EQUIVALENT TARGETS, DIFFERENT SAMPLING COSTS: PRECISION AND SCALE GEOMETRY OF NOISY SCORES

Anonymous authors
Paper under double-blind review

ABSTRACT

Statistically equivalent targets can have asymptotically different sampling costs under the same score interface. We establish this through an oracle that returns the full normalized smoothed-score vector at one shared scale, with independent Gaussian observation noise and a precision cap. For geometrically spaced Fourier and spatial families, constant target capacity coexists with an optimal cost of $\Theta(\sqrt{d})$ queries at precision $B = d$ and fixed, sufficiently small average total variation error. Preserving the entire statistical experiment, we then construct Blackwell-equivalent spatial targets with a diverging ratio of optimal costs at the same dimension, precision, and accuracy. Matching bounds across all precision levels explain this separation at fixed relative TV accuracy in a local perturbation regime under a uniform prior. Precision determines the feasible coordinate block length; at high precision, the optimal gain over coordinate thinning is of the order of its square root. A blockwise converse quantifies the dependence required by an accurate joint output. Shared activation attains the bound by matching coordinate marginals, cancelling single-coordinate error terms, and controlling the remaining interactions. Nonadaptive samplers match lower bounds allowing adaptive queries and random stopping.

1 INTRODUCTION

The cost of score-based sampling depends on what each vector evaluation makes available to the generator. Standard diffusion models evaluate a noise-prediction vector at a common noise level, and fast solvers seek to reduce the number of these evaluations (Ho et al., 2020; Lu et al., 2022). A full vector can expose coordinates with very different sensitivities at that level. This raises a question about the fundamental value of a query: when two targets contain the same statistical information, can their generation costs still differ under the same observation rules?

Recent query lower bounds make these rules central to sampling complexity. Xun & Price (2026) identify scale search as a source of difficulty under global L^p score guarantees; Liu (2026) develops a channel-synthesis converse for finite-bit smoothed-score communication. We study how the placement of parameter dependence across smoothing scales changes the optimal number of finite-precision vector queries. Our comparison preserves the target experiment while varying this scale layout, separating its effect from the statistical information to be generated.

Let a hidden sign vector Θ select a target P_Θ, and let $s_{\Theta,\sigma}$ be its score after Gaussian smoothing at scale σ. At time t, the algorithm chooses a point x_t, one scalar scale σ_t shared across coordinates, and precision $0 < b_t \le B$, and receives

$$Y_t = \sigma_t s_{\Theta,\sigma_t}(x_t) + b_t^{-1/2} G_t, \qquad G_t \sim \mathcal{N}(0, I_d). \tag{1}$$

This observation model gives precision a common meaning across scales. The normalized score is the negative conditional noise prediction: $-\sigma s_{\theta,\sigma}(x) = \mathbb{E}[Z \mid X + \sigma Z = x]$ for independent $X \sim P_\theta$ and $Z \sim \mathcal{N}(0, I_d)$ (Vincent, 2011; Ho et al., 2020). Fresh independent Gaussian observation noise permits precision to accumulate under repetition. The algorithm may adapt queries and stopping; the cap B fixes the quality available from any one call.

For fixed $q > 1$, the Fourier family perturbs Gaussian coordinates at frequencies $k_j = q^j$, and the spatial family perturbs Gaussian mixture weights at radii $r_j = q^j$. The families have different

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.

score-response decay profiles but the same sampling complexity law. With $\varepsilon = \alpha/\sqrt{d} \le 1/4$, $0 < \alpha \le 1$, we minimize expected calls subject to average TV error at most $\delta = \kappa\alpha$ under the uniform prior on Θ, for sufficiently small fixed $\kappa > 0$.

Target capacity, the maximum mutual information between the hidden parameter and one target sample over all priors, is $\Theta(\alpha^2)$. Nevertheless, at $\alpha = 1$ and $B = d$, the optimal query count is $\Theta(\sqrt{d})$ (Corollary 2). We next preserve the complete target experiment, retaining statistical detail that capacity alone can miss. Public transformations independent of the hidden parameter convert samples from either spatial layout into samples from the other. These layouts are Blackwell equivalent, yet their optimal query costs differ under the same rules. Three results quantify this distinction.

- Theorem 3 compares clustered and separated spatial layouts with the same ambient dimension, precision cap, and accuracy requirement. Their optimal cost ratio grows as the square root of the effective block length under the stated spacing and precision conditions. A concrete sequence gives a $\sqrt{\log m}$ separation with m active parameters.
- Theorem 1 gives the rates d/B, $d/\sqrt{B}$, and $\alpha\sqrt{d/\ell_B}$, where $\ell_B = \min\{d, 1 + \log(1 + \varepsilon\sqrt{B})\}$. Repetition, coordinate thinning, and block activation respectively attain these rates for both families, with constants depending only on q, κ.
- At high precision, the optimal saving relative to the coordinate-thinning construction is of order $\sqrt{\ell_B}$. A blockwise converse and a shared-activation sampler give matching bounds, including against adaptive queries and random stopping (Lemma 5 and Proposition 10).

At high precision, bounded witnesses turn one global TV guarantee into aggregate requirements on how the output law changes with blocks of hidden parameters. Blockwise sensitivity bounds what each response can supply. A matching sampler shares a random activation decision across coordinates to save calls. This can introduce dependence between generated coordinates. Marginal matching cancels all single-coordinate error terms; controlling the remaining interactions determines the activation frequency needed for joint accuracy (Figure 1).

These nonadaptive samplers attain all three rates. The constructions and transcript bounds also give optimal rates for threshold observations and positive linear combinations of expected calls and cumulative precision.

2 Model and optimal query complexity

Fix $q > 1$ and put $k_j = r_j = q^j$ for $j \in [d] = \{1, \dots, d\}$. A hidden sign vector $\theta \in \{\pm 1\}^d$ indexes the target families. Let ϕ_v be the density of $\mathcal{N}(0, v)$ and $\phi = \phi_1$. The Fourier family encodes signs in oscillatory perturbations:

$$h_j^{\mathsf{F}}(x) = \cos(k_j x) - e^{-k_j^2/2}, \qquad p_\theta^{\mathsf{F}}(x) = \prod_{j=1}^{d} \phi(x_j)\big(1 + \varepsilon\theta_j h_j^{\mathsf{F}}(x_j)\big). \tag{2}$$

The spatial family encodes signs in mixture weights, with each factor obtained by rescaling a common base:

$$\bar{p}_j(x) = \tfrac{1}{2}\phi_{r_j^2}(x - r_j) + \tfrac{1}{2}\phi_{r_j^2}(x + r_j), \qquad h_j^{\mathsf{S}}(x) = \tanh(x/r_j), \tag{3}$$

$$p_\theta^{\mathsf{S}}(x) = \prod_{j=1}^{d} \bar{p}_j(x_j)\big(1 + \varepsilon\theta_j h_j^{\mathsf{S}}(x_j)\big). \tag{4}$$

For either family $\mathsf{a} \in \{\mathsf{F}, \mathsf{S}\}$, each factor is $\mu_j^{\mathsf{a}}(dx)(1 + \varepsilon\theta_j h_j^{\mathsf{a}}(x))$, where $\mu_j^{\mathsf{F}} = \mathcal{N}(0, 1)$ and μ_j^{S} has density $\bar{p}_j$. In both families, $\int h_j^{\mathsf{a}}(x)\, \mu_j^{\mathsf{a}}(dx) = 0$, so every coordinate factor integrates to one. Throughout,

$$0 < \varepsilon = \frac{\alpha}{\sqrt{d}} \le \frac{1}{4}, \qquad 0 < \alpha \le 1. \tag{5}$$

Since $\|h_j^{\mathsf{a}}\|_\infty \le 2$, all perturbation factors lie in $[1/2, 3/2]$; this keeps the score and divergence denominators positive. We use $\mathrm{TV}(P, Q) = \frac{1}{2}\int |dP - dQ|$. The notation $f \asymp_{q,\kappa} g$ denotes two-sided bounds with positive constants depending only on q, κ; generic proof constants may change between lines.

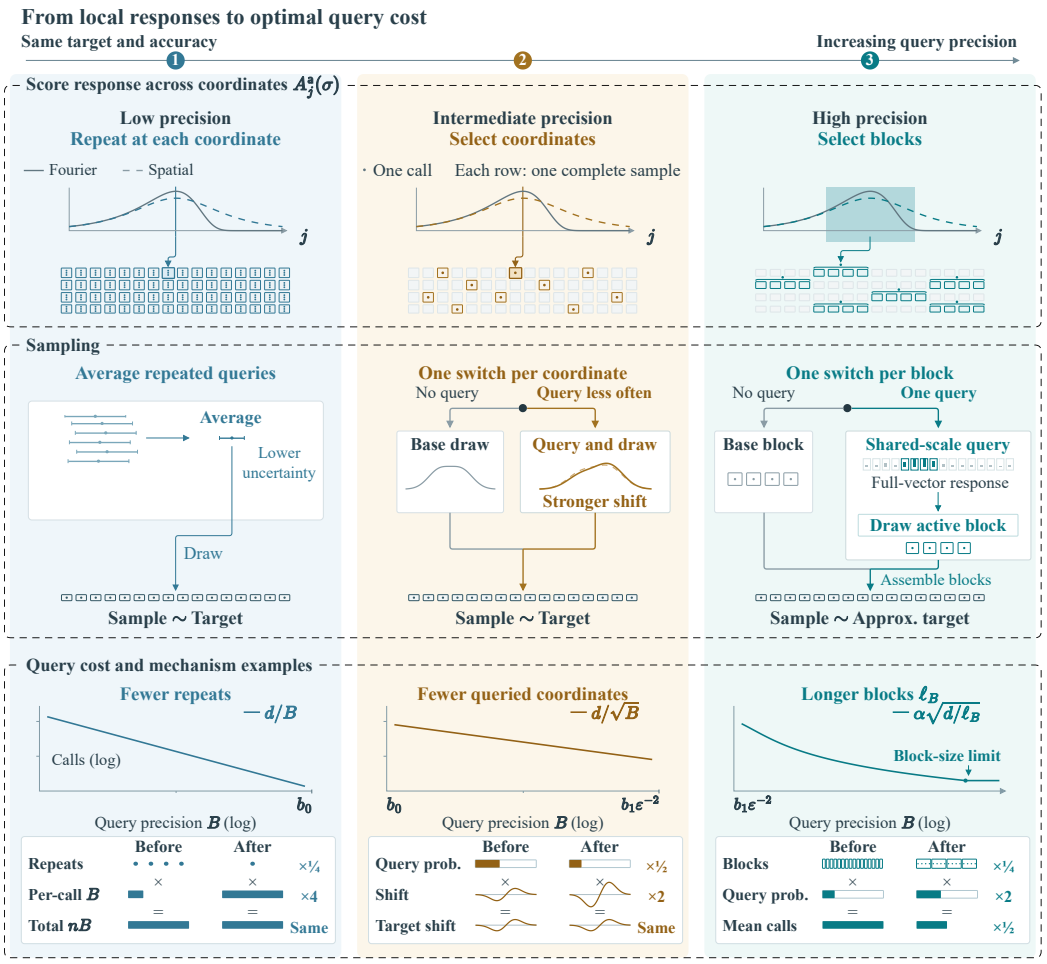


Figure 1: Precision changes the optimal use of a full-vector query: repetition improves reliability, coordinate thinning reduces query frequency, and block activation shares calls while controlling joint error. For the fixed targets and accuracy of Theorem 1, the first two procedures are exact; the third preserves coordinate marginals and controls joint TV. Curves are schematic. The examples illustrate $nB = \text{constant}$, $ps = \varepsilon$, and $\mathbb{E}T = Mp$ for feasible $L \asymp_{q,\kappa} \ell_B$; comparison constants depend only on q, κ.

For $\sigma > 0$, let $p^{\mathsf{a}}_{\theta,\sigma} = p^{\mathsf{a}}_{\theta} * \mathcal{N}(0, \sigma^2 I_d)$ and $s^{\mathsf{a}}_{\theta,\sigma} = \nabla \log p^{\mathsf{a}}_{\theta,\sigma}$. A legal protocol uses the oracle in (1), choosing $x_t \in \mathbb{R}^d$, one shared $\sigma_t > 0$, and $0 < b_t \le B$ predictably from its history. The Gaussian noises are independent of the uniform parameter $\Theta \sim \mathrm{Unif}\{\pm 1\}^d$ and the algorithmic seed. Its stopping time T is almost surely finite for every parameter and has finite Bayes mean. The output law Q_θ is obtained from the transcript by a randomized Markov kernel independent of θ. Expectations average over the prior, oracle noise, and algorithmic randomness. Define

$$T^*_{\mathsf{a}}(d, \varepsilon, q, B, \delta) := \inf_{\substack{\mathcal{A} \text{ legal:} \\ 2^{-d}\sum_\theta \mathrm{TV}(Q_\theta, P^{\mathsf{a}}_\theta) \le \delta}} 2^{-d} \sum_\theta \mathbb{E}_\theta T. \tag{6}$$

For the fixed relative error $\delta = \kappa\alpha$, set

$$\ell_B := \min\{d, 1 + \log(1 + \varepsilon\sqrt{B})\}. \tag{7}$$

with natural logarithms. For thresholds $b_0, b_1 > 0$ satisfying $b_0 \le 16 b_1$, write

$$\mathcal{R}_d(B) = \begin{cases} d/B, & 0 < B \le b_0, \\ d/\sqrt{B}, & b_0 < B \le b_1 \varepsilon^{-2}, \\ \alpha\sqrt{d/\ell_B}, & B > b_1 \varepsilon^{-2}. \end{cases} \tag{8}$$

Theorem 1 (Query complexity for the Fourier and spatial families)**.** *For every fixed $q > 1$, there are constants $\kappa_0(q) > 0$ and $b_0(q) > 0$. For every $0 < \kappa < \kappa_0(q)$, there is a constant $b_1(q, \kappa) > 0$ satisfying $b_0(q) \le 16 b_1(q, \kappa)$. These constants can be chosen so that the following holds for all integers $d \ge 1$, $0 < \alpha \le 1$, and $B > 0$ with $\varepsilon := \alpha/\sqrt{d} \le 1/4$. Set $\delta := \kappa\alpha$ and define $\mathcal{R}_d$ by* (8) *with these thresholds. Then*

$$T^*_{\mathsf{F}}(d, \varepsilon, q, B, \kappa\alpha) \asymp_{q,\kappa} T^*_{\mathsf{S}}(d, \varepsilon, q, B, \kappa\alpha) \asymp_{q,\kappa} \mathcal{R}_d(B). \tag{9}$$

The three branches are achieved by repetition, randomized coordinate thinning, and randomized block activation. Their accuracy guarantees are uniform in θ, and all comparison constants are uniform in d, α, B.

The high precision branch measures the gain from jointly generating a feasible block: its length grows logarithmically until $\ell_B = d$, where the cost has order α. Nonadaptive samplers attain lower bounds covering adaptive queries and random stopping, so adaptation cannot improve the order of this cost.

The capacity $C(P^{\mathsf{a}}) = \sup_\Pi I_\Pi(\Theta; X)$, with $X \mid \Theta = \theta \sim P^{\mathsf{a}}_\theta$ and the supremum over all priors, is $\Theta_q(\alpha^2)$ (Proposition 4, Appendix B). Combining this with the theorem gives a direct separation between target information and query cost.

Corollary 2 (Constant target capacity and growing query complexity)**.** *Under the assumptions of Theorem 1, for each $\mathsf{a} \in \{\mathsf{F}, \mathsf{S}\}$ and every $d \ge 16$, if $\alpha = 1$, $\varepsilon = d^{-1/2}$, $B = d$, and $\delta = \kappa$, then*

$$C(P^{\mathsf{a}}) = \Theta_q(1), \qquad T^*_{\mathsf{a}}(d, d^{-1/2}, q, d, \kappa) = \Theta_{q,\kappa}(\sqrt{d}). \tag{10}$$

The comparison constants depend only on the fixed values of q and κ.

3 Joint sampling requirements and query lower bounds

An accurate output must retain parameter dependence supplied by the query transcript. Comparing this output demand with response sensitivity turns the geometry of the observation interface into a sampling lower bound. Coordinatewise comparisons determine the low and intermediate precision rates; a blockwise comparison captures the collective dependence required when one query serves several coordinates. The transcript bounds allow adaptive queries and random stopping.

Dependence required by the output. For an output channel $Q = \{Q_\theta\}$, measure sensitivity to one sign by

$$\bar{v}_j(Q) = 2^{-(d-1)} \sum_{\eta \in \{\pm 1\}^{d-1}} \mathrm{TV}(Q_{+,\eta}, Q_{-,\eta}). \tag{11}$$

Here $Q_{\pm,\eta}$ denotes the output law with $\theta_j = \pm 1$ and the remaining signs fixed at η. For $L \in [d]$, take $M = \lfloor d/L \rfloor$ disjoint consecutive blocks $\mathcal{B}_r$ of length L. For each block r, write $\zeta = \theta_{\mathcal{B}_r} \in \{\pm 1\}^L$ and $\eta = \theta_{[d]\setminus\mathcal{B}_r} \in \{\pm 1\}^{d-L}$. Averaging uniformly over ζ and η, define

$$J_r(Q) = \mathbb{E}_{\eta,\zeta}\, \mathrm{TV}\left(Q_{\zeta,\eta}, 2^{-L} \sum_{\zeta'} Q_{\zeta',\eta} \right). \tag{12}$$

These are changes with the parameter, not dependence among target coordinates. For any output satisfying $2^{-d} \sum_\theta \mathrm{TV}(Q_\theta, P^{\mathsf{a}}_\theta) \le \delta = \kappa\alpha$, with κ as in Theorem 1, Lemma 5 gives

$$\sum_{j=1}^{d} \bar{v}_j(Q) \gtrsim_q \varepsilon d, \qquad \sum_{r=1}^{M} J_r(Q) \gtrsim_q \alpha\sqrt{M}. \tag{13}$$

The permitted error $\delta = \kappa\alpha$ can exceed the perturbation ε of an individual coordinate. A normalized witness combines the weak coordinate biases before applying this global error allowance. With the

prior fixed, average conditional TV equals the TV distance between the joint laws of parameter and sample. A witness bounded by one therefore loses at most 2δ in expectation, independently of the number of coordinates it aggregates. Here $\mathbb{E}_P$ and $\mathbb{E}_Q$ average over the uniform parameter and the corresponding conditional sample law.

For the coordinate witness

$$F_{\mathrm{c}}(\theta, x) = \tanh\left(d^{-1/2} \sum_{j=1}^{d} \theta_j \operatorname{sign} h_j^{\mathsf{a}}(x_j)\right),$$

the target expectation is at least $c_{\mathrm{f}}(q)\alpha$. Averaging each parameter sign in turn changes the expectation by at most $\bar{v}_j(Q)/\sqrt{d}$; after all signs are averaged, symmetry makes it zero. Consequently,

$$c_{\mathrm{f}}(q)\alpha - 2\delta \le \mathbb{E}_Q F_{\mathrm{c}} \le \frac{1}{\sqrt{d}} \sum_{j=1}^{d} \bar{v}_j(Q).$$

To measure demand on whole parameter blocks, first aggregate within each block and then across blocks:

$$W_r(\theta, x) = \tanh\left(L^{-1/2} \sum_{j \in \mathcal{B}_r} \theta_j \operatorname{sign} h_j^{\mathsf{a}}(x_j)\right),$$

$$F_{\mathrm{b}}(\theta, x) = \tanh\left(M^{-1/2} \sum_{r=1}^{M} W_r(\theta, x)\right).$$

Averaging whole parameter blocks gives

$$c_{\mathrm{b}}(q)\alpha_L - 2\delta \le \mathbb{E}_Q F_{\mathrm{b}} \le \frac{4}{\sqrt{M}} \sum_{r=1}^{M} J_r(Q), \qquad \alpha_L = \varepsilon\sqrt{ML}.$$

Since $ML \ge d/2$, the block witness has target bias of order α. Sufficiently small relative error leaves a positive output bias in both comparisons, yielding the coordinatewise and blockwise demands from the same global TV guarantee. Appendices C and F give the full witness arguments.

Sensitivity supplied by a shared scale. The two normalized-score profiles are

$$A_j^{\mathsf{F}}(\sigma) = \frac{k_j \sigma}{1+\sigma^2} \exp\left\{-\frac{k_j^2 \sigma^2}{2(1+\sigma^2)}\right\}, \tag{14}$$

$$A_j^{\mathsf{S}}(\sigma) = \frac{r_j \sigma}{r_j^2 + \sigma^2}. \tag{15}$$

At a fixed query action (x, σ, b), flipping θ_j changes only the jth coordinate of the response mean, with absolute difference at most $(8/3)\varepsilon A_j^{\mathsf{a}}(\sigma)$, uniformly over query points. For the Fourier family, setting $u = k_j\sigma/\sqrt{1+\sigma^2}$ gives $A_j^{\mathsf{F}} = (1+\sigma^2)^{-1/2} u e^{-u^2/2}$; for the spatial family, $u = \sigma/r_j$ gives $A_j^{\mathsf{S}} = u/(1+u^2)$. Their different tails obey the same geometric summability bounds. These bounds account for both strong responses and weak tail responses at a common scale, and support comparable logarithmic block lengths. Lemma 8 records them in Appendix E; exact scores are in Appendix A.

Let H_T denote the complete stopped transcript, including the algorithmic seed, query actions, responses, and stopping indicators. Compare transcript laws under two fixed parameters at the same realized history. The predictable action is identical, so Gaussian shift bounds apply at each active step. A TV hybrid bound and a stopped KL chain rule accumulate them through T (Lemma 7, Appendix D). Writing $\mathsf{H} = \{\mathrm{Law}_\theta(H_T)\}_\theta$, the resulting bounds are

$$\mathbb{E} \sum_{t \le T} b_t \gtrsim_{q,\kappa} d,$$

$$\mathbb{E} \sum_{t \le T} \log(1 + C_q \varepsilon \sqrt{b_t}) \gtrsim_{q,\kappa} \varepsilon d, \tag{16}$$

$$\sum_{r=1}^{M} J_r(\mathsf{H}) \le C_q \mathbb{E} \sum_{t \le T} \min\left\{M, 1 + \frac{\log(1+\varepsilon\sqrt{b_t})}{L}\right\}. \tag{17}$$

The first two inequalities combine coordinate demands with the squared and truncated profile sums. The last bounds the parameter dependence that a shared-scale response can supply across whole blocks, accumulated over the stopped transcript. Lemma 9 and Appendix G give the general resource bounds.

From these bounds to the three rates. Under $b_t \le B$, the coordinate inequalities give

$$\mathbb{E}T \ge c_{q,\kappa} \max\left\{ \frac{d}{B}, \frac{\varepsilon d}{\log(1 + C_q \varepsilon \sqrt{B})} \right\}. \tag{18}$$

This yields d/B at low precision and $d/\sqrt{B}$ while $\varepsilon\sqrt{B} = O_{q,\kappa}(1)$. At high precision, take $L \asymp \ell_B$ before saturation and $L = d$ after saturation. The per-query block budget in (17) is then $O_q(1)$. Since output data processing gives $J_r(Q) \le J_r(\mathsf{H})$, the block demand yields

$$\mathbb{E}T \ge c_{q,\kappa} \alpha \sqrt{d/\ell_B}. \tag{19}$$

At high precision before saturation, the coordinatewise TV term in (18) has order $\alpha\sqrt{d}/\ell_B$. The blockwise demand strengthens this by a factor of order $\sqrt{\ell_B}$: an accurate joint output must retain collective dependence on whole parameter blocks. At saturation, $\ell_B = d$ yields a lower bound of order α.

These bounds establish the converse in Theorem 1; Appendix J handles the fixed transition intervals. The constructions below show that the stronger blockwise requirement determines the attainable cost of sharing queries.

4 Optimal sampling across precision regimes

All three samplers convert a noisy response directly into the required perturbation of a base distribution. For $0 \le s \le 1/4$, write $P^{\mathsf{a}}_{s,\theta_j,j}(dx) = \mu^{\mathsf{a}}_j(dx)(1 + s\theta_j h^{\mathsf{a}}_j(x))$, with $P^{\mathsf{a}}_{0,j} = \mu^{\mathsf{a}}_j$. At a public query point and scale, threshold response coordinate j at the public midpoint of its two possible means. After orienting the bit by the known sign of the mean gap, this gives $Z \in \{\pm 1\}$ with $\mathbb{E}[Z \mid \theta_j] = \rho\theta_j$. If $\rho \ge 2s$, independent sign flips reduce its correlation to $2s$. The kernel

$$R^{\mathsf{a}}_j(dx \mid z) = \mu^{\mathsf{a}}_j(dx)\big(1 + \tfrac{1}{2} z h^{\mathsf{a}}_j(x)\big) \tag{20}$$

then produces exactly $P^{\mathsf{a}}_{s,\theta_j,j}$. At the target amplitude $s = \varepsilon$, the required correlation is only $2\alpha/\sqrt{d}$, so a response with vanishing dependence on the hidden sign can still generate the exact coordinate law. The constructions allocate queries to supply this prescribed dependence: repetition combines weak responses, while random activation uses a stronger response less often.

Repetition and coordinate thinning. For each coordinate j, a public query point and scale can be chosen so that the two possible means of response coordinate j are separated by at least $c_q\varepsilon$, where $c_q > 0$ depends only on q. We call such a query matched to coordinate j; Appendix H gives the choices for both families. At low precision, repeat a matched query $n \asymp_q B^{-1}$ times per coordinate. Averaging yields precision $nB = \Theta_q(1)$, sufficient for amplitude $s = \varepsilon$. Independent coordinate procedures generate the exact target in $O_q(d/B)$ calls. At intermediate precision, a single matched response supports a larger amplitude. Activate each coordinate independently with probability $p \asymp_{q,\kappa} B^{-1/2}$, generate amplitude $s = \varepsilon/p$ when active, and otherwise draw from the base. Then

$$(1 - p)P^{\mathsf{a}}_{0,j} + pP^{\mathsf{a}}_{s,\theta_j,j} = P^{\mathsf{a}}_{\varepsilon,\theta_j,j}, \tag{21}$$

so the product output is exact with expected cost $O_{q,\kappa}(d/\sqrt{B})$. Appendix H gives the matched scales and feasible activation probabilities.

Block activation and joint accuracy. At high precision, a common scale can serve an entire block. Activating its coordinates together saves queries; choosing the activation probability also controls the resulting joint distribution. Partition $[d]$ into $M = \lceil d/L \rceil$ consecutive blocks, including a shorter final block if needed. Fix $\Gamma \ge 4$ sufficiently large in terms of q, κ, and set

$$p = \min\{1, \Gamma\varepsilon\sqrt{L}\}, \qquad s = \varepsilon/p. \tag{22}$$

Blocks are activated independently, so the full output is a product over blocks. An inactive block outputs its base law; an active block uses one shared-scale query and independent postprocessing to generate the product perturbation of amplitude s.

For this synthesis kernel, every coordinate's binary channel must support correlation at least $2s$. At suitable query points, a sufficient condition on the common scale is

$$\varepsilon\sqrt{B}\min_{j\in\mathcal{B}_r} A_j^{\mathsf{a}}(\sigma) \geq C_q^{\mathrm{syn}} s,$$

for a sufficiently large constant C_q^{syn}. Increasing precision compensates for weaker responses within a wider scale window, allowing a longer block to be generated by one query.

In the high precision interval there is a feasible integer block length $L_{B,\mathsf{a}}^{\mathrm{up}}$ with

$$L_{B,\mathsf{a}}^{\mathrm{up}} \asymp_{q,\kappa} \min\{d, 1+\log(1+\varepsilon\sqrt{B})\} = \ell_B \tag{23}$$

Proposition 10 and Appendix J give the family-specific precision thresholds and their inversion. With this choice,

$$\mathbb{E}T = Mp = O_{q,\kappa}\left(\alpha\sqrt{d/\ell_B}\right). \tag{24}$$

This matches the blockwise lower bound in query cost. The remaining step is to establish joint accuracy under shared activation, completing the high precision construction in Figure 1.

The identity $ps = \varepsilon$ preserves every coordinate marginal. Its effect on the joint law is seen by expanding the block densities relative to their product base and writing $u_j = \theta_j h_j^{\mathsf{a}}(x_j)$. The coefficient of u_j is ps in the output and ε in the target, while the coefficients of $u_i u_j$, for $i \neq j$, are ps^2 and ε^2. Hence

$$ps - \varepsilon = 0, \qquad ps^2 - \varepsilon^2 = p(1-p)s^2.$$

The remaining terms therefore involve at least two coordinates. The joint error estimate controls these interactions, including all higher-order terms.

Write $P_{s,\theta,r}^{\mathsf{a}} = \bigotimes_{j\in\mathcal{B}_r} P_{s,\theta_j,j}^{\mathsf{a}}$, $P_{0,r} = \bigotimes_{j\in\mathcal{B}_r} \mu_j^{\mathsf{a}}$, and, for fixed θ, $P_{\varepsilon,r}^{\mathsf{a}} = P_{\varepsilon,\theta,r}^{\mathsf{a}}$. For the block law $Q_r = (1-p)P_{0,r} + pP_{s,\theta,r}^{\mathsf{a}}$, the remaining interactions satisfy

$$\mathrm{TV}(Q_\theta, P_\theta^{\mathsf{a}})^2 \leq \sum_r \chi^2(Q_r \| P_{\varepsilon,r}^{\mathsf{a}}) \leq C_q\alpha^2/\Gamma^2. \tag{25}$$

Here $\chi^2(Q\|P) = \int (dQ/dP - 1)^2 dP$; the first inequality follows from multiplication of Hellinger affinities across independent blocks.

The activation probability keeps the total interaction strength within each block controlled as its size grows. Let $m_r = |\mathcal{B}_r|$. When $p < 1$, the choice $s = (\Gamma\sqrt{L})^{-1}$ gives $m_r s^2 \leq \Gamma^{-2}$. After the single-coordinate terms cancel, the remaining interactions satisfy

$$\chi^2(Q_r \| P_{\varepsilon,r}^{\mathsf{a}}) \leq C_q p^2 m_r^2 s^4 \leq \frac{C_q}{\Gamma^2}\varepsilon^2 m_r.$$

The residual error is therefore bounded in proportion to the number of coordinates in each block. Summing over $\sum_r m_r = d$ gives the global bound, so increasing the feasible block size preserves the required control of the joint output.

The general bound (142) also explains the activation probability. For any $0 < p < 1$ satisfying the shared-scale synthesis condition above, put $s = \varepsilon/p$. If $s \leq 1/4$ and $Ls^2 \leq 1$, it yields

$$\mathbb{E}T \asymp \frac{dp}{L}, \qquad \mathrm{TV}(Q_\theta, P_\theta^{\mathsf{a}})^2 \leq C_q \frac{\varepsilon^4 dL}{p^2}.$$

At fixed feasible L, increasing p raises query cost while reducing this error bound. Choosing $p = \Gamma\varepsilon\sqrt{L} < 1$ gives TV error at most $C_q\alpha/\Gamma$ and expected cost of order $\Gamma\alpha\sqrt{d/L}$. At high precision, the coordinate-thinning construction uses $\Theta_{q,\kappa}(\varepsilon d)$ calls (Appendix H). For feasible $L \asymp_{q,\kappa} \ell_B$, shared activation saves a factor of order $\sqrt{L}$ relative to this cost; the blockwise converse makes the resulting query rate optimal under joint TV accuracy. Appendix I derives the complete cancellation identity.

A fixed $\Gamma = \Gamma(q,\kappa)$ chosen sufficiently large ensures TV error below $\kappa\alpha$, uniformly in θ; when $p = 1$, the sampler is exact. All actions and activation decisions can be drawn before observing the oracle. A run with no activated block outputs the base law at zero cost, allowing the expected number of queries to fall below one.

5 EQUIVALENT TARGETS WITH DIFFERENT QUERY COSTS

The preceding bounds quantify the cost of jointly generating weak perturbations. We now vary their scale layout while preserving the full target experiment. Placing the same active signs in clustered and separated spatial layouts isolates how a shared smoothing scale affects this cost. The comparison fixes the ambient dimension, precision cap, and accuracy requirement.

For an active layout $S = \{j_1 < \cdots < j_m\} \subset [D]$, define

$$p^{\mathsf{S}}_{\theta,S}(x) = \prod_{r=1}^{m} \bar{p}_{j_r}(x_{j_r})\big(1 + \varepsilon\theta_r h^{\mathsf{S}}_{j_r}(x_{j_r})\big) \prod_{j \notin S} \bar{p}_j(x_j), \tag{26}$$

for $\theta \in \{\pm 1\}^m$. Let $T^*_{\mathsf{S}}(S; D, m, \varepsilon, q, B, \delta)$ denote the corresponding analog complexity in ambient dimension D, and define $T^*_{\mathsf{S},\mathrm{cmp}}(S; D, m, \varepsilon, q, B, \delta)$ analogously for the coordinatewise comparator oracle defined in Appendix L.

Theorem 3 (Query complexity for equivalent scale layouts)**.** *Fix $q > 1$ and $0 < \kappa < \kappa_0(q)$. There are positive constants $b_2(q,\kappa)$ and $C_{\mathrm{sep}}(q,\kappa)$ such that the following holds. For integers $m, G \geq 1$, set*

$$D = 1 + (m-1)G, \quad S_{\mathrm{cl}} = [m], \quad S_{\mathrm{sp}} = \{1, 1+G, \ldots, 1+(m-1)G\}, \tag{27}$$

$0 < \alpha \leq 1$, $\varepsilon = \alpha/\sqrt{m} \leq 1/4$, $\delta = \kappa\alpha$, and

$$w_B = 1 + \log(1 + \varepsilon\sqrt{B}), \qquad \ell_B^{(m)} = \min\{m, w_B\}. \tag{28}$$

If $B \geq b_2\varepsilon^{-2}$ and $G \geq C_{\mathrm{sep}} w_B$, then

$$T^*_{\mathsf{S}}(S_{\mathrm{cl}}; D, m, \varepsilon, q, B, \delta) \asymp_{q,\kappa} \alpha\sqrt{m/\ell_B^{(m)}}, \qquad T^*_{\mathsf{S}}(S_{\mathrm{sp}}; D, m, \varepsilon, q, B, \delta) \asymp_{q,\kappa} \alpha\sqrt{m}. \tag{29}$$

The same rates hold with T^*_{S} replaced by $T^*_{\mathsf{S},\mathrm{cmp}}$. Moreover, the target experiments $\{P^{\mathsf{S}}_{\theta,S_{\mathrm{cl}}}\}_\theta$ and $\{P^{\mathsf{S}}_{\theta,S_{\mathrm{sp}}}\}_\theta$ are Blackwell equivalent. If $w_B \to \infty$ along a sequence with $w_B \leq m$, then

$$\frac{T^*_{\mathsf{S}}(S_{\mathrm{sp}}; D, m, \varepsilon, q, B, \delta)}{T^*_{\mathsf{S}}(S_{\mathrm{cl}}; D, m, \varepsilon, q, B, \delta)} \asymp_{q,\kappa} \sqrt{w_B},$$

and the same ratio holds for the comparator complexities.

The equivalence follows from a common canonical experiment. Dividing each active coordinate by its public radius and discarding inactive coordinates produces independent factors

$$\nu_{\theta_r} = \frac{1+\varepsilon\theta_r}{2}\mathcal{N}(1,1) + \frac{1-\varepsilon\theta_r}{2}\mathcal{N}(-1,1), \qquad r \in [m].$$

Rescaling these factors and independently generating inactive coordinates reconstructs either target, giving exact parameter-independent kernels in both directions. These transformations preserve the target experiment. The observation geometry changes: normalizing coordinate j by r_j turns the common smoothing scale σ into σ/r_j, while each legal query must still use one scale across coordinates. Equivalence of target samples therefore leaves room for different query costs.

The cost difference is quantified by the response profiles on the separated grid. Uniformly over $\sigma > 0$ and $z \geq 0$,

$$\sum_{r=1}^{m} \min\{1, zA^{\mathsf{S}}_{1+(r-1)G}(\sigma)\} \leq C_q \min\left\{m, 1 + \frac{\log(1+z)}{G}\right\}.$$

The right side accounts for both the profiles above the truncation threshold and all weaker responses in the geometric tails. With $b_t \leq B$ and $G \geq C_{\mathrm{sep}} w_B$, each call contributes only $O_{q,\kappa}(1)$ to the summed face-separation budget. The fixed-parameter transcript argument then converts the output demand of order εm into the lower bound of order $\alpha\sqrt{m}$.

The common ambient dimension does not change this budget. Conditional on a public action, inactive score coordinates have a known law, including their Gaussian noise, and can be simulated internally. Accordingly, the converse applies to the entire D-dimensional response vector. Independent coordinate thinning attains the separated-layout bound, completing the comparison with the clustered block construction.

The spacing condition separates active scales beyond the width w_B that precision can compensate for. Clustering retains the block-sharing gain, while separation removes it even from the optimal adaptive protocol. Appendix K gives the proof, including the weak responses between active scales.

For a concrete instance, fix $\alpha = 1$, take $\varepsilon = m^{-1/2}$ and $B = m^3$, and set $G = \lceil C_{\mathrm{sep}}(q,\kappa)[1 + \log(1+m)]\rceil$. For sufficiently large m, all conditions of Theorem 3 hold, $w_B = 1 + \log(1+m)$, and both layouts have ambient dimension $D = 1 + (m-1)G \asymp_{q,\kappa} m \log m$. Writing T_{cl} and T_{sp} for their respective optimal query costs, we obtain

$$T_{\mathrm{cl}} \asymp_{q,\kappa} \sqrt{\frac{m}{\log m}}, \qquad T_{\mathrm{sp}} \asymp_{q,\kappa} \sqrt{m}.$$

Thus arranging the same statistical information across scales produces a $\sqrt{\log m}$ gap in optimal cost at the same precision and accuracy.

Threshold observations and cumulative precision. For the two d-dimensional families of Theorem 1, the same rates hold when each query returns one threshold bit per coordinate after Gaussian noise is added (Corollary 11, Appendix L). Under the same target and accuracy assumptions, let $N = \mathbb{E}T$ and $P = \mathbb{E}\sum_{t \le T} b_t$ for a legal δ-accurate protocol without a precision cap. Corollary 13 gives $N \gtrsim_{q,\kappa} \mathcal{R}_d(P/N)$ when $0 < N < \infty$ and $P < \infty$; a fixed precision selected for each $\lambda > 0$ attains the optimal order of $N + \lambda P$ (Appendix M). Appendix N analyzes four alternative Fourier interfaces admitting exact one-query simulation.

6 Related work

Xun & Price (2026) establish query lower bounds through scale search under global L^p score guarantees; Liu (2026) studies Gaussian sampling and adaptive finite-bit communication through channel synthesis. Diffusion analyses also quantify parallel depth, score approximation, and discretization (Zhou & Sugiyama, 2025; Chen et al., 2023b;a). We characterize expected calls across observation precisions under a shared-scale interface. Table 3 compares the models, resources, and methods.

Blackwell comparison and Le Cam deficiency formalize preservation of statistical experiments (Blackwell, 1953; Le Cam, 1964), while channel synthesis connects sampling to communication (Cuff, 2013; Harsha et al., 2010). Information constraints also yield stochastic-gradient sampling bounds and interactive estimation bounds (Chatterji et al., 2022; Chen et al., 2026; Castro, 2014; Acharya et al., 2023). Here, bounded witnesses translate joint sampling accuracy into blockwise requirements matched to score sensitivity, and shared activation attains the resulting query bounds.

Representation and forward-noise geometry offer complementary ways to shape generation. Guth et al. (2022) use normalized wavelet conditionals to improve score conditioning and sampling efficiency. Scimeca et al. (2025) shape forward covariance in the Fourier basis to alter diffusion paths and spectral inductive biases. Our layout comparison isolates the effect of scale arrangement on optimal query cost while preserving the entire target experiment. Our lower bounds cover every legal accurate protocol, and explicit nonadaptive constructions attain their order, making the separation independent of a particular sampling trajectory or solver.

7 Conclusion

Under the precision and spacing conditions of Theorem 3, our spatial constructions preserve the complete target experiment while allowing the ratio of optimal query costs to diverge as the effective block length grows. Each comparison fixes the ambient dimension, oracle interface, precision cap, and accuracy requirement. For the Fourier and spatial families in Theorem 1, the feasible block length has order ℓ_B in the high precision regime, with ℓ_B capped at d. At the prescribed joint TV accuracy, nonadaptive block activation uses a factor of order $\sqrt{\ell_B}$ fewer expected queries than the coordinate thinning construction and attains the optimal order. Adaptive queries and random stopping cannot improve this order.

AI USE STATEMENT

AI tools assisted with wording and proofreading of the manuscript.

REFERENCES

Jayadev Acharya, Clément L. Canonne, Ziteng Sun, and Himanshu Tyagi. Unified lower bounds for interactive high-dimensional estimation under information constraints. In *Advances in Neural Information Processing Systems*, volume 36, pp. 51133–51165, 2023. doi: 10.52202/075280-2226. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/a07e87ecfa8a651d62257571669b0150-Abstract-Conference.html.

Ben Adcock, Anders C. Hansen, Clarice Poon, and Bogdan Roman. Breaking the coherence barrier: A new theory for compressed sensing. *Forum of Mathematics, Sigma*, 5:e4, 2017. doi: 10.1017/fms.2016.32. URL https://doi.org/10.1017/fms.2016.32.

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272, 1953. doi: 10.1214/aoms/1177729032. URL https://doi.org/10.1214/aoms/1177729032.

Rui M. Castro. Adaptive sensing performance lower bounds for sparse signal detection and support estimation. *Bernoulli*, 20(4):2217–2246, 2014. doi: 10.3150/13-BEJ555. URL https://doi.org/10.3150/13-BEJ555.

Rui M. Castro and Ervin Tánczos. Adaptive compressed sensing for support recovery of structured sparse sets. *IEEE Transactions on Information Theory*, 63(3):1535–1554, 2017. doi: 10.1109/TIT.2017.2653802. URL https://doi.org/10.1109/TIT.2017.2653802.

Niladri S. Chatterji, Peter L. Bartlett, and Philip M. Long. Oracle lower bounds for stochastic gradient sampling algorithms. *Bernoulli*, 28(2):1074–1092, 2022. doi: 10.3150/21-BEJ1377. URL https://doi.org/10.3150/21-BEJ1377.

Fan Chen, Sinho Chewi, Constantinos Daskalakis, and Alexander Rakhlin. High-accuracy log-concave sampling with stochastic queries. In Steve Hanneke and Tor Lattimore (eds.), *Proceedings of the Thirty-Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 1341–1372. PMLR, 2026. URL https://proceedings.mlr.press/v336/chen26g.html.

Hongrui Chen, Holden Lee, and Jianfeng Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 4735–4763. PMLR, 2023a. URL https://proceedings.mlr.press/v202/chen23q.html.

Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R. Zhang. Sampling is as easy as learning the score: Theory for diffusion models with minimal data assumptions. In *The Eleventh International Conference on Learning Representations*. OpenReview.net, 2023b. URL https://openreview.net/forum?id=zyLVMgsZ0U_.

Paul Cuff. Distributed channel synthesis. *IEEE Transactions on Information Theory*, 59(11):7071–7096, 2013. doi: 10.1109/TIT.2013.2279330. URL https://doi.org/10.1109/TIT.2013.2279330.

Anming Gu, Kevin Tian, Hubert Yang, and Yusong Zhu. The tractability landscape of sampling with inexact scores, 2026. URL https://arxiv.org/abs/2607.19004.

Florentin Guth, Simon Coste, Valentin De Bortoli, and Stéphane Mallat. Wavelet score-based generative modeling. In *Advances in Neural Information Processing Systems*, volume 35, pp. 478–491. Curran Associates, Inc., 2022. doi: 10.52202/068431-0035. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/03474669b759f6d38cdca6fb4eb905f4-Abstract-Conference.html.

Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. *IEEE Transactions on Information Theory*, 56(1):438–449, 2010. doi: 10.1109/TIT.2009.2034824. URL https://doi.org/10.1109/TIT.2009.2034824.

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

Felix Krahmer and Rachel Ward. Stable and robust sampling strategies for compressive imaging. *IEEE Transactions on Image Processing*, 23(2):612–622, 2014. doi: 10.1109/TIP.2013.2288004. URL https://doi.org/10.1109/TIP.2013.2288004.

Lucien Le Cam. Sufficiency and approximate sufficiency. *The Annals of Mathematical Statistics*, 35 (4):1419–1455, 1964. doi: 10.1214/aoms/1177700372. URL https://doi.org/10.1214/aoms/1177700372.

Jingbo Liu. Smoothed score queries and the complexity of sampling, 2026. URL https://arxiv.org/abs/2605.27769.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural Information Processing Systems*, volume 35, pp. 5775–5787. Curran Associates, Inc., 2022. doi: 10.52202/068431-0418. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/260a14acce2a89dad36adc8eefe7c59e-Abstract-Conference.html.

Sirin Nitinawarat, George K. Atia, and Venugopal V. Veeravalli. Controlled sensing for multihypothesis testing. *IEEE Transactions on Automatic Control*, 58(10):2451–2464, 2013. doi: 10.1109/TAC.2013.2261188. URL https://doi.org/10.1109/TAC.2013.2261188.

Luca Scimeca, Thomas Jiralerspong, Berton Earnshaw, Jason Hartford, and Yoshua Bengio. Learning what matters: Steering diffusion via spectrally anisotropic forward noise, 2025. URL https://arxiv.org/abs/2510.09660.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011. doi: 10.1162/NECO_a_00142. URL https://doi.org/10.1162/NECO_a_00142.

Zhiyang Xun and Eric Price. Query lower bounds for diffusion sampling, 2026. URL https://arxiv.org/abs/2604.10857.

Huanjian Zhou and Masashi Sugiyama. Parallel simulation for log-concave sampling and score-based diffusion models. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 79192–79225. PMLR, 2025. URL https://proceedings.mlr.press/v267/zhou25x.html.

The proofs connect the dependence required by an accurate sample to the dependence available through a shared-scale observation. Appendices A–G establish the score profiles, target requirements, and adaptive transcript bounds. Appendices H–J construct samplers that attain the resulting rates, including the joint error control needed for shared activation. Appendix K applies these bounds to Blackwell-equivalent layouts, keeping the statistical content fixed while separating their optimal query costs. The remaining appendices treat threshold observations, precision costs, alternative interfaces, and related models.

For calculations concerning the Fourier family, we abbreviate $h_j = h_j^{\mathsf{F}}$, $p_\theta = p_\theta^{\mathsf{F}}$, and $A_j = A_j^{\mathsf{F}}$.

A Exact formulas for targets and smoothed scores

For $\eta \in \{\pm 1\}^{d-1}$, let $\theta^{j,+}(\eta)$ and $\theta^{j,-}(\eta)$ be the vectors in $\{\pm 1\}^d$ with jth coordinates $+1$ and -1, respectively, and common complementary coordinates η. Define the normalized mean gap

$$D_j^{\mathsf{a}}(x, \sigma; \eta) := \sigma s^{\mathsf{a}}_{\theta^{j,+}(\eta),\sigma,j}(x) - \sigma s^{\mathsf{a}}_{\theta^{j,-}(\eta),\sigma,j}(x).$$

For the product families above this quantity is independent of η, so we write it as $D_j^{\mathsf{a}}(x, \sigma)$.

We first record the algebra behind Lemma 8. Write $m_j = e^{-k_j^2/2}$, $\tau^2 = 1 + \sigma^2$, and

$$a_{j,\sigma} = \exp\left\{-\frac{k_j^2\sigma^2}{2\tau^2}\right\}, \quad \omega_j(\sigma) = \frac{k_j}{\tau^2}, \quad H_{j,\sigma}(x) = a_{j,\sigma}\cos(\omega_j(\sigma)x) - m_j. \tag{30}$$

The Gaussian characteristic function gives

$$\big[\phi(\cdot)\cos(k_j\cdot)\big] * \phi_{\sigma^2}(x) = \phi_{\tau^2}(x) a_{j,\sigma}\cos(\omega_j(\sigma)x). \tag{31}$$

Consequently the smoothed target remains a coordinate product:

$$p_{\theta,\sigma}(x) = \prod_{j=1}^{d} \phi_{\tau^2}(x_j)\big(1 + \varepsilon\theta_j H_{j,\sigma}(x_j)\big). \tag{32}$$

Differentiation gives the exact normalized score

$$\sigma s_{\theta,\sigma,j}(x) = -\frac{\sigma x_j}{\tau^2} - \frac{\varepsilon\theta_j A_j(\sigma)\sin(\omega_j(\sigma)x_j)}{1 + \varepsilon\theta_j H_{j,\sigma}(x_j)}, \tag{33}$$

where A_j is defined in (14). Thus flipping only θ_j changes only the jth mean, and the absolute gap is exactly

$$|D_j(x,\sigma)| = \frac{2\varepsilon A_j(\sigma)|\sin(\omega_j(\sigma)x_j)|}{1 - \varepsilon^2 H_{j,\sigma}(x_j)^2}. \tag{34}$$

Since $|H_{j,\sigma}| \le a_{j,\sigma} + m_j \le 2$ and $\varepsilon \le 1/4$, $1 - \varepsilon^2 H_{j,\sigma}^2 \ge 3/4$. This proves (74); it also shows that the gap is at least $2\varepsilon A_j|\sin(\omega_j x_j)|$.

For every j and σ, the algorithm can choose x_j so that $|\sin(\omega_j x_j)| = 1$. The midpoint between the two score means is known from $(j, x_j, \sigma, \varepsilon)$ and does not require knowing θ_j. Explicitly,

$$\bar{\mu}_j(x,\sigma) = -\frac{\sigma x_j}{\tau^2} + \frac{\varepsilon^2 A_j(\sigma) H_{j,\sigma}(x_j)\sin(\omega_j(\sigma)x_j)}{1 - \varepsilon^2 H_{j,\sigma}(x_j)^2}. \tag{35}$$

Let Φ denote the standard normal distribution function. Thresholding $N(\mu_{\theta_j}, b^{-1})$ at this midpoint yields a BSC with correlation

$$\rho_j = 2\Phi\left(\frac{\sqrt{b}|D_j|}{2}\right) - 1. \tag{36}$$

Multiplication by an independent random sign of prescribed mean realizes every weaker correlation in $[0, \rho_j]$. The sign of the mean gap is public, so the algorithm complements the threshold bit when necessary. A coordinate with zero sine phase yields a fair parameter-independent bit and is discarded.

A.1 SCORE OF THE SPATIAL MIXTURE

Coordinate j of the spatial target is equivalently

$$\frac{1+\varepsilon\theta_j}{2}\mathcal{N}(r_j, r_j^2) + \frac{1-\varepsilon\theta_j}{2}\mathcal{N}(-r_j, r_j^2). \tag{37}$$

Fix one spatial scale $r > 0$ and put $\tau^2 = r^2 + \sigma^2$. Convolution of the two mixture components in (37) gives

$$p^{\mathsf{S}}_{\theta,r,\sigma}(x) = \frac{1+\varepsilon\theta}{2}\phi_{\tau^2}(x-r) + \frac{1-\varepsilon\theta}{2}\phi_{\tau^2}(x+r). \tag{38}$$

Writing $t = \tanh(rx/\tau^2)$, direct differentiation yields

$$\sigma s^{\mathsf{S}}_{\theta,r,\sigma}(x) = -\frac{\sigma x}{\tau^2} + \frac{r\sigma}{\tau^2}\frac{t+\varepsilon\theta}{1+\varepsilon\theta t}. \tag{39}$$

The absolute flip gap is therefore exactly

$$|D^{\mathsf{S}}_r(x,\sigma)| = 2\varepsilon A^{\mathsf{S}}_r(\sigma)\frac{1-t^2}{1-\varepsilon^2 t^2}, \qquad A^{\mathsf{S}}_r(\sigma) = \frac{r\sigma}{r^2+\sigma^2}. \tag{40}$$

Since $\varepsilon \le 1/4$, this is at most $(32/15)\varepsilon A^{\mathsf{S}}_r < (8/3)\varepsilon A^{\mathsf{S}}_r$, uniformly in x. At $x = 0$, equality in the profile factor is attained:

$$\sigma s^{\mathsf{S}}_{\theta,r,\sigma}(0) = \varepsilon\theta A^{\mathsf{S}}_r(\sigma). \tag{41}$$

Thus threshold zero gives a BSC of correlation $2\Phi(\varepsilon A^{\mathsf{S}}_r(\sigma)\sqrt{b}) - 1$. The profile depends only on σ/r. This scale covariance uses component variance r^2; replacing it by a fixed variance destroys the localized profile.

B TARGET OVERLAP, CAPACITY, AND RECOVERY

$$C(P^{\mathsf{a}}) := \sup_{\Pi} I_{\Theta\sim\Pi}(\Theta; X), \qquad X \mid \Theta = \theta \sim P^{\mathsf{a}}_\theta, \tag{42}$$

where the supremum is over all priors on $\{\pm 1\}^d$.

Proposition 4 (Capacity and recovery of all parameters from one target sample)**.** *There are constants $0 < c_q < C_q < \infty$ such that*

$$c_q\alpha^2 \le C(P^{\mathsf{a}}) \le C_q\alpha^2, \qquad \mathsf{a} \in \{\mathsf{F}, \mathsf{S}\}. \tag{43}$$

Under the uniform prior,

$$\sup_{\widehat{\Theta}:\,\mathbb{R}^d\to\{\pm1\}^d \text{ measurable}} \mathbb{P}\{\widehat{\Theta}(X) = \Theta\} \le 2^{-d}\exp(C_q\varepsilon d) = \exp\{-d\log 2 + O_q(\alpha\sqrt{d})\}. \tag{44}$$

We prove Proposition 4 with explicit constants. Gaussian moments give

$$\mathbb{E}_\phi h_j = 0, \qquad v_j := \mathbb{E}_\phi h_j^2 = \frac{(1-e^{-k_j^2})^2}{2}, \qquad m_j^{\mathrm{abs}} := \mathbb{E}_\phi|h_j| \ge \frac{v_j}{2}. \tag{45}$$

The last inequality follows from $\mathbb{E}h_j^2 \le \|h_j\|_\infty \mathbb{E}|h_j|$ and $\|h_j\|_\infty < 2$. In particular

$$v_j \ge \frac{(1-e^{-q^2})^2}{2}, \qquad m_j^{\mathrm{abs}} \ge \frac{(1-e^{-q^2})^2}{4}. \tag{46}$$

Every coordinate factor integrates to one and is at least $1 - 2\varepsilon \ge 1/2$, proving normalization and common support.

Under the uniform prior, the output marginal is $\phi^{\otimes d}$ and the channel factors by coordinate. With

$$g(u) = \frac{(1+u)\log(1+u) + (1-u)\log(1-u)}{2}, \tag{47}$$

we have

$$I_{\mathrm{unif}}(\Theta; X) = \sum_{j=1}^{d} \mathbb{E}_\phi g(\varepsilon h_j(G)). \tag{48}$$

Now $g(0) = g'(0) = 0$ and $g''(u) = (1-u^2)^{-1}$. Since $|\varepsilon h_j| \le 1/2$,

$$\frac{u^2}{2} \le g(u) \le \frac{2u^2}{3} \qquad (|u| \le 1/2). \tag{49}$$

It follows that

$$I_{\mathrm{unif}}(\Theta; X) \ge \frac{(1-e^{-q^2})^2}{4}\alpha^2. \tag{50}$$

For an arbitrary prior Π, use the Gaussian product base $P_0 = \phi^{\otimes d}$. The KL-radius identity, coordinate additivity, and $D_{\mathrm{KL}}(P_{\theta_j}\|\phi) \le \chi^2(P_{\theta_j}\|\phi) = \varepsilon^2 v_j$ imply

$$I_\Pi(\Theta; X) \le \mathbb{E}_{\Theta\sim\Pi} D_{\mathrm{KL}}(P_\Theta \| P_0) \tag{51}$$

$$\le \varepsilon^2 \sum_{j=1}^{d} v_j \le \frac{1}{2}\alpha^2. \tag{52}$$

Together, (50) and (52) prove (43) for the Fourier family.

For the spatial family, the rescaling $Y = X/r_j$ maps every base factor $\bar{p}_j$ to the same symmetric law

$$\bar{\mu} = \tfrac{1}{2}\mathcal{N}(1,1) + \tfrac{1}{2}\mathcal{N}(-1,1), \qquad h^{\mathsf{S}}(Y) = \tanh Y. \tag{53}$$

Put $v_{\mathsf{S}} = \mathbb{E}_{\bar{\mu}} \tanh^2 Y > 0$ and $m_{\mathsf{S}} = \mathbb{E}_{\bar{\mu}} |\tanh Y| > 0$. The uniform output marginal is the product base, so the same calculation with h_j replaced by $\tanh Y$ gives

$$\frac{1}{2} v_{\mathsf{S}} \alpha^2 \le I_{\mathrm{unif}}(\Theta; X) \le C(P^{\mathsf{S}}) \le v_{\mathsf{S}} \alpha^2. \tag{54}$$

The uniform posterior factors coordinatewise. The optimal success probability for bit j is $(1 + \varepsilon m_j^{\mathrm{abs}})/2$, so the exact full-vector MAP success is

$$\sup_{\widehat{\Theta}} \mathbb{P}\{\widehat{\Theta}(X) = \Theta\} = 2^{-d} \prod_{j=1}^{d} (1 + \varepsilon m_j^{\mathrm{abs}}) \le 2^{-d} \exp\left(\alpha\sqrt{d/2}\right), \tag{55}$$

where $m_j^{\mathrm{abs}} \le \sqrt{v_j} \le 1/\sqrt{2}$.

For the spatial family, the exact uniform-prior MAP success is

$$2^{-d}(1 + \varepsilon m_{\mathsf{S}})^d \le 2^{-d} \exp(m_{\mathsf{S}} \varepsilon d). \tag{56}$$

This proves the recovery statement for both families.

The targets also have quantitative overlap. If $m = d_H(\theta, \theta')$, then

$$D_{\mathrm{KL}}(P_\theta^{\mathsf{a}} \| P_{\theta'}^{\mathsf{a}}) \le 4\varepsilon^2 m \le 4\alpha^2. \tag{57}$$

This follows coordinatewise from $\mathbb{E}h_j = 0$, $|\varepsilon h_j| \le 1/2$, and the standard quadratic bound on the log likelihood ratio. Moreover, coordinatewise Hellinger affinity and $\sqrt{1-u} \ge \exp(-2u/3)$ for $0 \le u \le 1/4$ give

$$\int \sqrt{dP_\theta^{\mathsf{a}} dP_{\theta'}^{\mathsf{a}}} \ge e^{-2\varepsilon^2 m/3} \ge e^{-2\alpha^2/3}. \tag{58}$$

C Bounded test functions and coordinatewise sensitivity

For a channel Q, the coordinate and block sensitivities are defined in (11)–(12). For the target itself,

$$\bar{v}_j(P^{\mathsf{a}}) = \varepsilon \mathbb{E}_{\mu_j^{\mathsf{a}}} |h_j^{\mathsf{a}}|, \qquad \sum_{j=1}^{d} \bar{v}_j(P^{\mathsf{a}})^2 = \Theta_q(\alpha^2). \tag{59}$$

Lemma 5 (TV accuracy implies coordinatewise and blockwise separation)**.** *For every fixed $q > 1$, there are constants $c_{\mathrm{f}}(q), c_{\mathrm{b}}(q) > 0$ such that the following holds. Suppose $2^{-d}\sum_\theta \mathrm{TV}(Q_\theta, P_\theta^{\mathsf{a}}) \le \delta$ for $\mathsf{a} \in \{\mathsf{F}, \mathsf{S}\}$. Then:*

(i) *For every nonempty $S \subseteq [d]$ with $n = |S|$ and $\varepsilon\sqrt{n} \le 1$,*

$$\sum_{j\in S} \bar{v}_j(Q) \ge \sqrt{n}\,[c_{\mathrm{f}}(q)\varepsilon\sqrt{n} - 2\delta]_+. \tag{60}$$

(ii) *For $L \in [d]$, let $M = \lfloor d/L \rfloor$ and $\mathcal{B}_r = \{(r-1)L + 1, \dots, rL\}$ for $r \in [M]$, and put $\alpha_L = \varepsilon\sqrt{ML}$. Then*

$$\sum_{r=1}^{M} J_r(Q) \ge \frac{\sqrt{M}}{4}[c_{\mathrm{b}}(q)\alpha_L - 2\delta]_+. \tag{61}$$

We prove part (i) below; the proof of part (ii) is in Appendix F.

For nonempty $S \subseteq [d]$, put $n = |S|$ and take $\mathrm{sign}(0) = 0$. The coordinate witness is

$$F_S(\theta, x) = \tanh\left(n^{-1/2}\sum_{j\in S}\theta_j \,\mathrm{sign}\, h_j^{\mathsf{a}}(x_j)\right). \tag{62}$$

For $S = [d]$, this is the coordinate witness F_{c} used in Section 3.

Average conditional TV error is exactly joint TV when the parameter marginal is fixed:

$$\mathrm{TV}\big(2^{-d}\sum_\theta \delta_\theta Q_\theta, 2^{-d}\sum_\theta \delta_\theta P_\theta^{\mathsf{a}}\big) = 2^{-d}\sum_\theta \mathrm{TV}(Q_\theta, P_\theta^{\mathsf{a}}). \tag{63}$$

Thus a witness bounded by one loses at most 2δ in expectation.

We use the following elementary estimate for biased Rademacher variables.

Lemma 6 (A tanh statistic of biased Rademacher variables)**.** *Let $n \ge 1$, $0 < a \le 1$, and $0 < \varepsilon \le n^{-1/2}$. Let $Z_1, \dots, Z_n$ be independent $\{\pm 1\}$-valued random variables with $a\varepsilon \le \mathbb{E}Z_j = \beta_j \le \varepsilon$. Then there is a universal constant $c > 0$ such that*

$$\mathbb{E}\tanh\left(n^{-1/2}\sum_{j=1}^{n} Z_j\right) \ge c\,a\varepsilon\sqrt{n}. \tag{64}$$

Proof. Interpolate independent $\{\pm 1\}$-valued variables with means $t\beta_j$, $t \in [0, 1]$, and call the expectation $f(t)$. For the jth summand, set $W_{-j} = \sum_{i\ne j} Z_i(t)$. Differentiating the finite product law,

$$f'(t) = \sum_{j=1}^{n}\frac{\beta_j}{2}\mathbb{E}\left[\tanh\left(\frac{W_{-j}+1}{\sqrt{n}}\right) - \tanh\left(\frac{W_{-j}-1}{\sqrt{n}}\right)\right]. \tag{65}$$

Uniformly in t and j, $\mathbb{E}W_{-j}^2 \le n + (\sum_i \beta_i)^2 \le 2n$. Hence $\mathbb{P}(|W_{-j}| \le 2\sqrt{n}) \ge 1/2$. On this event, the bracket in (65) is at least $2\,\mathrm{sech}^2(3)/\sqrt{n}$. Therefore $f'(t) \ge cn^{-1/2}\sum_j \beta_j \ge ca\varepsilon\sqrt{n}$. Since $f(0) = 0$, integration proves the claim. □

Under the ideal joint experiment, the variables

$$Z_j = \Theta_j \,\mathrm{sign}\, h_j(X_j) \tag{66}$$

are independent Rademachers and $\mathbb{E}Z_j = \varepsilon m_j^{\mathrm{abs}} \ge a_q\varepsilon$ by (46). Lemma 6 therefore gives

$$\mathbb{E}_P F_S(\Theta, X) \ge c_{\mathrm{f}}(q)\varepsilon\sqrt{n}. \tag{67}$$

For the spatial family, replace h_j by h_j^{S} and use $\mathbb{E}|h_j^{\mathsf{S}}| = m_{\mathsf{S}} > 0$ from (53); the same bound holds with an absolute constant. Every subsequent symmetrization step is measure-theoretic and unchanged.

It remains to relate F_S to coordinatewise TV sensitivity for an arbitrary output channel. For a fixed complementary parameter η, replacing the two j-face laws by their average changes the expectation by the exact signed-measure term

$$\frac{1}{4}\int (F_+ - F_-)\, d(Q_{+,\eta} - Q_{-,\eta}). \tag{68}$$

Because $\tanh$ is one-Lipschitz, $\|F_+ - F_-\|_\infty \le 2/\sqrt{n}$; hence the absolute value of (68) is at most $\mathrm{TV}(Q_{+,\eta}, Q_{-,\eta})/\sqrt{n}$. Sequentially average all bits in S. Previously averaged bits can only decrease the remaining coordinatewise TV sensitivity by convexity. The fully averaged channel has zero witness expectation by the pairing $\theta_S \leftrightarrow -\theta_S$. Thus

$$|\mathbb{E}_Q F_S| \le \frac{1}{\sqrt{n}} \sum_{j\in S} \bar{v}_j(Q). \tag{69}$$

Combining (63), (67), and (69) proves (60) and completes the proof of Lemma 5(i) for both families.

D Adaptive transcripts with random stopping

The score bounds must remain valid when the algorithm selects its next query using previous responses. We establish this by comparing transcript laws under two fixed parameters. The transcript includes the random seed, predictable actions, responses, and stopping indicators. After stopping, it records a common cemetery symbol. At any common realized history, both response laws use the same action, so the score bounds apply to their comparison.

Lemma 7 (TV hybrid bound and KL chain rule for stopped adaptive transcripts)**.** *Consider a legal adaptive protocol, and let* $\mathsf{H}_\theta := \mathrm{Law}_\theta(H_T)$*. For fixed parameters* θ, θ'*, let* $K^h_{\theta,t}$ *and* $K^h_{\theta',t}$ *be the active response kernels at the same realized history* h*. Then*

$$\mathrm{TV}(\mathsf{H}_\theta, \mathsf{H}_{\theta'}) \le \mathbb{E}_\theta \sum_{t\le T} \mathrm{TV}(K^{H_{t-1}}_{\theta,t}, K^{H_{t-1}}_{\theta',t}), \tag{70}$$

$$D_{\mathrm{KL}}(\mathsf{H}_\theta \| \mathsf{H}_{\theta'}) = \mathbb{E}_\theta \sum_{t\le T} D_{\mathrm{KL}}(K^{H_{t-1}}_{\theta,t} \| K^{H_{t-1}}_{\theta',t}). \tag{71}$$

The second equality is interpreted in $[0,\infty]$.

We prove Lemma 7. Put the parameter-independent random seed, predictable actions, observations, and stopping indicator into a canonical sequence; after stopping, every coordinate is a common cemetery symbol. For a deterministic horizon N, let $\mathsf{H}_{\theta,N} := \mathrm{Law}_\theta(H_N)$. The ordinary hybrid inequality gives

$$\mathrm{TV}(\mathsf{H}_{\theta,N}, \mathsf{H}_{\theta',N}) \le \mathbb{E}_\theta \sum_{t=1}^{N} \mathbf{1}\{t \le T\}\, \mathrm{TV}(K^{H_{t-1}}_{\theta,t}, K^{H_{t-1}}_{\theta',t}). \tag{72}$$

Indeed, replace one active response kernel at a time while holding the same realized history and predictable action fixed. Likewise, the conditional KL chain rule gives

$$D_{\mathrm{KL}}(\mathsf{H}_{\theta,N} \| \mathsf{H}_{\theta',N}) = \mathbb{E}_\theta \sum_{t=1}^{N} \mathbf{1}\{t \le T\} D_{\mathrm{KL}}(K^{H_{t-1}}_{\theta,t} \| K^{H_{t-1}}_{\theta',t}). \tag{73}$$

Let $\mathcal{F}_N$ be the sigma-field generated by the first N canonical coordinates and $\mathcal{F}_N \uparrow \mathcal{F}_\infty$. Relative to the mixture of the two full transcript laws, $R = (\mathsf{H}_\theta + \mathsf{H}_{\theta'})/2$, the projected likelihood ratios are uniformly integrable martingales and converge in $L^1(R)$ to the full likelihood ratios. Hence the TV distances of the projections increase to the TV of the complete transcripts. Relative entropy is the supremum of relative entropies over increasing finite-coordinate sub-sigma-fields, by its variational representation. Letting $N \to \infty$ in (72)–(73) and using Tonelli/monotone convergence on the nonnegative costs proves (70)–(71). Predictability and $b_t < \infty$ ensure that, at a common history, the two Gaussian kernels have the same positive-definite covariance.

E GEOMETRIC SCALE ENVELOPES

Lemma 8 (Bounds on score differences for the Fourier and spatial families)**.** *For every fixed $q > 1$, there are constants $C_q < \infty$ and $a_q > 0$ such that the following holds. Flipping θ_j changes only coordinate j of the response mean, and*

$$|D_j^{\mathsf{a}}(x, \sigma)| \le \frac{8}{3}\varepsilon A_j^{\mathsf{a}}(\sigma). \tag{74}$$

Moreover,

$$\sup_{\sigma>0} \sum_{j=1}^{d} A_j^{\mathsf{a}}(\sigma)^2 \le C_q, \tag{75}$$

$$\sup_{\sigma>0} \sum_{j=1}^{d} \min\{1, zA_j^{\mathsf{a}}(\sigma)\} \le C_q \log(1+z) \qquad \textit{for every } z \ge 0. \tag{76}$$

For $L \in [d]$, let $M = \lfloor d/L \rfloor$ and $\mathcal{B}_r = \{(r-1)L + 1, \dots, rL\}$ for $r \in [M]$, and define $E_r^{\mathsf{a}}(\sigma) = (\sum_{j\in\mathcal{B}_r} A_j^{\mathsf{a}}(\sigma)^2)^{1/2}$. Then

$$\sup_{\sigma>0} \sum_{r=1}^{M} \min\{1, zE_r^{\mathsf{a}}(\sigma)\} \le \min\left\{M, C_q\left(1 + \frac{\log(1+z)}{L}\right)\right\} \qquad \textit{for every } z \ge 0. \tag{77}$$

For every contiguous block $\mathcal{B} \subseteq [d]$ of length L,

$$\sup_{\sigma>0} \min_{j\in\mathcal{B}} A_j^{\mathsf{F}}(\sigma) \ge a_q \sqrt{L}\, q^{-L}, \qquad \sup_{\sigma>0} \min_{j\in\mathcal{B}} A_j^{\mathsf{S}}(\sigma) \ge a_q q^{-L/2}. \tag{78}$$

Let

$$\lambda = \frac{\sigma}{\sqrt{1+\sigma^2}} \in (0,1), \qquad g(u) = ue^{-u^2/2}. \tag{79}$$

Then

$$A_j(\sigma) = \sqrt{1-\lambda^2}\, g(k_j\lambda) \le g(k_j\lambda). \tag{80}$$

We repeatedly use the geometric-grid estimate: for a nonnegative unimodal f,

$$\sum_{m\in\mathbb{Z}} f(aq^m) \le 2\sup_{u>0} f(u) + \frac{1}{\log q}\int_0^\infty f(u)\frac{du}{u}. \tag{81}$$

To prove it, split the grid at a maximizer. On either monotone side, compare each rectangle in logarithmic coordinates with its adjacent integral; at most one endpoint on each side costs $\sup f$.

Taking $f = g^2$ in (81) yields

$$\sum_j A_j^2 \le \frac{2}{e} + \frac{1}{2\log q} =: C_{2,q}. \tag{82}$$

Taking $f(u) = \min\{1, zg(u)\}$ gives, for all $z \ge 0$,

$$\sum_j \min\{1, zA_j\} \le C_q \log(1+z). \tag{83}$$

For $0 \le z < \sqrt{e}$, one has $zg(u) < 1$ for all u, and hence the left side is at most $z\sum_j A_j = O_q(z) = O_q(\log(1+z))$. For $z \ge \sqrt{e}$, split at the two solutions of $zg(u) = 1$ (counting the double solution at equality). The saturated logarithmic interval contains $O_q(1 + \log z)$ grid points and both unsaturated tails sum geometrically.

We need two block forms. For any contiguous length-L block $\mathcal{B}$, choose λ so that the largest $u_j = k_j\lambda$ in the block is $\sqrt{2(L-1)\log q}$ when $L \ge 2$; for $L = 1$, take $u_j = 1$. Since g is unimodal, its minimum on the block occurs at an endpoint. Balancing the linear low-frequency endpoint with the Gaussian high-frequency endpoint gives

$$\sup_{\sigma>0} \min_{j\in\mathcal{B}} A_j(\sigma) \ge a_q \sqrt{L}\, q^{-L}. \tag{84}$$

For $L \geq 2$, if the final index of the block is $m \geq L$, then

$$\lambda = \frac{\sqrt{2(L-1)\log q}}{q^m} \leq \frac{\sqrt{2(L-1)\log q}}{q^L} \leq \frac{1}{q\sqrt{e}}. \tag{85}$$

For $L = 1$, the choice $u_j = 1$ gives $\lambda = 1/k_j \leq 1/q$. Thus the factor $\sqrt{1-\lambda^2}$ is bounded below by a positive q-dependent constant in both cases.

For the upper block envelope, set $E_r = (\sum_{j\in\mathcal{B}_r} A_j^2)^{1/2}$. Here the grid is restricted to the retained coordinates $j \in [ML]$. Let r_0 be the block containing the last grid point with $k_j\lambda \leq 1$; if no such block exists, take the first block, and if all grid points lie below one, take the last. The norm of r_0 and either adjacent crossing block are absorbed into C_q. If a block lies to the low-frequency side, $g(u) \leq u$ and geometric summation within the block gives $E_r \leq C_q u_{r,+}$, where $u_{r,+}$ is its largest $k_j\lambda$. Moving one block farther left multiplies this bound by q^{-L}. If a block lies to the high-frequency side, its first term dominates up to C_q, since for $u \geq 1$,

$$\frac{g(q^m u)}{g(u)} = q^m \exp\{-\tfrac{1}{2}(q^{2m}-1)u^2\}. \tag{86}$$

Moving one block farther right therefore contracts its norm by at most ρ_q^L, where

$$q^{2L} - 1 \geq L(q^2-1), \qquad \rho_q := q\exp\{-(q^2-1)/2\} < 1; \tag{87}$$

after the first such block the decay is super-exponential. It follows that the set $\{r : zE_r \geq 1\}$ contains at most $C_q(1+\log(1+z)/L)$ blocks: on the low side this is the number of factors q^L needed to grow from $1/z$ to one, and the high side is no larger. The two unsaturated tails have total $zE_r = O_q(1)$ by the geometric ratios just proved. More explicitly, after reindexing distance from r_0, the two tails are bounded, for every existing $m \geq 1$, by

$$E_{r_0-m} \leq C_q q^{-L(m-1)}, \qquad E_{r_0+m} \leq C_q \rho_q^{L(m-1)}, \tag{88}$$

so each saturation inequality allows at most $O_q(1+\log(1+z)/L)$ integers m, while the remaining geometric series, multiplied by z, are $O_q(1)$. This establishes

$$\sum_r \min\{1, zE_r\} \leq \min\left\{M, C_q\left(1 + \frac{\log(1+z)}{L}\right)\right\}. \tag{89}$$

The displayed tails are asserted only for existing blocks on the corresponding side of r_0; a missing side contributes zero. Hence the same argument applies at either finite-grid endpoint, since deleting grid points can only decrease the left side.

E.1 Spatial profile

For the spatial family, put $u_j = \sigma/r_j$. Then

$$A_j^{\mathsf{S}}(\sigma) = a(u_j), \qquad a(u) = \frac{u}{1+u^2} = a(u^{-1}), \qquad u_{j+1} = q^{-1}u_j. \tag{90}$$

The function a is unimodal and

$$\int_0^\infty a(u)^2 \frac{du}{u} = \frac{1}{2}, \qquad \int_0^\infty a(u)\frac{du}{u} = \frac{\pi}{2}. \tag{91}$$

The geometric-grid estimate (81) therefore gives $\sum_j (A_j^{\mathsf{S}})^2 \leq C_q$. Applying the same estimate to $\min\{1, za(u)\}$ gives

$$\sum_j \min\{1, zA_j^{\mathsf{S}}\} \leq C_q \log(1+z). \tag{92}$$

Indeed, for $z \leq 2$ the sum is $O_q(z)$; for $z > 2$, the saturated interval between the two roots of $za(u) = 1$ has logarithmic width $O(\log z)$ and both unsaturated tails are geometric.

For a contiguous block with endpoint scales r_- and $r_+ = q^{L-1}r_-$, choose $\sigma = \sqrt{r_- r_+}$. The endpoint values agree, and unimodality gives

$$\min_{j\in\mathcal{B}} A_j^{\mathsf{S}}(\sigma) = \frac{q^{(L-1)/2}}{1+q^{L-1}} \geq a_q q^{-L/2}. \tag{93}$$

Finally, $a(u) \le \min\{u, u^{-1}\}$. Once a length-$L$ block lies wholly on either side of the crossing $u = 1$, moving one block farther away contracts its ℓ_2 profile norm by at least a factor q^{-L}, up to a q-dependent endpoint constant. On the low side this follows explicitly from

$$\frac{a(u/q^L)}{a(u)} = q^{-L}\frac{1+u^2}{1+q^{-2L}u^2} \le 2q^{-L}, \qquad 0 < u \le 1, \tag{94}$$

and the high side follows from $a(u) = a(u^{-1})$. The saturation count and two geometric tails used for (89) then give

$$\sum_r \min\{1, zE_r^{\mathsf{S}}(\sigma)\} \le \min\left\{M, C_q\left(1 + \frac{\log(1+z)}{L}\right)\right\}. \tag{95}$$

Equations (92)– (95), together with the Fourier calculation, prove Lemma 8.

F Nested test functions and blockwise transcript bounds

Fix $\mathsf{a} \in \{\mathsf{F}, \mathsf{S}\}$ and write $m_j^{\mathsf{a}} = \mathbb{E}_{\mu_j^{\mathsf{a}}}|h_j^{\mathsf{a}}|$. The preceding moment calculations give $\inf_j m_j^{\mathsf{a}} > 0$. For complete blocks $\mathcal{B}_1, \dots, \mathcal{B}_M$, define

$$\begin{aligned} W_r(\theta, x) &= \tanh\left(L^{-1/2}\sum_{j\in\mathcal{B}_r}\theta_j \operatorname{sign} h_j^{\mathsf{a}}(x_j)\right), \\ F(\theta, x) &= \tanh\left(M^{-1/2}\sum_{r=1}^{M} W_r(\theta, x)\right). \end{aligned} \tag{96}$$

Here F is the block witness F_{b} used in Section 3.

Under the ideal joint experiment, interpolate all ML Rademacher biases from zero to their values $\beta_j = \varepsilon m_j^{\mathsf{a}}$. Let $Z_j(t)$ be independent Rademachers with mean $t\beta_j$ and define

$$\Psi(t) = \mathbb{E}\tanh\left(M^{-1/2}\sum_{r=1}^{M}\tanh\left(L^{-1/2}\sum_{j\in\mathcal{B}_r} Z_j(t)\right)\right). \tag{97}$$

Fix a coordinate j in block r and write

$$U = L^{-1/2}\sum_{i\in\mathcal{B}_r\setminus\{j\}} Z_i, \qquad V = M^{-1/2}\sum_{s\ne r} W_s. \tag{98}$$

At every interpolation time, $\mathbb{E}U^2 \le 2$: its variance is at most one and its mean is at most $\varepsilon\sqrt{L} \le \alpha_L$. Also $|\mathbb{E}W_s| \le L^{-1/2}\sum_{i\in\mathcal{B}_s}\beta_i \le \varepsilon\sqrt{L}$. This bound follows by interpolating the inner tanh and using its $L^{-1/2}$ coordinate Lipschitz constant. Independence across blocks then gives $\mathbb{E}V^2 \le 2$. The two remainders are independent, so

$$\mathbb{P}\{|U| \le 3,\ |V| \le 3\} \ge (1 - 2/9)^2. \tag{99}$$

On this event, flipping Z_j changes the inner tanh by at least $2\operatorname{sech}^2(4)/\sqrt{L}$; the resulting change in the outer tanh is at least another factor $\operatorname{sech}^2(4)/\sqrt{M}$. Differentiating the finite product law exactly as in (65) gives

$$\Psi'(t) = \sum_{j=1}^{ML}\frac{\beta_j}{2}\mathbb{E}\big[F(Z_j = 1, Z_{-j}(t)) - F(Z_j = -1, Z_{-j}(t))\big]. \tag{100}$$

The central event therefore makes every summand at least $c\beta_j/\sqrt{ML}$. Since $\sum_j \beta_j \ge a_q\varepsilon ML$, integration over the bias path gives

$$\mathbb{E}_P F(\Theta, X) \ge c_{\mathrm{b}}(q)\varepsilon\sqrt{ML} = c_{\mathrm{b}}(q)\alpha_L. \tag{101}$$

Average one whole parameter block ζ at a time. For fixed complement η, put

$$\bar{Q}_\eta = 2^{-L} \sum_{\zeta'} Q_{\zeta',\eta}, \qquad \bar{F}_\eta(x) = 2^{-L} \sum_{\zeta'} F(\zeta', \eta, x). \tag{102}$$

The exact change at this symmetrization step is

$$2^{-L} \sum_{\zeta} \int \big(F_{\zeta,\eta} - \bar{F}_\eta\big)\, d\big(Q_{\zeta,\eta} - \bar{Q}_\eta\big). \tag{103}$$

The oscillation of the outer tanh when $W_r \in [-1, 1]$ changes is at most $2/\sqrt{M}$; TV duality therefore bounds the absolute average of (103) by $4J_r(Q)/\sqrt{M}$. Averaging earlier blocks can only reduce later J_r by joint convexity. After all complete blocks are averaged, the expectation is zero: for every fixed leftover parameter and x, pairing the complete block vector ζ with $-\zeta$ negates every W_r and hence negates F. Thus

$$|\mathbb{E}_Q F| \le \frac{4}{\sqrt{M}} \sum_{r=1}^{M} J_r(Q). \tag{104}$$

Combining (63), (101), and (104) proves (61).

We now prove the adaptive contraction without conditioning on a parameter posterior. Let $\mathsf{H}_\theta = \mathrm{Law}_\theta(H_T)$ and $\mathsf{H} = \{\mathsf{H}_\theta\}_\theta$. Writing $\theta = (\zeta, \eta)$ for the block decomposition, convexity gives

$$J_r(\mathsf{H}) \le \mathbb{E}_{\eta,\zeta,\zeta'}\, \mathrm{TV}(\mathsf{H}_{\zeta,\eta}, \mathsf{H}_{\zeta',\eta}) \tag{105}$$

$$\le \mathbb{E}_{\eta,\zeta,\zeta'} \mathbb{E}_{\zeta,\eta} \sum_{t \le T} \mathrm{TV}(K^{H_{t-1}}_{\zeta,\eta,t}, K^{H_{t-1}}_{\zeta',\eta,t}), \tag{106}$$

where the second line is Lemma 7 for each fixed complete parameter pair. At the same history, the action and covariance coincide. Only coordinates in $\mathcal{B}_r$ can have different means. Put $E^{\mathsf{a}}_r(\sigma) = (\sum_{j \in \mathcal{B}_r} A^{\mathsf{a}}_j(\sigma)^2)^{1/2}$. Then

$$\|\mu_{\zeta,\eta} - \mu_{\zeta',\eta}\|_2 \le C\varepsilon E^{\mathsf{a}}_r(\sigma_t). \tag{107}$$

The TV distance between equal-covariance Gaussian shifts is therefore at most

$$\min\{1, C\varepsilon \sqrt{b_t} E^{\mathsf{a}}_r(\sigma_t)\}. \tag{108}$$

This bound is independent of ζ'. Averaging the source parameter and its path law yields

$$J_r(\mathsf{H}) \le \mathbb{E} \sum_{t \le T} \min\{1, C\varepsilon \sqrt{b_t} E^{\mathsf{a}}_r(\sigma_t)\}. \tag{109}$$

If $Q_\theta = \mathsf{H}_\theta K$ for a common output kernel K, TV data processing and convexity give $J_r(Q) \le J_r(\mathsf{H})$. Summing (109) over blocks and applying (89) for F or (95) for S proves (113).

G Coordinatewise bounds and the lower bound

Table 1 summarizes the converse and attaining constructions.

Table 1: Each precision regime has a matching sampling requirement and attaining construction. Repetition supplies sufficient precision, coordinate thinning supplies weak dependence at lower query frequency, and block activation supplies joint dependence with a shared query. Rates hold up to constants depending on q and κ; boundaries are specified in (8).

Precision	Lower bound	Sampling procedure	Expected queries
Low	Accumulated precision, (111)	Repetition	d/B
Intermediate	Coordinatewise TV sensitivity, (112)	Coordinate thinning	$d/\sqrt{B}$
High	Blockwise TV sensitivity, (61) and (113)	Block activation	$\alpha\sqrt{d/\ell_B}$

For an adjacent pair θ, θ' differing only in coordinate j and a realized action (x, σ, b), Gaussian shift formulas and Lemma 8 give

$$\mathrm{TV}_j(x,\sigma,b) \le \min\{1, C\varepsilon\sqrt{b}\, A_j^{\mathsf{a}}(\sigma)\}, \qquad D_{\mathrm{KL},j}(x,\sigma,b) \le C\varepsilon^2 b\big(A_j^{\mathsf{a}}(\sigma)\big)^2. \tag{110}$$

Averaging over parameter pairs gives coordinatewise transcript bounds. For blocks, convexity bounds the distance to the uniform mixture of block signs. Lemma 7 then accumulates these contributions along the adaptive transcript.

Lemma 9 (Coordinatewise and blockwise bounds for adaptive transcript laws)**.** *For every fixed $q > 1$, there are constants $c_q, C_q > 0$ such that the following holds. For either family, let H_T be any legal transcript, set $\mathsf{H} = \{\mathsf{H}_\theta := \mathrm{Law}_\theta(H_T)\}_\theta$, and let $\gamma = [c_{\mathrm{f}}(q) - 2\delta/\alpha]_+$. Suppose there is a parameter-independent Markov kernel K such that $Q_\theta = \mathsf{H}_\theta K$ and $2^{-d}\sum_\theta \mathrm{TV}(Q_\theta, P_\theta^{\mathsf{a}}) \le \delta$. Then*

$$\mathbb{E}\sum_{t\le T} b_t \ge c_q d\gamma^2, \tag{111}$$

$$\mathbb{E}\sum_{t\le T} \log(1 + C_q\varepsilon\sqrt{b_t}) \ge c_q\varepsilon d\gamma. \tag{112}$$

For every $L \in [d]$, let $M = \lfloor d/L \rfloor$ and let $\mathcal{B}_1, \dots, \mathcal{B}_M$ be the blocks defined in Lemma 5(ii)*. Then*

$$\sum_{r=1}^{M} J_r(\mathsf{H}) \le C_q\mathbb{E}\sum_{t\le T} \min\left\{M, 1 + \frac{\log(1+\varepsilon\sqrt{b_t})}{L}\right\}. \tag{113}$$

Take $S = [d]$ in (60). With $\gamma = [c_{\mathrm{f}}(q) - 2\delta/\alpha]_+$,

$$V := \sum_{j=1}^{d} \bar{v}_j(Q) \ge \varepsilon d\gamma. \tag{114}$$

Let $\mathsf{H} = \{\mathsf{H}_\theta := \mathrm{Law}_\theta(H_T)\}_\theta$. Output data processing bounds this by the corresponding separations between transcript laws. For each fixed adjacent pair, Lemma 7 and the Gaussian shift TV formula give

$$\bar{v}_j(\mathsf{H}) \le \mathbb{E}_{\eta,+}\sum_{t\le T} \min\{1, C\varepsilon\sqrt{b_t}A_j^{\mathsf{a}}(\sigma_t)\}. \tag{115}$$

After summing in j, the change from the directed average over parameter pairs to the uniform Bayes average costs only a factor two. Explicitly, for either nonnegative one-step cost $R_{j,t}$ used below,

$$\sum_j \mathbb{E}_{\eta,+}\sum_{t\le T} R_{j,t} = 2\mathbb{E}_\Theta \sum_j \mathbf{1}\{\Theta_j = +1\}\sum_{t\le T} R_{j,t} \le 2\mathbb{E}_\Theta \sum_j \sum_{t\le T} R_{j,t}. \tag{116}$$

Equation (83) or (92) therefore proves (112).

For the sum of squared coordinatewise separations, Pinsker and Jensen give

$$\bar{v}_j(\mathsf{H})^2 \le \frac{1}{2}\mathbb{E}_{\eta,+} D_{\mathrm{KL}}(\mathsf{H}_{+,\eta} \| \mathsf{H}_{-,\eta}) \tag{117}$$

$$\le C\varepsilon^2 \mathbb{E}_{\eta,+}\sum_{t\le T} b_t (A_j^{\mathsf{a}}(\sigma_t))^2. \tag{118}$$

Summing, applying the corresponding ℓ_2 envelope, and using $\sum_j \bar{v}_j^2 \ge V^2/d$ proves (111).

When $\delta = \kappa\alpha$ and $\kappa_0(q)$ is chosen below $c_{\mathrm{f}}(q)/4$, γ is a positive q-dependent constant. The cap $b_t \le B$ gives (18); in the intermediate interval, $\log(1 + C\varepsilon\sqrt{B}) \le C'\varepsilon\sqrt{B}$. For the block lower bound, take L to be an integer within a fixed factor of ℓ_B when $\ell_B < d$. Then $ML \ge d/2$ and $[c_{\mathrm{b}}(q)\alpha_L - 2\delta]_+ = \Omega_q(\alpha)$ after additionally requiring $\kappa_0(q) \le c_{\mathrm{b}}(q)/(4\sqrt{2})$. By this choice of L, $1 + \log(1 + \varepsilon\sqrt{B})/L = O(1)$. The per-query factor on the right-hand side of (113) is therefore $O_q(1)$, giving $\mathbb{E}T \gtrsim_{q,\kappa} \alpha\sqrt{d/\ell_B}$. If $\ell_B = d$, use the single block $L = d$ and the $M = 1$ cap in (113). This proves (19) and the full converse.

H REPETITION AND COORDINATE THINNING

For $s \in [0, 1/4]$ and $\theta_j \in \{\pm 1\}$, write

$$P^{\mathsf{a}}_{s,\theta_j,j}(dx) := \mu^{\mathsf{a}}_j(dx)\big(1 + s\theta_j h^{\mathsf{a}}_j(x)\big). \tag{119}$$

For Fourier coordinate j, choose

$$\frac{\sigma_j}{\sqrt{1+\sigma_j^2}} = \frac{1}{k_j}. \tag{120}$$

Then $A_j(\sigma_j) = e^{-1/2}\sqrt{1 - k_j^{-2}} \geq a_q > 0$. Choose $x_{t,j}$ at a unit sine phase and set $x_{t,\ell} = 0$ for $\ell \neq j$; the other response means are parameter-independent and can be discarded. For spatial coordinate j, choose $\sigma_j = r_j$, query at $x_j = 0$, and discard the other response coordinates. Equation (41) gives $A^{\mathsf{S}}_j = 1/2$. Thus both families provide a matched mean half-gap at least $a_q\varepsilon$.

With n repetitions at precision B, the sufficient statistic is an equal-variance Gaussian shift of effective precision nB. By (36), $n = \lceil C_q/B \rceil$ makes its BSC correlation at least 2ε. Applying (20) gives one exact target coordinate. Repeating for every j costs $O_q(d/B)$ when $B \leq b_0(q)$.

For thinning, independently activate each coordinate with probability

$$p_{\mathrm{c}} = \max\{C_q/\sqrt{B}, 4\varepsilon\}, \tag{121}$$

where C_q is first chosen large enough for the correlation bound at a matched query below. Then choose $b_0(q) \geq C_q^2$. In the thinning interval, $B > b_0(q)$ and $4\varepsilon \leq 1$ ensure $p_{\mathrm{c}} \leq 1$. An active coordinate has target amplitude $s = \varepsilon/p_{\mathrm{c}} \leq 1/4$. One precision-$B$ matched query has correlation at least $2s$. To verify this uniformly, put $w = \varepsilon\sqrt{B}$ and $f(w) = 2\Phi(a_q w) - 1$, decreasing a_q if necessary to absorb the constant in the exact score difference. Choose C_q so that $f(C_q/4) \geq 1/2$. Concavity and monotonicity give

$$f(w) \geq \begin{cases} 2w/C_q, & w \leq C_q/4, \\ 1/2, & w \geq C_q/4. \end{cases} \tag{122}$$

If $p_{\mathrm{c}} = C_q/\sqrt{B}$, then $2s = 2w/C_q$; if $p_{\mathrm{c}} = 4\varepsilon$, then $2s = 1/2$. Thus the response simulates $P^{\mathsf{a}}_{s,\theta_j,j}$ in both cases. An inactive coordinate outputs $P^{\mathsf{a}}_{0,j}$. Identity (21) then gives the exact target coordinate, independently across j, and

$$\mathbb{E}T = dp_{\mathrm{c}} \leq C_q d/\sqrt{B} + 4\varepsilon d. \tag{123}$$

For $b_0(q) < B \leq b_1(q,\kappa)\varepsilon^{-2}$ the second term is absorbed by the first, since

$$4\varepsilon d \leq 4\sqrt{b_1(q,\kappa)}\, d/\sqrt{B}. \tag{124}$$

This also covers any constant-width gap between the repetition and block constructions.

I BLOCK SAMPLING AND TOTAL VARIATION ACCURACY

For a constant $K_q > 0$ specified in Proposition 10, define the block precision thresholds

$$B^{(\Gamma)}_{L,\mathsf{F}} = K_q \frac{q^{2L}}{L\min\{1, \Gamma^2\varepsilon^2 L\}}, \tag{125}$$

$$B^{(\Gamma)}_{L,\mathsf{S}} = K_q \frac{q^{L}}{\min\{1, \Gamma^2\varepsilon^2 L\}}. \tag{126}$$

Proposition 10 (Randomized block simulator)**.** *For every fixed $q > 1$, there are constants $K_q, C_q \in (0,\infty)$ such that the following holds. Fix $\kappa > 0$ and $\Gamma \geq \max\{4, 2C_q/\kappa\}$, let $L \in [d]$, $M = \lceil d/L \rceil$, and $p = \min\{1, \Gamma\varepsilon\sqrt{L}\}$. For either $\mathsf{a} \in \{\mathsf{F}, \mathsf{S}\}$, if $B \geq B^{(\Gamma)}_{L,\mathsf{a}}$ with the thresholds in* (125)–(126) *defined using this K_q, then there is a nonadaptive legal sampler with*

$$\mathbb{E}T = Mp \asymp_{q,\kappa} \min\{M, \Gamma\alpha\sqrt{M}\} \tag{127}$$

and, uniformly in θ,

$$\mathrm{TV}(Q_\theta, P^{\mathsf{a}}_\theta) \leq C_q\alpha/\Gamma < \kappa\alpha. \tag{128}$$

When $p = 1$, the sampler is exact.

Fix $\mathsf{a} \in \{\mathsf{F}, \mathsf{S}\}$ and partition the coordinates into $M_+ = \lceil d/L \rceil$ consecutive blocks. A full block uses (84) for F and (93) for S. Embed a final short block in an adjacent length-L scale window, choose that window's scale, and use only the short-block coordinates. This extra call has independent noise, so the emitted block outputs remain conditionally independent.

Fix $\Gamma \geq 4$ and p, s as in (22). Choose $K_q < \infty$ to absorb the constants for matched score differences, thresholding, and the final shorter block. On an active Fourier block choose phases with sine equal to one; on an active spatial block query at $x = 0$. Every coordinate half-gap is at least $a'_q \varepsilon m_{L,\mathsf{a}}$, where

$$m_{L,\mathsf{F}} = \sqrt{L}\, q^{-L}, \qquad m_{L,\mathsf{S}} = q^{-L/2}. \tag{129}$$

The condition $B \geq B^{(\Gamma)}_{L,\mathsf{a}}$ gives

$$\sqrt{B}\, a'_q \varepsilon m_{L,\mathsf{a}} \geq a'_q \sqrt{K_q}\, s, \tag{130}$$

both when $s = (\Gamma\sqrt{L})^{-1}$ for $p < 1$ and when $s = \varepsilon$ for $p = 1$. Choose K_q large enough that the resulting BSC correlation is at least $2s$ for every $s \leq 1/4$, by the same concavity argument as in (122). Degrade independently to exactly $2s$ and apply (20). The active output is precisely

$$P^{\mathsf{a}}_{s,\theta,r}(dx) = \prod_{j \in \mathcal{B}_r} \mu^{\mathsf{a}}_j(dx_j)(1 + s\theta_j h^{\mathsf{a}}_j(x_j)). \tag{131}$$

Write $P^{\mathsf{a}}_{\theta,r} := P^{\mathsf{a}}_{\varepsilon,\theta,r}$ for the ideal target block. In the error calculation below we suppress the fixed family superscript. Independently draw activation indicators $A_r \sim \mathrm{Bernoulli}(p)$ before the interaction. If $A_r = 0$, emit the base block $P^{\mathsf{a}}_{0,r}$; if $A_r = 1$, make the one query above. Thus

$$Q_{\theta,r} = (1-p)P^{\mathsf{a}}_{0,r} + pP^{\mathsf{a}}_{s,\theta,r}, \qquad Q_\theta = \bigotimes_{r=1}^{M_+} Q_{\theta,r}, \qquad T = \sum_{r=1}^{M_+} A_r. \tag{132}$$

All actions are predictable and T is almost surely finite. Moreover $\mathbb{E}T = M_+ p$. When $p < 1$,

$$M_+ p = \Gamma\varepsilon\sqrt{L}\, M_+ \asymp \Gamma\alpha\sqrt{M_+}, \tag{133}$$

because $d \leq M_+ L \leq 2d$. When $p = 1$, $\Gamma\varepsilon\sqrt{L} \geq 1$ implies $\Gamma\alpha\sqrt{M_+} \gtrsim M_+$, so $M_+ \asymp \min\{M_+, \Gamma\alpha\sqrt{M_+}\}$. This proves (127) in both branches.

This construction explains the high precision example in Figure 1. The expected query count is the number of blocks times their activation probability. Suppose both L and $4L$ are feasible at the given precision and both activation probabilities are below one, that is, $2\Gamma\varepsilon\sqrt{L} < 1$. Increasing the block length from L to $4L$ then doubles the activation probability. If $4L$ divides d, the number of blocks is quartered and the expected query count is exactly halved. With a shorter final block, the same comparison holds up to the rounding in $M = \lceil d/L \rceil$.

We use $\chi^2(Q\|P) := \int (dQ/dP - 1)^2\, dP$ for $Q \ll P$. We prove the uniform global error. Fix a block, write $u_j = \theta_j h^{\mathsf{a}}_j$ and $f_a = \prod_j (1 + a u_j)$ relative to the product base $\bigotimes_j \mu^{\mathsf{a}}_j$, and

$$v_j = \mathbb{E}_{\mu^{\mathsf{a}}_j} \frac{u_j^2}{1 + \varepsilon u_j}. \tag{134}$$

Positivity and $\|h^{\mathsf{a}}_j\|_\infty < 2$ give the uniform bound

$$0 \leq v_j \leq \frac{\mathbb{E}_{\mu^{\mathsf{a}}_j}(h^{\mathsf{a}}_j)^2}{1 - \varepsilon\|h^{\mathsf{a}}_j\|_\infty} \leq C. \tag{135}$$

Under P_ε, put $A = f_0/f_\varepsilon = 1/f_\varepsilon$ and $D = f_s/f_\varepsilon$. Their coordinate factors are $A^{(0)}_j = (1 + \varepsilon u_j)^{-1}$ and $D^{(s)}_j = (1 + s u_j)/(1 + \varepsilon u_j)$. Coordinate factorization gives

$$\mathbb{E}_{P_\varepsilon}(A^{(0)}_j)^2 = 1 + \varepsilon^2 v_j, \tag{136}$$

$$\mathbb{E}_{P_\varepsilon}(D^{(s)}_j)^2 = 1 + (s - \varepsilon)^2 v_j, \tag{137}$$

$$\mathbb{E}_{P_\varepsilon} A^{(0)}_j D^{(s)}_j = 1 - \varepsilon(s - \varepsilon) v_j. \tag{138}$$

Since the likelihood ratio of Q is $(1-p)A + pD$, expanding the products in (138), using $\varepsilon = ps$ and $s - \varepsilon = (1-p)s$, yields

$$\chi^2(Q\|P_\varepsilon) = \sum_{\substack{S\subseteq\mathcal{B}\\|S|\geq 2}} p^2(1-p)^2 s^{2|S|}\left[p^{|S|-1} + (-1)^{|S|}(1-p)^{|S|-1}\right]^2 \prod_{j\in S} v_j. \tag{139}$$

For $|S| = 1$, the coefficient vanishes because

$$(1-p)(-ps) + p(1-p)s = 0. \tag{140}$$

The terms involving a single coordinate cancel exactly. Since the bracket in (139) has absolute value at most one for $|S| \geq 2$,

$$\chi^2(Q\|P_\varepsilon) \leq p^2\left[(1+Cs^2)^{|\mathcal{B}|} - 1 - C|\mathcal{B}|s^2\right] \tag{141}$$

$$\leq Cp^2|\mathcal{B}|^2 s^4 e^{C|\mathcal{B}|s^2}. \tag{142}$$

When $p < 1$, $p = \Gamma\varepsilon\sqrt{L}$ and $s = (\Gamma\sqrt{L})^{-1}$, so

$$\chi^2(Q_{\theta,r}\|P_{\theta,r}) \leq \frac{C}{\Gamma^2}\varepsilon^2|\mathcal{B}_r|. \tag{143}$$

When $p = 1$, $Q_{\theta,r} = P_{\theta,r}$ exactly.

Let $a_r = \int\sqrt{dP_{\theta,r}dQ_{\theta,r}}$. Then $2(1-a_r) \leq \chi^2(Q_{\theta,r}\|P_{\theta,r})$ and product affinity is $a = \prod_r a_r$. Hence

$$\mathrm{TV}(P_\theta, Q_\theta)^2 \leq 1 - a^2 \leq 2(1-a) \leq \sum_r \chi^2(Q_{\theta,r}\|P_{\theta,r}). \tag{144}$$

Using $\sum_r |\mathcal{B}_r| = d$ proves (128). Taking $\Gamma \geq 2C_q/\kappa$ makes the error smaller than $\kappa\alpha$.

J Feasible block lengths and regime transitions

Set $z = \varepsilon\sqrt{B}$ and $R_B = \min\{d, 1 + \log(1+z)\}$. We show that in the high precision interval

$$L_{B,\mathsf{a}}^{\mathrm{up}} := \max\{1 \leq L \leq d : B \geq B_{L,\mathsf{a}}^{(\Gamma)}\} \asymp_{q,\Gamma} R_B. \tag{145}$$

If the feasible set is empty, its value is defined to be zero and the coordinate construction is used instead.

For F, feasibility of L implies

$$q^{2L} \leq K_q^{-1}\Gamma^2 z^2 L^2. \tag{146}$$

Taking logarithms and using $\log L \leq (\log q)L/2 + C_q$ gives

$$L \leq C_{q,\Gamma}[1 + \log(1+z)], \tag{147}$$

and of course $L \leq d$. This proves the upper half of (145).

Conversely, choose $L_0 = \max\{1, \lfloor c_q R_B\rfloor\}$ with c_q small enough that $2c_q \log q \leq 1/2$. Then $q^{2L_0} \leq C_q(1+z)^{1/2}$. Since $1/\min\{1,a\} \leq 1 + 1/a$,

$$B_{L_0,\mathsf{F}}^{(\Gamma)} \leq K_q q^{2L_0}\left(\frac{1}{L_0} + \frac{1}{\Gamma^2\varepsilon^2 L_0^2}\right). \tag{148}$$

After division by $B = z^2/\varepsilon^2$,

$$\frac{B_{L_0,\mathsf{F}}^{(\Gamma)}}{B} \leq K_q q^{2L_0}\left[\frac{\varepsilon^2}{L_0 z^2} + \frac{1}{\Gamma^2 L_0^2 z^2}\right]. \tag{149}$$

The right side is at most $C_{q,\Gamma}(1+z)^{1/2}/z^2$. Choose $z_0 = z_0(q,\Gamma,K_q)$ so this quantity is at most one for every $z \geq z_0$, and choose the theorem's fixed high precision threshold $b_1(q,\kappa) \geq z_0^2$ (recall that Γ is fixed from κ). Then the high interval has $z = \varepsilon\sqrt{B} \geq z_0$, so L_0 is feasible. The same argument applies when $R_B = d$ and gives $L_0 \asymp d$. This proves (145) for F without a monotonicity assumption.

For S, feasibility gives

$$q^L \le K_q^{-1}\Gamma^2 z^2 L, \tag{150}$$

and hence the same upper bound $L \le C_{q,\Gamma}[1+\log(1+z)]$. For the same $L_0 = \max\{1, \lfloor c_q R_B \rfloor\}$, now choose c_q so $c_q \log q \le 1/2$. Then

$$\frac{B^{(\Gamma)}_{L_0,\mathsf{S}}}{B} \le K_q q^{L_0}\left[\frac{\varepsilon^2}{z^2} + \frac{1}{\Gamma^2 L_0 z^2}\right] \le \frac{C_{q,\Gamma}(1+z)^{1/2}}{z^2}. \tag{151}$$

Increasing the same fixed z_0 makes L_0 feasible. Thus (145) holds for both profiles.

For the converse, $M_- := \lfloor d/L \rfloor$ complete blocks satisfy $M_- L \ge d/2$ for all $1 \le L \le d$; for the upper bound, $M_+ := \lceil d/L \rceil$ and $d \le M_+ L \le 2d$. These distinct roundings only change constants. The short final upper block is handled as in Section I.

Choose $b_0(q)$ so repetition and thinning overlap at constant precision. If the high precision inversion starts at $z = \varepsilon\sqrt{B} \ge z_0(q,\Gamma)$, choose $b_1(q,\kappa) \ge \max\{z_0(q,\Gamma)^2, b_0(q)/16\}$. At $B \asymp 1$, $d/B \asymp d/\sqrt{B}$. At $B \asymp \varepsilon^{-2}$,

$$d/\sqrt{B} \asymp \varepsilon d = \alpha\sqrt{d}, \tag{152}$$

This matches the high precision branch because $\ell_B = \Theta(1)$. At the boundary between partial and full activation, $\Gamma^2\varepsilon^2 L = 1$, the two formulas in (125) both equal $K_q q^{2L}/L$, while those in (126) equal $K_q q^L$; the expected number of active blocks changes continuously up to constants. Finally, when $\ell_B = d$, inversion gives $L^{\mathrm{up}}_{B,\mathsf{a}} \ge c_{q,\kappa} d$ and hence $M_+ = O_{q,\kappa}(1)$, but it need not give exactly one block. When $p < 1$,

$$M_+ p = \Gamma\varepsilon\sqrt{L^{\mathrm{up}}_{B,\mathsf{a}}}M_+ = O_{q,\kappa}(\alpha). \tag{153}$$

When $p = 1$, $\Gamma\varepsilon\sqrt{L^{\mathrm{up}}_{B,\mathsf{a}}} \ge 1$ implies $\alpha \ge c_{q,\kappa} > 0$, so $M_+ = O_{q,\kappa}(1) = O_{q,\kappa}(\alpha)$. This completes the upper-bound stitching and the proof of Theorem 1.

K Proof for equivalent scale layouts

Theorem 3 in Section 5 compares two Blackwell-equivalent target experiments under the same precision cap. The proof separates their query costs by applying the block construction to the clustered layout and the coordinatewise converse to the separated layout, then verifies equivalence by public rescaling.

In the proof below, we suppress the common arguments $(D, m, \varepsilon, q, B, \delta)$ from the query complexity notation for each layout. We prove Theorem 3. Inactive coordinates in (26) have a known law independent of θ. Their score responses, including fresh noise, can be simulated from the public action (x, σ, b) using internal randomness, and their target coordinates can be generated after the interaction. They are therefore ancillary for both lower and upper bounds. Choose b_2 at least as large as the high precision threshold b_1 in Theorem 1 and the constant required by the thinning construction in (121).

First consider $S_{\mathrm{cl}} = [m]$. After discarding the ancillary coordinates, the experiment is exactly the m-dimensional spatial experiment of Theorem 1, with $\varepsilon = \alpha/\sqrt{m}$. Its high precision branch gives

$$T^*_{\mathsf{S}}(S_{\mathrm{cl}}) \asymp_{q,\kappa} \alpha\sqrt{m/\ell_B^{(m)}}. \tag{154}$$

For the spread layout, the active profile values lie on a grid of ratio q^G. Repeating the proof of (92) with q replaced by q^G gives, uniformly in $\sigma > 0$ and $z \ge 0$,

$$\sum_{r=1}^{m} \min\{1, zA^{\mathsf{S}}_{1+(r-1)G}(\sigma)\} \le C_q \min\left\{m, 1 + \frac{\log(1+z)}{G}\right\}. \tag{155}$$

Indeed, the indices satisfying $za(u) \ge 1$ occupy a logarithmic interval of width $O(\log(1+z))$, containing $O_q(1+\log(1+z)/G)$ points of the sparse grid. On either side, $a(u) \le \min\{u, u^{-1}\}$ bounds the remaining contributions by geometric series with ratio q^{-G} and total $O_q(1)$. If the interval is empty, the same bound follows by splitting the grid at $u = 1$.

For the output demand, discard inactive coordinates and map each active coordinate to $x'_r = q^r x_{j_r}/r_{j_r}$. This parameter-independent map takes the target to the m-dimensional spatial family and contracts both the output TV error and each face separation. Applying Lemma 5(i) to the transformed output, with $\delta = \kappa\alpha$ and admissible κ, therefore gives

$$\sum_{r=1}^{m} \bar{v}_r(Q) \geq c_{q,\kappa}\varepsilon m. \tag{156}$$

Let $\mathsf{H} = \{\mathsf{H}_\theta := \mathrm{Law}_\theta(H_T)\}_\theta$ for the layout experiment. The fixed-pair path argument in (115), followed by (155), bounds the separation between transcript laws by

$$\sum_{r=1}^{m} \bar{v}_r(\mathsf{H}) \leq C_q \mathbb{E} \sum_{t \leq T} \left(1 + \frac{\log(1 + C_q\varepsilon\sqrt{b_t})}{G}\right). \tag{157}$$

This calculation uses the complete D-dimensional Gaussian mean vector; all inactive coordinates are common to each adjacent parameter pair. Since $b_t \leq B$ and $G \geq C_{\mathrm{sep}} w_B$, choosing C_{sep} sufficiently large in terms of (q, κ) makes the parenthesis $O_{q,\kappa}(1)$. Data processing and (156) therefore yield

$$\mathbb{E}T \geq c_{q,\kappa}\varepsilon m = c_{q,\kappa}\alpha\sqrt{m}. \tag{158}$$

For the matching upper bound, independently activate every active coordinate with the probability p_{c} in (121), query it at its matched scale when active, and sample every inactive target coordinate from its public base law. This produces the target exactly. The condition $B \geq b_2\varepsilon^{-2}$ implies $p_{\mathrm{c}} \leq C_{q,\kappa}\varepsilon$, and hence

$$\mathbb{E}T = m p_{\mathrm{c}} \leq C_{q,\kappa}\varepsilon m = C_{q,\kappa}\alpha\sqrt{m}. \tag{159}$$

All achieving procedures use only midpoint comparisons, while every comparator transcript is a garbling of the analog transcript. Thus the same rates hold for the comparator oracle.

It remains to verify equivalence of the two target experiments. Under either layout, map each active coordinate to $z_r = x_{j_r}/r_{j_r}$ and discard the inactive coordinates. Conditional on θ_r, this gives the canonical factor

$$\frac{1 + \varepsilon\theta_r}{2}\mathcal{N}(1, 1) + \frac{1 - \varepsilon\theta_r}{2}\mathcal{N}(-1, 1). \tag{160}$$

Conversely, rescale each canonical factor by the public r_{j_r} and generate the inactive base coordinates independently. These parameter-independent kernels are inverse Blackwell garblings. Equations (154), (158), and (159) prove Theorem 3.

Along any sequence with $G = o(w_B)$ and $\min\{m, w_B/G\} \to \infty$, the spread layout admits a sampler with expected query cost $o(\alpha\sqrt{m})$ at the prescribed TV accuracy.

Choose $K = \lfloor c_q \min\{m, w_B/G\} \rfloor$ and partition the active indices into $\lceil m/K \rceil$ consecutive groups. Each full group has K active factors and physical span $W \leq KG \leq c_q w_B$; the final group is shorter. Choose $c_q \log q \leq 1$. With $z = \varepsilon\sqrt{B}$, this gives $q^W \leq C_q(1 + z)$, so along the stated sequence the block construction is eventually feasible with $p = \min\{1, \Gamma\varepsilon\sqrt{K}\}$ and $B \gtrsim_q q^W / \min\{1, \Gamma^2\varepsilon^2 K\}$; inactive positions are public ancillary factors. The same cancellation of terms involving a single coordinate controls the global TV error. Its expected call count is

$$\left\lceil \frac{m}{K} \right\rceil \min\{1, \Gamma\varepsilon\sqrt{K}\} = O_{q,\kappa}\left(\alpha\sqrt{m/K}\right) = O_{q,\kappa}\left(\alpha \max\left\{1, \sqrt{mG/w_B}\right\}\right), \tag{161}$$

Since $K \to \infty$, this cost is $o(\alpha\sqrt{m})$. The cost of order $\alpha\sqrt{m}$ in Theorem 3 therefore cannot persist along these sequences. The construction also covers the range $w_B > m$.

L THRESHOLD OBSERVATIONS

A companion interface applies a coordinatewise comparator after the Gaussian noise is added. The algorithm also chooses a predictable threshold vector $\tau_t \in \mathbb{R}^d$ and observes

$$Z_{t,j} = \mathbf{1}\{Y_{t,j} \geq \tau_{t,j}\}, \qquad Z_t \in \{0, 1\}^d. \tag{162}$$

Define $T^*_{\mathsf{a},\mathrm{cmp}}$ by (6) for this interface. The analogous definition applies to the active-layout families of Section 5.

Corollary 11 (Coordinatewise threshold observations)**.** *Under the assumptions of Theorem 1, for each* $\mathsf{a} \in \{\mathsf{F}, \mathsf{S}\}$ *the comparator oracle has the same rate:*

$$T^*_{\mathsf{a},\mathrm{cmp}}(d, \varepsilon, q, B, \kappa\alpha) \asymp_{q,\kappa} \mathcal{R}_d(B). \tag{163}$$

Each query returns a vector of d *threshold bits. At* $\alpha = 1$ *and* $B = d$*, the optimal expected number of queries is* $\Theta_{q,\kappa}(\sqrt{d})$*, and the target capacity is* $\Theta_q(1)$*.*

For the analog oracle, a precision-b_t response with $b_t \le B$ can be simulated from a precision-B response by adding conditionally independent Gaussian noise

$$\left(b_t^{-1} - B^{-1}\right)^{1/2} G'_t, \qquad G'_t \sim \mathcal{N}(0, I_d). \tag{164}$$

For the comparator oracle, the corresponding simulation randomizes the threshold before the query, as shown in the proof below. We retain predictable b_t because the same bounds also yield the cumulative-precision result in Appendix M.

We first record the binary aggregation needed at low precision.

Lemma 12 (Majority amplification of weak binary symmetric channels)**.** *Fix* $a > 0$ *and a finite* $\bar{B} > 0$*. There is a constant* $0 < C(a, \bar{B}) < \infty$ *such that the following holds. Let* $0 < \varepsilon \le 1/4$*,* $0 < B \le \bar{B}$*, and let* n *be the smallest positive odd integer at least* $C(a, \bar{B})/B$*. Let* $\Theta \in \{\pm 1\}$ *and, conditionally on* Θ*, let* $Z_1, \dots, Z_n \in \{\pm 1\}$ *be independent with*

$$\mathbb{E}[Z_i \mid \Theta] = \rho_B(\varepsilon)\Theta, \qquad \rho_B(\varepsilon) = 2\Phi(a\varepsilon\sqrt{B}) - 1. \tag{165}$$

The majority vote is an output of a binary symmetric channel. Its correlation with Θ *is at least* 2ε*.*

Proof. Write $n = 2m + 1$ and let $g_n(\rho)$ be the correlation of majority applied to n independent bits of correlation ρ. Direct differentiation of the binomial law gives

$$g'_n(\rho) = n\binom{2m}{m}\left(\frac{1-\rho^2}{4}\right)^m. \tag{166}$$

Hence g_n is increasing and concave on $[0, 1]$. The map $\varepsilon \mapsto \rho_B(\varepsilon)$ is also increasing and concave, so their composition $G(\varepsilon) = g_n(\rho_B(\varepsilon))$ is concave and $G(0) = 0$. The continuous extension at zero makes

$$c_{a,\bar{B}} := \inf_{0 < B \le \bar{B}} \frac{2\Phi(a\sqrt{B}/4) - 1}{\sqrt{B}} > 0. \tag{167}$$

Choose $C(a, \bar{B})$ so that

$$n\rho_B(1/4)^2 \ge 2\log 4. \tag{168}$$

This choice is uniform because $\rho_B(1/4) \ge c_{a,\bar{B}}\sqrt{B}$ and $n \ge C(a, \bar{B})/B$. Hoeffding's inequality makes the majority error at most $1/4$, and therefore $G(1/4) \ge 1/2$. Concavity now gives $G(\varepsilon) \ge 4\varepsilon G(1/4) \ge 2\varepsilon$ for every $0 \le \varepsilon \le 1/4$. □

Proof of Corollary 11. For a predictable action $a = (x, \sigma, b, \tau)$, coordinatewise thresholding is a parameter-independent Markov kernel from the AWGN response to $\{0, 1\}^d$. An adaptive comparator protocol can therefore be simulated recursively from the AWGN oracle: query according to its virtual bit history, threshold the Gaussian response, and reveal only the resulting bits to the virtual protocol. The transcript, stopping time, and output laws agree for every parameter. Thus

$$T^*_{\mathsf{a}}(d, \varepsilon, q, B, \delta) \le T^*_{\mathsf{a},\mathrm{cmp}}(d, \varepsilon, q, B, \delta), \tag{169}$$

and all three AWGN lower bounds transfer by Blackwell data processing.

Precision B also simulates any requested $b < B$ in the comparator model. Before the call, add a fresh coordinatewise threshold dither $(b^{-1} - B^{-1})^{1/2}G'$ with $G' \sim \mathcal{N}(0, I_d)$ to τ and do not reveal G' to the virtual protocol. Comparing a precision-B response with this randomized threshold has exactly the law of comparing a precision-b response with τ.

For the Fourier upper bound, choose the public midpoint (35) and orient the bit using the known sign of the gap. For the spatial upper bound, use threshold zero at $x = 0$. Equations (34) and (41) give a

matched BSC of correlation at least $2\Phi(a_q\varepsilon\sqrt{B}) - 1$ in both cases. In the low precision construction, take the smallest odd $n \geq C_q/B$, invoke Lemma 12 with $\bar{B} = b_0(q)$, apply majority, degrade to correlation exactly 2ε, and apply (20). This costs $O_q(B^{-1})$ calls per coordinate. The thinning and block constructions already use one midpoint threshold per active coordinate followed by the same independent degradation and synthesis kernel; their proofs in Appendices H and I therefore apply without further change. This proves all three upper rates and (163). Each call returns a vector of d bits. The construction uses only coordinatewise thresholding after AWGN. □

M CUMULATIVE PRECISION AND SCALARIZATION

A protocol without a precision cap allows any finite predictable $b_t \in (0, \infty)$, with the stopping-time and output-kernel requirements of Section 2.

Corollary 13 (Query cost and cumulative precision)**.** *Fix* $\mathsf{a} \in \{\mathsf{F}, \mathsf{S}\}$ *under the assumptions of Theorem 1, and set* $\delta = \kappa\alpha$*. For a legal protocol without a precision cap* $\mathcal{A}$*, let*

$$N(\mathcal{A}) := \mathbb{E}T, \qquad P(\mathcal{A}) := \mathbb{E}\sum_{t\leq T} b_t, \qquad \bar{B}(\mathcal{A}) := P(\mathcal{A})/N(\mathcal{A}). \tag{170}$$

If $\mathcal{A}$ *is* δ*-accurate,* $0 < N(\mathcal{A}) < \infty$*, and* $P(\mathcal{A}) < \infty$*, then*

$$N(\mathcal{A}) \geq c_{q,\kappa}\mathcal{R}_d\big(\bar{B}(\mathcal{A})\big). \tag{171}$$

Conversely, for every $B > 0$ *there is a* δ*-accurate protocol* $\mathcal{A}$ *with* $N(\mathcal{A}) \leq C_{q,\kappa}\mathcal{R}_d(B)$ *and* $P(\mathcal{A}) = BN(\mathcal{A})$*. Consequently, for every* $\lambda > 0$*,*

$$\inf_{\substack{\mathcal{A}\text{ cap-free, legal:}\\ \mathcal{A}\ \delta\text{-accurate}}} \big\{N(\mathcal{A}) + \lambda P(\mathcal{A})\big\} \asymp_{q,\kappa} \inf_{B>0}(1+\lambda B)\mathcal{R}_d(B). \tag{172}$$

For each cost weight $\lambda > 0$*, a suitable precision* B*, chosen for that weight and held fixed across queries, attains the optimal order of* $N + \lambda P$*.*

We prove Corollary 13. Fix a protocol without a precision cap $\mathcal{A}$ and abbreviate $N = N(\mathcal{A})$, $P = P(\mathcal{A})$, and $\bar{B} = \bar{B}(\mathcal{A})$. Define the finite measure on $(0, \infty)$

$$\Lambda(A) = \mathbb{E}\sum_{t\leq T} \mathbf{1}\{b_t \in A\}. \tag{173}$$

Its mass is N and its first moment is P. The functions

$$f(b) = \log(1 + C_q\varepsilon\sqrt{b}), \qquad g_L(b) = \min\left\{M, 1 + \frac{\log(1+\varepsilon\sqrt{b})}{L}\right\} \tag{174}$$

are increasing and concave. If $N > 0$, Jensen's inequality for the normalized occupation measure gives

$$\int f\,d\Lambda \leq Nf(P/N), \qquad \int g_L\,d\Lambda \leq Ng_L(P/N). \tag{175}$$

For g_L, concavity is preserved because the increasing concave logarithmic branch meets a horizontal cap, so its slope can only drop to zero. The energy law gives $P \geq c_{q,\kappa}d$. Combining this fact and (175) with the coordinatewise and blockwise bounds yields the cap-B lower-bound calculation with B replaced by $\bar{B} = P/N$. Hence

$$N \geq c_{q,\kappa}\mathcal{R}_d(\bar{B}). \tag{176}$$

The target witness also rules out $N = 0$ in the stated error regime.

Every upper construction uses precision B on each realized call. Thus its expected cumulative precision is exactly $P = BN$, including randomized zero-call branches, and Theorem 1 gives the reverse parametric bound. Finally, for $\lambda > 0$, any protocol with finite scalarized cost has finite P and

$$N + \lambda P = N(1 + \lambda\bar{B}) \geq c_{q,\kappa}(1 + \lambda\bar{B})\mathcal{R}_d(\bar{B}). \tag{177}$$

Taking the infimum over $\bar{B}$ proves the lower bound in (172). For each $\lambda > 0$, choosing a suitable precision B for that weight and using it on every realized call gives the matching upper bound. The optimization ranges over all $B > 0$, including repetition at low precision and saturation of the effective block length at high precision.

N ALTERNATIVE OBSERVATION INTERFACES

Table 2 compares three observation interfaces. The main model adds independent Gaussian noise to the normalized score. Its precision parameter is comparable across smoothing scales, and independent observations accumulate precision under repetition.

Table 2: Score interfaces and their roles.

Interface	Response	Useful feature	Role
Normalized AWGN	$\sigma s_{\theta,\sigma}(x) + b^{-1/2}G$	Inverse noise variance with a common normalization across scales	Main model; independent Gaussian noise at each query
Fixed-range quantization	$Q(\sigma s_{\theta,\sigma}(x))$	explicit finite representation	Sensitivity depends on range and encoder; Liu (2026) analyzes a broader adaptive finite-bit model
General observation channel	$Y \sim K_{\sigma,\lambda}(\cdot \mid s_{\theta,\sigma}(x))$	Comparison of experiments and data processing	Bounds on the observation channel connect output accuracy to query cost

The results below vary one feature of the observation model at a time. They quantify the roles of a shared smoothing scale, geometric spacing, score normalization, and the precision cap in the Fourier construction.

Proposition 14 (Exact one-query simulation under relaxed observation models)**.** *For the Fourier construction* (2)*, each of the following four modifications, considered separately, admits an exact one-query simulator with independent response noise across coordinates:*

1. *a query chooses a different smoothing scale in every coordinate;*
2. *the frequency grid is dense, specifically $q_d = e^{1/d}$;*
3. *the oracle returns $Y = s_{\theta,\sigma}(x) + \nu G$, with fresh independent $G \sim \mathcal{N}(0, I_d)$ and noise standard deviation $\nu > 0$, and only the raw precision ν^{-2} is capped; or*
4. *normalized precision has no finite per-call cap.*

In cases 1 and 3, a precision bound depending only on the original fixed q suffices; in case 2, a universal constant precision suffices. In case 4, a finite precision depending on d and q suffices for exact one-query simulation; a noiseless score response is not required.

Proof of Proposition 14. We first record a common one-query construction. Suppose one vector response has independent coordinates and, after subtracting a known midpoint, the two means in coordinate j have gap at least $2a\varepsilon$, where $a > 0$ is independent of j, d, ε. Midpoint thresholding at precision b gives a binary channel $Z_j \in \{\pm 1\}$ with

$$\mathbb{E}[Z_j \mid \theta_j] = \rho_j \theta_j, \qquad \rho_j \geq 2\Phi(a\varepsilon\sqrt{b}) - 1. \tag{178}$$

The right side is concave in $\varepsilon \geq 0$ and vanishes at zero. Therefore a sufficiently large constant $b = b(a)$ makes it at least 2ε simultaneously for $0 < \varepsilon \leq 1/4$. Independently degrade every coordinate to correlation exactly 2ε and output from

$$R_j(dx \mid z) = \phi(x)\big(1 + \tfrac{1}{2} z h_j(x)\big)\, dx. \tag{179}$$

This is a probability kernel because $\mathbb{E}_\phi h_j = 0$ and $\|h_j\|_\infty < 2$, and averaging over Z_j gives exactly $\phi(x)(1+\varepsilon\theta_j h_j(x))dx$. Independence across response noises and the auxiliary degradations produces the full product target in one call.

For coordinatewise scales, choose

$$\lambda_j := \frac{\sigma_j}{\sqrt{1+\sigma_j^2}} = \frac{1}{k_j}. \tag{180}$$

Then

$$A_j(\sigma_j) = e^{-1/2}\sqrt{1-k_j^{-2}} \geq e^{-1/2}\sqrt{1-q^{-2}} =: a_q > 0. \tag{181}$$

Choose each query coordinate at a phase with sine equal to one. The exact gap (34) is at least $2a_q\varepsilon$, so the common construction applies at constant normalized precision.

For a dense frequency grid, take $q_d = e^{1/d}$, so $k_j = e^{j/d} \in [e^{1/d}, e]$, and use the single scalar $\lambda = e^{-1}$. Then

$$A_j = \sqrt{1-e^{-2}}\, g(k_j/e), \qquad g(u) = ue^{-u^2/2}, \tag{182}$$

and $\inf_{u\in[e^{-1},1]} g(u) > 0$. A single common scale thus gives the uniform gap required by (178).

For the raw-score oracle, fix the original geometric grid and choose $\lambda = 1/k_d$. The raw score gap is the normalized gap divided by σ, while

$$\frac{A_j(\sigma)}{\sigma} = k_j(1-\lambda^2)\exp\{-\tfrac{1}{2}(k_j\lambda)^2\} \geq q(1-q^{-2})e^{-1/2}. \tag{183}$$

Thus a constant cap on raw precision ν^{-2} supplies every coordinate with a constant multiple of ε in the same call. Under this observation model, the common construction gives an exact sample with one query.

Finally, retain a common scalar scale and fixed q, but remove the cap on b. At any fixed $\sigma > 0$, all finitely many $A_j(\sigma)$ are positive. Choosing phases with sine equal to one and taking b large enough as a function of $\min_{j\leq d} A_j(\sigma)$ makes every thresholded coordinate at least as strong as correlation 2ε. The same garbling is exact. With an exact real response, the sign of every nonzero centered mean shift reveals the corresponding bit directly. □

O COMPARISON OF OBSERVATION MODELS AND RESULTS

Table 3 compares the observation model, charged resource, statistical task, and proof method. These distinctions locate the present comparison of equivalent targets: preserving a target experiment leaves open how efficiently a restricted score interface can supply the dependence needed to generate it.

Table 3: Related results compared by observation model, statistical task, and main method.

Work	**Observation model and cost**	**Statistical task**	**Main method**
Xun & Price (2026)	Globally L^p-accurate smoothed-score fields; calls	Sampling from approximate smoothed scores	Planted-versus-null testing over informative scale intervals
Liu (2026)	Adaptive finite-bit smoothed-score transcripts; bits	Gaussian sampling from smoothed scores	Channel-synthesis converse from channel coding
Acharya et al. (2023)	General interactive channels; information constraints	High-dimensional parameter estimation	Interactive information contraction
Chatterji et al. (2022)	Fresh stochastic gradients; calls	Sampling with a stochastic gradient oracle	Le Cam comparison and adjacent-pair KL
Chen et al. (2026)	Stochastic gradients; calls	High-accuracy log-concave sampling	First-order rejection and proximal sampling; information-theoretic converse
Gu et al. (2026)	Deterministic inexact scores; score error class	Tractability of sampling with score approximation	Indistinguishable Gaussian targets under one approximate score field
Multiscale and controlled sensing	Linear or controlled observations; samples or sensing effort	Signal recovery and hypothesis testing	Local coherence, structured sampling, and controlled sensing
This work	Normalized scores with a shared scale and Gaussian noise; calls and precision	Sampling overlapping product targets at prescribed TV accuracy	Blockwise TV converse and randomized activation; equivalent-layout comparison

Multiscale and controlled sensing: Adcock et al. (2017), Krahmer & Ward (2014), Castro & Tánczos (2017), and Nitinawarat et al. (2013).

Theorem 1 fixes $q > 1$ and the relative accuracy δ/α for the two product families. Its blockwise bounds quantify the gain from query sharing under a shared scale and a finite precision cap. Theorem 3 then preserves the full target experiment while varying the scale layout. Proposition 14 shows how four individual changes to the Fourier observation model permit exact one-query generation; Corollary 13 instead retains the shared scale and charges cumulative precision. Together these results identify how the allowed observation resources enter the sampling cost.