent, so their boundedness must be proved together with path localization. Cauchy–Schwarz gives

$$\left\|\mathbf{Q}_{n,I}^{s,t}\right\|_F \le \|\mathbf{X}_n^s\|_n \left\|\mathbf{X}_n^t\right\|_n, \qquad \left\|\mathbf{M}_{n,I}^t\right\|_F \le \|\mathbf{A}_n\|_n \left\|\mathbf{X}_n^t\right\|_n. \tag{103}$$

Because the core and channel dimensions are fixed and $\sup_I \alpha_{n,I} < \infty$, the explicit polynomial (10) also gives

$$\left\|\mathcal{K}_{n,I}(\mathbf{E})\right\|_F \le C \prod_{m \ne n} (1 + \|\mathbf{E}_m\|_F), \tag{104}$$

$$\left\|\mathbf{\Omega}_{n,I}^t\right\|_F \le C \sum_{m \ne n} \left\|\overline{\mathbf{J}}_{m,I}^t\right\|_F \prod_{k \ne n,m} (1 + \left\|\mathbf{Q}_{k,I}^{t-1,t}\right\|_F). \tag{105}$$

On a bounded iterate history, (103)–(104) bound every covariance input. The Jacobian bound (30) then bounds $\overline{\mathbf{J}}_{m,I}^t$, and (105) bounds the empirical reaction matrix.

The finite-system factorization in Proposition 4.1 rewrites the actual field as

$$\mathbf{B}_n^t = \mathbf{A}_n \mathcal{K}_{n,I}(\mathbf{M}_I^t) + \mathcal{H}_{n,I}(\mathbf{X}^t) - \mathbf{X}_n^{t-1}\mathbf{\Omega}_{n,I}^t. \tag{106}$$

Assumption 4.2 gives $\max_n\{\|\mathbf{A}_n\|_n + \left\|\mathbf{X}_n^0\right\|_n\} = O_{\mathbb{P}}(1)$. Suppose inductively that the iterate histories through time t are tight. For any prescribed error probability, choose a deterministic R that bounds these histories and the planted arrays outside an event of that probability. On this event, the simultaneous Gaussian bound (39), applied to $\mathbf{X}^t$ and zero, and (103)–(106) give $\max_n \left\|\mathbf{B}_n^t\right\|_n \le C_{T,R}$ with probability tending to one. The row-map linear-growth bound then gives

$$\left\|\mathbf{X}_n^{t+1}\right\|_n \le L_{T,C_{T,R}}(1 + \left\|\mathbf{B}_n^t\right\|_n).$$

This closes the induction for the actual AMP using the finite-system coefficient bounds established above. Separately, the population recursion and the same linear-growth bounds give finite deterministic SE coefficients through the fixed horizon. The preceding induction therefore also applies to the finite and fixed-interaction oracles. Moreover, g/g_I is bounded for all large I. Consequently, for each fixed T,

$$\max_{\substack{n \le N \\ t \le T+1}} \{\left\|\mathbf{X}_n^t\right\|_n + \left\|\overline{\mathbf{X}}_n^t\right\|_n + \left\|\mathbf{X}_n^{\circ,t}\right\|_n\} + \max_{\substack{n \le N \\ t \le T}} \{\left\|\mathbf{B}_n^t\right\|_n + \left\|\overline{\mathbf{B}}_n^t\right\|_n + \left\|\mathbf{B}_n^{\circ,t}\right\|_n\} = O_{\mathbb{P}}(1). \tag{107}$$

Equation (107) is therefore a pathwise fixed-horizon estimate established independently of the empirical state-evolution conclusion.

E.3 Interaction-scale transfer

Couple the two oracle recursions through the same symmetric Gaussian tensor and initialization. Put

$$\delta_I = \left| \frac{g}{g_I} - 1 \right|, \tag{108}$$

which tends to zero by (45) and (47). On the localization event, (39) bounds the operator-scale discrepancy by $C_T \delta_I$. With

$$d_t^{\mathrm{coef}} = \max_n \left\| \overline{\mathbf{X}}_n^t - \mathbf{X}_n^{\circ,t} \right\|_n, \qquad e_t^{\mathrm{coef}} = \max_n \left\| \overline{\mathbf{B}}_n^t - \mathbf{B}_n^{\circ,t} \right\|_n,$$

the field equations and row-map bound give

$$e_t^{\mathrm{coef}} \le C_T \{ d_t^{\mathrm{coef}} + d_{t-1}^{\mathrm{coef}} + \delta_I \}, \qquad d_{t+1}^{\mathrm{coef}} \le C_T e_t^{\mathrm{coef}}. \tag{109}$$

Fixed-horizon induction gives convergence of both discrepancies to zero in probability. Moment localization and the pseudo-Lipschitz difference bound therefore transfer the source state evolution from the fixed-interaction oracle to the finite oracle.

E.4 Coupling recurrence

Throughout this subsection, hatted quantities are evaluated on the finite oracle trajectory $(\overline{\mathbf{B}}, \overline{\mathbf{X}})$. In particular, define its overlaps by

$$\widehat{\mathbf{M}}_{n,I}^t = \frac{1}{I_n} \mathbf{A}_n^\top \overline{\mathbf{X}}_n^t, \qquad \widehat{\mathbf{Q}}_{n,I}^{s,t} = \frac{1}{I_n} (\overline{\mathbf{X}}_n^s)^\top \overline{\mathbf{X}}_n^t, \tag{110}$$

Evaluate the signal coefficient, covariance input, average Jacobian, and reaction matrix on that same finite oracle trajectory:

$$\begin{aligned}
\widehat{\mathbf{\Theta}}_{n,I}^t &= \mathcal{K}_{n,I}(\widehat{\mathbf{M}}_I^t), \\
\widehat{\mathbf{K}}_{n,I}^{t,t} &= \mathcal{K}_{n,I}(\widehat{\mathbf{Q}}_I^{t,t}), \\
\widehat{\overline{\mathbf{J}}}_{n,I}^t &= \frac{1}{I_n} \sum_{i_n=1}^{I_n} D_{\boldsymbol{b}} f_{n,t-1}\big([\overline{\mathbf{B}}_n^{t-1}]_{i_n,\cdot}, \mathbf{K}_n^{t-1,t-1}\big), \qquad t \ge 1, \\
\widehat{\mathbf{\Omega}}_{n,I}^t &= \sum_{m \ne n} D_m \mathcal{K}_{n,I}(\widehat{\mathbf{Q}}_I^{t-1,t})[\widehat{\overline{\mathbf{J}}}_{m,I}^t], \qquad t \ge 1.
\end{aligned} \tag{111}$$

Thus $\widehat{\mathbf{\Theta}}$ and $\widehat{\mathbf{K}}$ are the finite-system signal and covariance coefficients formed from finite-oracle overlaps, while $\widehat{\overline{\mathbf{J}}}$ uses the deterministic covariance $\mathbf{K}$ that actually appears in the finite-oracle update (102). Oracle state evolution, the bounded Lipschitz Jacobian test, and $\alpha_{n,I} \to \alpha_n$ imply

$$\begin{aligned}
r_{I,T} := \max_{\substack{n \le N \\ 0 \le t \le T}} \left\| \widehat{\mathbf{M}}_{n,I}^t - \mathbf{M}_n^t \right\| + \max_{\substack{n \le N \\ 0 \le s,t \le T}} \left\| \widehat{\mathbf{Q}}_{n,I}^{s,t} - \mathbf{Q}_n^{s,t} \right\| + \max_{\substack{n \le N \\ 0 \le t \le T}} \left\| \widehat{\mathbf{\Theta}}_{n,I}^t - \mathbf{\Theta}_n^t \right\| \\
+ \max_{\substack{n \le N \\ 0 \le t \le T}} \left\| \widehat{\mathbf{K}}_{n,I}^{t,t} - \mathbf{K}_n^{t,t} \right\| + \max_{\substack{n \le N \\ 1 \le t \le T}} \left\{ \left\| \widehat{\overline{\mathbf{J}}}_{n,I}^t - \overline{\mathbf{J}}_n^t \right\| + \left\| \widehat{\mathbf{\Omega}}_{n,I}^t - \mathbf{\Omega}_n^t \right\| \right\} \xrightarrow{\mathbb{P}} 0.
\end{aligned} \tag{112}$$

Define

$$d_t^{\mathrm{cpl}} = \max_n \left\| \mathbf{X}_n^t - \overline{\mathbf{X}}_n^t \right\|_n, \qquad e_t^{\mathrm{cpl}} = \max_n \left\| \mathbf{B}_n^t - \overline{\mathbf{B}}_n^t \right\|_n. \tag{113}$$

On the localization event, overlap stability, the row-map bounds, Lemma C.1, and (39) control the three field differences separately. The reaction estimate involves discrepancies in both the current and preceding iterates. Uniform moment localization gives

$$\max_m \left\| \mathbf{Q}_{m,I}^{t-1,t} - \widehat{\mathbf{Q}}_{m,I}^{t-1,t} \right\| \le C_T (d_{t-1}^{\mathrm{cpl}} + d_t^{\mathrm{cpl}}), \tag{114}$$

$$\max_m \left\| \mathbf{K}_{m,I}^{t-1,t-1} - \mathbf{K}_m^{t-1,t-1} \right\| \le C_T(d_{t-1}^{\mathrm{cpl}} + r_{I,T}), \tag{115}$$

$$\max_m \left\| \overline{\mathbf{J}}_{m,I}^t - \widehat{\overline{\mathbf{J}}}_{m,I}^t \right\| \le C_T(e_{t-1}^{\mathrm{cpl}} + d_{t-1}^{\mathrm{cpl}} + r_{I,T}). \tag{116}$$

The last line uses the Lipschitz dependence of the row Jacobian on both its field and covariance arguments. Applying the polynomial derivative bound in Lemma C.1 to (12) and (111) now gives

$$\max_n \left\| \mathbf{\Omega}_{n,I}^t - \widehat{\mathbf{\Omega}}_{n,I}^t \right\| \le C_T(d_t^{\mathrm{cpl}} + d_{t-1}^{\mathrm{cpl}} + e_{t-1}^{\mathrm{cpl}} + r_{I,T}). \tag{117}$$

The reaction difference decomposes as

$$\begin{aligned} &\mathbf{X}_n^{t-1}\mathbf{\Omega}_{n,I}^t - \overline{\mathbf{X}}_n^{t-1}\mathbf{\Omega}_n^t \\ &\quad = (\mathbf{X}_n^{t-1} - \overline{\mathbf{X}}_n^{t-1})\mathbf{\Omega}_{n,I}^t + \overline{\mathbf{X}}_n^{t-1}(\mathbf{\Omega}_{n,I}^t - \widehat{\mathbf{\Omega}}_{n,I}^t) + \overline{\mathbf{X}}_n^{t-1}(\widehat{\mathbf{\Omega}}_{n,I}^t - \mathbf{\Omega}_n^t). \end{aligned} \tag{118}$$

Equations (112) and (117) bound its normalized norm by $C_T(d_t^{\mathrm{cpl}} + d_{t-1}^{\mathrm{cpl}} + e_{t-1}^{\mathrm{cpl}} + r_{I,T})$. In particular, d_t^{cpl} enters through the current factor of the mixed-time overlap, and d_{t-1}^{cpl} enters through both its previous factor and the preceding-iterate multiplier.

The signal term is at most $C_T(d_t^{\mathrm{cpl}} + r_{I,T})$ after multiplication by the tight normalized norm of $\mathbf{A}_n$, and the Gaussian-operator term is at most $C_T d_t^{\mathrm{cpl}}$. Applying the row-map bound after these three estimates gives

$$\begin{aligned} e_t^{\mathrm{cpl}} &\le C_T(d_t^{\mathrm{cpl}} + d_{t-1}^{\mathrm{cpl}} + e_{t-1}^{\mathrm{cpl}} + r_{I,T}), \\ d_{t+1}^{\mathrm{cpl}} &\le C_T(e_t^{\mathrm{cpl}} + d_t^{\mathrm{cpl}} + r_{I,T}), \end{aligned} \tag{119}$$

with $d_{-1}^{\mathrm{cpl}} = e_{-1}^{\mathrm{cpl}} = 0$. Since the coupled initializations agree, fixed-horizon induction gives

$$\max_{t \le T+1} d_t^{\mathrm{cpl}} + \max_{t \le T} e_t^{\mathrm{cpl}} \xrightarrow{\mathbb{P}} 0. \tag{120}$$

For a pseudo-Lipschitz order-2 test, Cauchy–Schwarz bounds the difference between the two empirical averages by the normalized history discrepancy times a tight second-moment factor. Equations (107) and (120) prove Theorem 4.3, including (20) and (21).

F Representation Consequences and Frozen-Map Selection

This appendix gives the moment formulas and guarantees summarized in Section 5.1, followed by the frozen-map, finite-family, and terminal-readout consequences of the joint row law. All channel dimensions and the horizon remain fixed.

F.1 Moment-based statistics

Fix a mode n and define the centered population moments

$$\begin{aligned} \boldsymbol{\mu}_{a,n} &= \mathbb{E}\boldsymbol{a}_n, & \boldsymbol{\mu}_{x,n}^t &= \mathbb{E}\boldsymbol{x}_n^t, \\ \mathbf{\Sigma}_{aa,n} &= \mathbb{E}[(\boldsymbol{a}_n - \boldsymbol{\mu}_{a,n})^\top(\boldsymbol{a}_n - \boldsymbol{\mu}_{a,n})], & \mathbf{\Sigma}_{xx,n}^t &= \mathbb{E}[(\boldsymbol{x}_n^t - \boldsymbol{\mu}_{x,n}^t)^\top(\boldsymbol{x}_n^t - \boldsymbol{\mu}_{x,n}^t)], \\ \mathbf{\Sigma}_{ax,n}^t &= \mathbb{E}[(\boldsymbol{a}_n - \boldsymbol{\mu}_{a,n})^\top(\boldsymbol{x}_n^t - \boldsymbol{\mu}_{x,n}^t)]. \end{aligned} \tag{121}$$

Hats denote empirical row-average quantities formed with normalization I_n^{-1} and empirical row means.

Lemma F.1 (Centered-moment transfer)**.** *Under the assumptions of Theorem 4.3, jointly for every fixed $t \le T+1$,*

$$(\widehat{\boldsymbol{\mu}}_{a,n}, \widehat{\boldsymbol{\mu}}_{x,n}^t, \widehat{\mathbf{\Sigma}}_{aa,n}, \widehat{\mathbf{\Sigma}}_{xx,n}^t, \widehat{\mathbf{\Sigma}}_{ax,n}^t) \xrightarrow{\mathbb{P}} (\boldsymbol{\mu}_{a,n}, \boldsymbol{\mu}_{x,n}^t, \mathbf{\Sigma}_{aa,n}, \mathbf{\Sigma}_{xx,n}^t, \mathbf{\Sigma}_{ax,n}^t). \tag{122}$$

Proof. Coordinate projections are Lipschitz and coordinate products are pseudo-Lipschitz of order two. Equation (19) therefore gives convergence of every entry of the row means and uncentered second moments. For example,

$$\widehat{\mathbf{\Sigma}}^t_{ax,n} = I_n^{-1}\mathbf{A}_n^\top \mathbf{X}_n^t - (\widehat{\boldsymbol{\mu}}_{a,n})^\top \widehat{\boldsymbol{\mu}}^t_{x,n},$$
$$\widehat{\mathbf{\Sigma}}^t_{xx,n} = I_n^{-1}(\mathbf{X}_n^t)^\top \mathbf{X}_n^t - (\widehat{\boldsymbol{\mu}}^t_{x,n})^\top \widehat{\boldsymbol{\mu}}^t_{x,n}.$$

Finite dimensionality and the continuous mapping theorem give (122). □

Corollary F.2 (Canonical correlations and population affine-probe risk)**.** *Under the assumptions of Theorem 4.3, the following conclusions hold for every fixed* $0 \le t \le T+1$.

1. *If* $\mathbf{\Sigma}_{aa,n} \succeq \lambda \mathbf{I}_{P_n}$ *and* $\mathbf{\Sigma}^t_{xx,n} \succeq \lambda \mathbf{I}_{P_n}$ *for some* $\lambda > 0$*, every empirical canonical correlation converges to the corresponding singular value of* $\mathbf{\Sigma}_{aa,n}^{-1/2}\mathbf{\Sigma}^t_{ax,n}(\mathbf{\Sigma}^t_{xx,n})^{-1/2}$.

2. *Fix* $\boldsymbol{w}_\star \in \mathbb{R}^{P_n}$ *and the latent-linear label* $y = (\boldsymbol{a}_n - \boldsymbol{\mu}_{a,n})\boldsymbol{w}_\star + \varepsilon$*, where* $\mathbb{E}[\varepsilon \mid \boldsymbol{a}_n, \boldsymbol{x}_n^t] = 0$ *and* $\mathbb{E}\varepsilon^2 = \sigma_\varepsilon^2 < \infty$*. The minimum population risk over affine predictors of* y *from* $\boldsymbol{x}_n^t$ *is*

$$\sigma_\varepsilon^2 + \boldsymbol{w}_\star^\top \mathbf{\Sigma}_{aa,n}\boldsymbol{w}_\star - \boldsymbol{w}_\star^\top \mathbf{\Sigma}^t_{ax,n}(\mathbf{\Sigma}^t_{xx,n})^\dagger(\mathbf{\Sigma}^t_{ax,n})^\top \boldsymbol{w}_\star, \tag{123}$$

where $\dagger$ *denotes the Moore–Penrose pseudoinverse. If* $\mathbf{\Sigma}^t_{xx,n} \succeq \lambda \mathbf{I}_{P_n}$ *for some* $\lambda > 0$*, the empirical moment plug-in version converges in probability to* (123)*.*

Proof. By Lemma F.1, the empirical centered moments converge jointly to their population counterparts. Under the lower spectral bounds in the CCA statement, both empirical covariance matrices are positive definite with probability tending to one. The inverse square-root map is continuous on the positive-definite cone bounded away from zero. Matrix multiplication and ordered singular values are continuous in fixed dimensions, proving the convergence of the canonical correlations.

For the population-probe claim, put $\boldsymbol{c} = (\mathbf{\Sigma}^t_{ax,n})^\top \boldsymbol{w}_\star$. If $\boldsymbol{v} \in \ker(\mathbf{\Sigma}^t_{xx,n})$, then $(\boldsymbol{x}_n^t - \boldsymbol{\mu}^t_{x,n})\boldsymbol{v} = 0$ almost surely, and hence $\boldsymbol{v}^\top \boldsymbol{c} = 0$. Thus $\boldsymbol{c} \in \mathrm{range}(\mathbf{\Sigma}^t_{xx,n})$. After centering the representation, the population normal equations admit the coefficient

$$(\mathbf{\Sigma}^t_{xx,n})^\dagger(\mathbf{\Sigma}^t_{ax,n})^\top \boldsymbol{w}_\star. \tag{124}$$

The orthogonal-projection identity and the conditional mean-zero assumption on ε give (123). Under the stated positive-definiteness condition, inversion is continuous. The plug-in conclusion follows from Lemma F.1. □

Centering accounts for the potentially nonzero means of the AMP iterates. A sufficient centered special case is invariance of the planted–initial law under a global sign flip together with row maps that are odd in the field input and have no bias.

F.2 Independently selected frozen row maps

Corollary F.3 (Independently selected frozen row maps)**.** *Suppose a finite-dimensional parameter* $\widehat{\theta}$*, independent of the evaluated finite-system sequence and held fixed across ambient sizes, determines jointly Borel measurable row maps satisfying Assumption 4.2 almost surely. Conditional on* $\widehat{\theta}$*, Theorem 4.3 and Corollary F.2 hold. Any statistic* $\widehat{F}_I$ *covered by those results satisfies*

$$\widehat{F}_I - F(\widehat{\theta}) \xrightarrow{\mathbb{P}} 0, \tag{125}$$

where $F(\theta)$ *denotes the corresponding conditional SE limit at parameter value* θ.

Proof of Corollary F.3. For almost every admissible realization θ, take a regular conditional law given $\widehat{\theta} = \theta$ and apply Theorem 4.3 with the resulting deterministic maps. For any $\varepsilon > 0$, the conditional probability

$$\rho_I(\theta) = \mathbb{P}\{\left\|\widehat{F}_I - F(\theta)\right\| > \varepsilon \mid \widehat{\theta} = \theta\} \tag{126}$$

tends to zero for almost every θ and is bounded by one. Dominated convergence gives $\mathbb{E}[\rho_I(\widehat{\theta})] \to 0$, which is exactly (125). The same argument applies to CCA, the probe, or another nondegenerate functional on the almost-sure parameter event where its corresponding condition holds. A spectral lower bound may depend on the parameter realization. □

Corollary F.3 applies to any frozen row maps satisfying Assumption 4.2. The following construction gives a convenient sufficient class of fixed-dimensional neural row maps. Fix a twice continuously differentiable activation $\sigma : \mathbb{R} \to \mathbb{R}$ with bounded first and second derivatives, applied coordinatewise.

For a symmetric matrix $\mathbf{S} \in \mathbb{S}_+^{P_n}$, a representative covariance-aware fixed-width map is

$$f_\theta(\boldsymbol{b}, \mathbf{S}) = \sigma\big(\boldsymbol{b}\mathbf{W}_b + \mathrm{vech}(\mathbf{S})\mathbf{W}_S + \boldsymbol{c}_h\big)\,\mathbf{W}_o + \boldsymbol{c}_o, \tag{127}$$

where $\mathrm{vech}(\mathbf{S}) \in \mathbb{R}^{P_n(P_n+1)/2}$ stacks the lower-triangular entries of $\mathbf{S}$, including the diagonal, in a fixed order. For fixed hidden width h, the weights have dimensions $\mathbf{W}_b \in \mathbb{R}^{P_n \times h}$, $\mathbf{W}_S \in \mathbb{R}^{P_n(P_n+1)/2 \times h}$, and $\mathbf{W}_o \in \mathbb{R}^{h \times P_n}$, with biases $\boldsymbol{c}_h \in \mathbb{R}^h$ and $\boldsymbol{c}_o \in \mathbb{R}^{P_n}$.

Lemma F.4 (Admissibility of fixed-parameter neural row maps)**.** *Suppose $f_{m,t}(\boldsymbol{b}, \mathbf{S}) = g_{m,t}(\boldsymbol{b})$, where $g_{m,t}$ is a fixed finite composition of affine maps and coordinatewise σ, with an optional final affine map, and σ satisfies the conditions above. If all weights and biases are finite and fixed, the map is globally Lipschitz in its row input, continuously differentiable, has bounded Lipschitz input Jacobian, and has at most linear growth. Its covariance Lipschitz constant is zero. It therefore satisfies the row-map conditions in Assumption 4.2. The covariance-aware map (127), with fixed finite parameters and hidden width, satisfies the same conditions while allowing dependence on the covariance input.*

Proof. Bounded σ' makes σ globally Lipschitz, while bounded σ'' makes σ' globally Lipschitz. A fixed finite composition with fixed affine maps is therefore globally Lipschitz and continuously differentiable. The chain rule expresses its Jacobian as a finite product of fixed weights and bounded diagonal derivative matrices. Taking differences of these products and using the Lipschitz continuity of σ' shows that the Jacobian is bounded and Lipschitz. Global Lipschitz continuity also gives at most linear growth.

For a vector $\boldsymbol{v}$, write $\mathrm{diag}(\boldsymbol{v})$ for the diagonal matrix with $\boldsymbol{v}$ on its diagonal. The row-convention input Jacobian of the covariance-aware map is

$$D_{\boldsymbol{b}} f_\theta(\boldsymbol{b}, \mathbf{S}) = \mathbf{W}_b \,\mathrm{diag}\Big(\sigma'\big(\boldsymbol{b}\mathbf{W}_b + \mathrm{vech}(\mathbf{S})\mathbf{W}_S + \boldsymbol{c}_h\big)\Big)\,\mathbf{W}_o. \tag{128}$$

Linearity of vech and the bounds on σ' and σ'' give the required covariance Lipschitz bound and joint Lipschitz continuity of the input Jacobian on bounded covariance sets. They also give at most linear growth and a uniform bound on $f_\theta(0, \mathbf{S})$ over each bounded covariance set. □

The activation $\sigma = \tanh$ satisfies these conditions, and the resulting class contains the neural row maps used in Section 6.

F.3 Finite-family loss convergence and selection

For candidate j, a pseudo-Lipschitz order-two loss ℓ_n on each mode-n row history and nonnegative weights w_n define

$$\widehat{L}_{I,j} = \sum_{n=1}^{N} \frac{w_n}{I_n} \sum_{i_n=1}^{I_n} \ell_n\Big([\mathbf{A}_n]_{i_n,\cdot}, [\mathbf{X}_{n,j}^{0:T+1}]_{i_n,\cdot}, [\mathbf{B}_{n,j}^{0:T}]_{i_n,\cdot}\Big),$$
$$L_j = \sum_{n=1}^{N} w_n \mathbb{E}\ell_n(\boldsymbol{a}_n, \boldsymbol{x}_{n,j}^{0:T+1}, \boldsymbol{b}_{n,j}^{0:T}), \tag{129}$$

where the second line uses the candidate-specific SE. When its minimizer $j^\star$ is unique, set

$$\Delta = \min_{j \neq j^\star}(L_j - L_{j^\star}) > 0. \tag{130}$$

Proof of Corollary 5.1. For each fixed pair (j, n), apply Theorem 4.3 to the candidate-j trajectory with the test function ℓ_n. Summing over the finitely many modes gives $\widehat{L}_{I,j} \xrightarrow{\mathbb{P}} L_j$. Since N_{cand} is fixed, a finite union bound gives (23). The candidate trajectories may share the same tensor and need not be independent.

Put $E_I = \max_j \left|\widehat{L}_{I,j} - L_j\right|$. On $\{E_I < \Delta/2\}$, for every $j \neq j^\star$,

$$\widehat{L}_{I,j} - \widehat{L}_{I,j^\star} \geq L_j - L_{j^\star} - 2E_I \geq \Delta - 2E_I > 0. \tag{131}$$

This proves the unique-minimizer statement. □

Finite-family extensions. Write $\mathcal{J}_\star = \arg\min_j L_j$. If

$$\min_{j \notin \mathcal{J}_\star} \left(L_j - \min_k L_k \right) > 0$$

with the minimum over the empty set interpreted as $+\infty$, every empirical minimizer lies in $\mathcal{J}_\star$ with probability tending to one. A gap $L_{(k)} < L_{(k+1)}$, where $L_{(k)}$ is the kth smallest population loss, gives consistent top-k recovery. Pairwise distinct population losses give recovery of the full ranking.

These conclusions also hold for a panel parameterized by a finite-dimensional random element $\widehat{\theta}_{\mathrm{panel}}$ that is independent of all model and initialization randomness used for evaluation and frozen across the ambient-size sequence. Assume the parameterized maps are jointly Borel measurable and admissible almost surely. The corresponding conditional population losses are denoted by $L_j(\widehat{\theta}_{\mathrm{panel}})$.

Finally, suppose independent Monte Carlo SE scores with sample size S satisfy

$$\max_{1 \leq j \leq N_{\mathrm{cand}}} \left|\widetilde{L}_{S,j} - L_j\right| \xrightarrow{\mathbb{P}} 0 \qquad \text{as } S \to \infty. \tag{132}$$

If the fixed finite family has a unique minimizer, then it has a positive gap. With measurable tie-breaking, $\widetilde{j}_S \in \arg\min_j \widetilde{L}_{S,j}$ is consistent and, along every sequence $I_q, S_q \to \infty$,

$$\mathbb{P}\left(\widetilde{j}_{S_q} = \widehat{j}_{I_q} = j^\star\right) \to 1. \tag{133}$$

Condition (132) is additional to Theorem 4.3.

Proof. The set-valued, top-k, and full-ranking statements use the preceding margin argument with their finite collections of positive gaps.

For a random panel, condition on its finite-dimensional parameter $\widehat{\theta}_{\mathrm{panel}}$. Independence leaves the model and initialization assumptions unchanged. For almost every admissible realization, the deterministic argument applies, and joint Borel measurability makes the conditional failure probability measurable. It converges to zero almost surely and is bounded by one, so dominated convergence gives the unconditional result.

The Monte Carlo minimizer statement follows from the same finite-family and margin argument with $\widetilde{L}_{S,j}$ in place of $\widehat{L}_{I,j}$. Finally, along any $I_q, S_q \to \infty$, the union bound

$$\mathbb{P}\left\{(\widetilde{j}_{S_q}, \widehat{j}_{I_q}) \neq (j^\star, j^\star)\right\} \leq \mathbb{P}(\widetilde{j}_{S_q} \neq j^\star) + \mathbb{P}(\widehat{j}_{I_q} \neq j^\star) \tag{134}$$

proves (133). □

F.4 Terminal history readouts

Shorter feature sets, such as $(\boldsymbol{x}_n^0, \boldsymbol{b}_n^0)$, are represented by setting unused rows of the full readout matrix to zero.

Corollary F.5 (Terminal history readouts)**.** *Under the assumptions of Theorem 4.3, fix a finite collection of matrices* $\mathbf{W}_{n,j}^{\mathrm{out}} \in \mathbb{R}^{(T+2)P_n \times P_n}$, $n \in [N]$, $j \in [N_{\mathrm{cand}}]$, *and define*

$$\widehat{L}_{I,j}^{\mathrm{out}} = \frac{\sum_{n=1}^N I_n^{-1} \left\| \mathbf{H}_{n,I} \mathbf{W}_{n,j}^{\mathrm{out}} - \mathbf{A}_n \right\|_{\mathrm{F}}^2}{\sum_{n=1}^N P_n},$$

$$L_j^{\mathrm{out}} = \frac{\sum_{n=1}^{N} \mathbb{E}\left\| \boldsymbol{h}_n \mathbf{W}_{n,j}^{\mathrm{out}} - \boldsymbol{a}_n \right\|_2^2}{\sum_{n=1}^{N} P_n}. \tag{135}$$

Then $\max_j |\widehat{L}_{I,j}^{\mathrm{out}} - L_j^{\mathrm{out}}| \xrightarrow{\mathbb{P}} 0$. *The same conclusion holds conditionally on a jointly Borel measurable collection of almost-surely finite readout matrices selected independently of the evaluated model and initialization sequence and held fixed across ambient sizes. A unique population minimizer gives consistent empirical selection. Independent MC-SE selection additionally requires the score consistency condition* (132) *for these losses.*

Proof. For each fixed matrix, the loss in (135) is a quadratic function of the finite-dimensional planted row and history, hence PL(2). Theorem 4.3, summation over finitely many modes, and a finite union bound give the stated convergence. Conditioning on independently fitted matrices leaves the evaluation law unchanged. Almost-sure conditional convergence of the failure probabilities and bounded convergence give the unconditional result as in Corollary F.3. The selection claims use the same finite-family margin argument proved above. The same row-law argument gives limits of the readout Gram matrices and their overlaps with the planted factors. The tensor-error factorization in Appendix C then gives normalized reconstruction-error limits whenever the limiting signal energy is positive. □

The ridge formula (25) also applies to a shorter history before zero-padding. For nonzero means, center the features and response and add the corresponding intercept.

G Specializations and Separating Examples

This appendix first specializes the covariance and Onsager reaction maps to a diagonal CP core, then explains the additional constraints required to represent a general core through a diagonal lift. The remaining subsections treat the Bayes-matched reduction, off-Nishimori non-closure, a one-step overlap bound, and a mixed-time separating example. The final subsection relates the reaction formula to bilinear and linear AMP through a common covariance potential.

G.1 Diagonal CP core

Assume $P_1 = \cdots = P_N$ throughout this subsection. For the diagonal CP core

$$\boldsymbol{\mathcal{C}}_{p_1,\ldots,p_N} = \mathbf{1}\{p_1 = \cdots = p_N\}, \tag{136}$$

the two core indicators in $\mathcal{K}_n$ force $p_m = p_n$ and $r_m = r_n$ for every $m \neq n$. This proves (22). Differentiating this reduced map leaves the previous and current channel indices distinct, giving

$$D_m \mathcal{K}_n(\mathbf{Q})[\mathbf{H}] = \alpha_n \mathbf{H} \odot \bigodot_{k \neq n,m} \mathbf{Q}_k. \tag{137}$$

Setting $\mathbf{H} = \overline{\mathbf{J}}_m^t$ gives the population diagonal-core Onsager reaction without requiring a diagonal Jacobian. Replacing these objects by their finite-system counterparts gives

$$\mathbf{\Omega}_{n,I}^t = \alpha_{n,I} \sum_{m \neq n} \overline{\mathbf{J}}_{m,I}^t \odot \bigodot_{k \neq n,m} \mathbf{Q}_{k,I}^{t-1,t}. \tag{138}$$

Table 3 summarizes the diagonal-core specialization, including the additional Bayes-matched reduction under Corollary G.2.

G.2 Original coordinates versus a diagonal-core lift

Fix the finite CP representation

$$\boldsymbol{\mathcal{C}}_{p_1,\ldots,p_N} = \sum_{\ell=1}^{L} \prod_{n=1}^{N} \mathbf{U}_n[p_n, \ell], \qquad \mathbf{U}_n \in \mathbb{R}^{P_n \times L}. \tag{139}$$

Table 3: Diagonal-core specialization of the general-core objects.

General-core object	Diagonal-core specialization
mode-specific channel dimension P_n	equal dimensions $P_1 = \cdots = P_N$
core $\boldsymbol{\mathcal{C}}$	equality indicator (136)
mean coefficient $\mathcal{K}_n(\mathbf{M})$	$\alpha_n \bigodot_{m\neq n} \mathbf{M}_m$
covariance $\mathcal{K}_n(\mathbf{Q})$	$\alpha_n \bigodot_{m\neq n} \mathbf{Q}_m$
SE order parameters	$(\mathbf{M}, \mathbf{Q}^{s,t})$
derived reaction input	$\overline{\mathbf{J}}$
Bayes-matched state under Corollary G.2	$\mathbf{M}_n^t = \mathbf{Q}_n^{t,t}$

Write $\widetilde{\mathcal{K}}_n(\widetilde{\mathbf{Q}}) = \alpha_n \bigodot_{m\neq n} \widetilde{\mathbf{Q}}_m$ for the rank-L diagonal-core covariance map and set $\widetilde{\mathbf{Q}}_m = \mathbf{U}_m^\top \mathbf{Q}_m \mathbf{U}_m$. A lifted row has the constrained form $\widetilde{\boldsymbol{x}}_n = \boldsymbol{x}_n \mathbf{U}_n$. Preserving the original-coordinate row update on every attainable lifted field requires

$$\widetilde{f}_n(\widetilde{\boldsymbol{b}}, \widetilde{\mathbf{S}}) = f_n(\widetilde{\boldsymbol{b}}\mathbf{U}_n^\top, \mathbf{U}_n \widetilde{\mathbf{S}} \mathbf{U}_n^\top)\mathbf{U}_n. \tag{140}$$

Proposition G.1 (Recursion preservation under a diagonal-core lift)**.** *Suppose a finite CP representation of $\boldsymbol{\mathcal{C}}$ uses factor maps $\mathbf{U}_n \in \mathbb{R}^{P_n \times L}$. Projecting the rank-$L$ diagonal-core covariance map and its derivatives through the $\mathbf{U}_n$ gives $\mathcal{K}_n$ and $D_m\mathcal{K}_n$. Preserving arbitrary original-coordinate row histories additionally requires $\operatorname{rank}(\mathbf{U}_n) = P_n$ for a lossless state representation, together with the denoiser consistency in (140) on attainable lifted fields. A trajectory confined to a proper subspace may admit a lower-rank representation.*

Proof. The projected covariance and derivative identities are

$$\mathcal{K}_n(\mathbf{Q}) = \mathbf{U}_n \widetilde{\mathcal{K}}_n\big((\mathbf{U}_m^\top \mathbf{Q}_m \mathbf{U}_m)_{m\neq n}\big)\mathbf{U}_n^\top, \tag{141}$$

$$D_m\mathcal{K}_n(\mathbf{Q})[\mathbf{H}] = \mathbf{U}_n D_m \widetilde{\mathcal{K}}_n(\widetilde{\mathbf{Q}})[\mathbf{U}_m^\top \mathbf{H} \mathbf{U}_m]\mathbf{U}_n^\top. \tag{142}$$

Consequently, with $\widetilde{\mathbf{J}}_m = \mathbf{U}_m^\top \overline{\mathbf{J}}_m \mathbf{U}_m$, projecting the lifted reaction by $\mathbf{U}_n(\cdot)\mathbf{U}_n^\top$ gives exactly $\sum_{m\neq n} D_m\mathcal{K}_n(\mathbf{Q})[\overline{\mathbf{J}}_m]$.

Expanding $\mathbf{U}_m^\top \mathbf{Q}_m \mathbf{U}_m$ in (141), the two sums over CP terms reconstruct the two copies of $\boldsymbol{\mathcal{C}}$ in the definition of $\mathcal{K}_n$. Differentiating this identity gives (142) and the reaction statement.

The row-input Jacobian of (140) is

$$D_{\widetilde{\boldsymbol{b}}} \widetilde{f}_n(\widetilde{\boldsymbol{b}}, \widetilde{\mathbf{S}}) = \mathbf{U}_n^\top D_{\boldsymbol{b}} f_n(\widetilde{\boldsymbol{b}}\mathbf{U}_n^\top, \mathbf{U}_n \widetilde{\mathbf{S}} \mathbf{U}_n^\top)\mathbf{U}_n,$$

as used in Proposition G.1. These covariance, derivative, and Jacobian identities also hold at the finite-system scale $\alpha_{n,I}$. At $\widetilde{\mathbf{X}}_m^t = \mathbf{X}_m^t \mathbf{U}_m$, expanding (139) in the Gaussian contraction gives

$$\widetilde{\mathcal{H}}_{n,I}\big((\mathbf{X}_m^t \mathbf{U}_m)_m\big)\mathbf{U}_n^\top = \mathcal{H}_{n,I}(\mathbf{X}^t),$$

where the tilde denotes the diagonal-core contraction of the same Gaussian tensor. The same identity holds for the signal contraction. Together with $\mathbf{U}_n \widetilde{\boldsymbol{\Omega}}_{n,I}^t \mathbf{U}_n^\top = \boldsymbol{\Omega}_{n,I}^t$ and (140), this yields

$$\widetilde{\mathbf{B}}_n^t \mathbf{U}_n^\top = \mathbf{B}_n^t, \qquad \widetilde{\mathbf{X}}_n^{t+1} = f_{n,t}(\mathbf{B}_n^t, \mathbf{K}_{n,I}^{t,t})\mathbf{U}_n = \mathbf{X}_n^{t+1}\mathbf{U}_n,$$

with $f_{n,t}$ applied rowwise. Starting from $\widetilde{\mathbf{X}}_n^0 = \mathbf{X}_n^0 \mathbf{U}_n$, induction proves these finite-trajectory identities.

If the lift must retain arbitrary original-coordinate row histories injectively, then $\operatorname{rank}(\mathbf{U}_n) = P_n$ and hence $L \geq P_n$. Under this rank condition, $L > P_n$ introduces $L - P_n$ independent linear consistency constraints. For $L < P_n$ no injective lift exists. Even for a square full-rank lift, the denoiser must be conjugated as in (140). An unconstrained diagonal-core AMP with coordinatewise lifted denoisers does not, in general, preserve the original-coordinate recursion. □

G.3 Bayes-matched reduction and Gaussian calculation

Corollary G.2 (Bayes-matched diagonal-core reduction)**.** *Under the assumptions of Theorem 4.3 and for the diagonal CP core (136), suppose the planted rows are drawn from the priors used by matched posterior-mean denoisers and $\mathbf{M}_n^0 = \mathbf{Q}_n^{0,0}$ for every mode. Then*

$$\mathbf{M}_n^t = \mathbf{Q}_n^{t,t}, \qquad \mathbf{\Theta}_n^t = \mathbf{K}_n^{t,t} = \alpha_n \bigodot_{m \neq n} \mathbf{Q}_m^{t,t}. \tag{143}$$

If, in addition, the priors factorize across channels and the single-time order parameters are diagonal in that channel basis, the resulting single-time recursion has the same coordinatewise Gaussian-channel form as the dynamic mean-field equations of Kadmon & Ganguli (2018), with the aspect-ratio normalization given by (22). The mode-n Onsager reaction matrix in the rigorous trajectory still uses the mixed overlaps $(\mathbf{Q}_m^{t-1,t})_{m \neq n}$.

Proof. Suppose $\mathbf{M}_m^t = \mathbf{Q}_m^{t,t}$ for every mode. Because the same core-induced covariance map $\mathcal{K}_n$ governs the signal and covariance, it gives $\mathbf{\Theta}_n^t = \mathbf{K}_n^{t,t} =: \mathbf{K}$. Thus

$$\boldsymbol{b}_n^t = \boldsymbol{a}_n \mathbf{K} + \boldsymbol{u}_n^t, \qquad \boldsymbol{u}_n^t \sim \mathcal{N}(0, \mathbf{K}),$$

is the matched Gaussian channel for the planted row. If $\boldsymbol{x}_n^{t+1} = \mathbb{E}[\boldsymbol{a}_n \mid \boldsymbol{b}_n^t]$, the tower property gives

$$\mathbb{E}[\boldsymbol{a}_n^\top \boldsymbol{x}_n^{t+1}] = \mathbb{E}[(\boldsymbol{x}_n^{t+1})^\top \boldsymbol{x}_n^{t+1}]. \tag{144}$$

The initialization condition $\mathbf{M}_n^0 = \mathbf{Q}_n^{0,0}$ therefore propagates by induction. Under the additional factorization and diagonal assumptions, (22) and the posterior-mean map act coordinatewise, which gives the stated Gaussian-channel reduction. □

For a Gaussian calculation with any fixed core, suppose $\boldsymbol{a}_n \sim \mathcal{N}(0, \mathbf{\Lambda}_n^{-1})$ and, inductively, that $\mathbf{M}_m^t = \mathbf{Q}_m^{t,t}$ for every mode. The shared core-induced covariance map then gives $\mathbf{\Theta}_n^t = \mathbf{K}_n^{t,t} =: \mathbf{K}$. In the effective channel

$$\boldsymbol{b} = \boldsymbol{a}_n \mathbf{K} + \boldsymbol{u}, \qquad \boldsymbol{u} \sim \mathcal{N}(0, \mathbf{K}), \tag{145}$$

the Gaussian posterior mean is $\boldsymbol{x} = \boldsymbol{b}(\mathbf{K} + \mathbf{\Lambda}_n)^{-1}$. Hence

$$f_n(\boldsymbol{b}, \mathbf{K}) = \boldsymbol{b}(\mathbf{K} + \mathbf{\Lambda}_n)^{-1}, \qquad \mathbf{M}_n^{t+1} = \mathbf{Q}_n^{t+1,t+1} = \mathbf{\Lambda}_n^{-1} - (\mathbf{\Lambda}_n + \mathbf{K}_n^{t,t})^{-1}. \tag{146}$$

Direct calculation gives

$$\mathbb{E}[\boldsymbol{a}_n^\top \boldsymbol{x}] = \mathbf{\Lambda}_n^{-1} \mathbf{K} (\mathbf{K} + \mathbf{\Lambda}_n)^{-1} = \mathbf{\Lambda}_n^{-1} - (\mathbf{K} + \mathbf{\Lambda}_n)^{-1}, \tag{147}$$

$$\begin{aligned} \mathbb{E}[\boldsymbol{x}^\top \boldsymbol{x}] &= (\mathbf{K} + \mathbf{\Lambda}_n)^{-1} (\mathbf{K} \mathbf{\Lambda}_n^{-1} \mathbf{K} + \mathbf{K})(\mathbf{K} + \mathbf{\Lambda}_n)^{-1} \\ &= \mathbf{\Lambda}_n^{-1} - (\mathbf{K} + \mathbf{\Lambda}_n)^{-1}. \end{aligned} \tag{148}$$

This proves the induction and (146). Under the diagonal core, diagonal initial order parameters, and priors whose $\mathbf{\Lambda}_n$ are diagonal in the same channel basis, each component evolves independently, giving the diagonal-core Bayes single-time dynamic mean-field update.

G.4 Off-Nishimori non-closure

An explicit admissible construction shows that signal overlaps alone do not close the recursion. Take $P_n = 1$ for every mode and the scalar core $\boldsymbol{\mathcal{C}} = 1$. Take all A_k, V_k to be independent $\mathcal{N}(0,1)$. For the first initialization set $X_k^0 = A_k$ in every mode. Fix one mode m and, for the second, set $\widetilde{X}_m^0 = A_m + V_m$ and $\widetilde{X}_k^0 = A_k$ for $k \neq m$. Both laws have $M_k^0 = 1$ for every k. The first has $Q_k^{0,0} = 1$ in every mode, whereas the second has $\widetilde{Q}_m^{0,0} = 2$ and $\widetilde{Q}_k^{0,0} = 1$ for $k \neq m$. Thus all signal means agree while, for every $n \neq m$, $\widetilde{K}_n^{0,0} = 2K_n^{0,0}$, so the effective Gaussian field laws differ.

The identity denoiser $f(b, S) = b$ makes the first-iterate laws differ in every mode $n \neq m$. It is globally Lipschitz, has constant bounded Jacobian, and is independent of the covariance argument. The construction therefore lies within the assumptions of Theorem 4.3 and establishes the claimed non-closure.

G.5 One-step behavior near zero signal overlap

Corollary G.3 (One-step superlinear overlap bound)**.** *Under Assumption 4.2, suppose* $\mathbb{E}[\boldsymbol{a}_n] = 0$ *for every mode and write* $\epsilon_t = \max_n \left\|\mathbf{M}_n^t\right\|_{\mathrm{F}}$. *Along any covariance path admitted by the state-evolution recursion, there is* $C_t < \infty$ *such that*

$$\epsilon_{t+1} \le C_t \epsilon_t^{N-1}. \tag{149}$$

Thus zero signal overlap is invariant. For $\epsilon_t > 0$, $\epsilon_{t+1}/\epsilon_t \le C_t \epsilon_t^{N-2}$, *so for* $N \ge 3$ *the signal-overlap amplitude strictly decreases in the next step whenever* $\epsilon_t < C_t^{-1/(N-2)}$ *(immediately if* $C_t = 0$*), even when the covariance state does not vanish.*

Proof. Multilinearity of $\mathcal{K}_n$ gives $\left\|\mathbf{\Theta}_n^t\right\|_{\mathrm{F}} \le C \prod_{m \ne n} \left\|\mathbf{M}_m^t\right\|_{\mathrm{F}}$. Since $\boldsymbol{u}_n^t$ is independent of $\boldsymbol{a}_n$, centering and the row-map Lipschitz bound imply

$$\begin{aligned}\left\|\mathbf{M}_n^{t+1}\right\|_{\mathrm{F}} &= \left\|\mathbb{E}\Big[(\boldsymbol{a}_n)^\top \{f_{n,t}(\boldsymbol{a}_n \mathbf{\Theta}_n^t + \boldsymbol{u}_n^t, \mathbf{K}_n^{t,t}) - f_{n,t}(\boldsymbol{u}_n^t, \mathbf{K}_n^{t,t})\}\Big]\right\|_{\mathrm{F}} \\ &\le C_t \left\|\mathbf{\Theta}_n^t\right\|_{\mathrm{F}} \le C_t \epsilon_t^{N-1}.\end{aligned}$$

Taking the maximum over modes proves Corollary G.3. □

G.6 An admissible mixed-time discrepancy

Consider a scalar three-mode CP channel with $c_n = \nu = 1$, so $\alpha_n = 1$ for every mode. Take the $A_n \sim \mathcal{N}(0,1)$ to be independent across modes, initialize $X_n^0 = A_n$, and use the admissible first denoiser

$$f_{n,0}(b, S) = \frac{b}{\sqrt{2}}. \tag{150}$$

Choose any admissible denoisers at later times. The Gaussian planted-initial pair has moments of every order, while (150) is globally Lipschitz, has a constant bounded Jacobian, and is independent of its covariance argument. The construction therefore satisfies Assumption 4.2. The state-evolution recursion has $M_n^0 = Q_n^{0,0} = 1$, and hence $\Theta_n^0 = K_n^{0,0} = 1$. Therefore

$$B_n^0 = A_n + U_n^0, \qquad X_n^1 = \frac{A_n + U_n^0}{\sqrt{2}}, \qquad U_n^0 \sim \mathcal{N}(0,1), \quad U_n^0 \text{ is independent of } A_n. \tag{151}$$

It follows directly that

$$Q_n^{1,1} = 1, \qquad Q_n^{0,1} = \frac{1}{\sqrt{2}}, \qquad \overline{J}_n^1 = \frac{1}{\sqrt{2}}. \tag{152}$$

For any target mode n, the exact mixed-time coefficient and its current–current substitute at $t = 1$ are consequently

$$\Omega_n^1 = \sum_{m \ne n} \overline{J}_m^1 \prod_{k \ne n,m} Q_k^{0,1} = 1, \tag{153}$$

$$\Omega_{n,\mathrm{cc}}^1 = \sum_{m \ne n} \overline{J}_m^1 \prod_{k \ne n,m} Q_k^{1,1} = \sqrt{2}. \tag{154}$$

Thus the substitution changes the reaction by $\sqrt{2} - 1$ along a population SE trajectory covered by Theorem 4.3, while the corresponding finite-system AMP converges to this trajectory. The numerical experiment additionally measures the induced field-mean and covariance errors along a longer AMP trajectory.

G.7 Covariance–response form across interaction orders

The derivative in (12) can be expressed through a scalar covariance potential, paralleling the covariance-derivative form of scalar-overlap multi-species AMP (Huang & Sellke, 2024). This gives a common algebraic description of the reaction coefficients for multilinear factorization, its two-mode specialization, and the two message streams of linear AMP.

Covariance potential and reaction. Recall $L_I = \sum_{k=1}^N I_k$ and set $\lambda_{n,I} = I_n/L_I$. For an arbitrary matrix tuple $\mathbf{E} = (\mathbf{E}_k)_{k=1}^N$, define

$$\Psi_I(\mathbf{E}) = \frac{\prod_{k=1}^N I_k}{L_I \nu_I} \sum_{\boldsymbol{p},\boldsymbol{r}} \mathcal{C}_{\boldsymbol{p}} \mathcal{C}_{\boldsymbol{r}} \prod_{k=1}^N \mathbf{E}_k[p_k, r_k]. \tag{155}$$

All matrix entries are treated as independent coordinates when differentiating: $[\nabla_n \Psi_I(\mathbf{E})]_{pr} = \partial \Psi_I(\mathbf{E})/\partial \mathbf{E}_n[p,r]$. In particular, mixed-time overlaps need not be symmetric.

Proposition G.4 (Covariance potential representation)**.** *For every $N \geq 2$, the finite-system covariance map and reaction satisfy*

$$\mathcal{K}_{n,I}(\mathbf{E}) = \lambda_{n,I}^{-1} \nabla_n \Psi_I(\mathbf{E}), \tag{156}$$

$$\mathbf{\Omega}_{n,I}^t = \lambda_{n,I}^{-1} \sum_{m \neq n} D_m \nabla_n \Psi_I(\mathbf{Q}_I^{t-1,t})[\overline{\mathbf{J}}_{m,I}^t], \qquad t \geq 1. \tag{157}$$

Thus the reaction is a contraction of the mixed Hessian blocks of Ψ_I with the average denoiser Jacobians.

Proof. Differentiating (155) with respect to $\mathbf{E}_n[p,r]$ removes its mode-n factor. Multiplication by $\lambda_{n,I}^{-1}$ changes the prefactor to $\prod_{k \neq n} I_k/\nu_I = \alpha_{n,I}$, giving (10). Substitution into (12) proves (157). □

The potential has a direct Gaussian interpretation. For deterministic query arrays $\mathbf{X}$, set

$$H_I(\mathbf{X}) = \frac{1}{\sqrt{\nu_I}} \sum_{\boldsymbol{i}} [\mathcal{G}]_{\boldsymbol{i}} \sum_{\boldsymbol{p}} \mathcal{C}_{\boldsymbol{p}} \prod_{k=1}^N \mathbf{X}_k[i_k, p_k]. \tag{158}$$

Its gradient with respect to $\mathbf{X}_n$ is the Gaussian field $\mathcal{H}_{n,I}(\mathbf{X})$ in (32). Independence of the entries of $\mathcal{G}$ and factorization over ambient indices give

$$\mathbb{E}[H_I(\mathbf{X}) H_I(\widetilde{\mathbf{X}})] = L_I \Psi_I\left(\left(I_k^{-1} \mathbf{X}_k^\top \widetilde{\mathbf{X}}_k \right)_{k=1}^N \right). \tag{159}$$

Hence Ψ_I records the covariance of the Gaussian interaction. Its mixed Hessian blocks describe how a perturbation of one mode's overlap changes another mode's field covariance, and the Jacobian in (157) supplies the denoiser response.

Bilinear factorization. For $N = 2$, write the core as a matrix $\mathbf{C} \in \mathbb{R}^{P_1 \times P_2}$. Equations (156) and (157) become

$$\mathcal{K}_{1,I}(\mathbf{E}) = \alpha_{1,I} \mathbf{C} \mathbf{E}_2 \mathbf{C}^\top, \qquad \mathcal{K}_{2,I}(\mathbf{E}) = \alpha_{2,I} \mathbf{C}^\top \mathbf{E}_1 \mathbf{C}, \tag{160}$$

$$\mathbf{\Omega}_{1,I}^t = \alpha_{1,I} \mathbf{C} \overline{\mathbf{J}}_{2,I}^t \mathbf{C}^\top, \qquad \mathbf{\Omega}_{2,I}^t = \alpha_{2,I} \mathbf{C}^\top \overline{\mathbf{J}}_{1,I}^t \mathbf{C}. \tag{161}$$

There is no explicit overlap factor left in these derivatives, although the Jacobians can still depend on the trajectory. For a diagonal CP core of order N, each summand in (138) instead contains $N-2$ mixed-time overlap factors. For example, at $N = 3$, with $\{n,m,k\} = \{1,2,3\}$,

$$\mathbf{\Omega}_{n,I}^t = \alpha_{n,I} \big(\overline{\mathbf{J}}_{m,I}^t \odot \mathbf{Q}_{k,I}^{t-1,t} + \mathbf{Q}_{m,I}^{t-1,t} \odot \overline{\mathbf{J}}_{k,I}^t \big). \tag{162}$$

The interaction order therefore determines which additional overlaps enter the reaction, rather than changing the derivative–response rule.

Linear inverse problems as two message streams. The linear model $\boldsymbol{y} = \mathbf{W}_{\text{lin}} \boldsymbol{x}_\star + \boldsymbol{\varepsilon}$ has a known sensing matrix $\mathbf{W}_{\text{lin}} \in \mathbb{R}^{d_y \times d_x}$ with independent $\mathcal{N}(0, 1/d_y)$ entries and aspect ratio $\delta_{\text{lin}} = d_y/d_x$. For column-vector queries $\boldsymbol{x} \in \mathbb{R}^{d_x}$ and $\boldsymbol{z} \in \mathbb{R}^{d_y}$, the two Gaussian operators $\mathbf{W}_{\text{lin}} \boldsymbol{x}$ and $\mathbf{W}_{\text{lin}}^\top \boldsymbol{z}$ are the gradients of the bilinear interaction $H(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{z}^\top \mathbf{W}_{\text{lin}} \boldsymbol{x}$. Fixed query pairs have overlaps $q_x = d_x^{-1} \boldsymbol{x}^\top \widetilde{\boldsymbol{x}}$ and $q_z = d_y^{-1} \boldsymbol{z}^\top \widetilde{\boldsymbol{z}}$. With $L_{\text{lin}} = d_x + d_y$ and $\lambda_x = d_x/L_{\text{lin}}$, its normalized

covariance is $\Psi(q_x, q_z) = \lambda_x q_x q_z$, since $\mathbb{E}[H(\boldsymbol{x}, \boldsymbol{z})H(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{z}})] = L_{\text{lin}}\Psi(q_x, q_z)$. The coordinatewise query covariances and their derivatives are consequently

$$K_z = \delta_{\text{lin}}^{-1} q_x, \qquad K_x = q_z, \qquad \partial_{q_x} K_z = \delta_{\text{lin}}^{-1}, \qquad \partial_{q_z} K_x = 1. \tag{163}$$

To recover the classical AMP iteration, use a differentiable coordinatewise input denoiser $\eta_{\text{lin},t}$ and the output map $g_{\text{out}}(p, y) = y - p$. Starting from $\boldsymbol{z}^{-1} = 0$ and $\beta_{\text{lin},0} = 0$, the output-side Onsager coefficient at $t \geq 1$ is $\beta_{\text{lin},t} = \delta_{\text{lin}}^{-1}\langle \eta'_{\text{lin},t-1}(\boldsymbol{r}_{\text{lin}}^{t-1})\rangle$, where brackets denote the empirical average over the relevant coordinates. The two derivative–response corrections give

$$\boldsymbol{p}_{\text{lin}}^t = \mathbf{W}_{\text{lin}}\boldsymbol{x}^t - \beta_{\text{lin},t}\boldsymbol{z}^{t-1}, \qquad \boldsymbol{z}^t = g_{\text{out}}(\boldsymbol{p}_{\text{lin}}^t, \boldsymbol{y}) = \boldsymbol{y} - \boldsymbol{p}_{\text{lin}}^t, \tag{164}$$

$$\begin{aligned} \boldsymbol{r}_{\text{lin}}^t &= \mathbf{W}_{\text{lin}}^\top \boldsymbol{z}^t - \boldsymbol{x}^t \langle \partial_p g_{\text{out}}(\boldsymbol{p}_{\text{lin}}^t, \boldsymbol{y})\rangle \\ &= \boldsymbol{x}^t + \mathbf{W}_{\text{lin}}^\top \boldsymbol{z}^t, \qquad \boldsymbol{x}^{t+1} = \eta_{\text{lin},t}(\boldsymbol{r}_{\text{lin}}^t). \end{aligned} \tag{165}$$

Eliminating $\boldsymbol{p}_{\text{lin}}^t$ yields $\boldsymbol{z}^t = \boldsymbol{y} - \mathbf{W}_{\text{lin}}\boldsymbol{x}^t + \beta_{\text{lin},t}\boldsymbol{z}^{t-1}$, the usual residual update (Donoho et al., 2009; Bayati & Montanari, 2011). The $+\boldsymbol{x}^t$ term in the input field arises from the output response $\partial_p g_{\text{out}} = -1$. Thus linear AMP fits the same covariance–response calculus through the two alternating message streams of asymmetric AMP (Gerbelot & Berthier, 2023).

Additivity and SE scope. For a sum of independent centered Gaussian interactions on the same collection of query arrays, the covariance potentials add under a common normalization. Their covariance derivatives, and hence the corresponding reaction contributions for fixed denoiser responses, add as well. These are algebraic identities. Matrix and graph-based SE results (Bayati & Montanari, 2011; Gerbelot & Berthier, 2023) address the corresponding Gaussian-matrix recursions. Theorem 4.3 here retains the assumption $N \geq 3$. Its coercivity argument uses Lemma D.2, whose full-space lower bound fails at degree two because of skew-symmetric null directions.

H Experimental Settings and Supplementary Results

The subsections first document the fixed-horizon SE and reaction comparisons from Section 6 and an unequal-channel covariance-dependent recursion, followed by the implementation consequence of the factorized reaction identity. We then give the representation-calibration, terminal-readout, and frozen-panel ranking settings, the one-step calculation near zero overlap, an order-four tensor experiment, and an out-of-model comparison of AMP and alternating least squares (ALS) on the measured amino-acid fluorescence tensor using all observed entries.

All numerical experiments were run on a MacBook Pro with a 12-core Apple M2 Pro processor and 16 GB of unified memory under macOS 15.2. Except for the frozen-denoiser surrogate-ranking pipeline, AMP and SE evaluations used MATLAB R2024b Update 3. For that pipeline, candidate generation, Monte Carlo SE scoring, and aggregation used Python 3.10.12, whereas finite-system evaluation used MATLAB R2024b Update 3.

In every synthetic experiment generated from (4), the planted-factor entries are i.i.d. $\mathcal{N}(0, 1)$ across modes, rows, and channels, independently of the observation noise. Thus the planted-row covariance is $\mathbf{I}_{P_n}$ in each mode. All tensor entries are observed in these synthetic experiments.

Coefficient and overlap calibration uses pooled relative Frobenius errors, factor recovery uses per-entry MSE, and tensor recovery uses NMSE normalized by reference-tensor energy. Selection is evaluated using the same recovery loss and its regret. For relative calibration errors, we pool squared discrepancies and reference energies before dividing. If the pooled reference root mean square (RMS) per entry is at most 10^{-6}, we report the square root of the mean squared blockwise Frobenius error instead and mark the relative error unavailable. Zero-target moments use the nonzero covariance scales specified below.

In Table 4, $\mathbf{E}_{\text{init}}$ denotes an i.i.d. standard Gaussian initialization-noise matrix of the corresponding ambient dimension, independent of the planted factors and observation noise. The five-size grid is

$$\mathcal{I}_5 = \{(36, 45, 54), (48, 60, 72), (72, 90, 108), (96, 120, 144), (120, 150, 180)\}.$$

The fixed initialization and denoiser matrices are

$$\mathbf{P}_c = \begin{bmatrix} 0.08 & 0.62 \\ -0.48 & 0.12 \end{bmatrix}, \qquad \mathbf{W}_c = \begin{bmatrix} 0.82 & 0.48 \\ -0.38 & 0.74 \end{bmatrix},$$

$$\mathbf{P}_r = \begin{bmatrix} 0.60 & 0.20 \\ -0.15 & 0.40 \end{bmatrix}, \qquad \mathbf{W}_r = \begin{bmatrix} 0.68 & 0.28 \\ -0.22 & 0.60 \end{bmatrix},$$
$$\mathbf{P}_g = \begin{bmatrix} 0.55 & 0.18 \\ -0.12 & 0.35 \end{bmatrix}, \qquad \mathbf{W}_g = \begin{bmatrix} 0.62 & 0.24 \\ -0.18 & 0.56 \end{bmatrix}.$$

All displayed $\tanh$ maps act elementwise.

Table 4: Configurations for the trajectory and unequal-channel validation studies. Monte Carlo (MC) row counts are per mode.

Study	Core, noise, horizon	Ambient dimensions	Initialization; denoiser	Repetitions
Off-Nishimori separation	non-diagonal $P_n = 2$ core; $\gamma = 1$; $\nu = 0.55$; $T = 5$	$(72, 90, 108)$	$\mathbf{A}\mathbf{P}_c + 0.95\mathbf{E}_{\text{init}}$; $1.15\tanh(\mathbf{B}\mathbf{W}_c)$	AMP seeds: 11–40 MC rows: 100,000 SE seed: 201
Reaction ablation	non-diagonal $P_n = 2$ core; $\gamma = 0.7$; $\nu = 0.5$; $T = 5$	$\mathcal{I}_5$	$\mathbf{A}\mathbf{P}_r + 0.5\mathbf{E}_{\text{init}}$; $1.35\tanh(\mathbf{B}\mathbf{W}_r)$	AMP seeds: 131–160 MC rows: 100,000 SE seed: 301
Non-diagonal-core scaling	$P_n = 2$; $\gamma \in \{0.2, 0.7, 1\}$; $\nu = 0.6$; $T = 5$	$\mathcal{I}_5$	$\mathbf{A}\mathbf{P}_g + 0.55\mathbf{E}_{\text{init}}$; $\tanh(\mathbf{B}\mathbf{W}_g)$	AMP seeds: 371–400 MC rows: 120,000 SE seed: 501
Unequal-channel calibration	$2 \times 3 \times 4$ core (Appendix H.2); $\nu = 0.6$; $T = 3$	$(36, 45, 54)$, $(72, 90, 108)$, $(120, 150, 180)$	(168) with $q_{\text{init}} = 0.6$; $\boldsymbol{b}(\mathbf{I}_{P_n} + \mathbf{S})^{-1}$	AMP seeds: 22201–22230 Deterministic moment SE MC rows: 120,000 MC seeds: 22101, 22102

The off-Nishimori and reaction amplitudes were selected on search seeds disjoint from the reported seeds. The strong-core study includes two additional sizes, each evaluated with eight seeds.

H.1 Trajectory-level settings and diagnostics

For each trajectory, define the residual $\mathbf{R}_n^t = \mathbf{B}_n^t - \mathbf{A}_n \mathcal{K}_{n,I}(\mathbf{M}_I^t)$. For each configuration, we evaluate the following empirical moments:

$$\mathbf{M}_{n,I}^t, \quad \mathbf{Q}_{n,I}^{s,t}, \quad \frac{1}{I_n}(\mathbf{R}_n^s)^\top \mathbf{R}_n^t, \quad 0 \le s \le t \le T, \tag{166}$$

for every mode. The plots show seed-aggregated means with one standard error.

For Figure 1, the pooled relative Frobenius error for one seed is

$$e_{\text{hist}} = \left\{ \frac{\sum_{n=1}^{N} \sum_{t=0}^{T} \left\| \mathbf{M}_{n,I}^t - \mathbf{M}_{n,\text{MC}}^t \right\|_{\text{F}}^2 + \sum_{n=1}^{N} \sum_{t=0}^{T} \sum_{s=0}^{t} \left\| \mathbf{Q}_{n,I}^{s,t} - \mathbf{Q}_{n,\text{MC}}^{s,t} \right\|_{\text{F}}^2}{\sum_{n=1}^{N} \sum_{t=0}^{T} \left\| \mathbf{M}_{n,\text{MC}}^t \right\|_{\text{F}}^2 + \sum_{n=1}^{N} \sum_{t=0}^{T} \sum_{s=0}^{t} \left\| \mathbf{Q}_{n,\text{MC}}^{s,t} \right\|_{\text{F}}^2} \right\}^{1/2},$$

where $\mathbf{M}_{n,\text{MC}}^t$ and $\mathbf{Q}_{n,\text{MC}}^{s,t}$ are the Monte Carlo SE references. The right panel of Figure 1 reports its mean and one standard error over 30 seeds. The smallest pooled reference RMS per entry across these settings is 0.110, so nearly zero individual overlaps do not cause division by a nearly zero scale.

For the off-Nishimori separation comparison in Figure 4, write $\widehat{\mathbf{D}}$ for an empirical coefficient, $\mathbf{D}$ for its SE target, and $\widetilde{\mathbf{D}}$ for the forced-$\mathbf{M} = \mathbf{Q}$ target. Both comparisons use the same full-SE trajectory

scale. At iteration t, the pooled error is

$$\left\{ \frac{\sum_{n=1}^{N} \sum_{D\in\{M,Q,\Theta,K\}} \left\| \widehat{\mathbf{D}}_n^t - \mathbf{D}_n^t \right\|_{\mathrm{F}}^2}{(T+1)^{-1} \sum_{u=0}^{T} \sum_{n=1}^{N} \sum_{D\in\{M,Q,\Theta,K\}} \|\mathbf{D}_n^u\|_{\mathrm{F}}^2} \right\}^{1/2},$$

where Q and K denote the single-time blocks. For the forced-$\mathbf{M} = \mathbf{Q}$ comparison, only the numerator target is replaced by $\widetilde{\mathbf{D}}_n^t$. The denominator is fixed across iterations and both comparisons, avoiding normalization by a vanishing late-iteration target. Curves show means and one standard error over seeds. The forced target is a deliberately misspecified baseline.

For the reaction ablation, the current–current substitute replaces $\mathbf{Q}^{t-1,t}$ in (12) by $\mathbf{Q}^{t,t}$ while retaining the same Jacobian and sum over $m \neq n$. The diagonal substitute keeps only the diagonal entries of the exact mixed-time reaction, and the no-reaction variant sets it to zero.

For the reaction diagnostics, write $\widehat{\mathbf{K}}_n^{s,t} = I_n^{-1}(\mathbf{R}_n^s)^\top \mathbf{R}_n^t$. The operator $\mathrm{Tr}(\cdot)$ denotes the matrix trace. For $t \geq 1$ define

$$e_\mu = \frac{\left\| I_n^{-1} \sum_{i_n=1}^{I_n} [\mathbf{R}_n^t]_{i_n,\cdot} \right\|_2}{\sqrt{\mathrm{Tr}(\mathbf{K}_n^{t,t})}}, \qquad e_A = \frac{\left\| I_n^{-1} \mathbf{A}_n^\top \mathbf{R}_n^t \right\|_{\mathrm{F}}}{\sqrt{\left\| I_n^{-1} \mathbf{A}_n^\top \mathbf{A}_n \right\|_{\mathrm{F}} \left\| \mathbf{K}_n^{t,t} \right\|_{\mathrm{F}}}},$$

$$e_0 = \frac{\left\| \widehat{\mathbf{K}}_n^{t,t} - \mathbf{K}_n^{t,t} \right\|_{\mathrm{F}}}{\left\| \mathbf{K}_n^{t,t} \right\|_{\mathrm{F}}}, \qquad e_1 = \frac{\left\| \widehat{\mathbf{K}}_n^{t-1,t} - \mathbf{K}_n^{t-1,t} \right\|_{\mathrm{F}}}{\sqrt{\left\| \mathbf{K}_n^{t-1,t-1} \right\|_{\mathrm{F}} \left\| \mathbf{K}_n^{t,t} \right\|_{\mathrm{F}}}}. \tag{167}$$

All four denominators exceed 2.2 over the reported reaction settings and $t \geq 1$. The zero mean and planted cross-moment targets are scaled by covariance magnitudes, and the lag error uses single-time covariance scales rather than the possibly zero lag target. The dimensionless RMS of the residual-moment errors is $\{(e_\mu^2 + e_A^2 + e_0^2 + e_1^2)/4\}^{1/2}$. At $t = 0$ the unavailable lag component is set to zero.

Figure 3 compares the exact mixed-time reaction with the current–current, diagonal, and no-reaction substitutes. At $(120, 150, 180)$, the arithmetic mean of the modewise residual-moment RMS over the three modes, iterations $t = 1, \ldots, 5$, and 30 realizations is 0.230 for the exact reaction and 1.165 for current–current.

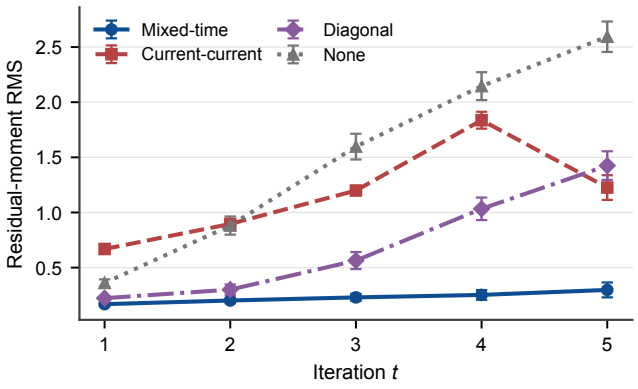

Figure 3: Reaction-substitute ablation at $\gamma = 0.7$, $\nu = 0.5$, and $(I_1, I_2, I_3) = (120, 150, 180)$. At each iteration, curves average the modewise residual-moment RMS. Bars show one standard error across 30 seedwise mode averages.

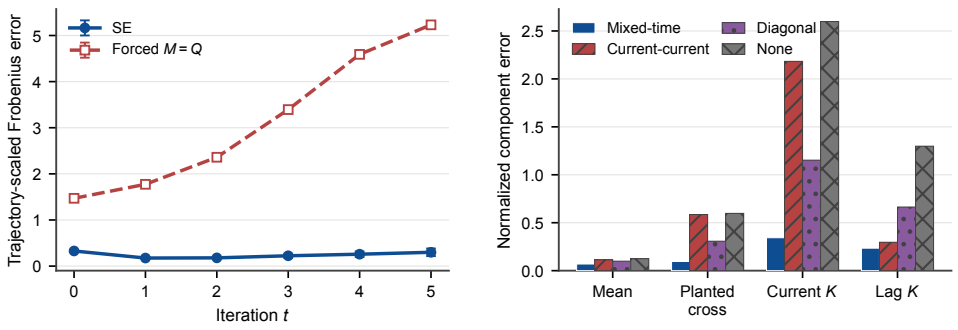

Figure 4: Supplementary off-Nishimori separation and reaction diagnostics. Left: SE versus the deliberately misspecified forced-$\mathbf{M} = \mathbf{Q}$ baseline. Right: the four normalized residual components for the reaction variants in Figure 3.

Figure 5 provides the pooled relative coefficient error for each seed:

$$\left\{ \frac{\sum_{C \in \mathcal{C}} \left\| \widehat{C} - C \right\|_{\mathrm{F}}^2}{\sum_{C \in \mathcal{C}} \|C\|_{\mathrm{F}}^2} \right\}^{1/2},$$

where $\mathcal{C}$ is the finite collection of SE coefficient blocks being pooled and $\widehat{C}$ is the empirical counterpart of C. For $\mathbf{K}$, $\mathcal{C}$ contains all modes and $0 \le s \le t \le T$. And for $\mathbf{\Omega}$, it contains all modes and $1 \le t \le T$. The off-diagonal curves apply the same formula after zeroing every diagonal. Pooling before division prevents a single nearly zero target entry from dominating the scale.

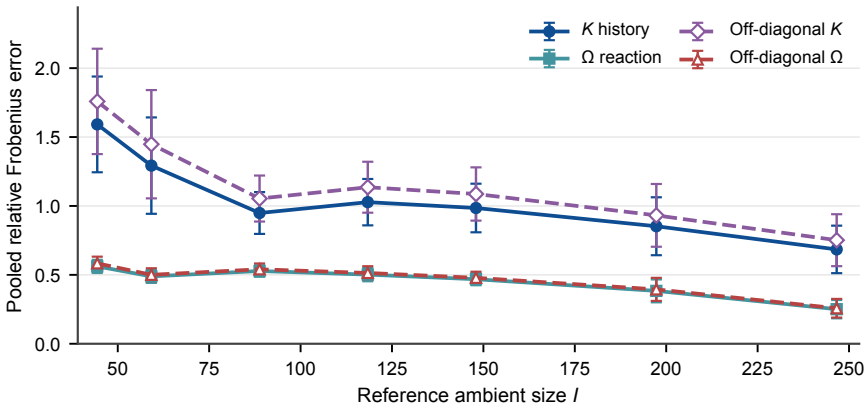

Figure 5: Strong-core pooled coefficient scaling at $\gamma = 1$, the largest mixing value in the experimental family. Curves show the all-entry and off-diagonal errors for $\mathbf{K}$ and $\mathbf{\Omega}$ across seven ambient sizes at $T = 5$. Bars indicate one standard error over 30 seeds for the five smaller sizes and eight seeds for the two larger sizes.

At the largest reaction size, averaging within each mode over iterations $t = 1, \ldots, 5$ and 30 realizations gives residual-moment RMS values $0.242, 0.229, 0.219$ for the mixed-time reaction and $0.780, 0.713, 0.634$ for the best altered reaction, which is the diagonal variant for this diagnostic. At the largest non-diagonal-core size, the pooled relative history-overlap errors for $\gamma = 0.2, 0.7, 1$ are 0.203, 0.285, and 0.501, respectively. At $\gamma = 1$, the largest-size pooled relative errors for $\mathbf{K}$, $\mathbf{\Omega}$, off-diagonal $\mathbf{K}$, and off-diagonal $\mathbf{\Omega}$ are respectively 0.985, 0.468, 1.087, and 0.478. At the two extended sizes, the same four errors are respectively $0.853/0.384/0.932/0.395$ and $0.685/0.251/0.752/0.258$. The pooled $\mathbf{Q}$-history and normalized residual-moment RMS also decrease from 0.535 to 0.333 and from 0.578 to 0.385. Together, these trends provide finite-size evidence that both the trajectory overlaps and the exact reaction coefficients approach their SE values, while the strongest core has the largest history-overlap errors on the common size grid.

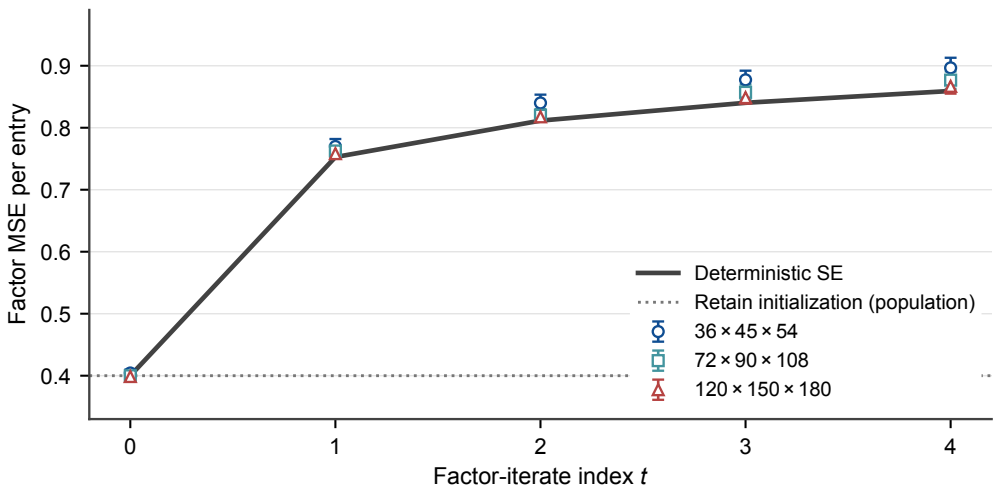

Figure 6: Fixed-horizon factor-MSE prediction for the unequal-channel Gaussian updates in Table 4. With $T = 3$, the displayed iterates are $t = 0, \dots, 4$. Markers show empirical means and one standard error over 30 realizations per size. The solid curve is the deterministic SE prediction. The dotted line is the population risk of retaining the initialization.

H.2 Unequal-channel covariance-dependent updates

For the unequal-channel study in Table 4, the core is obtained by reshaping $(\sin k + 0.35\cos(2k))_{k=1}^{24}$ into a $2 \times 3 \times 4$ array in column-major order, then rescaling it to squared Frobenius norm 2.5.

The matched Gaussian side-channel initialization is

$$\mathbf{X}_n^0 = q_{\text{init}}\mathbf{A}_n + \sqrt{q_{\text{init}}(1 - q_{\text{init}})}\,\mathbf{E}_{n,\text{init}}, \tag{168}$$

where the initialization-noise entries are independent standard Gaussians, independent of the planted arrays and observation noise. The parameter q_{init} is the initial signal overlap: $\mathbf{M}_n^0 = \mathbf{Q}_n^{0,0} = q_{\text{init}}\mathbf{I}_{P_n}$ in SE. The row map in Table 4 is the Gaussian posterior mean for the matched effective channel, as derived in Appendix G.3. It depends on the covariance input and has the required bounded-set regularity on the entire positive- semidefinite cone. The finite recursion uses its empirical covariance inputs and exact mixed-time Onsager reaction, with synchronous updates and no damping or parameter learning.

The seed list is shared across sizes, and comparisons between sizes are therefore paired. Since the row maps are linear at fixed SE coefficients, Gaussian second-moment identities evaluate the full SE history without Monte Carlo sampling. Two independent 120,000-row-per-mode MC-SE estimates are also computed for comparison. Table 2 uses the deterministic moment recursion as its reference.

The pooled relative Frobenius errors use the following time ranges. Signal and iterate overlaps include the final iterate: $0 \le t \le T + 1$ for $\mathbf{M}$ and $0 \le s \le t \le T + 1$ for $\mathbf{Q}$. Covariance blocks use $0 \le s \le t \le T$, average Jacobians use $1 \le t \le T + 1$, and reactions use $1 \le t \le T$. Here $\mathbf{K}_{n,I}^{s,t}$ is computed from the iterate overlaps, not from sample residual covariances.

Figure 6 complements the coefficient and overlap errors in Table 2 with the complete factor-MSE trajectory and its deterministic SE prediction. At the largest size, the mean final-factor MSE is 0.867 ± 0.012, compared with the SE prediction 0.859 and initial population MSE 0.4. Bayes matching here refers to the current effective Gaussian channel. The agreement shows that SE also predicts deterioration from an informative initialization for this particular update rule.

H.3 Exact factorized implementation of the Onsager reaction

Proposition 4.1 gives an exact finite-system factorization of the row-explicit chain-rule reaction into cross-time iterate overlaps and mode-averaged denoiser Jacobians. We verify this identity numerically

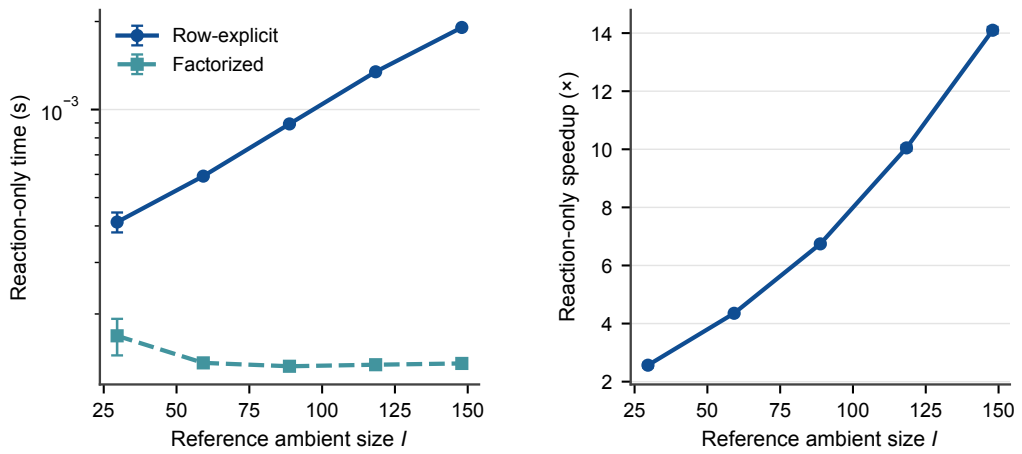

Figure 7: Exact factorization of the fully observed chain-rule Onsager reaction. Left: reaction-only runtime. Right: row-explicit-to-factorized speedup. The row-explicit reference materializes the Cartesian ambient-index contraction, whereas the factorized path uses cross-time overlaps and mode-averaged denoiser Jacobians. Points show means and bars show one standard error over twelve independent realizations per dimension tuple. Reaction-only timings exclude field contraction, denoising, and data generation.

and compare its two implementations using the exact Rademacher posterior-mean denoiser

$$\boldsymbol{f}(\boldsymbol{b};\mathbf{K}) = \mathbb{E}_q[\boldsymbol{a} \mid \boldsymbol{b}], \qquad q(\boldsymbol{a} \mid \boldsymbol{b}) \propto \exp\left\{\boldsymbol{b}\boldsymbol{a}^\top - \frac{1}{2}\boldsymbol{a}\mathbf{K}\boldsymbol{a}^\top\right\}, \quad \boldsymbol{a} \in \{-1,1\}^2.$$

Its row Jacobian is the posterior covariance $\mathrm{Cov}_q(\boldsymbol{a} \mid \boldsymbol{b})$, which generally varies with $\boldsymbol{b}$.

The comparison uses $N = 3$, $P_n = 2$, the non-diagonal core with $\gamma = 0.7$, $\nu = 0.6$, and the asymmetric dimension tuples

$$\{(24,30,36),(48,60,72),(72,90,108),(96,120,144),(120,150,180)\}.$$

Because the identity holds for arbitrary finite-system iterates and rowwise Jacobians, the benchmark isolates the reaction calculation from field contraction and denoising. Paired correlated Gaussian iterate arrays generate common inputs, and both implementations receive the same fields and rowwise Jacobians. The row-explicit reference materializes the Cartesian ambient-index contributions in (34). The factorized path forms the cross-time iterate overlaps and average denoiser Jacobians before evaluating (12).

For fixed N and channel dimensions, the row-explicit contraction costs $O(\sum_n I_{\setminus n}) = O(I^{N-1})$ under proportional growth. Forming the overlaps and average Jacobians for the factorized path costs $O(\sum_m I_m P_m^2) = O(I)$.

To quantify equality and nontrivial row dependence, define $\overline{\mathbf{J}}_m = I_m^{-1} \sum_{i_m=1}^{I_m} \mathbf{J}_{m,i_m}$ and report

$$e_\Omega = \left(\frac{\sum_n \|\boldsymbol{\Omega}_n^{\mathrm{row}} - \boldsymbol{\Omega}_n^{\mathrm{fact}}\|_{\mathrm{F}}^2}{\sum_n \|\boldsymbol{\Omega}_n^{\mathrm{fact}}\|_{\mathrm{F}}^2}\right)^{1/2}, \qquad d_J = \left(\frac{\sum_m \sum_{i_m} \|\mathbf{J}_{m,i_m} - \overline{\mathbf{J}}_m\|_{\mathrm{F}}^2}{\sum_m \sum_{i_m} \|\mathbf{J}_{m,i_m}\|_{\mathrm{F}}^2}\right)^{1/2}.$$

Figure 7 shows twelve independent realizations per dimension tuple, with execution order alternated after warmup. Across all runs, the maximum reaction discrepancy is $e_\Omega = 1.65 \times 10^{-15}$, while the minimum pooled relative within-mode Jacobian dispersion is $d_J = 0.451$. At the largest dimension tuple, the mean and median reaction-only speedups are $14.10\times$ and $14.08\times$, respectively. At the largest size, storing the mode-averaged Jacobians uses 12 entries, compared with 1800 for all rowwise Jacobians.

Shared settings for Table 5. The representation, terminal-readout, and frozen-denoiser studies in Table 5 share $N = 3$, $P_n = 2$, the non-diagonal core with $\gamma = 0.7$, and the exact mixed-time reaction. Their initialization is $\mathbf{X}_n^0 = \mathbf{A}_n\mathbf{P}_r + 0.5\mathbf{E}_{n,\mathrm{init}}$, with $\mathbf{P}_r$ as specified above and standard Gaussian initialization noise independent of the planted factors and observation noise.

Table 5: Configurations for representation calibration, terminal history readouts, and frozen-denoiser ranking. S, M, and L denote ambient dimensions $(48, 60, 72)$, $(72, 90, 108)$, and $(120, 150, 180)$, respectively. Finite-system repetitions are per noise–size setting. MC-SE sample sizes are rows per mode, and multipliers count independent samples.

Study	Noise, sizes, horizon	Base row map	Finite runs	MC-SE samples
Representation calibration	$\nu = 0.5$; $\mathcal{I}_5$; $T = 5$	$1.35\tanh(\boldsymbol{b}\mathbf{W}_r)$	12	Reference: 120,000
Fixed-base-map readouts	$\nu = 0.5, 0.2, 0.1, 0.05$; S/M/L; $T = 3$	Fixed map (169)	30	Fit: 100,000 Select: $2 \times 65{,}536$ Validate: $2 \times 120{,}000$
Noise-specific readouts	$\nu \in \{0.5, 0.6\}$; S/M/L; $T = 3$	Selected maps, one per noise level	30	Same sample sizes as fixed-base-map readouts, with disjoint samples
Frozen-denoiser ranking	$\nu \in \{0.5, 0.6\}$; S/M/L; $T = 3$	64 frozen $\kappa\tanh(\boldsymbol{b}\mathbf{W})$ maps	8	Score: $3 \times 65{,}536$

H.4 Representation calibration

For the representation study in Table 5, the probe task uses $\boldsymbol{w}_\star = (0.8, -0.6)^\top$ and $\sigma_\varepsilon^2 = 0.25$ in every mode, and no averaging over label directions is performed. The top-CCA error is absolute because its target can be zero and correlations already lie on a fixed scale. Probe-risk error is relative to the SE magnitude, which is at least the label-noise variance 0.25. A spectral threshold of 10^{-6} marks either quantity invalid when its nondegeneracy condition fails.

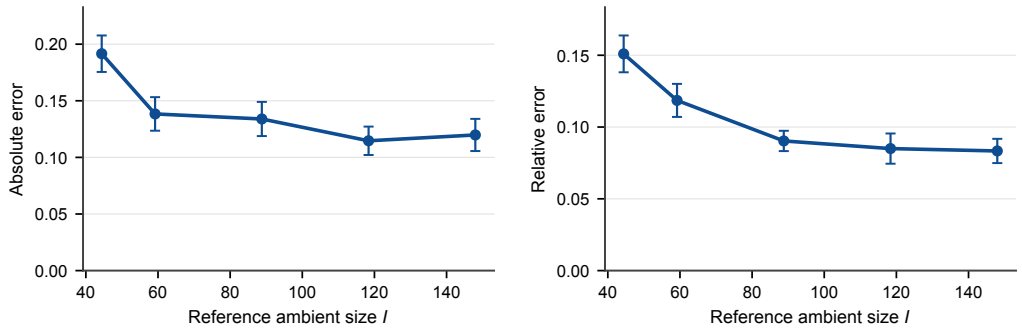

Figure 8: Empirical–SE calibration of representation observables under the exact covariance-derivative reaction. Left: absolute error in the top canonical correlation. Right: relative error in the population probe risk. For this experiment, $N = 3$, $P_n = 2$, $\gamma = 0.7$, $\nu = 0.5$, and $T = 5$. Each point first averages the valid error over modes and iterations within a finite-system realization and then over 12 seeds; bars show one standard error across seeds. The SE reference uses 120,000 rows per mode. Both endpoint errors decrease over the displayed size range.

Figure 8 tests the two main representation consequences of Corollary F.2. From the smallest to largest size, the empirical–SE errors in top canonical correlation and population probe risk decrease from 0.192 and 0.151 to 0.120 and 0.0833.

H.5 SE-CALIBRATED TERMINAL HISTORY READOUTS

We detail the fixed-base-map experiment of Section 6.2, then compare readouts under additional noise-specific base maps.

H.5.1 BASE RECURSION AND CALIBRATION

The readout configurations in Table 5 use the history $X^0, \dots, X^4$ and $B^0, \dots, B^3$. The fixed-base-map experiment uses

$$f(\boldsymbol{b}, \mathbf{S}) = 0.86127937 \tanh\left(\boldsymbol{b} \begin{bmatrix} -0.28597478 & -0.33447121 \\ 0.73263150 & 0.81488320 \end{bmatrix}\right), \tag{169}$$

with parameters displayed to eight decimal places. This untrained map is shared across modes and iterations. Only the noise coefficient varies in the fixed-base-map AMP recursion. Each ν remains fixed across ambient sizes, preserving the proportional-growth scaling. For the standard Gaussian prior, the ratio of expected signal and noise energies is $3.47/(\nu I^2)$, approximately $-35.0, -31.0, -28.0, -25.0$ dB at the largest size for $\nu = 0.5, 0.2, 0.1, 0.05$, respectively.

For each feature set X^0, $[X^0, B^0]$, and $[X^0, B^{0:3}]$, we fit (25) at $\lambda \in \{0.001, 0.01, 0.1, 1\}$. At each known noise level, independent SE samples of the sizes in Table 5 are used for fitting, selection, and validation. The fitting sample estimates the readout moments. The two selection samples score the twelve readouts by factor MSE. Their average selects a ridge value within each feature set and an overall readout, with ties resolved by feature order and ascending ridge. The two validation samples give held-out loss predictions. The fitted matrices are frozen across all three sizes before finite-system evaluation, paired across readouts, noises, and sizes. Finite-system planted factors are used only for evaluation.

Besides the raw initialization and final iterate, the principal baseline is the analytic Gaussian estimator based on X^0 alone:

$$\widehat{\mathbf{A}}_{n,\mathrm{init}} = \mathbf{X}_n^0 (\mathbf{P}_r^\top \mathbf{P}_r + 0.25 \mathbf{I}_2)^{-1} \mathbf{P}_r^\top. \tag{170}$$

Its population factor MSE per entry is $\frac{1}{2}\operatorname{Tr}[(\mathbf{I}_2 + 4\mathbf{P}_r\mathbf{P}_r^\top)^{-1}] = 0.48150$, compared with 0.54125 for retaining X^0. The tables instead use same-realization empirical baselines.

Factor MSE is the fitting and selection objective. Tensor NMSE is evaluated separately:

$$\frac{\left\| \widehat{\boldsymbol{\mathcal{Z}}} - \boldsymbol{\mathcal{Z}} \right\|_{\mathrm{F}}^2}{\|\boldsymbol{\mathcal{Z}}\|_{\mathrm{F}}^2}, \qquad [\widehat{\boldsymbol{\mathcal{Z}}}]_{\boldsymbol{i}} = \sum_{\boldsymbol{p}} \boldsymbol{\mathcal{C}}_{\boldsymbol{p}} \prod_{n=1}^{N} [\widehat{\mathbf{A}}_n]_{i_n p_n}. \tag{171}$$

We compute this ratio per realization before averaging, using the finite factorization in Appendix C for the readout factors.

H.5.2 FIXED-BASE-MAP NOISE SWEEP

The finite-system evaluation samples are independent of those used for fitting and selection. Each of the 360 trajectories is computed once and evaluated under all frozen readouts. Table 6 reports the factor losses. Every noise level selects $\lambda = 0.001$ within the first-field and full-history groups, with full history selected overall.

At the largest size, full-history factor MSE is $0.334, 0.202, 0.136, 0.091$ for $\nu = 0.5, 0.2, 0.1, 0.05$, respectively, below the calibrated-initialization MSE of 0.479 throughout. The improvements are 57.8%, 71.7%, and 81.1% at the three lower noise levels. Paired full-history minus first-field mean differences at those noises are $-0.0685, -0.0628, -0.0473$, with one-sided 95% Student-t upper limits $-0.0627, -0.0594, -0.0445$. At the same settings, full-history readouts also reduce tensor NMSE relative to both baselines (Table 7).

H.5.3 ADDITIONAL BASE-MAP COMPARISONS

The noise-specific readout experiments in Table 5 use base maps selected in Appendix H.6. They use the same calibration procedure with disjoint samples. First-field readout improves factor MSE over calibrated initialization at all six settings. Full history adds factor information at the largest size, but does not consistently improve tensor reconstruction.

Table 6: Factor MSE for the fixed-base-map noise sweep. Empirical entries are means with one standard error over all 30 realizations. The last column is the independent held-out SE prediction for full history.

ν	Size	Calibrated X^0	Raw X^4	First field	Full history	History SE MSE
0.5	S	0.485 ± 0.007	0.869 ± 0.022	0.367 ± 0.006	0.512 ± 0.075	0.310
0.5	M	0.481 ± 0.005	0.896 ± 0.018	0.362 ± 0.004	0.400 ± 0.043	0.310
0.5	L	0.479 ± 0.003	0.893 ± 0.015	0.360 ± 0.003	0.334 ± 0.005	0.310
0.2	S	0.485 ± 0.007	0.808 ± 0.018	0.287 ± 0.006	0.266 ± 0.014	0.184
0.2	M	0.481 ± 0.005	0.794 ± 0.010	0.278 ± 0.004	0.221 ± 0.010	0.184
0.2	L	0.479 ± 0.003	0.774 ± 0.007	0.271 ± 0.003	0.202 ± 0.005	0.184
0.1	S	0.485 ± 0.007	0.786 ± 0.012	0.230 ± 0.005	0.176 ± 0.007	0.121
0.1	M	0.481 ± 0.005	0.798 ± 0.008	0.213 ± 0.006	0.149 ± 0.006	0.121
0.1	L	0.479 ± 0.003	0.791 ± 0.006	0.198 ± 0.003	0.136 ± 0.003	0.121
0.05	S	0.485 ± 0.007	0.792 ± 0.012	0.186 ± 0.007	0.125 ± 0.006	0.075
0.05	M	0.481 ± 0.005	0.804 ± 0.008	0.161 ± 0.008	0.106 ± 0.006	0.075
0.05	L	0.479 ± 0.003	0.797 ± 0.006	0.138 ± 0.002	0.091 ± 0.002	0.075

Table 7: Tensor NMSE in the same noise sweep, computed by (171). Entries are means with one standard error over all 30 realizations. The final column gives the full-history SE prediction.

ν	Size	Calibrated X^0	First field	Full history	History SE NMSE
0.5	S	0.875 ± 0.006	0.744 ± 0.015	16.116 ± 8.424	0.647
0.5	M	0.867 ± 0.004	0.739 ± 0.008	5.300 ± 3.455	0.647
0.5	L	0.867 ± 0.002	0.739 ± 0.007	0.859 ± 0.069	0.647
0.2	S	0.875 ± 0.006	0.674 ± 0.020	1.222 ± 0.323	0.385
0.2	M	0.867 ± 0.004	0.648 ± 0.035	0.706 ± 0.160	0.385
0.2	L	0.867 ± 0.002	0.608 ± 0.009	0.484 ± 0.017	0.385
0.1	S	0.875 ± 0.006	0.724 ± 0.079	0.722 ± 0.149	0.254
0.1	M	0.867 ± 0.004	0.616 ± 0.108	0.440 ± 0.077	0.254
0.1	L	0.867 ± 0.002	0.485 ± 0.010	0.324 ± 0.013	0.254
0.05	S	0.875 ± 0.006	0.909 ± 0.201	0.561 ± 0.117	0.159
0.05	M	0.867 ± 0.004	0.643 ± 0.219	0.339 ± 0.063	0.159
0.05	L	0.867 ± 0.002	0.377 ± 0.013	0.231 ± 0.011	0.159

H.5.4 Finite-size prediction and selection errors

For the fixed-base-map sweep, the smallest-size full-history factor-MSE prediction gaps are $0.2014, 0.0820, 0.0547, 0.0504$ across $\nu = 0.5, 0.2, 0.1, 0.05$. At $\nu = 0.5$, the middle-size gap is 0.0899. First-field readout also has a 0.0693 gap at the smallest size and $\nu = 0.05$. Estimation gains therefore need not imply small SE prediction errors at these sizes.

The selected readout has zero observed regret among the twelve fitted candidates at all six middle/largest lower-noise settings. Smallest-size regrets are $39.43\%, 1.54\%, 0.188\%, 0\%$ across the four noise levels, and the $\nu = 0.5$ middle/largest regrets are 10.62% and 0%. These sample-mean comparisons do not establish a population uniqueness gap.

Tensor NMSE can be substantially larger despite moderate factor MSE. At $\nu = 0.5$, the fixed-base full-history readout has tensor NMSE 16.116 and 5.300 at the smallest and middle sizes. At $\nu = 0.2$ and the smallest size, full-history tensor NMSE also exceeds the calibrated baseline. At $\nu = 0.05$ and the smallest size, first-field tensor NMSE exceeds that baseline. The additional noise-specific base-map comparisons show still larger full-history tensor errors (Table 8). For a frozen linear readout, factor loss is a linear combination of empirical second moments of the planted rows and history features. Moment discrepancies therefore affect the loss prediction, with sensitivity set by the readout weights. Tensor reconstruction multiplies readout factors across modes and can further amplify their

Table 8: Additional noise-specific base-map comparisons. Entries are means with one standard error over 30 realizations. The four central columns provide factor MSE, and the last provides tensor NMSE for the full-history readout.

ν	Size	Calibrated X^0	Raw X^4	First field	Full history	History tensor NMSE
0.5	S	0.481 ± 0.008	0.832 ± 0.023	0.361 ± 0.006	0.604 ± 0.137	49.003 ± 33.818
0.5	M	0.483 ± 0.005	0.860 ± 0.024	0.361 ± 0.005	0.397 ± 0.032	4.351 ± 2.238
0.5	L	0.489 ± 0.005	0.906 ± 0.016	0.372 ± 0.004	0.347 ± 0.008	0.997 ± 0.219
0.6	S	0.481 ± 0.008	0.899 ± 0.017	0.374 ± 0.006	0.680 ± 0.190	94.396 ± 76.537
0.6	M	0.483 ± 0.005	0.931 ± 0.015	0.376 ± 0.005	0.392 ± 0.025	3.123 ± 1.481
0.6	L	0.489 ± 0.005	0.970 ± 0.009	0.387 ± 0.004	0.355 ± 0.006	0.906 ± 0.126

fluctuations. Quantifying these effects requires per-realization, modewise analysis. Calibration uses the known core, noise coefficient, planted-row law, and informative initialization law, so these results concern model-informed estimation from a fixed-horizon trajectory.

Independent held-out SE samples differ by at most 0.0043 in factor MSE in the fixed-base-map sweep.

H.6 Frozen-denoiser surrogate-ranking settings

For candidate j in the frozen-denoiser study of Table 5, the ranked finite-system and SE population losses are

$$\mathrm{MSE}_{I,j} = \frac{\sum_{n=1}^{N} I_n^{-1} \left\| \mathbf{X}_{n,j}^{T+1} - \mathbf{A}_n \right\|_{\mathrm{F}}^2}{\sum_{n=1}^{N} P_n}, \qquad \mathrm{MSE}_{\mathrm{SE},j} = \frac{\sum_{n=1}^{N} \mathbb{E} \left\| \boldsymbol{x}_{n,j}^{T+1} - \boldsymbol{a}_n \right\|_2^2}{\sum_{n=1}^{N} P_n}. \tag{172}$$

The panel consists of 64 randomly sampled, untrained row maps of the form $\kappa \tanh(\boldsymbol{b}\mathbf{W})$, with i.i.d. uniform entries of $\mathbf{W}$ on $[-1, 1]$ and κ drawn uniformly from $[0.55, 1.75]$. The complete panel and its ordering are frozen before any loss is evaluated. Independent MC-SE samples of the sizes in Table 5 score each candidate. The three MC scores are averaged before selection. The smallest average selects one candidate per noise level, shared across the three sizes. The $\nu = 0.6$ setting uses the same panel and protocol with only the noise coefficient changed from $\nu = 0.5$.

Figure 9 and Table 9 show calibration and selection for this fixed panel across all six size–noise settings. These losses concern the raw final iterates, distinct from the terminal readouts above.

The population per-entry MSEs of retaining the initialization and using the zero estimator are, respectively,

$$L_{\mathrm{init}} = \tfrac{1}{2} \|\mathbf{P}_r - \mathbf{I}_2\|_{\mathrm{F}}^2 + 0.25 = 0.54125, \qquad L_{\mathrm{zero}} = 1.$$

At $\nu = 0.5$, MC-SE selects a map whose mean finite-system losses at the three increasing sizes are 0.8503, 0.8224, and 0.8221. At $\nu = 0.6$, the selected map has losses 0.9558, 0.9481, and 0.9525.

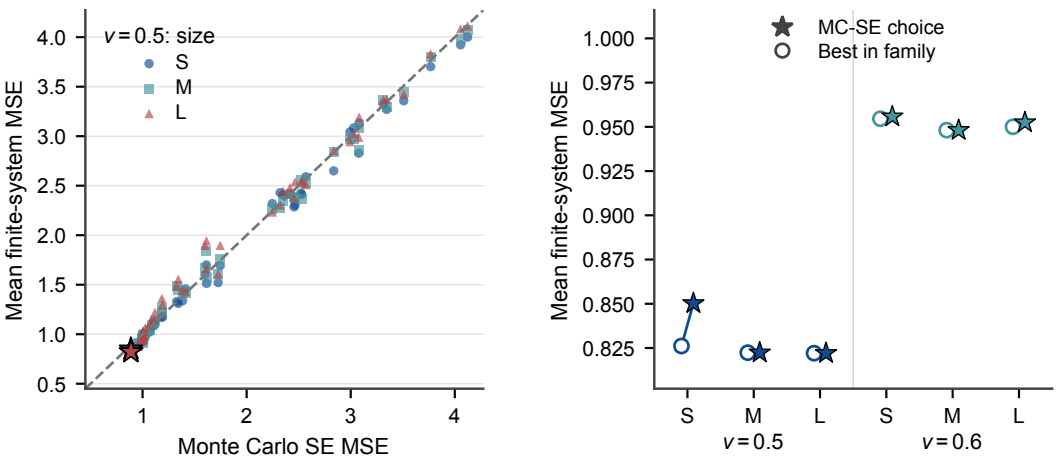

Figure 9: Calibration and within-family selection. Left: $\nu = 0.5$ calibration with MC-SE choices starred. Right: selected and best-in-family mean losses at both noise levels. S/M/L denote the three increasing sizes.

Table 9: Frozen-row-map surrogate evaluation. Rank is the finite-system rank of the candidate minimizing the mean MC-SE selection MSE, and regret is its relative mean finite-system loss above the best candidate under the same size–noise setting. Zero means no excess loss is observed under this sample-mean benchmark.

ν	Ambient dimensions	Pearson	Spearman	Top-8 overlap	Rank of MC-SE choice	Observed regret (%)
0.5	$(48, 60, 72)$	0.9975	0.9764	6/8	3	2.92
0.5	$(72, 90, 108)$	0.9977	0.9871	7/8	1	0
0.5	$(120, 150, 180)$	0.9960	0.9743	6/8	1	0
0.6	$(48, 60, 72)$	0.9774	0.9755	6/8	2	0.124
0.6	$(72, 90, 108)$	0.9984	0.9787	6/8	1	0
0.6	$(120, 150, 180)$	0.9975	0.9769	6/8	2	0.257

H.7 One-step state evolution near zero overlap

We evaluate one population update from Corollary G.3. Take $N = 3$, $P_n = 2$, mode ratios 4:5:6, $\nu = 0.6$, the non-diagonal core with $\gamma = 0.7$, and the row map $f(\boldsymbol{b}, \mathbf{K}) = \tanh(\boldsymbol{b}\mathbf{W}_g)$ from the non-diagonal-core scaling study. Set $\mathbf{H} = \mathbf{I}_2/\sqrt{2}$ and, for each $\epsilon_0 \in (0, 0.3]$, set

$$\mathbf{M}_n^0 = \epsilon_0 \mathbf{H}, \qquad \mathbf{Q}_n^{0,0} = \mathbf{I}_2, \qquad n = 1, 2, 3. \tag{173}$$

This is an admissible family: for independent $\boldsymbol{a}_n, \boldsymbol{z}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$, the row initialization

$$\boldsymbol{x}_n^0 = \epsilon_0 \boldsymbol{a}_n \mathbf{H} + \boldsymbol{z}_n \big(\mathbf{I}_2 - \epsilon_0^2 \mathbf{H}^\top \mathbf{H}\big)^{1/2} \tag{174}$$

has exactly the moments in (173). Along this family, $\mathbf{K}_n^{0,0} = \mathcal{K}_n(\mathbf{I}_2, \mathbf{I}_2, \mathbf{I}_2)$ is independent of ϵ_0, with $\min_n \lambda_{\min}(\mathbf{K}_n^{0,0}) = 1.361$.

For $N = 3$, $\mathcal{K}_n$ is bilinear in the two complementary overlap arguments, so $\mathbf{\Theta}_n^0(\epsilon_0) = \epsilon_0^2 \mathcal{K}_n(\mathbf{H}, \mathbf{H}, \mathbf{H})$. Gaussian integration by parts then evaluates the next overlap without a small-signal Monte Carlo subtraction:

$$\mathbf{M}_n^1(\epsilon_0) = \mathbf{\Theta}_n^0(\epsilon_0)\, \mathbb{E}[D_{\boldsymbol{b}} f(\boldsymbol{B}_n, \mathbf{K}_n^{0,0})], \quad \boldsymbol{B}_n \sim \mathcal{N}\Big(\mathbf{0}, \mathbf{K}_n^{0,0} + \mathbf{\Theta}_n^0(\epsilon_0)^\top \mathbf{\Theta}_n^0(\epsilon_0)\Big). \tag{175}$$

The scalar $\operatorname{sech}^2(x) = 1 - \tanh^2(x)$ terms in the Jacobian are computed by deterministic adaptive quadrature.

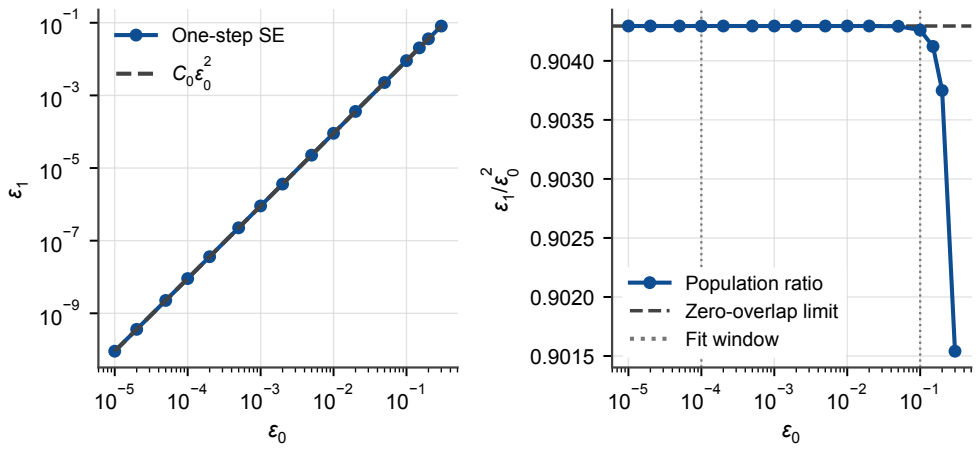

Figure 10: One-step SE illustration near zero overlap. Writing $\epsilon_1 = \max_n \left\|\mathbf{M}_n^1\right\|_{\mathrm{F}}$, the left panel compares the population map with its quadratic asymptotic guide. Right: ϵ_1/ϵ_0^2 and its zero-overlap limit $C_0 = 0.904$. Dotted lines mark the prespecified window $10^{-4} \le \epsilon_0 \le 10^{-1}$.

The log–log fit in the left panel of Figure 10 has slope 2.000, and $\lim_{\epsilon_0 \downarrow 0} \epsilon_1/\epsilon_0^2 = 0.904$. The ratio varies by 3.8×10^{-5} relative over the fit window. These values are consistent with the exponent $N - 1 = 2$ in (149) while the covariance remains nonzero. The corollary itself supplies an upper bound. The nonzero leading coefficient here uses the selected smooth denoiser and core.

H.8 Order-four tensor validation

To illustrate the theorem beyond order three, we run one compact $N = 4$ experiment with $P_n = 2$. The core has $\mathcal{C}_{1111} = \mathcal{C}_{2222} = 1$ and $\mathcal{C}_{1222} = \mathcal{C}_{2122} = \mathcal{C}_{2212} = \mathcal{C}_{2221} = 0.7$, with all other entries zero. We use the same coupled denoiser and initialization as the non-diagonal-core scaling experiment, $\nu = 0.6$, $T = 3$, eight seeds, and asymmetric ambient-dimension tuples $(24, 28, 32, 36)$, $(32, 38, 44, 50)$, and $(40, 48, 56, 64)$. Each tuple uses its own SE reference with aspect ratios I_n/I and coefficients $I/(\nu I_n)$. Each reference uses 120,000 rows per mode.

The separate one-step calculation fixes mode ratios 5:6:7:8, $\nu = 0.6$, and $\mathbf{Q}_n^{0,0} = \mathbf{I}_2$, $\mathbf{M}_n^0 = \epsilon_0 \mathbf{I}_2/\sqrt{2}$ for $n \in [4]$, using the same core and row map. Here $\epsilon_1 = \max_n \left\|\mathbf{M}_n^1\right\|_{\mathrm{F}}$ is the resulting signal-overlap amplitude.

In the bottom-left panel of Figure 11, the pooled residual-covariance history error decreases from 0.529 ± 0.083 to 0.371 ± 0.034, while the $\mathbf{\Omega}$ error decreases from 1.691 ± 0.660 to 0.476 ± 0.098. The one-step population map in the bottom-right panel of Figure 11 has fitted log–log slope 3.000, limiting ratio $\epsilon_1/\epsilon_0^3 = 0.697$, and $\min_n \lambda_{\min}(\mathbf{K}_n^{0,0}) = 1.501$. For this core and row map, the nonzero limiting ratio identifies a cubic leading term, consistent with the $N - 1 = 3$ exponent in (149).

H.9 Out-of-model AMP–ALS illustration on amino-acid fluorescence data

We also compare rank-three AMP and classical ALS (Kolda & Bader, 2009) on the amino-acid fluorescence data described by Bro (1997). We use the `amino.mat` file from the official data archive.[1] The dataset comprises a $5 \times 201 \times 61$ tensor with sample, emission, and excitation modes, together with a 5×3 concentration matrix for tryptophan, tyrosine, and phenylalanine. We use every tensor entry and normalize the measured reference tensor $\mathcal{Z}_{\mathrm{ref}}$ by its entrywise RMS. At each signal-to-noise ratio $\mathrm{SNR} \in \{10, 20, 30\}$ dB, the input is

$$\mathcal{Y} = \mathcal{Z}_{\mathrm{ref}} + 10^{-\mathrm{SNR}/20}\mathcal{G}, \qquad [\mathcal{G}]_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1). \tag{176}$$

This experiment compares decomposition performance on a real measured signal under synthetic additive noise.

[1] Data source: Amino acid fluorescence data archive.

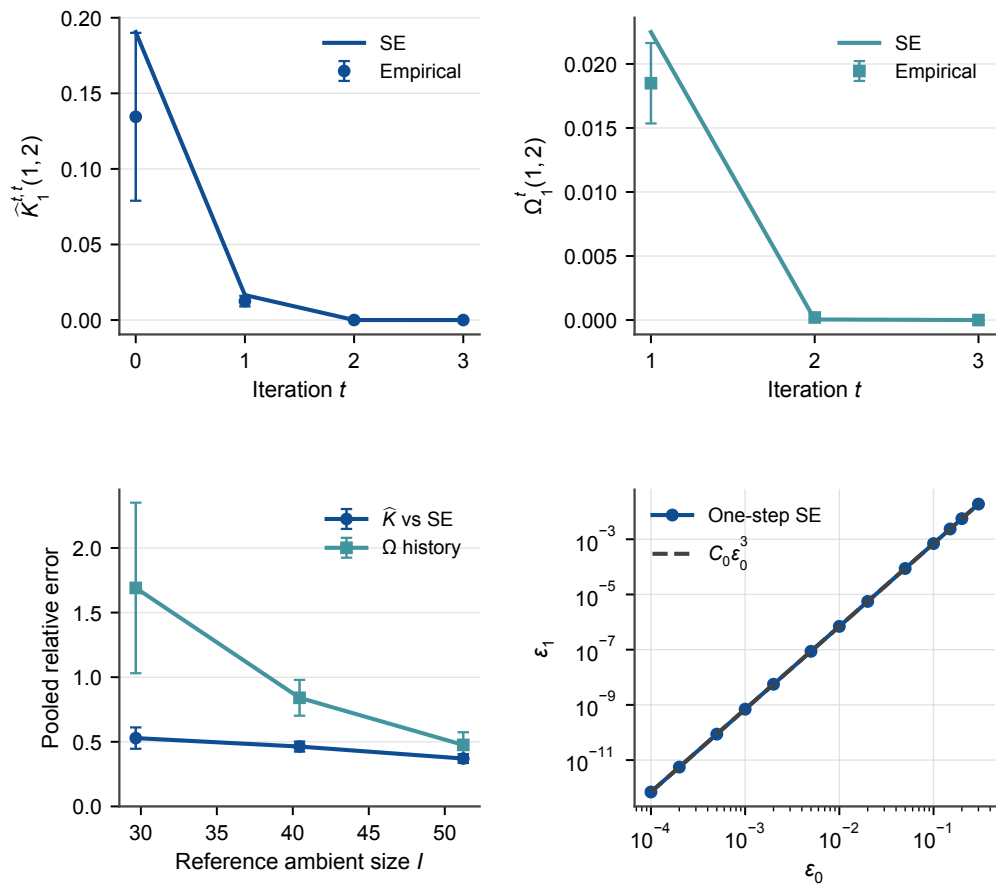

Figure 11: Order-four tensor validation. Top left: empirical residual covariance $\widehat{\mathbf{K}}_1^{t,t}(1,2)$ and its SE value. Top right: Onsager reaction $\mathbf{\Omega}_1^t(1,2)$ and its SE value. Both use the largest geometry, with one-standard-error bars over eight seeds. Bottom left: pooled full-history errors across three increasing rectangular geometries, each with its own aspect-ratio SE reference. Bottom right: one-step SE calculation at fixed ratios 5:6:7:8 with its cubic guide.

At each noise level, AMP and ALS are evaluated on 20 paired noisy tensors. Table 10 compares their algorithm settings. Each iteration is one complete sweep over the three modes. The ALS updates and normalization follow Tensor Toolbox (Bader & Kolda, 2006), version 2.5[2], with its fit-change stopping test replaced by the same reconstruction-change test used for AMP.

The AMP row update is $\boldsymbol{f}(\boldsymbol{b};\mathbf{K}) = \boldsymbol{b}(\mathbf{K}+\mathbf{\Lambda})^{-1}$. Write $\widehat{\mathbf{A}}_n^{(k)}$ for the mode-n factor mean after iteration k and $\mathbf{V}_n^{(k)}$ for its mode-shared approximate posterior row covariance. The implementation updates the diagonal precision entries as

$$[\mathbf{\Lambda}^{(k+1)}]_{pp} = \frac{\sum_{n=1}^{3} I_n}{\max\left\{\sum_{n=1}^{3}\left(\left\|[\widehat{\mathbf{A}}_n^{(k)}]_{\cdot,p}\right\|_2^2 + I_n[\mathbf{V}_n^{(k)}]_{pp}\right), \varepsilon_{\text{mach}}\right\}}, \qquad p \in [3], \tag{177}$$

where $\varepsilon_{\text{mach}}$ is machine precision. Both methods stop at the first iteration $k \geq 10$ satisfying

$$\frac{\|\widehat{\boldsymbol{\mathcal{Z}}}^{(k)} - \widehat{\boldsymbol{\mathcal{Z}}}^{(k-1)}\|_F^2}{\|\widehat{\boldsymbol{\mathcal{Z}}}^{(k-1)}\|_F^2 + 10^{-10}} \leq 10^{-6}, \tag{178}$$

[2] https://www.tensortoolbox.org/

Table 10: AMP–ALS settings for the fluorescence comparison. Both methods receive the same noise realization and paired initial factor directions, and use the same early stopping criterion.

Setting	AMP	ALS
Rank	Fixed rank-three diagonal CP core	Fixed rank three
Update	Gaussian row map with known added-noise variance; diagonal precision update (177)	Tensor Toolbox 2.5 ALS updates (Bader & Kolda, 2006) with the common stopping test
Initialization	Two Gaussian starts; $\mathbf{\Lambda}^{(1)} = \mathbf{I}_3$	Two paired Gaussian starts
Start selection	Smaller observed-tensor relative residual	Same criterion
Iteration limit	1000	1000
Early stopping	Relative squared reconstruction change (178), from iteration 10	Same criterion and starting iteration

Table 11: Amino-acid fluorescence results with a common early stopping criterion. Entries are mean $\pm$ one sample standard deviation over 20 paired noisy tensors. Runtime and iteration counts sum over both starts. NMSE and correlation describe the selected fit.

SNR (dB)	Method	NMSE	Matched correlation	Runtime (s)	Iterations
10	AMP	0.002007 ± 0.000079	0.99866 ± 0.00033	0.0315 ± 0.0072	69 ± 16
10	ALS	0.002045 ± 0.000080	0.99853 ± 0.00034	0.034 ± 0.010	67 ± 19
20	AMP	0.000854 ± 0.000032	0.99856 ± 0.00030	0.0298 ± 0.0095	68 ± 24
20	ALS	0.000866 ± 0.000026	0.99852 ± 0.00032	0.0340 ± 0.0098	70 ± 22
30	AMP	0.000757 ± 0.000021	0.99851 ± 0.00034	0.0299 ± 0.0074	69 ± 19
30	ALS	0.000756 ± 0.000021	0.99855 ± 0.00040	0.0350 ± 0.0096	73 ± 22

where $\widehat{\boldsymbol{\mathcal{Z}}}^{(k)}$ is the reconstruction in the RMS-normalized data coordinates. Both implementations evaluate this statistic using the same CP inner-product calculation, accounting for AMP's internal scaling. Runtime includes internal setup and stopping calculations for both starts, but excludes common data preparation and final evaluation metrics. The tensor NMSE is measured with respect to the unperturbed measured tensor $\boldsymbol{\mathcal{Z}}_{\mathrm{ref}}$:

$$\mathrm{NMSE}(\widehat{\boldsymbol{\mathcal{Z}}}, \boldsymbol{\mathcal{Z}}_{\mathrm{ref}}) = \frac{\|\widehat{\boldsymbol{\mathcal{Z}}} - \boldsymbol{\mathcal{Z}}_{\mathrm{ref}}\|_F^2}{\|\boldsymbol{\mathcal{Z}}_{\mathrm{ref}}\|_F^2}.$$

Write $\widehat{\mathbf{A}}_1 \in \mathbb{R}^{5\times 3}$ for the recovered sample-mode factor matrix and $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3 \in \mathbb{R}^5$ for the measured concentration columns. For each fit, the matched correlation is

$$\max_{\pi \in S_3} \frac{1}{3} \sum_{r=1}^{3} \left| \mathrm{corr}\left([\widehat{\mathbf{A}}_1]_{\cdot,\pi(r)}, \boldsymbol{c}_r \right) \right|,$$

where Pearson correlation is centered over the five samples and a constant recovered factor is assigned correlation zero. The maximization removes the component-permutation ambiguity, the absolute value removes the sign ambiguity, and correlation is invariant to nonzero componentwise rescaling. Table 11 summarizes the results.

All 240 candidate runs meet the shared stopping criterion. The paired mean differences in NMSE (AMP minus ALS) are -3.81×10^{-5}, -1.23×10^{-5}, and 6.58×10^{-7} at 10, 20, and 30 dB. At 30 dB, both methods have mean NMSE about 0.00076, compared with 0.00062743 for the rank-three ALS reference fitted directly to the unperturbed measured tensor $\boldsymbol{\mathcal{Z}}_{\mathrm{ref}}$. At 10, 20, and 30 dB, the ratios of mean ALS runtime to mean AMP runtime are 1.09, 1.14, and 1.17, respectively. The two implementations thus have similar iteration counts and a modest measured runtime difference under the common stopping rule. This rule controls successive reconstruction changes, while accuracy is assessed separately by the reported NMSE and correlations.

Beyond point reconstruction, this Bayesian AMP implementation maintains approximate posterior row covariances and uses posterior second moments to update the prior precisions in (177), functionality absent from the ALS baseline used here. These covariance estimates provide a basis for uncertainty quantification, whose calibration remains to be assessed.