# RANDOM-FEATURE PRECISION FOR WEAK DIFFUSION RESIDUALS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Random Fourier features can reduce the cost of kernel diagnostics, but their approximation error must be assessed together with the noise in the observations. This paper establishes sharp frequency budgets for a signed Gaussian-kernel U-statistic of weak diffusion residuals. The analysis considers a bounded class of scalar periodic velocity loops whose own endpoint law is Gaussian. For $n$ observations and unknown squared RKHS residual norm $q = \mathrm{WSD}^2 \geq b/n$, controlling additional mean-square error at level $\varepsilon^2 q^2$ uniformly over this class requires $\Theta(\sqrt{\log n}/\varepsilon^2)$ frequencies. Preserving the exact-kernel estimator's risk within a factor of $1 + \varepsilon^2$ instead requires $\Theta(n/\varepsilon^2)$. Both budgets are optimal over admissible iid importance-sampling laws shared across the signal class, and are attained by defensive Gaussian mixtures. The key analytical ingredient is a uniform data-risk profile that tracks signal-dependent variance and the interpolation-noise floor. Together with control of the data–frequency interaction, this profile yields a sharp Gaussian proposal-width threshold. Below it, intermediate signals determine the worst-case budget, which a boundary-only analysis misses. Population-risk calculations and noisy-statistic simulations examine the predicted budget gap and difficult signal scales on finite panels. At fixed tolerance, the two precision criteria lead to nearly linear and quadratic scalar evaluation costs, respectively.

## 1 INTRODUCTION

The frequency budget for a diffusion residual estimate depends on the precision required of the statistic. Controlling extra error relative to the signal and preserving the exact estimator's mean-square error are different objectives. I show that, at fixed tolerance, this choice changes the optimal budget from $\sqrt{\log n}$ to $n$ for the same target, signed estimator, and model class. The difference comes from how feature error compares with data risk across signal scales.

The target is Weak Schedule Deviation (WSD), an RKHS norm of the weak continuity residual. A diffusion velocity can generate a probability path that differs from the stochastic interpolation of its own endpoint law; Schedule Deviation measures this inconsistency (Pfrommer et al., 2025). Kernel test functions make its weak residual observable without estimating a density score or a neural-field divergence. Random features can reduce the evaluation cost, but their additional error must be compared with the sampling error already present in the exact-kernel estimator.

Write $q = \mathrm{WSD}^2$, let $R_\infty$ be the exact estimator's MSE, and let $\Delta R$ be the additional MSE from sampling frequencies. To study this comparison, I construct periodic velocity loops with common velocity and derivative bounds. The loops return to their own Gaussian endpoint while allowing the residual to approach zero. A single importance proposal must cover all unknown signals $q \geq b/n$ in this class. Table 1 summarizes the two optimal budgets for its signed U-statistic.

Theorems 1 and 2 establish both rows. At fixed precision, signal-relative error control permits nearly linear scalar evaluation. Preserving exact-kernel MSE uniformly requires a linear frequency count, so this $O(nm)$ evaluation becomes quadratic in $n$, excluding sampling and velocity costs.

At $q \asymp n^{-1}$ the two reference scales $q^2$ and $R_\infty$ have the same order, so this boundary alone does not explain the difference between the uniform budgets. The data-risk profile derived below tracks the signal-dependent variance and the noise floor across all carrier frequencies. Their balance identifies the scale at which an insufficient Gaussian width becomes costly. Below a sharp width



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.

Table 1: **Changing the precision target changes the optimal budget.** Both rows use the same endpoint-compatible periodic class and the same signed estimator. The guarantees are uniform over $q \geq b/n$; lower bounds range over all common admissible iid importance proposals; $0 < \varepsilon < 1$.

| Required precision | $m$ (optimal order) | Attaining mixture |
|---|---|---|
| $\Delta R \leq \varepsilon^2 q^2$ | $\Theta(\sqrt{\log n}/\varepsilon^2)$ | Broad, varying with $n$ |
| $R_\infty + \Delta R \leq (1+\varepsilon^2) R_\infty$ | $\Theta(n/\varepsilon^2)$ | Fixed width, all $n$ |

threshold, the worst-case order is attained at an *intermediate* signal, strictly above the class boundary (Proposition 1). For $h = 1$, ordinary Gaussian features need order $n^{6/5}/(\log n)^{4/5}$ uniformly, whereas a fixed variance inflation of at least $3/2$ achieves order $n$. In comparison, the ordinary-Gaussian boundary budget is order $n^{4/5}$ (Theorem 3, which also characterizes risk-ratio recovery). Thus a budget chosen only at the boundary can miss the cost of intermediate signals.

Weak continuity objectives and kernel embeddings provide the residual formulation (Nakano, 2026; Berman et al., 2026; Sriperumbudur et al., 2011). Earlier feature-budget analyses study discrepancies and testing (Huggins & Mackey, 2018; Choi & Kim, 2024). I establish matching bounds on the *joint* data–frequency risk for compatible noisy-row models. The analysis gives a uniform risk profile and identifies the intermediate-signal width transition. For exact-risk precision, a two-signal argument gives the linear lower bound. For the matching upper bound, I control data variance and feature interaction across all signal scales with one fixed proposal. The proofs combine U-statistic variance (Serfling, 1980) with derivative-feature sampling, defensive mixtures and multi-target proposal design (Minh, 2016; Hesterberg, 1995; Roy & Evangelou, 2024). Appendix E gives the detailed comparison with prior work.

The experiments compare the hardest signals on a finite panel with those at the boundary, then quantify their costs under both precision criteria. Noisy U-statistic simulations cover 246 proposal–budget conditions, using paired intervals to compare observed losses with population risks. Section 5 reports where these comparisons support the theory, where the tested range leaves width behavior unresolved, and how rare spectral events limit precision.

## 2 Model, Estimators, and Precision Criteria

The reference endpoint and interpolation determine which residual is observable. I use a compatible class that holds this reference fixed while allowing the residual's scale and spectral shape to vary, so that frequency budgets can be compared across signals.

### 2.1 Weak residual and signed estimator

Fix a conditioning value and suppress it in the notation. For a declared endpoint $X_0 \sim \nu_0$ and independent $\epsilon \sim N(0, I)$, let

$$X_t = \alpha_t X_0 + \sigma_t \epsilon, \qquad T_t = \dot{\alpha}_t X_0 + \dot{\sigma}_t \epsilon, \qquad \Delta_t = v_\theta(X_t, t) - T_t. \tag{1}$$

The reference velocity is $v_t^*(x) = \mathbb{E}[T_t \mid X_t = x]$ (Albergo & Vanden-Eijnden, 2023; Lipman et al., 2023). If $P_t$ is the reference marginal, the weak continuity residual is $R_t = -\nabla \cdot [(v_\theta - v_t^*) P_t]$. Under the finite-measure and integrability assumptions in Appendix A.7, $\int \varphi \, \mathrm{d}R_t = \mathbb{E}[\nabla \varphi(X_t)^\top \Delta_t]$. For model self-consistency, $\nu_0$ must be the model's own endpoint; holding an unrelated endpoint fixed gives only a reference comparison.

For a differentiable RKHS kernel, let $Y = \nabla_1 k(X_t, \cdot)^\top \Delta_t$ and $q = \|\mathbb{E}Y\|^2 = \mathrm{WSD}_k^2$. An exact differentiable $C_0$-universal kernel identifies the residual under the stated residual, kernel-regularity, measurability, and integrability assumptions (Appendix A.7); this uses signed-measure embedding and derivative reproduction (Sriperumbudur et al., 2011; Zhou, 2008). It does not identify the velocity itself or certify a finite bank.

For $n \geq 2$ iid rows the exact estimator is

$$U_n = \frac{1}{n(n-1)} \sum_{i \neq j} \langle Y_i, Y_j \rangle, \qquad \mathbb{E}U_n = q. \tag{2}$$

I approximate the kernel with paired sine–cosine features (Rahimi & Recht, 2007; Sutherland & Schneider, 2015). With $m$ iid frequencies from $\nu$ and weights $\pi/\nu$, the same off-diagonal statistic is $U^{\nu}_{n,m}$, jointly unbiased for $q$. Conditional on a realized bank, it estimates that bank's kernel target. The estimator retains raw signed values without clipping or square roots. The statistic costs $O(nm)$ in the scalar models below. Appendix B gives the complete definitions, including fixed time grids.

### 2.2 Compatible model and precision criteria

Fix $h, T > 0$, $s = h^2$, and $\tau = \pi/(2T) > 0$. For independent $X_0, \varepsilon \sim N(0,1)$, the interpolation $X_t = \cos(\tau t)X_0 + \sin(\tau t)\varepsilon$ has zero reference velocity; $X_t \sim N(0,1)$ and its derivative $T_t \sim N(0,\tau^2)$ are independent. Choose $\rho \in C^1([0,T])$ with $\rho(0) = \rho(T) = 0$, $\rho'(t_0) = 1$ at an interior time, and set $v_{r,g,t}(x) = \rho'(t)u_{r,g}(x)$, $u_{r,g}(x) = g(rx)/r$, $r \geq 1$. For fixed $L, a_- > 0$, let $\mathcal{G}(L, a_-)$ be any nonempty class of real, $2\pi$-periodic $C^1$ functions satisfying

$$\int_0^{2\pi} g(y)\ \mathrm{d}y = 0, \qquad \|g'\|_\infty \leq L, \qquad a(g) := 2\left|\frac{1}{2\pi}\int_0^{2\pi} g(y)e^{-iy}\ \mathrm{d}y\right| \geq a_-. \tag{3}$$

Higher harmonics have unrestricted signs and phases. If $\Psi_a$ is the autonomous flow of $u_{r,g}$, the model flow is $\Psi_{\rho(t)}$, returning to the identity at both endpoints. The Gaussian reference therefore uses each model's own endpoint. At $t_0$, evaluation rows are $(X, \Delta)$ with $\Delta = u_{r,g}(X) - T_{t_0}$ and $\mathbb{E}\Delta^2 \leq \tau^2 + \pi^2 L^2$.

Use $k_h(x,y) = \exp\{-(x-y)^2/(2s)\}$, its spectral density $\pi_s = N(0, 1/s)$, and paired derivative features $g_\omega = \omega\Delta(-\sin(\omega X), \cos(\omega X))$. Then

$$z_{r,g}(\omega) = \|\mathbb{E}g_\omega\|^2 = \omega^2|\mathbb{E}[u_{r,g}(X)e^{i\omega X}]|^2, \qquad q_{r,g} = \mathbb{E}_{\pi_s} z_{r,g} = \mathrm{WSD}^2_{k_h}. \tag{4}$$

Take $n \geq 2$ iid rows and $m \geq 1$ iid frequencies independent of them. For a density $\nu$, use $w = \pi_s/\nu$ and concatenate $\sqrt{w(\omega)}g_\omega$ with normalization $m^{-1/2}$. The resulting signed off-diagonal U-statistic is $U^{\nu}_{n,m}$; the exact-kernel statistic $U_n$ uses $Y = \Delta\partial_x k_h(X, \cdot)$ instead. Raw importance weights, used without sampled-sum normalization, preserve joint unbiasedness of both statistics for $q$. The construction combines paired Fourier features and importance weighting (Rahimi & Recht, 2007; Sutherland & Schneider, 2015; Szabó & Sriperumbudur, 2019; Minh, 2016).

Write $\mu = \mathbb{E}Y$, $C = \mathrm{Cov}(Y)$, $A = \langle \mu, C\mu\rangle$, $D = \|C\|^2_{\mathrm{HS}}$, $\mathcal{R}^{\nu}_{n,m} = \mathbb{E}(U^{\nu}_{n,m} - q)^2$, $\mathcal{R}_{n,\infty} = \mathbb{E}(U_n - q)^2$, and $V_\nu = \mathrm{Var}_\nu(wz)$. U-statistic variance and orthogonal centering give

$$\begin{aligned} \mathcal{R}_{n,\infty} &= \frac{4A}{n} + \frac{2D}{n(n-1)}, \\ \Delta\mathcal{R}^{\nu}_{n,m} &:= \mathcal{R}^{\nu}_{n,m} - \mathcal{R}_{n,\infty} = \frac{V_\nu + \mathcal{J}^{\nu}_n}{m}, \qquad \mathcal{J}^{\nu}_n \geq 0. \end{aligned} \tag{5}$$

The interaction $\mathcal{J}^{\nu}_n$ captures noisy-row variation not included in feature variance. These two contributions and the exact-data risk give the total MSE (Serfling, 1980; Choi & Kim, 2024). Precise moment conditions and the transfer identity appear in Appendix A.1 and Lemma A.1.

For $b > 0$, let the unknown-scale class and admissible densities be

$$\begin{aligned} \mathcal{F}_n(b) &= \{(r,g) : r \geq 1,\ g \in \mathcal{G}(L, a_-),\ q_{r,g} \geq b/n\}, \\ \mathfrak{N}_s &= \left\{\nu : \pi_s \ll \nu, \quad \int_{\mathbb{R}} \frac{\pi_s(\omega)^2\omega^4}{\nu(\omega)}\ \mathrm{d}\omega < \infty\right\}. \end{aligned} \tag{6}$$

A proposal may depend on $n, m, h$ and known class constants, but is common to all unknown $(r,g)$ and declared before evaluation rows. I compare two uniform extra-risk criteria:

$$\mathfrak{E}_{n,m}(\nu) = \sup_{\mathcal{F}_n(b)} \frac{\Delta\mathcal{R}^{\nu}_{n,m}}{q^2_{r,g}}, \qquad \mathfrak{D}_{n,m}(\nu) = \sup_{\mathcal{F}_n(b)} \frac{\Delta\mathcal{R}^{\nu}_{n,m}}{\mathcal{R}_{n,\infty}(r,g)}. \tag{7}$$

Positive noise makes the second denominator positive. The exact-risk criterion uses the full data risk rather than a fixed $n^{-2}$ baseline.

## 3 TWO SHARP PRECISION BUDGETS

Signal-relative precision requires covering the possible signal spectra; exact-risk precision also requires comparison with the exact estimator's shrinking data risk. I establish matching budgets for both criteria on the same estimator and unknown-scale class. For exact-risk precision, the matching upper bound must cover fixed, intermediate and vanishing signals with one proposal, as analyzed in Section 4. The criteria are defined in (7), with proofs in Appendix A.

In both budget theorems, $c, C > 0$ and $n_0 \geq 2$ depend only on $s, \tau, L, a_-, b$, not on $n, m$ or precision $\varepsilon$. Let $\phi_v$ denote the $N(0, v)$ density.

**Theorem 1** (Signal-relative budget)**.** *Set* $\ell_n = \sqrt{\log(en)}$ *and* $\bar{\nu}_n = \frac{1}{2}\pi_s + \frac{1}{2}\phi_{\log(en)/s}$. *For all* $n \geq n_0$ *and all integers* $m \geq 1$*,*

$$\frac{c\ell_n}{m} \leq \inf_{\nu \in \mathfrak{N}_s} \mathfrak{E}_{n,m}(\nu) \leq \mathfrak{E}_{n,m}(\bar{\nu}_n) \leq \frac{C\ell_n}{m}. \tag{8}$$

*For* $0 < \varepsilon < 1$*, the smallest common-proposal budget for* $\Delta\mathcal{R} \leq \varepsilon^2 q^2$ *throughout* $\mathcal{F}_n(b)$ *lies between* $\lceil c\ell_n/\varepsilon^2 \rceil$ *and* $\lceil C\ell_n/\varepsilon^2 \rceil$.

**The $\sqrt{\log n}$ cost is spectral coverage.** The uniform signal profile $q_{r,g} \asymp e^{-sr^2/(s+2)}$ confines $q \geq b/n$ to a carrier interval of length $O(\sqrt{\log n})$. The lower-bound construction uses that many disjoint intervals, each carrying a fixed positive amount of normalized spectral mass for a different signal. One common proposal must assign little mass to at least one such interval. The broad component of $\bar{\nu}_n$ covers this interval, while its defensive component bounds importance weights by two. Together, the two components control both $V_\nu/q^2$ and the noisy-row interaction $\mathcal{J}_n^\nu/q^2$; the resulting budget therefore applies to the complete noisy statistic. The proposal combines multi-integrand sampling (Roy & Evangelou, 2024, Sections 3.1–3.2) with a defensive mixture (Hesterberg, 1995, Section 6.1).

**Theorem 2** (Exact-risk budget)**.** *Set* $\kappa_s^\star = \max\{1, (s+2)/(2s)\}$ *and* $\nu_s^\star = \frac{1}{2}\pi_s + \frac{1}{2}\phi_{\kappa_s^\star/s}$. *For all* $n \geq n_0$ *and all integers* $m \geq 1$*,*

$$\frac{cn}{m} \leq \inf_{\nu \in \mathfrak{N}_s} \mathfrak{D}_{n,m}(\nu) \leq \mathfrak{D}_{n,m}(\nu_s^\star) \leq \frac{Cn}{m}. \tag{9}$$

*The upper bound holds over all* $r \geq 1$*,* $g \in \mathcal{G}$*, without the signal cutoff. For* $0 < \varepsilon < 1$*, the smallest budget guaranteeing* $\mathcal{R}_{n,m}^\nu \leq (1+\varepsilon^2)\mathcal{R}_{n,\infty}$ *throughout* $\mathcal{F}_n(b)$ *is* $\Theta(n/\varepsilon^2)$. *Uniformly bounded risk ratios are attainable iff* $m_n = \Omega(n)$*; uniform ratio recovery is attainable iff* $m_n/n \to \infty$.

**Feature variance relative to data risk.** At a fixed nonzero signal, $q^2$ stays fixed while the exact data risk shrinks as $n^{-1}$. For two fixed, nonproportional normalized spectra, one common proposal leaves a feature variance bounded away from zero in at least one model, uniformly over the proposal. That variance divided by $m$ must compete with a vanishing data baseline, giving the $n/m$ lower bound through the importance-sampling variance comparison. For the upper bound, feature variance and the noisy-row interaction must both be controlled relative to data risk across signal scales. The uniform data-risk profile in Section 4 provides the denominator for this comparison; the resulting bounds determine which Gaussian widths suffice.

**The cost of preserving precision.** For this raw importance-weighted U-statistic with a common iid frequency law, the two uniform budgets give scalar-statistic evaluation costs $O(n\sqrt{\log n}/\varepsilon^2)$ and $O(n^2/\varepsilon^2)$, excluding sampling and velocity evaluation. The exact-risk guarantee therefore entails quadratic scalar-statistic evaluation at fixed tolerance.

## 4 INTERMEDIATE SIGNALS AND THE WIDTH THRESHOLD

Data projection variance and interpolation noise compete across carrier scales. When Gaussian width is insufficient, the largest asymptotic exact-risk cost occurs near this balance, beyond the local boundary. Comparing feature error with both sources of data risk also determines which widths suffice across carrier scales.

**The data-risk balance.** Lemma A.3 gives the following profile, uniformly over $r \geq 1$ and $g \in \mathcal{G}$:

$$\mathcal{R}_{n,\infty}(r,g) \asymp \frac{r^2 e^{-\gamma r^2}}{n} + \frac{1}{n(n-1)}, \qquad \gamma = \frac{s}{s+1}. \tag{10}$$

The first term captures data projection variance; the second is a persistent interpolation-noise term of order $n^{-2}$ as the carrier grows. Comparing feature error to this changing data risk requires tracking both terms. Uniformity over signed profiles requires controlling all higher harmonics and ruling out cancellation at bounded carriers through a positive lower bound. The $n^{-2}$ floor is essential when projection variance decays along a sequence of carriers.

For the fixed defensive proposal in Theorem 2, the proof establishes $V_{\nu_s^\star} \leq CA$ and $\mathcal{J}_n^{\nu_s^\star} \leq C/n$. The exact risk bounds both $A/n$ and $n^{-2}$ from below, so $(V_{\nu_s^\star} + \mathcal{J}_n^{\nu_s^\star})/\mathcal{R}_{n,\infty} \leq Cn$ uniformly. The bound establishes attainability for the complete estimator, including the data–frequency interaction.

For a narrower Gaussian, feature error relative to exact risk can be largest near the balance between the two data-risk terms, where $nr^2 e^{-\gamma r^2} \asymp 1$. The corresponding signal lies above $b/n$. This intermediate region determines the width threshold and the larger budget required below it.

**Proposition 1** (Width phase transition)**.** *Fix $\kappa \geq 1$ and take either $\nu = \phi_{\kappa/s}$ or $\nu = \frac{1}{2}\pi_s + \frac{1}{2}\phi_{\kappa/s}$. Put $t_\kappa = s(2-\kappa^{-1})$, $\rho_\kappa = 2t_\kappa/[\gamma(t_\kappa+4)]$. For sufficiently large $n$, uniformly in $m \geq 1$,*

$$\mathfrak{D}_{n,m}(\nu) \asymp \frac{B_\kappa(n)}{m}, \qquad B_\kappa(n) = \begin{cases} n, & \kappa \geq \kappa_s^\star, \\ n^{2-\rho_\kappa}\{\log(en)\}^{-\rho_\kappa}, & 1 \leq \kappa < \kappa_s^\star. \end{cases} \tag{11}$$

*Constants may depend on fixed $\kappa$. The second case requires $s < 2$ and has $\rho_\kappa < 1$; its lower bound occurs at $r_n^2 = (\log n + \log\log n + O(1))/\gamma$, with $q_{r_n,g} \asymp (n\log n)^{-(s+1)/(s+2)} \gg n^{-1}$.*

At $h = 1$, ordinary features need $B_1(n) = n^{6/5}/\{\log(en)\}^{4/5}$ under exact-risk precision, whereas fixed $\kappa \geq 3/2$ gives $B_\kappa(n) = n$. The proposal choices use Gaussian variance inflation and defensive sampling (Minh, 2016; Hesterberg, 1995), while preserving the same RBF target and unnormalized importance weights.

### 4.1 The local boundary benchmark

At $q \asymp n^{-1}$, $q^2$ and $\mathcal{R}_{n,\infty}$ have the same order, making this boundary a comparison point for the uniform budgets. I use $\cos$ as a boundary benchmark independently of membership in the specified $\mathcal{G}$. Write $\mathcal{R}_{n,m} = \mathcal{R}_{n,m}^{\pi_s}$ for ordinary Gaussian features. For the cosine sequence below, put $V_n = \mathrm{Var}_{\pi_s} z_{r_n,\cos}$ and $\mathcal{J}_n = \mathcal{J}_n^{\pi_s}$.

**Theorem 3** (Local Gaussian budget)**.** *For $g(y) = \cos y$, choose $r_n$ with $q_{r_n} \asymp n^{-1}$ and put $\eta_h = 4/(h^2+4)$. For explicit $A_h, B_{h,\tau} > 0$ in (22),*

$$\begin{gathered} V_n \sim A_h q_{r_n}^{2-\eta_h}, \qquad \mathcal{R}_{n,\infty} \sim B_{h,\tau} n^{-2}, \\ \Delta\mathcal{R}_{n,m}^{\pi_s} = (V_n/m)\{1 + o(1)\}, \end{gathered} \tag{12}$$

*where the remainder is uniform in integers $m \geq 1$. Hence $\mathcal{R}_{n,m} \asymp n^{-2} + n^{-2+\eta_h}/m$: $\mathcal{R}_{n,m_n} = O(n^{-2})$ iff $m_n = \Omega(n^{\eta_h})$, and $\mathcal{R}_{n,m_n}/\mathcal{R}_{n,\infty} \to 1$ iff $m_n/n^{\eta_h} \to \infty$. If $q_{r_n} \sim c/n$ and $m_n/n^{\eta_h} \to \lambda \in (0,\infty]$, then $n^2\mathcal{R}_{n,m_n} \to B_{h,\tau} + A_h c^{2-\eta_h}/\lambda$ with $1/\infty = 0$.*

The proof controls $\mathcal{J}_n = o(V_n)$ and identifies the complete noisy estimator's limiting risk, extending feature-moment budgets (Huggins & Mackey, 2018; Choi & Kim, 2024). Its periodic extension is Proposition A.1; Table 2 compares the local proposals. Since zero output has $O(n^{-2})$ MSE at $q \asymp n^{-1}$, risk-ratio recovery provides a finer measure of precision on this scale. For the unbiased signed estimator, the theorem determines when its risk ratio to the exact estimator tends to one.

## 5 Experiments: Difficult Signals and Precision Costs

I use controlled loops to locate difficult signals, compare the two precision budgets, and check the predicted risks of the noisy U-statistic. Before formal sampling, I fixed the profiles, proposals and comparison protocol.

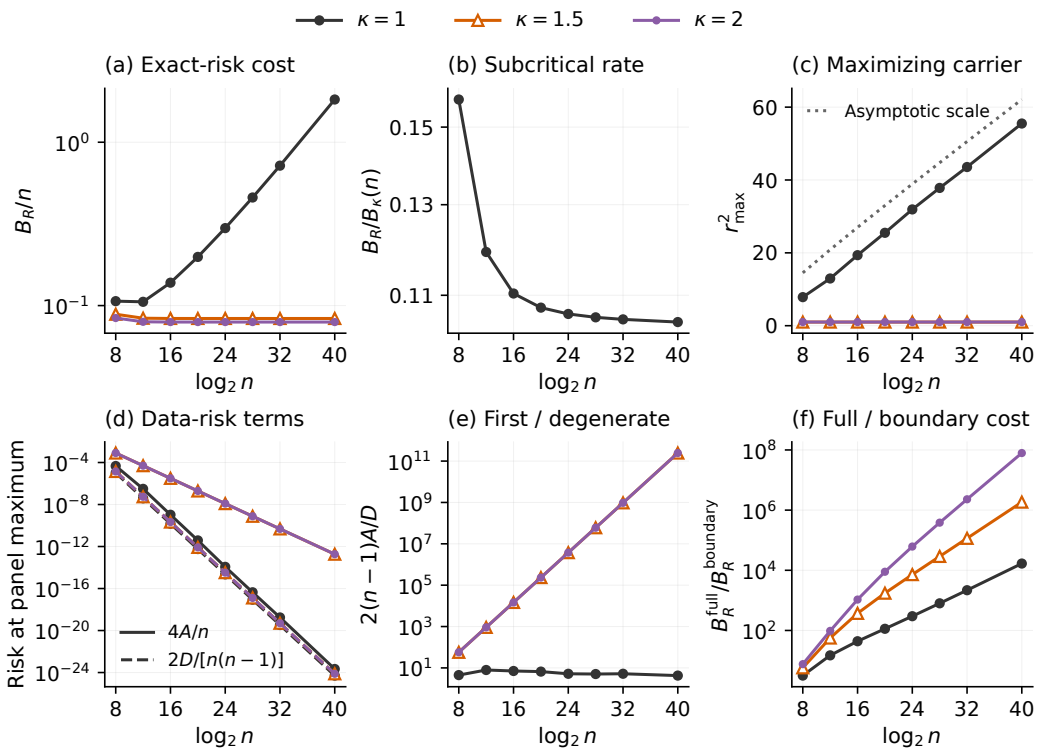



Figure 1: **Intermediate signals can be harder than the class boundary.** At $s = 1$, across all four profiles: (a,b) panel budgets and unfitted theoretical normalization; (c) maximizing carrier and an asymptotic reference scale; (d,e) the two exact-data risk terms at each panel maximum; (f) full-panel versus boundary-only budget. All panels use population calculations at each $n$ over the specified finite panel.

The loops use $T = \pi/2, t_0 = T/2$, independent $X, T_{t_0} \sim N(0,1)$, and $\Delta = g(rX)/r - T_{t_0}$. Four profiles include a cosine, a phase shift and two signed harmonic combinations. For each method at a given $(n, h)$, the proposal law is shared across all carriers and profiles: Ordinary is $\pi_s$, Broad-defensive is the proposal of Theorem 1, and Fixed-critical is that of Theorem 2. The Local-shifted mixture serves as a local-scale control. The laws and protocols are in Appendix G.

For $q \geq 0.1/n$, define $B_q^{\text{panel}} = \max_{\text{panel}}\{(V_\nu + J_n^\nu)/q^2\}$ and $B_R^{\text{panel}} = \max_{\text{panel}}\{(V_\nu + J_n^\nu)/R_\infty\}$ over the same finite panel. The scan covers eight $n$ values from $2^8$ to $2^{40}$, 64 model domains and 7,850 eligible carrier points. Independent moment calculations and two grid resolutions check sensitivity. Large $n$ enters population integrals only: $2^{40}$ does not denote a simulated data-set size. These maxima describe the specified finite panel; the continuous-class guarantees come from the theory.

**The weakest signals miss the width obstruction.** Ordinary at $s = 1, n = 2^{40}$ attains its largest $B_R$ at $g_2, r = 7.45$, where $q$ is about 13,196 times the signal cutoff. The full-panel budget is about 16,754 times the boundary-only value (Figure 1). Inspecting only $q \approx 0.1/n$ would therefore miss the difficult region. Panels (d,e) relate the maximizer to the balance between the two exact-data risk terms in Proposition 1.

At $s = 0.5$, however, widths $\kappa = 2$ and 2.5 both remain dominated by $r = 1$ through $n = 2^{40}$. Their predicted asymptotic separation is unresolved in this finite window; I report both curves without fitting rate exponents. The earlier ordinary local benchmark and its adverse total-relative-MSE results are in Appendix G.5.

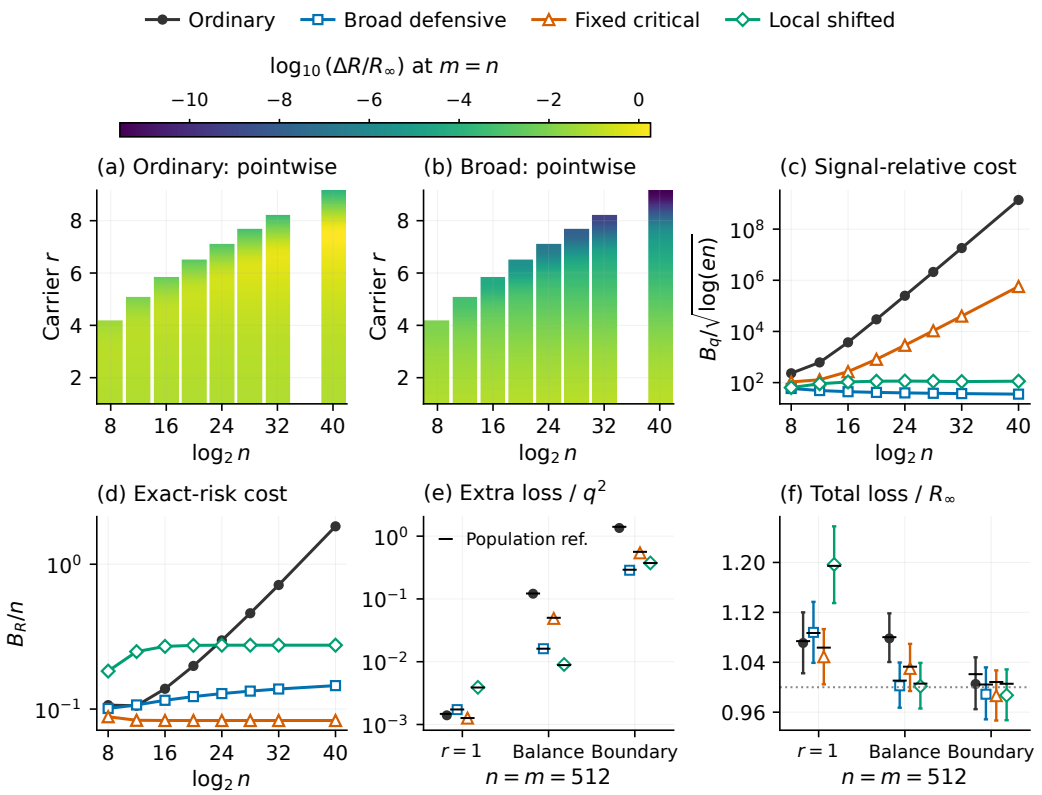



Figure 2: **The two precision requirements have different costs.** (a,b) Pointwise population additional risk relative to exact risk at $m = n$ for $g_1, s = 1$. (c,d) Maxima over all four profiles. (e,f) Actual-statistic losses at the specified $n = m = 512, g_1, s = 1$ fixed ($r = 1$), risk-balance and boundary anchors: (e) additional loss divided by $q^2$; (f) total loss divided by population $R_\infty$. Thus these panels show different precision criteria, not just a denominator change. Colored markers and error bars are means and pointwise 95% paired bootstrap intervals; black ticks are population references. Only (e,f) use simulated data sets; the other panels use large-$n$ population calculations.

**Different precision targets, different budgets.** At $s = 1, n = 2^{40}$, the same Broad-defensive proposal has $B_q^{\text{panel}} = 189.700$ but $B_R^{\text{panel}}/n = 0.145198$. For $\varepsilon = 0.5$, this means 759 frequencies for the signal-relative criterion versus approximately $6.386 \times 10^{11}$ for the exact-risk criterion (Table 3). These budgets apply to this proposal over the evaluated finite panel. At its largest signal-relative risk, the interaction $J_n^\nu$ contributes 93.64% of the full additional-risk coefficient, so a feature-variance-only calculation would omit most of the additional risk at this point.

**Checking the noisy statistic.** To check the population calculations with noisy observations, I simulate 246 proposal–budget conditions at 22 fixed, intermediate and local anchors, with $n \in \{128, 512, 2048\}$ and profiles $g_1, g_3$. An independent pilot fixed 55,808 outer groups, each containing four data sets paired across methods and frequency prefixes. The exact kernel is evaluated once per data set. The loss $(U^\nu - U^{\text{exact}})^2$ estimates additional risk; observed total MSE includes the finite-repetition cross term. The pointwise 95% paired intervals use 2,000 resamples of whole outer groups with common indices, and are not simultaneous guarantees.

Figure 2(e,f) compares the two precision criteria through the actual losses of the noisy estimator. In six prespecified Broad-defensive and Fixed-critical comparisons, increasing $m$ from 22 to 88 gives additional-risk ratios of 0.244–0.270; every individual interval contains $1/4$ (Table 5). These paired comparisons support the finite-sample inverse-budget relation for the complete statistic.

Across all conditions, the mean additional loss divided by its population reference ranges from 0.876 to 1.141. Empirical MCSE/reference meets the 10% planning target in 244/246 conditions. However, nine predeclared pressure flags apply and 15 percentile intervals miss their numerical references. Mean agreement supports the finite-sample risk calculation, while true tail precision remains uncertain. The accompanying supplementary data package gives every condition and its tail summaries.

## 6 Discussion

Uniform signal-relative error control permits nearly linear scalar evaluation. Requiring the same signed statistic to preserve exact-kernel MSE uniformly changes that cost to $O(n^2/\varepsilon^2)$. The balance between signal-dependent projection variance and the persistent interpolation-noise floor also explains where a Gaussian proposal below the width threshold is most costly: the difficult signals occur at an intermediate scale. A width at or above the threshold attains the optimal linear frequency count.

The matching bounds apply to common iid importance proposals in a scalar periodic class with nonzero interpolation noise. The proof suggests a route to broader spectra: compare feature variance with the data projection variance, then control the remaining data–frequency interaction. Adaptive proposals, dependent features, quadrature and other estimators require separate budget analyses. The analytic experiments test the noisy statistic's losses, with finite-range width behavior and spectral-tail uncertainty limiting what the comparisons establish. Experiments with learned diffusion fields would test how these mechanisms appear beyond the analytic class.

WSD measures the weak continuity residual in the chosen RKHS. A distributional path bound additionally requires the initial mismatch and the integrated bounded-Lipschitz dual norm of the residual. The examples in Appendix C show why velocity identification and sample quality require information beyond this residual. The statistical and computational scope is detailed in Appendix D.

### Reproducibility statement

Section 2 specifies the model and precision criteria; Appendix A proves the results of Sections 3 and 4. Appendix G specifies all profiles, proposals, grids, repetitions, paired analyses, and precision limitations. The accompanying implementation uses raw signed estimates and independent random streams. It includes numerical checks and the data for every figure and table. Appendix F describes the build and verification procedure.

### AI use statement

I used large language models for literature searches, guidance on theoretical directions, experimental design, code development and debugging, and manuscript drafting. Synthetic data and numerical results were obtained through the documented sampling and computation procedures using AI-assisted code. I take responsibility for the final manuscript, including its mathematical arguments, code, reported results, and attribution.

## References

Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=li7qeBbCR1t.

Aloisio Araujo and Evarist Giné. *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley, 1980.

Jules Berman, Tobias Blickhan, and Benjamin Peherstorfer. Leveraging gauge freedom for learning non-gradient population dynamics of stochastic systems, 2026. URL https://arxiv.org/abs/2605.25107v1. Version 1, 24 May 2026.

Ting Chen. On the importance of noise scheduling for diffusion models. *arXiv preprint arXiv:2301.10972*, 2023.

Ikjun Choi and Ilmun Kim. Computational-statistical trade-off in kernel two-sample testing with random Fourier features, 2024. URL https://arxiv.org/abs/2407.08976v2. Version 2, 20 May 2026; first posted 12 July 2024.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012. URL https://www.jmlr.org/papers/v13/gretton12a.html.

Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Transactions on Information Theory*, 51(4): 1261–1282, 2005. doi: 10.1109/TIT.2005.844072.

Seung-Yeal Ha and Jian-Guo Liu. A simple proof of the Cucker–Smale flocking dynamics and mean-field limit. *Communications in Mathematical Sciences*, 7(2):297–325, 2009. doi: 10.4310/CMS.2009.v7.n2.a2.

Tim Hesterberg. Weighted average importance sampling and defensive mixture distributions. *Technometrics*, 37(2):185–194, 1995. doi: 10.1080/00401706.1995.10484303. URL https://www.stat.cmu.edu/technometrics/90-00/vol-37-02/v3702185.pdf.

Jonathan Ho, Ajay N. Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Christian Horvat and Jean-Pascal Pfister. On gauge freedom, conservativity and intrinsic dimensionality estimation in diffusion models, 2024. URL https://arxiv.org/abs/2402.03845v1. Version 1, 6 February 2024.

Jonathan H. Huggins and Lester Mackey. Random feature Stein discrepancies. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper/2018/file/0f840be9b8db4d3fbd5ba2ce59211f55-Paper.pdf.

Wittawat Jitkrittum, Wenkai Xu, Zoltán Szabó, Kenji Fukumizu, and Arthur Gretton. A linear-time kernel goodness-of-fit test. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://papers.nips.cc/paper/6630-a-linear-time-kernel-goodness-of-fit-test.pdf.

Yuhao Li, Haokun Lu, and Xiaojun Song. Kernel minimum distance estimation and testing with conditional moment restrictions: A unified framework, 2026. URL https://arxiv.org/abs/2607.16605v1. Version 1, 18 July 2026.

Shanchuan Lin, Bingchen Liu, Jiashi Li, and Xiao Yang. Common diffusion noise schedules and sample steps are flawed. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 5404–5411, January 2024.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Qiang Liu, Jason Lee, and Michael Jordan. A kernelized Stein discrepancy for goodness-of-fit tests. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 276–284. PMLR, 2016. URL https://proceedings.mlr.press/v48/liub16.html.

Hà Quang Minh. Operator-valued Bochner theorem, Fourier feature maps for operator-valued kernels, and vector-valued learning, 2016. URL https://arxiv.org/abs/1608.05639v1. Version 1, 19 August 2016.

Krikamol Muandet, Wittawat Jitkrittum, and Jonas Kübler. Kernel conditional moment test via maximum moment restriction. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*, volume 124 of *Proceedings of Machine Learning Research*, pp. 41–50. PMLR, 2020. URL https://proceedings.mlr.press/v124/muandet20a.html.

Yumiharu Nakano. Continuum-marginal optimal transport: a mesh-free kernel method, 2026. URL https://arxiv.org/abs/2604.24226v3. Version 3, 26 August 2026.

Daniel Pfrommer, Zehao Dou, Christopher Scarvelis, Max Simchowitz, and Ali Jadbabaie. Is your diffusion model actually denoising? In *Advances in Neural Information Processing Systems*, volume 38, pp. 94380–94420, 2025. doi: 10.52202/085713-3157. URL https://papers.nips.cc/paper_files/paper/2025/hash/8790ba7a741c9389383575bc3e907768-Abstract-Conference.html.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems*, volume 20, pp. 1177–1184, 2007.

Vivekananda Roy and Evangelos Evangelou. Selection of proposal distributions for multiple importance sampling. *Statistica Sinica*, 34(1):27–46, 2024. doi: 10.5705/ss.202021.0049. URL https://www3.stat.sinica.edu.tw/statistica/oldpdf/A34n102.pdf.

Amirmojtaba Sabour, Sanja Fidler, and Karsten Kreis. Align your steps: Optimizing sampling schedules in diffusion models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 42947–42975, 2024.

Robert J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley, 1980.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Bharath K. Sriperumbudur, Kenji Fukumizu, and Gert R. G. Lanckriet. Universality, characteristic kernels and RKHS embedding of measures. *Journal of Machine Learning Research*, 12: 2389–2410, 2011.

Danica J. Sutherland and Jeff Schneider. On the error of random Fourier features, 2015. URL https://arxiv.org/abs/1506.02785v1. Version 1, 9 June 2015; published at UAI 2015.

Zoltán Szabó and Bharath K. Sriperumbudur. On kernel derivative approximation with random Fourier features. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 827–836. PMLR, 2019. URL https://proceedings.mlr.press/v89/szabo19a.html.

Ding-Xuan Zhou. Derivative reproducing properties for kernel methods in learning theory. *Journal of Computational and Applied Mathematics*, 220(1–2):456–463, 2008. doi: 10.1016/j.cam.2007.08.023.

# A PRECISION BUDGETS: PROOFS AND SUPPORTING RESULTS

The model, proposals and risk criteria are those of Section 2.2. After the shared risk and spectral tools, the proofs follow the main results: the signal-relative budget (Theorem 1), the exact-risk budget and width transition (Theorem 2 and Proposition 1), and the local Gaussian budget (Theorem 3). Supporting results are stated next to their derivations; standard facts are cited with the conditions needed for this observation model.

## A.1 STANDARD ESTIMATOR IDENTITIES AND JOINT-RISK TRANSFER

All statements condition on a standard-Borel pre-evaluation record fixing the geometry, bandwidth, weights and, where specified, the feature bank. Rows are conditionally iid. For continuously distributed conditions $Z$, statements concern almost every $z$ under a regular conditional law unless a pointwise version and its assumptions are fixed.

**Hilbert and U-statistic facts.** For iid Bochner-measurable $Y_i$ in a fixed separable real Hilbert space, $\mathbb{E}\|Y\|^2 < \infty$, $\mu = \mathbb{E}Y$, $C = \mathrm{Cov}(Y)$ and $q = \|\mu\|^2$, standard Hilbert CLT and Hoeffding theory (Araujo & Giné, 1980; Serfling, 1980) give

$$\mathbb{E}U_n = q, \qquad \mathrm{Var}(U_n) = \frac{4\langle \mu, C\mu\rangle}{n} + \frac{2\|C\|_{\mathrm{HS}}^2}{n(n-1)}, \qquad \mathbb{E}\|\widehat{\mu}_n - \mu\|^2 = \frac{\operatorname{tr} C}{n}. \tag{13}$$

The assumptions make $C$ trace class and the centered pair kernel $L^2$; thus only the second data moment is needed. With $G \sim N(0, C)$, $\sqrt{n}(\widehat{\mu}_n - \mu) \Rightarrow G$. At $\theta = \|\mu\| > 0$, the norm delta method has variance $\langle \mu/\theta, C\mu/\theta\rangle$; at $\mu = 0$ its limit is $\sqrt{n}\|\widehat{\mu}_n\| \Rightarrow \|G\|$ instead. For $U_n$, $\sqrt{n}(U_n - q) \Rightarrow N(0, 4\langle \mu, C\mu\rangle)$ is nondegenerate only when that variance is positive, even if $q > 0$. At $q = 0$, $nU_n \Rightarrow \sum_j \lambda_j(Z_j^2 - 1)$, where $\lambda_j$ are the eigenvalues of $C$ and $Z_j$ are iid standard normals; the series is $L^2$ since $\sum_j \lambda_j^2 < \infty$. The associated plug-in bias is the elementary identity

$$\mathbb{E}\|\widehat{\mu}_n\|^2 = q + \operatorname{tr} C/n. \tag{14}$$

For exact-kernel rows on a fixed finite grid, deterministic $w_j \geq 0$ and paired joint rows, apply (13) to $\widetilde{Y}_i = (\sqrt{w_j}Y_{i,j})_{j=1}^J$ whenever $\sum_j w_j \mathbb{E}\|Y_{i,j}\|^2 < \infty$. This gives $\mathbb{E}I_{n,J} = Q_J^{\mathrm{RBF}}$ and the same variance formula with $\widetilde{\mu}, \widetilde{C}$, including all cross-time covariance blocks. For frozen-feature rows the conditional target is $Q_{J,m}^{\mathrm{bank}}$ instead. Zero-weight times remain unidentified; this is a finite-grid identity, not a quadrature rate.

**Paired Gaussian features with noisy rows.** For fixed $\bar{X} = A(X - c) \in \mathbb{R}^p$, $\bar{\Delta} = A\Delta$ and $h > 0$, use $g_\omega = (\omega^\top \bar{\Delta})(-\sin(\omega^\top \bar{X}), \cos(\omega^\top \bar{X}))$ with $\omega \sim \pi = N(0, I_p/h^2)$ independent of the rows. Write

$$M = \mathbb{E}\bar{\Delta}\bar{\Delta}^\top, \quad F = \operatorname{tr} M < \infty, \quad H = \operatorname{tr}(M^2), \quad D_m = (F^2 + 2H/m)/h^4. \tag{15}$$

The standard paired map and derivative construction (Rahimi & Recht, 2007; Sutherland & Schneider, 2015; Szabó & Sriperumbudur, 2019) have $2m$ real coordinates and unit base-kernel diagonal. With $z(\omega) = \|\mathbb{E}_{\mathrm{data}} g_\omega\|^2$, $q_m^{\mathrm{bank}} = m^{-1}\sum_j z(\omega_j)$ and $q = \mathrm{WSD}_k^2$,

$$\mathbb{E}_{\mathrm{data}}[U_{n,m} \mid \omega_{1:m}] = q_m^{\mathrm{bank}}, \quad \mathbb{E}_\omega q_m^{\mathrm{bank}} = q, \quad \mathbb{E}(U_{n,m} - q)^2 \leq D_m\left(\frac{1}{n} + \frac{2}{n(n-1)}\right) + \frac{D_1}{m}. \tag{16}$$

Indeed $\|g_\omega\|^2 = (\omega^\top \bar{\Delta})^2$ gives the integrable Fubini envelope $F/h^2$. Jensen gives $z \leq Q = \omega^\top M\omega$; Gaussian moments give $\mathbb{E}s_m^2 = D_m$ for $s_m = m^{-1}\sum_j Q(\omega_j)$ and $\mathbb{E}Q^2 = D_1$. Conditional covariance bounds in (13), followed by total variance, yield (16). At $q = 0$, nonnegativity forces $z = 0$ almost surely, so

$$\mathbb{E}U_{n,m}^2 \leq \frac{2D_m}{n(n-1)}. \tag{17}$$

Thus the joint error is $O_p(n^{-1/2} + m^{-1/2})$ generally and $U_{n,m_n} = O_p(n^{-1})$ at the exact null for every deterministic $m_n \geq 1$. No fourth data moment or dimension-free constant is claimed; the exact-null implication does not identify alternatives from a finite bank.

**Lemma A.1** (Joint-risk transfer). *Let $\pi \ll \nu$, $w = \pi/\nu$, $Z = wz$, $Q_\nu = wQ$, $V_\nu = \mathrm{Var}_\nu Z$, $W_\nu = \mathrm{Var}_\nu Q_\nu$, and $S = \mathbb{E}_\nu Q_\nu$. Draw $m$ iid frequencies from $\nu$ independently of the rows. If $\mathbb{E}_\nu Q_\nu^2 < \infty$, then for $n \geq 2, m \geq 1$,*

$$\mathcal{R}^\nu_{n,m} = \mathcal{R}_{n,\infty} + (V_\nu + \mathcal{J}^\nu_n)/m, \qquad \mathcal{J}^\nu_n \geq 0, \tag{18}$$

$$\mathcal{J}^\nu_n \leq \frac{4}{n}\{qS + \sqrt{V_\nu W_\nu}\} + \frac{2(S^2 + W_\nu)}{n(n-1)}, \tag{19}$$

$$\mathcal{R}^\nu_{n,m} \leq \frac{4qS}{n} + \frac{2S^2}{n(n-1)} + \frac{6W_\nu}{mn(n-1)} + \frac{2V_\nu}{m}. \tag{20}$$

*These hold separately for every predeclared proposal, including changing ones. For ordinary Gaussian features, $W = 2H/h^4 \leq 2S^2$; hence under uniformly bounded $S$ and $q_n = O(n^{-1})$,*

$$\mathcal{R}_{n,m_n} = O(n^{-2}) \quad \Longleftrightarrow \quad V_n/m_n = O(n^{-2}). \tag{21}$$

*Proof.* For the dataset $D$, put $A(D,\omega) = w(\omega)U^\omega_{n,1}$. Distinct rows and $\mathbb{E}_\nu Q_\nu^2 < \infty$ ensure $A \in L^2$. Importance weighting gives $\mathbb{E}_\nu[A \mid D] = U_n(D)$ and $\mathbb{E}_D[A \mid \omega] = Z(\omega)$. The remainder $I = A - U_n - Z + q$ has both conditional means zero. Orthogonality on the product space gives (18) with $\mathcal{J}^\nu_n = \mathbb{E}I^2$; this is the standard two-source variance decomposition underlying random-feature analyses (Choi & Kim, 2024). For $q_m = m^{-1}\sum_j Z_j$ and $s_m = m^{-1}\sum_j Q_{\nu,j}$, (13) bounds conditional risk by $4q_m(s_m - q_m)/n + 2(s_m - q_m)^2/[n(n-1)]$. Use $\mathbb{E}q_m s_m = qS + \mathrm{Cov}(Z, Q_\nu)/m$, $\mathbb{E}s_m^2 = S^2 + W_\nu/m$ and $\mathrm{Var}(q_m) = V_\nu/m$. Cauchy–Schwarz and $4\sqrt{V_\nu W_\nu}/(nm) \leq V_\nu/m + 4W_\nu/(n^2 m)$ give (20). At $m = 1$, subtracting the nonnegative exact risk gives (19). The lower decomposition and upper transfer bound prove (21). □

### A.2 Shared spectral calculations for the periodic class

Use the model, class and target of (3)–(4). Here $\pi_t$ denotes the density of $N(0, 1/t)$ and $\mu_{r,g} = \mathbb{E}\{\Delta\, \partial_x k_h(X, \cdot)\}$ is the exact RKHS witness, also viewed as its evaluation function. Write $\beta = s/(s+2)$, $\eta_h = 4/(s+4)$ and

$$\begin{aligned} C_1(s) &= \frac{2\sqrt{s/(s+2)}}{(s+2)^2}, \qquad C_2(s) = \frac{32\sqrt{s/(s+4)}}{(s+4)^4}, \\ A_h &= C_2(s)C_1(s)^{\eta_h - 2}, \qquad B_{h,\tau} = 2\tau^4\sqrt{\frac{s}{s+4}}\,\frac{s^2+4s+12}{s^2(s+4)^2} > 0. \end{aligned} \tag{22}$$

All uniform constants below depend on fixed $s, L, a_-$ (and $\tau$ for data moments), not on the profile or carrier.

**Gaussian and Fourier formulas.** The Gaussian calculations use one identity: for $t, p > 0$, $j = 0, 1, 2$,

$$\int \omega^{2j} e^{-p(\omega - y)^2/2}\pi_t(\omega)\,\mathrm{d}\omega = \sqrt{\frac{t}{t+p}}\,e^{-pty^2/[2(t+p)]}P_{2j}\left(\frac{py}{t+p}, \frac{1}{t+p}\right), \tag{23}$$

where $P_0 = 1$, $P_2(\mu, v) = \mu^2 + v$ and $P_4(\mu, v) = \mu^4 + 6\mu^2 v + 3v^2$. This follows by completing the square and taking the indicated normal moment. For $c_k = (2\pi)^{-1}\int_0^{2\pi} g(y)e^{-iky}\,\mathrm{d}y$, zero mean, Parseval and integration by parts give $c_0 = 0$, $c_{-k} = \overline{c_k}$, $\sum_k k^2|c_k|^2 \leq L^2$ and $|c_k| \leq L/|k|$. Cauchy–Schwarz gives $\sum_k |c_k| < \infty$, so Gaussian integration of the absolutely and uniformly convergent series yields

$$F_{r,g}(\omega) = \frac{\omega}{r}\sum_{k \neq 0} c_k e^{-(\omega + kr)^2/2}, \qquad z_{r,g} = |F_{r,g}|^2. \tag{24}$$

Let $F^{(1)}$ retain $k = \pm 1$ and set $H = F - F^{(1)}$. For $p = 2, 4$, (23) bounds $(\omega/r)e^{-(\omega+kr)^2/2}$ in $L^p(\pi_t)$ norm by $C_{s,p}|k|e^{-tk^2r^2/[2(t+p)]}$, uniformly for $t \in [s, 2s]$, $r \geq 1$. Minkowski and $|c_k| \leq L/|k|$ give

$$\|H\|_{L^p(\pi_t)} \leq C_{s,p}L\sum_{|k| \geq 2} e^{-tk^2r^2/[2(t+p)]} \leq C'_{s,p}Le^{-2tr^2/(t+p)}. \tag{25}$$

**Lemma A.2** (Uniform signal, weighted moments and spectral peaks). *For $a = a(g)$, uniformly in the class as $r \to \infty$,*

$$q_{r,g} \sim C_1(s)a^2 e^{-\beta r^2}, \qquad \mathbb{E}_{\pi_s} z_{r,g}^2 \sim C_2(s)a^4 e^{-2sr^2/(s+4)}. \tag{26}$$

*For $\nu_\kappa = \phi_{\kappa/s}$, $\kappa \geq 1$, put $t_\kappa = s(2 - \kappa^{-1})$. Then*

$$\mathbb{E}_{\nu_\kappa}(w z_{r,g})^2 \sim \frac{32a^4\sqrt{\kappa s}}{(t_\kappa + 4)^{9/2}} e^{-2t_\kappa r^2/(t_\kappa+4)}, \tag{27}$$

*with relative error uniform in $g$ and $t_\kappa \in [s, 2s]$. There are $c_*, C_* > 0$ such that*

$$c_* e^{-\beta r^2} \leq q_{r,g} \leq C_* e^{-\beta r^2} \quad (r \geq 1,\ g \in \mathcal{G}). \tag{28}$$

*On every bounded carrier interval, both $q_{r,g}$ and $\mathbb{E}\{\mu'_{r,g}(X)^2\}$ have uniform positive minima. Finally, for $p_{r,g} = \pi_s z_{r,g}/q_{r,g}$, fixed $A, d > 0$ and $R_*$ can be chosen so that, uniformly in $g$,*

$$I_r = [2r/(s+2) - A,\ 2r/(s+2) + A], \qquad p_{r,g}(I_r) \geq d \quad (r \geq R_*). \tag{29}$$

*More precisely, $p_{r,g}$ approaches in $L^1$ the equal mixture of $N(\pm 2r/(s+2), 1/(s+2))$, uniformly in $g$.*

*Proof.* The first harmonic is $a\cos(y + \theta)$ with $a_- \leq a \leq 2L$. For its squared and fourth-power feature norms, the two unmixed lobes give the constants $C_1(t)$ and $C_2(t)$ by (23), with relative error $O(r^{-2})$. The squared cross term is $O(a^2 r^{-2} e^{-r^2})$; mixed fourth powers have exponents $-2r^2 + (2j-4)^2 r^2/[2(t+4)]$, $j = 1, 2, 3$, strictly below the unmixed exponent $-2tr^2/(t+4)$. Phase factors have modulus one. The exponential gap in (25), the lower bound on $a$ and the triangle and reverse triangle inequalities therefore preserve both equivalents, uniformly for $t \in [s, 2s]$. The identity $\pi_s^2/\nu_\kappa = \sqrt{\kappa s/t_\kappa}\,\pi_{t_\kappa}$ then proves (27).

For bounded $r$, enlarge the class to its compact closure of continuous periodic $L$-Lipschitz profiles with zero mean and $2|c_1| \geq a_-$. Arzelà–Ascoli applies since $\|g\|_\infty \leq \pi L$. Both claimed functionals are continuous on this closure times $[1, R]$: use dominated convergence and the bounded Gaussian mixed derivatives. If $q = 0$, the continuous Fourier transform of the integrable function $u_{r,g}\phi_1$ vanishes off zero, hence everywhere. Fourier uniqueness and continuity imply $g = 0$, a contradiction. If $\mathbb{E}\mu'(X)^2 = 0$, continuity makes $\mu$ constant, and its Gaussian-kernel integral tends to zero at infinity; thus $\mu = 0$ and $q = 0$, also impossible. This proves compact positivity and, with the large-carrier equivalent, (28).

Square completion turns each first-harmonic lobe of $\pi_s|F^{(1)}|^2$ into a normal density centered at $\pm 2r/(s+2)$ with variance $1/(s+2)$, multiplied by $\omega^2$. After normalization the latter factor $\omega^2/\mathbb{E}\omega^2$ tends to one in $L^1$ under that normal. The two masses tend to $1/2$, while the mixed term is negligible. Finally $\int \pi_s \big||F|^2 - |F^{(1)}|^2\big| \leq 2\|F^{(1)}\|_2\|H\|_2 + \|H\|_2^2 = o(q)$ uniformly. This proves the asserted $L^1$ approximation and the peak-mass bound. □

**The common noise floor.** Put $B_x = \partial_x k_h(x, \cdot)$ and $C_N = \tau^2\mathbb{E}(B_X \otimes B_X)$. For independent $X, X' \sim N(0,1)$, $G = X - X' \sim N(0, 2)$ gives

$$\|C_N\|_{\mathrm{HS}}^2 = \tau^4 \mathbb{E}\left[\left(\frac{1}{s} - \frac{G^2}{s^2}\right)^2 e^{-G^2/s}\right] = B_{h,\tau}/2 > 0. \tag{30}$$

Weighting this normal by $e^{-G^2/s}$ gives normalization $\sqrt{s/(s+4)}$ and variance $2s/(s+4)$; its second and fourth moments give $s^{-2}\{1 - 4/(s+4) + 12/(s+4)^2\}$, establishing the constant.

### A.3 Signal-relative budget

Use the model, class and risks in (3)–(7). Throughout this subsection, constants depend only on $s, \tau, L, a_-, b$, with $\tau > 0$; thresholds $n_0$ and constants do not depend on $m$ or precision $\varepsilon$. Write $\beta = s/(s+2)$, $M = \mathbb{E}\Delta^2$, $Q(\omega) = M\omega^2$, $S = M/s$ and $S_0 = (\tau^2 + \pi^2 L^2)/s$. The periodic Lipschitz bound gives $\tau^2 \leq M \leq \tau^2 + \pi^2 L^2$.

*Proof of Theorem 1.* The uniform signal bounds (28), including bounded carriers, imply $r^2 \le \beta^{-1}\log(C_* n/b) = O(\log(en))$ on $\mathcal{F}_n(b)$. Put $\kappa = \log(en)$, $t = s(2-\kappa^{-1})$ and $\gamma_\kappa = \phi_{\kappa/s}$. The common proposal $\bar\nu_n = (\pi_s + \gamma_\kappa)/2$ dominates each component divided by two. In particular it belongs to $\mathfrak{N}_s$, and

$$\mathbb{E}_{\bar\nu_n}(wQ)^2 \le 2\mathbb{E}_{\pi_s}Q^2 = 6S^2, \qquad W_{\bar\nu_n} \le 5S^2 \le 5S_0^2. \tag{31}$$

The Gaussian change of measure $\pi_s^2/\gamma_\kappa = \sqrt{\kappa s/t}\,\pi_t$ and the uniform fourth moment (27) give, at large $r$,

$$\frac{\mathbb{E}_{\bar\nu_n}(wz)^2}{q^2} \le C\sqrt{\kappa}\exp\left\{\frac{4sr^2}{(s+2)\{\kappa(2s+4)-s\}}\right\} \le C'\ell_n. \tag{32}$$

Uniformity of that moment for $t \in [s, 2s]$ permits the changing proposal. On bounded carrier intervals, compact positivity of $q$, $z \le Q$ and (31) give the same bound. Thus $V_{\bar\nu_n} \le C\ell_n q^2$ on the whole class. The interaction bound (19) now yields

$$\frac{\mathcal{J}_n^{\bar\nu_n}}{q^2} \le C\left\{\frac{1+\sqrt{\ell_n}}{nq} + \frac{1}{n^2q^2}\right\} \le C'\ell_n, \tag{33}$$

using $nq \ge b$ and $n \ge 2$. Equation (18) therefore proves the full joint-risk upper bound.

For the lower bound, fix any $g_* \in \mathcal{G}$; a cosine need not belong to the specified nonempty class. By (29), the normalized density $p_r = \pi_s z_{r,g_*}/q_{r,g_*}$ puts mass at least $d > 0$ in $I_r = [2r/(s+2) - A, 2r/(s+2) + A]$ for $r \ge R_*$, with uniform constants. Take carriers spaced by a fixed $\Delta_r > (s+2)A$ between $R_*$ and $\sqrt{\beta^{-1}\log(c_* n/b)}$. They all lie in $\mathcal{F}_n(b)$ by (28), and their $K_n \asymp \ell_n$ intervals are disjoint. Every common $\nu$ gives some interval mass at most $1/K_n$, so

$$1 + \frac{V_\nu(r_j, g_*)}{q^2_{r_j,g_*}} = \int \frac{p_{r_j}^2}{\nu} \ge \frac{p_{r_j}(I_{r_j})^2}{\nu(I_{r_j})} \ge d^2 K_n. \tag{34}$$

Subtract one and use $\mathcal{J}_n^\nu \ge 0$. The bad index may depend on $\nu$, establishing the stated infimum–supremum order even when $\nu$ depends on $n, m, \varepsilon$. Together with the upper bound this proves (8) for every integer $m \ge 1$; solving the two inequalities and rounding up gives the stated precision bounds with the same $n_0$. □

**Testing consequence.** For $\alpha, \zeta \in (0,1)$, reject when $U_{n,m}^{\bar\nu_n} > t_\alpha$, where

$$t_\alpha = S_0\sqrt{\frac{2(1+5/m)}{\alpha n(n-1)}}. \tag{35}$$

Under the same scalar reference, every exact null $q = 0$ with $\mathbb{E}\Delta^2 \le sS_0$ has joint level at most $\alpha$ for $n \ge 2$. There are $K_{\alpha,\zeta}, C_{\alpha,\zeta} > 0$, depending also on the fixed class parameters, such that for $n \ge n_0$, alternatives in $\mathcal{F}_n(b)$ with $q \ge K_{\alpha,\zeta}S_0/n$ have joint power at least $1-\zeta$ when $m \ge C_{\alpha,\zeta}\ell_n$. The null (for example $u = 0$) is separate from the nonzero periodic class; this sufficient guarantee is not a testing lower bound.

*Proof.* At $q = 0$, the conditional feature mean is zero almost surely. With $S_m = m^{-1}\sum_j w(\omega_j)Q(\omega_j)$, the null variance is at most $2\mathbb{E}S_m^2/[n(n-1)]$, where $\mathbb{E}S_m^2 \le S_0^2(1+5/m)$ by (31). Chebyshev proves level. Under an alternative with $q \ge KS_0/n$, the exact energy bound and the theorem give

$$\frac{\mathcal{R}_{n,m}^{\bar\nu_n}}{q^2} \le \frac{4}{K} + \frac{4}{K^2} + \frac{C\ell_n}{m}, \qquad \frac{t_\alpha}{q} \le \frac{\sqrt{24/\alpha}}{K}.$$

Choose $K$ with $4/K + 4/K^2 \le \zeta/8$ and $\sqrt{24/\alpha}/K \le 1/2$, then $m$ with $C\ell_n/m \le \zeta/8$. Joint unbiasedness and Chebyshev give nonrejection probability at most $\mathcal{R}_{n,m}^{\bar\nu_n}/(q - t_\alpha)^2 \le \zeta$. Both probabilities average the bank and evaluation rows. □

### A.4 EXACT-RISK BUDGET AND WIDTH PHASE TRANSITION

**Lemma A.3** (Uniform periodic exact-data risk)**.** *In the nonempty class* (3)*, with the fixed independent noise $\tau > 0$, let $A_{r,g} = \langle \mu_{r,g}, C_{r,g}\mu_{r,g}\rangle$ and $\gamma = s/(s+1)$. Uniformly in $g$ as $r \to \infty$,*

$$A_{r,g} \sim \frac{\tau^2 a(g)^2 s}{2(s+1)^{9/2}\sqrt{s+3}}\, r^2 e^{-\gamma r^2}. \tag{36}$$

*Uniformly over $r \geq 1$, $g \in \mathcal{G}$ and $n \geq 2$,*

$$\mathcal{R}_{n,\infty}(r,g) \asymp \frac{r^2 e^{-\gamma r^2}}{n} + \frac{1}{n(n-1)}, \tag{37}$$

*with comparison constants depending only on $s, \tau, L, a_-$.*

*Proof.* Set $H_{r,g} = \mathbb{E}[\mu'_{r,g}(X)^2]$. Independent mean-zero target noise gives the exact decomposition

$$\begin{aligned} A_{r,g} &= \tau^2 H_{r,g} + \mathrm{Var}\{u_{r,g}(X)\mu'_{r,g}(X)\}, \\ \tau^2 H_{r,g} &\leq A_{r,g} \leq (\tau^2 + \pi^2 L^2/r^2) H_{r,g}. \end{aligned} \tag{38}$$

Put $d = s+1$ and $\xi = rx/d + \theta$. Gaussian convolution and two derivatives give, for the first harmonic $a\cos(rx+\theta)/r$,

$$\begin{aligned} (\mu^{(1)}_{r,g})'(x) &= \frac{a\sqrt{s/d}}{rd^2} e^{-(x^2+sr^2)/(2d)} \big[(r^2 + d - x^2)\cos\xi - 2rx\sin\xi\big], \\ \|(\mu^{(1)}_{r,g})'\|_2^2 &\sim \frac{a^2 s}{2d^{9/2}\sqrt{d+2}} r^2 e^{-\gamma r^2}. \end{aligned} \tag{39}$$

Here $\|\cdot\|_2$ is in $L^2(N(0,1))$. The second line uses $\mathbb{E}e^{-X^2/d} = \sqrt{d/(d+2)}$; the leading cosine square averages to $1/2$ with an exponentially small phase-uniform error, and lower-degree terms are $o(r^4)$ before the prefactor $r^{-2}$. For each Fourier term $c_k e^{ikrx}/r$, the same differentiation yields $\|(\mu^{(k)})'\|_2 \leq C_s|c_k|k^2 r e^{-\gamma k^2 r^2/2}$. Absolute Fourier convergence and these summable derivative bounds justify termwise convolution in $L^2$. With $|c_k| \leq L/|k|$,

$$\|(\mu_{r,g} - \mu^{(1)}_{r,g})'\|_2 \leq C_s L r \sum_{|k|\geq 2} |k| e^{-\gamma k^2 r^2/2} \leq C'_s L r e^{-2\gamma r^2}.$$

The first harmonic has $a \geq a_-$, so the remainder is uniformly negligible, proving (36).

On bounded carrier intervals, Lemma A.2 gives uniform positive minima for $H$ on the compact Lipschitz enlargement. Thus $A \asymp r^2 e^{-\gamma r^2}$ for all $r \geq 1$. Finally $C = C_N + C_u$ with both summands positive and the fixed noise operator satisfying $\mathrm{tr}\, C_N = \tau^2/s > 0$. Consequently $0 < \|C_N\|_{\mathrm{HS}}^2 \leq \|C\|_{\mathrm{HS}}^2 \leq S_0^2$. The exact identity (5) proves the risk bound. □

*Proof of Theorem 2.* For fixed $\kappa \geq 1$, set $d_\kappa = 2t_\kappa/(t_\kappa + 4)$, where $t_\kappa = s(2 - \kappa^{-1})$. The moment (27) gives $\mathbb{E}_{\phi_{\kappa/s}}(wz)^2 \asymp e^{-d_\kappa r^2}$ uniformly at large $r$, and $d_\kappa \geq \gamma$ iff $\kappa \geq (s+2)/(2s)$. The defensive mixture dominates $\phi_{\kappa/s}/2$. At $\kappa = \kappa_s^\star$, the preceding lemma therefore gives $V_{\nu_s^\star} \leq CA$ at large carriers; bounded moments and compact positivity of $A$ extend this to all $r \geq 1$. The bounded $q, S, V, W$ and (19) give $\mathcal{J}_n^{\nu_s^\star} \leq C/n$. Since $\mathcal{R}_{n,\infty} \geq 4A/n$ and $\mathcal{R}_{n,\infty} \geq c/n^2$,

$$\frac{V_{\nu_s^\star} + \mathcal{J}_n^{\nu_s^\star}}{\mathcal{R}_{n,\infty}} \leq Cn.$$

Together with (18), this proves the upper bound without a signal cutoff, including the critical width equality.

For the lower bound, fix any $g_* \in \mathcal{G}$. The uniform two-peak approximation behind (29) supplies two fixed carriers $r_1, r_2$ with normalized spectra $p_1, p_2$ satisfying $\mathrm{TV}(p_1, p_2) \geq d_0 > 0$: choose $r_1$ with fixed positive mass in $I_{r_1}$, then $r_2$ so large that both peaks and their $L^1$ remainder put less than half

that mass there. These choices and their positive $q_i$ bounds are uniform in $g_*$. For every common admissible density, the total-variation triangle inequality and $\mathrm{TV}(p_i,\nu) \le \sqrt{\chi^2(p_i\|\nu)}/2$ give

$$\max_{i=1,2} \frac{V_\nu(r_i,g_*)}{q_i^2} = \max_i \chi^2(p_i\|\nu) \ge d_0^2. \tag{40}$$

Both fixed signals belong to $\mathcal{F}_n(b)$ for all sufficiently large $n$ and have exact risk at most $C/n$. Thus one model has $V_\nu/\mathcal{R}_{n,\infty} \ge cn$. Nonnegative interaction proves the lower bound for every common $\nu$, including choices depending on $n, m, \varepsilon$. The matching bounds give $\Theta(n/\varepsilon^2)$ for $0 < \varepsilon < 1$ and the two stated iff conditions for bounded ratios and ratio recovery. □

*Proof of Proposition 1.* For fixed $\kappa \ge 1$, the density inequalities $\pi_s \le \sqrt{\kappa}\,\phi_{\kappa/s}$ and

$$\tfrac{1}{2}\phi_{\kappa/s} \le \tfrac{1}{2}\pi_s + \tfrac{1}{2}\phi_{\kappa/s} \le \tfrac{1+\sqrt{\kappa}}{2}\phi_{\kappa/s}$$

make the two specified proposals' spectral moments comparable. Since $d_\kappa < 2\beta$ for every finite $\kappa$, $q^2 = o(\mathbb{E}_\nu(wz)^2)$ and $V_\nu \asymp e^{-d_\kappa r^2}$ uniformly at large $r$. Both proposals have bounded row moments, so $\mathcal{J}_n^\nu/\mathcal{R}_{n,\infty} \le Cn$. For $d_\kappa \ge \gamma$, the preceding upper argument and the two-spectra lower bound give order $n$, including equality.

For $\rho = d_\kappa/\gamma < 1$, necessarily $s < 2$. Lemma A.3 yields, at large carriers,

$$\frac{V_\nu}{\mathcal{R}_{n,\infty}} \asymp f_n(x) := \frac{n^2 e^{-\rho\gamma x}}{1 + nxe^{-\gamma x}}, \qquad x = r^2. \tag{41}$$

Let $x_n$ be the larger solution of $nx_n e^{-\gamma x_n} = 1$; then $x_n = (\log n + \log\log n + O(1))/\gamma$. For $x \ge x_n$, $f_n(x) \le n^{2-\rho}x_n^{-\rho}$. For $1 \le x \le x_n$, $f_n(x) \le ne^{(1-\rho)\gamma x}/x$, whose maximum is at an endpoint because its only interior critical point is a minimum. The endpoint values are $O(n)$ and $n^{2-\rho}x_n^{-\rho}$; bounded carriers also contribute at most $Cn$. At $x_n$, equality up to a factor two gives $f_n(x_n) = n^{2-\rho}x_n^{-\rho}/2$. For any fixed $g_* \in \mathcal{G}$, (28) gives $q_{\sqrt{x_n},g_*} \asymp (nx_n)^{-\beta/\gamma}$, where $\beta/\gamma = (s+1)/(s+2) < 1$; hence $nq \to \infty$ and the lower example lies in $\mathcal{F}_n(b)$ eventually. This proves the intermediate carrier and signal orders in the statement. The resulting $B_\kappa(n)$ dominates $n$, so interaction cannot change the order. Equation (18) supplies the factor $1/m$ for every integer $m \ge 1$. Constants and the sufficiently-large-$n$ threshold may also depend on fixed $\kappa$, never on $m$. □

### A.5 Local Gaussian budget and changes of proposal

*Proof of Theorem 3.* For $u_r(x) = \cos(rx)/r$, the bounded, globally Lipschitz autonomous flow $\Psi_a$ exists for all $a$. The prescribed flow is $\Psi_{\rho(t)}$, so $\rho(0) = \rho(T) = 0$ makes both endpoint maps the identity. At the test time, $\Delta = u_r(X) - T$ with independent $X \sim N(0,1)$ and $T \sim N(0,\tau^2)$. In particular $\mathbb{E}\Delta^2 = \tau^2 + (1 + e^{-2r^2})/(2r^2)$; along the whole path it is bounded by $\tau^2 + \|\rho'\|_\infty^2$. The weak residual has integrable density $\{\sin(rx) + x\cos(rx)/r\}\phi_1(x)$. Equation (24) specializes to $z_r = \omega^2\{e^{-(\omega-r)^2/2} + e^{-(\omega+r)^2/2}\}^2/(4r^2)$. The Gaussian identity (23) gives the finite-carrier formula used in the experiments:

$$q_r = \frac{\sqrt{s/(s+2)}}{2r^2}\left[\left\{\frac{1}{s+2} + \frac{4r^2}{(s+2)^2}\right\}e^{-sr^2/(s+2)} + \frac{e^{-r^2}}{s+2}\right]. \tag{42}$$

For completeness, its exact spectral second moment is

$$\mathbb{E}_{\pi_s} z_r^2 = \frac{\sqrt{s/(s+4)}}{16r^4}\sum_{j=0}^{4}\binom{4}{j}P_4\left(\frac{(2j-4)r}{s+4}, \frac{1}{s+4}\right)e^{-2r^2+(2j-4)^2r^2/[2(s+4)]}. \tag{43}$$

The two outer terms and $q_r^2 = o(\mathbb{E}z_r^2)$ yield $V_r \sim A_h q_r^{2-\eta_h}$, also directly from Lemma A.2. Hence $q_{r_n} \asymp n^{-1}$ gives $r_n^2 = \beta^{-1}\{\log n + O(1)\}$ and $V_n \asymp n^{-2+\eta_h}$. Here $S = \mathbb{E}\Delta^2/s$ and $W = 2S^2$ are bounded, so (19) implies $\mathcal{J}_n/V_n = O(n^{-\eta_h} + n^{-\eta_h/2}) = o(1)$.

For the exact risk, $C_r = C_N + C_{u,r}$ with positive summands and $\operatorname{tr} C_{u,r} \le 1/(sr^2) \to 0$. Thus $\|C_r\|_{\mathrm{HS}}^2 \to B_{h,\tau}/2$ by (30), with the finite-sample lower bound $\mathcal{R}_{n,\infty} \ge B_{h,\tau}/[n(n-1)]$.

Lemma A.3 gives $A_r = O(r^2 e^{-sr^2/(s+1)}) = o(q_r)$, so $nA_{r_n} \to 0$. The exact variance (5) now gives $n^2\mathcal{R}_{n,\infty} \to B_{h,\tau}$. Substitution in (18) proves all risk orders, both equivalences and the stated crossover constant. The excess-risk relative error is independent of $m$, hence uniform for integers $m \geq 1$. For $q_{r,m} = m^{-1}\sum_j z_r(\omega_j)$, independence also gives

$$\frac{\mathbb{E}(q_{r,m} - q_r)^2}{q_r^2} = \frac{V_r}{mq_r^2} \sim \frac{A_h q_r^{-\eta_h}}{m}, \qquad \mathcal{R}_{n,m} \geq V_r/m \quad (n \geq 2), \tag{44}$$

where the comparison constants are uniform in $m, n$ for large $r$. Although $q_{r,m} > 0$ almost surely even at $m = 1$, its relative accuracy therefore deteriorates along weak alternatives. □

**Proposition A.1** (Uniform extension to periodic profiles)**.** *For the nonempty class* (3)*, the same own-endpoint construction has uniformly bounded velocities, spatial Lipschitz constants and energies. Uniformly in $g$ as $r \to \infty$,*

$$V_{r,g} \sim A_h a(g)^{2\eta_h} q_{r,g}^{2-\eta_h}. \tag{45}$$

*For arbitrary $g_n \in \mathcal{G}$ and $q_{r_n,g_n} \asymp n^{-1}$, the risk-order statement and both budget equivalences of Theorem 3 hold uniformly, as does $\Delta\mathcal{R}^{\pi_s}_{n,m} = (V_{r_n,g_n}/m)\{1 + o(1)\}$ uniformly in $m \geq 1$. The explicit crossover constant stated there concerns the cosine sequence.*

*Proof.* Zero mean and periodic Lipschitz continuity give $\|u_{r,g}\|_\infty \leq \pi L/r$ and $\|u'_{r,g}\|_\infty \leq L$. The preceding flow argument applies, with energy at most $\tau^2 + \pi^2 L^2 \|\rho'\|^2_\infty$ and finite signed residual $-(u_{r,g}\phi_1)'$. The two moment equivalents in Lemma A.2 give (45); the powers of $a$ are $4 - 2(2 - \eta_h) = 2\eta_h$. The uniform energy bound and Lemma A.3 supply the same interaction and exact-risk estimates as above. Since $a_- \leq a(g) \leq 2L$, all constants and remainders are uniform in the profile. □

**Changing the frequency law.** Trace sampling, broader Gaussian derivative features and defensive importance mixtures are established constructions (Minh, 2016; Huggins & Mackey, 2018; Hesterberg, 1995). The following calculation keeps the target and noisy data law fixed and determines their joint-risk budgets in this compatible local class.

**Proposition A.2** (Local proposal budgets)**.** *Fix $s, \tau, L, a_- > 0$ and $0 < c_- < c_+ < \infty$. Consider*

$$g_n \in \mathcal{G}(L, a_-), \qquad c_-/n \leq q_{r_n,g_n} \leq c_+/n. \tag{46}$$

*For the proposals in Table 2, with raw weights $w = \pi_s/\nu$, uniformly over this class and $m \geq 1$,*

$$\Delta\mathcal{R}^{\nu}_{n,m} \asymp \frac{b_\nu(n)}{mn^2}, \qquad \mathcal{R}^{\nu}_{n,m} \asymp n^{-2}\{1 + b_\nu(n)/m\}. \tag{47}$$

*Consequently $O(n^{-2})$ risk holds iff $m = \Omega(b_\nu(n))$, and $\mathcal{R}^{\nu}_{n,m}/\mathcal{R}_{n,\infty} \to 1$ iff $m/b_\nu(n) \to \infty$. For every deterministic $\kappa_n \geq 1$, with constants independent of this choice,*

$$\Delta\mathcal{R}^{N(0,\kappa_n/s)}_{n,m} \geq c\sqrt{\log n}/(mn^2) \tag{48}$$

*for all sufficiently large $n$ and $m \geq 1$. The shifted proposal in the table, for fixed $\epsilon \in (0, 1)$, is*

$$\begin{aligned} v_s &= (s+2)^{-1}, \qquad a_n = 2\sqrt{\log n}/\sqrt{s(s+2)}, \\ \nu_n(\omega) &= \epsilon\pi_s(\omega) + \frac{1-\epsilon}{2}\{\phi_{v_s}(\omega - a_n) + \phi_{v_s}(\omega + a_n)\}. \end{aligned} \tag{49}$$

*It depends on $n, h, \epsilon$, not on the carrier, profile, target value or observations. Constants in* (47) *may depend on the fixed class parameters, fixed $\kappa$ or $\epsilon$, and the fixed comparison bounds for $\kappa_n/\log n$.*

*Proof.* Throughout, $r^2 = \beta^{-1}\{\log n + O(1)\}$ uniformly by Lemma A.2. Put $M_r = \mathbb{E}\Delta^2 = \tau^2 + O(r^{-2})$, $Q = M_r\omega^2$ and $S = M_r/s$. The ordinary row follows from the preceding proposition. For Energy, $w_E = (s\omega^2)^{-1}$ almost surely makes $w_E Q = S$ constant, so $W_E = 0$. Applying (23) with the resulting quadratic rather than quartic factor gives

$$\mathbb{E}_{\nu_E}(w_E z)^2 \sim \frac{2a(g)^4}{\sqrt{s}(s+4)^{5/2}r^2} e^{-2sr^2/(s+4)}. \tag{50}$$

Table 2: Sharp local budgets for (46). Each row obeys (47); risk-ratio recovery requires $m/b_\nu(n) \to \infty$. All banks are iid and independent of rows.

| Frequency proposal | Budget scale $b_\nu(n)$ |
|---|---|
| Ordinary $\pi_s = N(0, 1/s)$ | $n^{\eta_h}$ |
| Energy $\nu_E = s\omega^2\pi_s$ | $n^{\eta_h}/\log n$ |
| Fixed $N(0, \kappa/s)$, $\kappa > 1$ | $n^{\eta_\kappa}$, $\eta_\kappa = 4/\{\kappa(2s+4) - s\}$ |
| Growing $N(0, \kappa_n/s)$, $\kappa_n \asymp \log n$ | $\sqrt{\log n}$ |
| Shifted defensive mixture (49) | 1 |

Uniformity follows as in (25): the norm of one weighted harmonic is at most $C_s|c_k||k|^{1/2}r^{-1/2}e^{-sk^2r^2/[2(s+4)]}$, and summing $|k| \geq 2$ is exponentially smaller than the first harmonic. For a Gaussian proposal, use (27) and the exact row moment

$$S^2 + W_\kappa = E_{\nu_\kappa}(w_\kappa Q)^2 = \frac{3M_r^2\sqrt{\kappa s}}{t_\kappa^{5/2}}, \qquad t_\kappa = s(2 - \kappa^{-1}). \tag{51}$$

For fixed $\kappa$, the exponent identity $2t_\kappa/(t_\kappa + 4) = \beta(2 - \eta_\kappa)$ gives $V_\kappa \asymp n^{-2}b_\kappa$; similarly $V_E \asymp n^{-2}b_E$. Since both budget scales diverge and $W$ is bounded, (19) gives $\mathcal{J}_n/V = O(b^{-1} + b^{-1/2}) = o(1)$.

If $\kappa_n \asymp \log n$, the same uniform Gaussian moment gives

$$\frac{\mathbb{E}_{\nu_{\kappa_n}}(wz)^2}{q^2} \asymp \sqrt{\kappa_n}\exp\left\{\frac{4sr^2}{(s+2)\{\kappa_n(2s+4) - s\}}\right\} \asymp \sqrt{\log n}.$$

Thus $V \asymp n^{-2}\sqrt{\log n}$ and $W \asymp \sqrt{\log n}$. Keeping this growing row moment in (19) gives $\mathcal{J}_n = O(n^{-2}\sqrt{\log n})$; it need not be $o(V)$. For the centered-class lower bound, $p = \pi_s z/q$ assigns mass at least $d$ to the fixed-width interval $I_r$ of (29). There $|\omega| \geq c_s r$ eventually, whereas every centered normal density satisfies $\phi_{\sigma^2}(\omega) \leq 1/(\sqrt{2\pi e}|\omega|)$. Hence $\nu(I_r) \leq C/r$ for every variance, and Cauchy–Schwarz gives $1 + V/q^2 = \int p^2/\nu \geq p(I_r)^2/\nu(I_r) \geq cr$. Equation (18) proves (48).

For the shifted mixture, $w \leq \epsilon^{-1}$ implies $S^2 + W \leq \epsilon^{-1}\mathbb{E}_{\pi_s}Q^2 = O(1)$. Write $\mu_r = 2r/(s+2)$ and $v = v_s$. The local condition gives $|\mu_r - a_n| = O(r^{-1})$. Square completion bounds the first-harmonic spectrum by $\pi_s|F^{(1)}|^2 \leq Ca^2e^{-\beta r^2}(\omega^2/r^2)\{\phi_v(\omega - \mu_r) + \phi_v(\omega + \mu_r)\}$. The identity

$$\frac{\phi_v(\omega - \mu)^2}{\phi_v(\omega - a)} = e^{(\mu-a)^2/v}\phi_v(\omega - (2\mu - a))$$

and its fourth Gaussian moment give $\int \pi_s^2|F^{(1)}|^4/\nu_n = O(q^2)$. The defensive part and (25) give $\int \pi_s^2|H|^4/\nu_n \leq \epsilon^{-1}\|H\|^4_{L^4(\pi_s)} = O(e^{-8sr^2/(s+4)}) = o(q^2)$. Minkowski in $L^4(\pi_s^2/\nu_n)$ yields $V = O(q^2)$, and (19) gives $\mathcal{J}_n = O(n^{-2})$. For the strict lower bound, choose a fixed-width union $U_r$ around both peaks with $p(U_r) \geq 1 - \epsilon/4$, using the uniform $L^1$ limit in Lemma A.2. Then $\pi_s(U_r) \to 0$ and $\nu_n(U_r) \leq 1 - \epsilon + o(1)$, so $1 + V/q^2 \geq (1 - \epsilon/4)^2/(1 - \epsilon + o(1)) > 1 + c$. Thus $V \asymp q^2$. In all rows, the nonnegative decomposition (18) and $\mathcal{R}_{n,\infty} \asymp n^{-2}$ now prove (47) and both equivalences. □

At $s = 1$, Energy saves a logarithmic factor over $n^{4/5}$ and fixed $\kappa = 2$ gives $n^{4/11}$, without implying finite-sample numerical dominance. The centered-Gaussian bound restricts that proposal class. The shifted law changes with $n$ even for fixed $m$: it gives $O(n)$ scalar-statistic cost at fixed $m$, while $m = \lceil \log\log(n+e^e) \rceil$ recovers the exact-risk ratio with $O(n \log\log n)$ cost, excluding sampling and velocity evaluation. These are local-class MSE conclusions, not finite-bank identification, uniform power for all $q \geq c_-/n$, or guarantees for all bounded Lipschitz velocities. The zero mean and nonvanishing first harmonic are essential to the specified spectral class.

### A.6 Moment budgets and residual-preserving rotations

All expectations condition on the fitted geometry, time and reference law; evaluation rows and sampled frequencies are independent. The paired maps and signed-statistic identities are those of Ap-

pendix A.1. The residual $\bar{\Delta}$ includes interpolation noise. Unless stated otherwise, only its second moment is assumed finite.

**Covariance and fixed-reference local alternatives.** For a separable real RKHS with strongly measurable derivative features, jointly $C^2$ kernel, bounded diagonal and derivative reproduction, suppose $M_k^2 = \sup_x \operatorname{tr}[\nabla_x \nabla_y k(x,y)]_{y=x} < \infty$. Write $B_x a = \sum_j a_j \partial_{x_j} k(x,\cdot)$, so $\|B_x\|_{\mathrm{op}} \le M_k$. For $v^\star(X) = \mathbb{E}[T \mid X]$, $\delta = v_\theta - v^\star$ and $\mathbb{E}\|T\|^2 + \mathbb{E}\|\delta(X)\|^2 < \infty$, total covariance gives the positive trace-class identity

$$\operatorname{Cov}\{B_X(v_\theta(X) - T)\} = \operatorname{Cov}\{B_X \delta(X)\} + \mathbb{E}[B_X \operatorname{Cov}(T \mid X) B_X^*]. \tag{52}$$

Indeed, the centered target has conditional mean zero, eliminating the cross terms; the noise term's expected trace is at most $M_k^2 \mathbb{E}\|T\|^2$. If $\delta = 0$ almost surely, the first term vanishes; a zero residual measure need not make it vanish.

For a fixed reference and kernel, put $Y_0 = B_X(v_0(X) - T)$, $H_u = B_X u(X)$, $b_u = \mathbb{E}H_u$ and $C_0 = \operatorname{Cov}(Y_0)$. If $\mathbb{E}Y_0 = 0$ and $\mathbb{E}\|Y_0\|^2 + \mathbb{E}\|H_u\|^2 < \infty$, the statistic for $v_n = v_0 + n^{-1/2}u$ satisfies

$$\mathbb{E}U_n = \frac{\|b_u\|^2}{n}, \qquad nU_n \Rightarrow \|G + b_u\|^2 - \operatorname{tr} C_0, \quad G \sim N_{\mathcal{H}_k}(0, C_0). \tag{53}$$

The Hilbert CLT and law of large numbers give $n^{-1/2}\sum_i (Y_{0,i} + n^{-1/2}H_{u,i}) \Rightarrow G + b_u$ and $n^{-1}\sum_i \|Y_{0,i} + n^{-1/2}H_{u,i}\|^2 \to \operatorname{tr} C_0$; the signed diagonal identity proves the limit (Araujo & Giné, 1980). If additionally $u \in L^1(P)$ and $S = -\nabla \cdot (uP)$ is a finite signed Radon measure, Appendix A.7 identifies $b_u$ with its kernel mean. Thus $b_u = 0$ iff $S = 0$ under (86), but not for an arbitrary finite bank. Changing the model's own endpoint changes the reference and is not covered by this fixed-reference limit; no universal power claim follows.

**Gaussian spectral envelope and complete risk.** For $\pi = N(0, I_p/h^2)$, let $g_\omega$ be the unnormalized two-coordinate paired derivative contraction. Set $M = \mathbb{E}\bar{\Delta}\bar{\Delta}^\top$, $F = \operatorname{tr} M$, $S = F/h^2$, $q = \mathrm{WSD}_k^2$ and $D_m = h^{-4}\{F^2 + 2\operatorname{tr}(M^2)/m\}$. Then $Q = \omega^\top M \omega$, $z = \|\mathbb{E}g_\omega\|^2$ satisfy $0 \le z \le Q$, $\mathbb{E}_\pi z = q \le S = \mathbb{E}_\pi Q$. For $S > 0$, define

$$K_S(0) = 0, \qquad K_S(q) = \min\{D_1,\, 4Sq\log(2\sqrt{2}S/q)\} \quad (q > 0). \tag{54}$$

The signal-adaptive moment and risk bounds are

$$\mathbb{E}_\pi z^2 \le \mathbb{E}_\pi[zQ] \le K_S(q), \qquad \operatorname{Var}_\pi(z) \le K_S(q) - q^2, \tag{55}$$

$$\mathbb{E}(U_n - q)^2 \le \frac{4q(S-q)}{n} + \frac{2(S-q)^2}{n(n-1)}, \tag{56}$$

$$\mathbb{E}(U_{n,m} - q)^2 \le \frac{4}{n}\left[(1 - m^{-1})qS + \frac{K_S(q)}{m}\right] + \frac{2D_m}{n(n-1)} + \frac{K_S(q) - q^2}{m}. \tag{57}$$

Here $n \ge 2, m \ge 1$; the last risk averages both rows and bank. The feature-target term is $O(Sq\{1 + \log(S/q)\}/m)$ as $q/S \downarrow 0$. At $q = 0$, the exact and paired risks are at most $2S^2/[n(n-1)]$ and $2D_m/[n(n-1)]$, with no feature-target error. If $S = 0$, the transformed residual and both estimators vanish almost surely. These moment bounds do not imply finite-bank identification.

*Verification.* The paired norm is $(\omega^\top \bar{\Delta})^2$. Gaussian quadratic moments give $\mathbb{E}Q^2 = D_1$ and, writing $Q/S = \sum_j p_j G_j^2$, convexity of $xe^{x/4}$ gives $\mathbb{E}[Qe^{Q/(4S)}] \le 2\sqrt{2}S$. Jensen under the tilted probability $(z/q)\pi$ therefore yields $\mathbb{E}[zQ] \le 4Sq\log(2\sqrt{2}S/q)$; also $z^2 \le zQ \le Q^2$. For the exact kernel, $\operatorname{tr} C = S - q$; covariance positivity and (13) give (56). For a bank, $q_m = m^{-1}\sum_j z_j$ and $s_m = m^{-1}\sum_j Q_j$ obey

$$\mathbb{E}(q_m s_m) = (1 - m^{-1})qS + m^{-1}\mathbb{E}(zQ), \qquad \mathbb{E}s_m^2 = D_m. \tag{58}$$

Apply the same variance identity with $\langle \mu_m, C_m \mu_m \rangle \le q_m s_m$, $\|C_m\|_{\mathrm{HS}}^2 \le s_m^2$, then add $\operatorname{Var}(q_m) = \operatorname{Var}(z)/m$. At the null, nonnegativity forces $z = q_m = 0$ almost surely. The Gaussian fourth moment here is spectral, not a fourth data moment. □

**Known-budget testing.** Let $S_0 > 0$ be a valid deterministic upper bound on $S$, declared before evaluation. An empirical energy estimate alone does not supply this assumption. For $0 \le q \le S_0$, set $K_0(0) = 0$, $K_0(q) = \min\{3S_0^2, 4S_0 q \log(2\sqrt{2}S_0/q)\}$ for $q > 0$, and

$$B_{n,m}(q; S_0) = \frac{4}{n}[(1 - m^{-1})qS_0 + K_0(q)/m] + \frac{2S_0^2(1 + 2/m)}{n(n-1)} + \frac{K_0(q)}{m}, \tag{59}$$

$$t_\alpha^{\mathrm{exact}} = \frac{\sqrt{2}S_0}{\sqrt{\alpha n(n-1)}}, \qquad t_\alpha^{\mathrm{RFF}} = S_0\sqrt{\frac{2(1 + 2/m)}{\alpha n(n-1)}}. \tag{60}$$

For $\alpha, \beta \in (0, 1)$, rejection above the corresponding threshold has null probability at most $\alpha$. For $q > t_\alpha^{\mathrm{RFF}}$,

$$\Pr(U_{n,m} \le t_\alpha^{\mathrm{RFF}}) \le \frac{B_{n,m}(q; S_0)}{(q - t_\alpha^{\mathrm{RFF}})^2}. \tag{61}$$

Thus $q \ge 2t_\alpha^{\mathrm{RFF}}$ and $B_{n,m} \le \beta q^2/4$ give power at least $1 - \beta$. Use (56) with $S_0$ for the exact test. A conservative explicit choice, with $r = S_0/q$, is

$$n \ge \left\lceil \max\{64r/\beta, 8r/\sqrt{\alpha}\} \right\rceil, \qquad m \ge \left\lceil 64r \log(2\sqrt{2}r)/\beta \right\rceil. \tag{62}$$

To verify these statements, use $D_m \le S_0^2(1 + 2/m)$, monotonicity of $4Sq\log(2\sqrt{2}S/q)$ in $S \ge q$, and Chebyshev. The displayed counts give $t_\alpha^{\mathrm{RFF}}/q \le 1/2$ and $B_{n,m}/q^2 \le \beta/4$ by bounding its four terms by $\beta/16$ each. Taking $r = S_0/q_{\min}$ covers every $q \ge q_{\min}$: the sufficient orders are $n = O(S_0/q)$ and $m = O((S_0/q)\log(2S_0/q))$, or $O(n \log n)$ features at a sufficiently large $S_0/n$ alternative. The signal-independent $D_1/m$ bound gives the more conservative $O(n^2)$ sufficient count there. These are joint-bank guarantees, not per-bank power statements, minimax results or necessary testing budgets.

**A known matrix envelope and importance features.** For a nonzero predeclared $\mathsf{H} \succeq M$, possibly singular, define

$$B_0 = \frac{\operatorname{tr}\mathsf{H}}{h^2}, \quad Q_{\mathsf{H}} = \omega^\top \mathsf{H}\omega, \quad \nu_{\mathsf{H}} = \frac{Q_{\mathsf{H}}}{B_0}\pi, \quad w_{\mathsf{H}} = \frac{B_0}{Q_{\mathsf{H}}}. \tag{63}$$

Use $Y_m^{\mathsf{H}} = m^{-1/2}(\sqrt{w_1}g_{\omega_1}, \ldots, \sqrt{w_m}g_{\omega_m})$ with iid $\nu_{\mathsf{H}}$ frequencies. Trace sampling is established (Minh, 2016, Theorem 9, Corollary 10, Example 4); importance-weighted feature discrepancies are studied by Huggins & Mackey (2018). Here the envelope controls the noisy residual's conditional moments. For almost every bank, with $q_m = \|\mathbb{E}_{\mathrm{data}} Y_m^{\mathsf{H}}\|^2$ and $s_m = \mathbb{E}_{\mathrm{data}}\|Y_m^{\mathsf{H}}\|^2$, and with $c_n = (n-2)(n-3)/[n(n-1)]$ for integers $n \ge 2, m \ge 1$,

$$\mathbb{E}_{\mathrm{data}}[U_{n,m}^{\mathsf{H}} \mid \omega_{1:m}] = q_m, \quad \mathbb{E}_{\nu_{\mathsf{H}}} q_m = q, \quad 0 \le q_m \le s_m \le B_0, \tag{64}$$

$$\operatorname{Var}_{\mathrm{data}}(U_{n,m}^{\mathsf{H}} \mid \omega_{1:m}) \le \frac{4q_m(B_0 - q_m)}{n} + \frac{2(B_0 - q_m)^2}{n(n-1)}, \tag{65}$$

$$V_m := \operatorname{Var}_{\nu_{\mathsf{H}}}(q_m) \le \frac{q(B_0 - q)}{m}, \tag{66}$$

$$\mathbb{E}(U_{n,m}^{\mathsf{H}} - q)^2 \le \frac{4q(B_0 - q)}{n} + \frac{2(B_0 - q)^2}{n(n-1)} + c_n\frac{q(B_0 - q)}{m}. \tag{67}$$

For $\mathsf{H} = M \ne 0$, $s_m = B_0$ for every admissible bank. The weighted kernel is $k_m^{\mathsf{H}}(x, y) = m^{-1}\sum_j w_j \cos(\omega_j^\top(x - y))$: its expectation is $k_h$, but its diagonal is $m^{-1}\sum_j w_j$, not one. Self-normalizing the weights would generally change this target.

*Verification and sampling.* The nullspace of nonzero $\mathsf{H} \succeq 0$ is Gaussian-null, so $w_{\mathsf{H}}\, d\nu_{\mathsf{H}} = d\pi$ on a full-measure set. Even for singular $\mathsf{H}$, derivative averaging is integrable since $\mathbb{E}_{\nu_{\mathsf{H}}} w_{\mathsf{H}}\|\omega\|^2 = p/h^2$. The paired norm gives $w_{\mathsf{H}} Q_M \le B_0$ and $Z = w_{\mathsf{H}} z \in [0, B_0]$ with $\mathbb{E}Z = q$. Thus $V_m \le q(B_0 - q)/m$; the standard conditional variance gives the bank bound. Averaging it and adding $V_m$ gives a coefficient $1 - 4/n + 2/[n(n-1)] = c_n \ge 0$ for every integer $n \ge 2$, proving the joint bound. Also $\bar{\Delta} \in \operatorname{range}\mathsf{H}$ almost surely and $(\omega^\top\bar{\Delta})^2 \le Q_{\mathsf{H}}\bar{\Delta}^\top\mathsf{H}^\dagger\bar{\Delta}$, so no inverse is needed in the estimator and the moment condition still includes target noise. To sample exactly, diagonalize $\mathsf{H} =$

$V\,\mathrm{diag}(\lambda_j)V^\top$, choose $J$ with probabilities $\lambda_j/\operatorname{tr}\mathsf{H}$, and set $\omega = VG/h$: independent coordinates of $G$ are standard normal except $G_J$, which has independent uniform sign and magnitude $\chi_3$. This follows from the density tilt $\sum_j(\lambda_j/\operatorname{tr}\mathsf{H})G_j^2$. □

At the exact null, $q_m = 0$ almost surely, so rejection above

$$t_\alpha^{\mathsf{H}} = \frac{\sqrt{2}B_0}{\sqrt{\alpha n(n-1)}} \tag{68}$$

has conditional-bank level at most $\alpha$. For joint power $1-\beta$, with $r = B_0/q$, sufficient counts are

$$n \geq \left\lceil \max\{64r/\beta, 4r/\sqrt{\alpha}\} \right\rceil, \qquad m \geq \lceil 16r/\beta \rceil. \tag{69}$$

Indeed $t_\alpha^{\mathsf{H}}/q \leq 1/2$ and the normalized risk is at most $4r/n + 4r^2/n^2 + r/m < \beta/4$; Chebyshev applies jointly over rows and bank. If only a scalar bound $F_0 > 0$, $F_0 \geq F$ is known, $\mathsf{H} = F_0 I_p$ is valid but $B_0 = pF_0/h^2$ incurs a dimension factor. An empirical matrix is not automatically an upper bound. Neither this design nor its conditional null calibration implies finite-bank identification or a general power advantage.

**Gaussian loops: unchanged residual, changing variance.** Fix $d \geq 3$, $h, T > 0$, $\tau = \pi/(2T)$ and nonzero $A^\top = -A$ with $Ae_1 = 0$; set $F_A = \operatorname{tr}(A^\top A)$ and $H_A = \operatorname{tr}[(A^\top A)^2]$. The VP interpolation $X_t = \cos(\tau t)X_0 + \sin(\tau t)\varepsilon$ has independent $X_t \sim N(0, I_d)$, $T_t \sim N(0, \tau^2 I_d)$. For $b, c \in C^1$ vanishing at both endpoints and $L \geq 0$,

$$v_t^{(L)}(x) = b'(t)e_1 + Lc'(t)Ax, \tag{70}$$

$$\Phi_{t,0}^{(L)}(x) = b(t)e_1 + e^{Lc(t)A}x, \quad q_t^{(L)} = N(b(t)e_1, I_d), \tag{71}$$

$$R_t^{(L)} = b'(t)x_1\phi_d(x)\,dx. \tag{72}$$

The reference is the interpolation of each model's own Gaussian endpoint, in either integration direction; the marginal path and residual do not vary with $L$. With $C_{\mathrm{rot}} = \mathrm{Cov}(B_X AX)$, $s = h^2$,

$$\|C_{\mathrm{rot}}\|_{\mathrm{HS}}^2 = \left(\frac{s}{s+4}\right)^{d/2} \frac{3F_A^2 + (s^2+4s+6)H_A}{s^2(s+4)^2} > 0, \tag{73}$$

and for fixed $n \geq 2$, $c'(t_0) \neq 0$,

$$\lim_{L\to\infty} \frac{\mathrm{Var}(U_n^{(L)})}{L^4} = \frac{2c'(t_0)^4\|C_{\mathrm{rot}}\|_{\mathrm{HS}}^2}{n(n-1)}. \tag{74}$$

If $b'(t_0) \neq 0$, the common WSD is positive; for example $b = c = \sin^2(\pi t/T)$ and $t_0 = T/4$. Thus WSD and $n$ alone cannot bound this estimator's MSE; this family has increasing energy, not an all-estimator or testing lower bound. Its exact first-order term stays fixed by (78) below. Marginal-preserving rotations are established (Berman et al., 2026; Horvat & Pfister, 2024).

*Verification.* Differentiating the flow and using $Ae_1 = 0$ verifies the ODE; $e^{LcA}$ is orthogonal and the endpoint map is the identity. $\nabla\cdot(Ax\phi_d) = (\operatorname{tr}A - x^\top Ax)\phi_d = 0$ gives the residual. Hence $W = B_X AX$ has mean zero and $\operatorname{tr}C_{\mathrm{rot}} = F_A/h^2 > 0$. At $t_0$, $Y^{(L)} = Y_0 + Lc'(t_0)W$, so its covariance divided by $L^2$ converges in trace norm to $c'(t_0)^2 C_{\mathrm{rot}}$; (13) gives the limit. For the explicit coefficient, let $R = (X - X')/\sqrt{2}$, $S' = (X + X')/\sqrt{2}$ for independent standard Gaussians. The derivative kernel contraction is

$$\langle B_X AX, B_{X'} AX'\rangle = e^{-\|R\|^2/s}\left\{\frac{\|AS'\|^2 - \|AR\|^2}{2s} - \frac{(R^\top AS')^2}{s^2}\right\}. \tag{75}$$

Squaring tilts $R$ to covariance $cI_d$, $c = s/(s+4)$, with factor $c^{d/2}$. Put $R = \sqrt{c}V$, $Q(x) = x^\top A^\top Ax$ and $Z = (V^\top AS')^2$. Gaussian moments give $\mathbb{E}[Q(S') - cQ(V)]^2 = (1-c)^2F_A^2 + 2(1+c^2)H_A$, $\mathbb{E}[Q(S')Z] = \mathbb{E}[Q(V)Z] = F_A^2 + 2H_A$, and $\mathbb{E}Z^2 = 3(F_A^2 + 2H_A)$. Expanding $c^{d/2}\mathbb{E}\{[Q(S') - cQ(V)]/(2s) - cZ/s^2\}^2$ proves (73). □

**Finite-bank leakage with bounded coefficients.** Take $L = 1$, $b = a\rho$, $c = \ell\rho$, where $\rho \in C^1([0,T])$, $t_0 \in (0,T)$, $\rho(0) = \rho(T) = 0$, $\rho'(t_0) = 1$, and $a, \ell$ are real. At $t_0$,

$$\Delta^{(\ell)} = ae_1 + \ell AX - T_{t_0}, \qquad F = a^2 + \ell^2 F_A + d\tau^2. \tag{76}$$

Thus $a^2 + \ell^2 F_A \le F_0 - d\tau^2$ gives a common energy class when $F_0 > d\tau^2$. Along the reference path, the residual energy is at most $\|\rho'\|_\infty^2(a^2 + \ell^2 F_A) + d\tau^2$ and the spatial Lipschitz constant is at most $|\ell|\|\rho'\|_\infty \|A\|_{\mathrm{op}}$. For the exact RBF kernel,

$$\mu(x) = \frac{ah^d x_1 e^{-\|x\|^2/[2(h^2+1)]}}{(h^2+1)^{d/2+1}}, \quad q = \frac{a^2 h^d}{(h^2+2)^{d/2+1}}, \tag{77}$$

$$\langle \mu, B_x Ax\rangle = 0 \ (\forall x), \qquad \langle \mu, C_\ell \mu\rangle = \langle \mu, C_0 \mu\rangle. \tag{78}$$

For iid $\omega_j \sim N(0, I_d/h^2)$ and the normalized paired map $B_{m,x}$, write $Y_m^{(\ell)} = B_{m,X}\Delta^{(\ell)}$, $W_m = B_{m,X}AX$, $\mu_m = \mathbb{E}Y_m^{(\ell)}$, $C_{0,m} = \mathrm{Cov}(Y_m^{(0)})$, $C_{R,m} = \mathrm{Cov}(W_m)$ and $\Gamma_m = D_{0R,m} + D_{0R,m}^*$ with $D_{0R,m} = \mathbb{E}[(Y_m^{(0)} - \mu_m) \otimes W_m]$. Then $\mathbb{E}W_m = 0$ and

$$q_m = \|\mu_m\|^2 = \frac{a^2}{m}\sum_j \omega_{j,1}^2 e^{-\|\omega_j\|^2} \tag{79}$$

is independent of $\ell$. For every bank, let $\mathcal{V}_{n,m}(\ell) = \mathrm{Var}_{\mathrm{data}}(U_{n,m}^{(\ell)} \mid \omega_{1:m})$. Its complete nonnegative increment is

$$\begin{aligned}\mathcal{V}_{n,m}(\ell) - \mathcal{V}_{n,m}(0) = \frac{4\ell^2}{n}\langle \mu_m, C_{R,m}\mu_m\rangle + \frac{2\ell^4}{n(n-1)}\|C_{R,m}\|_{\mathrm{HS}}^2 \\ + \frac{2\ell^2}{n(n-1)}\{2\langle C_{0,m}, C_{R,m}\rangle_{\mathrm{HS}} + \|\Gamma_m\|_{\mathrm{HS}}^2\}.\end{aligned} \tag{80}$$

With

$$\kappa_A = \frac{F_A}{2h^4}\left\{\left(\frac{h^2}{h^2+2}\right)^{d/2+2} + \left(\frac{h^2}{h^2+6}\right)^{d/2+2}\right\} > 0, \tag{81}$$

the projection and joint-risk increments satisfy

$$\mathbb{E}_\omega\langle \mu_m, C_{R,m}\mu_m\rangle = a^2\kappa_A/m, \tag{82}$$

$$\mathbb{E}_\omega[\mathcal{V}_{n,m}(\ell) - \mathcal{V}_{n,m}(0)] \ge \frac{4a^2\ell^2\kappa_A}{nm} + \frac{2\ell^4}{n(n-1)}\mathbb{E}_\omega\|C_{R,m}\|_{\mathrm{HS}}^2, \tag{83}$$

where $\mathbb{E}_\omega\|C_{R,m}\|_{\mathrm{HS}}^2 \ge \|C_{\mathrm{rot}}\|_{\mathrm{HS}}^2 > 0$. Since $q_m$ is unchanged, this is also the difference of the two joint MSEs relative to $q$. For fixed nonzero $a, \ell$, an $O(n^{-2})$ increment requires $m = \Omega(n)$; this is not a requirement for total MSE or consistency, since its baseline can be $O(n^{-1})$. For $a_n = O(n^{-1/2})$, the displayed first-order contribution is $O(n^{-2}m^{-1})$.

*Verification.* Gaussian convolution gives (77); its gradient is a scalar multiple of $e_1 - x_1 x/(h^2+1)$, orthogonal to $Ax$. For a bank the mean has pairs $am^{-1/2}(0, \omega_{j,1}e^{-r_j/2})$, $r_j = \|\omega_j\|^2$, because $\omega_j^\top X$ and $\omega_j^\top AX$ are independent centered Gaussians. In sine/cosine blocks, $C_{0,m}$ and $C_{R,m}$ are block diagonal, $\Gamma_m$ is off-diagonal and $\mu_m$ is in the cosine block. Target noise contributes $\tau^2\mathbb{E}[B_{m,X}B_{m,X}^*]$ with the same diagonal blocks and zero cross covariance with functions of $X$. Hence the linear/cubic terms vanish when (13) is applied to $C_{0,m} + \ell\Gamma_m + \ell^2 C_{R,m}$, giving the full polynomial. Positivity gives $\langle C_{0,m}, C_{R,m}\rangle_{\mathrm{HS}} \ge 0$. For the first projection, set $f_m(x) = \frac{a}{m}\sum_j \omega_{j,1}e^{-r_j/2}(\omega_j^\top Ax)\cos(\omega_j^\top x)$. Each summand has zero frequency mean by exact cancellation, so $\mathbb{E}_\omega\mathbb{E}_X f_m^2$ retains only diagonal-frequency terms. The Gaussian identities

$$\mathbb{E}_X[(\omega^\top AX)^2\cos^2(\omega^\top X)] = \tfrac{1}{2}\|A^\top\omega\|^2(1 + e^{-2\|\omega\|^2}),$$

$$\mathbb{E}_\omega[\omega_1^2\|A^\top\omega\|^2 e^{-u\|\omega\|^2}] = \frac{F_A}{h^4}\left(\frac{h^2}{h^2+2u}\right)^{d/2+2}$$

at $u = 1, 3$ give (82). Finally, for $H_m(X, X') = \langle B_{m,X}AX, B_{m,X'}AX'\rangle$, frequency averaging recovers the exact derivative contraction. Jensen yields $\mathbb{E}_\omega\mathbb{E}_{X,X'}H_m^2 \ge \|C_{\mathrm{rot}}\|_{\mathrm{HS}}^2$; Gaussian data and frequency moments justify all interchanges. □

### A.7 STANDARD RESIDUAL IDENTIFICATION AND TRANSPORT FACTS

**Weak representation and kernel identification.** Fix $(z,t)$ in the declared path of Appendix B. Assume a differentiable schedule with finite derivatives, $\mathbb{E}[\|X_0\| \mid z] + \mathbb{E}\|\varepsilon\| < \infty$, $\delta v_t \in L^1(P_t^z)$, and a finite signed Radon measure $R_t^z = -\nabla \cdot (\delta v_t P_t^z)$. Integration by parts and total-variation duality give

$$\int \varphi \, \mathrm{d}R_t^z = \mathbb{E}[\nabla\varphi(X_t)^\top \Delta_t \mid z], \qquad \mathrm{SD} = \sup_{\substack{\varphi \in C_c^\infty \\ \|\varphi\|_\infty \le 1}} \left|\mathbb{E}[\nabla\varphi(X_t)^\top \Delta_t \mid z]\right|. \tag{84}$$

The first identity holds for bounded $C^1$ tests with bounded gradient: cutoffs have error $O(R^{-1})\|\varphi\|_\infty \|\delta v_t\|_{L^1}$, and conditional expectation replaces $v_t^\star$ by $T_t$. Mollification preserves the unit sup-norm bound in the TV duality. These assumptions, when valid there, also cover endpoints.

Let $k$ be real, symmetric, positive definite and jointly $C^2$, with bounded diagonal, first-derivative reproduction and $M_k^2 = \sup_x \mathrm{tr}[\nabla_x \nabla_y k(x,y)]_{y=x} < \infty$. Use Borel/strongly measurable versions of the observations and features. Derivative reproduction (Zhou, 2008) and $\mathbb{E}[\|\Delta_t\| \mid z] \le \|\delta v_t\|_{L^1} + 2\mathbb{E}[\|T_t\| \mid z]$ make the derivative feature Bochner integrable. The RKHS unit-ball inclusion then gives the standard restriction

$$\mathrm{WSD}_k \le \mathrm{SD} \quad \text{if } \sup_x k(x,x) \le 1. \tag{85}$$

If additionally $\mathcal{H}_k \subset C_0(\mathbb{R}^d)$ is sup-norm dense in $C_0$, injectivity of the finite-signed-measure embedding (Sriperumbudur et al., 2011, Proposition 2) gives

$$\mathrm{WSD}_k = 0 \quad \Longleftrightarrow \quad R_t^z = 0. \tag{86}$$

Indeed (84) identifies the derivative mean with $\int k(x,\cdot) R_t^z(\,\mathrm{d}x)$; bounded sections and finite $|R_t^z|$ verify its Bochner integrability. This uses $C_0$-universality for signed measures, not characteristicness for probabilities alone. It concerns the exact kernel: neither finite banks nor noninjective projections inherit identification, and a zero residual need not imply a zero velocity error.

The other standard conditional-expectation identity, for square-integrable $T_t$ and $\delta v_t(X_t, z)$, is

$$\mathbb{E}[\|\Delta_t\|^2 \mid z] = \|v_\theta - v_t^\star\|^2_{L^2(P_t^z)} + \mathbb{E}[\mathrm{tr}\,\mathrm{Cov}(T_t \mid X_t, z) \mid z]. \tag{87}$$

The cross term vanishes upon conditioning on $X_t$. In an interior scalar Gaussian channel the noise term is a schedule factor times conditional MMSE; its entropy-rate expression additionally requires the conditional I–MMSE hypotheses (Guo et al., 2005). No identification of arbitrary KL, curvature or trajectory diagnostics follows.

**Conditional transport comparison.** Use $\|f\|_{\mathrm{BL}} = \max\{\|f\|_\infty, \mathrm{Lip}(f)\}$ and $d_{\mathrm{BL}}(q,p) = \|q - p\|_{\mathrm{BL}^*}$. On $[0,T]$, assume a Carathéodory velocity with global spatial Lipschitz envelope $L_t$ and $\|v_{\theta,t}(0)\| \le a_t$, where $L, a \in L^1_+(0,T)$. Let $R_t$ be weak-* measurable with $\int_0^T \|R_t\|_{\mathrm{TV}} \,\mathrm{d}t < \infty$, and $q, p^\star$ weak-* continuous finite signed-measure solutions, uniformly TV-bounded, of $\partial_t q + \nabla \cdot (v_\theta q) = 0$ and $\partial_t p^\star + \nabla \cdot (v_\theta p^\star) = -R$, with their stated initial traces. The TV bound is automatic for probability curves. Writing $E_{s,t} = \exp(\int_s^t L_r \,\mathrm{d}r)$, the standard Duhamel–Gronwall argument for BL transport (Ha & Liu, 2009) yields

$$d_{\mathrm{BL}}(q_t, p_t^\star) \le E_{0,t} d_{\mathrm{BL}}(q_0, p_0^\star) + \int_0^t E_{s,t} \|R_s\|_{\mathrm{BL}^*} \,\mathrm{d}s. \tag{88}$$

For completeness, the required weak-solution adaptation is as follows. The global flow $\Phi_{t,s}$ is proper and $E_{s,t}$-Lipschitz by forward and inverse growth bounds. Testing with $f \circ \Phi_{t,s}$ gives $\eta_t = (\Phi_{t,0})_\# \eta_0 + \int_0^t (\Phi_{t,s})_\# R_s \,\mathrm{d}s$, where $\eta = q - p^\star$. For time-measurable velocities, smooth approximation on the common compact support of a terminal test has adjoint error bounded by $C \sup_s \|\eta_s\|_{\mathrm{TV}} \int_0^t \sup_{x \in K} \|v_s - v_s^\epsilon\| \,\mathrm{d}s \to 0$. Properness, strong $C_0$-measurability of the composed tests and the TV-integrable source justify the signed-measure integral and passage to the limit. Finally $\|T_\# \nu\|_{\mathrm{BL}^*} \le \max\{1, \mathrm{Lip}(T)\} \|\nu\|_{\mathrm{BL}^*}$ gives (88). The constant may depend on dimension through $L$; replacing the source norm by RBF-WSD requires a separate bridge.

Geometric-Cauchy numerical tails require a contraction bound valid for all future refinements. A fixed-panel rank certificate requires strictly separated outward-rounded intervals carrying data, feature, grid and reference errors; three refinement levels or overlapping intervals are only diagnostic.

## B REFERENCE AND ESTIMATOR DEFINITIONS

**Declared path.** Fix $Z = z$, an endpoint law $\nu_0^z$, and independent $X_0 \sim \nu_0^z$, $\varepsilon \sim N(0, I_d)$. Define

$$X_t = \alpha_t X_0 + \sigma_t \varepsilon, \qquad T_t = \dot{\alpha}_t X_0 + \dot{\sigma}_t \varepsilon, \tag{89}$$

$$v_t^\star(x, z) = \mathbb{E}[T_t \mid X_t = x, Z = z]. \tag{90}$$

For model self-consistency, $\nu_0^z$ is the evaluated model's own law obtained by integrating its well-defined flow from $T$ to 0. With $(\alpha_0, \sigma_0) = (1, 0)$ and $(\alpha_T, \sigma_T) = (0, 1)$, the terminal law is $N(0, I_d)$, and the model and reference agree at both endpoints. The question concerns intermediate marginals. A fixed-reference comparison holds $\nu_0^z$ fixed while varying the velocity and need not retain this property. Neither endpoint is assumed to equal the data law. Changing the reference law, terminal or time map, or truncated initial law changes the estimand; approximate model references require a separate error bound.

Write $P_t^z = \mathcal{L}(X_t \mid Z = z)$ and $\delta v_t = v_\theta - v_t^\star$. For integrable $\delta v_t$, the finite vector measure $J_t^z(A) = \int_A \delta v_t \, \mathrm{d}P_t^z$ defines

$$R_t^z = -\nabla \cdot J_t^z, \tag{91}$$

$$\Delta_t = v_\theta(X_t, z, t) - T_t. \tag{92}$$

The first identity is distributional. When $R_t^z$ is a finite signed Radon measure, $\mathrm{SD}(v_\theta; z, t) = \|R_t^z\|_{\mathrm{TV}}$, with no factor $1/2$. For admissible tests, integration by parts and conditional expectation give

$$\int \varphi \, \mathrm{d}R_t^z = \mathbb{E}[\nabla\varphi(X_t)^\top \Delta_t \mid Z = z]. \tag{93}$$

This first-moment identity uses the observed $T_t$ and requires neither a density score nor a pointwise neural-field divergence. Replacing $T_t$ by its conditional mean preserves this moment but changes the observation model and second moments: with $\eta_t = T_t - v_t^\star(X_t, z)$, $\Delta_t = \delta v_t(X_t, z) - \eta_t$ and $\mathbb{E}[\eta_t \mid X_t, Z = z] = 0$. Equation (87) retains the conditional target variance; it is nonnegative, not necessarily positive.

**Kernel and finite targets.** For a differentiable kernel with RKHS $\mathcal{H}_k$, set $Y_t = \nabla_1 k(X_t, \cdot)^\top \Delta_t$ and $\mu_t^z = \mathbb{E}[Y_t \mid Z = z]$. Then

$$\mathrm{WSD}_k(z, t) = \|\mu_t^z\|_{\mathcal{H}_k}. \tag{94}$$

Under the derivative-reproduction, bounded derivative-section, measurability and integrability conditions of Equation (85), $\sup_x k(x, x) \le 1$ implies $\mathrm{WSD}_k \le \mathrm{SD}$. This population comparison is not a realized finite-sample or finite-feature lower-bound certificate. For a declared finite grid and nonnegative weights, define

$$Q_J^{\mathrm{RBF}} = \sum_{j=1}^{J} w_j \, \mathrm{WSD}_{k_j}^2(z, t_j), \tag{95}$$

$$Q_{J,m}^{\mathrm{bank}} = \sum_{j=1}^{J} w_j \, \mathrm{WSD}_{k_{m,j}}^2(z, t_j). \tag{96}$$

Here $k_{m,j}$ is the kernel of the complete ordered, frozen paired sine–cosine bank: $m$ frequency draws produce $2m$ real coordinates. Importance features multiply each pair by the square root of its density-ratio weight. The weights are not divided by their sampled sum; the weighted kernel generally has a nonunit diagonal. Frequency averaging under the declared proposal preserves the exact RBF target. Conditioning on one bank instead fixes its distinct target in (96).

For conditionally iid evaluation rows independent of the bank, the signed statistic is

$$U_{n,j} = \frac{\sum_{i \ne r} \langle Y_{i,j}, Y_{r,j} \rangle}{n(n-1)}, \qquad I_{n,J} = \sum_j w_j U_{n,j}. \tag{97}$$

Its conditional expectation is $Q_J^{\mathrm{RBF}}$ for exact kernels and $Q_{J,m}^{\mathrm{bank}}$ for frozen features. Cross-time dependence within a row changes variance, not this expectation. Raw negative values are retained without clipping or square roots. In contrast, $V_n = \|n^{-1} \sum_i Y_i\|^2$ has bias $\mathrm{tr}(C)/n$ for covariance $C$ (Equation (14)); deleting the diagonal removes that term, not the remaining target-noise variance.

### B.1 Relations among estimands

The declared residual $R_t^z$ determines a population embedding and a signed estimator of the exact-kernel or frozen-bank target. A declared grid defines its own estimand; interpreting it as an unbiased quadrature rule requires a specified integral and a separate justification. Path discrepancy $d_{\mathrm{BL}}(q_t, P_t^z)$ requires the separate transport theorem's *initial mismatch and integrated bounded-Lipschitz dual norm* of the residual. A sampler response is a further empirical object, requiring independent pools, controlled numerical tails and a same-law floor. Data, frequency, grid, reference and numerical-sampler errors must be justified separately; availability of a finite statistic does not establish the later relations.

## C Explicit non-implications

**No unconditional RBF-to-BL bridge.** Let $\phi_s$ denote the $N(0, s^2)$ density and $k_h(x, y) = e^{-(x-y)^2/(2h^2)}$. For $n \geq 1$, $A_n > 0$ and $D = 1 + 2s^2/h^2$, direct Gaussian integration yields

$$\mathrm{d}\mu_n(x) = A_n \sin(nx)\phi_s(x)\ \mathrm{d}x, \tag{98}$$

$$\|\mu_n\|_{k_h}^2 = \frac{A_n^2}{2\sqrt{D}}\big(e^{-s^2n^2/D} - e^{-s^2n^2}\big), \tag{99}$$

$$\|\mu_n\|_{\mathrm{BL}^*} \geq \frac{A_n}{2n}\big(1 - e^{-2s^2n^2}\big). \tag{100}$$

The last line tests with $\sin(nx)/n$, whose supremum and Lipschitz norms are at most one. With $A_n = 1$, uniformly bounded TV does not give a universal linear reverse constant. With $A_n = n$, RBF energy vanishes while the BL lower bound tends to $1/2$, excluding a vanishing distribution-free modulus over all finite zero-mass residuals. This second sequence lacks a uniform TV or regularity envelope, so it does not exclude bounds that assume and track one. These are admissible residuals: $j_n(x) = -\int_{-\infty}^{x} A_n \sin(ny)\phi_s(y)\ \mathrm{d}y$ is integrable, and $\delta v_n = j_n/\phi_s$ gives $-\partial_x(\phi_s \delta v_n) = \mu_n$. Thus stationary RBF-WSD cannot replace the BL source norm without additional assumptions; this does not say every small-WSD model has large path error.

**Positive residual, identical endpoints.** Take the stationary Gaussian VP reference $\alpha_t = \cos(\pi t/(2T))$, $\sigma_t = \sin(\pi t/(2T))$, so $P_t = N(0, 1)$ and $v_t^\star = 0$. Let $b(t) = a\sin^2(\pi t/T)$, $v_t(x) = b'(t)$ and $a > 0$. Then $q_t = N(b(t), 1)$ and $R_t = b'(t)x\phi_1(x)\ \mathrm{d}x$. Writing $K_h^2 = h/(h^2+2)^{3/2} = \|x\phi_1(x)\ \mathrm{d}x\|_{k_h}^2$,

$$\int_0^T \mathrm{WSD}_{k_h}(t)\ \mathrm{d}t = 2aK_h > 0, \qquad \int_0^T \mathrm{WSD}_{k_h}^2(t)\ \mathrm{d}t = \frac{a^2\pi^2K_h^2}{2T} > 0. \tag{101}$$

Both endpoints nevertheless equal $N(0, 1)$; at $T/2$ the model mean is $a$. Later transport can undo an interior deviation, so positive WSD gives no endpoint lower bound and does not contradict a path upper bound.

**Finite-time and finite-feature blind spots.** For any finite grid $t_j \in [0, T]$, choose $b(t) = t(T - t)\prod_j (t - t_j)^2$, $g = b'$ and $v_t(x) = g(t)$. Multiple roots give $g(t_j) = 0$, including grid endpoints, while $b(0) = b(T) = 0$. The model's own Gaussian endpoint therefore induces the same stationary reference, and $\mathrm{WSD}_{k_h}^2(t) = g(t)^2K_h^2$. All finite-grid targets vanish, but for every continuous $w \geq 0$ positive on a set of positive measure, $K_h^2 \int g(t)^2 w(t)\ \mathrm{d}t > 0$: the nonzero polynomial $g$ has only finitely many roots. This limits continuous-time reinterpretations, not the validity of a declared grid estimand.

For a finite map $\Phi : \mathcal{X} \to \mathbb{R}^M$, choose $M + 2$ distinct points. The matrix with columns $(1, \Phi(x_i))$ has a nonzero null vector $c$, so $\mu = \sum_i c_i \delta_{x_i} \neq 0$ has zero mass and zero feature embedding; strict positive definiteness of the Gaussian RBF gives $\|\mu\|_{k_h}^2 = c^\top Kc > 0$. For a smooth, endpoint-compatible realization of a fixed paired bank, choose $0 < s < 1$ and, with $\widehat{f}(\xi) = \int e^{-i\xi x} f(x)\ \mathrm{d}x$, prescribe

$$\widehat{f}(\xi) = \prod_{\ell=1}^{m} (\xi^2 - \omega_\ell^2)^2 e^{-s^2\xi^2/2}, \qquad \widehat{\mu}(\xi) = \xi^2 \widehat{f}(\xi). \tag{102}$$

The latter vanishes at zero and all bank frequencies but has positive exact-RBF energy. Here $\mathrm{d}\mu = -f''\,\mathrm{d}x$; $f$ is a polynomial times $\phi_s$, and $u = f'/\phi_1$ is smooth with bounded derivatives and $-(\phi_1 u)' = -f''$. For $a \in C^1([0,T])$ with $a(0) = a(T) = 0$ and $a' \not\equiv 0$, the velocity $a'(t)u(x)$ has global flow $\Psi_{a(t)}$, where $\Psi$ is the autonomous flow of $u$. It is the identity at both endpoints, and its own-endpoint reference has $R_t = a'(t)\mu$. Every bank target vanishes; the exact target is positive when $a'(t) \neq 0$. The residual here is selected *after* the bank, not independently of a randomly drawn bank. All these statements concern population targets; an observed $U = 0$ establishes none of their nulls. Exact-kernel identification at each time, or almost-everywhere zero residual with matching initial law under the path theorem's assumptions, remains valid.

## D SCOPE OF THE GUARANTEES

**Identification and stability.** The signed-measure and kernel conditions identify $R_t^z$, not the velocity: weighted divergence-free discrepancies remain invisible. Rank-deficient projections identify only a pushed-forward residual. Equation (74) shows that unchanged marginals and WSD can coexist with growing estimator variance. Path control additionally requires the initial mismatch, the integrated $\mathrm{BL}^*$ source norm and an integrable global Lipschitz envelope. The associated Gronwall factor may be unavailable or vacuous; local refinements and sampler discretization bounds are separate results, not assumed here.

**Inference and moment budgets.** The exact-null U-statistic is degenerate, so ordinary nondegenerate Gaussian intervals do not apply. Replication SD is not an SE. Feature-energy budgets include conditional target noise, may have large constants, and are single-time assumptions: an integrated energy bound alone is insufficient. Equation (61) and the importance-feature test require a valid bound declared before evaluation; an independent fitted matrix or an unvalidated sample estimate is not automatically an upper bound. The scalar substitute for a matrix bound incurs a dimension factor. Conditional-bank null control, including the realized-envelope Gaussian threshold in (110), differs from joint MSE and power guarantees that average over independent banks and data.

**Which budgets are sharp.** The ordinary-Gaussian $n^{4/(h^2+4)}$ count matches the local exact-risk *order* within the oscillatory and specified periodic classes; recovering its risk ratio requires a larger count. The periodic assumptions include zero mean, bounded derivative and a nonvanishing first harmonic. The larger energy class's sufficient $O(n \log n)$ bound has no matching lower bound here. Within the local proposal comparison, the $\sqrt{\log n}$ lower bound restricts centered Gaussian variance inflation. A constant-count shifted mixture uses $c_-/n \leq q_n \leq c_+/n$ and a law changing with $n$; stronger signals and other spectral shapes require separate analysis. A zero output has $O(n^{-2})$ MSE on that local scale. The shifted-mixture statistic also provides joint unbiasedness for the exact target without the local-scale restriction and exact-risk-ratio recovery within the local class whenever $m_n \to \infty$. These properties give a more informative comparison than the local risk order alone; they concern the specified estimator.

For unknown $q \geq b/n$, Theorem 1 controls additional joint risk relative to $q^2$, preserving the data-error floor. Theorem 2 instead compares with each signal's exact-kernel risk and requires $\Theta(n/\varepsilon^2)$ frequencies, hence quadratic scalar $O(nm)$ evaluation cost at fixed precision. Its lower bound uses two fixed signals. Proposition 1 concerns fixed widths, with width-dependent constants and a harder intermediate scale below the threshold. Both uniform results optimize *common admissible iid importance proposals for this U-statistic*. They do not optimize all estimators, pilot-adaptive proposals, dependent features or quadrature, nor give simultaneous guarantees for every signal under one realized bank. The sufficient power result in Equation (35) needs a valid predeclared budget and sufficiently large $n$; MSE lower bounds are not necessary power budgets. Independent interpolation noise and the periodic assumptions are essential here; general spectra and high-dimensional learned fields are outside these results.

**What the experiments resolve.** The population scans use four scalar Gaussian profiles and three values of $h^2$; noisy simulations use two profiles. The earlier rotation and energy studies use one or three dimensions, while the affine and mixture diagnostics have distinct protocols. Finite profiles and numerical carrier scans do not verify class-wide lower bounds or continuous suprema; large $n$ in an

integral is not a large-data simulation. No experiment establishes trained-checkpoint performance, sample quality, sampler disagreement prediction, cross-model ranking or high-dimensional scaling.

The experimental appendices also report the cases where a proposal performs worse. Energy sampling is not variance-optimal: it loses to ordinary features in the rotation controls and to broader Gaussian features at the two largest oscillator stress frequencies. The isolated first-order rotation term contributes 0.85–7.11% of total MSE on the $a = 1$, $\ell \in \{1, 2\}$ grid; bounding that component by $O(n^{-2})$ at a fixed signal says nothing comparable about total risk (locally it is $O(n^{-2}/m)$). Earlier local-sequence relative-MSE intervals stay above one even for the exact kernel; null tests are conservative, and scalar inflation of $\mathsf{H}$ raises thresholds without changing the statistic. Twelve earlier spectral conditions have analytic relative MCSE above 10%; bootstrap can miss a reference when rare losses are absent.

The unknown-scale experiment separately retains nine predeclared warnings: seven from pilot variance projections and two from spectral necessary bounds, not nine confirmed precision failures. Empirical MCSE/reference is at most 10% in 244/246 conditions, without certifying true relative SE. Fifteen pointwise percentile intervals miss their additional-risk references; this descriptive count is neither coverage calibration nor a simultaneous test. The two spectral-pressure cases use 4096 outer groups against necessary counts about 21,042 and 14,032; close observed means or resampling cannot remove missing-tail uncertainty. These necessary bounds use checked numerical moments, not interval arithmetic. Four of 704 integer budgets are sensitive to recorded numerical perturbations; integral and grid refinement certify neither every root nor the continuous supremum. At $h^2 = 0.5$, widths 2 and 2.5 remain unseparated through $n = 2^{40}$, with maxima at $g_1, r = 1$, limiting finite-window support for the asymptotic threshold.

## E Related work and attribution

**Generative paths and consistency.** DDPM and score-SDE models connect denoising to reverse dynamics (Ho et al., 2020; Song et al., 2021); Flow Matching and stochastic interpolants learn velocities for prescribed paths (Lipman et al., 2023; Albergo & Vanden-Eijnden, 2023). Schedule Deviation studies model-own-endpoint consistency: Pfrommer et al. (2025, Theorem 1) gives a global time-integrated TV upper bound and a local converse under additional regularity. WSD evaluates the weak residual of this reference using observable $T_t$. It is a residual statistic, not a sampler, score estimator or sample-quality metric. Noise schedules and discretization can themselves cause train–inference mismatch (Lin et al., 2024; Chen, 2023); optimizing a path-space sampling surrogate need not optimize endpoint divergence (Sabour et al., 2024).

**Weak residuals and statistical inference.** Kernel methods express weak continuity objectives through RKHS residuals and paired Fourier tests (Nakano, 2026; Berman et al., 2026); Berman et al. estimate temporal moment derivatives from time slices, whereas WSD uses $T_t$ at a fixed time. Velocity nonuniqueness permits marginal-preserving Gaussian rotations (Lipman et al., 2023; Horvat & Pfister, 2024); the rotation construction here quantifies their effect on a specified estimator's variance. WSD uses the RKHS-norm/U-statistic structure familiar from MMD and Stein discrepancies (Gretton et al., 2012; Liu et al., 2016), with a different operator: a kernel derivative contracted with $\Delta_t$, eliminating the target density score. Identification relies on finite-signed-measure universality (Sriperumbudur et al., 2011) and derivative reproduction (Zhou, 2008). Hilbert-space and U-statistic theory give the limit laws, diagonal correction and null degeneracy (Araujo & Giné, 1980; Serfling, 1980). The interior Gaussian-channel specialization of target variance uses I–MMSE only under its additional conditions (Guo et al., 2005).

Conditional-moment U-tests distinguish null and alternative limits (Muandet et al., 2020); their $\mathbb{E}[\Delta \mid X] = 0$ condition is stronger than WSD's residual null. Related covariance/power analyses and local quadratic-form limits are available (Jitkrittum et al., 2017; Li et al., 2026). Total covariance gives the observable covariance identity, and the fixed-reference local limit specializes those limit arguments to the WSD observation model. Optimal detection would require further analysis. The BL path bound uses flow-stability arguments (Ha & Liu, 2009); its source norm cannot be replaced by stationary RBF-WSD without the additional bridge excluded above.

**Frequency approximation and proposal design.** Random features (Rahimi & Recht, 2007), paired real maps (Sutherland & Schneider, 2015) and derivative approximation (Szabó & Sriperumbudur, 2019) supply the computational tools. Feature-moment control and budget conversion appear in Huggins & Mackey (2018, Section 3.2) and Choi & Kim (2024, Equation (10), Lemma 17, Section 4.2); the latter also derives exact-statistic/feature decompositions and $1/(nm)$ interactions (Appendix B.5–B.6). The WSD calculations control the noisy derivative contraction, realize bounded endpoint-compatible oscillations, and obtain matching joint-risk budgets, a cosine-family constant, and distinct order/ratio criteria. The rotation example further isolates finite features' loss of an isotropic witness's pointwise cancellation. At an exact spectral null, zero feature-target error follows directly from the nonnegative spectral representation.

Minh (2016, Theorem 9, Corollary 10, Example 4) gives trace sampling and the Gaussian derivative proposal proportional to $\|\omega\|^2$ times its spectral law. Its Section 3.3, Equation (39), also uses broader Gaussians; Huggins & Mackey (2018, Theorem 3.8) analyzes overdispersion. The matrix-envelope transform applies those designs to noisy WSD with conditional-bank moments, joint risk and a known-budget test. Defensive mixtures control the weights, while separate shifted components cover two important tails (Hesterberg, 1995, Sections 6.1, 6.3). With these proposals and unnormalized importance weights, the WSD risk analysis gives fixed-width exponents, the centered-Gaussian local optimum and the local mixture's constant count while controlling the changing data–feature interaction.

Proposal selection can optimize worst-case relative precision (Roy & Evangelou, 2024, Sections 3.1–3.2). For the endpoint-compatible periodic class, uniform spectral coverage bounds control additional *joint* risk. Disjoint spectral peaks and Cauchy–Schwarz give the coverage lower bound. For exact-risk-relative precision, two fixed spectra give the linear count lower bound. A uniform periodic witness derivative estimate yields the matching upper bound and intermediate-scale width transition. Optimization is over common iid importance proposals for the specified estimator.

# F Reproducibility

The manuscript test suite checks implementations, raw signed observations and a machine-readable provenance manifest. The unknown-scale study uses separate validation, pilot, simulation and analysis random streams (master seed 20260911). Each of 55,808 independent outer groups contains four data sets; 2000 bootstrap resamples retain whole groups and all paired methods/prefixes. Audits replay every bank and data identity and check all raw-group summaries. U-statistic replay covers each anchor's first and last outer groups at data repeat zero, for every proposal and prefix. Cross terms and deduplicated exact references are retained.

The experimental appendices specify profiles, proposals, grids, replications, precision warnings and distinct protocols; observations are not pooled. The older affine/mixture diagnostic archive checks reported values against fixed records; a bitwise rerun from one Git revision is not guaranteed. `README.md` and `paper/README.md` give configurations, seeds, requirements and verification commands; `src/` and `tests/` contain implementations and tests. From the repository root,

```
make -C paper
```

builds `paper/build/main.pdf` using PDFLaTeX, BibTeX and `latexmk`. The LaTeX build does not rerun the experiments.

`supplementary_data.zip` contains complete analyzed CSV/JSON tables and figures, with member hashes in `RESULTS_MANIFEST.json` and paths in its README. Under `paper/generated/`, the `supplementary_` CSV files retain all 246 conditions, 704 budgets, scans, comparisons, timing and pressure records. Complete-table TeX is included: `budget_appendix_tables.tex` contains all 96-bank, matrix/slack and tail tables, and `budget_timing_table.tex` all 28 earlier timing anchors. Profile and control plots are in `paper/figures/`; both asset manifests are in `paper/`. The ZIP is not a full code or raw-sample archive. The repository's frozen runs `results/supplementary_stage2/stage2_20260911_v1/` and `results/budget_stage2/stage2_20260910_v1/` retain raw signed arrays, stream keys and provenance.

# G Experimental Protocols and Additional Evidence

I analyze the unknown-scale study, earlier fixed-proposal controls and legacy diagnostics separately, using each study's specified estimators, references and replication units. Their observations are not pooled. The accompanying `supplementary_data.zip` contains the original project-relative paths, all result tables and profile plots, machine-readable results, and source manifests, including every condition and adverse result. Raw signed estimates are unclipped; the legacy square-root display below serves only as a scaling diagnostic.

## G.1 Unknown-scale models and proposals

Set $T = \pi/2$, $t_0 = T/2$, and $\rho(t) = -T\sin(2\pi t/T)/(2\pi)$. For independent $X_0, \epsilon \sim N(0,1)$, the reference at evaluation is $X = (X_0+\epsilon)/\sqrt{2}$, $T_{t_0} = (-X_0+\epsilon)/\sqrt{2}$. Thus $X, T_{t_0} \sim N(0,1)$ independently and $\Delta = g(rX)/r - T_{t_0}$ for the model $v_t(x) = \rho'(t)g(rx)/r$, $r \geq 1$. The model loop returns to its own Gaussian endpoint; its intermediate marginals need not equal the stationary reference. The four profiles specified before sampling are

$$\begin{aligned} &g_1(y) = \cos y, && g_2(y) = \cos(y+\pi/2),\\ &g_3(y) = \{\cos(y+\pi/4) - 0.3\cos(2y-0.7) + 0.2\cos(3y+1.2)\}/2.2,\\ &g_4(y) = \{\cos y + 0.2\sin(2y) - 0.1\cos(4y)\}/1.8. \end{aligned} \tag{103}$$

They are smooth, zero-mean, $2\pi$-periodic, with $\|g'\|_\infty \leq 1$, first-harmonic amplitude at least $0.4$, and $\|g\|_\infty \leq 1$. Hence $\mathbb{E}\Delta^2 \leq 2$ for this finite panel, not necessarily for the entire theoretical class.

Write $s = h^2$, let $\varphi_v$ denote the $N(0,v)$ density, and put $\pi_s = \varphi_{1/s}$. The four proposals are

$$\begin{aligned} \nu_{\text{ordinary}} &= \pi_s, \qquad \nu_{\text{broad}} = \tfrac{1}{2}\pi_s + \tfrac{1}{2}\varphi_{\log(en)/s},\\ \nu_{\text{critical}} &= \tfrac{1}{2}\pi_s + \tfrac{1}{2}\varphi_{\kappa_*/s}, \qquad \kappa_* = \max\{1, (s+2)/(2s)\},\\ \nu_{\text{local}}(\omega) &= \tfrac{1}{2}\pi_s(\omega) + \tfrac{1}{4}\varphi_{1/(s+2)}(\omega - a_n) + \tfrac{1}{4}\varphi_{1/(s+2)}(\omega + a_n), \quad a_n = \frac{2\sqrt{\log n}}{\sqrt{s(s+2)}}. \end{aligned} \tag{104}$$

Broad-defensive and Local-shifted name these complete mixtures. Fixed-width comparisons replace $\kappa_*$ by $\{1, 1.5, 2\}$ at $s = 1$, $\{1, 2, 2.5, 3\}$ at $s = 0.5$, and $\{1, 2\}$ at $s = 2$; identical actual laws are deduplicated. The same density at fixed $(n, h, \kappa)$ applies to every carrier and profile, without fitted signal or evaluation data. Weights are the full ratio $w = \pi_s/\nu$, without clipping or self-normalization; $m$ frequencies give $2m$ real paired coordinates. These existing sampling principles are attributed in Section E. In particular, Broad-defensive is not the pure broader Gaussian used in the earlier study below.

## G.2 Population panels and noisy-statistic replication

For each model, the two precision criteria use the same exact identity

$$\mathcal{R}^\nu_{n,m} = R_\infty + (V_\nu + J^\nu_n)/m, \qquad R_\infty = 4A/n + 2D/[n(n-1)], \tag{105}$$

and finite-panel coefficients

$$B^{\text{panel}}_q = \max_{q \geq 0.1/n} (V_\nu + J^\nu_n)/q^2, \qquad B^{\text{panel}}_R = \max_{q \geq 0.1/n} (V_\nu + J^\nu_n)/R_\infty. \tag{106}$$

All bandwidths use $n = 2^k$, $k \in \{8, 12, 16, 20, 24, 28, 32, 40\}$: four profiles at $s = 1$ give 32 domains, and $g_1, g_3$ at each of $s = 0.5, 2$ give 16 more each. The 64 domains include eligible $r = 1, 1.5, 2$, evenly spaced carriers, roots of $q = 0.1c/n$ for $c = 1, 4, 16$, and roots balancing the two exact-data risk terms. A finite-harmonic tail envelope bounds the carrier search; the root checks do not assume monotonicity or certify all continuous roots. Spacings $0.1$ and $0.05$ gave maxima differing by at most $0.5664\%$; the predeclared $2\%$ trigger for spacing $0.025$ was not reached. Of 8,412 archived points, 7,882 satisfy the signal condition, including 32 original boundary-root diagnostics; 7,850 qualified points enter the maxima. The records include those diagnostics and their distinct interior guards, 192 signal-root and 64 risk-balance candidate checks, and 177 independent moment checks using real-space and spectral calculations. Integration refinement is a sensitivity check, not interval arithmetic. Large $n$ enters population formulas only, not simulated data sets.

For $\varepsilon = 1/2, 1/4$, the reported budgets are

$$m_q = \max\{1, \lceil B_q^{\mathrm{panel}}/\varepsilon^2 \rceil\}, \qquad m_R = \max\{1, \lceil B_R^{\mathrm{panel}}/\varepsilon^2 \rceil\}. \tag{107}$$

The full and boundary-only panels comprise 176 reported budget rows and 704 integers; four integers change under the recorded numerical perturbations. The data package gives all sensitivity ranges. These calculations concern the finite panel; they provide neither continuous-class budget certification nor measured runtimes at $n = 2^{40}$.

The main noisy grid uses $s = 1$, $n = 128, 512, 2048$, $g_1, g_3$, fixed/intermediate/local regions, all four proposals, and $m = 16, \lceil 8\sqrt{\log(en)} \rceil, n$. The fixed anchor is $r = 1$. The intermediate anchor is the largest eligible risk-balance root, or the prespecified middle-signal root if none is eligible; the records distinguish those roles. Local anchors are qualified roots near $q = 0.1/n$. At $n = 512, g_1$, Broad-defensive and Fixed-critical also use $4\lceil 8\sqrt{\log(en)} \rceil = 88$ in each region. At $s = 0.5$, $n = 512, 2048$, middle/local $g_1$ anchors compare Ordinary with defensive widths $2.5, 3$ at $m = n/4, n$. I selected the model anchors from reference formulas before simulation, without comparing proposal performance.

This gives 22 anchors and 246 proposal–budget conditions. I used an independent pilot to fix $B \in \{512, 1024, 2048, 4096\}$ outer groups per anchor, each containing four independent noisy data sets. Rows are paired across methods and prefixes; each actual law supplies one largest bank, and the exact kernel is evaluated once per data set. Different anchors have independent banks and rows. Totals are 55,808 outer groups, 223,232 data sets, 219,648 actual banks, and 2,666,496 signed feature estimates, excluding the pilot. For the primary additional-loss observation,

$$G_b = \tfrac{1}{4} \sum_{j=1}^{4} (U_{b,j}^{\nu} - U_{b,j}^{\mathrm{exact}})^2, \quad \widehat{\Delta R} = B^{-1} \sum_b G_b, \quad \widehat{\mathrm{MCSE}} = \sqrt{\widehat{\mathrm{Var}}(G_b)/B}, \tag{108}$$

the sample variance uses $B - 1$ and the reference is $(V_\nu + J_n^\nu)/m$. Stored total, exact, conditional and spectral losses include their finite-repetition cross terms. Observed total-minus-exact loss is not substituted for this paired additional loss.

Pointwise 95% percentile intervals use 2000 resamples of complete outer groups, with shared indices for all methods, prefixes and metrics within an anchor. Four data sets stay together in every resample. Empirical ratios are ratios of mean losses with shared bootstrap denominators; population $q^2$ and $R_\infty$ denominators are labeled separately. Each loss has its own MCSE. Neither per-row ratios nor averaged chunk MCSE are used. Master seed 20260911 has disjoint pilot, formal and analysis phases 211, 307 and 503. I neither added nor replaced observations, nor stopped sampling based on outcomes.

### G.3 Unknown-scale findings and precision limits

Figures 1 and 2 are the representative mechanism displays. At $s = 1, n = 2^{40}$, Broad-defensive has $B_q^{\mathrm{panel}} = 189.700$ but $B_R^{\mathrm{panel}}/n = 0.145198$; at $\varepsilon = 0.5$ the budgets are 759 and approximately $6.386 \times 10^{11}$ (Table 3). At its largest signal-relative risk, the interaction contributes 93.64% of $V_\nu + J_n^\nu$. For local $g_1, n = m = 512, s = 1$, Ordinary's observed additional risk$/q^2$ is $1.35\,[1.27, 1.43]$, whereas total MSE$/R_\infty$ is $1.01\,[0.965, 1.05]$: the latter is a population, not empirical, exact-risk denominator. Six prespecified $m = 22 \to 88$ comparisons give paired additional-risk ratios 0.243613–0.269806; each individual interval contains $1/4$. This supports an inverse-budget mechanism, not simultaneous coverage or the continuous-class lower bounds.

Across all 246 conditions, additional-loss mean/reference ranges from 0.875555 to 1.141417. MCSE/reference is at most 10% in 244 conditions; MCSE/observed mean exceeds 10% in three. All deviations are below three empirical MCSE, but 15 pointwise intervals exclude their numerical references. These counts describe this run; they provide no guarantee of true precision or calibrated simultaneous coverage. All nine pressure flags fixed before formal sampling apply: seven pilot-variance requests exceeded 4096 groups, and two numerical spectral necessary bounds did so. The former are planning warnings, not proven precision failures. For local Ordinary $g_1, n = 2048, s = 1, m = 16, 24$, the numerical necessary counts are about 21,042 and 14,032. At $B = 4096$, their necessary relative MCSE bounds are 22.67% and 18.51%, while empirical MCSE/reference is 18.905% and 18.416%. The respective reference-normalized intervals

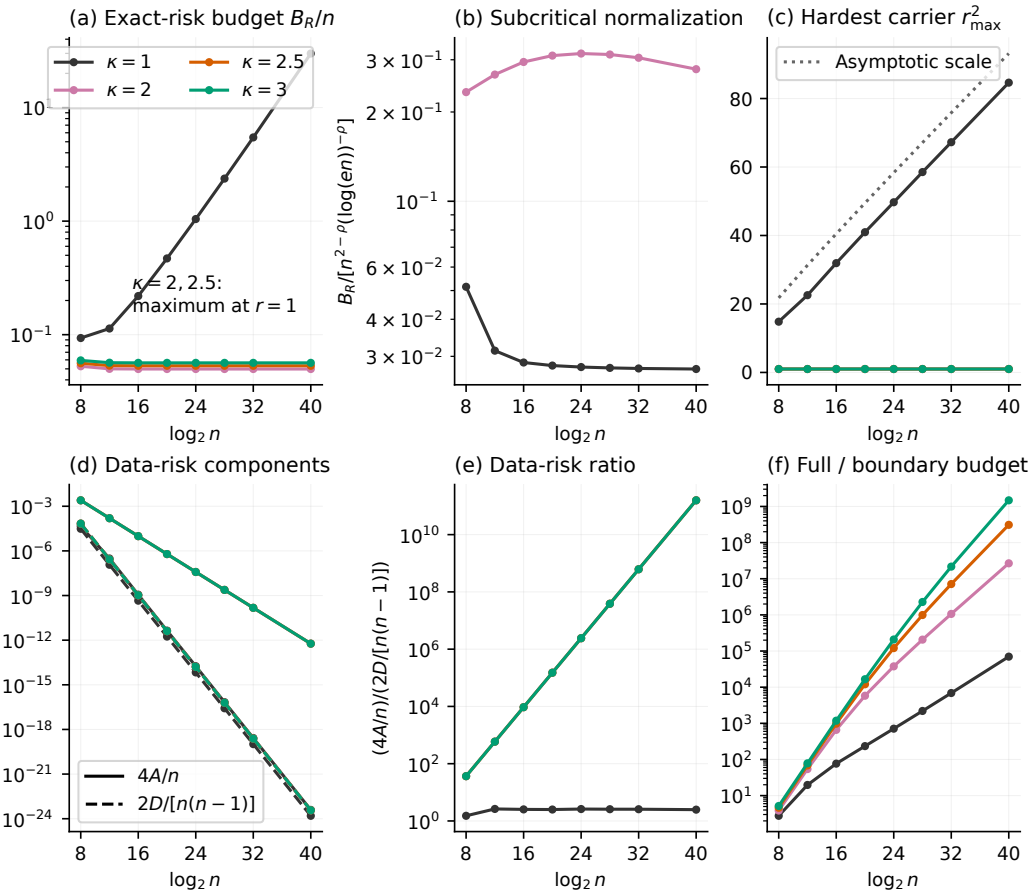



Figure 3: The $s = 0.5$ width panel across $g_1, g_3$ shows the unresolved separation of $\kappa = 2$ and 2.5: both remain dominated by $r = 1$. Normalizations use fixed theoretical exponents, not fitted rates. All other bandwidth and four-profile plots are in `supplementary_data.zip` under `paper/figures/`.

$[0.6383, 1.3640]$ and $[0.7023, 1.4083]$ do not repair this limitation. The largest 1% of individual losses contributes 76.25% and 73.45% of summed paired additional loss; these correlated within-bank tail summaries are descriptive.

Ordinary's $s = 1, n = 2^{40}$ maximum occurs at $g_2, r = 7.45$, with $q$ about 13,196 times the cutoff, first-order data risk about 4.25 times the degenerate term, and a full-panel budget about 16,754 times the local-boundary value. The width comparison also has a clear finite-window limit: at $s = 0.5$, both $\kappa = 2$ and 2.5 remain maximized at $g_1, r = 1$ through $n = 2^{40}$, with $B_R/n \approx 0.049841, 0.053217$. Their predicted asymptotic separation is not observed. Width one has $B_R/n \approx 29.920$ near $r = 9.2$; I do not select a window to fit exponents.

The pressure and fourfold-comparison tables below include every case specified in advance for those categories. The timing table reports an independent single-thread benchmark at $g_1, r = 1, s = 1$: medians of five repeats after three warmups, evaluating all declared prefixes up to the maximum count. Fresh includes construction and weights; reuse excludes setup. Data generation and population-reference calculations are excluded. The package contains all 22 timing anchors and individual repeats.

Table 3: Finite-panel frequency budgets at $\varepsilon = 0.5$, $s = 1$, across all four profiles. The targets are extra risk $\Delta R \le \varepsilon^2 q^2$ ($m_q$) and $\Delta R \le \varepsilon^2 R_\infty$ ($m_R$). These are rounded population-formula calculations, not certified continuum budgets. No runtime is extrapolated to the large-$n$ rows; measured costs at the actual simulation sizes are reported in Table 6.

| $n$ | Proposal | $m_q$ | $m_R$ |
|---|---|---|---|
| $2^8$ | Ordinary | 2,371 | 109 |
| $2^8$ | Broad defensive | 604 | 104 |
| $2^8$ | Fixed critical | 1,088 | 91 |
| $2^8$ | Local shifted | 654 | 188 |
| $2^{20}$ | Ordinary | 457,812 | 832,896 |
| $2^{20}$ | Broad defensive | 639 | 511,533 |
| $2^{20}$ | Fixed critical | 12,574 | 349,717 |
| $2^{20}$ | Local shifted | 1,751 | 1,154,908 |
| $2^{40}$ | Ordinary | 29,505,820,678 | 8,032,029,686,434 |
| $2^{40}$ | Broad defensive | 759 | 638,589,186,014 |
| $2^{40}$ | Fixed critical | 12,225,844 | 366,698,602,203 |
| $2^{40}$ | Local shifted | 2,438 | 1,214,551,781,123 |

Table 4: All seven empirical-planning and two spectral-necessary pressure cases, with their unchanged frozen repetitions. The latter are ordinary sampling at $n = 2048, g_1$, local, $m = 16, 24$; numerical necessary MCSE lower bounds are not interval-arithmetic certificates. All nine remain not certified. Top-1% contributions use correlated within-bank loss records descriptively; the inference unit is the full outer group. Bootstrap does not repair rare-tail undercoverage.

| $n$/profile/role | Proposal /$m$ | Extra/reference [95%] | MCSE/ ref. | Necessary MCSE/ref. | Top 1% |
|---|---|---|---|---|---|
| 128/g1/ middle | Ordinary/16 | $1.08\,[0.982, 1.19]$ | 5.53% | 0.97% | 34.80% |
| 128/g3/local | Broad defensive/ 128 | $0.954\,[0.893, 1.02]$ | 3.20% | 0.00% | 28.69% |
| 2048/g1/ middle | Ordinary/16 | $1.02\,[0.835, 1.22]$ | 9.66% | 8.45% | 55.66% |
| 2048/g1/ middle | Ordinary/24 | $0.936\,[0.805, 1.07]$ | 6.97% | 6.98% | 44.13% |
| 2048/g1/local | Ordinary/16 | $0.975\,[0.638, 1.36]$ | 18.90% | 22.67% | 76.25% |
| 2048/g1/local | Ordinary/24 | $1.02\,[0.702, 1.41]$ | 18.42% | 18.51% | 73.45% |
| 2048/g3/local | Local shifted/ 2048 | $1.01\,[0.945, 1.09]$ | 3.66% | 0.00% | 31.05% |
| 512/g1/local | Broad defensive/ 512 | $0.981\,[0.922, 1.05]$ | 3.30% | 0.00% | 27.38% |
| 512/g3/ middle | Ordinary/16 | $0.928\,[0.857, 1]$ | 3.58% | 0.52% | 30.32% |

### G.4 Earlier fixed-proposal study: models and design

This separate study fixes $h = 1$, identity geometry, the same explicitly defined $\rho, T, t_0$ above, and independent $X, T_{t_0} \sim N(0, I_d)$. The three-dimensional rotation uses $A_{23} = -1, A_{32} = 1$, all other entries zero, $v(X) = ae_1 + \ell AX$, $|a| \le 1$, $|\ell| \le 2$. Its moment matrix is $M = \mathrm{diag}(a^2 + 1, \ell^2 + 1, \ell^2 + 1)$, target $q = a^2/3^{5/2}$ and common envelope $\mathsf{H} = \mathrm{diag}(2, 5, 5)$. The scalar oscillator uses $v(X) = \gamma \cos(rX)/r$, $0 \le \gamma \le 1$, $r \ge 1$, with $M = 1 + \gamma^2(1 + e^{-2r^2})/(2r^2) \le 2$ and $\mathsf{H} = [2]$. Its exact target is $\gamma^2$ times (42). Actual energy varies within these common bounds. All model loops return to their own Gaussian endpoint.

Gaussian uses $\pi_s$; Energy uses the known-envelope trace proposal of Equation (63); pure Broad uses $N(0, 2I_d/h^2)$ with weight $2^{d/2} \exp(-h^2\|\omega\|^2/4)$ (Minh, 2016). Neither weights

Table 5: Paired actual-$U$ comparisons at the six predeclared $n = 512, g_1, s = 1$ anchor/proposal pairs. Budgets increase from $m = 22$ to 88 using the same frequency prefixes and full bootstrap groups. Intervals describe the ratio of empirical extra-risk means, not a population $1/m$ curve substituted for a random experiment.

| Anchor | Proposal | Observed extra-risk ratio [95%] | Population ratio |
|---|---|---|---|
| local | Broad defensive | $0.249\,[0.234, 0.265]$ | 0.25 |
| local | Fixed critical | $0.244\,[0.228, 0.26]$ | 0.25 |
| middle | Broad defensive | $0.245\,[0.232, 0.26]$ | 0.25 |
| middle | Fixed critical | $0.246\,[0.231, 0.262]$ | 0.25 |
| fixed | Broad defensive | $0.257\,[0.231, 0.284]$ | 0.25 |
| fixed | Fixed critical | $0.27\,[0.243, 0.301]$ | 0.25 |

Table 6: Prespecified single-thread timings at the $g_1, r = 1, s = 1$ anchors, collected in an independent benchmark: medians of five repeats after three warmups. Feature timings evaluate all declared prefixes up to the displayed maximum, not only one budget. Fresh includes bank construction and weights; reuse excludes bank setup. Data generation and population-reference computation are excluded. The CSV retains all 22 anchors, methods and individual repeats. These are measured scalar-statistic evaluation costs, not an end-to-end training benchmark.

| $n$ | Proposal | Max $m$ | Fresh (ms) | Reuse (ms) | Exact (ms) |
|---|---|---|---|---|---|
| 128 | Ordinary | 128 | 0.422 | 0.236 | 0.263 |
| 128 | Broad defensive | 128 | 0.498 | 0.289 | 0.263 |
| 128 | Fixed critical | 128 | 0.451 | 0.227 | 0.263 |
| 128 | Local shifted | 128 | 0.554 | 0.304 | 0.263 |
| 512 | Ordinary | 512 | 3.468 | 3.157 | 5.295 |
| 512 | Broad defensive | 512 | 4.254 | 3.969 | 5.295 |
| 512 | Fixed critical | 512 | 3.593 | 3.281 | 5.295 |
| 512 | Local shifted | 512 | 4.570 | 4.295 | 5.295 |
| 2048 | Ordinary | 2048 | 62.590 | 61.861 | 93.213 |
| 2048 | Broad defensive | 2048 | 75.554 | 74.473 | 93.213 |
| 2048 | Fixed critical | 2048 | 63.868 | 63.171 | 93.213 |
| 2048 | Local shifted | 2048 | 81.002 | 80.008 | 93.213 |

nor finite-kernel diagonals are renormalized. For a realized bank, target formulas are $q_m^{\mathrm{rot}} = a^2 m^{-1} \sum_j w_j \omega_{j,1}^2 e^{-\|\omega_j\|^2}$ and $q_m^{\mathrm{osc}} = \gamma^2 (4mr^2)^{-1} \sum_j w_j \omega_j^2 [e^{-(\omega_j - r)^2/2} + e^{-(\omega_j + r)^2/2}]^2$. The row second moment is $m^{-1} \sum_j w_j \omega_j^\top M \omega_j$. The population error decomposition is

$$\mathbb{E}(U - q)^2 = \mathbb{E}_{\mathrm{bank}} \operatorname{Var}_{\mathrm{data}}(U \mid \mathrm{bank}) + \mathbb{E}_{\mathrm{bank}}(q_m - q)^2. \tag{109}$$

Empirically, $(U - q)^2 = (U - q_m)^2 + (q_m - q)^2 + 2(U - q_m)(q_m - q)$ retains the cross term. At $q = 0$, only absolute errors and rejection rates are reported.

There are 95 nested finite-feature records and 25 unique exact references; 96 fixed null banks and two exact controls; 36 power conditions and six exact controls. Each exact null control uses its own 8000 data sets. Power amplitudes satisfy $\gamma^2 = cB_0/(nq_1) \le 1$ with $B_0 = 2$ and $q_1$ the unit-amplitude target at $r = 1$. Master seed 20260910 uses separate validation, pilot, formal and bootstrap streams. Data keys exclude proposal, $m$, H, and rotation $\ell$; banks are shared across $n$ and oscillator carriers at fixed proposal/$m$. Energy additionally keys the normalized envelope shape. Nested losses average eight data sets before 2000 paired outer-group bootstrap resamples; intervals are pointwise 95%, and ratios compare mean losses. Exact references are deduplicated, and scalar-slack comparisons reuse the same estimates.

### G.5 Local risk, importance sampling and calibration

At $\gamma = 1$, roots of $q = 1/n$ are $(2.57704, 2.94045, 3.26790, 3.56744, 3.84483)$ for the five sample sizes; $\lceil n^{0.8} \rceil = (28, 49, 85, 148, 256)$. Gaussian's fixed-$m = 32$ relative total MSE

Table 7: Earlier controlled-study grids. Nested units are 1024 outer groups with eight data sets each. Exact controls have no feature count; reused records are not independent replications.

| Study | Fixed grid | Independent units |
|---|---|---|
| Rotation | $a = 1$; $\ell = 0, 1, 2$; $n = 128, 512$; $m = 16, 64, 128$; Gaussian/exact | $1024 \times 8$ data sets |
| Rotation controls | $a = 0, n^{-1/2}$; $\ell = 0, 2$; $n = 128, 512$; $m = 64$; Gaussian/exact | $1024 \times 8$ data sets |
| Local oscillator | $q = 1/n$; $n = 64, 128, 256, 512, 1024$; $m = 32, \lceil n^{0.8} \rceil, n$; three proposals | $1024 \times 8$ data sets |
| Fixed oscillator | $r = 2, 4$; $n = 64, 256, 1024$; $m = 64$; Gaussian/exact | $1024 \times 8$ data sets |
| Matrix bound | $a = 1$; $\ell = 0, 2$; $n = 512, m = 64$; $\mathsf{H}, 12I, M$; three proposals | Shared nested records |
| Projection | $a = 1$; $m = 4, 8, 16, 32, 64, 128$; Gaussian | 256 banks $\times$ 2048 positions |
| Spectral error | $r = 1, \ldots, 6$; $m = 16, 64, 256, 1024$; three proposals | $B = 65536, 16384, 4096, 1024$ |
| Conditional null | Zero oscillator or $a = 0, \ell = 1$ rotation; $n = 128$; $m = 16, 128$ | 8 banks $\times$ 1000 data sets |
| Joint power | Oscillator $r = 1$; $q = cB_0/n$; $c = 2, 8, 32$; $n = 512, 1024$; $m = 64, n$ | 1000 fresh bank/data pairs |
| Scalar slack | Same power rows at $n = 512, m = 64$; all $c$; $\mathsf{H} \times 1, 2, 4$ | No new draws |

Table 8: Analytic feature-budget requirements and observed total risk are distinct. For the local $n = 1024$, $q = 1/n$ case ($r \approx 3.84483$), the two middle columns give the exact minimum integer $m$ needed for *relative feature MSE* at most 1 or $1/4$, using frozen independent spectral moments. The final column reports *total* relative MSE at $m = 32$ (exact: no features), with pointwise 95% outer-bank bootstrap intervals.

| Estimator | Feature risk $\leq 1$ | Feature risk $\leq 1/4$ | Total MSE$/q^2$ at $m = 32$ |
|---|---|---|---|
| Exact RBF | — | — | 1.259 [1.194, 1.332] |
| Gaussian RFF | 73 | 289 | 4.195 [3.574, 4.916] |
| Energy proposal | 7 | 25 | 1.504 [1.433, 1.584] |
| Broad Gaussian | 7 | 28 | 1.570 [1.484, 1.664] |

rises from 2.473 to 4.195, and its analytic feature component from 0.234 to 2.257. The $n^{0.8}$ rule keeps analytic relative feature error at 0.267–0.282. At $n = 1024$, total relative MSE is $4.195[3.574, 4.916]$, $1.588[1.503, 1.686]$, and $1.360[1.289, 1.439]$ for $m = 32, 256, 1024$; exact gives $1.259[1.194, 1.332]$. Every local interval, including exact, lies above one. These five sample sizes support the budget mechanism, not an asymptotic constant or a relative-error guarantee of one. For fixed Gaussian $m = 64$, $r = 2$ total relative MSE falls from 0.808 to 0.094 as $n = 64 \to 1024$, toward the analytic feature floor 0.0498. At $r = 4$ it falls from 426.7 to 4.410, but the final data component 2.678 still exceeds the empirical feature component 1.776.

Energy/Gaussian MSE ratios are 0.358–0.940 across all 15 local conditions, with pointwise paired intervals below one. At $n = 1024, m = 32$, Energy gives $1.504[1.433, 1.584]$ and paired ratio $0.358[0.307, 0.421]$. Its ratio to Broad at $m = 1024$ is $1.001[0.992, 1.010]$, establishing neither superiority nor equivalence. For rotation $n = 512, m = 64$, Energy/Gaussian ratios instead are 2.22–3.74 in six matrix-bound conditions, all pointwise intervals above one. Even $\mathsf{H} = M$ does not optimize spectral or total variance.

All proposals use the same known-envelope conditional-null rule

$$U_{n,m} > \frac{\sqrt{2}\,\overline{Q}_{\mathsf{H}}}{\sqrt{\alpha n(n-1)}}, \qquad \overline{Q}_{\mathsf{H}} = \frac{1}{m}\sum_j w_j \omega_j^\top \mathsf{H} \omega_j, \quad \alpha = 0.05. \tag{110}$$

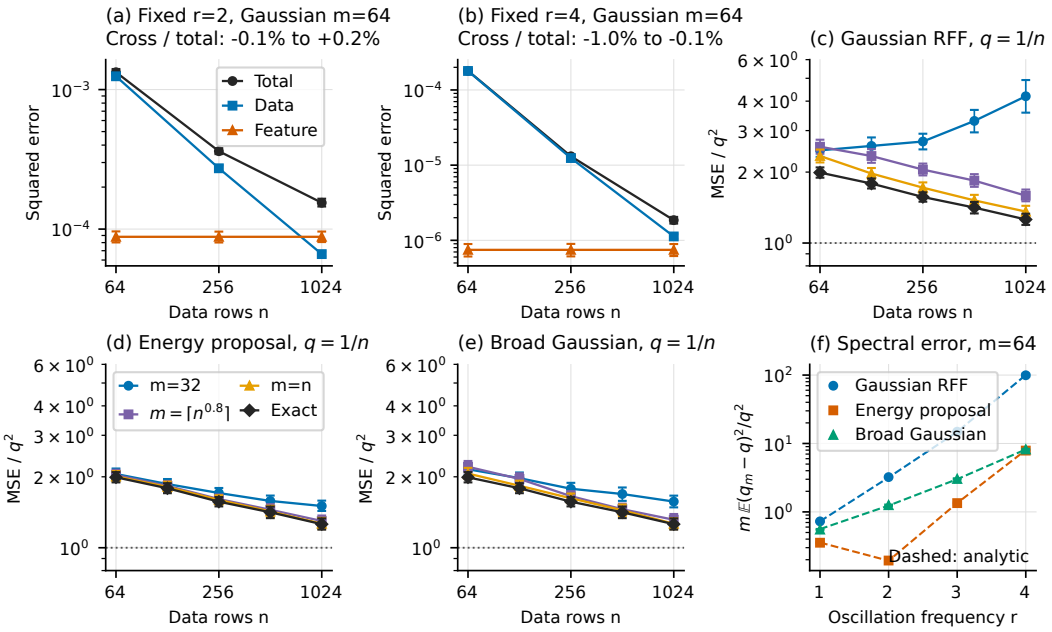



Figure 4: **Weak residuals require separate data and feature error accounting.** (a,b) At fixed oscillation frequency, Gaussian RFF ($m = 64$) shows a feature error floor as $n$ grows. The signed cross term is retained and its observed range relative to total MSE is annotated. (c–e) Along the prespecified $q = 1/n$ sequence, all three frequency budgets are compared with the same exact reference. The dotted line marks relative MSE one; it is not a guarantee achieved by these experiments. (f) Markers and pointwise bootstrap intervals show $m\,\mathbb{E}(q_m - q)^2/q^2$ at $m = 64$; dashed curves are independent analytic spectral moments. Only the primary $r \leq 4$ range is shown here. Nested intervals resample 1,024 outer bank groups (eight data sets each); spectral intervals resample their separate banks.

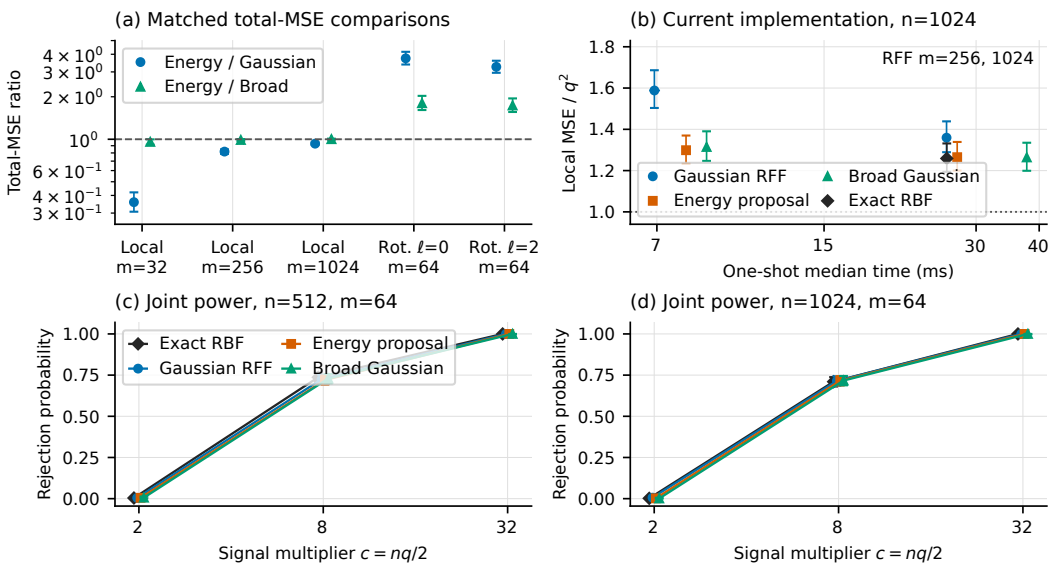



Figure 5: **Importance sampling improves selected budgets but is not uniformly superior.** (a) Paired total-MSE ratios compare the energy proposal with Gaussian and broad Gaussian proposals at identical $n, m, H$ within each case: local $n = 1024$ cases and rotation $n = 512, m = 64$ cases are distinct comparisons. (b) Error versus measured current-implementation cost at $n = 1024$, $m = 256, 1024$, and the unique exact reference. Vertical bars are pointwise MSE intervals; horizontal bars are interquartile ranges of 20 one-shot timings. Timing inputs use $r = 1$, while errors use the local-signal case at the same $d, n, m$; this is not a proposal-level algorithm speed ranking. (c,d) Joint power uses 1,000 independent bank/data pairs per point and the same known-$H$ conditional-null threshold rule for every method ($m = 64$). Bars are pointwise 95% Clopper–Pearson intervals; horizontal offsets distinguish overlapping points.

Energy makes $\overline{Q}_{\mathsf{H}} = B_0$; Gaussian and Broad use valid bank-dependent envelopes under the same calibration rule. All 96 fixed banks have zero rejections in 1000 independent data sets. The 98-comparison Bonferroni 95% Clopper–Pearson upper bounds, including exact controls, are 0.824% per finite bank and 0.103% per exact control. This shows conservatism, not error exactly zero or calibration near 5%. The archive also includes the 96-bank family. Joint power uses 1000 independent bank/data pairs per condition: rates are 0.1–0.6% at $c = 2$, 71.0–73.7% at $c = 8$, and 100% at $c = 32$ (individual interval lower endpoint 99.63%). No consistent Energy power advantage is established.

### G.6 Rotation, bound sensitivity and spectral tails

Changing $\ell$ at fixed $a$ preserves the whole model marginal path, $q$, and every population Fourier moment. For $a = 1$, the measured projection $\langle \mu_m, B_{m,X} AX \rangle$ is squared and averaged over 2048 positions in each of 256 banks. Its bank mean divided by $\kappa_A/m$ ranges from 0.954 to 1.075, with all six pointwise intervals containing one. The exact projection vanishes by $Ae_1 = 0$, an analytic identity, not another numerical observation. The predicted first-order increment $4a^2\ell^2\kappa_A/(nm)$ contributes only 0.85–7.11% of total MSE for $a = 1, \ell = 1, 2$, about 0.058% for $a = n^{-1/2}$, and zero for $a = 0$. Total error nevertheless rises in all 24 paired rotation comparisons, including exact and zero-translation controls; higher-order terms matter.

At $\ell = 2$, $M$ equals the common envelope and those labels share estimates. Tightening $12I$ to $M$ reduces the stated bound yet increases Energy's empirical MSE. The joint bound is 8.9–53.0 times observed MSE in the local oscillator grid and 13.3–113.5 times it in the matrix grid. Scalar multiples of H leave Energy's proposal, weights and estimates unchanged but raise thresholds: at $n = 512, c = 8, m = 64$, multipliers $1, 2, 4$ reduce its power from 72.0% to 6.4% to zero. Other proposals show the same qualitative threshold cost.

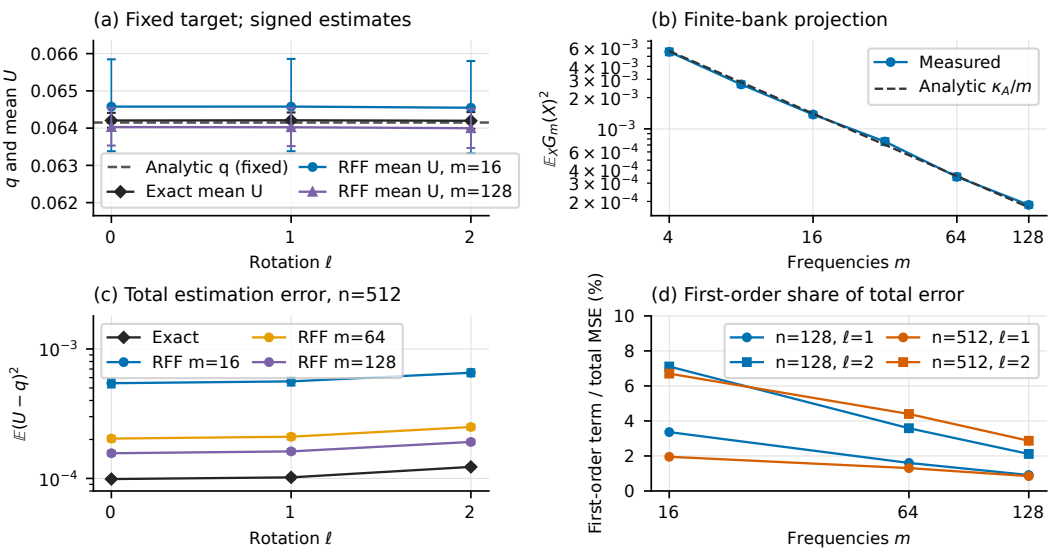



Figure 6: **Finite features reveal a projection that the exact isotropic kernel cancels.** (a) The analytic squared target $q = 0.06415$ is constant across rotation strengths; markers show signed estimator means ($a = h = 1$, $n = 512$). (b) Measured conditional projection second moments agree with the analytic $\kappa_A/m$ prediction; the exact projection is zero by the identity $Ae_1 = 0$, not a numerical measurement. (c) Total MSE includes all variance and feature-target terms. (d) The analytic first-order increment divided by measured total MSE is at most 7.11% on the displayed $a = 1$ grid; this term does not dominate total error. Error bars in (a)–(c) are pointwise 95% outer-bank bootstrap intervals; (d) is a descriptive ratio without an interval.

The 72 spectral conditions use approximately $2^{20}$ frequency draws per proposal/$m$, sharing banks across the six carriers. All 48 primary $r \leq 4$ conditions contain analytic MSEs in their reported intervals and have analytic relative MCSE at most about 5%. Gaussian at $r = 5, 6$ and Energy at $r = 6$ exceed the prespecified 10% criterion at every $m$. Gaussian at $r = 6, m = 256$ yields relative feature MSE $32.725[10.352, 71.391]$, below analytic 81.009, with analytic relative MCSE about 82%. Four nested Gaussian feature components also exceed 10%: local $n = 1024, m = 32$ and all three fixed-$r = 4$ cases. These are feature-component flags, not total-MSE precision lower bounds. Broad has smaller analytic pure feature MSE than Energy at $r = 5, 6$; this is not a complete-estimator power ranking. Second/fourth centered spectral moments, loss quantiles, top-1% shares, batch and cumulative means are archived. I added no repetitions after examining the results.

**Implementation timing.** I timed the earlier 28 anchors after experimental workers exited, using one worker/BLAS thread, three warmups and 20 fresh-bank repeats. Reuse is one setup plus 100 data-set evaluations divided by 100, not 100 independent timing observations. Data generation is excluded. At $n = 512, m = 32$, Energy takes 0.589 ms fresh and 0.283 ms reused, versus exact 7.082 and 7.043 ms; at $n = m = 1024$, Gaussian 26.220 ms is close to exact 26.249 ms. Timing matches dimension/$n, m$ but uses different signals from the error comparisons. Broad uses complex exponentials, whereas Gaussian/Energy use real sine–cosine features; these are implementation costs, not a proposal-algorithm speed ranking. Identical peak-RSS records include process setup and pre-generated data, so do not establish comparative memory efficiency.

### G.7 Legacy diagnostic studies and their distinct targets

Under the exact Gaussian strong null, median $\sqrt{|U_n|}$ has log–log slope $-0.528$ over $n = 32, 64, 128, 256, 512$. This is a labeled scaling display, not a WSD estimator. Across 40 settings, the mean rescaled diagonal gap $n(V_n - U_n)$ varies by a factor 1.002 over these sample sizes. Plug-in correlations with a positive weighted trace-MMSE baseline stay at 0.897–0.872, whereas signed-U correlations range from $-0.052$ to 0.062, isolating the diagonal term in Equation (14). The origi-

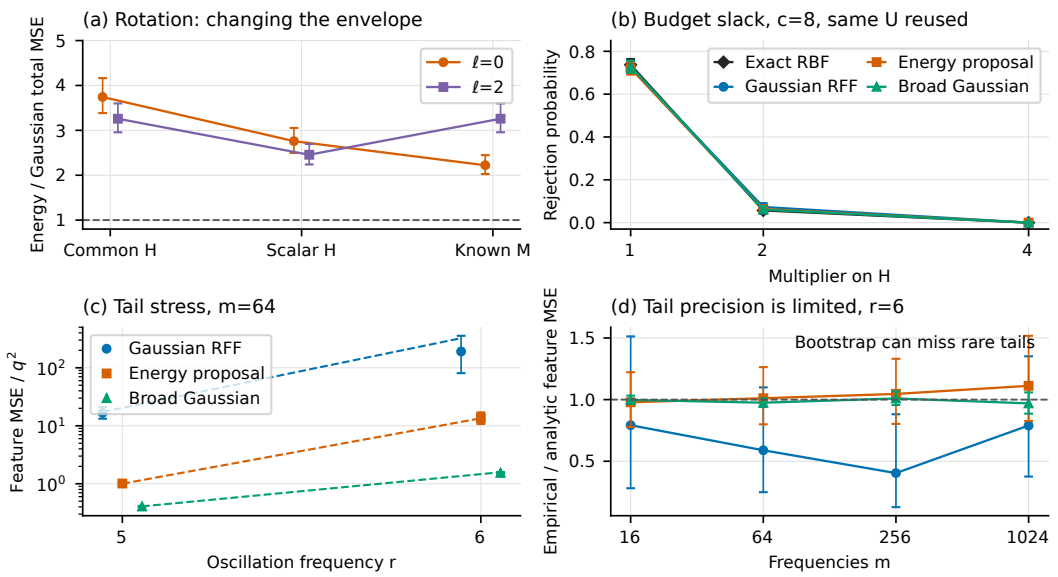



Figure 7: **Controls and tail stress delimit the empirical conclusions.** (a) At $n = 512, m = 64$, neither a scalar envelope nor the exact population second-moment matrix makes Energy uniformly preferable for rotation; ratios have paired pointwise bootstrap intervals. (b) Enlarging the scalar budget raises thresholds on the same statistics ($n = 512, m = 64, c = 8$); intervals are pointwise Clopper–Pearson. (c) At $r = 5, 6$, analytic feature-risk curves and estimates show Broad can outperform Energy. (d) At $r = 6$, empirical/analytic risk ratios expose limited tail precision. Bootstrap intervals can miss rare tails. The independent population-moment MCSE diagnostics are in `paper/generated/budget_appendix_tables.tex` within `supplementary_data.zip`. No repeats were appended.

Table 9: All affine sensitivity settings, changing one predeclared grid or bandwidth choice at a time. The convex mixture is not response-adaptive.

| Setting | Spearman $\rho$ | Conditional 95% interval |
|---|---|---|
| 3 log-SNR points | 0.710 | $[0.389, 0.868]$ |
| 5 log-SNR points | 0.856 | $[0.660, 0.946]$ |
| 9 log-SNR points | 0.777 | $[0.480, 0.929]$ |
| Bandwidth multiplier 0.5 | 0.563 | $[0.171, 0.809]$ |
| Bandwidth multiplier 1 | 0.856 | $[0.660, 0.946]$ |
| Bandwidth multiplier 2 | 0.763 | $[0.470, 0.914]$ |
| Fixed mixture of 0.5, 1, 2 | 0.779 | $[0.506, 0.925]$ |

nal diagnostic image is available as `results/figures/week2_propositions.png` in the package.

The affine ablation has 24 paired conditions in dimension 64, two data repeats and two feature repeats. Table 9 compares whitened signed integrated-WSD-squared to Monte Carlo integrated SD. Its 95% condition-bootstrap intervals are conditional on the fitted transform, bandwidth and recorded repeats; neither grid refinement nor bandwidth mixing gives monotone or invariant rank recovery.

The 16-condition, two-dimensional Gaussian-mixture stress test uses the recorded data-source reference D, not each velocity's generated endpoint. At $n = 192, m = 512$, its whitened signed random-feature integrated WSD-squared/analytic-reference Monte Carlo SD correlation is $0.515[-0.065, 0.899]$; the $n = 48, 96, 192$ sweep gives $0.621, 0.350, 0.515$. The interval includes zero and the sweep is non-monotone, so stable rank recovery is not established. Both affine and mixture studies use legacy random-phase maps, not the paired-bank estimator of Appendix A.1. These diagnostics provide no trained high-dimensional performance claim.