Why Learning Rediscovers the Closed-Form Diagonal Regularizer

Jeahn Han **Pyojin Kim***
Gwangju Institute of Science and Technology
jeahnhaan@gm.gist.ac.kr, pjinkim@gist.ac.kr
https://ugrm.github.io/diagonal-regularizer/

Abstract

Recovering modal amplitudes from a few sensors requires regularization, and the shape of the penalty decides how strongly each mode is shrunk. The best shape depends on the truncation noise that discarded modes add to every measurement, which is unobserved and could be room-specific. We identify a diagonal saturation principle: for a single snapshot, isotropic truncation noise makes the Bayes-optimal Tikhonov shape a closed-form power law $\Gamma_k \propto \lambda_k^{|s|}$, set by the prior alone and independent of the domain. The off-diagonal anisotropy equals the Herfindahl index H of the discarded-mode power exactly, Weyl's eigenvalue counting law fixes the number of discarded modes and hence H, and Berry's random-wave conjecture enters only through a diagonal term D, which we measure directly. Together they predict an approximately flat loss landscape over the penalty exponent, which leaves a learned diagonal regularizer little to gain from tuning the exponent per room. On simulated acoustic rooms the closed form stays within 5.83% of a per-room power-law oracle across observation windows, and three learned diagonal architectures track its error, the closest within a fraction of a point and none more than 3 pp above it, while learning qualitatively different spectra. In the truncation regime no tested method consistently improves on the closed form, including Learned Iterative Ridge, which couples modes across layers. The framework extends to heat diffusion through a known Green's-function correction with no new free parameters.

1 Introduction

Modal inverse problems on bounded domains arise in acoustic, thermal, and electromagnetic PDEs. A sensor array records finitely many numbers, so every estimator keeps finitely many modes, and the discarded modes enter each measurement as *truncation noise*. In room acoustics, our running example, a pressure field is a weighted sum of the room's eigenmodes whose weights are recovered from a few microphones. With $K{=}50$ retained modes and $M{=}8$ microphones, one snapshot gives 8 equations for 100 unknowns, and a typical room has over 300 modes in total, so truncation noise dominates the error budget.

The shape question. Regularization makes the problem solvable by penalizing amplitudes the data cannot constrain [Stuart, 2010, Kaipio and Somersalo, 2005]. The penalty has a scalar strength α, whose selection is well studied [Hansen, 1992, Golub et al., 1979, Morozov, 1966], and a diagonal shape Γ that decides how strongly each mode is shrunk. In physical systems the energy of mode k decays as λ_k^{-s}, where λ_k is the eigenvalue (roughly frequency squared) and s is set by the excitation; in our simulated data, excited by a broadband click, the population value measured from the modal

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

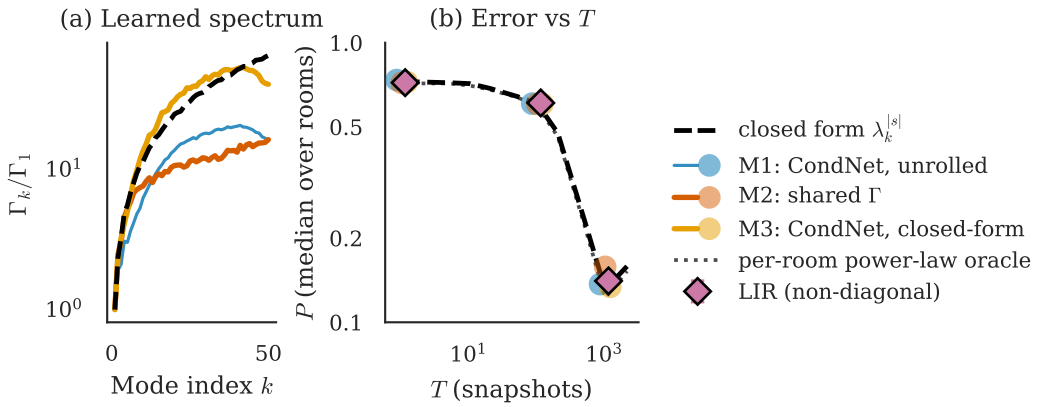


Figure 1: **Learned diagonal regularizers learn different spectra yet reach the closed form's error.** M3 follows the closed form, M1 and M2 learn flatter spectra that penalize high modes less (a, single runs). All three track the closed form's error, M2 slightly above at the longest window, and the mode-coupling LIR tracks it too (b). Lower P is better, 1 means no better than guessing zero. This is the saturation of Section 6, which Section 6.1 finds outside the diagonal family as well.

time series is $|s| = 1.13$ (Appendix C.1). The natural penalty is a power law $\Gamma_k = \lambda_k^p$: p=0 penalizes all modes equally, and p=2 suppresses high frequencies strongly. Which p is best, and does it depend on the room?

Why the noise decides. The optimal p depends on how signal energy decays across modes, which is easy to measure, and on how truncation error spreads across the retained modes, which is not. If the noise has room-specific structure, a learned regularizer, such as one trained by deep unrolling [Gregor and LeCun, 2010, Adler and Öktem, 2018, Aggarwal et al., 2018], could exploit it; if the noise is *isotropic*, then for a single snapshot $p^\star = |s|$, set by the signal alone. Our key insight is that the discarded modes are too numerous to leave exploitable structure: Weyl's eigenvalue counting law [Weyl, 1912, Ivrii, 2016] supplies hundreds in the median room (Table 4), so the truncation noise concentrates around its isotropic average. Berry's random-wave conjecture [Berry, 1977], which says that high-frequency eigenmodes of generic domains behave like random fields, controls only one term of the anisotropy, and we measure that term rather than assume it (Appendix A.2). Saturation is therefore a property of the truncation noise, not of the models: without discarded modes, freedom in the shape pays again (Appendix F.4).

Findings. Across rooms ranging from triangles to decagons, none of the per-mode regularizers we test, including regularizers learned by deep unrolling, gains robustly over the closed form within the diagonal family in the truncation regime (Figure 1). Learned Iterative Ridge, which couples modes across layers and so lies outside the diagonal family, does not consistently improve on it either (Section 6.1); the claim concerns the methods tested, not learning in general. In practice, $\Gamma_k = \lambda_k^{|s|}$ is set once per excitation regime and used for every room: what transfers is not the measured value of $|s|$ but the room-independence of the shape, which persists when K, M, and the excitation exponent change (Appendix D.4).

Contributions.

1. **Theory.** Under isotropic truncation noise, $p^\star = |s|$ exactly at T=1 (Proposition 1); Weyl's law and Berry's conjecture play separate, measurable roles in the isotropy argument (Section 4). Multi-snapshot results are empirical.
2. **Verification.** On 187 in-scope rooms ($K_{\text{total}} > K$), the median relative cost of p=$|s|$ against per-room power-law oracle tuning is at most 5.83% at every observation window, and at T=1000 two rooms in three lose less than 1.1 pp absolute (Section 5); the 10 rooms with $K_{\text{total}} \leq K$ are reported separately.

3. **Diagonal saturation.** Three learned diagonal architectures track the closed form's P-vs-T curve, all within 3 pp and M3 closest (Table 3), while learning qualitatively different spectra; $p{=}|s|$ is a principled choice inside a wide near-optimal basin (Section 6).
4. **Cross-PDE consistency.** For heat diffusion the theory predicts $\Gamma_k \propto \lambda_k^s e^{c\lambda_k}$ with $c = 2\kappa t$; per-room fits recover the rate to within a few percent (Section 7), a consistency check, not physical validation.

2 Related Work

Regularizer strength and shape. Tikhonov regularization has two parts: a scalar strength α and a matrix Γ that decides which modes to penalize [Stuart, 2010, Kaipio and Somersalo, 2005]. Strength selection is mature, as the L-curve [Hansen, 1992], generalized cross-validation [Golub et al., 1979], the discrepancy principle [Morozov, 1966], and Bayesian posterior contraction [Cavalier, 2008, Knapik et al., 2011] choose α reliably. For the shape, truncated SVD and Landweber iteration filter the operator's singular values [Engl et al., 1996, Hansen, 1998]; Pinsker's estimator gives minimax per-coordinate shrinkage [Pinsker, 1980, Johnstone, 2002], hierarchical Bayesian models derive per-component weights from data [Calvetti and Somersalo, 2025], and spectral Bayesian methods recover analogous structure on manifolds [Durastanti, 2026]; these three require training data or an assumed smoothness class. Alberti et al. [2021] prove that the MSE-optimal shape depends only on the signal covariance Σ_x, with $O(1/\sqrt{m})$ generalization bounds, and Leong et al. [2024] confirm this geometrically. In modal problems the truncation noise is approximately isotropic, which reduces the shape question to Σ_x, and a power-law excitation prior reduces Σ_x to one scalar, $|s|$.

Learned regularizers. Algorithm unrolling [Gregor and LeCun, 2010] turns iterative solvers into trainable networks and underlies many learned regularizers [Adler and Öktem, 2018, Sun et al., 2016, Aggarwal et al., 2018]. The learned answer is often the classical one: learned parameters converge to variational solutions [Kofler et al., 2023], bilevel optimization reduces to hyperparameter tuning [Kunisch and Pock, 2013], an untrained CNN matches a trained denoiser [Ulyanov et al., 2018], and plug-and-play regularizers collapse to the denoiser's spectral penalty [Hurault et al., 2022]. Instabilities in learned methods trace to an accuracy–stability tradeoff [Antun et al., 2020, Gottschling et al., 2025] that no algorithm can circumvent in general [Colbrook et al., 2022]. We give one concrete reason for this pattern in modal problems: the loss landscape over the penalty exponent is flat.

Berry's conjecture and Weyl's law. Berry's conjecture states that high-frequency eigenmodes of generic domains look like random waves [Berry, 1977]; quantum ergodicity supports it [Shnirel'man, 1974, Zelditch, 2005], and it fails for scarred states [Heller, 1984] and some symmetric geometries [Hassell and Hillairet, 2010]. Weyl's law counts such modes: in 2D the eigenvalue count grows linearly with λ [Weyl, 1912, Ivrii, 2016], a fact used in wave-chaotic compressive sensing [Del Hougne et al., 2020] and cavity acoustics [Tanner and Søndergaard, 2007]. We are not aware of prior work that uses either result to choose a regularizer.

Sound-field reconstruction. Sparse-microphone reconstruction uses Bayesian methods [Schmid et al., 2021], compressive sensing [Antonello et al., 2017], spherical arrays [Fernandez-Grande, 2016], physics-informed networks that penalize the PDE residual [Karakonstantis et al., 2024], and Matérn-kernel Gaussian-process priors, whose regularity parameter plays the role of $|s|$ [Rasmussen, 2003]. Sensor-placement methods ask where to put the microphones [Krause et al., 2008, Alexanderian et al., 2014]; we ask which penalty shape to use and whether it depends on the room.

Backward heat problems. The backward heat equation is a standard ill-posed problem: mode amplitudes decay as $e^{-\kappa\lambda_k t}$, so recovering them amplifies noise exponentially [Beck et al., 1985, Kaipio and Fox, 2011]. Minimax rates [Knapik et al., 2013] and variational source conditions [Hohage and Weidling, 2017] give the scaling but not the regularizer, and they do not involve κ or t.

3 Problem Setup

We consider modal inverse problems on a bounded 2D domain $\Omega \subset \mathbb{R}^2$ with eigenpairs (λ_k, φ_k) of the Laplacian. The framework is instantiated on the acoustic wave equation and validated on

187 random convex polygons, with eigenpairs computed by FEM and pressure fields simulated by a finite-difference time-domain solver (Section 5, Appendix C.8); the heat equation serves as a cross-PDE check (Section 7). The state at time t admits the modal expansion

$$u(x,t) = \sum_{k=1}^{\infty} a_k(t)\,\varphi_k(x), \tag{1}$$

with eigenfunctions L^2-normalized so that $\int_\Omega \varphi_k^2 = 1$ and $\mathbb{E}_x[\varphi_k^2(x)] = 1/|\Omega|$ for x uniform in Ω. Each eigenfunction is a spatial pattern, and the amplitude $a_k(t)$ says how much of that pattern is present at time t. Higher modes typically carry less energy: the variance of a_k decays as a power law in λ_k with an exponent s that depends on the excitation statistics and not on the room geometry. For the acoustic wave equation with uniform damping γ, each amplitude is a damped sinusoid,

$$a_k(t) = e^{-\gamma t}\big[c_k\cos(\omega_k t) + \beta_k \sin(\omega_k t)\big], \qquad \omega_k = c_s\sqrt{\lambda_k}. \tag{2}$$

A damping rate shared by every mode distinguishes acoustics from heat diffusion (Section 7). We write $\sigma_{a,k}^2$ for the variance of the random initial conditions (c_k, β_k).

Truncation and observation model. We retain $K = 50$ modes and observe through M microphones at positions $\{x_m\}_{m=1}^M$:

$$y_m(t) = \underbrace{\sum_{k=1}^{K} a_k(t)\,\varphi_k(x_m)}_{\text{retained signal}} + \underbrace{\sum_{n=K+1}^{K_{\text{total}}} a_n(t)\,\varphi_n(x_m)}_{\text{truncation noise } \eta_m^{\text{trunc}}(t)}. \tag{3}$$

Stacking M microphones over T snapshots gives

$$\tilde{\mathbf{y}} = \tilde{\Phi}\,\mathbf{a}_0 + \tilde{\boldsymbol{\eta}}, \qquad \tilde{\mathbf{y}} \in \mathbb{R}^{MT}, \quad \tilde{\Phi} \in \mathbb{R}^{MT\times 2K}, \tag{4}$$

where $\tilde{\Phi}$ combines the spatial measurement matrix $\Phi_{mk} = \varphi_k(x_m)$ with the temporal basis of (2), and $\mathbf{a}_0$ collects the initial amplitudes $(c_1, \beta_1, \ldots, c_K, \beta_K)$.

Tikhonov estimator. The estimate of $\mathbf{a}_0$ is the penalized least-squares solution

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}}\big\{\|\tilde{\mathbf{y}} - \tilde{\Phi}\,\mathbf{a}\|^2 + \alpha\,\mathbf{a}^\top\Gamma\,\mathbf{a}\big\} = \big(\tilde{\Phi}^\top\tilde{\Phi} + \alpha\,\Gamma\big)^{-1}\tilde{\Phi}^\top\tilde{\mathbf{y}}, \tag{5}$$

where $\alpha > 0$ is the strength and $\Gamma \succ 0$ is a diagonal matrix of per-mode penalties. Quality is the normalized modal MSE

$$P_{\text{modal}} = \frac{\sum_k \mathbb{E}[(\hat{a}_k - a_{0,k})^2]}{\sum_k \operatorname{Var}(a_{0,k})}, \tag{6}$$

with expectation and variance over the evaluated snapshots, so $P_{\text{modal}} = 0$ is perfect reconstruction and 1 matches predicting each mode's mean, which for the zero-mean acoustic amplitudes is guessing zero; we write P hereafter.

Terminology. *Ridge* means $\Gamma = I$ (p=0); the *closed form* means $\Gamma_k = \lambda_k^{|s|}$; the *per-room power-law oracle* means the best $(p^\star, \alpha^\star)$ of the scalar sweep $\Gamma_k = \lambda_k^p$ for a given room. Every oracle in this paper is of this type; no experiment searches over unrestricted per-mode weights.

4 Why the Prior Sets the Penalty Shape

For a single snapshot (T=1) with isotropic truncation noise, the optimal regularizer has the form $\Gamma_k^\star \propto \lambda_k^s$, where $s > 0$ is the prior spectral decay rate.[2] The argument has three steps: a linear-estimation calculation shows that isotropic noise fixes the shape, an exact identity controls the off-diagonal anisotropy without Berry's conjecture, and Weyl's mode count sets the size of that anisotropy. Appendix A gives the details.

[2]We write $s > 0$ in theoretical statements and $|s|$ in experiments, where the exponent is the absolute value of a fitted slope.

4.1 Step 1: isotropic noise fixes the shape

The truncation noise $\boldsymbol{\eta}_{\mathrm{trunc}}(t) = \sum_{n>K} a_n(t)\,\boldsymbol{\varphi}_n$ has covariance (at $t = 0$; uniform damping cancels)

$$R_{\mathrm{trunc}} = \sum_{n>K} \sigma_{a,n}^2\,\boldsymbol{\varphi}_n\boldsymbol{\varphi}_n^\top = \sigma_{\mathrm{trunc}}^2\big(I_M + E\big), \tag{7}$$

where $\sigma_{\mathrm{trunc}}^2 = M^{-1}\sum_{n>K}\sigma_{a,n}^2$ is the average noise power and E is the deviation from isotropy. When $E = 0$, the prior alone sets the shape:

Proposition 1 (Isotropy ⇒ power-law regularization, single snapshot)**.** *Let $T = 1$, so that $\mathbf{y} = \Phi\,\mathbf{a} + \boldsymbol{\eta}$ with $\Phi \in \mathbb{R}^{M\times K}$, $\Phi_{mk} = \varphi_k(x_m)$, and $\mathbf{a} \in \mathbb{R}^K$ the modal amplitudes at the observed snapshot. Assume $\mathbb{E}[\mathbf{a}] = 0$, $\mathbb{E}[\boldsymbol{\eta}] = 0$, $\mathrm{Cov}(\boldsymbol{\eta}) = \sigma^2 I_M$, $\mathrm{Cov}(\mathbf{a},\boldsymbol{\eta}) = 0$, and $\mathrm{Cov}(\mathbf{a}) = \Sigma_{\mathbf{a}}$ diagonal with $\Sigma_{kk} = c\,\lambda_k^{-s}$. Then the linear minimum-MSE estimator of $\mathbf{a}$ is the Tikhonov estimator (5) with $\alpha\,\Gamma = \sigma^2\,\Sigma_{\mathbf{a}}^{-1}$, so the optimal shape is $\Gamma_{kk}^\star \propto \lambda_k^s$, i.e., $p^\star = s$, independent of the domain. If $(\mathbf{a},\boldsymbol{\eta})$ are moreover jointly Gaussian, this estimator is MMSE-optimal among all estimators.*[3]

Proof. Among linear estimators, the mean-squared error depends on $(\mathbf{a},\boldsymbol{\eta})$ only through the second moments above, and its minimizer is $\hat{\mathbf{a}} = (\Phi^\top\Phi + \sigma^2\Sigma_{\mathbf{a}}^{-1})^{-1}\Phi^\top\mathbf{y}$ (Appendix A.1). Comparing with (5) gives $\alpha\,\Gamma = \sigma^2\,\Sigma_{\mathbf{a}}^{-1}$. Since $\Sigma_{\mathbf{a}}^{-1}$ is diagonal with entries $c^{-1}\lambda_k^s$, the shape $\Gamma_{kk} \propto \lambda_k^s$ is set by the prior, and the noise level sets only α. Under joint Gaussianity the same estimator is the posterior mean, hence MMSE among all estimators. □

Modes with less energy are penalized more. Alberti et al. [2021] show that the optimal shape requires Σ_x; in modal problems physics gives Σ_x analytically, so the remaining question is whether the truncation noise is isotropic.

Remark 1 (Departures from isotropy)**.** *Once isotropy is lost, the noise-whitening term is non-diagonal and the Bayes-linear estimator leaves the diagonal Tikhonov family, so Proposition 1 is exact only at $E = 0$. Write the covariance as $\sigma^2(I_M + \Delta)$ with Δ trace-free, let $\widetilde{P}_\Delta(p) = \min_{\alpha>0} P(p,\alpha;\Delta)$ be the profiled population risk, and let p_0, p_Δ minimize $\widetilde{P}_0$, $\widetilde{P}_\Delta$. If $\partial_p\widetilde{P}_\Delta$ is continuously differentiable jointly in (p,Δ) and p_0 is an interior minimum with $\partial_p^2\widetilde{P}_0(p_0) > 0$, the implicit function theorem gives $p_\Delta - p_0 = O(\|\Delta\|_{\mathrm{op}})$ and $\widetilde{P}_\Delta(p_0) - \widetilde{P}_\Delta(p_\Delta) = O(\|\Delta\|_{\mathrm{op}}^2)$; the constant depends on the room through $\partial_p^2\widetilde{P}_0(p_0)$, so this is a local sensitivity statement, not a uniform guarantee. Structured sensor arrays are outside this analysis (Section 8).*

A controlled sweep through and beyond the most anisotropic rooms observed leaves the median cost of the fixed exponent essentially unchanged at $T \le 100$; at T=1000 it stays small up to the mean anisotropy and rises by a few tenths of a point at the strongest levels (Appendix A.5, Table 6).

4.2 Step 2: the off-diagonal anisotropy is exactly H

Is the truncation noise isotropic? The covariance (7) is a weighted sum of rank-one matrices $\boldsymbol{\varphi}_n\boldsymbol{\varphi}_n^\top$, one per discarded mode, so the answer depends on the statistics of $\boldsymbol{\varphi}_n = [\varphi_n(x_1),\dots,\varphi_n(x_M)]^\top$ across modes. Berry's random-wave conjecture [Berry, 1977] predicts that high-frequency eigenfunctions of generic bounded domains behave like random superpositions of plane waves. We use only a sensor-level version:

Conjecture 1 (Sensor-averaged eigenfunction independence)**.** *For generic sensor placements $\{x_m\}_{m=1}^M$ drawn uniformly in $\Omega \subset \mathbb{R}^2$, the cross-correlations*

$$C_{kn} := \frac{1}{M}\sum_{m=1}^{M}\varphi_k(x_m)\,\varphi_n(x_m) \tag{8}$$

satisfy $\mathbb{E}[C_{kn}^2] \approx 1/M$ for $k \neq n$ under the discrete sensor normalization $\sum_m \varphi_n(x_m)^2 \approx 1$ (Appendix A.2).

[3] The multi-snapshot model ($T > 1$) has temporally correlated truncation noise (Appendix A.4); we state no $T > 1$ corollary, and every multi-snapshot result in Sections 5 and 6 is empirical. At T=1 the stacked model (4) has $2K$ unknowns with prior $\Sigma_{\mathbf{a}} \otimes I_2$, and the proposition applies verbatim with $\tilde{\Phi}$ for Φ.

Table 1: **Only the first of the three layers is a theorem.** The exact result holds for a single snapshot, whereas the saturation result is empirical and restricted to the tested diagonal methods in the truncation regime. The no-truncation control of Section F.4, where shape freedom pays again, is the explicit exception to the second layer.

Layer	Claim
Exact theory	Under exact isotropy at T=1, $\Gamma_k \propto \lambda_k^{\lvert s\rvert}$ is the optimal Tikhonov penalty shape (Proposition 1).
Tested diagonal saturation	In the truncation regime, the closed form is near the per-room power-law oracle and is not consistently improved by M1–M3, the learned diagonal models tested (Sections 5 and 6).
Beyond the diagonal family	LIR, which couples modes across layers, does not reach the oracle and does not consistently improve on the closed form (Section 6.1).

The off-diagonal part of the anisotropy does not need the conjecture. Let $\psi_n = \sqrt{|\Omega|}\,\varphi_n$ and $w_n = \sigma_{a,n}^2 / \sum_{m>K} \sigma_{a,m}^2$, so that $E_{ij} = \sum_{n>K} w_n\, \psi_n(x_i)\psi_n(x_j) - \delta_{ij}$ (population normalization, Appendix A.2). For sensors drawn independently and uniformly in Ω, the off-diagonal entries satisfy, exactly,

$$\mathbb{E}[E_{ij}^2] \;=\; \frac{\sum_{n>K}(\sigma_{a,n}^2)^2}{\left(\sum_{n>K}\sigma_{a,n}^2\right)^2} \;=\; H\,, \qquad i \neq j, \tag{9}$$

where H is the Herfindahl index of the noise power across discarded modes: the probability that two randomly drawn units of noise power come from the same mode. The identity uses only L^2-orthonormality and the independence of the sensor draws. The diagonal entries contribute $D = \mathrm{Var}_x\big[\sum_{n>K} w_n \psi_n(x)^2\big]$, the spatial variance of the discarded modes' summed intensity, so that $\mathbb{E}\|E\|_F^2 = M(M{-}1)H + MD$ exactly and $\mathbb{E}\|E\|_{\mathrm{op}}^2 \le \mathbb{E}\|E\|_F^2$. Berry's conjecture enters only through $D \approx 2H$, and at $M = 8$ the Berry-free off-diagonal term ($56H$) is the larger of the two. Measured on the in-scope rooms, $D/(2H)$ exceeds the Berry value by a moderate factor (Table 4), which raises $\mathbb{E}\|E\|_F^2$ above the Berry prediction; the excess is cross-mode covariance of squared intensities, not heavier single-mode marginals (Appendix A.2). By Jensen, $\mathbb{E}\|E\|_{\mathrm{op}} \lesssim \sqrt{M(M{-}1)H + MD}$, which with the measured H and D at $M = 8$ is an order-of-magnitude estimate consistent with the empirical median across the in-scope rooms (Table 4).

4.3 Step 3: Weyl's law sets H

How large is H? Weyl's law [Weyl, 1912, Ivrii, 2016] states that in two dimensions

$$N(\lambda) \sim \frac{|\Omega|}{4\pi}\,\lambda, \qquad \lambda \to \infty. \tag{10}$$

For our rooms this gives hundreds of discarded modes in the median room, a few dozen in the smallest (Table 4). With prior decay $\sigma_{a,n}^2 \propto \lambda_n^{-s}$ and $|s| = 1.13$,

$$H = \frac{\sum_{n>K} \lambda_n^{-2s}}{\left(\sum_{n>K} \lambda_n^{-s}\right)^2}, \tag{11}$$

a calculation from eigenvalues and prior weights alone, which for our rooms places H far below one (Table 4). The two classical results therefore play separate roles: Weyl's law fixes the number and density of the discarded modes, which with the prior decay determine H, and Berry's conjecture governs only D. At matched K_{total}, generic rooms and rectangles (where Berry's premise fails by construction) have the same H, their $D/(2H)$ separates without overlap, and their landscape-flatness medians are nearly identical (Table 35). The residual anisotropy is moderate, and the regularizer shape is insensitive to it (Remark 1): across the in-scope rooms at T=1000, higher anisotropy even correlates with lower cost (Figure 4; Table 4), a room-size confound that vanishes once area is controlled for (Appendix D.9).

Scope of the claims. The contribution has three layers (Table 1, Appendix A.1): an exact single-snapshot theorem; an empirical saturation result for the tested diagonal methods in the truncation regime; and a learned method outside the diagonal family, which does not consistently improve either. The no-truncation control of Appendix F.4, where shape freedom pays again, is the exception to the second layer.

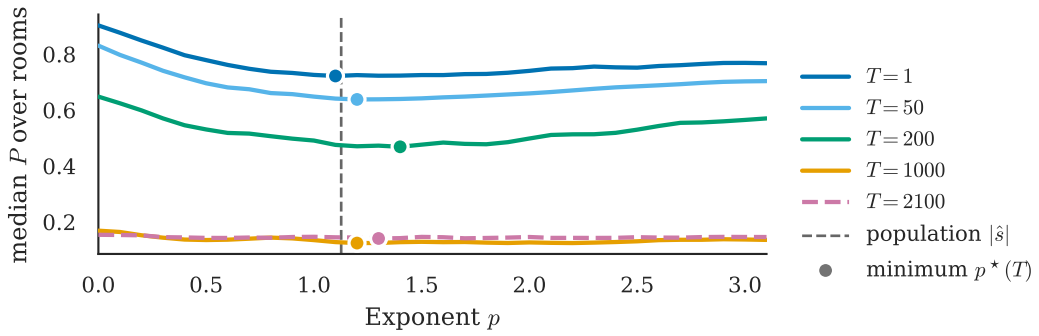


Figure 2: **The error landscape over the penalty exponent has a broad, shallow minimum that flattens as the window grows.** Each curve is the median error over the in-scope rooms for one window, where lower is better and 1 means no better than guessing zero. The minimum moves with T, but the basin widens faster than it moves, and the closed-form exponent stays inside every basin. This flatness is why $p = |s|$, exact only at $T{=}1$ by Proposition 1, costs little at longer windows.

5 Empirical Verification on Acoustic Rooms

The experiments go beyond Proposition 1: the truncation noise is only approximately isotropic, and for $T > 1$ snapshots are correlated in time. We evaluate $\Gamma_k = \lambda_k^{|s|}$ on the 187 rooms of a 197-room validation set that have truncation noise ($K_{\text{total}} > K$), with $K = 50$ and $M = 8$; the 10 rooms with $K_{\text{total}} \le K$ lie outside the regime of Proposition 1 and are reported separately (Appendix C.3). The FDTD data are excited by a broadband click, and the analysis models the excitation as $\sigma^2_{a,k} \propto \lambda_k^{-|s|}$; OLS in log space on the modal variances gives $|\hat{s}| = 1.13$ with a narrow bootstrap interval (Table 10; procedure and sensitivity in Appendix C.1). For each room the evaluation sweeps p over a 61-point grid on $[0, 6]$ and computes $P(p, T)$ at $T \in \{1, 5, 10, 20, 50, 100, 200, 500, 1000, 2100\}$, grid-searching α at each point. The pattern persists when the truncation rank, sensor count, and excitation exponent change ($K{=}100$, $M{\in}\{8, 16\}$, $|s|{=}1.29$; Appendix D.4), and under heavy-tailed (t_3) and correlated ($\rho{=}0.3$) priors, which leave $p^\star$ and the cost of the closed form unchanged in the truncation regime (Appendix F).

The landscape is flat. Figure 2 shows a broad, shallow minimum of the median $P(p)$ at every T. The per-room oracle $p^\star$ drifts upward as T grows, but the basin widens faster than the optimum moves: per-room exponents are most concentrated at intermediate windows (Table 12), and the discarded-mode count supplied by Weyl's law flattens $P(p)$ (Appendix G.4). Empirically, $p = |s|$ is therefore a principled choice inside a wide near-optimal basin rather than a sharply identified optimum: it is exact under isotropy at $T{=}1$, and the flatness explains why it costs little away from that idealization. The location is not arbitrary. Planted exponents are recovered where the landscape has curvature (Appendix D.7), a change of excitation moves $|s|$ and the formula follows (Appendix D.4), and heat diffusion requires the predicted power-law–exponential form (Section 7).

Isotropy check. Under Berry's prediction, $|\Omega|\,\varphi_k(x_m)^2$ follows a $\chi^2(1)$ distribution at uniformly random sensor positions. Pooling all rooms, the Kolmogorov–Smirnov statistic against $\chi^2(1)$ is small (Table 4); Appendix B gives per-room, boundary, and worst-room analyses.

The cost of using $|s|$. Table 2 summarizes the relative cost. At $T{=}1000$, two in-scope rooms in three lose less than 1.1 pp absolute, and the worst in-scope absolute cost over all T is several times that, in the prior-dominated regime (Table 12; per-room and per-T breakdowns in Appendices C.3 and C.2). As T grows the landscape compresses until per-room "optimal" exponents are barely defined. Section 6 asks whether a learned model finds any remaining structure.

Table 2: **Replacing per-room power-law oracle tuning with the closed form costs little at the population median, and most at long observation windows.** The relative cost δ is the extra modal error of the closed form as a fraction of the oracle error, reported as the largest median over the in-scope rooms within each regime. Table 11 gives the value at every window. This is the empirical cost of Proposition 1 beyond a single snapshot.

Regime	Largest median δ
$T \leq 50$ (prior-dominated)	0.61%
$T \leq 100$ (crossover)	1.46%
All T (maximum at $T{=}1000$)	5.83%

Table 3: **M3 matches the error of the closed form at every tested window and recovers little of the gap to the oracle, although its seeds learn different spectra.** P is the normalized modal error, lower is better, and M3 entries are mean and standard deviation over seeds. The gap is the closed form's error minus the oracle's, an upper bound on what per-room power-law tuning can recover. This is the diagonal saturation that Section 6 reports for the truncation regime.

T	$P(\lvert s\rvert)$	P_{M3}	P_{oracle}	Gap
1	0.7264	0.7245 ± 0.0021	0.7151	1.13 pp
100	0.6035	0.6051 ± 0.0027	0.5936	0.99 pp
1000	0.1308	0.1339 ± 0.0003	0.1224	0.84 pp

6 Learned Diagonal Regularizers Match the Closed Form

Proposition 1 makes $\Gamma_k \propto \lambda_k^{|s|}$ optimal within the diagonal family under exact isotropy at $T{=}1$. The gap of up to 5.83% between the closed form and the per-room power-law oracle (Table 2) leaves room for a learned diagonal regularizer to exploit per-room structure. Three architectures look for it.

Setup. All models are trained on $n{=}800$ rooms ($K{=}50$, $M{=}8$, five seeds each; specifications in Appendix D.1). M1 (4,949 parameters) maps a 10-dimensional per-mode feature vector to a diagonal Γ through a small network (CondNet) and solves the Tikhonov problem by $L{=}10$ unrolled gradient steps. M2 (70 parameters) learns one unconditional Γ shared by all rooms, with the same unrolled solver. M3 (4,930 parameters) uses the CondNet of M1 with a closed-form differentiable linear solve. The strength α is learned jointly with Γ and used at evaluation. If per-room adaptation helps, M1 and M3 should beat both M2 and the closed form.

Results. All three models converge smoothly over 500 epochs (Appendix D.2). M3 reaches P within a fraction of a percentage point of the closed form at every T, slightly below it at $T{=}1$ and slightly above at longer windows (Table 3), although its five seeds learn qualitatively different spectra (Table 15). The residual anisotropy therefore opens no gap that the tested models exploit, and the loss is flat across the shapes the seeds reach. Figure 1 shows the same saturation across architectures. At $T{=}1000$ M3's per-room exponents match theory in the median, while M2's shared exponent is about half of it; yet the median P at the two exponents is nearly the same (Table 16). For M3 this is non-identifiability on a flat basin rather than a capacity limit, since M3 recovers planted exponents where curvature exists (Appendix D.7); M2's parameterization recovers planted exponents only up to a ceiling below one (Table 16), so the basin reading is scoped to M3. M2 still converges to a power law by SGD without any physics (Table 16), and its shallower exponent costs nothing in P.

Across all three architectures, five training sizes, and three snapshot counts, every valid per-seed evaluation in the truncation regime shows $\Delta P \geq 0$ relative to the oracle (Appendix D.6). In the no-truncation control of Appendix F, outside the regime of Proposition 1, M3 checkpoints trained with the same protocol beat the best power law by a clear median margin at $T{=}100$ (Appendix F.4): saturation is a property of the truncation noise, not of the models. No room descriptor predicts $p^\star$ (Appendix D.3), and no geometric or noise diagnostic predicts the residual gap out of sample (Appendix D.9).

Fitting s per room does not help. The label-free alternative is a per-room slope s_{room} fitted to each room's spectrum. Its estimation error exceeds the standard error of the population median by an

order of magnitude (Table 10), and the deconvolved inter-room spread is effectively zero, so plugging in s_{room} adds more noise than signal: across T it is never clearly better than $|s|$ and is worse at most windows (Table 21).

6.1 Beyond the diagonal family

Can coupling modes through the full $A^\top A$ structure go below the per-room power-law oracle? Learned Iterative Ridge (LIR) runs L steps of learned gradient descent on the Tikhonov objective with per-layer (η_l, α_l, D_l), $52L$ parameters in total (Appendix D.8). At no tested depth does LIR reach the oracle (Figure 1b at L=10; Table 23). At L=10 it is slightly below the closed form at T=1 and above it at T=100 and T=1000, so no method tested, diagonal or not, consistently improves on the closed form. Under the approximate isotropy of our data the Bayes-linear estimator lies outside the diagonal family, so a non-diagonal estimator could in principle go below the oracle; LIR as trained here does not. Classical non-diagonal baselines (Wiener/LMMSE with a diagonal prior and isotropic noise, generalized Tikhonov, TSVD, early-stopped CGLS, Landweber) and per-coordinate shrinkage (Pinsker, empirical Bayes) apply fixed per-mode shrinkage profiles set by the prior and cannot improve on the closed form's shape (Appendix D.8).

7 Extension to Heat Diffusion

Does the framework depend on the wave equation? Heat diffusion is a different process, and it fits the same framework with one modification that the physics fixes. For heat the estimation target is the current-state amplitude $a_k(t) = a_k(0)\, e^{-\kappa\lambda_k t}$ at each terminal snapshot; recovering $a_k(0)$ is exponentially ill-posed and is not attempted (Appendix E.1).

The amplitude spectrum gains an exponential factor. In acoustics the amplitude variance is a power law, $\sigma^2_{a,k} \propto \lambda_k^{-|s|}$. For heat the Green's function multiplies it by $e^{-2\kappa\lambda_k t}$:

$$\sigma^2_{a,k}(t) \;\propto\; \lambda_k^{-s}\, e^{-c\,\lambda_k}, \qquad c = 2\kappa t. \tag{12}$$

A power law $\Gamma_k = \lambda_k^p$ cannot represent this roll-off. The optimal regularizer is $\Gamma_k \propto \lambda_k^s\, e^{c\,\lambda_k}$, the prior's exponential factor inverted, as $\Gamma \propto \Sigma_a^{-1}$ requires.

The second parameter is known. For each room and observation time, OLS on $\log \hat\sigma^2_{a,k}(t) = \beta_0 - |s| \log \lambda_k - c\,\lambda_k$ recovers $(\hat s, \hat c)$ with a close fit at all but the earliest observation times (Appendix E.2; Figure 3 shows the landscapes). Across five diagnostic rooms, per-room regression slopes of $\hat c$ against c_{theory} lie close to one with near-perfect fits and small intercepts (Table 27). The factor $c = 2\kappa t$ is not fitted: the experimenter chooses t, and κ is a material constant ($\kappa = 1$ in our synthetic units). Only $|s|$ is estimated from data; Appendix E.3 reports the one-parameter fallback.

Same framework, different physics. For the synthetic heat data $|s| \approx 1.0$ rather than 1.13, because the excitation statistics differ; what carries over is the structure, a power law from the initial conditions composed with a PDE-specific correction known from theory. The fitted spectra follow the predicted rate (Table 27), but the two-parameter regularizer with $c = 2\kappa t$ at each snapshot is not evaluated: the 2-D sweep uses one c for all snapshots of a window and finds its optimum near $c = 0$ with a steeper power law (Figure 3). The simulated heat data contain no discarded modes, so the late-time breakdown of isotropy that discarded heat modes would cause (Appendix E.5) is an analytic prediction these data do not test. The heat data are generated by the same Green's function the theory predicts, so this is a consistency check on the fitting procedure, not physical validation.

8 Conclusion

Under isotropic truncation noise, the optimal diagonal regularizer for a modal inverse problem is set by the excitation, not by the room: it is $\Gamma_k = \lambda_k^{|s|}$. The statement is exact at T=1 (Proposition 1) and empirical for $T > 1$, where the closed form stays within 5.83% of per-room power-law oracle tuning at the population median. The discarded modes explain why: Weyl's count makes the truncation noise nearly isotropic, Berry's conjecture affects only a measured diagonal term, and the resulting

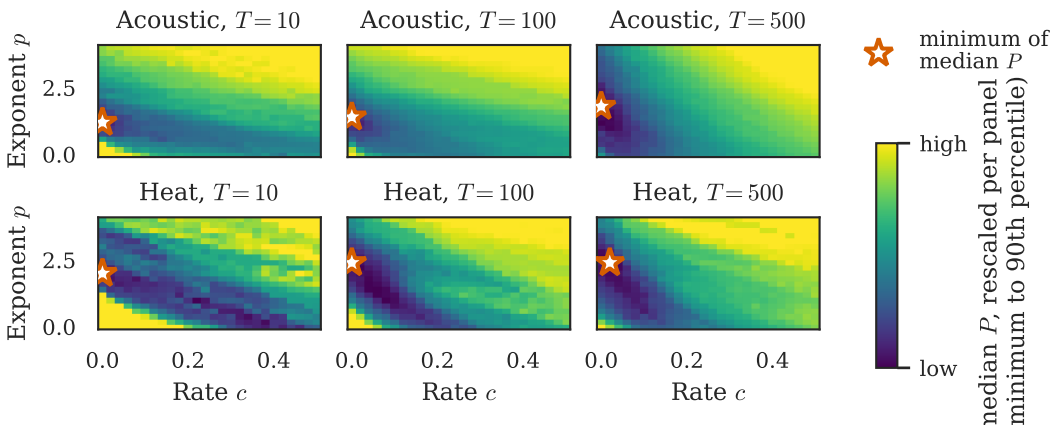


Figure 3: **Heat and acoustic landscapes differ in shape, not only in where the optimum lies.** Each map is the median error over p and c at one window, coloured from its own minimum upward so shapes, not levels, compare. Acoustic optima sit at $c = 0$, where a power law suffices (top). Heat optima sit near $c = 0$ at a larger p, a steeper power law standing in for the predicted exponential factor (bottom). This is the correction of Equation (12), whose rate Figure 18 checks.

landscape is flat. None of the learned architectures we tested consistently improves on the closed form in the truncation regime, including Learned Iterative Ridge, which couples modes across layers (Sections 6 and 6.1); this bounds the methods tested, not learning in general.

Scope and limitations. Our rooms are random convex 2D polygons; the Gaussian prior is tested against heavy-tailed and correlated alternatives (Appendix F). A 3D stress test on analytic boxes (Appendix G.5) finds H and landscape flatness in line with the 2D rooms at $T \leq 100$, but boxes are separable, so transfer to non-separable 3D rooms is an open hypothesis, and realistic sensing pipelines are untested. Sensors are drawn i.i.d. uniformly; structured arrays (linear, circular, spherical) may introduce eigenfunction correlations the analysis does not cover. Rectangular rooms, where Berry's premise fails analytically, are nearly as flat as generic rooms at matched discarded-mode counts (Appendix G). Electronic sensor noise only shifts α, and frequency-dependent damping composes with $|s|$; Appendix H gives diagnostics for when the closed form should be trusted. Real recordings are untested: a static compact array cannot resolve the retained modes at any room size the framework covers (Appendix I), so establishing the slope and the isotropy on real data needs a wider or moving array.

Acknowledgments and Disclosure of Funding

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (No.RS-2024-00358374) and (No.RS-2026-25504174, ERC-TRIPS).

References

J. Adler and O. Öktem. Learned primal-dual reconstruction. *IEEE transactions on medical imaging*, 37(6):1322–1332, 2018.

H. K. Aggarwal, M. P. Mani, and M. Jacob. Modl: Model-based deep learning architecture for inverse problems. *IEEE transactions on medical imaging*, 38(2):394–405, 2018.

G. S. Alberti, E. De Vito, M. Lassas, L. Ratti, and M. Santacesaria. Learning the optimal tikhonov regularizer for inverse problems. *Advances in Neural Information Processing Systems*, 34:25205–25216, 2021.

A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas. A-optimal design of experiments for infinite-dimensional bayesian linear inverse problems with regularized \ell_0-sparsification. *SIAM Journal on Scientific Computing*, 36(5):A2122–A2148, 2014.

N. Antonello, E. De Sena, M. Moonen, P. A. Naylor, and T. Van Waterschoot. Room impulse response interpolation using a sparse spatio-temporal representation of the sound field. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(10):1929–1941, 2017.

V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen. On instabilities of deep learning in image reconstruction and the potential costs of ai. *Proceedings of the National Academy of Sciences*, 117 (48):30088–30095, 2020.

J. V. Beck, B. Blackwell, and C. R. S. Clair. *Inverse heat conduction: Ill-posed problems*. James Beck, 1985.

M. V. Berry. Regular and irregular semiclassical wavefunctions. *Journal of Physics A: Mathematical and General*, 10(12):2083–2091, 1977.

D. Calvetti and E. Somersalo. Distributed tikhonov regularization for ill-posed inverse problems from a bayesian perspective. *Computational Optimization and Applications*, 91(2):541–572, 2025.

L. Cavalier. Nonparametric statistical inverse problems. *Inverse Problems*, 24(3):034004, 2008.

M. J. Colbrook, V. Antun, and A. C. Hansen. The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and smale's 18th problem. *Proceedings of the National Academy of Sciences*, 119(12):e2107151119, 2022.

P. Del Hougne, D. V. Savin, O. Legrand, and U. Kuhl. Implementing nonuniversal features with a random matrix theory approach: Application to space-to-configuration multiplexing. *Physical Review E*, 102(1):010201, 2020.

C. Durastanti. Spectral bayesian regression on the sphere. *arXiv preprint arXiv:2601.20528*, 2026.

H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

E. Fernandez-Grande. Sound field reconstruction using a spherical microphone array. *The Journal of the Acoustical Society of America*, 139(3):1168–1178, 2016.

G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.

N. M. Gottschling, V. Antun, A. C. Hansen, and B. Adcock. The troublesome kernel: On hallucinations, no free lunches, and the accuracy-stability tradeoff in inverse problems. *SIAM Review*, 67(1): 73–104, 2025.

K. Gregor and Y. LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th international conference on international conference on machine learning*, pages 399–406, 2010.

P. C. Hansen. Analysis of discrete ill-posed problems by means of the l-curve. *SIAM review*, 34(4): 561–580, 1992.

P. C. Hansen. *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*. SIAM, 1998.

A. Hassell and L. Hillairet. Ergodic billiards that are not quantum unique ergodic. *Annals of Mathematics*, pages 605–618, 2010.

E. J. Heller. Bound-state eigenfunctions of classically chaotic hamiltonian systems: scars of periodic orbits. *Physical Review Letters*, 53(16):1515, 1984.

T. Hohage and F. Weidling. Characterizations of variational source conditions, converse results, and maxisets of spectral regularization methods. *SIAM Journal on Numerical Analysis*, 55(2):598–620, 2017.

S. Hurault, A. Leclaire, and N. Papadakis. Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization. In *International Conference on Machine Learning*, pages 9483–9505. PMLR, 2022.

V. Ivrii. 100 years of weyl's law. *Bulletin of Mathematical Sciences*, 6(3):379–452, 2016.

W. James, C. Stein, et al. Estimation with quadratic loss. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 361–379. University of California Press, 1961.

I. M. Johnstone. Function estimation and gaussian sequence models. *Unpublished manuscript*, 2 (5.3):2, 2002.

J. P. Kaipio and C. Fox. The bayesian framework for inverse problems in heat transfer. *Heat Transfer Engineering*, 32(9):718–753, 2011.

J. P. Kaipio and E. Somersalo. *Statistical and computational inverse problems*. Springer, 2005.

X. Karakonstantis, D. Caviedes-Nozal, A. Richard, and E. Fernandez-Grande. Room impulse response reconstruction with physics-informed deep learning. *The Journal of the Acoustical Society of America*, 155(2):1048–1059, 2024.

B. T. Knapik, A. W. Van Der Vaart, and J. H. van Zanten. Bayesian inverse problems with gaussian priors. *The Annals of Statistics*, pages 2626–2657, 2011.

B. T. Knapik, A. W. Van Der Vaart, and J. H. van Zanten. Bayesian recovery of the initial condition for the heat equation. *Communications in Statistics-Theory and Methods*, 42(7):1294–1313, 2013.

A. Kofler, F. Altekrüger, F. Antarou Ba, C. Kolbitsch, E. Papoutsellis, D. Schote, C. Sirotenko, F. F. Zimmermann, and K. Papafitsoros. Learning regularization parameter-maps for variational image reconstruction using deep neural networks and algorithm unrolling. *SIAM Journal on Imaging Sciences*, 16(4):2202–2246, 2023.

A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9(2), 2008.

K. Kunisch and T. Pock. A bilevel optimization approach for parameter learning in variational models. *SIAM Journal on Imaging Sciences*, 6(2):938–983, 2013.

O. Leong, E. O'Reilly, and Y. S. Soh. The star geometry of critic-based regularizer learning. *Advances in Neural Information Processing Systems*, 37:71240–71276, 2024.

V. A. Morozov. Regularization of incorrectly posed problems and the choice of regularization parameter. *USSR Computational Mathematics and Mathematical Physics*, 6(1):242–251, 1966.

M. Pinsker. Optimal filtration of square-integrable signals in gaussian noise. *Prob. Info. Transmission*, 16(2):120–133, 1980.

C. E. Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.

J. M. Schmid, E. Fernandez-Grande, M. Hahmann, C. Gurbuz, M. Eser, and S. Marburg. Spatial reconstruction of the sound field in a room in the modal frequency range using bayesian inference. *The Journal of the Acoustical Society of America*, 150(6):4385–4394, 2021.

A. I. Shnirel'man. Ergodic properties of eigenfunctions. *Uspekhi Matematicheskikh Nauk*, 29(6): 181–182, 1974.

C. Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability, volume 1: Contributions to the theory of statistics*, volume 3, pages 197–207. University of California Press, 1956.

A. M. Stuart. Inverse problems: a bayesian perspective. *Acta numerica*, 19:451–559, 2010.

J. Sun, H. Li, Z. Xu, et al. Deep admm-net for compressive sensing mri. *Advances in neural information processing systems*, 29, 2016.

G. Tanner and N. Søndergaard. Wave chaos in acoustics and elasticity. *Journal of Physics A: Mathematical and Theoretical*, 40(50):R443–R509, 2007.

D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018.

H. Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen*, 71(4): 441–479, 1912.

S. Zelditch. Quantum ergodicity and mixing. *arXiv preprint math-ph/0503026*, 2005.

A Exact Theorem and Approximate Isotropy Argument

The argument of Section 4 has an exact part, Proposition 1 at $T = 1$ (Section A.1), and an approximate part, the isotropy of the truncation noise (Sections A.2 to A.3). Whether the approximation is adequate is an empirical question (Section 5, Appendix B, and Section A.5).

Notation. K retained modes, M microphones, eigenvalues λ_k, eigenfunction values $\varphi_k(x_m)$ collected in $\boldsymbol{\varphi}_k = [\varphi_k(x_1), \ldots, \varphi_k(x_M)]^\top \in \mathbb{R}^M$, and prior variances $\sigma_{a,k}^2 = c\,\lambda_k^{-s}$ with decay exponent $s > 0$.

A.1 Derivation of Proposition 1 at $T = 1$

At $T = 1$ the dimensions of the statement and the proof coincide: $\mathbf{y} \in \mathbb{R}^M$, $\Phi \in \mathbb{R}^{M\times K}$ with $\Phi_{mk} = \varphi_k(x_m)$, $\mathbf{a} \in \mathbb{R}^K$ holds the modal amplitudes at the observed snapshot, and $\boldsymbol{\eta} \in \mathbb{R}^M$ is the truncation noise at that snapshot. The proposition does not cover the multi-snapshot model of eq. (4), whose noise is temporally correlated (Section A.4).

Step (i): model and assumptions. The single-snapshot observation is

$$\mathbf{y} = \Phi\,\mathbf{a} + \boldsymbol{\eta}. \tag{13}$$

The prior has $\mathbb{E}[\mathbf{a}] = \mathbf{0}$ and diagonal $\mathrm{Cov}(\mathbf{a}) = \Sigma_{\mathbf{a}}$ with $(\Sigma_{\mathbf{a}})_{kk} = c\,\lambda_k^{-s}$: modes are uncorrelated, and low-frequency modes carry more energy than high-frequency ones. The noise has $\mathbb{E}[\boldsymbol{\eta}] = \mathbf{0}$, $\mathrm{Cov}(\boldsymbol{\eta}) = \sigma^2 I_M$, and $\mathrm{Cov}(\mathbf{a}, \boldsymbol{\eta}) = 0$, which holds when the discarded-mode amplitudes are uncorrelated with the retained ones. Isotropy is the key assumption (Section A.2); no distributional assumption enters.

Step (ii): best linear estimator. For a linear estimator $\hat{\mathbf{a}} = W\mathbf{y}$ the mean-squared error is

$$\mathbb{E}\|W\mathbf{y} - \mathbf{a}\|^2 = \mathrm{tr}\big[(W\Phi - I)\,\Sigma_{\mathbf{a}}\,(W\Phi - I)^\top\big] + \sigma^2\,\mathrm{tr}\big[WW^\top\big], \tag{14}$$

which depends on the joint law of $(\mathbf{a}, \boldsymbol{\eta})$ only through the second moments of Step (i). Setting its gradient with respect to W to zero gives the linear minimum-MSE (LMMSE) estimator

$$\hat{\mathbf{a}}_{\mathrm{LMMSE}} = \Sigma_{\mathbf{a}}\Phi^\top\big(\Phi\Sigma_{\mathbf{a}}\Phi^\top + \sigma^2 I_M\big)^{-1}\mathbf{y} = \big(\Phi^\top\Phi + \sigma^2\Sigma_{\mathbf{a}}^{-1}\big)^{-1}\Phi^\top\mathbf{y}, \tag{15}$$

where the second form follows from the push-through identity $\Sigma_{\mathbf{a}}\Phi^\top(\Phi\Sigma_{\mathbf{a}}\Phi^\top + \sigma^2 I_M)^{-1} = (\Phi^\top\Phi + \sigma^2\Sigma_{\mathbf{a}}^{-1})^{-1}\Phi^\top$.

Step (iii): connection to Tikhonov. At $T = 1$ the Tikhonov estimator of eq. (5) is $\hat{\mathbf{a}}_{\mathrm{Tikh}} = (\Phi^\top\Phi + \alpha\,\Gamma)^{-1}\Phi^\top\mathbf{y}$, which coincides with eq. (15) if and only if

$$\alpha\,\Gamma = \sigma^2\,\Sigma_{\mathbf{a}}^{-1}. \tag{16}$$

Since $(\Sigma_{\mathbf{a}}^{-1})_{kk} = c^{-1}\,\lambda_k^s$, this gives $\Gamma_{kk} = \sigma^2\lambda_k^s/(\alpha\,c)$. The prefactor is absorbed into α, so the shape is $\Gamma_{kk} \propto \lambda_k^s$, i.e., $p^\star = s$.

Step (iv): what Gaussianity adds. If $(\mathbf{a}, \boldsymbol{\eta})$ are jointly Gaussian, the posterior $p(\mathbf{a} \mid \mathbf{y})$ is Gaussian, and completing the square in $-\frac{1}{2\sigma^2}\|\mathbf{y} - \Phi\mathbf{a}\|^2 - \frac{1}{2}\mathbf{a}^\top\Sigma_{\mathbf{a}}^{-1}\mathbf{a}$ shows that its mean is eq. (15). The same estimator is then the posterior mean, hence minimum-MSE among all estimators, linear or not. Gaussianity is needed only to upgrade LMMSE-optimality to unrestricted MMSE-optimality: the shape conclusion is the same in both cases, and the second-moment version is what makes the population-risk comparisons of Appendix F exact.

Why the prior alone sets the shape. The forward operator $\Phi^\top\Phi$ couples modes in the estimator, but it does not enter the penalty shape, which is $\Sigma_{\mathbf{a}}^{-1}$; the noise level σ^2 and $\Phi^\top\Phi$ affect only α. □

Table 4: **The discarded modes are many, their Herfindahl index is small, and the measured anisotropy is moderate.** These population values back the isotropy argument of Section 4. Weyl's count sets the number of discarded modes and hence H, Berry's conjecture enters only through D, and the measured operator norm of E agrees in order of magnitude with the Jensen bound.

Quantity	Value
Discarded modes $K_{\mathrm{total}} - K$, median over the 197 validation rooms	263
Discarded modes, median over the 187 in-scope rooms	287
Herfindahl index H of Equation (11), median	0.005
$D/(2H)$, median [IQR]	2.16 [1.70, 2.72]
Excess of $\mathbb{E}\|E\|_F^2$ over the Berry value $D = 2H$	26%
Jensen bound on $\mathbb{E}\|E\|_{\mathrm{op}}$ at $M = 8$ (under $D = 2H$)	0.67 (0.60)
Empirical $\|E\|_{\mathrm{op}}$, median / mean / 95th percentile	0.58 / 0.88 / 2.42
Spearman $\rho(\|E\|_{\mathrm{op}}, \delta)$ at $T = 1000$	-0.30 ($p < 10^{-4}$)
Pooled KS statistic against $\chi^2(1)$ ($N = 77{,}968$ samples)	0.037
Monte Carlo check of the off-diagonal identity, $\mathbb{E}[E_{ij}^2]/H$ (50,000 i.i.d. Gaussian random-wave draws, 6 rooms)	1.00 ± 0.01
Single-mode part $S/(2H)$	0.877
$D/(2H)$ in separable domains: rectangles (Section G) / 3D boxes (Section G.5)	6.6–8.3 / 15–76
Sensor-draw Monte Carlo (32 rooms, 500 draws each): off-diagonal / diagonal term ratio	0.97 / 0.99
$\|E\|_{\mathrm{op}}^2/\mathbb{E}\|E\|_F^2$ at the median	0.51
Std of the trace factor between E_{tf} and $\bar{E}$ over sensor draws (median over 30 rooms)	0.08
Frobenius estimate against the empirical median $\|E\|_{\mathrm{op}}$	within 16%
Spearman $\rho(D/(2H), \|E\|_{\mathrm{op}})$	-0.16
$D/(2H)$ of the worst-Berry rooms (Section B.4)	1.4–2.5
Median per-room noise slope q, FEM-only / total residual (Figure 5)	-0.09 / $+0.12$

Scope of optimality. Under exact isotropy ($E = 0$) at $T = 1$ and a Gaussian power-law prior, the closed form is MMSE-optimal among all estimators (Step (iv)). In our rooms isotropy is approximate (Table 4), so the closed form is not exactly MMSE: the Bayes-linear estimator under the true covariance is non-diagonal and couples modes through E. Throughout the paper, the "per-room power-law oracle" is the best estimator within the scalar power-law family $(p^\star, \alpha^\star)$, not the unconstrained Bayes estimator or an unrestricted per-mode search. The saturation result of Section 6 is a statement about the diagonal family: the tested learned models do not improve on $\Gamma_k = \lambda_k^{|s|}$ in the truncation regime, and LIR (Section 6.1), a learned method outside that family, does not consistently improve on it either (Table 1).

A.2 Anisotropy: identities and bound

Proposition 1 requires isotropic noise. Write the truncation-noise covariance of eq. (7) as $R_{\mathrm{trunc}} = \sum_{n>K} \sigma_{a,n}^2 \boldsymbol{\varphi}_n \boldsymbol{\varphi}_n^\top = \sigma_{\mathrm{trunc}}^2 (I_M + E)$, with average noise power $\sigma_{\mathrm{trunc}}^2 = M^{-1} \sum_{n>K} \sigma_{a,n}^2$, so that E is the deviation from isotropy; the goal is a bound on $\|E\|_{\mathrm{op}}$. Most of the bound needs no input from Berry's conjecture: the off-diagonal second moments of E obey an exact identity, and the conjecture enters only through the diagonal term D.

The entries of E. The identities below use the *population* normalization, which divides R_{trunc} by $\bar{\sigma}^2 = M^{-1}\,\mathbb{E}[\mathrm{tr}\, R_{\mathrm{trunc}}] = |\Omega|^{-1} \sum_{n>K} \sigma_{a,n}^2$. With $\psi_n = \sqrt{|\Omega|}\,\varphi_n$ (unit second moment under uniform sensors) and weights $w_n = \sigma_{a,n}^2 / \sum_{m>K} \sigma_{a,m}^2$, the population-normalized anisotropy matrix $E = E^{\mathrm{pop}}$ has entries

$$E_{ij} = \sum_{n>K} w_n\, \psi_n(x_i)\, \psi_n(x_j) - \delta_{ij}. \tag{17}$$

Uniform sensors and L^2-orthonormality give the population identities $\mathbb{E}_x[\psi_n(x)^2] = 1$ and $\mathbb{E}_x[\psi_n(x)\,\psi_m(x)] = \delta_{nm}$.

Off-diagonal entries: an exact identity. For $i \neq j$,

$$\mathbb{E}[E_{ij}^2] = \sum_{n>K} \sum_{n'>K} w_n w_{n'} \, \mathbb{E}\big[\psi_n(x_i)\, \psi_{n'}(x_i)\big] \, \mathbb{E}\big[\psi_n(x_j)\, \psi_{n'}(x_j)\big] = \sum_{n>K} w_n^2 = H, \tag{18}$$

because x_i and x_j are independent draws (the expectation factorizes) and each factor equals $\delta_{nn'}$ by orthonormality. The identity uses no property of Berry's conjecture and holds for any L^2-orthonormal eigenbasis, including the separable sine products of Appendix G. Here H is the Herfindahl index

$$H := \frac{\sum_{n>K} (\sigma_{a,n}^2)^2}{\big(\sum_{n>K} \sigma_{a,n}^2\big)^2}. \tag{19}$$

A Monte Carlo with i.i.d. Gaussian random-wave values reproduces the identity to within its sampling error, and a sensor-draw check on the room eigenfunctions agrees (Table 4).

Diagonal entries: where Berry enters. Write $V(x) = \sum_{n>K} w_n \, \psi_n(x)^2$, so that $E_{ii} = V(x_i) - 1$ with $\mathbb{E}[V] = 1$ and

$$\mathbb{E}[E_{ii}^2] = \mathrm{Var}_x[V] =: D = \underbrace{\sum_{n>K} w_n^2 \, \mathrm{Var}_x[\psi_n^2]}_{S \text{ (single-mode)}} + \underbrace{\sum_{\substack{n \neq m \\ n,m>K}} w_n w_m \, \mathrm{Cov}_x\big(\psi_n^2, \psi_m^2\big)}_{C \text{ (cross-mode)}}. \tag{20}$$

Under Berry's random-wave statistics, $\psi_n(x)$ is a unit-variance Gaussian, so $\mathrm{Var}[\psi_n^2] = 2$, squared intensities of distinct modes are uncorrelated, and $D \approx 2H$; this is the only point at which the conjecture enters. Measured with exact degree-4 FEM quadrature across the in-scope rooms, the diagonal term is about twice Berry's value (Table 4). The excess comes from covariance between different modes, not from heavier single-mode marginals: the single-mode part sits slightly below Berry's value (Table 4). In separable domains, where Berry's premise fails by construction, $D/(2H)$ is several times larger for the rectangles of Appendix G and larger still for the 3D boxes of Appendix G.5, growing with mode count (Table 4).

Assembling the bound. Summing the $M(M-1)$ off-diagonal and the M diagonal entries gives, exactly,

$$\mathbb{E}\|E\|_F^2 = M(M-1)\, H + M\, D, \qquad \mathbb{E}\|E\|_{\mathrm{op}}^2 \leq \mathbb{E}\|E\|_F^2; \tag{21}$$

Berry's $D \approx 2H$ turns this into $M(M+1)H$. At $M = 8$ the Berry-free off-diagonal term ($56H$) is the larger one, so the measured D, about twice Berry's, raises the full Frobenius second moment by about a quarter (Table 4). A sensor-draw Monte Carlo reproduces both terms of eq. (21) closely and puts $\|E\|_{\mathrm{op}}^2$ at about half of $\mathbb{E}\|E\|_F^2$ at the median, which measures the looseness of the Frobenius relaxation (Table 4).

Normalization conventions. The identities above use the population normalization $\bar{\sigma}^2$. The reported operator norms (Table 4) use the *realized-trace* normalization instead: it divides by $\mathrm{tr}(R_{\mathrm{trunc}})/M$ at the actual sensor draw and yields the trace-free matrix $E_{\mathrm{tf}} = \bar{E}/\big(1 + \mathrm{tr}(E^{\mathrm{pop}})/M\big)$, where $\bar{E} = E^{\mathrm{pop}} - \mathrm{tr}(E^{\mathrm{pop}})\, I_M/M$ is the trace-free projection. For $\bar{E}$ the identity $\mathbb{E}\|\bar{E}\|_F^2 = (M-1)(MH + D)$ is exact, and E_{tf} differs from $\bar{E}$ only by the scalar trace factor, whose spread over sensor draws is small (Table 4). H, D, and every ratio built from them are invariant to the φ_n normalization.

From entries to operator norm. With $M = 8$ and the median H, Jensen's inequality gives the order-of-magnitude estimate $\mathbb{E}\|E\|_{\mathrm{op}} \lesssim \sqrt{M(M-1)H + MD}$ with the measured D, or $\sqrt{M(M+1)H}$ under $D = 2H$ (Table 4); because H varies across rooms, the estimate is tight only at the median. Across 187 rooms[4] (eigenfunctions evaluated at the sensor positions, realized-trace normalization E_{tf}), the empirical $\|E\|_{\mathrm{op}}$ has a median below one and a long right tail (Table 4). For the median room the Frobenius estimate is close (Table 4); for the worst rooms the Frobenius-to-operator-norm relaxation is loose. The right-tail rooms are small rooms with few discarded modes and large H (the extreme, a room with a single discarded mode, reaches the maximum $M-1$), not the rooms with the largest Berry violations: across rooms, $D/(2H)$ is nearly uncorrelated with $\|E\|_{\mathrm{op}}$, and the worst-Berry rooms of Section B.4 have ordinary $D/(2H)$ (Table 4).

[4]Ten rooms with $K_{\mathrm{total}} \leq 50$ have no discarded modes from which to compute $\|E\|_{\mathrm{op}}$ and are excluded here; all 197 rooms are retained for the per-room exponent comparison in Appendix D.5.

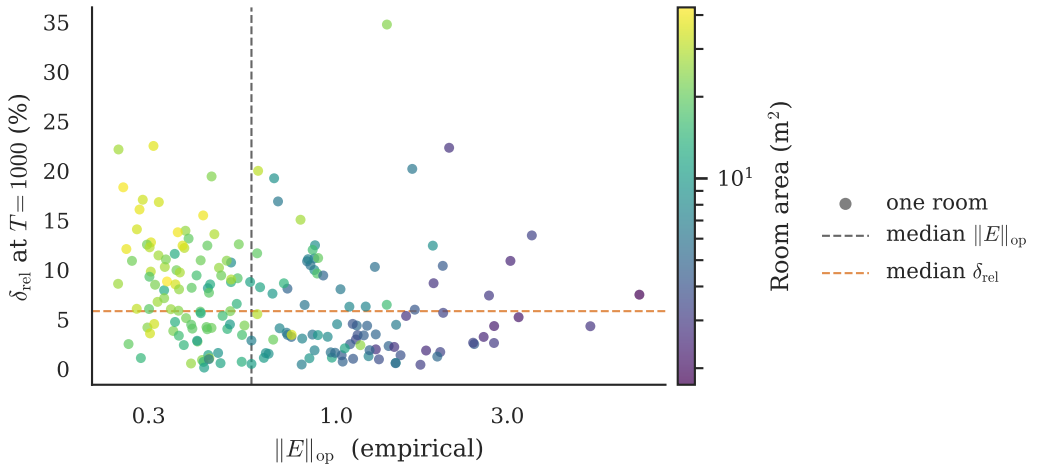


Figure 4: **Rooms with higher anisotropy pay *lower* cost, a confound of room size.** Each point is an in-scope room, $\|E\|_{\text{op}}$ is zero for isotropic truncation noise, and δ_{rel} is the relative cost of the closed form against the per-room power-law oracle. Large rooms have many discarded modes and hence low anisotropy, yet they tend to pay more than small rooms. Once room area is controlled for, no correlation survives (Section D.9), so anisotropy does not explain which rooms pay the cost.

Signal and noise in modal coordinates. The noise is moderately anisotropic, not negligibly so. In a decoupled modal model with prior variance $\lambda_k^{-|s|}$ and modal-noise variance $\propto \lambda_k^q$, the Wiener shrinkage is $1/(1 + c\,\lambda_k^{|s|+q})$, so the matching Tikhonov exponent is $p = |s| + q$; this is exact in the decoupled model and an approximation for the coupled forward operator. Figure 5 gives small slopes of either sign by two methods (Table 4), both an order of magnitude inside the basin width of Appendix C.6, so we apply no correction. Anisotropy across sensor directions does not enter this exponent at first order; its second-order cost is the subject of Remark 1 and Section A.5.

Cost, not drift, is the criterion. The oracle exponent $p^\star$ drifts upward as T grows (Table 12), but the criterion is whether staying at $|s|$ is costly. Because the landscape is flat (Section 5), the population-median relative cost peaks at 5.83% at $T = 1000$, and at $T = 1000$ two in-scope rooms in three lose less than 1.1 pp absolute (Table 12). Anisotropy does not explain the rooms with elevated δ: the worst-cost rooms are large rooms with low anisotropy, and the raw correlation of Figure 4 (Table 4) disappears once room area is controlled for (Appendix D.9). The controlled sweep of Section A.5 confirms directly that anisotropy up to the mean observed level leaves the median cost of the fixed exponent essentially unchanged, and that stronger anisotropy raises it noticeably only at the longest window (Table 6).

A.3 Herfindahl index: numerical values

For the discarded-mode weights $\sigma_{a,n}^2 = c\,\lambda_n^{-s}$ with $|s| = 1.13$, the Herfindahl index $H = \sum_{n>K} \lambda_n^{-2s} / (\sum_{n>K} \lambda_n^{-s})^2$ is an order of magnitude smaller in the large decagons, which discard about a thousand modes, than in the small triangles, which discard a few dozen (Table 5). For a representative octagon, the effective contributor count $1/H$ is a large fraction of its discarded modes.

H does not vanish as rooms grow: because $|s| > 1$, both $\sum n^{-2|s|}$ and $\sum n^{-|s|}$ converge as $K_{\text{total}} \to \infty$, so H tends to a small positive constant. Its smallness at our operating point is a finite-sample property of the eigenvalue distribution, not an asymptotic guarantee.

A.4 Temporal stacking

The analysis of Sections A.1 to A.2 applies at $T = 1$. For $T > 1$ the stacked noise covariance is $\text{Cov}(\tilde{\eta}) = \sum_{n>K} \sigma_{a,n}^2\, G_n \otimes (\boldsymbol{\varphi}_n \boldsymbol{\varphi}_n^\top)$, where G_n is a temporal rank-two matrix encoding the damped-sinusoidal evolution of mode n from eq. (2) (one outer product each for the cosine and sine

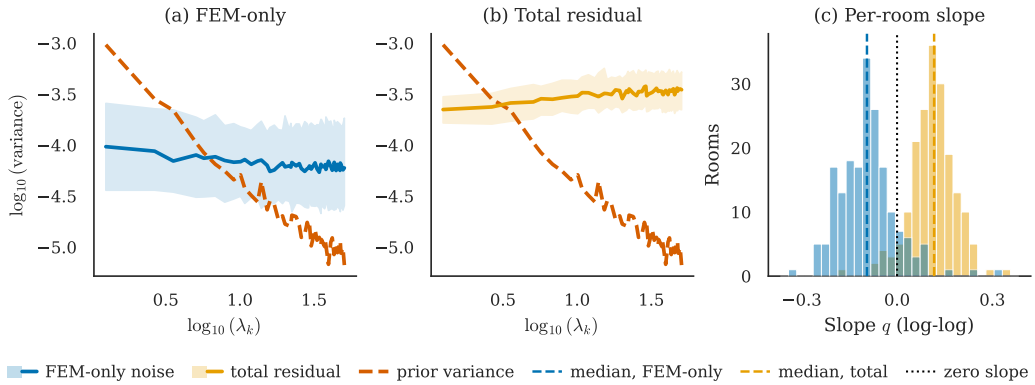


Figure 5: **Truncation noise is approximately isotropic across retained modes.** Estimated from the FEM residual alone, its median variance barely changes with eigenvalue while the prior variance decays (a). Estimated from the total residual against the spatial prior, it is also nearly flat (b). Per-room slopes of both estimates cluster around zero, the value for noise that is flat across modes (c). Both slopes lie well inside the basin width of Section C.6, so the closed form needs no correction.

Table 5: **The Herfindahl index is small in every room and smallest where the most modes are discarded.** Its spread across rooms follows the discarded-mode count, a representative room keeps a large effective contributor count, and the limit for infinitely many modes is positive but small. This is the quantity that Equation (9) ties to the off-diagonal anisotropy.

Quantity	Value
H across rooms, smallest (large decagons, about a thousand discarded modes)	0.002
H across rooms, largest among triangles (about thirty discarded modes)	0.03
H, median over in-scope rooms	0.005
Representative octagon (room 00850): K_{total} / discarded modes	483 / 433
Representative octagon: H / effective contributor count $1/H$	0.0038 / 260
Limit of H as $K_{\text{total}} \to \infty$ with $\lvert s\rvert = 1.13$ and $K = 50$	≈ 0.0003

components). The spatial factors still concentrate to I_M in aggregate by the argument of Section A.2, but a temporal structure remains: off-diagonal entries $\sum_n \sigma_{a,n}^2 \, g_n(t) g_n(t')$ couple snapshots through shared initial conditions. This temporal correlation shifts the optimal exponent $p^\star(T)$ away from $|s|$ as T grows; Section 5 measures its cost (peak median 5.83% at $T = 1000$, Table 2; over all rooms the median cost falls again at the longest window as the damped signal becomes uninformative regardless of regularizer, Table 11). We state no isotropic corollary for $T > 1$: Proposition 1 is a single-snapshot statement, and the multi-snapshot results of Sections 5 and 6 are empirical.

A.5 Controlled anisotropy sweep

Because Remark 1 is local and room-dependent, a controlled sweep measures the cost of anisotropy directly. The sweep replaces the empirical truncation-noise covariance of 24 in-scope rooms with positive-definite, fixed-power covariances $R = \sigma^2(I_M + a\,\hat{E})$, $\operatorname{tr}\hat{E} = 0$, and evaluates the exact population risk of the diagonal Tikhonov family on the paper's 61×12 (p, α) grid. The direction $\hat{E}$ comes from three families (each room's own empirical direction, a balanced spectrum, and a single spike that reaches a 9:1 eigenvalue ratio at $a = 3.5$) in random and forward-aligned orientations, and a rises from 0 through the empirical median, mean, and 95th-percentile markers of $\|E_{\text{tf}}\|_{\text{op}}$ (Table 4), up to 3.5 in the spike family.

Table 6 reports the cost of the fixed exponent $p = |s|$ relative to the isotropic baseline of the same rooms. Up to the mean marker the median cost barely moves at any T and the worst sampled orientation adds at most half a point; beyond it the same holds at $T \le 100$, while at $T = 1000$ the 95th-percentile and spike levels raise the median cost by a few tenths of a point and the worst sampled orientation by more than a point (Table 6). The optimal exponent stays within a tenth of $|s|$ in every cell that retains the full room set. Medians use those cells; in the empirical-direction

Table 6: **Anisotropy up to the observed mean barely moves the cost of the fixed exponent, and stronger anisotropy matters only at the longest window.** Entries are the largest increases over the sweep in the relative cost of $p = |s|$ against the per-room power-law oracle, measured from the isotropic baseline of the same in-scope rooms; medians use the cells with the full room set. The sweep tests the local sensitivity of Remark 1 directly, for sensor-space anisotropy with temporally white noise.

	Up to the mean marker		Up to the spike level	
T	Median	Sampled orientation	Median	Sampled orientation
1	$+0.01$ pp	0.05 pp	$+0.04$ pp	0.26 pp
100	$+0.01$ pp	0.04 pp	$+0.02$ pp	0.08 pp
1000	$+0.08$ pp	0.50 pp	$+0.21$ pp	1.42 pp
Optimal exponent over all cells with the full room set: $p^\star \in [1.10, 1.20]$				

family the positive-definiteness limit leaves only a few rooms at the largest a, and those reduced cells rise further. The sweep isolates sensor-space anisotropy with temporally white noise, so it does not replace the temporally correlated multi-snapshot experiments of Section 5, and structured sensor arrays remain outside it.

B Berry's Conjecture: Extended Validation

The isotropy argument uses Berry's conjecture only for the diagonal anisotropy term D (Section 4). The pooled test of Section 5 gives a small Kolmogorov–Smirnov statistic (Table 7), and the per-room, near-wall, and worst-room analyses below show that the deviations are real but do not translate into meaningful reconstruction cost. Two KS conventions appear, with different levels for the same room: squared values $|\Omega|\,\varphi_k(x_m)^2$ against $\chi^2(1)$ (Section B.1, Section B.2, Section B.4, and Appendix D.9) and signed values $\sqrt{|\Omega|}\,\varphi_k(x_m)$ against $\mathcal{N}(0,1)$ (Section B.3, and Section B.2 for comparison).

B.1 Eigenfunction value distribution

Under Berry's conjecture, $\varphi_k(x_m)$ at a uniformly random sensor position is a zero-mean Gaussian with variance $1/|\Omega|$, so $|\Omega|\,\varphi_k(x_m)^2$ follows $\chi^2(1)$. Pooling every retained mode ($k \le \min(K_{\text{total}}, 50)$), all $M = 8$ sensors, and all 197 validation rooms gives $N = 77{,}968$ samples.

Pooled distribution. Against $\chi^2(1)$ the pooled two-sided statistic is a few percent (Figure 6a; Table 7), which is how far the empirical CDF deviates from the Berry prediction at most. At this sample size even so small a deviation is significant (Table 7); what matters is its effect on the regularizer.

Tail behavior. The empirical distribution has a lighter upper tail than $\chi^2(1)$ (Figure 6a), consistent with finite-mode effects: the conjecture is asymptotic in frequency, and the lowest retained modes are not yet in that regime. Restricting the pool to $k \ge 10$ lowers the statistic further (Table 7).

Why the pass rate is non-monotonic. The pass rate (Figure 6b) peaks at moderate mode counts and declines on both sides (Table 7), and sample size explains neither decline. Every room with $K_{\text{total}} \ge K$ contributes the same number of samples, and the median D_{KS} rises with mode count across the three bins, crossing the critical value only in the highest (Table 7). The rooms with $K_{\text{total}} < K$ contribute fewer samples yet deviate more (Table 7): all their modes are low-order, where random-wave behavior is not expected. The regularizer is unaffected: a large K_{total} supplies enough discarded modes to flatten $P(p)$ even when individual modes conform imperfectly (Appendix G.4).

B.2 Per-room KS distribution

A pooled statistic can hide rooms that violate the conjecture badly, so D_{KS} is also computed for each of the 197 validation rooms separately, in both conventions.

Table 7: **Berry's prediction fits the pooled eigenfunction values closely, while per-room deviations are real, grow with mode count, and carry almost no reconstruction cost.** The pooled and per-room Kolmogorov–Smirnov statistics of Section B are collected here with their sample sizes and the cost of the worst-fitting rooms. Only the diagonal term D of Equation (21) depends on the conjecture, which is why these deviations leave the closed form unaffected.

Quantity	Value
Pooled D_{KS} against $\chi^2(1)$, all retained modes ($N = 77{,}968$)	0.037 ($p < 10^{-10}$)
Pooled D_{KS}, modes $k \geq 10$ only	0.016
Samples per room for $K_{\mathrm{total}} \geq K$ / critical value at the 0.05 level	400 / 0.067
Pass rate at the 0.05 level, $K_{\mathrm{total}} \in [50, 200)$ (the peak)	88.5%
Median per-room D_{KS} by bin, $[50, 200)$ / $[200, 500)$ / ≥ 500 (rooms)	0.054 (61) / 0.061 (54) / 0.068 (73)
Rooms with $K_{\mathrm{total}} < K$ (9 rooms): samples / median D_{KS}	136–376 / 0.089
Per-room D_{KS}, $\chi^2(1)$ convention: median [IQR] / maximum / rooms above 0.12	0.061 [0.046, 0.073] / 0.197 / 1
Per-room D_{KS}, signed convention: median [IQR] / maximum	0.054 [0.042, 0.066] / 0.205
Spearman correlation of room rankings between the two conventions	0.57
Rooms with $D_{\mathrm{KS}} > 0.08$ (tri / quad / pent / hex / hept / oct / non / dec)	38 (4 / 6 / 2 / 5 / 3 / 9 / 5 / 4)
Spearman $\rho(D_{\mathrm{KS}}, K_{\mathrm{total}} - K)$, in-scope (Pearson r) / all rooms	$+0.29$ ($+0.30$, $p < 10^{-4}$) / $+0.16$
Worst-five rooms: largest δ at $T \leq 100$ / $\delta(T{=}1000)$ relative to the in-scope median	3.4% / four within $1.5\times$, room 00835 at $3.9\times$
Near-wall vs interior sensors: pooled D_{KS} excess / per-room medians (100 samples per group)	22% / 0.084 vs 0.085
Rooms excluded from the worst-room test (fewer than 50 discarded modes) / eligible rooms	29 / 168
Worst-Berry rooms: vertices / K_{total} / H	6–10 / > 400 / 0.0025–0.0042
Room 00835: D_{KS} / $\delta(T{=}1000)$ / δ at $T \leq 100$	0.098 / 22.5% / $\leq 3\%$
Spearman $\rho(D_{\mathrm{KS}}, \delta(T{=}1000))$ over the eligible rooms	0.19 ($p = 0.012$)

Results. In the $\chi^2(1)$ convention the per-room values are concentrated just below the critical value, with a single outlying room (Table 7), and the distribution is unimodal with a slight right tail. In the signed convention the same rooms give somewhat lower levels, and the room rankings agree only moderately between the conventions (Table 7), so the levels are not interchangeable.

Geometry and mode count. The rooms with the largest D_{KS} span every polygon type (Table 7), so geometry alone does not predict D_{KS}. Mode count correlates with it, weakly but significantly across the in-scope rooms and more weakly across all rooms (Table 7), consistent with the boundary rooms ($K_{\mathrm{total}} \leq K$) forming a distinct regime. Every in-scope room contributes the same number of samples, so the correlation reflects geometry rather than statistical power; we do not read it causally. What matters is whether D_{KS} predicts reconstruction cost: the five rooms with the largest D_{KS} keep a small relative cost at $T \leq 100$, and at $T = 1000$ four of them stay close to the in-scope median while room 00835 lies several times above it (Table 7, Section B.4).

B.3 Boundary-stratified KS statistics

Berry's conjecture is known to degrade near boundaries. Ranking each room's sensors by normalized wall distance $d_m = \mathrm{dist}(x_m, \partial\Omega)/\mathrm{diam}(\Omega)$ and pooling the signed values of the bottom (near-boundary) and top (interior) quartiles, two sensors each, over the 187 in-scope rooms shows a slight degradation near walls, where the pooled D_{KS} is higher than in the interior (Table 8). Per room, with the same samples per group, both quartiles sit at the sampling-noise level and at nearly the same median (Table 7). The regularizer acts in modal space, not sensor space, and the aggregate $M \times M$ covariance mixes near-boundary and interior sensors (Section A.2).

B.4 Rooms with the worst Berry agreement

The five rooms with the highest per-room D_{KS} ($\chi^2(1)$ convention) test whether the largest Berry violations translate into reconstruction cost.

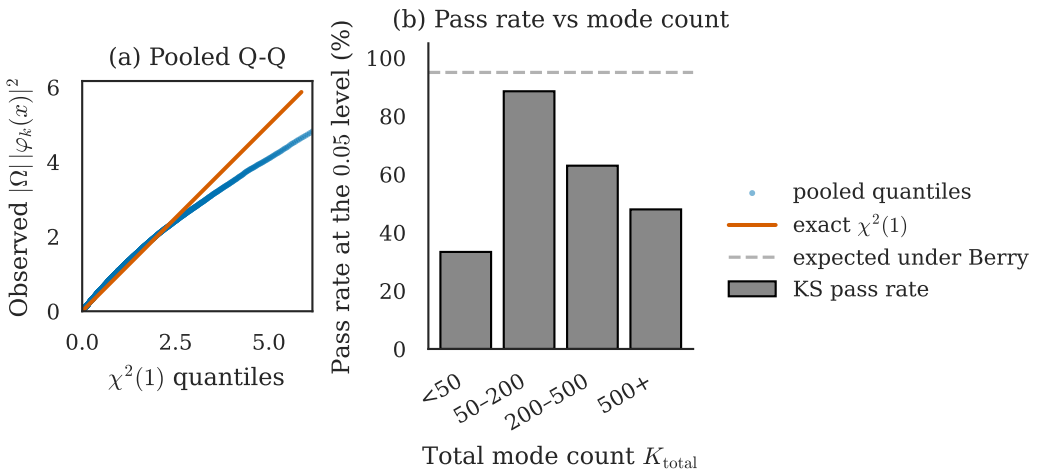


Figure 6: **Pooled eigenfunction values stay close to Berry's prediction, yet per-room tests detect real deviations.** Squared values follow $\chi^2(1)$ in the bulk but fall short in the upper tail (a). Per-room pass rates peak at moderate mode counts and stay below what exact agreement gives (b). Sample sizes are equal above K modes, so the decline at high counts is real. Berry's conjecture enters the bound only through D (Section A.2), and the deviations cost nothing in reconstruction (Section B.4).

Table 8: **Berry agreement degrades slightly near walls.** Lower D_{KS} means closer agreement with Berry's Gaussian prediction, here for signed eigenfunction values pooled over the in-scope rooms by quartile of sensor wall distance. The regularizer acts in modal space through the aggregate sensor covariance, which mixes near-wall and interior sensors (Section A.2).

Sensor group	Samples	Pooled D_{KS}
Interior (top quartile)	18,700	0.0198
Near-boundary (bottom quartile)	18,700	0.0242
All sensors	74,800	0.0203

Selection. Validation rooms with fewer than 50 discarded modes ($K_{\text{total}} - K < 50$) have too few for a meaningful per-room test and are excluded; Table 9 lists the top five of the rest (counts in Table 7).

Observations. The worst-Berry rooms are mid-to-large, many-sided rooms with well above the median mode count (Table 7), consistent with the positive correlation between D_{KS} and mode count (Section B.2). Their Herfindahl indices are near the population median, and their diagonal-anisotropy ratios are ordinary (Tables 7 and 4). Even room 00835, the worst of them, keeps a small cost at $T \leq 100$ (Table 7); its high cost at $T = 1000$ reflects oracle dispersion at long observation windows, which affects all rooms (Section C.4), not Berry failure.

Berry agreement and reconstruction cost. Across the eligible rooms, the correlation between D_{KS} and the cost at T=1000 is significant but weak (Table 7). Rooms with worse Berry agreement have slightly higher cost, yet even the worst-Berry rooms stay within the range reported in the main text. Berry violations are real but do not translate into meaningful reconstruction cost: the discarded-mode count, not Berry conformity, sets the flattening of the landscape (Appendix G.4), and anisotropy at the measured levels barely moves the cost of the fixed exponent (Appendix A.5).

C Acoustic Experiments: Extended Results

C.1 Estimating $|s|$ in practice

The population exponent $|s|$ is the one measured quantity the paper depends on.

Table 9: **The rooms with the worst Berry agreement are mid-to-large rooms with near-median H.** Rooms are ranked by D_{KS} in the $\chi^2(1)$ convention among those with enough discarded modes for a per-room test, and δ is the relative cost of the closed form against the per-room power-law oracle. Their costs stay modest except for one room at the longest window, where oracle dispersion rather than Berry failure drives the cost (Section C.4).

Room	Verts	Area	K_{total}	H	D_{KS}	$\delta(T{=}1)$	$\delta(T{=}100)$	$\delta(T{=}1000)$
00963	6	15.6	422	0.0042	0.103	0.0%	1.8%	5.4%
00924	9	18.8	508	0.0037	0.101	0.5%	0.1%	5.8%
00860	8	22.4	604	0.0033	0.100	0.0%	1.5%	3.8%
00835	10	36.8	987	0.0025	0.098	0.1%	2.8%	22.5%
00900	10	34.5	921	0.0026	0.097	3.2%	2.2%	8.5%

Procedure. For each room, the empirical variance $\hat{\sigma}^2_{a,k}$ of the k-th modal amplitude is computed from the amplitude time series across observation windows, and OLS fits

$$\log \hat{\sigma}^2_{a,k} = \beta_0 - |s| \log \lambda_k + \epsilon_k, \qquad k = 1, \ldots, K, \tag{22}$$

where β_0 absorbs the constant c; $|s|$ is the absolute value of the slope. The discarded modes are never observed, so the regression uses only the $K = 50$ retained modes and assumes that the power law $\sigma^2_{a,k} \propto \lambda_k^{-|s|}$ continues into the truncation band.

When the extrapolation holds. Two conditions suffice.

(a) **Dense eigenvalue spectrum.** Weyl's law gives approximately uniform eigenvalue spacing in 2D, $\lambda_k \sim k/|\Omega|$, so no gap separates the retained and discarded bands, and the fit over modes 1–50 extends to modes 51–313.

(b) **Stationary excitation statistics.** The excitation treats all modes alike, so the same power law holds above and below the truncation index; the broadband click of the simulations below has this property.

Excitation and prior. Each room is excited by a band-limited click from a point source, and the FDTD pressure field is projected onto the FEM eigenmodes to give the modal amplitudes; the decay $\lambda_k^{-|s|}$ is measured from these amplitudes, not imposed. The analysis models them with independent Gaussian amplitudes of variance $\lambda_k^{-|s|}$, the standard diffuse-field model and a reasonable approximation for reverberant rooms excited by broadband sources, where many incoherent sources or a broadband impulse leave modal amplitudes approximately independent. It is also the diagonal prior of Proposition 1; off-diagonal prior structure (e.g., adjacent-mode correlations from a localized source) would make the optimal Γ non-diagonal, a case Appendix F tests.

Population estimate. The ten rooms with $K_{\mathrm{total}} \leq 50$ have no discarded modes and hence no estimable noise profile; they are excluded from the fit but kept in all downstream evaluation. Table 10 summarizes the per-room fits of the remaining 187 rooms: the per-room standard deviation is the dispersion of fitted slopes, and the bootstrap quantities resample rooms to describe the population median, the value carried into all downstream experiments.

Sensitivity to estimation error. Re-estimating $|s|$ from 50 rooms shifts the exponent by up to about a tenth, and using such an estimate raises the median reconstruction error by less than half a point at every T (Table 13), because $P(p)$ is flat near its minimum (Figure 2). With more rooms the estimate of $|s|$ rises toward its full-sample value, which it approaches for the larger subsets, and the median cost ΔP falls over the same range (Figure 7).

Use in a new domain or a single room. Given a stable modal decomposition, a new domain or excitation regime uses the same procedure (estimate the retained-mode energies, regress log energy on log eigenvalue, take the absolute slope); a new excitation changes the value, not the procedure. The precondition is the real constraint: known or computed eigenmodes and enough spatial diversity (Appendix I). With a single room we would report the bootstrap interval (Table 10) and use the width of the $P(p)$ basin as a sensitivity check.

Table 10: **The population median of $|\hat{s}|$ is precise, although single-room slopes are not.** Statistics summarize the per-room fits of $|s|$ over the in-scope rooms, and bootstrap quantities resample rooms. The median is the exponent carried into all downstream experiments, and per-room slopes do not close the oracle gap (Section D.5).

Statistic	Value
Median $\|\hat{s}\|$	1.13
Mean $\|\hat{s}\|$	1.12
Std across rooms of the point estimates $\|\hat{s}\|$	0.25
Typical per-room bootstrap SE	0.27
Inter-room std after deconvolving the per-room SE	≈ 0 (Appendix D.5)
Bootstrap SE of the median	0.03
Bootstrap 95% CI on the median	$[1.08, 1.18]$
Median of the per-room fits, all 197 rooms (187 in-scope)	1.1243 (1.1266)

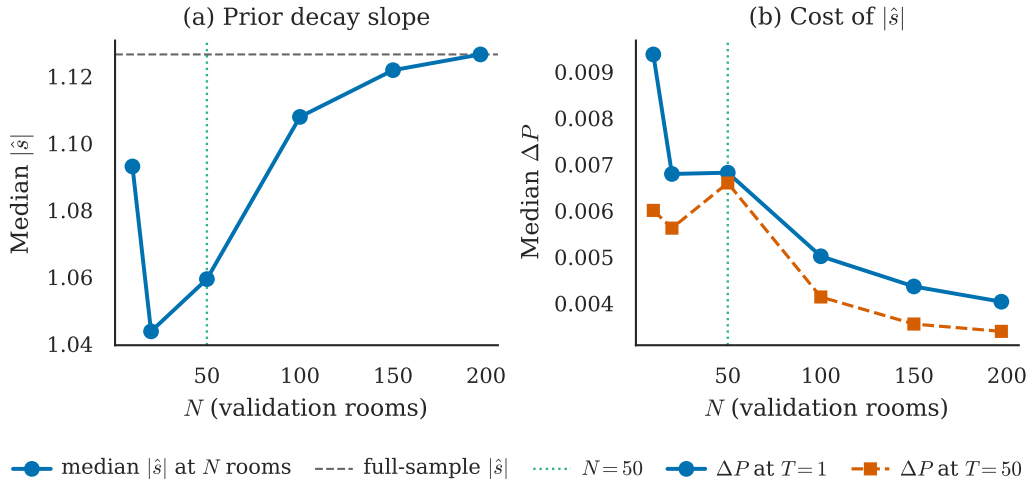


Figure 7: **The exponent of the closed form can be estimated from a few dozen rooms at little cost.** On subsets of the validation rooms the median fitted exponent sits below its full-sample value and climbs toward it as rooms are added (a). The median absolute cost of the resulting closed form stays small at every sample size, where ΔP is its extra error over the per-room power-law oracle (b). The flat landscape of Figure 2 keeps this estimation error cheap.

C.2 Full cost table

At large T the worst-room relative cost in Table 11 inflates because the oracle floor $P(p^\star, T)$ shrinks toward zero, while absolute gaps stay bounded at all T (Figure 8; Table 13); Section C.4 explains the non-monotone median.

C.3 Worst-case rooms

Worst absolute-cost room. Room 00806, a small triangle and a boundary room, has the largest absolute cost across all T, reached at $T = 100$, and even there most of its error is the oracle floor (Table 13). This is the expected failure mode at $K_{\text{total}} \leq K$: without discarded modes there is no truncation noise, and the per-room oracle picks an exponent far from the population $|s|$.

Worst in-scope rooms. Among in-scope rooms ($K_{\text{total}} > K$) the largest absolute cost occurs at a short window in a room with few discarded modes, and the worst in-scope gap at $T = 1000$ is smaller still, with almost all of that room's error in the oracle floor (Table 13).

Population view. In both worst cases most of the error is the oracle floor, the error that remains with the best per-room power-law regularizer and the limit of any static regularizer with $M = 8$

Table 11: **The median cost of the closed form stays small at every observation window, while worst-room relative costs inflate as the oracle floor shrinks.** The relative cost $\delta(T)$ is the extra modal error of the closed form as a fraction of the per-room power-law oracle error, over the in-scope and boundary rooms together. Table 2 summarizes the in-scope rooms alone, and Section C.4 explains why the median rises and then falls.

T	Median δ	IQR	95th pct	Worst room
1	0.5%	[0.2, 1.9]	4.8%	15.0%
5	0.5%	[0.1, 1.6]	5.6%	15.8%
10	0.5%	[0.1, 1.5]	4.9%	16.2%
20	0.5%	[0.1, 1.4]	3.9%	20.2%
50	0.6%	[0.2, 1.5]	4.5%	20.5%
100	1.5%	[0.6, 2.8]	6.2%	27.5%
200	2.2%	[0.7, 4.1]	7.9%	14.8%
500	5.1%	[2.7, 7.0]	12.3%	26.8%
1000	5.6%[†]	[2.8, 10.3]	16.9%	34.8%
2100	3.3%	[1.4, 5.7]	10.5%	44.8%

[†]Over the 187 in-scope rooms alone, without the 10 boundary rooms included here, this row reads 5.83% (Table 2).

Table 12: **The per-room oracle exponent drifts and spreads with the observation window, while the closed form's cost stays small for most rooms.** The exponent is concentrated at short and intermediate windows and barely defined at the longest, which is why the closed form costs little there despite the drift (Section 5). The last two rows summarize the absolute costs of Table 11 against the threshold used in Section 5.

T	1	50	100	500	1000	2100
Per-room $p^\star$, median over in-scope rooms	1.2	1.2	1.3	1.6	2.0	0.6
Per-room $p^\star$, std across in-scope rooms	0.85	0.26	0.30	0.42	0.99	1.67
Rooms below the 1.1 pp absolute-cost threshold at $T = 1000$: 69.0%						
Worst in-scope absolute cost over all T (prior-dominated regime): 7.61 pp						

sensors and $K = 50$ modes. At $T = 1000$, two rooms in three have absolute cost below 1.1 pp, with or without the boundary rooms (Table 13; Section 5), and the flat basin of the population median (Figure 2) holds room by room (Figure 9).

C.4 Non-monotonic relative cost

Table 11 shows $\delta(T)$ rising by an order of magnitude from $T = 1$ to $T = 1000$ and falling again at $T = 2100$. The median adaptation gap $\mathrm{gap}(T) = P(|s|, T) - P(p^\star, T)$ and the oracle floor $P(p^\star, T)$ reach their extrema at different T: the gap peaks earlier than the floor bottoms out (Section C.5; Table 13). Where the gap is largest the floor is still high (Table 14), which dilutes the ratio; by the time the floor bottoms out the gap has shrunk, which raises the ratio; beyond that window both push δ down. Signal decay does not drive the peak: more than half of the median signal energy remains at $T = 2100$ (Table 13), and decay is monotone in T, so it cannot produce a non-monotone $\delta(T)$.

C.5 Adaptation gap and oracle floor

As T grows, posterior concentration compresses the landscape: the posterior variance shrinks, Γ loses influence, and the median P range across $p \in [0, 3]$ falls by more than an order of magnitude from $T = 1$ to $T = 2100$ (Figure 2; Table 13). Table 14 tracks the two parts of the relative cost through this compression.

The adaptation gap is unimodal. The gap is flat for $T \leq 50$ (prior-dominated regime), roughly triples by $T = 500$ as the per-room oracle specializes, and falls back to its short-window level at $T = 2100$ as the landscape flattens (Table 14).

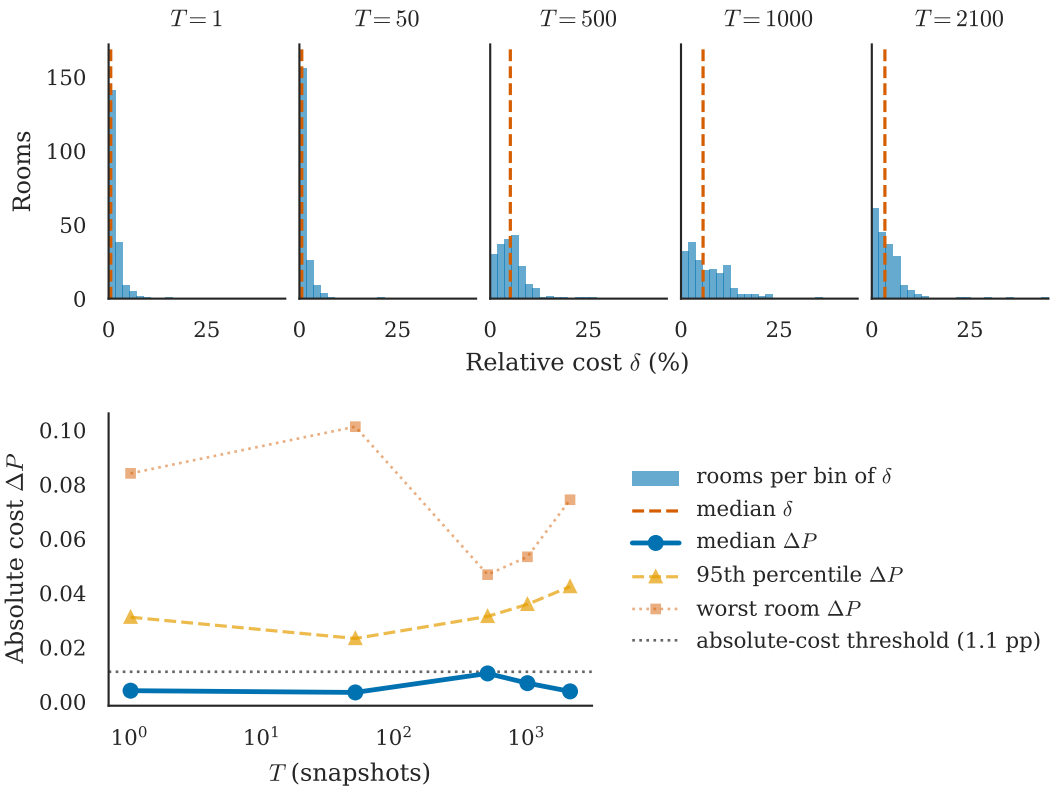


Figure 8: **The closed form costs little in most rooms at every window.** ΔP is the closed form's extra error over the per-room power-law oracle, and δ divides it by the oracle error. Relative costs pile up near zero at short windows and spread out at long ones (top). The median absolute cost stays at or below the threshold of Section 5, while the worst room loses several times that (bottom). It extends Table 2 to single rooms, boundary rooms included.

The oracle floor is U-shaped. The floor falls by a factor of six from $T = 1$ to $T = 1000$ and rebounds at $T = 2100$ (Table 14). At large T all exponents give nearly the same P, so a per-room "optimal" exponent is barely defined.

C.6 Grid resolution

The 61-point grid finds the global minimum reliably: the basin of $P(p)$ is an order of magnitude wider than the grid spacing (Table 13), so no continuous optimum falls between grid points.

C.7 Per-room landscape gallery

Figure 9 shows that the basin holds room by room: across six rooms with 3 to 10 vertices, per-room $p^\star$ varies by more than a factor of three (Table 13), yet $|\hat{s}| = 1.13$ lies inside the basin $P(p) \leq 1.05\, P(p^\star)$ in every room.

C.8 Dataset generation

Rooms are random convex 2D polygons with 3–10 vertices, with eigenpairs computed by FEM on a triangulated mesh; pressure fields are simulated on the same room by a finite-difference time-domain (FDTD) solver and projected onto the FEM eigenfunctions to give the modal amplitude time series. Of the 1,000 generated rooms, 800 are used for training (Section 6) and 197 for validation; three rooms (00905, 00913, 00921) have degenerate geometry and are excluded.

Table 13: **The closed form's worst cases are boundary rooms and short windows, and estimating its exponent from a few dozen rooms is cheap.** The table collects the sensitivity of the exponent estimate, the worst rooms, the two parts of the relative cost, and the grid resolution. Every worst case leaves most of the error in the oracle floor, the error that remains with the best per-room power law (Table 14).

Quantity	Value
Re-estimating $\|s\|$ from 50 rooms: deviation for the first 50 validation rooms / worst over 1,000 random subsets (95th percentile)	6% / 9.8% (6.6%)
Estimate from the first 50 validation rooms ($\|s\| = 1.06$): largest median per-room rise in P over T	0.38 pp
Worst absolute cost of all rooms (room 00806, boundary, $K_{\text{total}} = 19, T = 100$): $P(\|s\|)$ / $P(p^\star)$ / ΔP / floor fraction	0.491 / 0.385 / 10.6 pp / 78.4%
Worst in-scope absolute cost (room 00931, $K_{\text{total}} = 89$, $T = 5$)	7.61 pp
Worst in-scope gap at $T = 1000$ (room 00890, $K_{\text{total}} = 62$): ΔP / floor fraction	4.84 pp / 95.0%
Rooms below 1.1 pp absolute cost at $T = 1000$: all rooms / in-scope	67.0% / 69.0%
Window where the adaptation gap peaks / where the oracle floor is lowest	$T = 500$ / $T = 1000$
Median signal energy remaining at $T = 2100$	60%
Range of the median P over $p \in [0, 3]$: $T = 1$ / $T = 2100$ (ratio)	0.18 / 0.013 (14×)
Basin width of $P(p)$ (region within 5% of the minimum) vs grid spacing	$\Delta p \approx$ 1–2 vs 0.1
Per-room $p^\star$ in the gallery, decagon to quadrilateral (room 00930)	about 0.4 to 1.4

Table 14: **The adaptation gap and the oracle floor peak at different windows.** The gap is the closed form's extra error over the per-room power-law oracle, and the floor is the oracle's own error, boundary rooms included. Wide IQRs at short windows come from noise in the per-room oracle exponent, not from true dispersion of the gap. Relative cost is the gap divided by the floor, so it peaks where the floor is lowest, not where the gap is largest (Section C.4).

T	Median gap (pp)	IQR (pp)	Median $P(p^\star)$
1	0.40	$[0.12, 1.42]$	0.715
5	0.35	$[0.06, 1.07]$	0.715
10	0.34	$[0.10, 1.03]$	0.703
20	0.36	$[0.08, 0.98]$	0.679
50	0.34	$[0.13, 0.88]$	0.638
100	0.89	$[0.34, 1.53]$	0.594
200	0.91	$[0.33, 1.87]$	0.461
500	**1.04**	$[0.56, 1.65]$	0.220
1000	0.68	$[0.35, 1.34]$	**0.122**
2100	0.38	$[0.16, 1.21]$	0.152

D Learning Experiments: Extended Results

These checks ask whether the saturation of Section 6 reflects the models or the loss landscape.

Notation. The stacked state $\mathbf{a}_0 \in \mathbb{R}^{2K}$ pairs the cosine and sine amplitudes (c_k, β_k) of each mode; both share the prior variance $c\,\lambda_k^{-s}$, so a vector $\Gamma \in \mathbb{R}^K$ defines the tied penalty $\Gamma^{(2K)} = \Gamma \otimes I_2 = \operatorname{diag}(\Gamma_1, \Gamma_1, \ldots, \Gamma_K, \Gamma_K)$. All learned models parameterize Γ in this K-dimensional space; LIR acts on K-dimensional mode-pair vectors with the cosine/sine expansion implicit. $A = \tilde{\Phi}$ and $y = \tilde{\mathbf{y}}$ denote the stacked operator and measurements of eq. (4).

D.1 Architecture specifications

M1: CondNet with unrolled solver (4,949 parameters). The CondNet, a 3-layer MLP ($10 \to 64 \to 64 \to 1$, ReLU activations, softplus output), maps a 10-dimensional per-mode feature vector to a positive penalty per mode, giving the diagonal $\Gamma \in \mathbb{R}^K$. For mode k, with $G =$

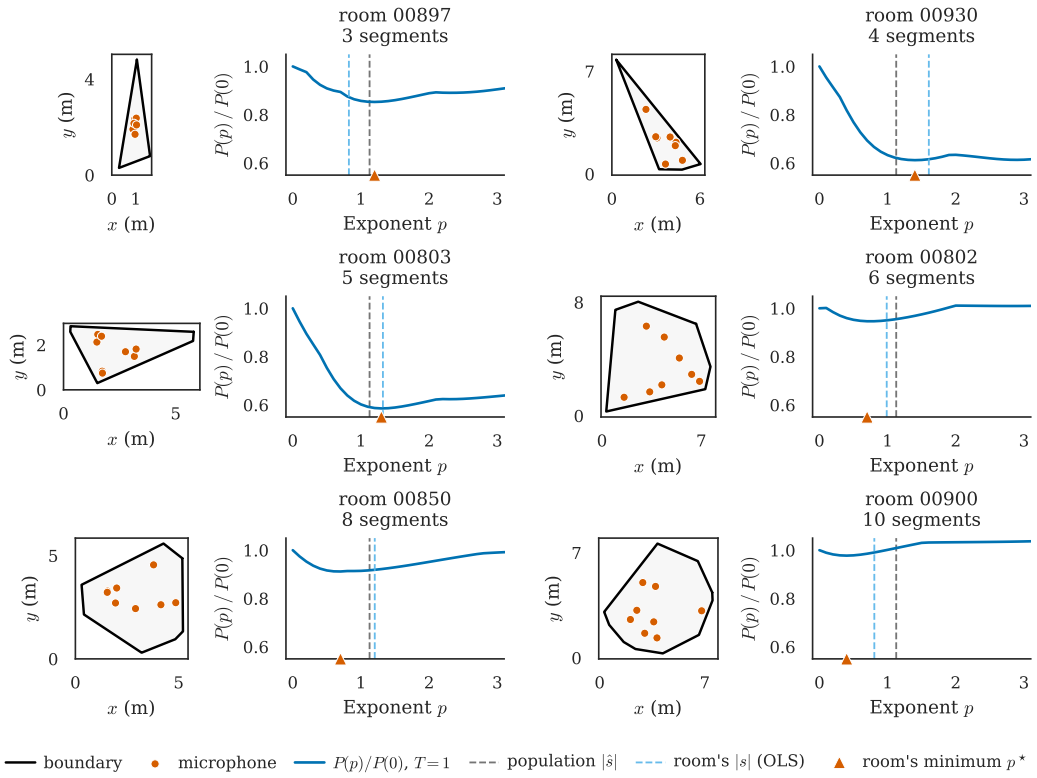


Figure 9: **A wide, flat basin appears in each of six room geometries.** Each row is one room, drawn with the eight microphones its curve uses (left). Its landscape at $T{=}1$ is divided by the ridge error, so values below 1 mean a gain over ridge (right). Per-room slopes and minima vary from room to room, yet the closed-form exponent stays inside every basin. The flat median landscape of Figure 2 holds room by room, not only on average.

$\tilde{\Phi}^\top\tilde{\Phi}/\lambda_{\max}(\tilde{\Phi}^\top\tilde{\Phi})$, d_k the mean of the two diagonal entries of G for mode k, and b_{kj} the Frobenius norm of the 2×2 block of G coupling modes k and $j\neq k$, the features are $\log\lambda_k$, $\log(c_s\sqrt{\lambda_k})$, $\log d_k$, $\log\max_j b_{kj}$, $\log\operatorname{mean}_j b_{kj}$, $2Kd_k/\operatorname{tr}G$, and $\max_j b_{kj}/d_k$, plus three entries shared by all modes: $\log M$, $\log T$, and the room damping rate γ. An $L{=}10$ unrolled gradient-descent solver minimizes $\frac{1}{2}\|Aa-y\|^2+\frac{1}{2}\alpha\, a^\top\operatorname{diag}(\Gamma)\,a$ with learned step sizes η_l and strengths α_l (10 of each); its conditioning input is fixed, so Γ is constant across iterations. The head acts on each mode independently, so M1 stays within the diagonal Tikhonov class.

M2: shared Γ with unrolled solver (70 parameters). One unconditional diagonal $\Gamma_k=\operatorname{softplus}(\theta_k)$, $\theta\in\mathbb{R}^{50}$, is shared by all rooms and solved with the same $L{=}10$ unrolled scheme and learned (η_l,α_l) as M1. M2 is the learned analogue of the closed form, fitted by SGD without physics; a non-power-law Γ from M2 that beat $|s|$ would reveal room-independent structure beyond the theory.

M3: CondNet with closed-form solver (4,930 parameters). M3 passes the CondNet output to a differentiable closed-form solve,

$$\hat{\mathbf{a}}(\Gamma)=\left(\tilde{\Phi}^\top\tilde{\Phi}+\alpha\operatorname{diag}(\Gamma^{(2K)})\right)^{-1}\tilde{\Phi}^\top\tilde{\mathbf{y}}, \tag{23}$$

where α is a single learned scalar; M1 and M3 differ only in this solve ($L{=}10$ gradient steps in M1) and are both differentiable end to end. The training loss is the per-sample $\|\hat{\mathbf{a}}-\mathbf{a}_0\|^2/\|\mathbf{a}_0\|^2$, with gradients through the inverse by implicit differentiation. M3 is the strongest diagonal baseline: it adapts Γ to each room and optimizes the evaluation metric directly.

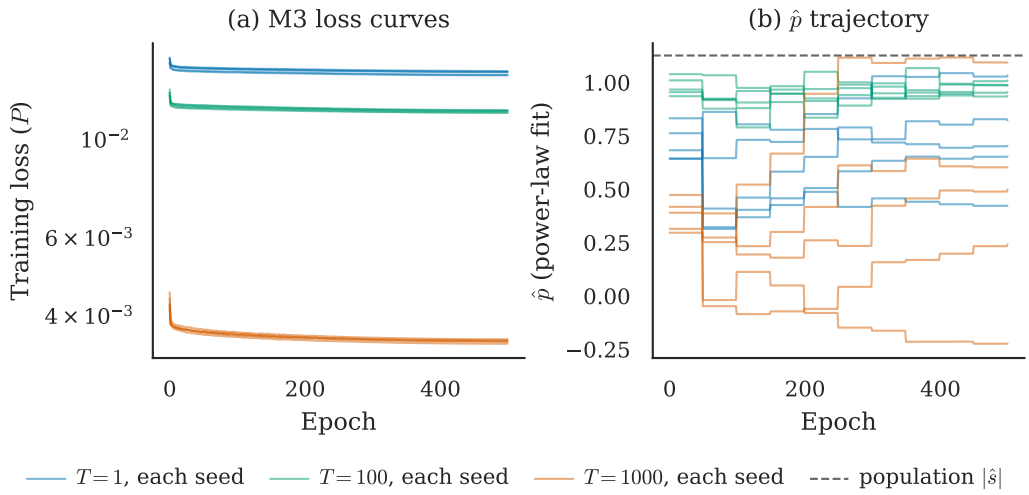


Figure 10: **M3 training converges while its seeds settle on different spectral exponents.** The training loss decreases smoothly and agrees across seeds at each observation window (a). The exponent of a power-law fit to the learned spectrum spreads across seeds, most widely at the longest window, where the spectra also fit a power law poorly (b). Equal loss at different exponents means the loss is nearly flat across the shapes the seeds reach, the basin that Section 6 reads as non-identifiability for M3.

Table 15: **M3's seeds end at different exponents at every window and agree least at the longest.** Each entry is the exponent of a power-law fit to one seed's learned spectrum, with the fit's R^2 below it. The spread widens and the fits worsen as the window grows, yet Table 3 shows equal error for every seed. This is the flat basin that Section 6 reads as non-identifiability for M3.

	seed 42	seed 43	seed 44	seed 45	seed 46
$\hat{p}$, T=1	0.65	1.03	0.42	0.70	0.82
$\hat{p}$, T=100	0.94	1.01	0.96	0.99	0.99
$\hat{p}$, T=1000	−0.22	0.24	0.50	0.60	1.09
R^2, T=1	0.76	0.79	0.63	0.65	0.91
R^2, T=100	0.80	0.82	0.83	0.83	0.78
R^2, T=1000	0.20	0.08	0.21	0.23	0.73

D.2 Training curves and learned spectra

Figure 10a shows that training converges: the loss decreases smoothly for all T and all 5 seeds, with no instability, overfitting, or collapse, and final losses agree closely across seeds. The learned spectra do not: the five seeds end at different exponents at every window, most widely at the longest (Table 15), yet all reach the same reconstruction error (Table 3). This is the most direct evidence of landscape flatness: the loss is nearly constant across the shapes the seeds reach, so nothing pulls them toward one shape. At the longest window the learned spectra also depart from a power law: the power-law fit worsens for every seed, sharply for four of the five (Table 15).

D.3 Room descriptors do not predict the oracle exponent

M3 might fail to beat $|s|$ because it lacks the right inputs; a model with room geometry (area, perimeter, vertex count) might predict $p^\star$. A regression of per-room $p^\star(T = 1000)$ on two tiers of features across 196 rooms (one room excluded for degenerate geometry) tests this: Tier 1 follows from the eigenvalue spectrum and is accessible at inference (Table 17); Tier 2 requires the room geometry (Table 18). No Tier 1 correlation survives Bonferroni correction across the nine features, and the strongest explains only a few percent of the variance (Table 17).

Table 16: **M3 and M2 settle on very different exponents at the longest window and reach almost the same error.** M3's per-room exponents match theory in the median, M2's shared exponent is about half of it, and the error at the two exponents is nearly equal. The capacity row, the limit of M2's parameterization (Section D.7), scopes the flat-basin reading of Section 6 to M3.

Quantity at $T = 1000$	Value
M3 per-room exponent $\hat{p}$, median [IQR] over the 187 in-scope rooms	1.13 [0.66, 1.39]
M2 shared exponent $\hat{p}$	0.62
R^2 of a power law fitted to the M2 spectrum	0.94
Difference in median P between the two exponents	0.009
Largest planted exponent that M2's parameterization recovers (Section D.7)	0.86
Smallest per-seed ΔP against the per-room oracle over all valid configurations	$+0.002$
Training configurations / numerical failures (all M3, $n \leq 400$) / failures at $T = 1000$	225 / 13 / 10

Table 17: **Tier 1 features, which are accessible at inference, do not predict $p^\star$.** Correlations are with the per-room oracle exponent $p^\star$ at T=1000, and none survives Bonferroni correction across both feature tiers. These descriptors therefore give the learned models of Section 6 little signal about the per-room exponent.

Feature	Spearman ρ	95% CI	p-value
Spectral gap ($\lambda_2 - \lambda_1$)	-0.15	$[-0.30,\ 0.00]$	0.03
Weyl exponent	$+0.07$	$[-0.09,\ 0.23]$	0.33
Per-room $\lvert s\rvert$	-0.04	$[-0.18,\ 0.11]$	0.58
λ_K	$+0.11$	$[-0.04,\ 0.25]$	0.13

Room area correlates significantly with $p^\star$ (Table 18), but area, K_{total}, mean spacing, and eigenvalue density are perfectly rank-correlated (Weyl's law: $K_{\text{total}} \propto \text{Area}$, spacing $\propto 1/\text{Area}$) and form one factor, room size (Figure 11). Even this strongest predictor explains only a few percent of the variance. A random forest on all nine features explains only a small fraction of the variance of $p^\star(T = 1000)$ (Table 18) and does worse than the mean at $T = 50$; the rest is noise or information that none of the features captures.

A K_{total}-adjusted exponent. The strongest simple rule, $\hat{p}_{\text{adj}} = \beta_0 + \beta_1 K_{\text{total}}$ fitted by OLS, trims the peak cost at $T = 1000$ by about a point (fit and evaluation both restricted to the in-scope rooms) but raises the cost at $T \leq 100$ by several points (Table 19); $|s|$ is the safer choice across operating conditions.

D.4 Robustness to dataset parameters

The main results use $K = 50$, $M = 8$, and $|s| = 1.13$. To test whether the pattern depends on this configuration, M3 is retrained on a parameter-varied dataset with $K = 100$, $M \in \{8, 16\}$, and an excitation regime with $|s| = 1.29$ (Table 20); at $K = 100$ a 10^{-8} diagonal jitter is added to M3's linear solve for conditioning. M3 trains on $n = 800$ rooms, and the medians run over 5 seeds and 197 validation rooms (167 in-scope with $K_{\text{total}} > K$ plus 30 boundary rooms). The one negative cell, at $M = 16$ and $T = 1$ (Table 20), reflects the gap between $|\hat{s}|$ and the per-room $p^\star$ at that operating point, not a shape advantage of the learned model. At $M = 16$ the overall P is lower than at $M = 8$, the landscape remains flat, and M3 still does not improve on the closed form; $p^\star$ depends weakly on M through the observation matrix, so exact agreement between $M = 8$ and $M = 16$ is not expected. $|s|$ itself is a per-regime quantity measured from data (1.13 and 1.29 here); what transfers is its role in $\Gamma_k = \lambda_k^{|s|}$.

Table 18: **Tier 2 features, which require room geometry, collapse to one room-size factor.** Correlations are with the per-room oracle exponent $p^\star$ at $T{=}1000$. All features except the segment count are perfectly rank-correlated, mean spacing inversely, as Weyl's law in Equation (10) implies. This factor is the strongest predictor of $p^\star$ among all tested features, yet it explains little of the variance and is unavailable at inference. Geometry would thus add only a weak room-size signal to the learned models of Section 6.

Feature	Spearman ρ	95% CI	p-value
Room area	$+0.21$	$[0.05,\ 0.37]$	0.003
K_{total}	$+0.21$	$[0.05,\ 0.37]$	0.003
Mean spacing	-0.21	$[-0.37,\ -0.05]$	0.003
Eigenvalue density	$+0.21$	$[0.05,\ 0.36]$	0.003
n_{segments}	$+0.14$	$[0.00,\ 0.29]$	0.05

Random forest on all nine features, cross-validated R^2 for $p^\star(T{=}1000)$: 0.14 (below the mean at $T{=}50$)

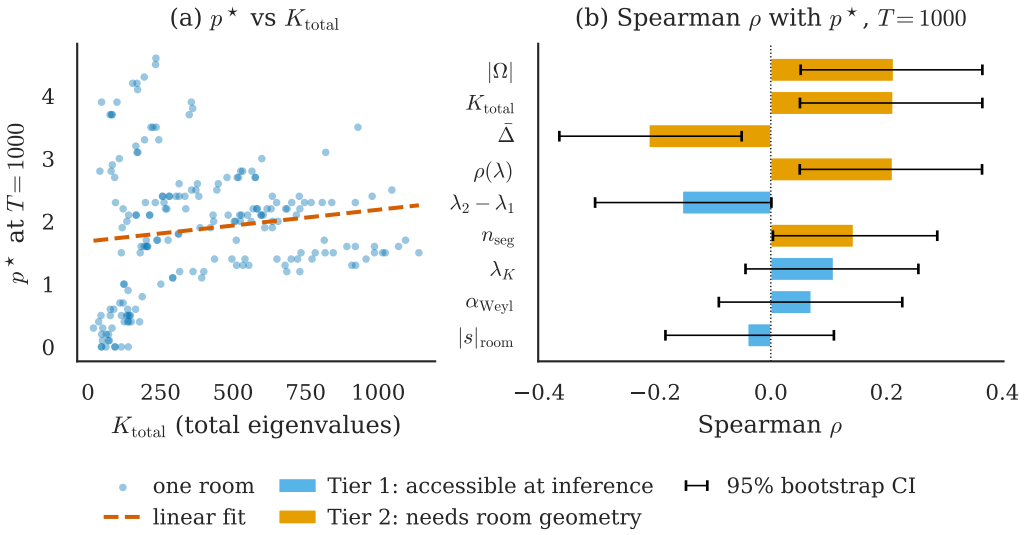


Figure 11: **No feature accessible at inference predicts $p^\star$.** The per-room oracle exponent rises only weakly with the total mode count (a). Correlations, with their confidence intervals, are weak for accessible features and largest for geometric proxies of room size (b). No tested input would therefore let the learned models of Section 6 track a room's oracle exponent.

D.5 Per-room versus population exponent

Does the oracle gap close if each room uses its own fitted exponent s_{room} instead of the population $|s|{=}1.13$? For all 197 rooms (boundary-inclusive, since the slope fit benefits from the largest sample), s_{room} is fitted from the retained modes ($K \le 50$) by the population log-log procedure, and $P(s_{\text{room}})$ is compared with $P(|s|)$ and with the per-room power-law oracle $P(p^\star)$.[5] The median per-room estimate is indistinguishable from the population value (Table 10). The point estimates spread across rooms by less than the median per-room bootstrap SE (Table 10), so the deconvolved inter-room spread is effectively zero;[6] this per-room error also exceeds the standard error of the population median by an order of magnitude.

[5]Ten rooms have $K_{\text{total}} \le 50$ and no discarded modes; they are excluded from the population fit (Appendix C.1), and their s_{room} is fitted on all K_{total} of their modes. Excluding them changes the correlations and the median costs only slightly and leaves the sign of $\delta(s_{\text{room}}) - \delta(|s|)$ unchanged at every T (Table 21).

[6]The method-of-moments estimate of the inter-room variance component is slightly negative under REML, a known boundary artifact when estimation noise exceeds the observed spread; it is reported as zero, consistent with per-room true values that are statistically indistinguishable from the population mean.

Table 19: **An exponent adjusted for K_{total} costs more than $|s|$ at every tested window except the longest.** The adjusted exponent is $\hat{p}_{\text{adj}} = \beta_0 + \beta_1 K_{\text{total}}$, and δ is the median relative cost against the per-room power-law oracle over the in-scope rooms, where lower is better. Its small gain at the longest window is outweighed by the losses at the other windows, so the closed form, whose cost Table 2 reports, remains the safer choice.

T	δ using $\lvert s\rvert$	δ using $\hat{p}_{\text{adj}}$
1	0.57%	2.9%
50	0.61%	4.1%
100	1.46%	5.0%
1000	5.83%	4.8%

Table 20: **On a dataset with more modes and a steeper excitation exponent, M3 beats the closed form in only one cell, and only slightly.** Entries are the median P over seeds and all validation rooms, boundary rooms included, and positive ΔP means M3 has higher error than the closed form. The one gain reflects the distance between $|\hat{s}|$ and the oracle exponent there, not a shape advantage. The saturation of Section 6 thus persists when K, M and the excitation exponent change.

	$T = 1$	$T = 100$	$T = 1000$
$M = 8$, M3	0.615	0.444	0.097
$M = 8$, Closed form ($\lvert\hat{s}\rvert$)	0.607	0.438	0.095
$M = 8$, ΔP	$+0.008$	$+0.006$	$+0.002$
$M = 16$, M3	0.473	0.254	0.078
$M = 16$, Closed form ($\lvert\hat{s}\rvert$)	0.477	0.251	0.076
$M = 16$, ΔP	-0.004	$+0.003$	$+0.002$

s_{room} carries some information about $p^\star$ at intermediate T (Table 21): the correlation is moderate at $T \in \{50, 100\}$, negligible at T=1, and weak at T=1000, a drop caused by the drift of $p^\star$ far from s_{room} (Table 12). Yet reconstruction with s_{room} improves on $|s|$ only marginally, at T=100 and T=500, and is worse at every other T, including T=50, where the correlation is strongest (Table 21); at T=1000 fewer than half of the rooms benefit from per-room tuning.

The mechanism is a noisy plug-in: $\hat{s}_{\text{room}}$ adds far more estimation noise than per-room signal, the setting in which shrinkage to the grand mean dominates per-unit estimation [James et al., 1961, Stein, 1956]; the analogy is qualitative because $P(p)$ is not quadratic in the exponent. At T=1000 a second effect adds to the first: $p^\star$ has drifted away from s_{room}, so even a noise-free s_{room} would target the wrong value. Per-room slope variability does not explain the oracle gap.

D.6 Per-seed sweep

Figure 12 compares every trained seed with both baselines. The main-grid baseline is the closed form at the grid value nearest $|\hat{s}|$ with oracle α, and against the per-room power-law oracle every per-seed ΔP is positive (Table 16). A small number of training configurations fail numerically, all in M3 at small training sizes and mostly at $T = 1000$ (Table 16), because $\tilde{\Phi}^\top \tilde{\Phi}$ is ill-conditioned; all main-text results use $n = 800$, where every seed succeeds.

D.7 Model capacity verification

Does M3 fail to beat $|s|$ because it cannot represent the optimal Γ? M3 is trained on synthetic data whose optimal Γ is a power law with known exponent $p_{\text{target}} \in \{0.5, 1.5, 2.5\}$ (prior $\sigma^2_{a,k} \propto \lambda_k^{-p_{\text{target}}}$). Figure 13 and Table 22 show that M3 recovers the targets monotonically, slightly below each, which is consistent with landscape flatness (a weak gradient near the minimum). M2 saturates below one because its softplus parameterization limits the reachable spectra, and M1 is non-monotonic, consistent with its weaker performance in the main experiments (Table 22).

M3 tracks power laws across the tested range, slightly below each target, so its failure to improve on $|s|$ is a property of the loss landscape, not of the architecture. Because M2 saturates and M1 is non-monotonic, the basin (non-identifiability) reading of Section 6 is scoped to M3.

Table 21: **The per-room exponent correlates with $p^\star$ most strongly at intermediate T but does not reliably cost less than the population $|s|$.** The relative cost δ against the per-room power-law oracle is a median over all validation rooms, boundary rooms included, and lower is better. Estimation noise in s_{room} outweighs its per-room signal, so per-room slopes do not close the oracle gap of Section 6.

T	$\rho_S(s_{\text{room}}, p^\star)$	p-value	R^2	$\delta(\lvert s\rvert)$	$\delta(s_{\text{room}})$
1	$+0.05$	0.45	$< 1\%$	0.55%	0.71%
50	$+0.40$	7×10^{-9}	16%	0.60%	0.88%
100	$+0.40$	7×10^{-9}	16%	1.46%	1.37%
500	$+0.14$	0.043	2%	5.13%	4.99%
1000	$+0.16$	0.026	3%	5.62%	5.84%
2100	$+0.18$	0.010	3%	3.33%	3.59%

Rooms that benefit from per-room tuning at T=1000: 88 of 197.
Excluding the ten boundary rooms changes each ρ_S by at most 0.03 and each median δ by at most 0.22 pp (ρ_S at T=1000: 0.14, $p = 0.065$).

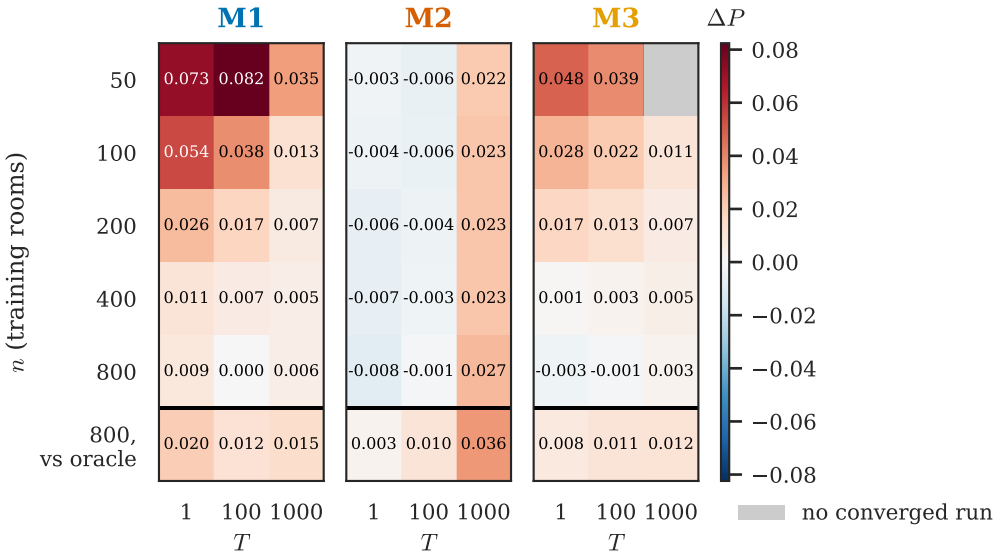


Figure 12: **No tested learned diagonal regularizer beats the power-law family.** Positive ΔP, a median over seeds, means higher error than the baseline. The baseline is the closed form with oracle α, except in the bottom row, which uses the per-room power-law oracle. Every valid trained seed is worse than the oracle, so small negative values against the closed form are no gain over the family. Grey cells had no converged run. This is the saturation of Section 6 in the truncation regime.

D.8 Learned Iterative Ridge (LIR)

Architecture. LIR parameterizes a linear estimator as L steps of learned gradient descent on a Tikhonov objective, starting from $a^{(0)} = 0$:

$$a^{(l)} = a^{(l-1)} - \eta_l\big(A^\top A\, a^{(l-1)} + \alpha_l\, D_l \odot a^{(l-1)} - A^\top y\big), \qquad l = 1, \dots, L, \tag{24}$$

where $\eta_l \in \mathbb{R}_+$ is a per-layer step size, $\alpha_l \in \mathbb{R}_+$ a per-layer strength, and $D_l \in \mathbb{R}_+^K$ a per-layer diagonal penalty shape on mode pairs. D_l is diagonal and positive, so that $\alpha_l D_l \odot a$ remains the gradient of a convex quadratic penalty ($52L$ parameters). Training uses Adam (learning rate 10^{-3}), 500 epochs (200 for $L = 1$), MSE loss on 800 training rooms, and 5 seeds $\in \{42, \dots, 46\}$.

Depth ablation. Table 23 and Figure 14 report P against depth L and window T. At $L = 1$ the map is a scaled adjoint $\hat{a} = \eta_1 A^\top y$, a constant filter that is worse than the oracle at all T; at $L \geq 5$ LIR is close to the closed form and above the oracle at all T, and only $L = 10$ at $T = 1$ is below the

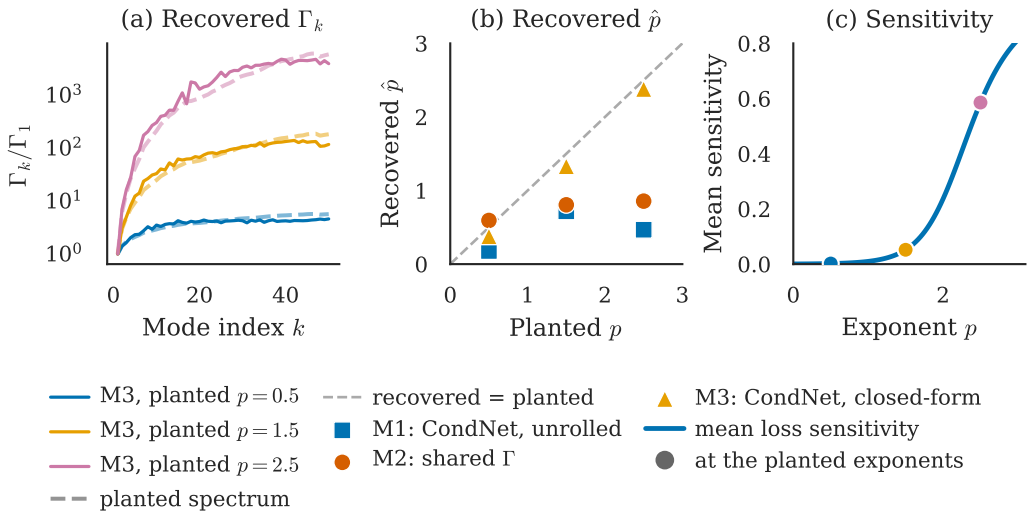


Figure 13: **M3 can represent the planted power laws, and M2 cannot.** With planted power-law priors, M3 learns spectra that follow the planted ones (a). M3's recovered exponent rises with the planted one, while M2 saturates and M1 is non-monotonic (b). M3 falls slightly short of every target, consistent with a loss that barely responds to the exponent (c). Its failure to improve on $|s|$ is a property of the landscape, not the architecture, so the basin reading of Section 6 applies to M3 only.

Table 22: **M3 recovers a planted exponent that rises with the target, M2 saturates, and M1 is not even monotone.** Each entry is the exponent of a power-law fit to the spectrum a model learns on synthetic data whose optimal shape is a known power law. M3 falls slightly short of every target, consistent with a flat loss, so the basin reading of Section 6 is scoped to M3.

Planted exponent p_{target}	0.5	1.5	2.5
Recovered by M1 (CondNet, unrolled)	0.18	0.72	0.47
Recovered by M2 (shared Γ)	0.59	0.81	0.86
Recovered by M3 (CondNet, closed-form)	0.38	1.33	2.38

closed form (Table 23). At $T = 1$, $L = 20$ has lower training loss than $L = 10$ but higher validation P; at $T = 100$ and $T = 1000$ the two depths agree within the spread over seeds.

Classical non-diagonal alternatives. Wiener/LMMSE with a diagonal prior and isotropic noise, generalized Tikhonov, TSVD, early-stopped CGLS, and Landweber iteration apply fixed per-mode shrinkage profiles in the modal basis, set by the prior, so they cannot improve on the closed form's

Table 23: **LIR does not reach the per-room power-law oracle at any tested depth, and improves on the closed form only at the shortest window.** Entries are the normalized modal error P (median over validation rooms) as mean and standard deviation over seeds, where lower is better. A single step is a constant filter and has the highest error in every column. This is the result outside the diagonal family that Section 6.1 reports.

L	$T = 1$	$T = 100$	$T = 1000$
1	0.908 ± 0.001	0.783 ± 0.000	0.368 ± 0.002
5	0.726 ± 0.002	0.624 ± 0.002	0.147 ± 0.000
10	0.720 ± 0.001	0.609 ± 0.003	0.141 ± 0.001
20	0.727 ± 0.004	0.607 ± 0.002	0.140 ± 0.001
Closed form	0.726	0.603	0.131
Oracle	0.715	0.594	0.122

Training loss at $T = 1$ (mean over seeds): $L = 10$ 0.709, $L = 20$ 0.702.

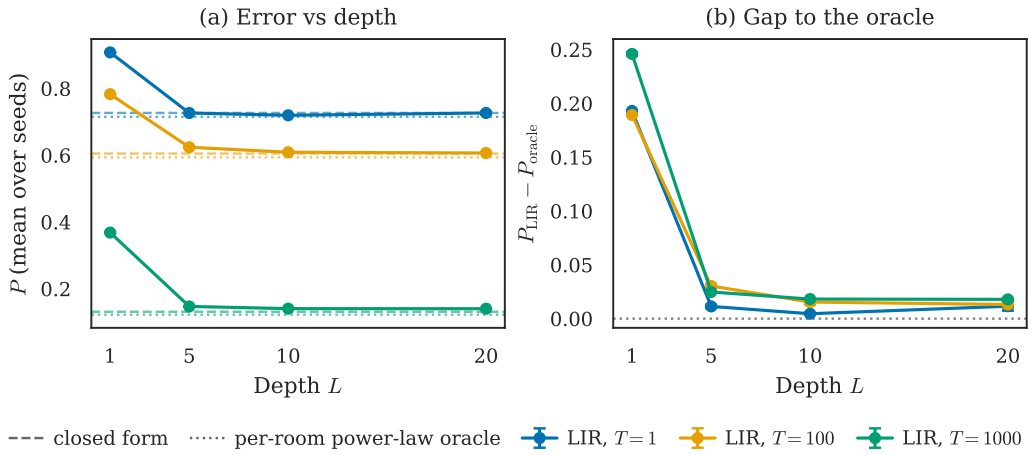


Figure 14: **LIR's error levels off by $L = 5$ near the closed form and stays above the oracle.** Depth beyond five layers changes the error little (a). Positive ΔP means LIR has higher error than the per-room power-law oracle (b). A single step is a constant filter, and composing several updates brings LIR to the closed form but not below the oracle.

Table 24: **Every geometry or noise descriptor that correlates with the gap at the longest window does so through room size, and none survives once area is controlled for.** Raw and area-adjusted Spearman correlations with the per-room relative gap over the in-scope rooms, with the mechanical reconstruction of the gap from drift and curvature and the out-of-sample prediction at shorter windows. This is why Figure 4 shows a confound rather than a mechanism.

Descriptor (in-scope rooms, $T = 1000$)	Raw Spearman ρ with δ	Partial ρ given area
Herfindahl index H	-0.42	-0.08
Room area and K_{total} (perfectly rank-correlated)	$+0.42$	–
$\|E\|_{\mathrm{op}}$	-0.30	$+0.10$
D_{KS} ($+0.19$ on the eligible subset of Section B.4)	$+0.16$	$+0.05$
Eigenvalue dynamic range $(\lambda_K/\lambda_1)^{\lvert s\rvert}$	-0.04	-0.03
Basin width on $p \in [0, 3]$	-0.64 ($p < 10^{-4}$)	–
All partial correlations have $p \geq 0.16$; $\rho(\text{area}, K_{\mathrm{total}}) = +1.00$ and $\rho(H, K_{\mathrm{total}}) = -1.00$.		
Drift-times-curvature reconstruction of δ, log-log R^2 at $T = 1$ / 50 / 100: 0.87 / 0.78 / 0.75.		
Random-forest cross-validated R^2 of the gap from all descriptors at $T \leq 100$: ≤ 0.		

shape: a signal-matched Wiener filter with prior $\Sigma_a = \mathrm{diag}(\lambda_k^{-s})$ recovers exactly $\Gamma_k = \lambda_k^{|s|}$, and Tikhonov soft shrinkage dominates TSVD [Hansen, 1998]. LIR instead composes L maps with $52L$ learnable coefficients and could in principle distribute regularization across coupled modes; as trained here, it does not consistently improve on the closed form (Table 23).

D.9 No single diagnostic predicts the residual gap

The regressions of Section D.3 target $p^\star$; what matters in practice is the per-room relative gap $\delta(T)$ to the per-room power-law oracle. On the 187 in-scope rooms the candidates are H, $\|E\|_{\mathrm{op}}$, the per-room Berry statistic D_{KS} ($\chi^2(1)$ convention, Section B.4), the eigenvalue dynamic range $(\lambda_K/\lambda_1)^{|s|}$, K_{total}, and area, plus two quantities that need no oracle argmin: the curvature of the α-minimized $P(p)$ at $p = |s|$ (second difference at $p \in \{1.0, 1.1, 1.2\}$; fully oracle-free) and the basin width $0.1 \times \#\{p \in [0, 3] : P(p) \leq 1.05 \min_p P\}$ (argmin-free, but it uses the minimum loss value). Mechanically, the gap is argmin drift converted by curvature: $\delta \approx \frac{1}{2}(P''/P)(p^\star)\,(p^\star - |s|)^2$ reconstructs the per-room δ well at $T \leq 100$ (Table 24), but it uses $p^\star$ and so explains without predicting.

At $T \leq 100$ no descriptor predicts the gap out of sample (Tables 25 and 24). At $T = 1000$ the raw correlations are moderate for H, area, and K_{total}, weaker for $\|E\|_{\mathrm{op}}$ and D_{KS}, and absent for

Table 25: **No accessible diagnostic predicts the residual gap.** Correlations are Spearman ρ with the per-room relative gap over in-scope rooms, and prediction is the cross-validated R^2 of a regression on all descriptors. Out of sample they predict nothing until the longest window, where only room size acts, and only the basin width, which needs the minimum loss, correlates strongly. The gap is thus a landscape property at long windows, not a diagnostic the learned models of Section 6 could exploit.

T	Geometry/noise descriptors	Curvature at $p = \lvert s\rvert$	Basin width on $p \in [0, 3]$
1	no out-of-sample prediction	$\rho \approx 0$	$\rho = +0.07$ ($p = 0.32$)
100	no out-of-sample prediction	$\rho = -0.17$ ($p = 0.02$), weak	$\rho = -0.04$ ($p = 0.55$)
1000	six descriptors give $R^2 = 0.18$ (relative gap) / 0.25 (absolute), driven by room size; same-size partial correlations are insignificant	$\rho = -0.04$ ($p = 0.61$)	$\rho = -0.64$ ($p < 10^{-4}$)

Table 26: **Heat modes die out many orders of magnitude faster than acoustic modes, a pure power law cannot follow them, and the two-parameter form recovers the predicted rate.** The table collects mode survival, fit quality of the two spectral forms, the oracle exponents each physics needs, absolute error levels, and how the heat truncation noise concentrates with time. They back the claim of Section 7 that the correction is predicted, not fitted.

Quantity	Value
Room 00840: energy retained by modes $k \geq 24$ at $t = 250$ ms / acoustic envelope at $t = 500$ ms	$< 10^{-9}$ / about 8%
Heat versus acoustic survival at $k = 49$, $t = 250$ / 495 ms (orders of magnitude)	17 / 34
OLS $\lvert\hat{s}\rvert$ on heat modal data, room 00840 / finite-k prediction for the five rooms (design value 1.0)	0.96 / 0.86–0.98
Spectral fit R^2, room 00840 at the three times of Figure 16: two-parameter form / pure power law	> 0.99 / 0.86–0.91
Residuals at late snapshots (log units): pure power law / two-parameter form	$\sim$30 / $\sim$0.3
Room 00840, pure power law at $t = 50$ / 250 / 495 ms: R^2 (fitted slope)	0.91 (3.6) / 0.87 (14.4) / 0.86 (27.5)
Pooled regression slope of $\hat{c}$ on c_{theory}, 95% CI	0.987 [0.953, 1.022]
Heat oracle exponent $p^\star$: $T = 1$ / $T \leq 200$ / $T \geq 1000$	2.5 / 2.5–2.8 / 3.0–3.2
Acoustic oracle exponent $p^\star$: $T = 1$ / $T = 1000$	1.4 / ≈ 2.2
P at $T = 2100$, $M = 8$: ridge / $\lambda_k^{1.13}$ / per-room one-parameter heat oracle (median $p^\star = 3.2$)	1.210 / 0.399 / 0.235
Heat Herfindahl index at the earliest observation time ($t \approx 12$ ms)	≈ 0.027
Heat effective contributors $1/H$: $t = 12$ ms / $t \approx 250$ ms	$\sim$20–80 / $\sim$1–6
Heat H at $t = 493$ ms: smallest room (00950, $K_{\mathrm{total}} = 93$) / median over the five rooms ($1/H$)	0.99 / 0.548 (≈ 1.8)

the dynamic range (Table 24). These are one correlation in several units, since area, K_{total}, and H are perfectly rank-correlated: after controlling for area none survives (Table 24), including the raw correlation of Figure 4. The only strong scalar at $T = 1000$ is the basin width (Table 24); the fully oracle-free curvature at $|s|$ is weak at every T. The residual gap is therefore a landscape property at long T, not a geometry or anisotropy diagnostic, and no single scalar predicts it across observation times; the learned models of Section 6 found none to exploit.

E Heat Equation: Extended Results

Section 7 predicts the two-parameter regularizer $\Gamma_k \propto \lambda_k^s \, e^{c\lambda_k}$ with $c = 2\kappa t$. All five-room analyses below use the diagnostic rooms 00805, 00826, 00840, 00880, and 00950.

E.1 Why heat diffusion differs from acoustics

In acoustics every mode decays at the same rate γ, so relative mode amplitudes stay fixed: the prior spectrum $\sigma_{a,k}^2 \propto \lambda_k^{-s}$ persists, and the optimal regularizer is a pure power law at every snapshot.

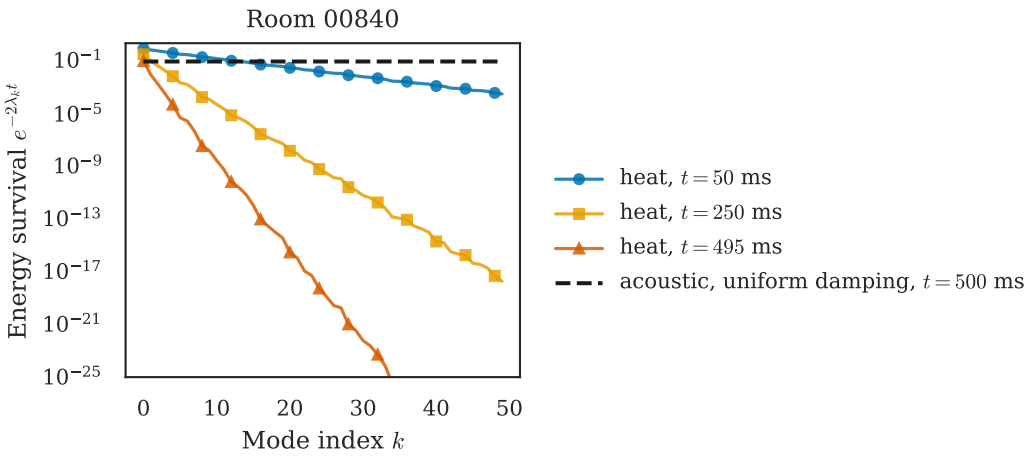


Figure 15: **High-frequency heat modes decay exponentially faster than acoustic modes, which all share one damping rate.** Survival is the fraction of a mode's initial energy left at time t, so on this log scale the steep heat curves mean that high modes carry almost no energy at late times. This mode-dependent decay is the exponential factor of Equation (12), which a pure power law cannot represent.

Heat diffusion breaks this: its Green's function damps mode k as $e^{-\kappa\lambda_k t}$, where κ is the thermal diffusivity, so high-frequency modes decay exponentially faster than low-frequency modes and soon carry negligible amplitude. The amplitude spectrum becomes a power law times an exponential,

$$\sigma^2_{a,k}(t) \propto \lambda_k^{-s} \cdot e^{-2\kappa t \lambda_k}, \tag{25}$$

where λ_k^{-s} carries the initial conditions, $e^{-2\kappa t\lambda_k}$ the temporal evolution, and the factor 2 comes from $\sigma^2_{a,k}$ being a variance. In room 00840, high heat modes retain almost none of their initial energy by the median observation time, while the acoustic envelope still holds a sizeable fraction at twice that time (Table 26). At high k, heat and acoustic survival differ by tens of orders of magnitude (Figure 15; Table 26); a one-parameter power law cannot suppress modes that decay this fast, so the heat regularizer needs the exponential factor.

Diffusivity convention. The synthetic experiments set $\kappa = 1$ in simulation units (eigenvalues in m^{-2}, time in s, so $\kappa\lambda t$ is dimensionless), a computational convenience rather than a material; realistic diffusivities range from $\sim 10^{-7}\,\mathrm{m}^2/\mathrm{s}$ (water) to $\sim 10^{-4}\,\mathrm{m}^2/\mathrm{s}$ (metals), and a physical deployment substitutes the material's κ into $c = 2\kappa t$. The predictions hold for any $\kappa > 0$; κ only shifts which snapshot indices fall in the regime where isotropy fails (Appendix E.5).

Heat exponent convention. The heat regularizer uses $|s| = 1.0$, the value implied by the synthetic initial-condition design $\mathrm{Var}(a_k(0)) \propto (1 + \lambda_k)^{-1}$. OLS fits on heat modal data return $|\hat{s}|$ slightly below one (Table 26); the deviation is the finite-k effect of the $(1 + \lambda_k)$ shift at small k, not an independent measurement. A physical deployment estimates $|s|$ from the system at hand with the log-log OLS procedure of the acoustic experiments (Section 5); the synthetic value does not transfer across excitation regimes (Section E.3).

E.2 The three-step fitting procedure

Step 1: Spectral fit. For each room and observation time t, OLS fits the empirical amplitude variances $\hat{\sigma}^2_{a,k}(t)$, $k = 1, \ldots, K$, in log space,

$$\log \hat{\sigma}^2_{a,k}(t) = \beta_0 - |s| \log \lambda_k - c\,\lambda_k + \epsilon_k, \tag{26}$$

with predictors $\log \lambda_k$ (power-law component) and λ_k (exponential component), yielding $\hat{s}$, $\hat{c}$, and R^2. The two-parameter model fits almost perfectly in the diagnostic rooms, a pure power law markedly worse (Figure 16; Table 26), and the power-law misfit is structural: its residuals are U-shaped at every

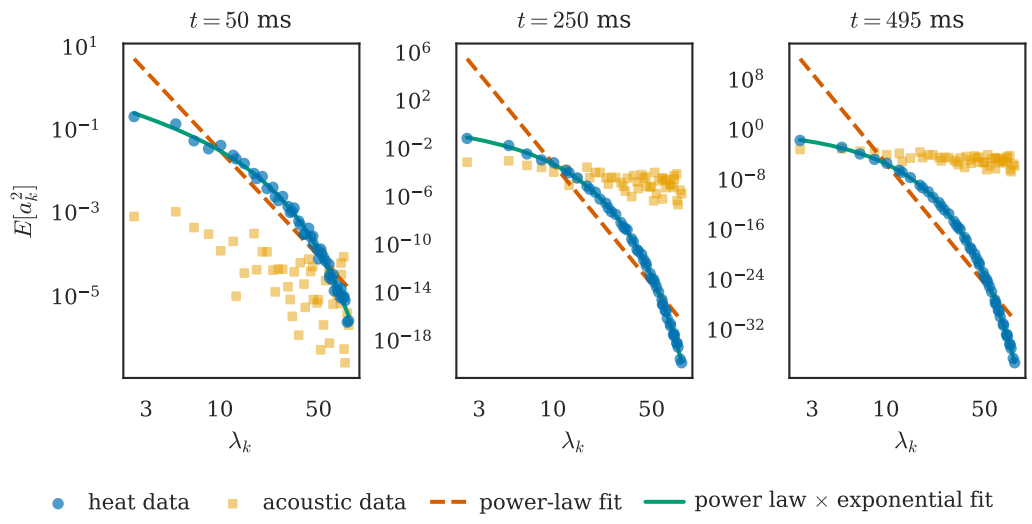


Figure 16: **An exponential-times-power-law form follows the heat amplitude spectrum, while a pure power law misses its roll-off at high eigenvalues.** Each panel plots mean squared modal amplitude against eigenvalue on log-log axes at one observation time, so a power law appears straight and the exponential factor bends it downward. The bend grows with time, as the rate $c = 2\kappa t$ predicts. This is the two-parameter spectrum of Equation (12), whose inverse is the heat regularizer.

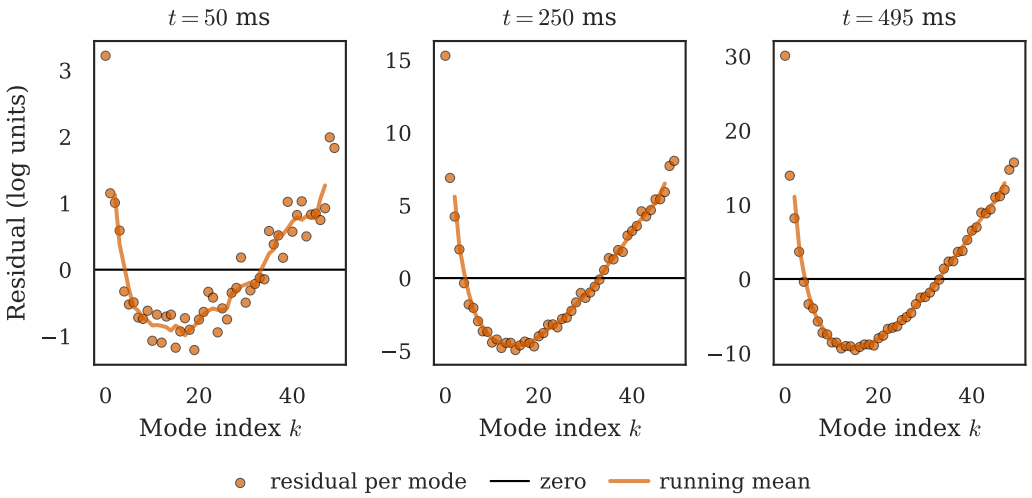


Figure 17: **A pure power law misfits the heat spectrum systematically, and the misfit grows with time.** Each point is the log-space residual of one mode, fit minus data, so zero is a perfect fit, and the curve is a running mean. The U-shape at every time is the signature of the missing exponential factor, which the power-law slope absorbs instead of estimating $|s|$. The two-parameter form of Equation (12) removes this structure (Figure 16).

snapshot and grow to tens of log-units at late snapshots, two orders of magnitude above those of the two-parameter model (Figure 17). In room 00840 the power-law fit worsens with observation time while its slope grows several-fold, absorbing the exponential factor instead of estimating the true $|s|$ (Table 26).

Step 2: Rate verification. The Green's function predicts the fitted rate without free parameters, $c_{\text{theory}} = 2\kappa t$, with κ known a priori (1 in the synthetic units; a material value in deployments) and t the observation time. For each diagnostic room, the estimates $\hat{c}(t)$ at multiple observation times are

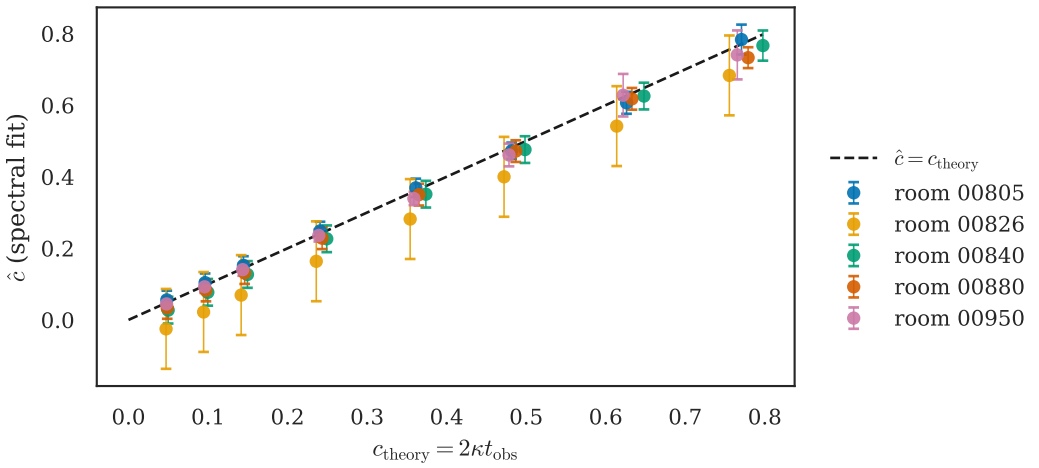


Figure 18: **The fitted exponential rate follows the prediction $c = 2\kappa t$ in every diagnostic room.** Each point is the rate fitted to one room at one observation time, with its standard error, so a point on the line means theory alone gives the rate. One room sits below the line by a constant offset that leaves its slope unchanged (Table 27). This supports setting the second parameter of Equation (12) from κ and t alone, a consistency check on synthetic data rather than physical validation.

Table 27: **The fitted rate follows c_{theory} in every diagnostic room.** Each room's row regresses $\hat{c}$ on c_{theory} across observation times, where a slope of one with an intercept near zero means agreement, and the pooled row fits all five rooms together. Figure 18 plots the underlying points of this check on the known second parameter of Equation (12).

Room	Slope α_1	Intercept α_0	R^2
00805	0.984	$+0.009$	0.9984
00826	1.000	-0.072	1.0000
00840	0.993	-0.021	0.9999
00880	0.973	-0.010	0.9990
00950	0.981	-0.002	0.9986
Pooled	**0.987**	-0.020	0.9880

regressed on the prediction,

$$\hat{c}(t) = \alpha_0 + \alpha_1 \cdot c_{\text{theory}}(t) + \epsilon, \tag{27}$$

and a correct theory gives $\alpha_1 \approx 1$ and $\alpha_0 \approx 0$. All five slopes lie close to one with near-perfect fits (Figure 18, Table 27), and the confidence interval of the pooled slope contains one (Table 26).

Step 3: Prediction, not post-hoc fit. The rate $c = 2\kappa t$ follows from κ and t before any data are collected; only $|s|$ is estimated from data. In these synthetic data $\hat{c} \approx 2\kappa t$ is expected, because the same eigenvalues enter the forward model and the modal basis, so the recovery is a consistency check on the fitting procedure; the independent test of the framework is the acoustic FDTD experiments of Section 5.

E.3 One-parameter fallback and two-parameter gain

If the optimal regularizer is $\Gamma_k \propto \lambda_k^s e^{c\lambda_k}$ but only $\Gamma_k = \lambda_k^p$ is searched, the fitted exponent rises above s to absorb the exponential roll-off at high modes. Figure 19(e) shows the rise: acoustic $p^\star$ stays near $|s| = 1.13$ up to $T = 200$, whereas heat $p^\star$ is already more than twice $|s|$ at $T = 1$ (the exponential is present even at a single snapshot) and rises further for $T \geq 1000$ (Table 26). The acoustic minima in panels (a) and (b) move up between $T = 1$ and $T = 2100$, the heat minima in panels (c) and (d) do the same, and in panel (e) heat $p^\star$ stays nearly constant for most $T \leq 200$ (Table 26).

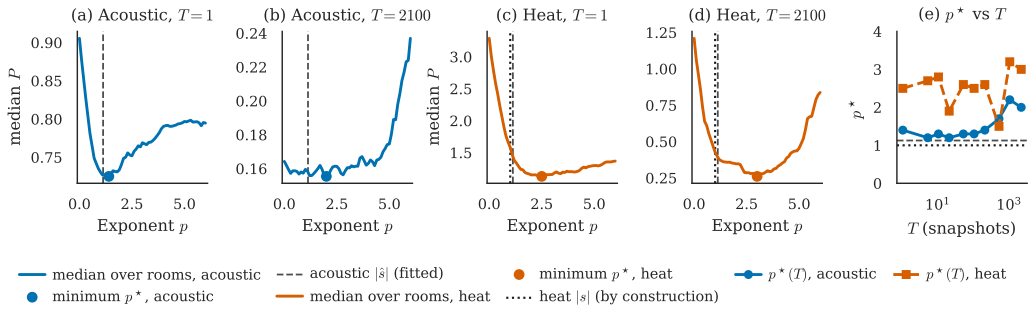


Figure 19: **Heat needs a steeper exponent than $|s|$ at every window, acoustics only at long ones.** Curves are room medians with their minima (a–d). The acoustic minimum moves up at a long window (b), and the heat minimum exceeds both references in a single snapshot (c) and moves up further at a long window (d). Both trends hold across all windows (e). The raised heat exponent absorbs the exponential factor of Equation (12), with the acoustic drift in Section A.4.

Table 28: **The second parameter barely lowers the population heat error, and its per-room gains are skewed.** $P^\star$ is the normalized modal error at the optimal parameters over the validation rooms; it exceeds one because heat amplitudes have a nonzero mean, which the variance in its denominator excludes. The improvement columns give the relative reduction when one c serves all rooms and its per-room median [IQR] when each room has its own parameters.

T	1-param $P^\star$	2-param $P^\star$	Pop. improv.	Per-room median [IQR]
1	1.275	1.256	1.5%	4.0% [0.1, 13.8]
10	1.244	1.244	0.0%	4.4% [0.2, 13.1]
100	0.871	0.871	0.0%	3.1% [0.0, 13.3]
500	0.738	0.727	1.4%	4.0% [0.0, 12.2]

Absolute performance ($T = 2100$, $M = 8$). Ridge ($p = 0$) has $P > 1$; the acoustic exponent $\Gamma_k = \lambda_k^{1.13}$ cuts that error to about a third, and the per-room one-parameter heat oracle lowers it by a further two fifths (Table 26). $|s|$ is system-specific ($|s|_{\text{heat}} \approx 1.0$ from the $t = 0$ spectrum, against $|s|_{\text{acoustic}} = 1.13$): the procedure transfers, the number does not.

Two-parameter gain. The population gain of the second parameter is small (Table 28) because one c must serve all rooms at fixed T; per room, the two-parameter oracle improves on the one-parameter oracle by a few percent in the median, with a skewed spread: many rooms gain nothing and a quarter gain more than a tenth. The difference reflects room-to-room variation in the optimal c, so for heat the per-room opportunity is not uniformly small.

E.4 Per-room spectral fits

Table 29 checks the rate prediction room by room. The slope $\hat{\alpha}_1$ and its fit quality $R^2_{\alpha_1}$ come from the Step 2 regression of Section E.2 across all 8 observation times; $\hat{c}$, c_{theory}, and $\hat{c}/c_{\text{theory}}$ are single-time spectral-fit estimates at the snapshot nearest the median observation time, where $c_{\text{theory}} = 2\kappa t$ with $\kappa = 1.0$ and t set by each room's own time step, so c_{theory} differs slightly from room to room. The across-time slope is close to one in every room. The single-time ratio is within a few percent of one in four rooms and lower in room 00826, whose single-time spectral fits are the least well determined, as the last column shows.

Two timescales. The *window duration* is $T \cdot \Delta t_{\text{sim}}$, with the FDTD step $\Delta t_{\text{sim}} \approx 24\,\mu$s set by the per-room CFL condition; the *observation time* is the snapshot index times $\Delta t_{\text{snap}} = 100 \cdot \Delta t_{\text{sim}} \approx 2.4$ ms. Each reported $P(T)$ averages over $\sim$190 valid snapshots per room spanning $t_{\text{obs}} \in [12, 493]$ ms with median $\sim$250 ms, a distribution roughly independent of T; the representative $t_{\text{obs}} \approx 250$ ms is that median.

Table 29: **Per-room spectral fits recover $\hat{c} \approx c_{\text{theory}}$ across observation times in all five diagnostic rooms, and at a single time in four of them.** A slope $\hat{\alpha}_1$ of one means that $\hat{c}(t)$ tracks c_{theory} across observation times, and a ratio of one means agreement at the single representative time. The last column is the R^2 of that single-time fit and accounts for the room that falls short. This room-level agreement backs Section 7, where the rate in Equation (12) is predicted rather than fitted.

Room	$\hat{\alpha}_1$	$R^2_{\alpha_1}$	c_{theory}	$\hat{c}$	$\hat{c}/c_{\text{theory}}$	R^2_{spec}
00805	0.98	0.9984	0.482	0.472	0.981	0.980
00826	1.00	1.0000	0.472	0.400	0.847	0.615
00840	0.99	0.9999	0.499	0.476	0.955	0.954
00880	0.97	0.9990	0.487	0.472	0.969	0.965
00950	0.98	0.9986	0.478	0.461	0.963	0.985

Table 30: **Heat diffusion concentrates the truncation noise into a few discarded modes as time grows, whereas acoustic noise stays spread over many modes at any time.** H is the Herfindahl index of Equation (11) with the decaying heat weights, $1/H$ counts the discarded modes that effectively contribute, and a small H means nearly isotropic noise. Proposition 1 needs isotropic noise, so with discarded modes present its diagonal optimality for heat would hold only at early observation times.

t_{obs}	Median H	H range (5 rooms)	$1/H$ range	Isotropy regime
12 ms	0.027	[0.013, 0.052]	[19, 76]	Holds (early)
250 ms	0.357	[0.156, 0.815]	[1.2, 6.4]	Degraded (median)
493 ms	0.548	[0.283, 0.987]	[1.0, 3.5]	Failed in smallest room
Acoustic	≈ 0.005	any t	≈ 200	Holds

E.5 Herfindahl index degradation under heat diffusion

In acoustics the Herfindahl index (Table 4) is time-independent because all modes share the damping rate γ. In heat diffusion the discarded-mode noise weights $w_n(t) \propto \lambda_n^{-s} \cdot e^{-2\kappa\lambda_n t}$ carry an exponential factor that suppresses high-frequency modes and concentrates the noise power into the lowest few discarded modes as t grows. Table 30 tracks this concentration through $H(t)$ and the effective contributor count $1/H(t)$. The simulated heat signals of Section 7 contain only the retained modes plus sensor noise, so they carry no truncation noise; the table is an analytic statement about the discarded modes a heat measurement would add. By $t_{\text{obs}} \approx 250\,\text{ms}$, about the median observation time, $1/H$ has dropped by an order of magnitude from its value at the earliest observation time, with the largest degradation in the smallest rooms (Table 26). At the latest observation time, H approaches one in the smallest diagnostic room, where the first discarded mode dominates because $e^{-2\kappa\lambda_{K+1}t_{\text{obs}}}$ is exponentially larger than every later factor; the median over the five rooms is about one half (Table 26), concentrated but not rank-1.

Why the spectral fit still works. The fits of Section 7 (Tables 27 and 26) verify that the signal-amplitude variance $\hat{\sigma}^2_{a,k}(t)$ follows $\lambda_k^{-s} \cdot e^{-2\kappa\lambda_k t}$, a property of the Green's function acting on the prior that does not require isotropic noise. The diagonal optimality of Proposition 1 does require isotropic noise and therefore holds at early t_{obs} only; at late t_{obs}, where H is concentrated, a non-diagonal estimator could in principle reach lower error than the diagonal regularizer.

What the heat experiments test. In the simulated heat data the only noise is sensor noise, which is isotropic at every observation time, so the late-time concentration of H does not enter them. With discarded modes present, the early snapshots, where H is still small (Table 26) and isotropy holds, carry the strongest signal and would dominate the average over the $\sim$190 snapshots of each $P(T)$; testing this needs heat data with truncation noise.

F Prior Robustness

The theory assumes a Gaussian, independent, power-law prior ($\mathbf{a} \sim \mathcal{N}(0, \Sigma_{\mathbf{a}})$ with $\Sigma_{kk} \propto \lambda_k^{-s}$), and the simulated data of the main experiments follow its power-law variance decay (Appendix C.1).

Table 31: **The alternative priors leave the noise at the sensors close to Gaussian, and shape freedom pays only once the truncation noise is removed.** The table collects how close the alternative priors stay to Gaussian noise, the largest movement under the correlated prior, and the one control in which the trained M3 beats the best power law. They support the scope statement of Table 1.

Quantity	Value
Rooms in the App. F test set / rooms among them with discarded modes	20 / 16
Largest KS gap between the noise marginals of the finite-variance alternatives and $\mathcal{N}(0,1)$ (187 in-scope rooms, 10 seeds)	0.016 (at most 0.017)
Empirical pooled ratio at $p = \lvert s\rvert$ under t_3 relative to the Gaussian prior	about 6% higher
Largest single-room movement under the correlated prior (M3 evaluation, $T = 1$, favors M3)	13.3 pp
No-truncation control: median gain of the trained M3 over the best power law at $T = 100$	2.8 pp

What happens when the prior is wrong? A heavy-tailed and a correlated alternative test whether $p^\star$, the cost of the closed form, and the landscape flatness change: with the discarded modes kept, so that truncation noise is present (Section F.2). Section F.4 repeats the comparison for the learned diagonal model M3.

F.1 Experimental setup

The three priors. All three share the marginal variance $\sigma_{a,k}^2 = \lambda_k^{-|s|}$ and differ in shape.

(a) **Gaussian (baseline).** $a_k \sim \mathcal{N}(0, \lambda_k^{-|s|})$, independent across modes: the prior the theory assumes.

(b) **Heavy-tailed: Student-t with $\nu = 3$.** $a_k \sim t_3(0, \lambda_k^{-|s|})$, independent across modes and scaled to the Gaussian variance, with infinite kurtosis; it models occasional extreme modal amplitudes, as from a localized impact. $\nu = 3$ is the most adversarial finite-variance choice: at $\nu = 2$ the variance is infinite, and at $\nu = 5$ the kurtosis is already 9.

(c) **Correlated: adjacent-mode correlation $\rho = 0.3$.** $\mathbf{a} \sim \mathcal{N}(0, \Sigma_{\mathrm{corr}})$ with diagonal entries $\lambda_k^{-|s|}$ and $(\Sigma_{\mathrm{corr}})_{jk} = \rho\sqrt{\lambda_j^{-|s|}\lambda_k^{-|s|}}$ for $|j-k| = 1$; it breaks the independence assumption and models a spatially extended source that excites neighboring modes together (larger ρ makes Σ_{corr} near-singular).

Protocol. For each prior, initial conditions are drawn for 20 rooms, and the closed-form Tikhonov solution sweeps p over the 61-point grid at $T \in \{1, 100\}$, recording the oracle exponent $p^\star$, the cost $\delta(|s|) = (P(|s|) - P(p^\star))/P(p^\star)$, and the flatness ratio $\max P / \min P$ over $p \in [0, 3]$. In the *primary* test, the prior drives all K_{total} modes of each room and the measurements are built from all of them, so the estimator, which models the retained $K = 50$, faces truncation noise as in Section 5; 16 rooms of the set have $K_{\mathrm{total}} > K$ and hence discarded modes (the other four have none). The primary test is evaluated as exact population risk (the risk of a fixed linear estimator depends on the prior only through its second moments, Proposition 1) and empirically over 24 independent seeds. In the *no-truncation control*, only $\min(K, K_{\mathrm{total}})$ modes are generated, so no truncation noise is present; the control uses one realization per room and lies outside the regime of Proposition 1. $T = 1000$ is omitted because without truncation noise $P \to 0$ for every p and δ is undefined in the control.

F.2 Results in the truncation regime (primary)

Table 32 gives the primary result, with $M = 8$ sensors. In population risk, all three priors give the same $p^\star$, equal to $|s|$ at grid resolution, with negligible $\delta(|s|)$ and the same flatness (Table 32). Empirically, the Gaussian and the t_3 prior give the same median $p^\star$ at $T = 100$ with a wide per-seed range, and the closed form's median oracle cost stays below one percent at both windows (Table 32).

Table 32: **In the truncation regime, a misspecified prior leaves $p^\star$ and the closed form's cost unchanged.** Entries are medians over in-scope rooms, where $\delta(|s|)$ is the closed form's relative extra error over the oracle exponent and a flatness ratio near one means p hardly matters. The upper block is exact population risk, the lower block empirical with per-seed ranges. Gaussian and variance-matched t_3 population rows coincide by construction, since a fixed linear estimator's risk depends only on second moments (Proposition 1).

Prior	T	$p^\star$	$P^\star$	$\delta(\|s\|)$	Flatness
Population risk					
Gaussian	1	1.1	0.659	0.1%	1.322
Gaussian	100	1.1	0.566	0.0%	1.274
Heavy-tail (t_3)	1	1.1	0.659	0.1%	1.322
Heavy-tail (t_3)	100	1.1	0.566	0.0%	1.274
Correlated	1	1.1	0.656	0.1%	1.329
Correlated	100	1.1	0.567	0.0%	1.285
Empirical, 24 seeds (median [range])					
Gaussian	1	1.0 [0.8, 1.4]	0.702	0.6% [0.0, 2.6]	1.253
Gaussian	100	1.1 [1.0, 1.4]	0.603	0.3% [−0.1, 1.6]	1.283
Heavy-tail (t_3)	1	0.9 [0.5, 1.6]	0.759	0.7% [0.0, 4.0]	1.182
Heavy-tail (t_3)	100	1.1 [0.6, 1.4]	0.658	0.7% [0.0, 3.5]	1.219
Correlated	1	1.0 [0.9, 1.4]	0.705	0.4% [0.0, 1.5]	1.252
Correlated	100	1.2 [1.0, 1.4]	0.601	0.3% [0.0, 1.2]	1.273

Normality and isotropy play different roles: in the generator, the truncation noise at the sensors is a linear combination of the discarded-mode amplitudes, and finite-variance alternatives stay close to Gaussian in marginal distribution (Table 31). One finite-sample effect remains: at the fixed estimator $p = |s|$ with α at the Gaussian population optimum, the empirical pooled ratio is somewhat higher under t_3 than under the Gaussian prior (Table 31). This is a property of the ratio under infinite kurtosis, not of the population risk, and it does not shift $p^\star$ toward zero.

F.3 Interpretation

Gaussian prior. With the truncation noise present, the population-risk $p^\star$ equals $|s|$ at grid resolution and the median δ is below one percent (Table 32).

Heavy-tailed prior. The variance-matched t_3 and Gaussian priors have identical population risk (Section F.1), so the heavy tail can change the closed form's cost only through finite-sample scatter, and empirically it leaves the median $p^\star$ and cost unchanged (Table 32).

Correlated prior. In the truncation regime (Table 32) and on all 187 in-scope rooms (Section F.4), the correlation moves the median cost of the closed form by a negligible amount (Table 33). The correlation $\rho = 0.3$ between adjacent modes adds mild off-diagonal structure to $\Sigma_{\mathbf{a}}$; the optimal Γ is no longer exactly diagonal, but the diagonal dominates and the diagonal power law remains a good approximation.

Summary. All three priors give flat landscapes, for three reasons. (i) Noise isotropy is a property of the noise (Weyl's mode count and the measured diagonal term), not of the prior. (ii) Landscape flatness is set by the discarded-mode count (Appendix G.4), not by the prior. (iii) Under isotropic noise at $T = 1$, the closed form is optimal for the Gaussian prior and remains the best *linear* estimator for any prior with the same second moments (Proposition 1); for non-Gaussian priors no quadratic Γ is Bayes-optimal among all estimators, and the flat landscape keeps that approximation error small. In the truncation regime these misspecifications leave $p^\star$ and the cost of the closed form unchanged (Section F.2). Priors with infinite-variance amplitudes, noise correlated with the signal, and strongly structured noise covariance are outside the scope of these tests.

Table 33: **Under a correlated prior the costs of the closed form and M3 move together, and the closed form stays below M3.** Entries are median relative costs against the per-room power-law oracle under exact population risk with truncation noise, with M3 as mean and standard deviation over seeds. Variance-matched t_3 rows are omitted because they coincide with the Gaussian rows by construction. M3 is no more robust to a wrong prior than the closed form, consistent with the saturation of Section 6.

Prior	T	Closed form $\delta(\lvert s\rvert)$	M3 shape δ
Gaussian	1	0.46%	$1.20 \pm 0.13\%$
Gaussian	100	1.41%	$2.74 \pm 0.13\%$
Correlated ($\rho = 0.3$)	1	0.41%	$1.11 \pm 0.15\%$
Correlated ($\rho = 0.3$)	100	1.35%	$2.78 \pm 0.20\%$

F.4 Learned diagonal models under misspecification

Are learned diagonal regularizers more robust to prior misspecification than the closed form? M3's conditioning features depend only on the room and the observation operator, so its emitted Γ is prior-independent, and the risk of any fixed linear estimator depends on the prior only through its covariance (Proposition 1). The closed form and M3 shapes trained with the protocol of Section 6 (five seeds at $n = 800$, evaluated under each prior rather than retrained for it, with α re-optimized over the paper's grid for every method) are therefore compared under exact population risk with truncation noise, on all 187 in-scope rooms at $T \in \{1, 100\}$ (Table 33); here the prior diagonal is each room's realized per-mode energy (the variance of its modal amplitudes) rather than $\lambda_k^{-|s|}$. Under the variance-matched t_3 prior both are unchanged by construction. Under the correlated prior ($\rho = 0.3$), the median costs of the closed form and of M3 each move by a negligible amount relative to the Gaussian prior, and the closed form's median cost stays below M3's at both T; the largest single-room movement is in M3's evaluation and favors M3 (Table 31). In these tests, learned diagonal models are not more robust to prior misspecification than the closed form: both move together. The learned models were not retrained under the perturbed priors or under the synthetic covariances of Section A.5; in naturally anisotropic rooms they do not beat the per-room power-law oracle (Section 6).

No-truncation control: where shape freedom pays. In the no-truncation control of Section F.1 (the same 20 rooms with the discarded modes removed synthetically), the same trained M3 checkpoints beat the best power law by a clear median margin at $T = 100$ (Table 31). That control has no discarded modes and hence no truncation noise, so it lies outside the regime of Proposition 1: removing the truncation noise restores landscape curvature (cf. the $K_{\text{total}} = K$ case of Section G.3), and shape freedom pays again. The saturation claim of Section 6 is therefore restricted to the truncation regime; this control is its one explicit exception.

G Rooms Where Berry's Premise Fails: Rectangles and 3D Boxes

Berry's conjecture is the only assumption in the isotropy argument that is neither an identity nor a calculation, so the direct test applies the closed form $\Gamma_k = \lambda_k^{|s|}$ where the conjecture fails. Rectangles with Dirichlet walls fail it by construction: they are integrable billiards whose eigenfunctions are products of sines, and their analytic eigenpairs give $P(p)$ without FEM discretization error. They also change the boundary condition of the main experiments (random convex polygons with Robin, or impedance, walls), so the test combines two departures from the main setting. The outcome: Weyl's discarded-mode count, not Berry conformity, flattens the landscape.

G.1 Setup

A rectangle $L_x \times L_y$ with Dirichlet walls has eigenfunctions and eigenvalues

$$\varphi_{mn}(x, y) = \frac{2}{\sqrt{L_x L_y}} \sin\Big(\frac{m\pi x}{L_x}\Big) \sin\Big(\frac{n\pi y}{L_y}\Big), \qquad m, n = 1, 2, 3, \ldots, \tag{28}$$

Table 34: **Rectangles reject the chaotic spacing statistics that Berry's conjecture presumes, yet a few dozen discarded modes flatten their landscapes nearly as far as those of generic rooms.** The table collects the spacing tests, the flattening of one rectangle with its mode count, the matched-count comparison, and the 3D box stress test. They separate the role of Weyl's count from the role of Conjecture 1.

Quantity	Value
NNSD of generic convex rooms (48 rooms, about 13,000 spacings): D_{KS} against GOE (p)	0.010 (0.17)
NNSD of rectangles (about 2,000 spacings), D_{KS} against GOE: 3×6 / 2×8 / 4×4 / 3×5 (all $p < 10^{-5}$)	0.40 / 0.34 / 0.58 / 0.20
3×6 room, flatness ratio at $K_{\mathrm{total}} = 50$ / 75 / 100 / 150 / 200 / 287 (generic reference)	3.41 / 1.29 / 1.26 / 1.24 / 1.23 / 1.22 (1.28)
3×6 room at $K_{\mathrm{total}} = 50$: condition number of $\tilde{\Phi}$ / discarded modes at its full K_{total} (Weyl estimate)	$\sim 10^8$ / 237
Discarded modes that bring the flatness ratio within 5% of the generic reference	about 25
Matched-count cohort: generic in-scope rooms with $K_{\mathrm{total}} \geq 300$	101
Flatness ratio with no discarded modes ($K_{\mathrm{total}} = 50$), median [range] over 10 seeds: generic / rectangles	2.34 [2.28, 2.35] / 2.24 [2.09, 2.64]
$D/(2H)$ at $K_{\mathrm{total}} = 300$: generic IQR upper bound / rectangle minimum	2.38 / 5.80
Flatness medians at $K_{\mathrm{total}} = 300$: difference / rectangle median inside the generic IQR	0.04 [0.004, 0.07] / four or five of six checkpoints
$D/(2H)$ over the full band: matched cohort / all in-scope rooms	2.24 / 2.16
3D boxes: volumes / K_{total} below 500 Hz / box-and-wall cells / sensor layouts / realizations per layout	18–126 m^3 / 169–1,921 / 8 / 10 / 20
3D worst cells over 10 layouts, relative cost (absolute): $T \leq 100$ / $T = 1000$	3.4% [2.7, 3.8] (2.8 pp) / 20.4% [15.6, 34.2] (1.7 pp)
2D boundary-inclusive maxima: $T \leq 100$, relative (absolute) of room 00806 / $T = 1000$, largest relative / largest absolute (different rooms)	27.5% (10.61 pp) / 34.8% / 5.33 pp

$$\lambda_{mn} = \pi^2\Big(\frac{m^2}{L_x^2} + \frac{n^2}{L_y^2}\Big). \tag{29}$$

Their nodal lines are straight and parallel to the walls, so at random sensor positions the cross-correlations C_{kn} follow the distribution of products of sines rather than a Gaussian. The four test rooms measure 3×6, 2×8, 4×4, and 3×5 m (areas 15–18 m^2; $K_{\mathrm{total}} = 287, 255, 255, 239$, Weyl's estimate of the count with $\lambda \leq 200\,\mathrm{m}^{-2}$); the 4×4 square is the most symmetric, with frequent eigenvalue degeneracies, and the 2×8 room has a 4:1 aspect ratio and a distinct eigenvalue distribution. Each room retains $K = 50$ modes, with $M = 8$ sensors placed uniformly at random, and $P(p)$ is evaluated at $T = 100$.

G.2 Eigenvalue spacing: Berry's premise fails

The standard diagnostic for "quantum chaos" is the nearest-neighbor spacing distribution (NNSD) of the eigenvalues, unfolded to mean spacing 1 to remove the density trend of Weyl's law. Here $s_i = (\lambda_{i+1} - \lambda_i)/\bar{\Delta}$ is the normalized spacing, as on the axis of Figure 20a, not the prior exponent. Integrable systems such as rectangles are expected to follow the Poisson density $f(s) = e^{-s}$, and chaotic systems such as generic convex polygons, the setting of Berry's conjecture, the GOE density $f(s) = \frac{\pi s}{2}\, e^{-\pi s^2/4}$.

Generic convex rooms match the GOE closely, consistent with Berry's conjecture (Figure 20a; Table 34). Rectangular rooms reject it decisively, every one of them, and two of the four lie closer to Poisson, while the 3×6 and 4×4 rooms are as far from Poisson as from GOE because their degeneracies produce exact-zero spacings (Table 34). The 4×4 square deviates most: its degeneracies produce level clustering, the opposite of GOE level repulsion.

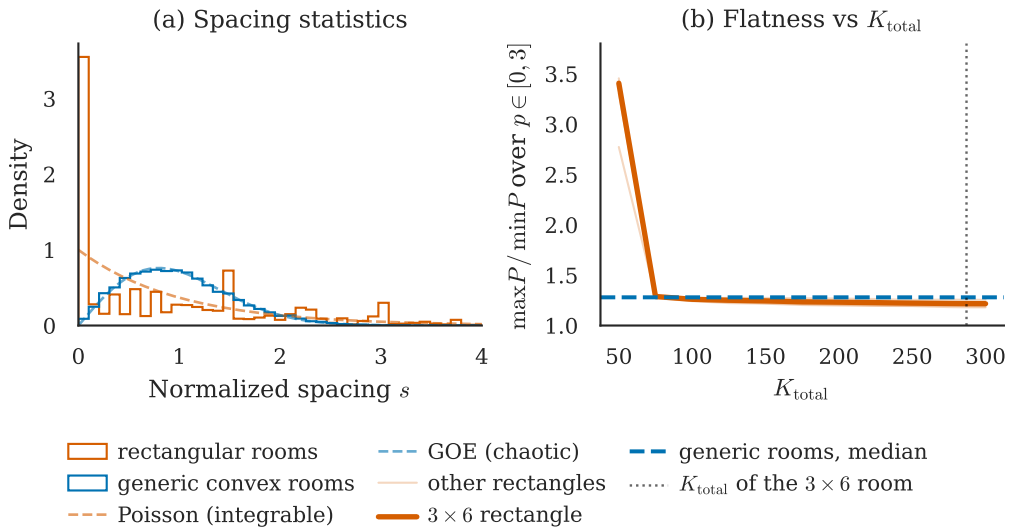


Figure 20: **Rectangles violate Berry's premise, yet discarded modes flatten their landscapes to the generic-room level.** Generic convex rooms show the GOE spacing of chaotic systems and rectangles do not (a). A ratio of 1 means the exponent does not matter (b). The rectangle curve is one click-excited realization, so its no-truncation anchor is not comparable cell by cell with Table 35. Weyl's law (Equation (10)) gives practical rooms hundreds of discarded modes, and this count, not conformity to Conjecture 1, sets the flattening.

G.3 Flatness versus discarded-mode count

The choice of p depends not on whether Berry holds but on whether the $P(p)$ landscape curves, which the flatness ratio $\max_p P / \min_p P$ over $p \in [0, 3]$ measures: near 1.0 the choice of p hardly matters.

Control experiment. For the 3×6 room, the mode set is truncated at progressively larger K_{total}. At $K_{\text{total}} = K = 50$ there is no truncation noise and the regularizer alone stabilizes the inverse; as K_{total} grows, the truncation noise grows and, with more modes contributing, becomes more isotropic.

Results. At $K_{\text{total}} = 50$ the landscape is strongly curved and $p^\star = 0$ (Table 34). With neither truncation nor measurement noise, shrinkage only costs accuracy: the best α is the smallest value of the α grid at every $p \leq 2$, and because every $\lambda_k > 1$, raising p strengthens the penalty in the directions that $\tilde{\Phi}$ resolves poorly ($\tilde{\Phi}$ has full column rank, $MT = 800 > 2K = 100$, but is badly conditioned; Table 34). This anchor therefore depends on the excitation and on the lower limit of the α grid, not only on the absence of discarded modes (Figure 20b). Adding a few dozen discarded modes brings the ratio most of the way down; by $K_{\text{total}} = 100$ it reaches the generic convex reference, and it keeps falling slowly up to the room's full mode count (Table 34).

Why the landscape flattens without Berry. A sum of many rank-one contributions concentrates even when its terms are neither Gaussian nor independent, provided there are enough of them with bounded cross-correlations; here there are hundreds (Table 34). Weyl's law makes $K_{\text{total}} - K$ grow with room area, so any room of practical size ($> 1\,\text{m}^2$) has hundreds of discarded modes, an order of magnitude more than the few dozen that bring the flatness ratio within a few percent of the generic reference (Table 34).

G.4 Weyl's count versus Berry's constants

The closed form works because the discarded-mode count supplied by Weyl's law flattens the landscape, and Berry conformity does not change the flattening rate (Figure 20b, Table 35). The two classical results play separate, measurable roles:

Table 35: **At matched K_{total}, Berry conformity separates generic rooms from rectangles, but landscape flatness does not.** Berry's conjecture predicts $D/(2H) = 1$, and a flatness ratio of 1 means the exponent does not matter. Entries are medians over a fixed cohort of in-scope generic rooms and the analytic rectangles of Section G.1 at $T = 100$; flatness is the median [range] over ten seeds. The two families share H, which Weyl's count fixes together with the prior, while Conjecture 1 enters only through D in Equation (21).

K_{total}	H	$D/(2H)$: generic / rectangle	Flatness: generic / rectangle
75	0.0407	1.92 / 1.63	1.33 [1.32, 1.35] / 1.37 [1.34, 1.43]
100	0.0210	1.31 / 2.59	1.30 [1.29, 1.32] / 1.30 [1.27, 1.39]
200	0.0081	1.82 / 4.85	1.29 [1.27, 1.30] / 1.31 [1.24, 1.34]
300	0.0055	1.89 / 6.09	1.28 [1.27, 1.29] / 1.29 [1.21, 1.34]

(i) Weyl's law fixes the number of discarded modes and, with the prior decay, the Herfindahl index H; at matched K_{total}, generic rooms and rectangles have the same H, as Weyl predicts.

(ii) The off-diagonal second moments of E equal H exactly for any orthonormal eigenbasis (Section A.2), so the sum of rank-one matrices concentrates whatever the distribution of its terms.

(iii) Berry's conjecture governs only the diagonal term D of $\mathbb{E}\|E\|_F^2 = M(M-1)H + MD$, through $D \approx 2H$. In rectangles $D/(2H)$ is several times its generic-room value (Table 4); the single-mode part is exactly $S/(2H) = 0.625 = (9/4 - 1)/2$, since the fourth moment $C_4 = \mathbb{E}_x[\psi_n(x)^4]$ equals $9/4$ for sine products.

Matched-count comparison. Table 35 applies one protocol (diffuse-field prior with $|s| = 1.13$, $K = 50$, $M = 8$, $T = 100$) to generic rooms and to the analytic rectangles at matched discarded-mode counts. The generic entries come from the in-scope rooms with $K_{\text{total}} \geq 300$ at every checkpoint (Table 34), so they describe one population; the rectangle entries have exact eigenpairs and $D/(2H)$. Most of the flattening happens before Berry conformity separates the families: with no discarded modes the two families are equally curved within the scatter over seeds, and a few dozen discarded modes flatten both to nearly the same level (Tables 35 and 34). At $K_{\text{total}} = 300$ the $D/(2H)$ ranges do not overlap, yet the flatness medians are nearly identical, and the rectangle median lies inside the generic IQR at nearly every checkpoint (Table 34). The generic $D/(2H)$ dip at $K_{\text{total}} = 100$ is reproducible across the cohort and absent in rectangles; we do not interpret it, and it leaves the flatness medians unchanged. The comparison caps the band at $K_{\text{total}} = 300$; over the full band, $D/(2H)$ on the cohort is close to its value on all in-scope rooms (Table 34, Section A.2). The no-truncation anchor of Figure 20b is not comparable with the table: it is a single room under click excitation, which places $p^\star$ at 0, and that anchor depends on the excitation, whereas the flattening with mode count does not.

Berry's conjecture pins the diagonal term at $D = 2H$, but the closed form does not need it: anisotropy up to the mean observed level leaves the cost of the fixed exponent essentially unchanged, and stronger anisotropy raises it noticeably only at the longest window (Section A.5). The mechanism does require $K_{\text{total}} - K \gg 1$; it fails for the smallest rooms, with $K_{\text{total}} \approx K$, which support too few modes below the frequency ceiling for meaningful acoustic reconstruction and lie outside the framework's intended use.

G.5 3D box stress test

All other experiments use 2D rooms. A targeted 3D stress test along the Berry-failure axis evaluates the closed form on analytic boxes of four volumes ($2.5{\times}3{\times}2.4$, $3{\times}4{\times}2.5$, $4{\times}5{\times}2.7$, and $6{\times}7{\times}3$ m; 18–126 m^3) under Dirichlet and Neumann walls: one cell per box and wall type, each with all modes below 500 Hz, $K = 50$, $M = 8$ uniform sensors, the diffuse-field prior with $|s| = 1.13$, and repeated realizations for each of ten sensor layouts under the same 61×12 oracle protocol (Table 34). Box eigenfunctions are separable, so Berry's premise fails by construction, with $D/(2H)$ an order of magnitude above generic rooms and growing with mode count (Table 4, Section A.2).

Table 36: **Separable 3D boxes have flatter landscapes than the 2D polygons except at the longest window.** Costs are the gap between the closed form and the per-room power-law oracle, in percentage points and in percent of the oracle error. The boxes use analytic modes in a lower band than the FDTD rooms of Section 5, so levels are not comparable. Boxes break Berry's premise like the rectangles of Section G.2, but as separable domains they do not stand in for general 3D rooms.

Quantity	2D polygons	3D boxes
Cases	187 in-scope rooms	8 cells × 10 layouts
Median K_{total}	337	530
Median H	0.0050	0.0027
Median landscape ratio, $T = 1/100/1000$	1.265 / 1.280 / 1.452	1.089 / 1.138 / 1.529
Cost, $T = 100$	0.87 pp (1.46%)	1.03 pp (1.45%)
Cost, $T = 1000$	0.62 pp (5.83%)	1.57 pp (13.58%)
3D cost range over layouts, $T = 100$ / 1000	–	0.69–1.43 / 1.39–2.10 pp

Table 36 compares the medians with the 2D rooms. The protocols differ in solver and band (analytic box modes at 500 Hz versus the 2D FDTD pipeline at 1000 Hz), so H, flatness, and the anisotropy decomposition transfer, but absolute cost levels are not directly comparable; absolute and relative medians are taken independently across cases, so the relative column is not the ratio of the absolute column to the median oracle floor. All eight 3D cells lie within the observed 2D room-to-room range for every sensor layout: the worst cell costs several times less than the worst 2D room at $T \leq 100$ and less, by up to a factor of two, at $T = 1000$ (Table 34). The 3D landscape is flatter than the 2D landscape at $T \leq 100$ but not at $T = 1000$, where the large-T oracle drift of Section C.4 recurs. The test does not replace numerical validation on non-separable 3D rooms, which remains open, as do structured arrays and realistic sensing.

H Sensor Noise and Model Mismatch

Real measurements add electronic sensor noise, calibration mismatch, and model mismatch to the truncation noise that the main text treats as dominant. None of them changes the recommended shape at typical operating points, and the Herfindahl index flags risky configurations before deployment.

H.1 Electronic sensor noise

Each microphone adds electronic noise $\epsilon_m \sim \mathcal{N}(0, \sigma^2_{\text{elec}})$, independent across sensors and time, so the total noise covariance is

$$R = R_{\text{trunc}} + \sigma^2_{\text{elec}}\, I_M = (\sigma^2_{\text{trunc}} + \sigma^2_{\text{elec}})\, I_M + \sigma^2_{\text{trunc}}\, E. \tag{30}$$

The electronic component is exactly isotropic and the truncation component approximately so, hence $R \approx (\sigma^2_{\text{trunc}} + \sigma^2_{\text{elec}})\, I_M$. By Proposition 1, the optimal single-snapshot shape under isotropic noise is $\Gamma_k \propto \lambda_k^s$ at any noise level; the level sets only the strength, $\alpha = (\sigma^2_{\text{trunc}} + \sigma^2_{\text{elec}})/c$. Sensor noise therefore calls for stronger regularization, not a different shape, whichever source dominates.

H.2 Mild sensor-noise anisotropy

Manufacturing variation, calibration drift, or age give microphones slightly different noise levels, $R_{\text{elec}} = \text{diag}(\sigma^2_{\text{elec},1}, \ldots, \sigma^2_{\text{elec},M})$, so $R = \sigma^2_{\text{trunc}}(I_M + E) + R_{\text{elec}}$. This anisotropy adds to that of E and could curve the $P(p)$ landscape if it were comparable to the truncation anisotropy. It is not while electronic noise stays well below the truncation noise, as the main text assumes: a calibration mismatch below 3 dB ($< 2\times$ in power) spreads the electronic noise levels by less than their mean, so it perturbs $\|E\|_{\text{op}}$ by less than the electronic-to-truncation noise ratio, which is then below the median truncation anisotropy (Table 4, Appendix A.2), where the controlled sweep of Section A.5 finds the cost of the fixed exponent essentially unchanged.

H.3 Frequency-dependent damping

Eq. (2) gives every mode the same damping γ, which preserves relative amplitudes over time. In real rooms absorption by walls, furniture, and air increases with frequency; the simple model

Table 37: **The noise structure decides whether the closed form suffices.** The Herfindahl index H follows from the eigenvalues before any data are collected, and a low value means the truncation noise is spread over many discarded modes. Low H keeps the off-diagonal anisotropy small by Equation (9), close to the isotropic noise that Proposition 1 assumes.

Diagnostic	Recommendation
Low H, many discarded modes, and small modal tilt q	Use the closed form and tune only its scalar strength α.
Modal noise varies approximately as λ^q	In a decoupled approximation, use exponent $\lvert s\rvert + q$ and validate it in the coupled model (Section A.2).
Few or no modes discarded, or strongly structured temporal/sensing covariance	Proposition 1 is not predictive; validate alternative estimators on calibration data.
Structured sensor arrays	Not covered by the current random-sensor analysis.

$\gamma_k = \gamma_0 + \gamma_1 \lambda_k$ gives the amplitude variance

$$\sigma^2_{a,k}(t) \propto \lambda_k^{-s}\, e^{-2\gamma_1 \lambda_k t}, \tag{31}$$

the heat-equation spectrum of eq. (12) with γ_1 in the role of κ. As in Section 7, the estimator then targets the current-state amplitudes $a_k(t)$ rather than $a_k(0)$, and the regularizer acquires the exponential correction

$$\Gamma_k \propto \lambda_k^{s}\, e^{2\gamma_1 \lambda_k t}. \tag{32}$$

If γ_1 is known from absorption measurements or material data, the correction is a prediction; otherwise the fallback is the one-parameter power law with an elevated $p^\star$, as for heat (Section E.3). Below $500\,\mathrm{Hz}$, the modal range of typical rooms, $\gamma_1 \lambda_K t \ll 1$ for the retained modes, so uniform damping is a reasonable approximation and $\Gamma_k = \lambda_k^{|s|}$ suffices; at higher frequencies or with strong frequency-dependent absorption the correction may become relevant.

H.4 Practical diagnostics

For a new (K, M) configuration, the Herfindahl index follows from the eigenvalues alone, before any data are collected:

$$H = \frac{\sum_{n>K} \lambda_n^{-2|s|}}{\left(\sum_{n>K} \lambda_n^{-|s|}\right)^2}. \tag{33}$$

Values $H < 0.01$ indicate noise power spread over many discarded modes, where $\Gamma_k = \lambda_k^{|s|}$ is expected to work; $H > 0.1$ indicates that a few discarded modes dominate and that the closed form needs validation before deployment. These markers are rough guides set by the range of H in the tested rooms at $M = 8$ (Table 5), not thresholds fitted to cost. The risky direction is increasing K toward K_{total}: H rises and isotropy weakens as the truncation band shrinks, and at $K = K_{\mathrm{total}}$ the framework does not apply. The pattern persists at $M \in \{8, 16\}$ (Appendix D.4). Small uncertainty in $|s|$ does not move this boundary, since a subset estimate of the kind tested in Section C.1 costs less than half a point (Table 13); Table 37 maps each noise structure to an action.

I Aperture Constraint

A static compact array cannot validate the framework in a real room (Section 8): its aperture is small against the retained wavelengths (Section I.1) at every room size compatible with the framework's assumptions (Section I.2). A moving sensor would lift the constraint (Section I.3).

I.1 Aperture-to-wavelength bound

Recovering K modal amplitudes from M microphones requires the spatial matrix $\Phi_{mk} = \varphi_k(x_m)$ to have effective rank K. For a compact array of aperture D_{ap} and shortest retained wavelength $\ell_{\min} = 2\pi/\sqrt{\lambda_K}$, each eigenfunction varies between any two microphone positions $x_m, x_{m'}$ of the array by at most

$$|\varphi_k(x_m) - \varphi_k(x_{m'})| \lesssim 2\pi\, \frac{D_{\mathrm{ap}}}{\ell_{\min}}\, \|\varphi_k\|_\infty. \tag{34}$$

Table 38: **A compact array's aperture-to-wavelength ratio stays small except in rooms too small for the framework.** Eigenvalues are those of a rigid-walled (Neumann) box, D_{ap}=12.6 cm, and K_{total} counts modes below 1 kHz. The last column is the largest difference a unit plane wave at the shortest retained wavelength can show across the aperture. By Equation (34), a small ratio concentrates the spatial matrix in a few directions, so no room in the framework suits a static compact array.

Room	Volume (m^3)	K_{total}	ℓ_{min} (m)	$D_{\mathrm{ap}}/\ell_{\mathrm{min}}$	Max amplitude variation
Large (3.45 × 7.20 × 2.45 m)	60.9	7017	2.07	6.1%	37.9%
Compact (1.33 × 2.10 × 2.47 m)	6.9	879	0.988	12.7%	78.0%
Closet (1.0 × 1.0 × 1.0 m)	1.0	141	0.535	23.6%	134.9%

When $D_{\mathrm{ap}}/\ell_{\mathrm{min}} \ll 1$, every column of Φ is approximately a constant vector times a mode-dependent factor, so the energy of Φ concentrates in a few singular directions regardless of M, and modal projection amplifies noise by the reciprocals of the small singular values. The limit is structural, not a matter of signal-to-noise ratio: more microphones within the same aperture or longer recordings do not supply spatial information that was never measured. The remedies are an aperture with $D_{\mathrm{ap}}/\ell_{\mathrm{min}} = O(1)$ or sampling at spatially distinct positions over time.

I.2 Values for three room scales

Table 38 evaluates the bound for K=50 retained modes in three room scales with a typical portable array (D_{ap}=12.6 cm). A compact array resolves modes only when the room is small enough for the retained wavelengths to approach the aperture. Shrinking the room further creates two problems: the source enters the array's near field, which violates the far-field assumption of the observation model, and K_{total} drops to where the mode-count argument no longer holds (the closet has fewer than three times the K=50 retained modes below the ceiling; Table 38). No room size therefore makes a static compact array a valid testbed for K=50 modal recovery within the framework's assumptions.

I.3 Resolution by sampling along a trajectory

A large array gives up portability; a moving sensor keeps the array compact and gains spatial diversity through motion, turning M static microphones into $M \cdot N$ measurement points over N positions along a trajectory. The isotropy argument needs only the sample average $\frac{1}{N}\sum_n \varphi_k(x_n)\,\varphi_l(x_n)$ to be small for $k \neq l$, and a sufficiently mixing trajectory in Ω induces a sampling distribution whose expectation converges to the orthogonality relation of the eigenfunctions, so the premise of Section 4 carries over from static random placements to trajectories. Which trajectory minimizes the reconstruction error under a path-length or duration budget, and its empirical validation, are left to follow-up work.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Abstract and §1 state four contributions, each backed by §4–§7. Isotropy is framed as approximate throughout.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: §8 covers 2D-only scope, Gaussian prior (Appendix F), Berry failure in rectangles (Appendix G), 3D as future work, the open non-diagonal estimator question, and physical validation via a compact-array real-data pilot (Appendix I).

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Proposition 1 lists assumptions and is proved in Appendix A. Berry is stated as a conjecture, verified in §5 and Appendix B.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: All settings specified: $K{=}50$, $M{=}8$, 800/197 train/val split, p-grid, α selection (§5); optimizer, lr, epochs, seeds (Appendix D.8); heat parameters (Appendix E); FEM data generation (§3, Appendix C.8); real-data pilot setup (Appendix I).

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: We will make the code public upon acceptance.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Data split, hyperparameters, optimizer, seeds, and loss are in §5 and Appendix D.8. The main formula has no free parameters beyond $|s|$.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Bootstrap 95% CI for $|s|$ (§5), mean $\pm$ std over 5 seeds for LIR (Appendix D.8), KS p-values (§5, Appendix B), IQR shading in figures.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: CPU: Intel i5-12400F, GPU: NVIDIA RTX 4090, 32 GB RAM. Closed-form Tikhonov runs on CPU in seconds. LIR training ($52L$ parameters) takes $\sim$5 min/seed on GPU. Total compute for all experiments: < 24 GPU-hours.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Synthetic FEM eigenmodes (main experiments) and author-recorded RIRs in an empty $6.9\,\mathrm{m}^3$ room (Appendix I). No human subjects, personal data, or dual-use concerns.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: Foundational signal-processing theory with no foreseeable negative societal impact.

    Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: Released assets are synthetic FEM eigenpairs, small networks, and RIRs from a small room. No misuse risk.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: All data authored by us: synthetic FEM eigenmodes plus RIRs from a $6.9\,\mathrm{m}^3$ room (Appendix I). No external datasets or licensed code.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [No]

    Justification: Code will be made public upon acceptance: FEM eigenpairs, real-data RIRs and sensor/source metadata, matched synthetic comparison, scripts, pipeline.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: No crowdsourcing or human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: No human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs used for writing and editing, not for experiments.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.