TEMPORAL POOLING SHAPES COVARIANCE GEOMETRY AND READOUT RISK IN RANDOM RECURRENT NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Temporal pooling determines the covariance seen by a downstream learner, yet this covariance is often reduced to total variance or participation dimension. We study a stationary linear random recurrent system with stability gap $\varepsilon = 1 - g^2$. In the ordered limit where width tends to infinity first, $T\sqrt{\varepsilon} \to \infty$ and $T\varepsilon \to 0$ define an intermediate observation window. For a broad class of normalized pooling kernels, every fixed moment of the variance-carrying part of the spectrum converges, and its covariance-size-biased limit is $Y_h = (M/2)|\widehat{h}(U_h)|^2$. Here M has the aspect-one Marchenko–Pastur law, while the independent filter-gain factor depends on the kernel. Pooling rules can therefore match trace and participation dimension while retaining different higher-order geometry.

To test whether this distinction matters for a common prediction task, we study 56 independent real Ginibre networks at widths 128 and 256. For each network, a calibration based only on feature covariances matches the trace and second spectral moment of boxcar and causal-exponential pooling to numerical precision. Both pooled views then predict the same future recurrent state. At width 256 and $T = \varepsilon^{-3/4}$, causal-exponential pooling reduced population normalized mean-squared error by 5.86–7.14 percentage points across three stability gaps. Aligning the kernels' mean amplitude-weighted sample ages reduced these differences to 1.71–2.43 points but did not eliminate them. Validation-selected finite-sample ridge differences were 1.93–3.91 points in the same direction, although two width-128 fits reached the edge of the penalty grid, making that comparison descriptive. A separate Gaussian-design analysis shows why no kernel ranks universally: higher-order geometry raises ridge bias but can lower variance enough to reverse the ordering under minimum-norm regression. These finite experiments measure task behavior at fixed parameters; they do not validate the ordered asymptotic law.

1 INTRODUCTION

Spikes are binned, hidden states are averaged, and sensor streams are tapered before a learner sees them. These operations change the covariance of a recurrent representation and therefore the directions available to a linear readout. Total variance and participation dimension summarize only the first two spectral moments. We ask what they miss when two pooling rules are applied to the same recurrent system.

Non-normal recurrence amplifies covariance and changes effective activity dimension (Hennequin et al., 2012; Clark et al., 2023), while the zero-frequency limit follows a different critical law (Hu & Sompolinsky, 2022). Temporal windows are known to reshape frequency weighting (Harris, 1978), and learned temporal covariance pooling is already used in representation learning (Gao et al., 2021). The exact boxcar identity of Calvo et al. (2026) shows that coarse-graining changes apparent critical scaling, but does not determine the noncommuting covariance spectrum for a general pooling kernel near the stability edge.

We first take the large-width circular limit at fixed gain. We then let the critical gap $\varepsilon = 1 - g^2$ decrease and the pooling scale T grow. The two dimensionless products

$$T\sqrt{\varepsilon} \to \infty, \qquad T\varepsilon \to 0$$

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.

define an intermediate-window regime. The recurrent dynamics fixes the two boundaries, but the limiting law inside the regime retains the distribution of filter power across frequency.

Covariance geometry alone cannot rank representations for every task. If C_h is the covariance of pooled features and K_h their cross-covariance with a shared target, then prediction risk depends on both C_h and K_h. We therefore separate two questions. The theorem identifies the covariance law; the finite experiment measures the combined effect of covariance geometry and feature–target alignment when two lower-moment-matched representations predict the same future state in a real-valued recurrent network.

Our contributions are:

1. We prove a covariance law for every fixed moment order. Its limit factors into a universal Marchenko–Pastur variable and a kernel-specific filter gain, whose lower spectral tail we classify.
2. We derive population denoising and linear-readout curves, together with an ordered Gaussian sample-Gram law, and construct kernels with equal trace and participation dimension but different higher moments and regularization curves.
3. We test the distinction in 56 finite real Ginibre networks. A per-network calibration matches the first two spectral moments while both pooled views predict the same future state; a separate bias–variance analysis shows why ridge and minimum-norm regression can rank the kernels differently.
4. We prove the signed-kernel localization needed by the main theorem and show that stable finite networks separate the two observation-time boundaries only at an eighth-root rate in width.

Together, these results show why equal participation dimension need not imply equal recoverability, while the finite-width analysis quantifies the slow approach to the strict intermediate regime. The theorem uses the complex circular model, whereas the shared-task experiment uses finite real matrices; the two ensembles share the leading first-order $*$-moment limit (Redelmeier, 2014). Our conclusions do not extend to a uniform real-edge transfer, a moving-rank law, a single contiguous training trajectory, or a universally optimal pooling kernel.

2 Finite-window recurrent covariance

Let c be a standard circular element in a tracial noncommutative probability space $(\mathcal{A}, \tau)$ and let $0 < g < 1$. The operator c is the large-width limit of a complex Ginibre recurrent matrix with entry variance $1/N$. Consider the stationary Ornstein–Uhlenbeck system

$$dx_t = -Ax_t\,dt + dB_t, \qquad A = 1 - gc, \qquad \varepsilon = 1 - g^2, \tag{1}$$

where the complex Brownian noise is normalized by $\mathbb{E}[dB_t\,dB_t^*] = 1\,dt$. For a real pooling kernel h, define

$$z_{T,h} = \int_{\mathbb{R}} T^{-1/2} h(s/T) x_s\,ds, \qquad C_{T,h} = \mathrm{Cov}(z_{T,h}), \tag{2}$$

where $\|h\|_2 = 1$. The boxcar choice $h_{\mathrm{box}} = \mathbf{1}_{[0,1]}$ gives the unit-energy interval integral used in earlier finite-window analyses. A temporal translation of h changes only Fourier phase and therefore leaves $C_{T,h}$ unchanged.

With the Fourier convention $\widehat{h}(u) = \int h(t)e^{-iut}dt$, write

$$\psi_h(u) = |\widehat{h}(u)|^2, \qquad Q_\omega = [(1 + i\omega) - gc]^{-1}[(1 - i\omega) - gc^*]^{-1}.$$

Stationarity and linear filtering give

$$C_{T,h} = \int_{\mathbb{R}} T\psi_h(T\omega) Q_\omega\,\frac{d\omega}{2\pi}. \tag{3}$$

Although each Q_ω is positive, Q_ω and $Q_{\omega'}$ generally do not commute. Thus the spectrum of $C_{T,h}$ is not an eigenvaluewise average of single-frequency spectra.

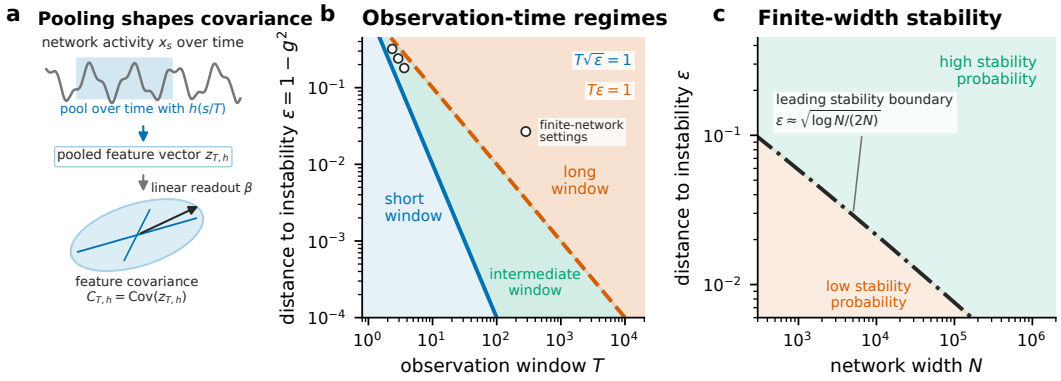

Figure 1: **Temporal pooling selects a covariance regime. a**, A normalized kernel combines network activity over time into the feature vector $z_{T,h}$. Its covariance $C_{T,h}$ determines the feature directions available to a linear predictor. **b**, With $\varepsilon = 1 - g^2$ measuring distance to instability, $T\sqrt{\varepsilon}$ and $T\varepsilon$ divide the ordered large-width limit into short, intermediate, and long observation windows. Open circles mark the finite-network settings used in Figure 3: $\varepsilon \in \{0.18, 0.24, 0.32\}$ and $T = \varepsilon^{-3/4}$. Their two-sided separations $\mathcal{S} = 1.33$–1.54 are not asymptotically large. **c**, A finite complex Ginibre network also requires ε above the leading stability boundary $\sqrt{\log N/(2N)}$ for high stability probability. Panel c is a feasibility boundary, not finite-width validation of panel b.

Let $\mu^{(h)}_{\varepsilon,T}$ denote the spectral probability measure of $C_{T,h}$. In the intermediate-window regime, define a locally finite measure on $(0, \infty)$ by

$$\nu^{(h)}_{\varepsilon,T}(B) = \frac{T}{\varepsilon}\, \mu^{(h)}_{\varepsilon,T}\left\{\lambda : \frac{\varepsilon^2}{T}\lambda \in B\right\}. \tag{4}$$

It resolves the rare spectral mass of order ε/T at eigenvalue scale T/ε^2. Multiplying this measure by y produces a probability limit that emphasizes directions in proportion to the covariance they carry. This covariance-weighted object controls the normalized trace functionals used below; it is neither the ordinary empirical spectral distribution nor a moving-rank limit.

Figure 1 summarizes the construction, the two observation-time boundaries, and the separate finite-width stability constraint.

3 A POOLING-KERNEL COVARIANCE LAW

Let

$$\phi_h(v) = \int_{\mathbb{R}} h(s+v)h(s)\, ds$$

be the autocorrelation of the pooling kernel. We use a sufficient regularity class that covers boxcar, causal-exponential, triangular, Hann, and Gaussian pooling: $h \in L^1 \cap L^2$ is real, $\|h\|_2 = 1$, $\phi_h \in W^{1,1} \cap W^{1,\infty}$, and $\int |v||\phi_h(v)|dv < \infty$. These are sufficient conditions for the argument, not necessary conditions.

Theorem 1 (Intermediate-window pooling law)**.** *Let h belong to the class above. Take the circular limit at fixed $g < 1$, followed by $\varepsilon \downarrow 0$ and $T \to \infty$ such that $T\sqrt{\varepsilon} \to \infty$ and $T\varepsilon \to 0$. For every fixed integer $p \geq 1$,*

$$\int y^p \nu^{(h)}_{\varepsilon,T}(dy) \longrightarrow B_p(h), \qquad B_p(h) = \frac{\Gamma(2p-1)}{2^{p-1}p!(p-1)!} J_p(h), \tag{5}$$

where

$$J_p(h) = \frac{1}{2\pi}\int_{\mathbb{R}} \psi_h(u)^p du. \tag{6}$$

Equivalently,

$$\tau(C^p_{T,h}) \sim B_p(h) T^{p-1} \varepsilon^{-(2p-1)}. \tag{7}$$

Parseval's identity gives $J_1(h) = 1$, so $B_1(h) = 1$. The first two moments give the participation dimension

$$d_h = N\frac{\tau(C_{T,h})^2}{\tau(C_{T,h}^2)} \sim \frac{2N\varepsilon}{J_2(h)T}, \tag{8}$$

which is a two-moment proxy for the number of covariance directions carrying substantial variance. This is an ordered-limit formula, not a uniform finite-N theorem.

The full moment sequence has a direct probability-law interpretation. Let M follow the aspect-one Marchenko–Pastur law (Marčenko & Pastur, 1967). Independently, draw U_h with density $f_{U_h}(u) = \psi_h(u)/(2\pi)$ and set $W_h = \psi_h(U_h)$.

Corollary 1 (Product law and transform)**.** *The eigenvalue-size-biased measures $\rho_{\varepsilon,T}^{(h)}(dy) = y\nu_{\varepsilon,T}^{(h)}(dy)$ converge weakly to the law of*

$$Y_h = \frac{M}{2}W_h. \tag{9}$$

The limiting transform is

$$\Psi_h(t) = \mathbb{E}\frac{1}{1+tY_h} = \frac{1}{2\pi t}\int_{\mathbb{R}}\left[\sqrt{1+2t\psi_h(u)}-1\right]du. \tag{10}$$

This factorization separates two sources of geometry: M comes from the circular recurrent dynamics, whereas W_h is entirely determined by the pooling rule. The transform in Eq. (10) therefore retains more information than participation dimension.

Proposition 1 (Lower tail of the rescaled critical law)**.** *Let F_h be the distribution function of W_h, and let $\nu_h(dy) = y^{-1}\mathcal{L}(Y_h)(dy)$ denote the locally finite limit. If $\widehat{h} \in L^1$, then the density of $\mathcal{L}(Y_h)$ obeys*

$$f_{Y_h}(y) \sim \frac{\sqrt{2}}{\pi}\mathbb{E}(W_h^{-1/2})y^{-1/2}, \qquad \mathbb{E}(W_h^{-1/2}) = \frac{1}{2\pi}\int_{\mathbb{R}}|\widehat{h}(u)|du. \tag{11}$$

If instead $F_h(w) \sim C\sqrt{w}$, then

$$f_{Y_h}(y) \sim \frac{C}{\sqrt{2}\pi}y^{-1/2}\log(1/y). \tag{12}$$

In either case, the density of ν_h is $f_{Y_h}(y)/y$. More generally, if F_h is regularly varying at zero with index $0 < \beta < 1/2$, then

$$\Pr(Y_h \le y) \sim \mathbb{E}[(M/2)^{-\beta}]F_h(y). \tag{13}$$

At the boundary $\beta = 1/2$, a generic slowly varying factor produces an integrated slowly varying term, not automatically a logarithm. We use Eq. (12) only when the stated square-root asymptotic has been proved. For boxcar pooling, $\psi_h(u) = \mathrm{sinc}^2(u/2)$ and $F_h(w) \sim (4/\pi^2)\sqrt{w}$, recovering $f_{Y_h}(y) \sim (2\sqrt{2}/\pi^3)y^{-1/2}\log(1/y)$. A causal exponential kernel has $F_h(w) \sim (\sqrt{2}/\pi)\sqrt{w}$ and therefore constant $1/\pi^2$ in Eq. (12). These are direct substitutions into the general factor $C/(\sqrt{2}\pi)$. A Gaussian kernel has $\widehat{h} \in L^1$ and therefore no logarithmic enhancement.

What ridge regularization sees. For a width-N covariance, define the ridge effective dimension

$$d_{\mathrm{eff},h}(\lambda) = \mathrm{tr}\{C_{T,h}(C_{T,h}+\lambda I)^{-1}\}.$$

At the intermediate eigenvalue scale $\lambda_r = rT/\varepsilon^2$, Corollary 1 gives

$$\frac{T}{N\varepsilon}d_{\mathrm{eff},h}(\lambda_r) \longrightarrow \frac{1}{r}\Psi_h(1/r). \tag{14}$$

Equation (14) describes how many directions remain after a given ridge penalty. It does not determine target-aligned risk, which also depends on the readout vector.

The distinction is exact for a shared prediction task. Let $z_h \in \mathbb{R}^N$ be a pooled feature, let $y \in \mathbb{R}^q$ be the same target for every pooling rule, and define $C_h = \mathrm{Cov}(z_h)$, $K_h = \mathrm{Cov}(z_h, y)$, and

$S_y = \mathrm{Cov}(y)$. Population ridge regression minimizes $\mathbb{E}\|y - W^\top z_h\|_2^2 + \lambda\|W\|_F^2$ and has $W_h = (C_h + \lambda I)^{-1}K_h$. Its unpenalized prediction error is

$$R_h(\lambda) = \operatorname{tr} S_y - \operatorname{tr}\{K_h^\top (C_h + \lambda I)^{-1} K_h\} - \lambda \operatorname{tr}\{K_h^\top (C_h + \lambda I)^{-2} K_h\}. \tag{15}$$

Thus the covariance law determines only one part of a shared-task comparison; the feature–target cross-covariance K_h records task alignment. We use this identity below to separate a covariance-only consequence from prediction of a common future recurrent state.

Although covariance does not determine arbitrary target risk, it does determine the optimal denoising risk for the representation itself.

Corollary 2 (Population denoising risk)**.** *For fixed (ε, T), let $C_{T,h}^{(N)}$ be deterministic positive semidefinite matrices whose empirical spectral measures converge weakly to that of the circular covariance $C_{T,h}$ and whose normalized traces converge to $\tau(C_{T,h})$ as $N \to \infty$. Here $\mathcal{CN}$ denotes the circularly symmetric complex Gaussian distribution. Draw $x_N \sim \mathcal{CN}(0, C_{T,h}^{(N)})$ and independent $\xi_N \sim \mathcal{CN}(0, I)$, observe $y_N = x_N + \sqrt{\lambda_r}\,\xi_N$ with $\lambda_r = rT/\varepsilon^2$, and use the posterior mean*

$$\widehat{x}_N = C_{T,h}^{(N)}(C_{T,h}^{(N)} + \lambda_r I)^{-1} y_N.$$

In the same ordered limit as Theorem 1,

$$\frac{\mathbb{E}\|x_N - \widehat{x}_N\|_2^2}{\mathbb{E}\|x_N\|_2^2} \longrightarrow \Psi_h(1/r). \tag{16}$$

The result follows from the finite-dimensional Gaussian posterior identity and Eqs. (14) and (7); Appendix A.5 gives the calculation. Figure 2 therefore also traces ordered population Bayes denoising risk. This consequence concerns noisy representation recovery, not finite recurrent trajectories or general supervised test risk.

What a supervised linear readout sees. The same product law determines two population learning curves. Draw a feature $z \sim \mathcal{CN}(0, C)$ and an independent isotropic teacher $\beta \sim \mathcal{CN}(0, I/N)$, with noiseless label $y = z^*\beta$. Population gradient flow from zero obeys $\dot{w}_s = -C(w_s - \beta)$, while population ridge uses $w_\lambda = (C + \lambda I)^{-1}C\beta$.

Corollary 3 (Population linear-readout curves)**.** *For any finite positive semidefinite C,*

$$\frac{\mathbb{E}|z^*(w_s - \beta)|^2}{\mathbb{E}|z^*\beta|^2} = \frac{\operatorname{tr}(Ce^{-2sC})}{\operatorname{tr} C}, \tag{17}$$

$$\frac{\mathbb{E}|z^*(w_\lambda - \beta)|^2}{\mathbb{E}|z^*\beta|^2} = \frac{\lambda^2 \operatorname{tr}\{C(C + \lambda I)^{-2}\}}{\operatorname{tr} C}. \tag{18}$$

For the covariance sequence in Corollary 2, set $s = u\varepsilon^2/T$ and $\lambda = T/(t\varepsilon^2)$. In the ordered limit,

$$\mathcal{G}_h(u) = \mathbb{E}e^{-2uY_h}, \tag{19}$$

$$\mathcal{H}_h(t) = \mathbb{E}(1 + tY_h)^{-2} = \Psi_h(t) + t\Psi_h'(t) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\psi_h(v)}{\sqrt{1 + 2t\psi_h(v)}}\,dv. \tag{20}$$

The finite identities follow by diagonalizing C and averaging over the isotropic teacher; Appendix A.6 gives the calculation. Their small-argument expansions are

$$\mathcal{G}_h(u) = 1 - 2B_2(h)u + 2B_3(h)u^2 + O(u^3), \qquad \mathcal{H}_h(t) = 1 - 2B_2(h)t + 3B_3(h)t^2 + O(t^3).$$

Participation matching therefore fixes the initial slope but not the curvature or later learning trajectory. For the matched boxcar and causal exponential kernels below, $\mathcal{G}_h(8)$ is respectively 0.312 and 0.356, while $\mathcal{H}_h(32)$ is 0.214 and 0.248. These are analytic ordered-limit values, not fitted finite-network results. The corollary averages over an isotropic population teacher and does not establish finite-sample empirical-risk minimization or performance for a target aligned with a particular eigenspace.

The higher-order distinction also survives Gaussian feature sampling. If $n/N \to \gamma\varepsilon/T$ after fixing (ε, T), the ordered empirical law of the $n \times n$ training Gram matrix has free cumulants $\kappa_p = \gamma^{p-1}B_p(h)$ and R-transform $R_{\gamma,h}(w) = \Psi_h(-\gamma w)$; see Appendix B. Thus the whole filter-power distribution, rather than participation dimension alone, enters the sampled feature geometry. This is an abstract Gaussian population-covariance result, not a theorem for one temporally correlated recurrent trajectory.

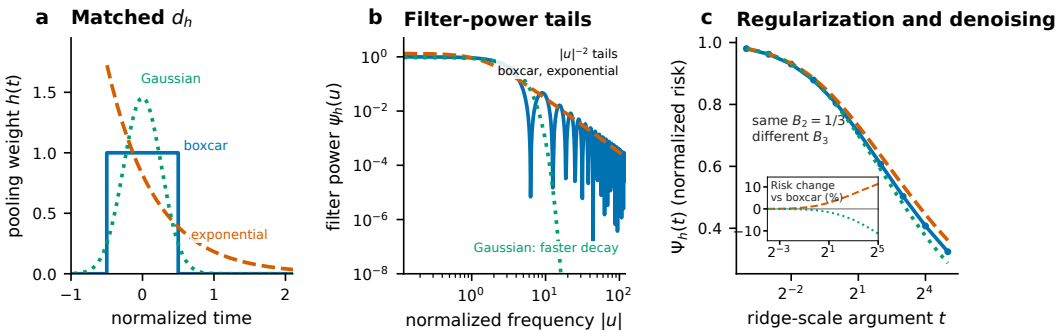

Figure 2: **Pooling kernels with equal participation dimension can have different critical covariance laws. a**, Unit-energy temporal kernels, translated only for visual alignment. **b**, Their filter power ψ_h; boxcar and causal-exponential power have quadratic tails, whereas Gaussian power decays faster than any power. **c**, The analytic transforms Ψ_h in Eq. (10); at argument $t = 1/r$, the same ordinate is the normalized Bayes-optimal denoising risk in Eq. (16). All three kernels have $J_2 = 2/3$, so their curves share the same first derivative at the origin, but their higher-order geometry and denoising risks differ. The inset reports the causal-exponential and Gaussian percentage changes in risk relative to boxcar.

Same participation dimension, different geometry. For $h_a(t) = a^{-1/2}h(t/a)$, one has $J_p(h_a) = a^{p-1}J_p(h)$. Table 1 compares three kernels scaled to the boxcar value $J_2 = 2/3$. They have the same leading participation dimension, but different B_3 and different transforms. The Gaussian taper also has a different lower tail in the rescaled, covariance-weighted limit. Thus matching total variance and participation dimension does not match the regularized covariance geometry, the supervised population curves in Eq. (19)–(20), or the optimal denoising risk in Eq. (16).

The Gaussian profile is a two-sided kernel suited to offline pooling; a strictly online representation requires a causal taper, such as the exponential example, or a separately analyzed delayed/truncated kernel.

Figure 2 shows the three pooling profiles, their power responses, and the distinct transform curves. At the largest displayed ridge-scale argument, $t = 32$, the normalized denoising risks are 0.361 for the causal exponential, 0.324 for boxcar, and 0.288 for the Gaussian. Equivalently, the first and third values differ from boxcar by $+11.4\%$ and -11.3%. These values come from Eq. (10) and deterministic quadrature, not from random matrices.

For the boxcar special case, Proposition 3 gives the equal-time singularity $y^{-7/4}$ in the same ordered normalization, while the zero-frequency law has singularity $y^{-5/3}$ (Hu & Sompolinsky, 2022). The intermediate boxcar law has the distinct logarithmically corrected exponent $3/2$. The exact second moment verifies both adjacent time boundaries, while the exact third moment independently verifies the intermediate constant and the suppression of macroscopically separated frequency sectors. Appendix D tabulates the three regimes.

Table 1: **Equal-participation pooling kernels remain spectrally distinct.** All kernels have $\|h\|_2 = 1$ and $J_2 = 2/3$, hence the same leading value in Eq. (8). The displayed density is that of the locally finite covariance law ν_h near zero.

kernel	dilation a	$B_3(h)$	ν_h density as $y \downarrow 0$
boxcar	1	$11/40$	$y^{-3/2}\log(1/y)$
causal exponential	$2/3$	$1/3$	$y^{-3/2}\log(1/y)$
Gaussian	$2/(3\sqrt{2\pi})$	$4/(9\sqrt{3})$	$y^{-3/2}$

4 PROOF STRATEGY

The proof converts the noncommuting frequency integral into a localized combinatorial series indexed by balanced run lengths. Its exact kernel coefficients are gamma averages of the autocorrelation ϕ_h. Absolute row, column, and displacement bounds localize these signed coefficients to $|j-k| = O(T)$, after which Fourier inversion supplies $T^{p-1}J_p(h)$. A balanced-word insertion argument couples the remaining count to the fixed-run noncrossing-pairing polynomial of Kemp et al. (2011); a rooted-tree recursion and Lagrange inversion give the simplex constant in Eq. (5). Boundary removal and Abelian summation at $n \asymp \varepsilon^{-1}$ finish the proof. Appendix A gives the exact series, localization estimates, pairing sandwich, and lower-tail argument. The new technical step is the uniform signed-kernel localization and its coupling to the pairing limit, rather than the imported pairing formula alone.

5 EXACT BOUNDARY CHECKS

The exact finite-window matrix identity of Calvo et al. (2026), translated to the complex notation used here, supplies the endpoint interpolation in Appendix D. In this section only, write $C_T = C_{T,h_{\mathrm{box}}}$. Our low-order spectral calculations begin with the exact first moment

$$\tau(C_T) = \varepsilon^{-1}\left[1 - \frac{1 - e^{-T\sqrt{\varepsilon}}}{T\sqrt{\varepsilon}}\right]. \tag{21}$$

For boxcar pooling, the second-moment two-frequency kernel is

$$\tau(Q_\omega Q_\nu) = \frac{(1+\omega^2)(1+\nu^2)}{(\varepsilon+\omega^2)(\varepsilon+\nu^2)[(\varepsilon+\omega\nu)^2+(\omega-\nu)^2]}. \tag{22}$$

This kernel contains two different frequency scales: a central width $\sqrt{\varepsilon}$ and a relative-frequency width ε. Consequently,

$$\tau(C_T^2) \sim \begin{cases} (3/32)T^2\varepsilon^{-5/2}, & T\sqrt{\varepsilon} \to 0, \\ (1/3)T\varepsilon^{-3}, & T\sqrt{\varepsilon} \to \infty, T\varepsilon \to 0, \\ \varepsilon^{-4}G_2(T\varepsilon), & T\varepsilon = O(1), \end{cases}$$

where $G_2(z) = 1 - 2/z + 2(1-e^{-z})/z^2$. These identities rule out a single crossover obtained by matching the endpoint powers. The third-moment calculation independently returns the boxcar value $B_3 = 11/40$; Appendix D records the kernels and constants.

6 FINITE-WIDTH CONSTRAINTS AND SHARED-TASK EVIDENCE

For a finite complex Ginibre matrix G_N with entries of variance $1/N$, set $J_N = \sqrt{1-\varepsilon_N}G_N$. Stationarity requires $\max \Re\,\mathrm{spec}(J_N) < 1$. The extremal-eigenvalue theorem of Cipolloni et al. (2022, Theorem 1 and Eq. (2), complex case) gives the following direct corollary; the centering γ_N below, including its $5\log\log N$ and $\log(2\pi^4)$ terms, is quoted verbatim from that result, and only its leading order $\gamma_N = \frac{1}{2}(1+o(1))\log N$ is used downstream.

Proposition 2 (Population-gain stability window)**.** *Let*

$$\gamma_N = \tfrac{1}{2}[\log N - 5\log\log N - \log(2\pi^4)]$$

and, where $\gamma_N > 0$*,*

$$s_N = \sqrt{4N\gamma_N}\left[(1-\varepsilon_N)^{-1/2} - 1 - \sqrt{\gamma_N/(4N)}\right].$$

If $s_N \to s$*, then the stability probability tends to* $\exp(-e^{-s})$*. The leading boundary is* $\varepsilon_N \asymp \sqrt{\log N/N}$*.*

Define the two-sided separation from the intermediate-window boundaries by

$$\mathcal{S}_N(T) = \min\{T\sqrt{\varepsilon_N}, (T\varepsilon_N)^{-1}\}.$$

Both terms must diverge for a strict intermediate window.

Corollary 4 (Eighth-root finite-width separation barrier)**.** *For fixed ε_N, the best possible separation over all $T > 0$ is*

$$\sup_{T>0} \mathcal{S}_N(T) = \varepsilon_N^{-1/4}, \qquad T_N^* = \varepsilon_N^{-3/4}. \tag{23}$$

Along any sequence whose stability probability tends to one at the leading Ginibre edge,

$$\sup_T \mathcal{S}_N(T) \le (1 + o(1)) \left(\frac{2N}{\log N} \right)^{1/8}. \tag{24}$$

Conversely, for any fixed $c > 1$, the explicit schedule $\varepsilon_N = c\sqrt{\log N/(2N)}$ and $T_N = \varepsilon_N^{-3/4}$ has stability probability tending to one and attains the same eighth-root order.

Thus achieving a two-sided separation $\mathcal{S}_N \ge \kappa$ near the stable edge requires, at leading order, $N \gtrsim (\kappa^8/2) \log N$. Ignoring the logarithmic correction, doubling the separation therefore costs a factor $2^8 = 256$ in width. This slow growth is a finite-width observability barrier rather than a failure of the ordered theorem. It also explains why modest-width simulations cannot by themselves validate the strict corridor. Appendix E gives the two-line optimization and edge substitution.

This proposition is a corollary of an existing edge theorem. Here it prevents an invalid finite-width extrapolation of the ordered free law. For any fixed admissible kernel, combining high-probability stability away from the edge with the two intermediate-window inequalities and Eq. (8) gives $d_h/\log N \to \infty$ for the ordered participation-dimension proxy. A bounded-dimensional intermediate schedule is therefore incompatible with that stability requirement unless a separate joint finite-N law changes the prediction.

The proposition does not provide a finite-width covariance law. Such a law would require uniform control as N, ε, and T move jointly; neither stability alone nor a fitted rank slope establishes the product limit.

Fixed-parameter covariance checks. Finite complex Ginibre covariances at widths 128–2048 approached exact circular moments at fixed (ε, T), but the limiting dilation $a = 2/3$ did not match lower moments at the accessible gaps. A separate fixed-parameter calibration nearly matched trace and participation at widths 256–1024 while leaving about a 2% third-moment difference. Appendix F reports both checks; neither tests the jointly moving limit.

6.1 Predicting the same future state

We next used real-valued finite networks and a common target. Let G_N have independent $\mathcal{N}(0, 1/N)$ entries and $A_N = I - \sqrt{1-\varepsilon}\, G_N$. For each stable base matrix, gap, and window, we computed the boxcar covariance C_{box} and chose a causal-exponential dilation a_i to solve

$$\frac{\mathrm{tr}(C_{\exp}(a_i))^2}{N \,\mathrm{tr}(C_{\exp}(a_i)^2)} = \frac{\mathrm{tr}(C_{\mathrm{box}})^2}{N \,\mathrm{tr}(C_{\mathrm{box}}^2)}.$$

We then scaled $C_{\exp}(a_i)$ to match $\mathrm{tr}(C_{\mathrm{box}})$. No target enters this calculation. It matches both the first and second spectral moments for each finite matrix. It is a mechanism control rather than one fixed kernel intended for deployment.

The pooled features share the target $y = x_{t+1}$, the full recurrent state one time unit after the pooling anchor. Instead of drawing unrelated feature ensembles, we sampled the boxcar feature, exponential feature, and target from their exact joint stationary Gaussian law. Population normalized mean-squared error (NMSE) uses the optimal linear predictor and is divided by $\mathrm{tr}\,\mathrm{Cov}(y)$. Finite-sample ridge learners use the same targets, training size, validation set, and test set for both kernels; each kernel selects its penalty from the same fixed grid. We write $d_{\mathrm{ref}} = \mathrm{tr}(C_{\mathrm{box}})^2/\,\mathrm{tr}(C_{\mathrm{box}}^2)$ for their exactly shared participation dimension. Recurrent matrices are the independent units for all intervals.

The experiment used 32 matrices at width 128 and 24 at width 256, $\varepsilon \in \{0.18, 0.24, 0.32\}$, and $T/\varepsilon^{-3/4} \in \{0.75, 1, 1.5\}$. Every matrix completed every condition without replacement. The largest relative trace or second-moment error was 1.9×10^{-14}, while the exponential-to-boxcar third-moment ratio at width 256 ranged from 1.0049 to 1.0382.

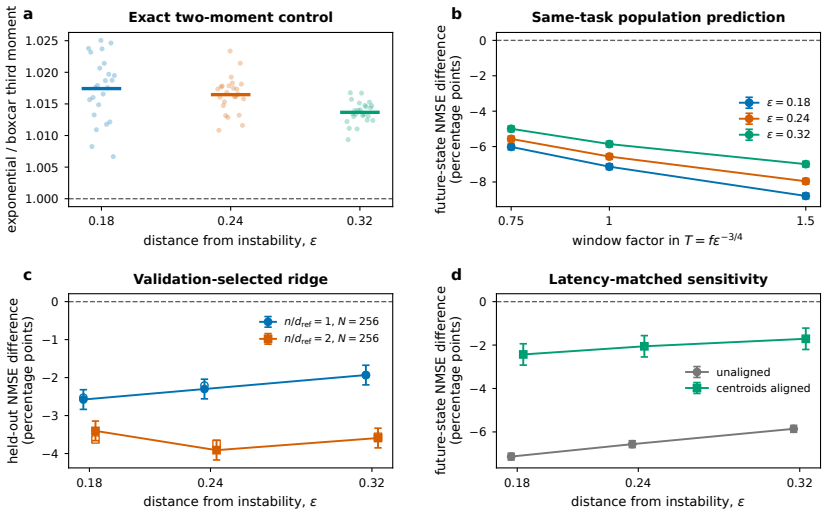

Figure 3: **Exactly matched finite covariances on a shared future-state task. a**, Third-moment ratios after exact per-network trace and second-moment matching (maximum relative error 1.9×10^{-14}). **b**, Population future-state NMSE at width 256. **c**, Validation-selected ridge differences (closed: width-256 simultaneous 95% intervals; open: width-128 means). **d**, Population sensitivity after amplitude-centroid alignment. Negative values favor exponential pooling.

At width 256, causal exponential pooling reduced population future-state NMSE in all nine gap-by-window cells by 4.99–8.79 percentage points. Central-window differences were 5.858–7.136 points and validation-selected width-256 ridge differences were 1.933–3.911 points, all in the same direction with simultaneous intervals excluding zero. Width 128 had the same mean sign, but two fits from one matrix and draw reached the penalty-grid edge, so this comparison is descriptive.

Temporal translation leaves feature covariance unchanged but alters feature–target cross-covariance. We therefore repeated the population comparison after adding a pure delay to the younger kernel so that boxcar and exponential pooling had the same mean amplitude-weighted sample age (amplitude centroid). This prespecified sensitivity reduced the three central-window differences to 1.71–2.43 points, with all simultaneous intervals still excluding zero. Temporal recency therefore explains 66–71% of the unaligned contrast; the residual is a joint feature-covariance and feature–target-alignment effect, not a spectrum-only estimate.

Estimator dependence. In a separate complex-Gaussian design at width 1024, matched exponential pooling raised total ridge risk by 0.242 points at $n/d_{\mathrm{ref}} = 1$, but lowered minimum-norm risk by 1.597 points at $n/d_{\mathrm{ref}} = 2$ because its variance reduction outweighed its bias increase. Appendix H reports the decomposition and simultaneous intervals.

Relation to prior covariance laws. Prior work gives equal-time, frequency-resolved, zero-frequency, and discrete-time recurrent covariance laws (Hennequin et al., 2012; Hu & Sompolinsky, 2022; Calvo et al., 2024; Zavatone-Veth, 2026), including an exact boxcar identity (Calvo et al., 2026). Finite-horizon Gramians and echo-state readouts are adjacent (Kürschner, 2018; Moakher et al., 2025). We identify the intervening kernel-dependent law and test its task consequences without calling the finite experiment an asymptotic validation.

Discussion and limitations. The theorem and experiments support a qualified conclusion: temporal pooling can change what a linear readout sees after trace and participation dimension are matched. The ordered law identifies kernel-specific higher-order geometry, and a residual difference remains on the finite common-target task after matching mean sample age. Timing explains most of that contrast, and estimator choice can reverse the ranking, so no kernel is uniformly best. Participation dimension is useful but incomplete without task alignment and the estimator. The scope is an ordered circular limit, independent stationary windows, and per-matrix matching used as a control.

AI USE STATEMENT

No generative AI tools or large language models were used in this work.

REFERENCES

Rubén Calvo, Carles Martorell, Guillermo B. Morales, Serena Di Santo, and Miguel A. Muñoz. Frequency-dependent covariance reveals critical spatiotemporal patterns of synchronized activity in the human brain. *Physical Review Letters*, 133(20):208401, 2024. doi: 10.1103/PhysRevLett.133.208401.

Rubén Calvo, Carles Martorell, Adrián Roig, and Miguel A. Muñoz. Robust scaling in human brain dynamics despite correlated inputs and limited sampling distortions. *Physical Review Letters*, 136 (6):068402, 2026. doi: 10.1103/36v9-wtm8.

Giorgio Cipolloni, László Erdős, Dominik Schröder, and Yuanyuan Xu. Directional extremal statistics for ginibre eigenvalues. *Journal of Mathematical Physics*, 63(10):103303, 2022. doi: 10.1063/5.0104290.

David G. Clark, L. F. Abbott, and Ashok Litwin-Kumar. Dimension of activity in random neural networks. *Physical Review Letters*, 131:118401, 2023. doi: 10.1103/PhysRevLett.131.118401.

Zilin Gao, Qilong Wang, Bingbing Zhang, Qinghua Hu, and Peihua Li. Temporal-attentive covariance pooling networks for video recognition. In *Advances in Neural Information Processing Systems*, volume 34, pp. 13587–13598, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/70efdf2ec9b086079795c442636b55fb-Abstract.html.

Fredric J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978. doi: 10.1109/PROC.1978.10837.

Guillaume Hennequin, Tim P. Vogels, and Wulfram Gerstner. Non-normal amplification in random balanced neuronal networks. *Physical Review E*, 86:011909, 2012. doi: 10.1103/PhysRevE.86.011909.

Yuan Hu and Haim Sompolinsky. The spectrum of covariance matrices of randomly connected recurrent neuronal networks with linear dynamics. *PLOS Computational Biology*, 18(7):e1010327, 2022. doi: 10.1371/journal.pcbi.1010327.

Todd Kemp, Karl Mahlburg, Amarpreet Rattan, and Clifford Smyth. Enumeration of non-crossing pairings on bit strings. *Journal of Combinatorial Theory, Series A*, 118(1):129–151, 2011. doi: 10.1016/j.jcta.2010.07.002.

Bernadette Krawczyk and Roland Speicher. Combinatorics of free cumulants. *Journal of Combinatorial Theory, Series A*, 90(2):267–292, 2000. doi: 10.1006/jcta.1999.3032.

Patrick Kürschner. Balanced truncation model order reduction in limited time intervals for large systems. *Advances in Computational Mathematics*, 44:1821–1844, 2018. doi: 10.1007/s10444-018-9608-6.

Vladimir A. Marčenko and Leonid A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457–483, 1967. doi: 10.1070/sm1967v001n04abeh001994.

Yessin Moakher, Malik Tiomoko, Cosme Louart, and Zhenyu Liao. Generalization in representation models via random matrix theory: Application to recurrent networks. *arXiv preprint arXiv:2511.02401*, 2025. URL https://arxiv.org/abs/2511.02401.

C. Emily I. Redelmeier. Real second-order freeness and the asymptotic real second-order freeness of several real matrix models. *International Mathematics Research Notices*, 2014(12):3353–3395, 2014. doi: 10.1093/imrn/rnt043. URL https://arxiv.org/abs/1101.0422.

Jacob A. Zavatone-Veth. Stationary covariance spectra of discrete-time non-normal random recurrent dynamics. *arXiv preprint arXiv:2606.31944*, 2026.

A PROOFS FOR THE POOLING-KERNEL LAW

This appendix fixes the normalization, supplies the signed-kernel coefficient argument used in Theorem 1, and proves the distributional statements in Corollary 1 and Proposition 1. Throughout, p is fixed before $\varepsilon \downarrow 0$.

The proof proceeds in five steps. First, the circular Wick rule gives the exact balanced-run series in Eq. (28). Second, the row, column, and displacement estimates Eqs. (29)–(33) localize its signed kernel coefficients and yield the convolution constant $J_p(h)$ in Eq. (35). Third, balanced-word insertion gives the two-sided pairing comparison Eqs. (36)–(37), from which the uniform local-balance estimate Eq. (38) follows. Fourth, Lemma 2 counts the remaining rooted forests and Lagrange inversion evaluates their simplex integral I_p. Finally, the boundary estimate following Eq. (42) and Abelian summation convert the coefficient asymptotic into Theorem 1. Each step uses the output of the preceding one; moment identification never replaces coefficient control.

A.1 FREQUENCY REPRESENTATION AND EXACT COEFFICIENTS

For $A = 1 - gc$, the stationary covariance is

$$\Sigma = \int_0^\infty e^{-As} e^{-A^* s} ds, \qquad A\Sigma + \Sigma A^* = 1.$$

For $C_{T,h}$ as in Eq. (2), stationarity gives

$$C_{T,h} = \int_{\mathbb{R}} T\psi_h(T\omega) Q_\omega \frac{d\omega}{2\pi}, \quad Q_\omega = [(1+i\omega) - gc]^{-1}[(1-i\omega) - gc^*]^{-1}.$$

Expanding each resolvent and applying the circular Wick rule produces balanced $2p$-run words. Let $\Phi(\mathbf{j}, \mathbf{k})$ be the number of noncrossing pairings of

$$+^{j_1} -^{k_1} +^{j_2} -^{k_2} \cdots +^{j_p} -^{k_p}, \qquad |\mathbf{j}| = |\mathbf{k}| = n.$$

This expansion is justified in noncommutative L_{2p} rather than operator norm. The fixed-run bound of Kemp et al. (2011) gives

$$\|c^j\|_{2p}^{2p} = \tau\big([(c^*)^j c^j]^p\big) = O_p((j+1)^{p-1}).$$

Consequently, uniformly for real ω,

$$\sum_{j \geq 0} \frac{g^j}{|1+i\omega|^{j+1}} \|c^j\|_{2p} < \infty.$$

The truncated geometric identity identifies the L_{2p} limit with $[(1+i\omega) - gc]^{-1}$. Noncommutative Hölder places products of the $2p$ analytic and anti-analytic factors in L_1. We first perform every frequency integral for finite truncations. The absolute coefficient bounds below, together with $g < 1$ and the fixed-run polynomial bound, then dominate the regrouped series. This justifies passage to the L_{2p} limit without assuming positivity of the autocorrelation or an operator-norm Neumann series near criticality.

Let $f_j(s) = e^{-s} s^j / j!\, \mathbf{1}_{s \geq 0}$. Fourier inversion and the Wiener–Khinchin identity give

$$b_{jk}^{(h)}(T) = \int_{\mathbb{R}} \frac{T\psi_h(T\omega)}{(1+i\omega)^{j+1}(1-i\omega)^{k+1}} \frac{d\omega}{2\pi} \tag{25}$$

$$= \int_0^\infty \int_0^\infty f_j(s) f_k(t) \phi_h((s-t)/T)\, ds\, dt \tag{26}$$

$$= \mathbb{E}\, \phi_h((S_j - S'_k)/T), \tag{27}$$

where S_j and S'_k are independent gamma variables with shape $j+1$ and $k+1$ and unit scale. Consequently,

$$\tau(C_{T,h}^p) = \sum_{n \geq 0} (1-\varepsilon)^n A_{p,n}^{(h)}(T), \qquad A_{p,n}^{(h)} = \sum_{|\mathbf{j}| = |\mathbf{k}| = n} \Phi(\mathbf{j}, \mathbf{k}) \prod_{r=1}^{p} b_{j_r k_r}^{(h)}(T). \tag{28}$$

The probability representation also implies

$$\sup_j \sum_k |b_{jk}^{(h)}| \le T m_0, \qquad \sup_k \sum_j |b_{jk}^{(h)}| \le T m_0, \qquad m_0 = \|\phi_h\|_1. \tag{29}$$

Indeed, $\sum_{k\ge 0} f_k(t) = 1$; extending the half-line integral to the real line and substituting $v = (s-t)/T$ proves the first bound, and the second is symmetric.

A.2 Uniform autocorrelation localization

Let $L = \|\phi_h'\|_\infty$. Jensen's inequality gives

$$|b_{jk}^{(h)}(T) - \phi_h((j-k)/T)| \le \frac{L\sqrt{j+k+2}}{T}. \tag{30}$$

Pointwise control is insufficient because $O(T)$ difference indices contribute. Put $m_{1/2} = \int |v|^{1/2}|\phi_h(v)|dv$ and $m_1 = \int |v||\phi_h(v)|dv$. If K_t is Poisson with mean t, then $\mathbb{E}|j - K_t| \le |j-t| + \sqrt{t}$. Using $|j-t| \le |j-s| + |s-t|$ and $\sqrt{t} \le \sqrt{s} + \sqrt{|s-t|}$ in Eq. (27) gives

$$\sum_{k\ge 0} |j-k|\, |b_{jk}^{(h)}(T)| \le T m_0 (1 + 2\sqrt{j+1}) + T^2 m_1 + T^{3/2} m_{1/2}. \tag{31}$$

Here we used $\mathbb{E}|j - S_j| \le 1 + \sqrt{j+1}$ and $\mathbb{E}\sqrt{S_j} \le \sqrt{j+1}$.

Lemma 1 (Uniform displacement-tail removal)**.** *Extend $b_{jk}^{(h)}$ by zero when $k < 0$. Fix $p \ge 2$ and an interior simplex compactum, let $j_r \asymp n$, and suppose $T/\sqrt{n} \to \infty$ and $T/n \to 0$. Then*

$$\lim_{R\to\infty} \limsup_{n\to\infty} T^{-(p-1)} \sum_{\substack{d_1+\cdots+d_p=0 \\ \max_r |d_r| > RT}} \prod_{r=1}^{p} \left| b_{j_r, j_r - d_r}^{(h)}(T) \right| = 0. \tag{32}$$

The same limit holds after multiplying each summand by $n^{-(p-1)}\Phi(\mathbf{j}, \mathbf{j}-\mathbf{d})$:

$$\lim_{R\to\infty} \limsup_{n\to\infty} \frac{1}{n^{p-1}T^{p-1}} \sum_{\substack{d_1+\cdots+d_p=0 \\ \max_r |d_r| > RT}} \Phi(\mathbf{j}, \mathbf{j}-\mathbf{d}) \prod_{r=1}^{p} \left| b_{j_r, j_r - d_r}^{(h)}(T) \right| = 0. \tag{33}$$

Proof. The normalized autocorrelation satisfies $|\phi_h(v)| \le \phi_h(0) = 1$, so Eq. (27) gives $|b_{jk}^{(h)}| \le 1$. For $a_r(d) = |b_{j_r, j_r - d}^{(h)}(T)|$, Eqs. (29) and (31) imply

$$\|a_r\|_1 \le T m_0, \qquad \|a_r\|_\infty \le 1,$$
$$\|a_r \mathbf{1}_{\{|d| > RT\}}\|_1 \le \frac{T}{R}\left\{ m_1 + \frac{m_{1/2}}{\sqrt{T}} + \frac{m_0(1 + 2\sqrt{j_r+1})}{T} \right\}.$$

The braces equal $m_1 + o(1)$ uniformly on the compactum. For a fixed coordinate r, retain the zero-sum restriction and discard only the remaining admissibility restrictions. Up to reversing harmless sign conventions, the resulting constrained sum is the convolution at zero

$$(a_1^{\mathrm{tail}} * a_2 * \cdots * a_p)(0).$$

Discrete Young's inequality now uses the tail factor in ℓ_1, one other factor in ℓ_∞, and the remaining $p-2$ factors in ℓ_1. The resulting bound is

$$\frac{T^{p-1}}{R}\{m_1 m_0^{p-2} + o(1)\}.$$

A union bound over the p possible tail coordinates, followed first by $n \to \infty$ and then by $R \to \infty$, proves Eq. (32). To obtain the weighted statement, use the upper balanced-insertion comparison proved below. Its symmetric fixed-run tree polynomial has degree at most $p-1$, while every enlarged run length is at most $2n$; hence, uniformly over all admissible $\mathbf{j}, \mathbf{d}$,

$$0 \le \Phi(\mathbf{j}, \mathbf{j}-\mathbf{d}) \le C_p (n+1)^{p-1}.$$

Multiplying the preceding tail estimate by this global bound proves Eq. (33). This is the form used in the coefficient sum $A_{p,n}^{(h)}$. □

Fix an interior simplex compactum, so $j \asymp n$. For any fixed $R > 1$, Eq. (30) controls $|j - k| \le RT$, while Markov's inequality and Eq. (31) control its complement. The sampled autocorrelation tail converges to $\int_{|v|>R} |\phi_h(v)|dv$; the $W^{1,1}$ assumption controls the lattice-sum error. If $T/\sqrt{n} \to \infty$ and $T/n \to 0$, then

$$\frac{1}{T}\sum_k |b_{jk}^{(h)}(T) - \phi_h((j-k)/T)| \longrightarrow 0 \tag{34}$$

uniformly on that compactum. Send $n \to \infty$ and then $R \to \infty$. Equation (32) removes the off-central part of the cyclic kernel product uniformly; Eq. (33) does so with the pairing count present in the exact coefficient formula. Telescoping Eq. (34) across the remaining p factors, using the absolute row/column bounds and discrete Young convolution, yields, for $p \ge 2$,

$$\sum_{d_1+\cdots+d_p=0} \prod_{r=1}^{p} b_{j_r, j_r-d_r}^{(h)}(T) = T^{p-1}\{J_p(h) + o(1)\}, \tag{35}$$

uniformly over interior $\mathbf{j}$. The constant is the lattice Riemann limit

$$J_p(h) = \int_{u_1+\cdots+u_p=0} \prod_{r=1}^{p} \phi_h(u_r) du_1 \cdots du_{p-1} = \frac{1}{2\pi}\int_{\mathbb{R}} \psi_h(u)^p du.$$

For $p = 1$, gamma concentration directly gives $b_{nn}^{(h)}(T) \to \phi_h(0) = 1$.

A.3 Local balance and the simplex constant

We record the pairing comparison that transfers the symmetric fixed-run asymptotics of Kemp et al. (2011) to locally imbalanced run lengths. We first use the following elementary insertion fact. If a balanced cyclic bit word v is obtained from another balanced word u by inserting contiguous balanced words into gaps of u, then $\Phi(u) \le \Phi(v)$. To see this, fix a noncrossing opposite-color pairing inside each inserted word. Extend any pairing of u by these internal chords. An inserted word occupies a contiguous cyclic interval containing no old vertex, so neither its chords nor chords from distinct inserted gaps can alternate with an old chord. Deleting the inserted vertices recovers the original pairing; the extension is therefore injective.

Traverse a balanced word and let $b_0 = 0$, $b_r = \sum_{i=1}^{r}(j_i - k_i)$ be its boundary heights, and $h_r = b_{r-1} + j_r = b_r + k_r$ its peaks. Put $B = \max_r b_r$, $b = \min_r b_r$, $s_r = h_r - B$, and $t_r = h_r - b$. Balanced-gap insertion always gives the upper comparison

$$\Phi(\mathbf{j}, \mathbf{k}) \le \Phi^*(\mathbf{t}). \tag{36}$$

where $\Phi^*(\mathbf{n}) = \Phi(\mathbf{n}, \mathbf{n})$ is the symmetric-run pairing count. If every s_r is nonnegative, the complementary lower comparison

$$\Phi^*(\mathbf{s}) \le \Phi(\mathbf{j}, \mathbf{k}) \tag{37}$$

also holds. For Eq. (37), set $q_r = B - b_r \ge 0$. Then $j_r = q_{r-1} + s_r$ and $k_r = s_r + q_r$, so the original word is obtained from the symmetric $\mathbf{s}$ word by inserting gaps $0^{q_r}1^{q_r}$. For Eq. (36), set $a_r = b_r - b \ge 0$. Since $t_r = a_{r-1} + j_r = k_r + a_r$, the symmetric $\mathbf{t}$ word is obtained from the original word by inserting $0^{a_r}1^{a_r}$. The insertion fact proves the two comparisons on their stated domains.

The lower comparison is used only on the central part of an η-interior simplex. There $j_r \ge \eta n$, $|j_r - k_r| \le RT$ for fixed R, and $T/n \to 0$. Hence $B - b \le pRT$ and $s_r = h_r - B \ge j_r - (B - b) \ge \eta n - pRT \ge 0$ for all sufficiently large n. Thus both comparisons apply on exactly the region where the local asymptotic is taken. The fixed-run leading polynomial then gives

$$\Phi(\mathbf{j}, \mathbf{k}) = n^{p-1}H_p(\mathbf{j}/n) + O_{p,\eta}((T+1)n^{p-2}) \tag{38}$$

uniformly on an η-interior simplex. Here is the uniformity argument in detail. On every weak-order chamber, the exact symmetric formula of Kemp et al. (2011, Theorem 3.12 and Eq. (3.13)) is a finite tree polynomial of degree at most $p - 1$; write H_p for its homogeneous degree-$(p-1)$ part. The chamber polynomials agree on a shared tie face. To see this without appealing only to pointwise continuity, parametrize the rational affine span of that face by its $p - 1$ free run lengths while fixing the strict order of the remaining coordinates. Its positive integer points contain arbitrarily large

Cartesian grids after a common integer dilation. Repeated univariate interpolation therefore makes the two polynomial restrictions identical on the whole face. Thus H_p is continuous and Lipschitz on every compact interior simplex. The same construction also gives

$$0 \le t_r - s_r = B - b = O_p(T), \qquad \left| s_r - \frac{j_r + k_r}{2} \right| + \left| t_r - \frac{j_r + k_r}{2} \right| = O_p(T).$$

The degree-$(p-2)$ remainder in the exact tree polynomial is uniform on that compactum. Applying it to both sides of Eq. (36) and using the Lipschitz bound squeezes the middle term, proving Eq. (38). This also shows that the argument needs only $o(n)$ local imbalance; the $O(T)$ form is the rate used below.

The remaining generating-function step has a simple combinatorial structure: root the balanced boundary walk at its unique minimum, delete the root to expose an ordered forest, and sum over root degree and subtree compositions. Lemma 2 formalizes this recursion.

Lemma 2 (Exponential-minimum recursion)**.** *Let $E_1, \dots, E_p$ be independent unit exponential variables and define*

$$F_p(t) = \mathbb{E} H_p(t + E_1, \dots, t + E_p), \qquad F(z,t) = \sum_{p \ge 1} F_p(t) z^p.$$

Then $F_1(t) = 1$ and, for $p \ge 2$,

$$F_p(t) = p e^{pt} \int_t^\infty e^{-pu} [z^{p-1}] \{ e^{uF(z,u)} - 1 \} \, du. \tag{39}$$

The unique triangular formal-power-series solution satisfies

$$F = z \exp(tF) + \frac{F^2}{2}. \tag{40}$$

Proof. The rooted plane-tree encoding is Eq. (3.10), and its chamberwise labeled version is Theorem 3.12 of Kemp et al. (2011); removing the root produces the unique ordered forest in their Lemma 3.8. The global minimum of p independent exponentials has density pe^{-pu}. Cyclic invariance moves the location of that unique minimum to the root. The p possible locations are already counted by the factor p in this density, so there is no second cyclic-symmetry factor. Conditional on the minimum, memorylessness leaves the other labels independent unit exponentials above the new floor. In the homogeneous degree-$(p-1)$ part of Eq. (3.13), a root of degree q contributes $u^q/q!$, while Lemma 3.8 makes the ordered sizes of its nonempty subtrees a composition of the remaining $p-1$ run pairs. Summing over the root degree and ordered subtree sizes gives $[z^{p-1}](e^{uF(z,u)} - 1)$, proving Eq. (39).

Differentiating Eq. (39) gives

$$F_p'(t) = pF_p(t) - p[z^{p-1}](e^{tF(z,t)} - 1), \qquad e^{-pt} F_p(t) \to 0.$$

Summing this triangular differential system over $p \ge 2$, and writing $E = e^{tF} - 1$, yields

$$F_t = zF_z - z - z(E + zE_z). \tag{41}$$

Implicit differentiation of Eq. (40) gives

$$F_z(1 - F - zte^{tF}) = e^{tF}, \qquad F_t(1 - F - zte^{tF}) = ze^{tF} F.$$

Substitution verifies Eq. (41). Each coefficient is a polynomial in t satisfying the terminal condition, so triangular uniqueness proves Eq. (40). □

Worked $p = 3$ check. Write $F = z + F_2(t)z^2 + F_3(t)z^3 + O(z^4)$. Equation (39) first gives

$$F_2(t) = 2e^{2t} \int_t^\infty u e^{-2u} \, du = t + \frac{1}{2}.$$

For $p = 3$,

$$[z^2](e^{tF} - 1) = tF_2(t) + \frac{t^2}{2} F_1(t)^2 = \frac{3}{2} t^2 + \frac{1}{2} t,$$

and hence

$$F_3(t) = 3e^{3t}\int_t^\infty e^{-3u}\left(\frac{3}{2}u^2+\frac{1}{2}u\right)du = \frac{3}{2}t^2+\frac{3}{2}t+\frac{1}{2}.$$

Thus $F_2(0) = F_3(0) = 1/2$, agreeing with $\mathrm{Cat}_{p-1}/2^{p-1}$ for $p = 2, 3$. This calculation exposes the ordered-forest multiplicities without relying on the closed generating function.

The simplex integral of H_p can now be evaluated without enumerating chambers. Rewriting Eq. (40) as $F = ze^{tF}/(1 - F/2)$, Lagrange inversion gives

$$F_p(t) = \frac{1}{p}[u^{p-1}]e^{ptu}(1-u/2)^{-p}.$$

At $t = 0$,

$$F_p(0) = \frac{\mathrm{Cat}_{p-1}}{2^{p-1}} = \Gamma(2p-1)\int_{\Delta_{p-1}} H_p(x)dx,$$

so

$$I_p := \int_{\Delta_{p-1}} H_p(x)dx = \frac{1}{2^{p-1}p!(p-1)!}. \tag{42}$$

As a separate low-order check, the chamber formula gives $H_3(x) = m^2/2 + mq$ when $m \le q \le r$, and integration over the six chambers gives $I_3 = 1/48$. Explicit integration over all $4! = 24$ weak-order chambers gives $I_4 = 1/1152$. For boxcar pooling, $J_4 = 151/315$, hence Eq. (5) gives $B_4 = 151/504$. The deterministic supplementary check `check_pairing_recursion.py` performs the chamber enumeration for $p = 1, 2, 3, 4$ without using Eq. (40) as its expected-value generator.

Combine Eqs. (35), (38), and (42). To remove the interior restriction, split weak compositions of n according to whether some coordinate is below ηn. The boundary set contains $O_p(\eta n^{p-1} + n^{p-2})$ compositions. On it the global fixed-run bound is $O_p((n+1)^{p-1})$, while repeated use of the absolute row/column bound in Eq. (29) makes the constrained product of the $b^{(h)}$ kernels at most $O_{p,h}(T^{p-1})$. Its normalized contribution is therefore $O_{p,h}(\eta) + o(1)$. The interior contribution is uniform by Eqs. (35) and (38). Sending $n \to \infty$ and then $\eta \downarrow 0$ gives

$$A^{(h)}_{p,n}(T) \sim I_p J_p(h) n^{2p-2} T^{p-1}. \tag{43}$$

The same estimates give $|A^{(h)}_{p,n}(T)| \le C_{p,h}(n+1)^{2p-2}T^{p-1}$. Apply Eq. (43) on a compact annulus in εn, use this absolute bound on both tails, and then remove the annulus. Since

$$\sum_{n\ge 0}(1-\varepsilon)^n n^{2p-2} \sim \Gamma(2p-1)\varepsilon^{-(2p-1)},$$

Theorem 1 follows.

A.4 Identification, transform, and small-value law

Because $B_1(h) = 1$, $\rho^{(h)}_{\varepsilon,T}(dy) = y\nu^{(h)}_{\varepsilon,T}(dy)$ is asymptotically a probability measure. The aspect-one Marchenko–Pastur moments are $\mathbb{E}M^q = \mathrm{Cat}_q$. With $f_{U_h}(u) = \psi_h(u)/(2\pi)$ and $W_h = \psi_h(U_h)$,

$$\mathbb{E}W_h^q = J_{q+1}(h), \qquad \mathbb{E}[(M/2)W_h]^q = \frac{\mathrm{Cat}_q}{2^q}J_{q+1}(h) = B_{q+1}(h).$$

Since $h \in L^1$, the product has compact support, so the Hausdorff moment problem is determinate. Next-moment bounds give uniform integrability and tightness, proving weak convergence of $\rho^{(h)}_{\varepsilon,T}$. For compactly supported continuous f on $(0,\infty)$, $f(y)/y$ is bounded; therefore $\nu^{(h)}_{\varepsilon,T}$ converges vaguely to $y^{-1}\mathcal{L}(Y_h)(dy)$.

Conditioning on W_h and using the Catalan transform

$$\mathbb{E}_M\left[(1+sM/2)^{-1}\right] = \frac{\sqrt{1+2s}-1}{s}$$

proves Eq. (10).

Write $X = M/2$ and let F_h be the distribution function of W_h. The density of X is

$$f_X(x) = \pi^{-1}\sqrt{(2-x)/x}, \qquad 0 < x < 2.$$

Conditioning on W_h gives the exact smoothed density

$$f_{Y_h}(y) = \frac{1}{\pi\sqrt{y}} \int_{[y/2, \|\widehat{h}\|_\infty^2]} \frac{\sqrt{2w-y}}{w}\, dF_h(w). \tag{44}$$

If $\mathbb{E}W_h^{-1/2} < \infty$, dominated convergence in Eq. (44) proves Eq. (11). Directly from the density of U_h,

$$\mathbb{E}W_h^{-1/2} = \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{h}(u)|du.$$

If $F_h(w) \sim C\sqrt{w}$, Stieltjes integration gives

$$\int_{[a,1]} w^{-1/2} dF_h(w) \sim \frac{C}{2}\log(1/a).$$

Replacing $\sqrt{2w-y}$ by $\sqrt{2w}$ in Eq. (44) changes only an $O(y^{-1/2})$ term and proves Eq. (12). If instead F_h is regularly varying with index $0 < \beta < 1/2$, a zero-tail version of Breiman's argument applies because $\mathbb{E}X^{-\beta} < \infty$, proving Eq. (13) at the distribution-function level.

For boxcar pooling, periodic averaging under $u = x/\sqrt{w}$ gives

$$F_h(w) \sim \frac{4}{\pi^2}\sqrt{w}. \tag{45}$$

The sinc pushforward density has accumulating caustics at noncentral lobe maxima, which is why we use the distribution function rather than differentiating Eq. (45). The resulting formulas are

$$f_{Y_h}(y) \sim \frac{2\sqrt{2}}{\pi^3} y^{-1/2}\log(1/y), \qquad \nu_h((y,\infty)) \sim \frac{4\sqrt{2}}{\pi^3} y^{-1/2}\log(1/y).$$

The boxcar large-t transform check is

$$\Psi_{h_{\mathrm{box}}}(t) \sim \frac{2\sqrt{2}}{\pi^2} t^{-1/2}\log t.$$

For $h_{\exp}(t) = \sqrt{2}e^{-t}\mathbf{1}_{t\geq 0}$, $\psi_{\exp}(u) = 2/(1+u^2)$. Thus U_h is standard Cauchy, $W_h = 2B$ with $B \sim \mathrm{Beta}(1/2, 1/2)$, and $Y_h = MB$. In this case,

$$J_p(h_{\exp}) = \frac{2^{p-1}}{\sqrt{\pi}}\frac{\Gamma(p-1/2)}{\Gamma(p)}, \qquad f_{Y_h}(y) \sim \pi^{-2} y^{-1/2}\log(1/y).$$

For $h_G(t) = \pi^{-1/4}e^{-t^2/2}$, $\psi_G(u) = 2\sqrt{\pi}e^{-u^2}$ and

$$J_p(h_G) = \frac{2^{p-1}\pi^{(p-1)/2}}{\sqrt{p}}, \qquad \mathbb{E}W_h^{-1/2} = \pi^{-1/4}.$$

Hence $f_{Y_h}(y) \sim \sqrt{2}\pi^{-5/4}y^{-1/2}$, without a logarithm. Finally, $h_a(t) = a^{-1/2}h(t/a)$ implies $J_p(h_a) = a^{p-1}J_p(h)$. The dilations and constants in Table 1 follow immediately.

A.5 Population denoising risk

We prove Corollary 2 directly. For any deterministic positive semidefinite matrix C and independent circular complex Gaussian vectors $x \sim \mathcal{CN}(0, C)$ and $\xi \sim \mathcal{CN}(0, I)$, the posterior mean from $y = x + \sqrt{\lambda}\,\xi$ is

$$\widehat{x} = C(C + \lambda I)^{-1} y.$$

Diagonalizing C shows that the posterior error covariance is

$$\mathrm{Cov}(x - \widehat{x}) = \lambda C(C + \lambda I)^{-1}. \tag{46}$$

Therefore

$$\frac{\mathbb{E}\|x-\widehat{x}\|_2^2}{\mathbb{E}\|x\|_2^2}=\frac{\lambda\,\mathrm{tr}\{C(C+\lambda I)^{-1}\}}{\mathrm{tr}\,C}. \tag{47}$$

For the approximants in Corollary 2, weak convergence of the empirical spectral measures at fixed (ε,T) applies to the bounded continuous function $x\mapsto x/(x+\lambda)$. Convergence of the first moments supplies the denominator in Eq. (47). First take $N\to\infty$ and then set $\lambda=\lambda_r=rT/\varepsilon^2$. Equations (14) and (7) with $p=1$ give

$$\frac{T}{N\varepsilon}\,\mathrm{tr}\{C(C+\lambda_r I)^{-1}\}\longrightarrow\frac{1}{r}\Psi_h(1/r),\qquad\frac{1}{N}\,\mathrm{tr}\,C\sim\varepsilon^{-1}.$$

Substitution into Eq. (47) yields $\Psi_h(1/r)$ in the ordered critical limit. No teacher vector or recurrent trajectory is introduced in this argument; the corollary is exactly the Bayes risk for noisy recovery of a Gaussian representation with the stated population covariance.

A.6 Population linear-readout curves

We first verify the shared-target identity in Eq. (15). For centered real vectors z and y, expand the penalized population objective as

$$\mathrm{tr}\,S_y-2\,\mathrm{tr}(W^\top K)+\mathrm{tr}(W^\top CW)+\lambda\,\mathrm{tr}(W^\top W).$$

Its stationary point is $W=(C+\lambda I)^{-1}K$. Writing $Q=(C+\lambda I)^{-1}$ and using $QCQ=Q-\lambda Q^2$ in the unpenalized prediction error gives

$$\mathrm{tr}\,S_y-2\,\mathrm{tr}(K^\top QK)+\mathrm{tr}(K^\top QCQK)=\mathrm{tr}\,S_y-\mathrm{tr}(K^\top QK)-\lambda\,\mathrm{tr}(K^\top Q^2K),$$

as stated. The same calculation with conjugate transposes covers complex features. In particular, two pooling rules can have the same first two spectral moments of C and still differ through either the rest of C or the cross-covariance K.

Let C be a deterministic positive semidefinite Hermitian matrix, $z\sim\mathcal{CN}(0,C)$, and $\beta\sim\mathcal{CN}(0,I/N)$ independently. With the noiseless population label $y=z^*\beta$, gradient flow from $w_0=0$ satisfies

$$\dot{w}_s=-C(w_s-\beta),\qquad w_s-\beta=-e^{-sC}\beta.$$

Conditioning on β and then averaging over the isotropic teacher gives

$$\mathbb{E}|z^*(w_s-\beta)|^2=\frac{1}{N}\,\mathrm{tr}(Ce^{-2sC}),\qquad\mathbb{E}|z^*\beta|^2=\frac{1}{N}\,\mathrm{tr}\,C,$$

which proves Eq. (17). Likewise, the population ridge solution satisfies

$$w_\lambda-\beta=-\lambda(C+\lambda I)^{-1}\beta.$$

Averaging the resulting quadratic form gives

$$\mathbb{E}|z^*(w_\lambda-\beta)|^2=\frac{\lambda^2}{N}\,\mathrm{tr}\{C(C+\lambda I)^{-2}\},$$

proving Eq. (18).

For eigenvalues $\lambda_i(C_{T,h}^{(N)})$, put $y_i=(\varepsilon^2/T)\lambda_i$. With $s=u\varepsilon^2/T$ and $\lambda=T/(t\varepsilon^2)$, the two normalized finite-dimensional errors become

$$\sum_i\frac{\lambda_i}{\sum_j\lambda_j}e^{-2uy_i},\qquad\sum_i\frac{\lambda_i}{\sum_j\lambda_j}(1+ty_i)^{-2}.$$

Both test functions are bounded and continuous on $[0,\infty)$, so weak convergence of the covariance-size-biased probability measures proves Eqs. (19) and (20). Finally, differentiate the bounded integrand in $\Psi_h(t)=\mathbb{E}(1+tY_h)^{-1}$ to obtain

$$\mathbb{E}(1+tY_h)^{-2}=\Psi_h(t)+t\Psi_h'(t).$$

Substitution of Eq. (10) yields the one-dimensional integral in Eq. (20). Expanding either bounded transform at the origin gives the displayed $B_2(h)$ and $B_3(h)$ coefficients in the main text. The argument is a population calculation averaged over the isotropic teacher; it does not replace a finite-sample generalization analysis.

B ORDERED FINITE-SAMPLE GAUSSIAN GRAM LAW

This secondary consequence records how the ordered population law enters a standard Gaussian sampling model. Fix an admissible pooling kernel h from Theorem 1, write $C_{T,h} = C_T$, and let $\rho_h(dy) = y\nu_h(dy) = \mathcal{L}(Y_h)(dy)$. Choose an intermediate-window sequence (ε_r, T_r), and let $H_{r,h}$ be the spectral probability law of the bounded operator $\varepsilon_r C_{T_r,h}$. At fixed r, choose deterministic positive semidefinite $N \times N$ matrices $A_{r,N}$ such that their empirical spectral measures converge weakly to $H_{r,h}$ and $\sup_N \|A_{r,N}\| < \infty$. A diagonal quantile discretization is one explicit construction. Draw an $n \times N$ complex Gaussian feature matrix whose rows have covariance $A_{r,N}/N$, and assume

$$\frac{n}{N} \longrightarrow \alpha_r := \frac{\gamma\varepsilon_r}{T_r}, \qquad K_{r,N} = X_{r,N}X_{r,N}^*.$$

The order is $N, n \to \infty$ at fixed r, followed by $r \to \infty$.

Corollary 5 (Pooled-feature Gaussian sampling law)**.** *The empirical spectral measure of $K_{r,N}$ converges in the stated order to a compactly supported probability law $\mu_{\gamma,h}$ with*

$$\kappa_p(\mu_{\gamma,h}) = \gamma^{p-1}B_p(h), \qquad R_{\gamma,h}(w) = \Psi_h(-\gamma w). \tag{48}$$

For every finite $\gamma > 0$, the support of $\mu_{\gamma,h}$ is separated from zero. If $t_{\gamma,h}$ is the unique solution of

$$t_{\gamma,h}\Psi_h(t_{\gamma,h}) = \gamma, \tag{49}$$

then $\int x^{-1}\mu_{\gamma,h}(dx) = t_{\gamma,h}/\gamma$. If, in addition, some $c_\gamma > 0$ satisfies

$$\lim_{r\to\infty} \liminf_{N\to\infty} \Pr\{\lambda_{\min}(K_{r,N}) \geq c_\gamma\} = 1, \tag{50}$$

then for independent labels $\eta \sim \mathcal{CN}(0, \sigma^2 I_n)$ and $\widehat{w}_0 = X^* K_{r,N}^{-1}\eta$,

$$\frac{1}{n\sigma^2}\mathbb{E}_\eta\|\widehat{w}_0\|_2^2 \xrightarrow{\Pr} \frac{t_{\gamma,h}}{\gamma}. \tag{51}$$

This is an abstract Gaussian population-covariance result, not a construction from finite recurrent trajectories or a held-out prediction-risk theorem. The no-small-outlier condition in Eq. (50) is additional to weak spectral convergence.

B.1 WICK EXPANSION AND THE LIMITING TRANSFORM

If $c = (1\,2\,\cdots\,p)$, the exact complex Gaussian Wick formula is

$$\mathbb{E}\frac{1}{n}\operatorname{Tr} K_{r,N}^p = \sum_{\sigma\in S_p} n^{\#(c\sigma)-1}N^{-p} \prod_{D\in\mathrm{cyc}(\sigma)} \operatorname{Tr}(A_{r,N}^{|D|}). \tag{52}$$

The permutation inequality $\#\sigma + \#(c\sigma) \leq p+1$ leaves precisely the genus-zero, noncrossing terms. Let $\mathrm{NC}(p)$ denote the set of noncrossing partitions of $\{1,\dots,p\}$. Gaussian concentration gives, for each fixed p,

$$m_p^{(r)} \longrightarrow \sum_{\pi\in\mathrm{NC}(p)} \prod_{B\in\pi} \alpha_r^{|B|-1} m_{|B|}(H_{r,h}). \tag{53}$$

The noncrossing moment–cumulant relation (Krawczyk & Speicher, 2000) therefore gives the fixed-r free cumulants $\kappa_p^{(r)} = \alpha_r^{p-1}m_p(H_{r,h})$. Theorem 1 gives

$$m_p(H_{r,h}) = \varepsilon_r^p\tau(C_{T_r,h}^p) \sim B_p(h)(T_r/\varepsilon_r)^{p-1},$$

so

$$\kappa_p^{(r)} \longrightarrow \gamma^{p-1}B_p(h). \tag{54}$$

Because $B_p(h) = \mathbb{E}_{\rho_h}Y^{p-1}$ and Y_h is bounded under the L^1 assumption on h, these cumulants have exponential growth and determine a compactly supported probability law $\mu_{\gamma,h}$. Its R-transform is therefore

$$R_{\gamma,h}(w) = \sum_{p\geq 1}\gamma^{p-1}B_p(h)w^{p-1} = \int \frac{1}{1-\gamma y w}\,\rho_h(dy) = \Psi_h(-\gamma w),$$

which proves Eq. (48). The first checks are $m_1 = 1$, $m_2 = 1 + \gamma B_2(h)$, and $m_3 = 1 + 3\gamma B_2(h) + \gamma^2 B_3(h)$. For boxcar pooling these reduce to $m_2 = 1 + \gamma/3$ and $m_3 = 1 + \gamma + (11/40)\gamma^2$.

B.2 LOWER EDGE AND MINIMUM-NORM SENSITIVITY

The Cauchy transform obeys

$$z = G_{\gamma,h}(z)^{-1} + R_{\gamma,h}(G_{\gamma,h}(z)).$$

For $z < 0$, write $G_{\gamma,h}(z) = -s$, $t = \gamma s$, and $q_h(t) = t\Psi_h(t)$. Then

$$z = \frac{q_h(t) - \gamma}{t}, \qquad q_h'(t) = \int \frac{1}{(1+ty)^2}\rho_h(dy) > 0. \tag{55}$$

We have $q_h(0) = 0$, while $\int y^{-1}\rho_h(dy) = \infty$: the aspect-one Marchenko–Pastur factor has an infinite inverse moment, and W_h is bounded away from zero on a set of positive probability. Monotone convergence therefore gives $q_h(t) \to \infty$. Hence Eq. (49) has a unique finite solution. For $0 < t \le t_{\gamma,h}$,

$$\frac{d}{dt}\frac{q_h(t) - \gamma}{t} = \frac{tq_h'(t) - q_h(t) + \gamma}{t^2} > 0.$$

Indeed, $q_h(t) - tq_h'(t) = t^2 \int y(1+ty)^{-2}\rho_h(dy) < q_h(t) \le \gamma$. At $t = t_{\gamma,h}$ the derivative is $q_h'(t_{\gamma,h})/t_{\gamma,h} > 0$. The analytic inverse-function theorem therefore continues the real Cauchy branch through a neighborhood of zero. Since $\mu_{\gamma,h}$ is supported on $[0,\infty)$, this continuation implies $\operatorname{dist}(0, \operatorname{supp}\mu_{\gamma,h}) > 0$ and

$$\int x^{-1}\mu_{\gamma,h}(dx) = -G_{\gamma,h}(0) = \frac{t_{\gamma,h}}{\gamma}. \tag{56}$$

For independent pure-noise labels and $\widehat{w}_0 = X^*K^{-1}\eta$, conditioning on X gives

$$\mathbb{E}_\eta\|\widehat{w}_0\|_2^2 = \sigma^2 \operatorname{Tr} K^{-1}.$$

Equation (56) is a statement about the limiting law. To pass to inverse traces of finite Gram matrices, we also impose Eq. (50). Then weak convergence and the lower bound agree on the bounded continuous function $x \mapsto x^{-1}$ over the realized spectra, so

$$\frac{1}{n}\operatorname{Tr} K_{r,N}^{-1} \xrightarrow{\Pr} \int x^{-1}\mu_{\gamma,h}(dx).$$

Combining this identity with the conditional label expectation proves Eq. (51). Without Eq. (50), weak convergence alone does not justify this inverse-trace step.

This consequence cannot be inferred from participation ratio alone. The flat match $\rho_{\mathrm{flat}} = \delta_{1/3}$ produces a mean-one Marchenko–Pastur law with aspect $\gamma/3$, whose inverse moment is $(1-\gamma/3)^{-1}$ for $\gamma < 3$ and whose zero atom appears at $\gamma \ge 3$. The exact law has no finite-γ zero edge. This qualitative contrast is not unique among smooth same-moment surrogates: a Gamma law matching B_2 and B_3 also has mass accumulating without bound near zero. Accordingly, we use Corollary 5 as an exact quantitative readout-geometry consequence, not as an exclusivity or held-out-risk claim.

C EQUAL-TIME ENDPOINT IN THE SAME NORMALIZATION

Let

$$\Sigma_g = \int_0^\infty e^{-2t}e^{gct}e^{gc^*t}\,dt$$

be the stationary equal-time covariance, and let μ_ε^Σ be its spectral probability measure. Define

$$\nu_\varepsilon^{(0)}(B) = \varepsilon^{-3/2}\mu_\varepsilon^\Sigma\{\lambda : \varepsilon^2\lambda \in B\}, \qquad B \subset (0,\infty).$$

Proposition 3 (Equal-time circular critical law)**.** *Take the circular limit at fixed $g < 1$, followed by $\varepsilon = 1 - g^2 \downarrow 0$. For every fixed integer $p \ge 1$,*

$$\int y^p \nu_\varepsilon^{(0)}(dy) \longrightarrow C_p := \frac{\Gamma(2p - 3/2)}{2^p\sqrt{\pi}\,p!(p-1)!}.$$

The probability measures $\rho_\varepsilon^{(0)}(dy) = 2y\nu_\varepsilon^{(0)}(dy)$ converge weakly to the law of $2UV$, where

$$U \sim \mathrm{Beta}(1/4, 3/4), \qquad V \sim \mathrm{Beta}(3/4, 5/4)$$

are independent. If h_0 is the density of the locally finite limiting measure $\nu^{(0)}$, then

$$h_0(y) \sim \frac{2^{1/4}}{3\sqrt{\pi}\,\Gamma(3/4)^2}y^{-7/4}, \qquad y \downarrow 0.$$

Table 2: **Three observation-time regimes for boxcar pooling.** The eigenvalue scales and masses describe the sector contributing to fixed positive moments. The endpoint rows are separate limiting sequences; the middle row is the regime of Theorem 1.

regime	condition	eigenvalue scale / mass	$\tau(C_{T,\mathrm{box}}^p)$
short	$T\sqrt{\varepsilon} \ll 1$	$T\varepsilon^{-2}$ / $\varepsilon^{3/2}$	$T^p\varepsilon^{-(2p-3/2)}$
intermediate	$T\sqrt{\varepsilon} \gg 1, T\varepsilon \ll 1$	$T\varepsilon^{-2}$ / ε/T	$T^{p-1}\varepsilon^{-(2p-1)}$
long	$T\varepsilon \gg 1$	ε^{-3} / ε^2	$\varepsilon^{-(3p-2)}$

Proof. Circular Wick expansion gives

$$\tau(\Sigma_g^p) = \sum_{n\geq 0} a_{p,n}(1-\varepsilon)^n,$$

where

$$a_{p,n} = 2^{-2n-p} \sum_{|\mathbf{j}|=|\mathbf{k}|=n} \Phi(\mathbf{j},\mathbf{k}) \prod_{r=1}^{p} \binom{j_r + k_r}{j_r}.$$

The weights without Φ sum to

$$\binom{2n+p-1}{p-1}\binom{2n}{n}.$$

Here is the corresponding concentration and boundary argument. Put $\mathbf{M} = \mathbf{J} + \mathbf{K}$. After normalization, $\mathbf{M}$ is uniform over weak compositions of $2n$, and, conditional on $\mathbf{M} = \mathbf{m}$,

$$\Pr(\mathbf{J} = \mathbf{j} \mid \mathbf{M} = \mathbf{m}) = \frac{\prod_r \binom{m_r}{j_r}}{\binom{2n}{n}}, \qquad |\mathbf{j}| = n.$$

Thus each J_r is hypergeometric with $\mathbb{E}(J_r \mid \mathbf{M}) = m_r/2$ and $\mathrm{Var}(J_r \mid \mathbf{M}) \leq n/4$. For fixed p, Chebyshev's inequality and a union bound give

$$\Pr\Big\{\max_r |J_r - m_r/2| > L\sqrt{n} \mid \mathbf{M}\Big\} \leq \frac{p}{4L^2}.$$

The proportion of weak compositions with some $m_r < 2\eta n$ is $O_p(\eta) + O_p(n^{-1})$. The global fixed-run polynomial bound dominates their normalized contribution by the same order. On the complementary interior, first let $n \to \infty$ at fixed L; then every run is nonnegative in the lower comparison and the imbalance is $O(L\sqrt{n}) = o(n)$. Finally send $L \to \infty$ and $\eta \downarrow 0$. The conditional concentration, the two comparisons in Eqs. (36)–(37), the simplex identity in Eq. (42), and Stirling's formula give

$$a_{p,n} \sim \frac{n^{2p-5/2}}{2^p\sqrt{\pi}\,p!(p-1)!}.$$

The coefficients are nonnegative and obey the corresponding polynomial bound, so Abelian summation proves the moment formula.

For $q \geq 0$, direct beta moments give

$$\mathbb{E}(2UV)^q = \frac{2^q(1/4)_q(3/4)_q}{(1)_q(2)_q} = 2C_{q+1}.$$

Compact support makes the moment problem determinate and proves weak convergence of the finite size-biased measures, whose total masses converge to one. Mellin convolution of the two beta densities then gives the stated small-y constant. This proposition supplies the equal-time endpoint used in the reference implementation; it is an ordered circular limit and does not imply a finite-width moving-rank law.

D EXACT LOW-ORDER CROSSOVER IDENTITIES

The time-domain identity

$$C_T = Q_0 - [B_T\Sigma + \Sigma B_T^*], \quad Q_0 = A^{-1}A^{-*}, \quad B_T = T^{-1}A^{-2}(1 - e^{-AT})$$

was derived by Calvo et al. (2026) for a real linear recurrent model. Their C^0, C^∞, and transpose correspond here to Σ, Q_0, and the adjoint, respectively. We use the identity only to recover the short- and long-window endpoints: it shows $C_T = T\Sigma + O(T^2)$ for short windows and $C_T = Q_0 + O(T^{-1})$ for long windows. The new calculations below evaluate the non-normal circular spectrum of the exact boxcar object in the critical scaling regimes. Circular expansion gives Eq. (21) exactly.

For $p = 2$, the circular word identity

$$\tau[c^j (c^*)^k c^\ell (c^*)^m] = \mathbf{1}_{j+\ell=k+m}[1 + \min(j, k, \ell, m)]$$

sums to Eq. (22). At the first boundary, $T = \theta/\sqrt{\varepsilon}$,

$$\tau(C_T^2) \sim \theta^2 \varepsilon^{-7/2} F_2(\theta), \tag{57}$$

where

$$F_2(\theta) = \frac{16\theta^3 - 144\theta + 315 - (12\theta^2 + 132\theta + 420)e^{-\theta} + (12\theta^2 + 66\theta + 105)e^{-2\theta}}{48\theta^4}. \tag{58}$$

It satisfies $F_2(0) = 3/32$ and $F_2(\theta) \sim 1/(3\theta)$. At the second boundary, $T = \zeta/\varepsilon$,

$$\tau(C_T^2) \sim \varepsilon^{-4} G_2(\zeta), \qquad G_2(\zeta) = 1 - \frac{2}{\zeta} + \frac{2(1 - e^{-\zeta})}{\zeta^2}. \tag{59}$$

For $p = 3$, the calculation can be displayed without a fitted ansatz. Put $s_i = 1 + \omega_i^2$, $S = s_1 s_2 s_3$, and $e_2 = \omega_1\omega_2 + \omega_1\omega_3 + \omega_2\omega_3$. Direct circular Wick summation gives the real, permutation-invariant kernel

$$\tau(Q_{\omega_1} Q_{\omega_2} Q_{\omega_3}) = \frac{S\{S - g^4(3 + e_2) + 2g^6\}}{\prod_{i=1}^{3}(\varepsilon + \omega_i^2) \prod_{i<j}\{(\varepsilon + \omega_i\omega_j)^2 + (\omega_i - \omega_j)^2\}}. \tag{60}$$

An equivalent six-run generating function is

$$\frac{1 - (a_1a_2b_2b_3 + a_2a_3b_3b_1 + a_3a_1b_1b_2) + 2a_1a_2a_3b_1b_2b_3}{\prod_{i,j}(1 - a_ib_j)}, \quad a_i = \frac{g}{1 + i\omega_i}, \quad b_i = \frac{g}{1 - i\omega_i}.$$

Under the middle rescaling, the two difference-frequency integrals equal $2\pi^2$ and $\int_{\mathbb{R}} \psi(x)^3 dx = 11\pi/10$, so Eq. (60) gives

$$\tau(C_T^3) \sim T^2 \varepsilon^{-5} \frac{(2\pi^2)(11\pi/10)}{(2\pi)^3} = \frac{11}{40} T^2 \varepsilon^{-5}.$$

On fixed macroscopic separation sectors, the two-plus-one and fully separated frequency contributions are smaller by $O(T\varepsilon)$ and $O((T\varepsilon)^2)$, respectively. These low-order identities verify both scales but do not assert a complete all-order boundary law.

E FINITE-WIDTH STABILITY CALCULATION

Let G_N be complex Ginibre with variance $1/N$ entries and $J_N = g_N G_N$. Stability is equivalent to $\alpha(G_N) < g_N^{-1}$, where α is the largest real part. Cipolloni et al. (2022) prove

$$\Pr\left\{\alpha(G_N) < 1 + \sqrt{\frac{\gamma_N}{4N}} + \frac{t}{\sqrt{4N\gamma_N}}\right\} = e^{-e^{-t}} + o(1).$$

Substituting g_N^{-1} proves the proposition in Section 6. Since $g_N^{-1} - 1 = \varepsilon_N/2 + O(\varepsilon_N^2)$, the leading stability scale is $\sqrt{\log N/N}$. The exact centering should not be used as a small-N calibration: γ_N becomes positive only at extremely large N.

For Corollary 4, put $q_1 = T\sqrt{\varepsilon}$ and $q_2 = (T\varepsilon)^{-1}$. Their product is $q_1 q_2 = \varepsilon^{-1/2}$, so $\min(q_1, q_2)$ is maximized when $q_1 = q_2$, namely at $T = \varepsilon^{-3/4}$, with value $\varepsilon^{-1/4}$. The leading edge in the proposition is $\varepsilon_N = (1 + o(1))\sqrt{\log N/(2N)}$. Substitution gives Eq. (24). If instead $\varepsilon_N =$

Table 3: **Paired fixed-parameter moment contrasts at $\varepsilon = 0.20$.** Entries are exponential-minus-boxcar relative differences in percent; brackets give 95% matrix-level paired-bootstrap intervals. The participation coefficient is $\tau(C)^2/\tau(C^2)$.

N	matrices	first moment	second moment	participation	third moment
256	48	0.053	0.175	−0.054	1.86 [1.14, 2.59]
512	32	0.019	0.137	−0.097	2.22 [1.82, 2.62]
1024	24	−0.005	−0.067	0.059	1.87 [1.52, 2.22]

$c\sqrt{\log N/(2N)}$ for fixed $c > 1$, the standardized edge excess s_N diverges to $+\infty$, so the stability probability tends to one, and $T_N = \varepsilon_N^{-3/4}$ attains the same order with constant $c^{-1/4}$.

This edge result identifies a necessary stability scale, not a simultaneous finite-N covariance theorem. In particular, it supplies neither uniform control of the covariance spectrum along a moving gap nor a transfer from the complex ensemble to real recurrent matrices. A finite-width numerical study would have to respect the edge scale and separately test the covariance law. The fixed-parameter experiments below do the latter at two fixed gaps; they do not establish a joint moving-gap theorem.

F Finite-width diagnostic details

Matrix model and numerical evaluation. For each independent unit, we generated a complex Ginibre matrix $G_N = (X + iY)/\sqrt{2N}$ with independent standard-normal entries in X and Y, and set $A_N = I - \sqrt{1-\varepsilon}\,G_N$. The same G_N was used for both pooling kernels and both fixed gaps. A complex Schur decomposition and a triangular Sylvester solve gave the stationary covariance. The boxcar covariance was evaluated from the exact finite-window matrix identity, and the causal-exponential covariance from $(A_N + (aT)^{-1}I)X = \Sigma_N$ and $C = s(X + X^*)$, where a is the dilation and s the scalar output normalization. Hermitian eigendecomposition then gave the first three normalized spectral moments and the finite population-readout expressions in Eqs. (17)–(18). Residual, Hermiticity, positive-semidefiniteness, and eigendecomposition checks were applied to every covariance. Every generated matrix satisfied these checks; none was discarded or replaced.

First experiment. We used $\varepsilon \in \{0.36, 0.20\}$, $T = \varepsilon^{-3/4}$, boxcar pooling, and the intermediate-window causal-exponential dilation $a = 2/3$. At widths $128, 256, 512, 1024, 2048$, we used 48, 32, 24, 16, and 8 independent matrices, respectively. For each moment order, Figure 4a reports the maximum over the two gaps and two kernels of the absolute relative difference between the sample mean and the exact fixed-parameter circular reference. At width 2048 these maxima were 0.18%, 0.62%, and 1.03% for orders one, two, and three. The circular reference was therefore approached, but the asymptotic two-moment comparison was not yet a fixed-parameter match. At $\varepsilon = 0.20$ and width 1024, the exponential-to-boxcar ratios were 1.039, 1.296, 0.832, and 1.645 for the first moment, second moment, participation coefficient, and third moment. This outcome motivated a separate fixed-parameter experiment; it was not used to tune that experiment's matrix samples or evaluation statistics.

Independent fixed-parameter experiment. Before generating a disjoint set of matrices, we solved for the exponential dilation a and scalar output normalization s that equated its exact circular first and second moments to those of the boxcar. The resulting constants were

(ε, T)	a	s
$(0.36, 2.151657)$	0.404600	1.277751
$(0.20, 3.343702)$	0.411750	1.262662.

The finite experiment used 48, 32, and 24 independent matrices at widths 256, 512, and 1024. Primary contrasts at $\varepsilon = 0.20$ were the paired relative differences $(q_{\mathrm{exp}} - q_{\mathrm{box}})/q_{\mathrm{box}}$ for the first three moments and participation coefficient, together with the direct exponential-minus-boxcar differences in gradient-flow error at $u = 8$ and ridge-bias error at $t = 32$. Confidence intervals were percentile intervals from 10,000 paired resamples of matrices.

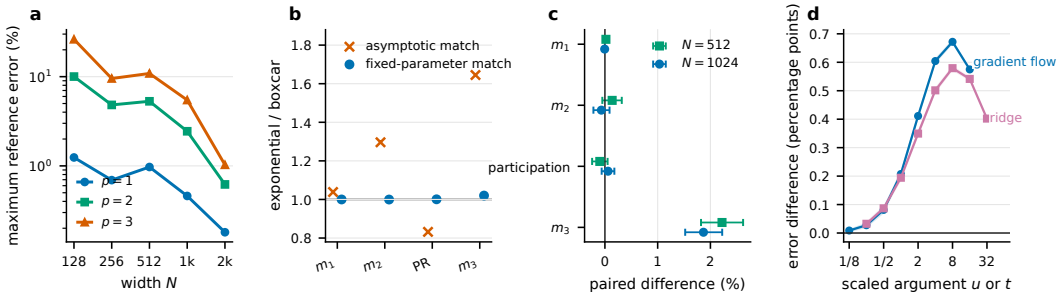

Figure 4: **Population-covariance checks at fixed parameters.** **a**, Maximum sample-mean error against the exact circular reference over two gaps, two kernels, and moments $p = 1, 2, 3$. **b**, Exponential-to-boxcar ratios at $\varepsilon = 0.20$, $T = \varepsilon^{-3/4}$, and $N = 1024$: direct intermediate-limit dilation (orange) and the independent fixed-parameter calibration (blue). The participation coefficient is $\tau(C)^2/\tau(C^2)$. **c**, Matrix-paired relative moment differences. **d**, Exponential-minus-boxcar population gradient-flow and ridge-bias curves at $N = 1024$; shading gives 95% paired-bootstrap intervals. Matrix counts are 48, 32, and 24 at $N = 256, 512, 1024$. These checks concern population covariances, not sampled predictors.

The prespecified gradient-flow grid was $u \in \{0.125, 0.25, 0.5, 1, 2, 4, 8, 16\}$, and the ridge-bias grid was $t \in \{0.25, 0.5, 1, 2, 4, 8, 16, 32\}$. At widths 256, 512, and 1024, the gradient-flow error differences at $u = 8$ were respectively 0.646, 0.661, and 0.672 percentage points, and the ridge-bias differences at $t = 32$ were 0.388, 0.396, and 0.402 percentage points. At width 1024, their 95% intervals were $[0.663, 0.681]$ and $[0.397, 0.407]$ percentage points. All 16 curve contrasts at that width had positive paired-bootstrap intervals; the maxima were 0.672 percentage points at $u = 8$ and 0.579 percentage points at $t = 8$. These are exact population linear-readout functionals of each finite covariance spectrum; they are not estimates from trained predictors or held-out examples. The experiment therefore supports the fixed-parameter statement that lower-moment matching does not determine the higher-order covariance or these isotropic population-readout curves. It does not establish the strict intermediate-window product law at jointly moving (N, ε, T).

G SHARED FUTURE-STATE FINITE-NETWORK EXPERIMENT

Real recurrent matrices and exact matching. The independent unit was a real Ginibre base matrix G_N with entries distributed as $\mathcal{N}(0, 1/N)$. The same base matrix was reused across the three gaps and three window factors, with $A_N = I - \sqrt{1-\varepsilon}G_N$. We generated 32 matrices at $N = 128$ and 24 at $N = 256$. All were stable at every tested gap and completed every required calculation; none was replaced.

For each matrix, gap, and window, let

$$q(C) = \frac{\operatorname{tr}(C)^2}{N \operatorname{tr}(C^2)}.$$

We found the causal-exponential dilation a_i satisfying $q(C_{\exp}(a_i)) = q(C_{\mathrm{box}})$ by a fixed log-grid bracket followed by Brent's method. Afterward, multiplying the exponential covariance by $s_i = \operatorname{tr}(C_{\mathrm{box}})/\operatorname{tr}(C_{\exp}(a_i))$ made the traces equal. Equality of q and trace then makes the second moments equal. The largest relative error in either moment across all records was 1.82×10^{-14} at width 128 and 1.81×10^{-14} at width 256. At width 256, the resulting dilations ranged from 0.371 to 0.502 over the complete grid, and the exponential-to-boxcar third-moment ratio ranged from 1.0049 to 1.0382.

Exact shared-window task. The target was $y = x_{t+1}$, and both pooled features ended at the same anchor t. For a forecast horizon r, the boxcar feature–target cross-covariance was evaluated as

$$K_{\mathrm{box}} = \Sigma e^{-A_N^\top r} A_N^{-\top} (I - e^{-A_N^\top T})/\sqrt{T}.$$

Table 4: **Shared future-state prediction at width 256.** Differences are matched causal exponential minus boxcar NMSE in percentage points, with simultaneous 95% matrix-bootstrap intervals. Population results use the optimal linear predictor at $T = \varepsilon^{-3/4}$ and horizon one. Finite results use validation-selected ridge penalties.

ε	population	finite, $n/d_{\mathrm{ref}} = 1$	finite, $n/d_{\mathrm{ref}} = 2$
0.18	-7.136 $[-7.293, -6.979]$	-2.578 $[-2.837, -2.319]$	-3.403 $[-3.662, -3.144]$
0.24	-6.563 $[-6.721, -6.406]$	-2.301 $[-2.560, -2.042]$	-3.911 $[-4.170, -3.652]$
0.32	-5.858 $[-6.015, -5.701]$	-1.933 $[-2.192, -1.674]$	-3.592 $[-3.851, -3.333]$

For the exponential decay $d = (a_i T)^{-1}$, define $X = \sqrt{2d}(A_N + dI)^{-1}\Sigma$. The matched exponential cross-covariance was

$$K_{\mathrm{exp}} = \sqrt{s_i}\, X^\top e^{-A_N^\top r}.$$

These expressions and the two feature covariances form the exact stationary Gaussian marginal of $(z_{\mathrm{box}}, z_{\mathrm{exp}}, y)$. The population error for either feature was

$$\mathrm{NMSE}(C, K) = \frac{\mathrm{tr}\{\Sigma - K^\top C^{-1} K\}}{\mathrm{tr}\,\Sigma}.$$

We evaluated $\varepsilon \in \{0.18, 0.24, 0.32\}$ and $T/\varepsilon^{-3/4} \in \{0.75, 1, 1.5\}$ at the primary horizon $r = 1$. Horizons $r \in \{0, 0.5, 2\}$ at the central window were secondary.

For finite-sample ridge, exact joint-Gaussian draws supplied paired features and identical targets. The reference dimension was the exactly shared participation dimension, and $n/d_{\mathrm{ref}} \in \{1, 2\}$. Each matrix contributed two training draws, each with a separate training sample, 512 validation examples, and 2048 test examples. Penalties were selected independently for each feature from the same fixed grid using validation NMSE, and the test set was evaluated once. The two draw-level differences were averaged within a matrix before inference. At the central window, d_{ref} ranged from 1.37 to 42.73 at width 128 and from 10.06 to 74.47 at width 256. The corresponding rounded training counts were 2–43 and 3–85 at width 128, and 10–74 and 20–149 at width 256, for ratios one and two, respectively.

Intervals and results. Each family of simultaneous 95% intervals uses 10,000 fixed-seed bootstrap resamples of recurrent matrices and one common maximum absolute-deviation radius. Table 4 reports the central-window results at width 256. All nine population differences over the full window-factor grid were negative, ranging from -4.99 to -8.79 percentage points. The secondary horizon analysis also stayed negative; at the central window, the magnitude fell from 14.7–16.0 points at horizon zero to 3.11–4.62 points at horizon two.

Across the six width-256 finite-sample contrasts in Table 4, the between-matrix standard deviation ranged from 0.047 to 0.664 percentage points (24 matrices per contrast). The simultaneous intervals, rather than these unadjusted standard deviations, remain the basis for inference.

The width-128 finite-sample means had the same signs, but one boxcar and one exponential fit from the same matrix and draw selected the smallest multiplier on the grid. The finite-sample comparison is therefore descriptive: the tested penalty range did not resolve every width-128 fit. All width-256 selections were interior.

Latency-matched sensitivity. Boxcar amplitude has centroid $T/2$, whereas the matched exponential has centroid $a_i T$. We aligned them by delaying whichever kernel had the younger centroid. A pure delay leaves C_{box} and C_{exp} unchanged but replaces the forecast horizon in K_h by the horizon plus the required delay. The same matrices, matching parameters, bootstrap unit, and population grid were reused. This analysis measures sensitivity to temporal alignment; it does not isolate a spectrum-only effect. At width 256 and the central window, the aligned differences for increasing ε were -2.432 ($[-2.924, -1.939]$), -2.056 ($[-2.548, -1.564]$), and -1.710 ($[-2.203, -1.218]$) percentage points, with simultaneous 95% intervals. These values are 66–71% smaller in magnitude than the unaligned contrasts, but every interval remains below zero. Temporal recency therefore accounts for most of the original contrast without accounting for all of it under this centroid-matched sensitivity. Across the complete three-gap by three-window grid at horizon one, all centroid-aligned simultaneous intervals also remained below zero. The mean differences ranged from -1.01 to -5.00 percentage points at width 256 and from -1.01 to -5.77 points at width 128.

Table 5: **Finite-sample risk at width 1024.** Ridge uses $a = 1$. Differences are matched causal exponential minus boxcar in percentage points. Rows marked † are the two primary contrasts; their simultaneous intervals are given in the text. The minimum-norm column is a prespecified descriptive analysis.

n/d_{ref}	n	ridge, boxcar	ridge, exponential	ridge difference	minimum-norm difference
$1/2$	66	0.821925	0.823030	0.110	-0.359
$1^{\dagger}$	132	0.718684	0.721106	0.242	-0.799
$2^{\dagger}$	264	0.593073	0.597160	0.409	-1.597
$10/3$	439	0.495319	0.500430	0.511	-2.468
5	659	0.418694	0.424252	0.556	-3.317

Scope. The experiment evaluates independent stationary trajectory windows from a real linear system. It is not a sample from one contiguous trajectory, a finite-width validation of Theorem 1, or evidence for nonlinear networks. The per-network dilation and scale are label-free controls designed to remove two lower-moment explanations, not a single deployable pooling rule.

H Finite-sample Gaussian readout experiment

Design. The finite-sample experiment used the $\varepsilon = 0.20$ fixed-parameter calibration from the preceding section: $T = 3.3437015$, causal-exponential dilation $a = 0.41174994$, and exponential output scale $s = 1.2626622$. Its exact circular reference moments were $m_1 = 2.4058574$ and $m_2 = 44.982213$, giving $q = m_1^2/m_2 = 0.1286764$ and $d_{\mathrm{ref}} = Nq$. New matrix, Gaussian-design, and bootstrap seed streams were disjoint from the population-covariance checks.

For each matrix and kernel, write $\bar{C} = C/m_1$. We drew two standard complex Gaussian arrays per matrix and transformed the same array by the square root of each paired covariance. Row prefixes gave $n/d_{\mathrm{ref}} \in \{1/2, 1, 2, 10/3, 5\}$, corresponding at $N = 1024$ to $n \in \{66, 132, 264, 439, 659\}$. For $S = X^*X/n$, $Q = (S + \lambda I)^{-1}$, an isotropic teacher $\beta \sim \mathcal{CN}(0, I/N)$, and unit label-noise variance, the conditional test risk divided by the realized signal variance is

$$\frac{\lambda^2 N^{-1} \operatorname{tr}(\bar{C}Q^2) + n^{-1} \operatorname{tr}(\bar{C}QSQ)}{N^{-1} \operatorname{tr}(\bar{C})}. \tag{61}$$

Teacher and noise randomness were integrated in Eq. (61); only the design was sampled. Ridge penalties were $\lambda = aN/n$ for $a \in \{1/4, 1, 4\}$, with $a = 1$ primary. The zero-penalty minimum-norm estimator used the exact row-space formula. Both primary contrasts, all secondary coordinates, and 10,000 matrix-level paired bootstrap resamples were specified before any of these experimental matrices or designs were generated.

Completion and primary inference. The fixed denominators were 48, 32, and 24 matrices at widths 256, 512, and 1024. All matrices and both design draws per matrix completed the covariance, Gram, and risk calculations. At width 1024 the two primary exponential-minus-boxcar differences were 0.002422 (simultaneous 95% interval $[0.002319, 0.002526]$) for $n/d_{\mathrm{ref}} = 1$ and 0.004087 ($[0.003984, 0.004190]$) for $n/d_{\mathrm{ref}} = 2$. Their width-512 means, 0.002506 and 0.004170, had the same sign. Each interval is its point estimate plus and minus one common radius: the 95th percentile, over the 10,000 paired resamples, of the maximum absolute deviation of the two bootstrap contrast means from their point estimates. The two primary contrasts and the width-512 sign check were specified before these experimental matrices were generated. Both simultaneous intervals excluded zero, and the width-512 means had the same sign.

Secondary geometry. At width 1024, the exponential-to-boxcar ratios of sample-Gram moments one, two, and three were 1.00005, 1.00048, and 1.00885 at $n/d_{\mathrm{ref}} = 1$, and 1.00004, 1.00046, and 1.01285 at $n/d_{\mathrm{ref}} = 2$. The paired population third-moment ratio was 1.02139, while trace, second moment, and participation ratios were within 5.1×10^{-4} of one. The complete ridge grid had positive mean contrasts at all widths, sample ratios, and positive penalties; the minimum-norm means had the opposite sign throughout. The planned global secondary bootstrap family mixed metrics with different numerical scales, so its common absolute max-deviation intervals were uninformative. For

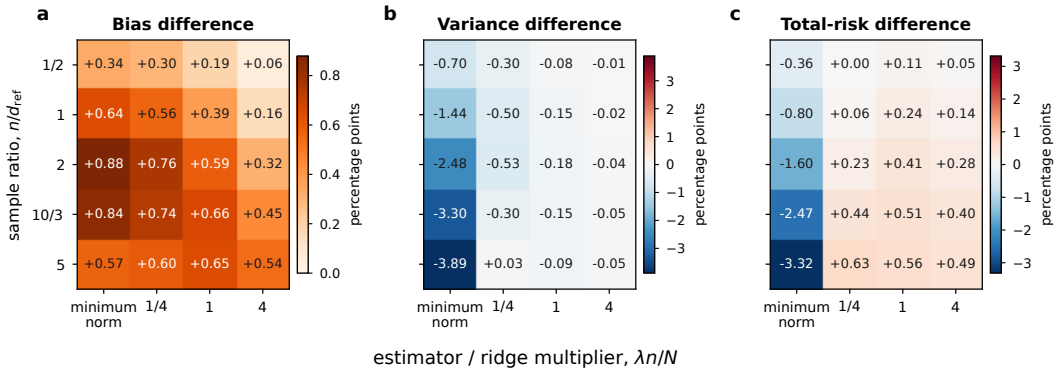

Figure 5: **Regularization changes the bias–variance balance.** Matched causal-exponential minus boxcar differences at width 1024, in percentage points. Each cell is the mean across 24 recurrent matrices after averaging the two paired Gaussian designs within a matrix. **a**, Bias is higher for the matched causal-exponential covariance throughout the grid. **b**, Its variance contribution is generally lower. **c**, Their sum changes sign between minimum norm and positive ridge penalties. This is a post hoc mechanism decomposition; studentized maximum-t intervals were constructed separately for the three component families.

the post hoc mechanism analysis, we instead formed studentized maximum-t intervals separately over the 20 bias, variance, and total-risk cells, with recurrent matrices as the resampling units. At $n/d_{\mathrm{ref}} = 1$ and $a = 1$, the bias difference was 0.388 percentage points ($[0.375, 0.401]$), the variance difference was -0.146 ($[-0.147, -0.145]$), and their sum was 0.242 ($[0.227, 0.257]$). At $n/d_{\mathrm{ref}} = 2$, the minimum-norm total difference was -1.597 points ($[-1.615, -1.580]$).

Figure 5 resolves the sign reversal into its two components. Across the grid, matched causal-exponential pooling increased conditional bias but usually reduced label-noise variance. Minimum norm gave the variance reduction enough weight to lower total risk, whereas every positive ridge penalty in the grid produced a positive mean total-risk contrast. This is an estimator-dependent bias–variance tradeoff, not a universal ranking of pooling kernels.

I Reproducibility Details

The supplementary implementation evaluates the boxcar intermediate, endpoint, and comparison laws through independent transform and probability-law calculations. A scalar checker verifies the exponential and Gaussian constants, product moments, equal-participation dilation, and Taylor coefficients. Two independent runs in the stated environment produced identical reference files; runs in other environments are compared using the reported tolerances. Running `python3 reproduce_all.py` regenerates the checks and figures.

The released records list every random seed, replicate, and curve argument. The package includes the complex-Ginibre generator, Lyapunov solver, both covariance constructions, reduced risk calculation, and a low-width direct-inversion check. It also includes the real-Ginibre shared-task construction, per-network two-moment matcher, exact joint stationary Gaussian sampler, and population future-state calculation. Automated checks validate each width-specific record, reproduce the reported interval families, and rebuild Figures 3, 5, and 4 directly from the released records.

J Expanded Limitations

For each admissible kernel, ν_h is locally finite but has infinite mass near zero in the examples studied here, whereas $\rho_h = y\nu_h$ is a probability measure. Vague convergence at fixed positive thresholds does not imply a moving-rank law, and fixed-moment convergence is not uniform when p, N, T, and ε^{-1} grow together. The product representation is an equality in distribution for the limiting size-biased spectrum, not a matrix factorization or a free multiplicative convolution.

The finite-width proposition concerns the unconditioned complex Ginibre ensemble; conditioning on stability or normalizing by the empirical edge changes the problem and requires a separate joint local law. The Gram corollary instead uses deterministic population matrices, independent Gaussian rows, and pure-noise labels. Its lower edge is asymptotic, and its minimum-norm claim assumes Eq. (50).

At fixed parameters, the covariance study checks convergence to exact circular moments and whether higher-order spectral differences survive two-moment matching. The independent-design study then tests sample-Gram geometry and ridge risk under a Gaussian design. The shared-task study adds finite real networks, exact matrixwise matching, and common future-state labels, but its examples are independent stationary windows rather than a single temporally correlated training trajectory. None of the experiments establishes nonlinear-network universality. Covariance spectra also do not determine target-aligned risk without the feature–target cross-covariance.