PHYSICAL INFORMATION, NUMERICAL RESOLUTION, AND LEARNED APPROXIMATION IN BORN SCATTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning wave-scattering operators presents three main challenges: resolving transmitted information, controlling discretization error, and approximating the operator from data. Using quadrature-weighted finite-aperture Born operators and DeepONets, we analyze their interplay. Operator spectra across frequency bands show a plateau and a steep tail, with a shifting knee that complicates global power-law fitting. Effective rank identifies key transmission directions. Receiver refinement shows first- or second-order convergence (rectangle or trapezoidal quadrature), making receiver requirements dependent on the integration rule and accuracy, while aperture extension affects physical coverage. In two-band studies, linear branches outperform MLPs; with width 64, wavelength matching cuts high-band linear-model errors by 66% and 52%. Refitting coefficients in learned bases removes most in-distribution error, but training-based POD with ridge prediction is more accurate at equal data, revealing an approximation gap. Full-wave checks find Born discrepancies of 4.5%–10.4% at nominal contrast, quantifying model accuracy but not neural full-wave generalization. Overall, we clarify the roles of physics, numerics, and learning, showing accuracy is distribution-dependent and not universally set by effective rank.

1 INTRODUCTION

Scientific machine learning accelerates repeated evaluations of physical models by learning maps between function spaces. DeepONet and Fourier Neural Operators provide frameworks for such surrogates (Lu et al., 2021; Li et al., 2021; Kovachki et al., 2023). In wave imaging, frequency band and aperture determine which medium variations are strongly represented in measurements, while discretization determines how accurately that acquisition is represented. These effects matter for receiver resolution, spectral benchmarks, and learned forward-map capacity.

We study finite-aperture, band-integrated Born operators mapping small medium perturbations to scattered acoustic measurements. Does changing the frequency band alter the singular-value decay law, or primarily the number of strongly transmitted directions? A widening plateau followed by a steep tail can require a larger approximation rank without a slower asymptotic decay law. Receiver density at fixed aperture refines a discretization; extending the aperture changes physical coverage. Learned approximation also depends on the input distribution, norm, architecture, and optimization.

Why the distinctions matter. Spectral truncation, reduced-order approximation, and neural-operator error analysis connect approximation quality to singular values and covariance spectra (Hansen, 1998; Lanthaler et al., 2022; Kovachki et al., 2023). A global power-law fit can confound a moving spectral knee with a changing exponent. Compactness alone does not specify an algebraic law; for a square-integrable Born kernel, an asymptotic law $\sigma_j \asymp j^{-\alpha}$ with $\alpha \leq 1/2$ would violate square summability. A finite-window slope cannot establish that asymptotic law. Moreover, the optimal worst-case error for a p-dimensional linear output space on the input unit ball is σ_{p+1}, whereas Frobenius truncation error involves $(\sum_{j>p} \sigma_j^2)^{1/2}$ and mean-square error for structured random inputs depends on output covariance. These are distinct objectives.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

A nominal Nyquist count likewise does not determine quadrature accuracy on a truncated receiver interval: endpoint effects remain relevant (Trefethen & Weideman, 2014). Inconsistent quadrature weights can confuse changes in the discrete norm with changes in physical information. Spectral fits, capacity arguments, and receiver requirements therefore need explicit operator, distribution, weighting, and geometry conventions.

Our approach and contributions. We combine quadrature-weighted, matrix-free spectral experiments with empirical POD and trained-network comparisons. Adjoint tests, dense-SVD comparisons, and refinement checks assess numerical reliability.

- **Spectral shape and physical dimension.** Local slopes reveal a moving knee and steep resolved tail across the tested bands. Effective rank increases from 51.5 to 319.5, while the phase-space scaling exponent 0.876 exposes departures from proportional mode counting.
- **Receiver fidelity and aperture coverage.** A fixed-mode, reference-checked experiment measures approximately first- and second-order receiver refinement for left rectangle and trapezoidal quadrature across two bands and two array layouts. We separate this numerical convergence from physical aperture extension.
- **Distribution-specific learning diagnostics.** We distinguish operator-SVD projection from empirical POD, real from complex rank, and learned-span from coefficient error. Frozen-basis refitting recovers nearly all ID coefficient error and improves tested OOD predictions, while a training-POD-plus-ridge baseline exposes a remaining basis gap at equal sample budget. Disjoint p/N_{eff} ranges prevent a neural collapse test; operator rank alone does not determine prediction or sample complexity. Separate grid-refined full-wave comparisons quantify the physical accuracy of the Born targets on held-out input subsets.

The paper is structured as follows to guide the reader through the context, methodology, and findings of this study. Sect. 2 reviews related work. Sect. 3 details the construction of the quadrature-weighted Born operator, introduces diagnostics based on effective rank, and outlines the numerical and analytical procedures used for verification. Sect. 4 formulates five key research questions and defines quantitative metrics to distinguish the contributions of physical information, numerical resolution, and learned approximation in the operator learning process. Sect. 5 presents a comprehensive set of experiments, analyzing spectral structure, effective-rank scaling, receiver convergence, and prediction accuracy. Sect. 6 summarizes the main findings, addresses study limitations, and proposes future directions.

2 Related Work

Neural operators and approximation limits. DeepONet (Lu et al., 2021) and FNO (Li et al., 2021) learn maps between function spaces, with analyses connecting error to approximation widths, regularity, and data complexity (Lanthaler et al., 2022; Kovachki et al., 2023; 2024). Linear-decoder restrictions connect operator learning to reduced-order modeling and POD (Pinkus, 1985; Benner et al., 2015). Neural approaches to full-waveform inversion and PDE inverse problems emphasize architectures and physics-constrained training (Jin et al., 2022; Cho & Son, 2025). Our focus is complementary: measuring the forward operator's effective dimension, calibrating learned-decoder errors against consistent projection references, and testing whether decoder width normalized by effective rank organizes prediction errors. We do not estimate sample-complexity scaling.

Scattering resolution and physical information dimension. Diffraction tomography and inverse scattering relate recoverable model wavenumbers to frequency and acquisition geometry (Devaney, 1982; Wu & Toksöz, 1987; Beylkin, 1985; Colton & Kress, 2019). These principles underpin seismic resolution analysis and full-waveform inversion (Virieux & Operto, 2009). The accessible wavenumber set depends on source–receiver geometry; an annular description requires additional idealizations. Our phase-space predictor uses this classical foundation to compare spectral knees across bands and apertures. For learned forward maps, we distinguish illumination from output covariance: the latter determines optimal mean-square approximation by a fixed-dimensional linear output space.

Spectral computation and discretization accuracy. Weighted discretizations relate discrete spectra to continuum operators (Hansen, 1998); randomized SVD and block Krylov methods provide scalable estimation (Halko et al., 2011; Musco & Musco, 2015). On truncated receiver arrays, accuracy also depends on finite-interval quadrature (Trefethen & Weideman, 2014), a different issue from internal-grid aliasing in FNO (Lanthaler et al., 2024). We separate physical information dimension from numerical receiver cost. Within a verified order-ν refinement regime, an error proportional to $n_r^{-\nu}$ implies a tolerance-dependent receiver requirement proportional to $\varepsilon^{-1/\nu}$.

3 METHODOLOGY

3.1 WEIGHTED ACQUISITION OPERATOR

For a homogeneous background c_0, let $G(x, y; \omega) = \frac{i}{4} H_0^{(1)}(\omega|x - y|/c_0)$ and $u_0(x, x_s; \omega) = S(\omega)G(x, x_s; \omega)$. Here S includes the Helmholtz source normalization. Linearizing in the velocity perturbation δc gives

$$\delta u(x_r, x_s, \omega) = \beta(\omega) \int_\Omega G(x_r, x; \omega) G(x, x_s; \omega) \delta c(x)\, dx, \qquad \beta(\omega) = -\frac{2\omega^2 S(\omega)}{c_0^3}. \tag{1}$$

Let m_j be model-cell weights and $d_{isr} = w_i q_s q_r$ the product of frequency, source, and receiver quadrature weights. With $B_{isr,j} = \beta_i G(x_r, x_j; \omega_i) G(x_j, x_s; \omega_i)$, we use $A = M_d^{1/2} B M_m^{1/2}$, where $M_m = \mathrm{diag}(m_j)$ and $M_d = \mathrm{diag}(d_{isr})$. Thus A maps $M_m^{1/2}\delta c$ to $M_d^{1/2}\delta u$, making its singular values comparable across discretizations (Hansen, 1998). Source windows remain separate from quadrature weights. Algorithm 1 streams frequency blocks, merging source and receiver indices into one matrix multiplication without storing the full A.

Algorithm 1 Weighted Born operator and adjoint actions

Require: Model-side block X (forward) or data blocks $\{Y_i\}$ (adjoint)
1: Initialize adjoint accumulator $Z = 0$, when needed
2: **for** each frequency i **do**
3: $\quad K_i[(s, r), j] \leftarrow G(x_r, x_j; \omega_i) G(x_j, x_s; \omega_i)$
4: $\quad D_i \leftarrow \beta_i\, \mathrm{diag}_{sr}(\sqrt{d_{isr}})$
5: $\quad$ **Forward:** emit $D_i K_i M_m^{1/2} X$
6: $\quad$ **Adjoint:** $Z \leftarrow Z + M_m^{1/2} K_i^{\mathsf{H}} D_i^{\mathsf{H}} Y_i$
7: **end for**
8: **return** emitted forward blocks or adjoint block Z

3.2 PHYSICAL DIMENSION AND EFFECTIVE RANK

In the high-frequency approximation, the local scattering covector is $\xi(x) = \frac{\omega}{c_0}(\widehat{x - x_s} + \widehat{x - x_r})$, where $\widehat{v} = v/\|v\|$. Consequently, $|\xi| \le 2\omega/c_0$; equality requires coincident source and receiver directions. For the continuum acquisition, let $\mathcal{S}(x)$ collect these covectors over all source positions, receiver positions, and frequencies. The phase-space counting heuristic is

$$N_{\mathrm{phase}} \approx \frac{1}{(2\pi)^2} \int_\Omega |\mathcal{S}(x)|\, dx. \tag{2}$$

This is a geometric dimension estimate, motivated by Weyl asymptotics (Hörmander, 1985), rather than an exact threshold-rank formula. Under ideal full-angle backscattering, $\mathcal{S}(x)$ is an annulus and the estimate reduces to $N_0 = |\Omega|(k_{\max}^2 - k_{\min}^2)/(4\pi)$, with $k = 2\omega/c_0$. For surface arrays, we retain N_0 as a reference scaling variable; the accessible set generally depends on position and opening angle. We define N_{eff} by the linearly interpolated crossing of $\sigma_j/\sigma_1 = 0.1$ and test $N_{\mathrm{eff}} \approx \gamma N_0$. The fitted $\gamma > 0$ includes geometry, amplitudes, and the threshold convention; it is not assumed to be a coverage fraction.

3.3 SPECTRAL ESTIMATION AND CONTROLLED DIAGNOSTICS

We draw a Gaussian sketch Z with $p + s$ columns and construct a model-side block Krylov basis (Musco & Musco, 2015). Its span is generated by $(A^{\mathsf{H}}A)^{\ell}Z$, $\ell = 0, \ldots, q$, where p is the target rank, s is oversampling, and q is the depth. Two-pass reorthogonalization and rank deflation stabilize basis construction. Algorithm 2 gives the streaming TSQR path (Demmel et al., 2008), which computes the projected spectrum without storing AQ or squaring its condition number. A Gram-based shortcut requires separate accuracy validation against this path.

Algorithm 2 Block Krylov spectrum with streaming TSQR

Require: Actions of A, A^{H}; target rank p, oversampling s, depth q
Draw Gaussian Z; $Q_0 \leftarrow \mathrm{orth}(Z)$; $Q \leftarrow Q_0$
for $\ell = 1, \ldots, q$ **do**
  $V \leftarrow A^{\mathsf{H}}(AQ_{\ell-1})$ ▷ accumulate frequency by frequency
  Apply $V \leftarrow V - Q(Q^{\mathsf{H}}V)$ twice
  $Q_\ell \leftarrow \mathrm{orth}(V)$, dropping numerically dependent columns
  Stop if Q_ℓ is empty; otherwise $Q \leftarrow [Q, Q_\ell]$
end for
Initialize empty R with as many columns as Q
for each streamed row block T of AQ **do**
  $R \leftarrow \mathrm{qr}_R([R; T])$ ▷ retain only the triangular factor
end for
return the largest p singular values of R

Diagnostics. We estimate $\alpha(j) = -d \log \sigma_j / d \log j$ by weighted least squares in sliding windows of half-width 0.15 decades, testing whether slopes remain constant across resolved modes. We fit γ and separately regress $\log N_{\mathrm{eff}}$ on $\log N_0$, testing the unit-slope hypothesis rather than imposing it. Window ablations compare rectangular, Hann, and Gaussian S at fixed band. Receiver refinement compares rectangle and trapezoidal quadrature at fixed geometry and window; a Hann receiver taper is a separate change of data weighting. A reproducible spectral-error definition uses fixed reference mode indices. For a reference spectrum σ^{ref} and a fixed index set $\mathcal{I}$ around its threshold crossing, we define

$$e_{\mathcal{I}} = \max_{j \in \mathcal{I}} \frac{|\widehat{\sigma}_j - \sigma_j^{\mathrm{ref}}|}{\sigma_j^{\mathrm{ref}}}. \tag{3}$$

In this definition the same indices and normalization apply throughout a refinement sequence, excluding reference values below the numerical accuracy floor. Section 5.3 applies this definition to the weighted real map $A_R = [\mathrm{Re}\, A; \mathrm{Im}\, A]$ and checks three reference resolutions; Appendix E specifies the fixed modes and fits. This separates convergence to a fixed weighted operator from changes induced by a different window or taper.

Verification. Before interpreting spectra, we check the adjoint identity and agreement with a per-source implementation using floating-point tolerances; compare against dense SVD on tractable instances as q increases; and verify convergence under grid refinement. The retained spectrum must extend below the rank threshold; otherwise N_{eff} is unresolved. TSQR controls the projected SVD's numerical stability, while increasing q tests whether the Krylov subspace captures the modes of interest. A complete spectral record requires (p, s, q), random seed, arithmetic precision, orthogonalization tolerance, reference resolution, and fitting range. Grid studies hold the physical aperture, frequency band, and source window fixed, so a change in the measured spectrum can be attributed to discretization rather than a change of acquisition.

4 RESEARCH QUESTIONS AND METRICS

We separate physical information, numerical discretization, and learned approximation using five questions and three principal metrics.

RQ1. Spectral structure: Does frequency change the decay law or primarily shift the spectral knee?

RQ2. Physical dimension: How do frequency band, source window, and aperture affect effective rank?

RQ3. Receiver resolution: How many receivers achieve a prescribed spectral accuracy at fixed aperture?

RQ4. Learned approximation: Can a trained linear decoder approach its distribution-specific POD reference?

RQ5. Capacity transfer: Does p/N_{eff} organize errors across bands and input ensembles?

M1: Spectral shape and effective rank (RQ1–RQ2). For the quadrature-weighted operator A, we measure the local slope $\alpha(j) = -\mathrm{d}\log\sigma_j/\mathrm{d}\log j$ and define N_{eff} by the interpolated crossing $\sigma_j/\sigma_1 = 0.1$. Constant local slopes support a power law on the tested interval; N_{eff} locates the knee. The threshold rank is an information proxy, not an exact physical dimension. Comparisons use fixed threshold, interpolation, and weighting conventions.

M2: Spectral discrepancy and receiver cost (RQ3). For a fixed set $\mathcal{I}$ of resolved reference modes, we report

$$e_{\mathrm{spec}}(n_r) = \max_{j\in\mathcal{I}} \frac{|\sigma_j(A_{n_r}) - \sigma_j(A_{\mathrm{ref}})|}{\sigma_j(A_{\mathrm{ref}})}, \qquad n_r^\star(\varepsilon) = \min\{n_r : e_{\mathrm{spec}}(n_r) \le \varepsilon\}. \tag{4}$$

This definition uses consistent weights without separately normalizing each spectrum by σ_1; $n_r^\star$ is the smallest tested passing count. If $e \approx C(\Delta x_r)^\nu$, then $n_r^\star - \delta \approx L(C/\varepsilon)^{1/\nu}$, where $\delta = 0$ for left rectangle and 1 for trapezoid. The new real-map receiver experiment uses this metric; the historical appendix curves retain their original convention. Finite-reference agreement does not certify continuum accuracy.

M3: Prediction error and POD gap (RQ4–RQ5). For centered inputs $x \sim \mu$ in weighted coordinates, define

$$E_\mu(\widehat{A}) = \left(\frac{\mathbb{E}_\mu\|\widehat{A}(x) - Ax\|_2^2}{\mathbb{E}_\mu\|Ax\|_2^2}\right)^{1/2}, \qquad E_{\mathrm{POD}}^2(p;\mu) = \frac{\sum_{j>p}\lambda_j(C_y)}{\mathrm{tr}(C_y)}, \quad C_y = AC_\mu A^{\mathsf{H}}. \tag{5}$$

For nonzero output energy, the covariance tail gives the population optimum over p-dimensional linear output spaces (Lanthaler et al., 2022). Finite-pool POD remains an empirical reference. The GPU measurements aggregate squared errors and output energies over the test set before taking their ratio and square root. They use real output coordinates, the same test samples and quadrature weights, and uncentered POD snapshots for zero-population-mean inputs (Appendix B). This convention differs from averaging samplewise relative errors.

Interpretation. The worst-case width σ_{p+1}, Frobenius tail $(\sum_{j>p}\sigma_j^2)^{1/2}$, and covariance-based POD error answer different questions. A truncated spectral tail cannot certify an upper error tolerance. Illumination is a separate diagnostic: $Az = 0$ implies $A(x+z) = Ax$, so invisible inputs impose no forward-error floor. An absence of shared normalized capacity prevents a collapse comparison; it is not evidence against a scaling law. Projection oracles and trained OOD predictions are evaluated separately.

5 Experiments

We address five questions separating the spectral structure of the Born operator, its physical information dimension, receiver discretization, learned approximation, and the transfer of capacity scaling across input distributions. The last two questions integrate the trained-network study with the spectral experiments.

Setup and numerical controls. The primary spectral study uses $c_0 = 3000\,\mathrm{m/s}$, an $800 \times 480\,\mathrm{m}$ domain at depths 50–530 m, a 10 m grid (3840 model coordinates), and 33 sources and 33 receivers. The bands 5–15, 15–30, and 30–50 Hz use 14, 21, and 27 frequency nodes. We compute 500 singular values of the quadrature-weighted operator A with Krylov depth $q = 2$. On dense verification instances, the maximum relative spectral discrepancy decreases from 0.67–0.83 at $q = 0$ to

2.2×10^{-14} at $q = 2$; adjoint and implementation checks agree to 1.3×10^{-15}. These verify the numerical implementation, not continuum accuracy. These primary spectral results and the smaller CPU learning study are retained from the original experiments; the audited GPU setup appears in Appendix B.

5.1 RQ1: Does frequency change the decay law or shift the spectral knee?

We compare global power-law fits with the local slope $\alpha(j) = -d \log \sigma_j / d \log j$. A power law requires approximately constant α on its claimed range. Instead, Figure 1 shows a plateau followed by a steep tail: local slopes vary by factors 63.6, 129.1, and 107.4 across the three bands. In the middle band, α increases from 0.16 at index 10 to 10.32 at index 300. The decreasing global exponents in Table 1 therefore do not identify a frequency-dependent power law; a moving knee can produce the same fitted trend. These finite spectra do not establish an asymptotic decay class. No common-window statistical model-selection result is established here.

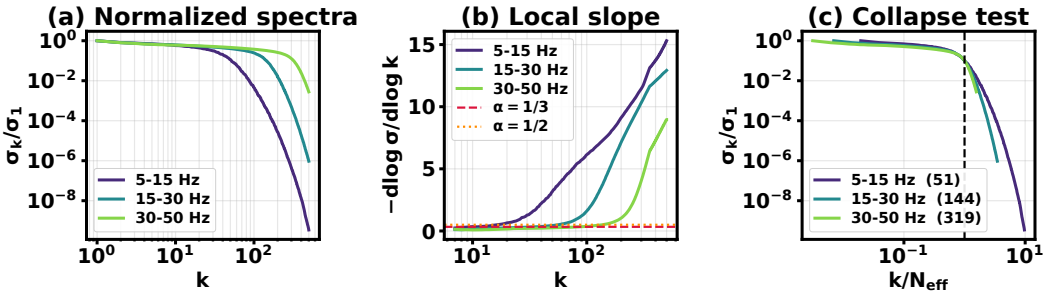

Figure 1: Spectral structure across frequency bands. (a) Normalized spectra. (b) Local slopes are strongly nonconstant. (c) Rescaling by the measured effective rank approximately aligns the knees; it does not establish identical tails. Legend ranks are rounded.

5.2 RQ2: How do band, window, and aperture control effective rank?

We compare N_{eff} with the phase-space-inspired reference $N_0 = |\Omega|(k_{\max}^2 - k_{\min}^2)/(4\pi)$, where $k_{\min,\max} = 4\pi f_{\min,\max}/c_0$. This idealized annular count is not the exact accessible-set volume of a surface array. The rank rises from 51.5 to 319.5, while the reported log–log scaling exponent is 0.876 and N_{eff}/N_0 decreases from 0.480 to 0.372 (Table 1). Thus the reference captures overall growth but leaves systematic band dependence; the three-band calibration is not an out-of-sample scaling test.

Table 1: Reported spectral diagnostics. The last column is a band-specific ratio, not an independently measured angular coverage fraction.

Band (Hz)	Global exponent	Slope ratio	N_{eff}	N_0	N_{eff}/N_0
5–15	7.043	63.6	51.5	107.2	0.480
15–30	4.493	129.1	143.8	361.9	0.397
30–50	1.388	107.4	319.5	857.9	0.372

At fixed 15–30 Hz band, Hann and Gaussian source windows reduce rank from 143.8 to 101.6 and 95.5, respectively (29.3% and 33.6%). The steep tail remains, although its slope changes. Separately, aperture extension at fine receiver spacing increases rank with diminishing gains over the largest tested apertures (Figure 7c in the appendix). Both controls change the physical acquisition operator. Effective rank therefore depends on spectral amplitudes and geometry as well as nominal frequency support; it is a threshold-based information proxy. The observed trend does not establish a universal aperture-to-depth threshold.

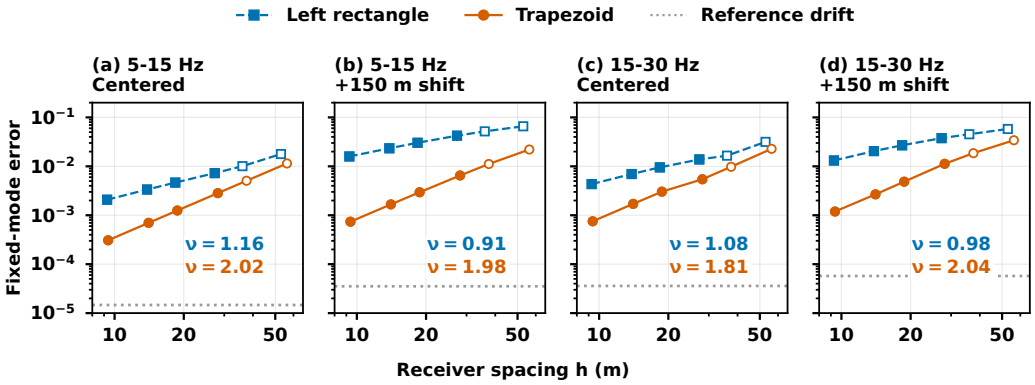

Figure 2: **Receiver convergence at fixed aperture.** Fixed-mode spectral error for two bands and centered or shifted receiver arrays. Filled markers enter the predeclared fits $e_{\mathcal{I}} \propto h^{\nu}$; open markers show additional measurements. Lines connect measured points. Dotted lines show the discrepancy between the two finest references (385 and 769 receivers), not continuum error bounds.

5.3 RQ3: What receiver count achieves prescribed spectral accuracy?

We test receiver refinement on the 40×24 model grid used in the GPU study, with 25 sources and a fixed 900 m receiver aperture. The bands are 5–15 and 15–30 Hz; receiver centers are 0 and 150 m, with geometry fixed within each sequence. Equation (3) compares the weighted real operator $A_R = [\mathrm{Re}\, A; \mathrm{Im}\, A]$ against a 769-receiver trapezoidal reference using fixed modes around its rank threshold. Across $n_r \in \{17, 25, 33, 49, 65, 97\}$, fits on the predeclared subset $\{33, 49, 65, 97\}$ give spacing orders $\nu = 0.907$–1.161 for left rectangle and 1.813–2.045 for trapezoidal quadrature (Figure 2). Thus the measured regime is approximately first versus second order, with variation across band and geometry.

All eight fitted curves pass the reference check: differences decrease under $193 \to 385 \to 769$ refinement, and fitted errors exceed the finest reference difference by at least a factor of ten. At $n_r = 17$, the fast Gram calculation agrees with direct TSQR/SVD on the diagnostic modes to 5.2×10^{-15}. For centered arrays, the smallest tested counts meeting 1% error are $33/49$ (low/high band) for left rectangle and $25/25$ for trapezoid. Shifted arrays require $33/49$ for trapezoid; left rectangle does not reach 1% by 97 receivers (Appendix E). With $h = L/(n_r - \delta)$ and observed $e \approx Ch^{\nu}$,

$$n_r^{\star}(\varepsilon) - \delta \approx L(C/\varepsilon)^{1/\nu}, \qquad \delta = 0 \text{ (left)}, \quad \delta = 1 \text{ (trapezoid)}. \tag{6}$$

Receiver requirements therefore depend on quadrature, geometry, and tolerance. These finite-reference fits do not certify continuum accuracy, exclude aliasing, or reproduce the historical experiment's different geometry.

5.4 RQ4: Can a trained linear decoder approach its distribution's optimum?

We first distinguish operator truncation from distribution-specific approximation. At rank $p = 100$, the reported relative Frobenius tails computed from 500 singular values are 0.004, 0.261, and 0.612. These truncated ratios are lower estimates of the full relative tails, not guarantees that a learned model meets an error tolerance. For centered inputs and a p-dimensional linear output space, the population optimum under Equation (5) is

$$E_{\mathrm{POD}}^2(p; \mu) = \frac{\sum_{j>p} \lambda_j(C_y)}{\mathrm{tr}(C_y)}. \tag{7}$$

This covariance-based reconstruction benchmark connects to operator-learning error analysis (Lanthaler et al., 2022). It differs from the worst-case unit-ball width $\sigma_{p+1}(A)$ and from the unweighted singular-value tail.

The small-scale CPU study trains an unstacked DeepONet with a nonlinear branch and a learned linear output basis, 800 training samples, and $p \in \{4, 8, 16, 32, 64\}$. The two bands have effective

ranks 34.6 and 92.8 on this separate, smaller discretization. A POD basis estimated from a separate pool under the training distribution is evaluated on the same test set as the network. The reported absolute relative-L_2 gap is below 0.01 for $p \leq 32$ and 0.03–0.05 at $p = 64$ (Figure 3a). This supports proximity to an empirical POD reference at the tested scale. Finite-pool POD is not an exact population bound, and the larger gap alone does not identify its optimization or sampling cause.

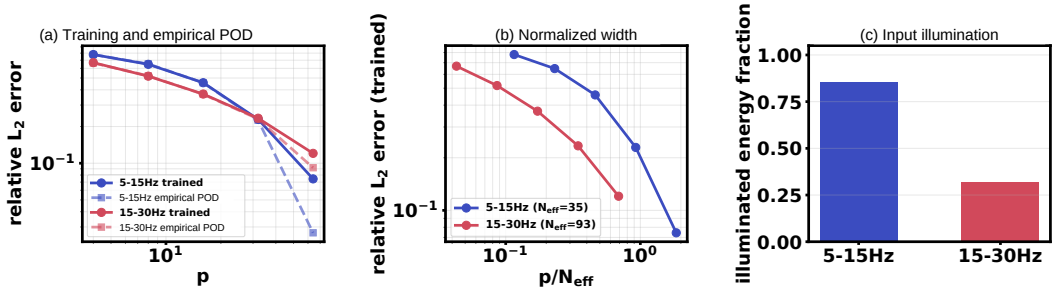

Figure 3: Reported trained-network results. (a) Prediction error approaches the empirical POD reference. (b) Rank normalization does not collapse the trained curves. (c) Illumination fractions differ, but this correlation does not establish the cause of the forward-prediction errors.

5.5 RQ5: Does normalized capacity transfer across input distributions?

The earlier CPU trained curves do not collapse under rank normalization (Figure 3b). This observation concerns that discretization and training protocol. The reported illumination fractions 0.85 and 0.32 characterize input visibility but do not establish the cause of forward error: $Az = 0$ implies $A(x + z) = Ax$. Invisible input energy therefore creates no irreducible error floor for predicting Ax. Distribution-specific linear approximation depends on the full output covariance, and the width sweeps do not estimate sample-complexity scaling.

GPU capacity and OOD transfer (RQ4–RQ5). We train linear- and MLP-branch models with free real p-dimensional output bases, 800 training samples, one seed, and a 1,000-epoch budget on an RTX 5090; validation selects checkpoints at $p \in \{32, 64\}$. Weighted relative L_2 error uses 512 samples per test split. Fixed-correlation and wavelength-matched inputs are compared at 5–15 and 15–30 Hz, with real-operator $N_{\text{eff}} = 98.12, 281.01$ (Appendix B). The linear branch outperforms the MLP in all 18 paired ID/OOD comparisons; at $p = 64$, its ID errors are 0.01127, 0.00917, and 0.02032 for low fixed, high fixed, and high matched inputs. Coefficient prediction accounts for 63.7%–85.2% of MLP ID squared error, while learned-span error accounts for 59.3%–96.1% for the linear branch (Appendix C). The normalized width ranges are disjoint across bands, so the capacity curves cannot test collapse (Figure 6 in the appendix). At $p = 64$, wavelength matching reduces high-band linear-model rough and anisotropic OOD errors by 65.8% and 51.8%, improving both error components; ID error on the matched distribution is higher. These single-seed results establish gains on the two tested OOD ensembles, not a universal capacity law or broad robustness claim.

Frozen-basis coefficient refitting. For the eight high-band checkpoints, we freeze each full learned output basis and replace its branch with a linear coefficient map fitted on the same 800 samples, selecting ridge regularization by ID validation only. Across $p \in \{32, 64\}$ and both ensembles, ID coefficient error falls to $(2.82\text{–}5.17) \times 10^{-8}$, and total error reaches the unchanged span error to reported precision. At $p = 64$, MLP ID error decreases from 0.03360 to 0.01292 for fixed inputs and from 0.11785 to 0.06348 for matched inputs. All 16 OOD comparisons improve by 8.3%–38.5%, removing at least 99.17% of coefficient squared error. Nevertheless, training-POD-plus-ridge predictors built from the same 800 pairs attain ID errors of 9.85×10^{-7} and 0.001884 at $p = 64$ (Figure 4). Thus coefficient error is recoverable in these tests, while the remaining gap reflects the learned output span; its training cause remains unresolved (Appendix F).

Full-wave validity of the Born target. Using eight held-out inputs per condition and split, we compare weighted Born and full-wave scattered fields at contrast multipliers $a \in \{0.25, 0.5, 1, 2\}$,

subdividing the same medium at factors $1, 2, 4, 8$ (Appendix G). All 288 reported cases have decreasing grid changes below 0.5%; 258 also keep the last change below 10% of the Born/full-wave discrepancy. The 30 failures occur at $a \leq 0.5$; all cases at $a = 1, 2$ pass. At nominal contrast, same-grid Born/full-wave errors are 4.47%–4.84% in the low band and 7.08%–10.39% in the high band, versus pooled grid changes of 0.052%–0.219%. At $a = 2$, high-band anisotropic error reaches 20.71%. These checks resolve the tested modeling discrepancies; they neither certify continuum accuracy nor evaluate neural predictions against full-wave targets. Shared high-band OOD inputs yield identical physical results across ensemble labels, not independent replicates.

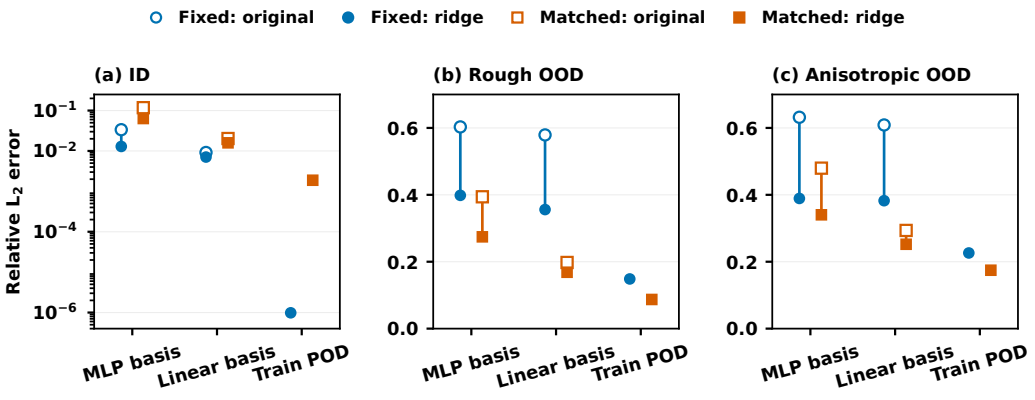

Figure 4: **Coefficient refitting improves prediction but leaves a basis gap.** High-band errors at $p = 64$. Open markers show original networks; filled markers show ridge predictors in the same frozen bases, joined within each checkpoint. Training POD uses the same 800 training outputs and has only a ridge predictor. ID uses a logarithmic error axis and each ensemble's own test distribution; OOD targets are shared. One training seed.

6 Conclusion

Quadrature-weighted Born spectra exhibit a plateau and steep resolved tail, whose moving knee can confound global decay-exponent fits. Effective rank, receiver discretization cost, and learned prediction accuracy describe different aspects of complexity. At fixed aperture, the tested receiver refinements are approximately first order for left rectangle and second order for trapezoid, making numerical requirements rule- and tolerance-dependent. The GPU follow-up shows distribution-specific OOD gains but cannot test cross-band collapse over disjoint normalized widths. Frozen-basis refitting recovers nearly all ID coefficient error, while equal-data training POD remains more accurate, exposing a learned-span gap. Separate full-wave checks resolve 4.47%–10.39% Born discrepancies at nominal contrast, distinguishing approximation of a linear target from its physical accuracy. Together, these results motivate controlled studies of output-basis learning and trained-network transfer to full-wave targets.

Limitations

Our results concern two-dimensional acoustics, a homogeneous reference background, and surface acquisition. Finite spectra and reference discretizations do not establish asymptotic decay or absolute continuum accuracy; receiver fits cover a finite interval with other discretizations fixed, and the Hann taper changes the target operator. Full-wave checks use eight inputs per condition and split in two bands; 30 of 288 reported cases at reduced contrast remain unresolved under the relative grid-change gate. They quantify Born-model discrepancy, not neural full-wave accuracy. Learning evidence remains small-scale and single-seed, with fixed query grids and disjoint normalized-width ranges. Independent POD projections are oracles; training-POD-plus-ridge is deployable. High-band refitting exploits linear Born targets and does not identify the cause of the learned-basis gap. Illumination differences do not establish causality, and Gaussian ensembles omit layered and faulted media. Broader geometry, medium coverage, and direct neural full-wave tests remain for future.

AI USE STATEMENT

We used ChatGPT/Codex (OpenAI) to assist with experiment planning and refinement, Python analysis and plotting code, interpretation of supplied numerical results, manuscript drafting and editing, and figure and table preparation. AI-assisted checks covered numerical-table consistency, recorded solver diagnostics, error aggregation, and LaTeX compilation and formatting. These checks did not independently rerun neural training or the supplied full-wave solves. The authors take responsibility for the final manuscript and accompanying code, including AI-assisted content.

REFERENCES

Peter Benner, Serkan Gugercin, and Karen Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4), 2015.

Gregory Beylkin. Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform. *Journal of Mathematical Physics*, 26(1):99–108, 1985.

Sung Woong Cho and Hwijae Son. Physics-informed deep inverse operator networks for solving PDE inverse problems. In *International Conference on Learning Representations*, 2025.

David Colton and Rainer Kress. *Inverse Acoustic and Electromagnetic Scattering Theory*. Springer, 4th edition, 2019.

James Demmel, Laura Grigori, Mark Hoemmen, and Julien Langou. Communication-optimal parallel and sequential QR and LU factorizations: theory and practice, 2008. URL https://arxiv.org/abs/0806.2159.

A. J. Devaney. A filtered backpropagation algorithm for diffraction tomography. *Ultrasonic Imaging*, 4(4):336–350, 1982.

Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53 (2):217–288, 2011.

Per Christian Hansen. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Society for Industrial and Applied Mathematics, 1998. doi: 10.1137/1.9780898719697.

Lars Hörmander. *The Analysis of Linear Partial Differential Operators III: Pseudo-Differential Operators*. Springer, 1985.

Peng Jin, Xitong Zhang, Yinpeng Chen, Sharon Xiaolei Huang, Zicheng Liu, and Youzuo Lin. Unsupervised learning of full-waveform inversion: Connecting CNN and partial differential equation in a loop. In *International Conference on Learning Representations*, 2022.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL https://www.jmlr.org/papers/v24/21-1524.html.

Nikola B. Kovachki, Samuel Lanthaler, and Hrushikesh Mhaskar. Data complexity estimates for operator learning, 2024. arXiv:2405.15992.

Samuel Lanthaler, Siddhartha Mishra, and George Em Karniadakis. Error estimates for DeepONets: A deep learning framework in infinite dimensions. *Transactions of Mathematics and Its Applications*, 6(1):tnac001, 2022. doi: 10.1093/imatrm/tnac001.

Samuel Lanthaler, Andrew M. Stuart, and Margaret Trautner. Discretization error of Fourier neural operators, 2024. arXiv:2405.02221.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2010.08895.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3:218–229, 2021. doi: 10.1038/s42256-021-00302-5.

Cameron Musco and Christopher Musco. Randomized block Krylov methods for stronger and faster approximate singular value decomposition. In *Advances in Neural Information Processing Systems*, volume 28, 2015. URL https://proceedings.neurips.cc/paper/2015/hash/1efa39bcaec6f3900149160693694536-Abstract.html.

Allan Pinkus. *n-Widths in Approximation Theory*. Springer, 1985.

Lloyd N. Trefethen and J. A. C. Weideman. The exponentially convergent trapezoidal rule. *SIAM Review*, 56(3):385–458, 2014. doi: 10.1137/130932132.

Jean Virieux and Stéphane Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26, 2009.

Ru-Shan Wu and M. Nafi Toksöz. Diffraction tomography and multisource holography applied to seismic imaging. *Geophysics*, 52(1):11–25, 1987.

A NOTATION

We present all the notation used throughout the paper in Table 2.

Table 2: Notation used in the main text and appendices. Unless stated otherwise, norms and singular values use quadrature-weighted coordinates. Reused symbols are qualified by context.

Symbol	Meaning
Scattering geometry and weighted operator	
Ω, x, x_j	Model domain, spatial point, and center of model cell j.
i, j, s, r (acquisition)	Frequency-node, model-cell, source-node, and receiver-node indices in the discrete Born operator.
x_s, x_r	Source and receiver positions; s, r also index their nodes.
c_0, δc, c	Homogeneous background speed, velocity perturbation, and full speed $c = c_0 + a\delta c$ in the full-wave study.
f, f_c, ω	Temporal frequency (Hz), band-center frequency, and angular frequency $\omega = 2\pi f$.
$S(\omega)$	Source spectrum, including the Helmholtz source normalization in the Born formula.
$G(x, y; \omega)$, G_0	Two-dimensional homogeneous Green function; G_0 denotes the same background kernel in the full-wave integral equation.
$H_0^{(1)}$	Order-zero Hankel function of the first kind in the background Green function.
u_0, δu	Background source field $S(\omega)G(x, x_s; \omega)$ and its first-order Born scattered field.
$\beta(\omega)$, β_i	Born velocity-linearization factor $-2\omega^2 S(\omega)/c_0^3$, evaluated at frequency node i. In the appendix Gram identity, β_i additionally absorbs stated quadrature and cell factors.
$B_{isr,j}$	Unweighted discrete Born kernel from model cell j to frequency–source–receiver datum (i, s, r).
m_j, M_m	Model-cell integration weight and diagonal model mass matrix $\mathrm{diag}(m_j)$.
w_i, q_s, q_r	Frequency, source, and receiver quadrature weights, respectively. Here q_s, q_r are weights, not Krylov depth or full-wave contrast.
d_{isr}, M_d	Product weight $w_i q_s q_r$ and diagonal data mass matrix $\mathrm{diag}(d_{isr})$.
$A = M_d^{1/2} B M_m^{1/2}$	Complex quadrature-weighted discrete Born operator.
A^{H}, $A_R^{\top}$	Hermitian adjoint of the complex map and ordinary transpose of the real map.
$x = M_m^{1/2}\delta c$	Real model input in weighted coordinates.
$A_R = [\mathrm{Re}\, A; \mathrm{Im}\, A]$, $y_R = A_R x$	Real-stacked weighted Born operator and output used in GPU learning and receiver-refinement studies.
K_i, D_i, Y_i (operator actions)	Frequency-i product-Green matrix, diagonal Born/data-weight factor, and incoming data block in the streaming adjoint action.
$G_{s,i}$, $G_{r,i}$, $\odot$	Source/receiver Green matrices at frequency i and elementwise matrix product in the appendix Gram identity.
Information dimension and spectral computation	
$\sigma_j(A)$, $\sigma_j(A_R)$	jth singular value of the complex or real-stacked weighted operator; the two spectra have different rank conventions.
$\xi(x)$, $\mathcal{S}(x)$	Local scattering covector and the set of covectors accessible at x over the acquisition.
$\widehat{v}$	Unit direction $v/\|v\|$ in the geometric covector expression.
k, $k_{\min}$, $k_{\max}$	Model-wavenumber radius $k = 2\omega/c_0 = 4\pi f/c_0$ and its band endpoints in the idealized full-angle estimate.

Table 2 continued

Symbol	Meaning
N_{phase}, N_0	Phase-space counting heuristic and ideal annular reference $\lvert\Omega\rvert(k_{\max}^2 - k_{\min}^2)/(4\pi)$.
N_{eff}	Interpolated index where σ_j/σ_1 crosses 0.1 for the specified operator.
γ	Fitted coefficient in the descriptive comparison $N_{\mathrm{eff}} \approx \gamma N_0$; not a measured angular coverage fraction.
$\alpha(j)$	Local log–log spectral slope $-\mathrm{d}\log\sigma_j/\mathrm{d}\log j$; distinct from ridge parameter α below.
p, s, q (Krylov)	Requested spectral rank, Gaussian oversampling, and block-Krylov depth. Elsewhere p is the learned basis width.
Z, Q_ℓ, Q (Krylov)	Gaussian starting sketch, new orthonormal block at depth ℓ, and accumulated model-side Krylov basis.
V (Krylov)	$A^{\mathsf{H}}AQ_{\ell-1}$ before reorthogonalization.
R (TSQR)	Triangular factor whose leading singular values estimate the spectrum of AQ; distinct from a receiver index.
$\mathcal{I}$, $e_{\mathcal{I}}$	Fixed reference-mode index set and maximum relative singular-value error over that set.
n_r, L, h	Receiver count, physical aperture length, and receiver-node spacing.
ν, C (quadrature)	Fitted receiver-convergence order and coefficient in $e_{\mathcal{I}} \approx Ch^{\nu}$.
ε, $n_r^\star(\varepsilon)$, δ	Spectral error tolerance, smallest *tested* receiver count attaining it, and offset in $h = L/(n_r - \delta)$ (0 left rectangle; 1 trapezoid).
$e_{\mathrm{spec}}(n_r)$	Fixed-mode spectral discrepancy relative to the finest receiver reference.
R^2	Coefficient of determination for the receiver-convergence log–log fit.
Learning, projections, and coefficient refitting	
μ, C_μ	Input distribution and its covariance in weighted model coordinates.
$C_y = AC_\mu A^{\mathsf{H}}$, $C_{y_R} = A_R C_\mu A_R^\top$	Complex or real-stacked output covariance under the specified input ensemble.
$\lambda_j(C_y)$	jth eigenvalue of output covariance, ordered nonincreasingly.
p (learning)	Width of the learned real output basis, or of a comparison projection; it need not equal N_{eff}.
$E_\mu(\widehat{A})$	Root mean-square relative forward-prediction error, pooled by squared error and target energy under μ.
$E_{\mathrm{POD}}(p;\mu)$	Population optimum over p-dimensional linear output spaces, given by the relative output-covariance tail; empirical POD uses a finite sample pool.
T, b_θ, $\widehat{y}_R$	Learned real output-basis matrix, linear or MLP branch coefficients, and prediction $Tb_\theta(x)$. The GPU trunk is a free matrix, not a coordinate network.
n_y	Real output dimension (number of rows of T).
Q, $P = QQ^\top$ (learning)	Orthonormal basis of the learned output span and orthogonal projector onto that span; distinct from Krylov Q.
y_i, $\widehat{y}_i$, D (learning)	True and predicted test outputs for sample i, and pooled target energy $D = \sum_i \lVert y_i\rVert_2^2$.
E_{tot}, E_{span}, E_{coef}	Pooled relative total, outside-learned-span, and within-span coefficient errors; their *squares* add for predictions in the learned span.
I, $\lVert\cdot\rVert_2$, $\lVert\cdot\rVert_F$, tr	Identity, Euclidean (or spectral) norm, Frobenius norm, and matrix trace; context distinguishes vector and matrix 2-norms.

Table 2 continued

Symbol	Meaning
X, Y, N, s_x	Row-wise training input and output matrices, number of training pairs, and training-only RMS input scale in coefficient refitting.
W, $\widehat{W}_\alpha$, α (ridge)	Input-to-basis coefficient matrix, fitted ridge solution, and validation-selected ridge penalty; distinct from spectral slope $\alpha(j)$.
$\widehat{Y}$	Predictions from the refitted coefficient map and frozen basis Q.
ℓ, ℓ_x, ℓ_z, k_x, k_z	Gaussian-field correlation lengths and Fourier wavenumbers in the horizontal/depth directions; these k_x, k_z differ from the scattering radius k.
Full-wave validation	
a, r (refinement)	Velocity-contrast multiplier and grid-subdivision factor $r \in \{1, 2, 4, 8\}$; here r is not a receiver index.
u, u_{inc}	Total and homogeneous incident wavefields in the Lippmann–Schwinger equation.
q, q_{lin} (scattering)	Full-wave contrast $\omega^2[(c_0 + a\delta c)^{-2} - c_0^{-2}]$ and its velocity-linearized Born version $-2\omega^2 a\delta c/c_0^3$; distinct from Krylov depth and quadrature weights.
$y^F_{r,i}$, $y^B_{r,i}$	Weighted full-wave and velocity-Born scattered outputs for sample i on refinement r. Here i indexes a sample, rather than a frequency node.
$y^{B,\mathrm{cache}}_{1,i}$	Original cached Born target on the training grid for sample i.
D (full wave)	Pooled energy $\sum_i \|y^F_{8,i}\|_2^2$ within the reported group; a different denominator from learning D.
E_B, E_{grid}, E_{paper}	Pooled errors for fine-grid Born versus full wave, refinement-4 versus refinement-8 full wave, and cached coarse Born versus fine full wave, all normalized by full-wave energy.

B GPU SETUP AND RANK CONVENTION

The GPU experiments use the weighted real map

$$x = M_m^{1/2}\delta c \in \mathbb{R}^{960}, \qquad A_R = \begin{bmatrix} \mathrm{Re}\, A \\ \mathrm{Im}\, A \end{bmatrix}, \qquad y_R = A_R x, \qquad C_{y_R} = A_R C_\mu A_R^\top. \tag{8}$$

Stacking real and imaginary parts preserves output energy but does not preserve the complex singular spectrum or the number of real basis directions. Both p and the GPU effective ranks in Table 3 refer to A_R. The primary spectral study uses the complex weighted operator A on a finer grid. Its ranks, the earlier CPU ranks, and the GPU ranks are not interchangeable across these settings.

Frequency quadrature is trapezoidal in Hz, source and receiver quadrature are trapezoidal, and model weights are midpoint cell masses; the source window is rectangular. The ensembles have zero population mean and use a velocity RMS setting of $30\,\mathrm{m/s}$. Gaussian fields are generated by Fourier filtering on a doubled periodic grid and cropping; deterministic ensemble scaling preserves Gaussianity, without per-sample normalization. The spectral power is proportional to $\exp[-(\ell_x^2 k_x^2 + \ell_z^2 k_z^2)/2]$. Wavelength matching sets $\ell = 250(10/f_c)\,\mathrm{m}$, where f_c is the band center in Hz. At 5–15 Hz this is the fixed ensemble, so its results are reused and do not form an independent replicate. The rough OOD ensemble uses $\ell = 60\,\mathrm{m}$; the anisotropic OOD ensemble uses $(\ell_x, \ell_z) = (250, 40)\,\mathrm{m}$.

The model is $\widehat{y}_R(x) = Tb_\theta(x)$ with a freely learned real matrix $T \in \mathbb{R}^{n_y \times p}$ and no separate output bias. The linear branch is a bias-free map $\mathbb{R}^{960} \to \mathbb{R}^p$; the MLP has three hidden layers of width 256 with GELU activations and a final p-dimensional layer. Scalar input and output normalizations are estimated from training data only. Training uses AdamW, learning rate 10^{-3}, weight decay 0, batch size 64, and validation every 5 epochs. The follow-up extends the original 150-epoch protocol to a 1,000-epoch budget without early stopping, retaining the lowest-validation-error checkpoint. Table 4 reports selected epochs. These models evaluate a fixed acquisition grid; the trunk is not a

Table 3: GPU setup. Both ensembles share each band's operator.

Quantity	Low band	High band
Frequency interval (Hz)	5–15	15–30
Frequency nodes	14	21
Background speed (m/s)	3000	3000
Model width $\times$ height (m)	800×480	800×480
Depth interval (m)	50–530	50–530
Grid cells	40×24	40×24
Cell spacing (m)	20×20	20×20
Sources / receivers	25 / 25	25 / 25
Source / receiver aperture (m)	900 / 900	900 / 900
Real input / output dimensions	960 / 17,500	960 / 26,250
$N_{\mathrm{eff}}(A_R)$	98.1153	281.0074
Fixed ID correlation length (m)	250	250
Matched ID correlation length (m)	250	111.1111
Training / validation samples	800 / 256	800 / 256
Samples per test split / independent POD pool	512 / 2048	512 / 2048

coordinate network. The free matrix can represent any discrete output span of dimension at most p, so the observed gap is not evidence of a coordinate-trunk expressivity restriction. Appendix F tests coefficient recovery at fixed learned spans; the cause of the remaining basis-training gap is not identified.

C LEARNED-SPAN AND COEFFICIENT-ERROR DECOMPOSITION

Let Q be an orthonormal basis for the learned trunk's column span and $P = QQ^\top$. With all sums over a test split and $D = \sum_i \|y_i\|_2^2$, define

$$E_{\text{tot}}^2 = D^{-1} \sum_i \|\widehat{y}_i - y_i\|_2^2, \tag{9}$$

$$E_{\text{span}}^2 = D^{-1} \sum_i \|(I - P)y_i\|_2^2, \qquad E_{\text{coef}}^2 = D^{-1} \sum_i \|\widehat{y}_i - Py_i\|_2^2. \tag{10}$$

When predictions lie in the learned span, orthogonality gives $E_{\text{tot}}^2 = E_{\text{span}}^2 + E_{\text{coef}}^2$. The coefficient share is $E_{\text{coef}}^2/E_{\text{tot}}^2$. Here "coefficient error" is the output-space discrepancy within the learned span, not the unweighted error in a possibly nonorthogonal branch coordinate system. The maximum normalized Pythagorean defect over the 36 recorded rows is 5.22×10^{-11}, and the maximum relative prediction leakage outside the span is 1.47×10^{-7}. These checks verify consistency of recorded metrics; they do not independently recompute checkpoint inference.

Table 4: ID error decomposition for validation-selected follow-up models. All errors are relative L_2 values; the last column is a squared-error share. "Low" and "high" denote 5–15 and 15–30 Hz.

Condition	Branch	p	Epoch	Total	Span	Coeff.	Coeff. (%)
Low fixed	MLP	32	625	0.05579	0.03262	0.04527	65.8
Low fixed	Linear	32	965	0.01872	0.01835	0.00368	3.9
Low fixed	MLP	64	425	0.04636	0.02791	0.03701	63.7
Low fixed	Linear	64	945	0.01127	0.00966	0.00580	26.5
High fixed	MLP	32	605	0.04153	0.02216	0.03512	71.5
High fixed	Linear	32	980	0.01600	0.01449	0.00678	18.0
High fixed	MLP	64	775	0.03360	0.01292	0.03102	85.2
High fixed	Linear	64	890	0.00917	0.00706	0.00585	40.7
High matched	MLP	32	350	0.13362	0.07566	0.11013	67.9
High matched	Linear	32	980	0.02854	0.02685	0.00968	11.5
High matched	MLP	64	245	0.11785	0.06348	0.09929	71.0
High matched	Linear	64	955	0.02032	0.01589	0.01266	38.8

Table 5: High-band OOD decomposition at $p = 64$. The target OOD ensemble is held fixed while the training ensemble changes.

Training	Branch	OOD test	Total	Span	Coeff.
Fixed	MLP	Rough	0.60305	0.39674	0.45417
Fixed	MLP	Anisotropic	0.63186	0.38706	0.49943
Fixed	Linear	Rough	0.57908	0.35367	0.45853
Fixed	Linear	Anisotropic	0.60887	0.37998	0.47575
Matched	MLP	Rough	0.39416	0.27442	0.28294
Matched	MLP	Anisotropic	0.47974	0.34027	0.33819
Matched	Linear	Rough	0.19778	0.16850	0.10357
Matched	Linear	Anisotropic	0.29355	0.25219	0.15022

For the linear branch, matching reduces rough-field total error from 0.57908 to 0.19778: span error falls by 52.4% and coefficient error by 77.4%. For anisotropic inputs, total error falls from 0.60887 to 0.29355, with span and coefficient reductions of 33.6% and 68.4%. The corresponding MLP total-error reductions are 34.6% and 24.1%. ID errors are evaluated on each training distribution's own ID set; the larger matched-ID errors are not a paired fixed-target comparison. The decomposition alone does not establish recoverability; Appendix F supplies the frozen-basis intervention. A dominant span share does not establish an architecture-level lower bound.

D PROJECTION REFERENCES AND NORMALIZED CAPACITY

Operator-SVD projection uses the leading real left singular subspace of A_R. Independent POD uses $2{,}048$ separate ID output snapshots for basis estimation; it uses an uncentered second moment because the input population mean is zero. Both methods project true test outputs and are oracles, not predictions from the input alone. An ID-fitted POD basis is also evaluated on OOD outputs; it is not an OOD-optimal basis. The projection calculations use cached targets at widths 16, 32, and 64, whereas the latest trained models only cover 32 and 64.

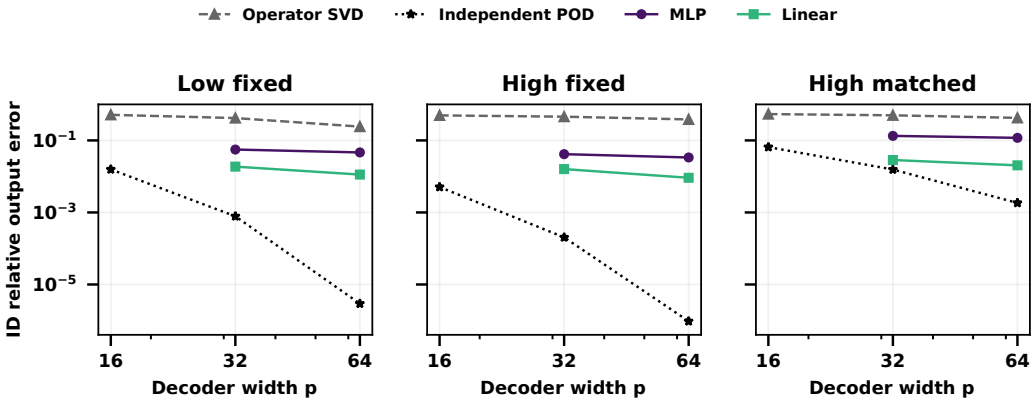

Figure 5: ID projection references and trained models. The logarithmic error axis exposes the small independent-POD errors. Operator-SVD and POD curves use true output coefficients; linear and MLP curves use validation-selected $1{,}000$-epoch-budget models. The low-band fixed condition also represents low-band wavelength matching.

Table 6: Cached ID projection errors. These empirical POD errors are finite-pool diagnostics, not exact population optima.

Condition	p	Operator SVD	Independent POD
Low fixed	16	0.5132807	0.01569483
Low fixed	32	0.4174134	0.0007821796
Low fixed	64	0.2416812	2.935905e-06
High fixed	16	0.4947487	0.005080534
High fixed	32	0.4556826	0.0002033405
High fixed	64	0.3826177	9.529125e-07
High matched	16	0.5381393	0.06521462
High matched	32	0.4979113	0.01556424
High matched	64	0.4222915	0.0018381

Descriptive alignment statistic. For each pair of band curves, we intersect their measured capacity ranges, interpolate error linearly against log capacity at 101 log-spaced points, and average the across-band population standard deviation divided by its mean. There is no extrapolation or statistical significance test. The latest neural ranges are $[0.32615, 0.65229]$ in the low band and $[0.11388, 0.22775]$ in the high band, with an empty intersection. For projection widths $\{16, 32, 64\}$, the normalized intersection is $[0.16307, 0.22775]$, compared with the raw interval $p \in [16, 64]$. Consequently, raw and normalized dispersion summaries compare different parts of the curves. They describe alignment only on the stated intervals.

Table 7: Descriptive cross-band projection dispersion on shared support. Lower values indicate closer alignment within the specified interval; raw and normalized intervals differ.

Training	Basis	Test split	Raw p (%)	p/N_{eff} (%)
Fixed	Operator SVD	ID	7.15	10.10
Fixed	Operator SVD	Rough OOD	7.79	12.97
Fixed	Operator SVD	Anisotropic OOD	9.21	14.82
Fixed	Independent POD	ID	55.20	99.24
Fixed	Independent POD	Rough OOD	33.69	60.93
Fixed	Independent POD	Anisotropic OOD	18.91	39.39
Matched	Operator SVD	ID	11.26	5.33
Matched	Operator SVD	Rough OOD	7.79	12.97
Matched	Operator SVD	Anisotropic OOD	9.21	14.82
Matched	Independent POD	ID	80.39	42.52
Matched	Independent POD	Rough OOD	44.12	69.41
Matched	Independent POD	Anisotropic OOD	23.14	46.39

For wavelength-matched ID outputs, operator-SVD dispersion decreases from 11.26% to 5.33%, while independent-POD dispersion decreases from 80.39% to 42.52%. For fixed ID outputs, normalization increases both dispersions. It also increases dispersion for both projection methods on the two OOD splits. Operator-SVD OOD results are identical under the two training-ensemble labels because the operator basis and OOD targets are unchanged. These comparisons support distribution dependence rather than universal normalized-capacity collapse. They do not independently measure trained-network OOD generalization.

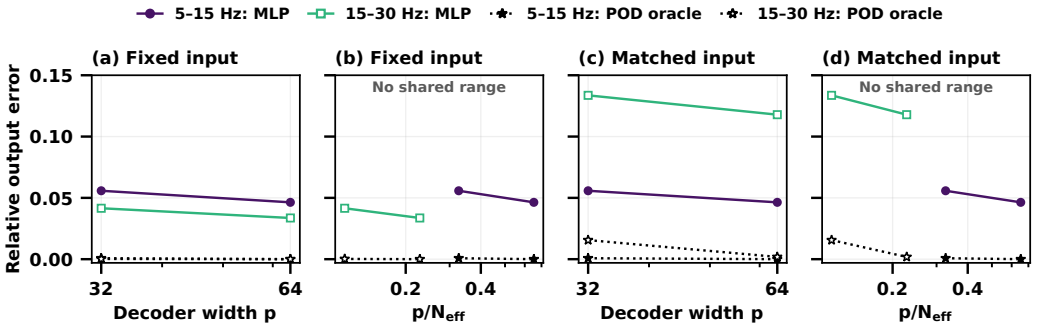

Figure 6: **Distribution-dependent accuracy in the $1{,}000$-epoch follow-up.** Validation-selected MLP errors for fixed (a,b) and wavelength-matched (c,d) inputs against p (a,c) and p/N_{eff} (b,d). Black dotted curves are independent POD oracles using $2{,}048$ basis snapshots and true test-output coefficients. The normalized ranges at $p \in \{32, 64\}$ do not overlap. The identical low-band condition is reused across ensembles. Lines connect measured widths; one training seed.

Scope of the learning evidence. The original audit contains 36 neural diagnostic rows and cached projection measurements. The real-checkpoint refit adds 100 evaluations of the two high-band ensembles (Appendix F); all numerical verification fixtures are excluded from the paper.

E FIXED-APERTURE RECEIVER CONVERGENCE

This completed experiment uses the $800 \times 480\,\mathrm{m}$ domain, 40×24 midpoint model grid, background speed, source geometry, and frequency nodes in Table 3. Sources remain centered and use 25 trapezoidal nodes over $900\,\mathrm{m}$. Receiver apertures also span $900\,\mathrm{m}$ and are centered at either 0 or $150\,\mathrm{m}$. Each layout defines a different physical operator, held fixed within its refinement sequence. The source window is rectangular, frequency weights are trapezoidal in Hz, and no receiver taper is applied.

Metric and fixed fitting protocol. Left rectangle uses n_r nodes on the left-closed, right-open aperture interval with spacing and weights $h = L/n_r$. Trapezoid includes both endpoints, has $h = L/(n_r - 1)$, and half endpoint weights. We retain absolute quadrature-weighted singular values of the real map A_R; individual spectra are not renormalized by their leading value. The finest reference uses 769 trapezoidal receiver nodes. Its 0.1-threshold effective rank fixes the one-based index interval $\mathcal{I} = [\lfloor 0.8 N_{\mathrm{eff}}^{\mathrm{ref}} \rfloor, \lceil 1.2 N_{\mathrm{eff}}^{\mathrm{ref}} \rceil]$ for all errors in that band/layout. We measure $n_r \in \{17, 25, 33, 49, 65, 97\}$ and fit $\log e_{\mathcal{I}} = \log C + \nu \log h$ only at the predeclared $\{33, 49, 65, 97\}$, without constraining ν to an integer.

Table 8: Measured spacing orders on the fixed fitting interval. R^2 refers to the log–log fit, not an uncertainty estimate.

Band (Hz)	Array center	ν_{rect}	R^2	ν_{trap}	R^2
5–15	0 m	1.161	1.000	2.020	1.000
5–15	150 m	0.907	0.999	1.980	1.000
15–30	0 m	1.078	0.997	1.813	0.993
15–30	150 m	0.976	0.996	2.045	1.000

Reference and implementation checks. Trapezoidal references at 193, 385, and 769 receivers halve the interval spacing successively. In every case their spectral differences decrease, and the smallest fitted error exceeds the 385-to-769 difference by at least tenfold. Table 9 lists this difference on the fixed modes. The centered finest-reference ranks are 98.0945 and 280.8814, versus 96.4250 and 275.4648 for the shifted arrays; they differ from the 25-receiver GPU ranks. The reference check assesses separation from the observed refinement drift; it is not a rigorous bound on continuum error.

Table 9: Reference diagnostics and smallest tested counts reaching 1% fixed-mode error. A dash means no tested count through 97 passes. The passing-count decisions agree for both finest references.

Band (Hz)	Center	Fixed modes	Ref. drift	$n^{\star}_{1\%,\mathrm{rect}}$	$n^{\star}_{1\%,\mathrm{trap}}$
5–15	0 m	78–118	1.46×10^{-5}	33	25
5–15	150 m	77–116	3.53×10^{-5}	—	33
15–30	0 m	224–338	3.59×10^{-5}	49	25
15–30	150 m	220–331	5.73×10^{-5}	—	49

For efficient computation, the weighted Green matrices obey

$$A^H A = \sum_i |\beta_i|^2 \big(G_{s,i}^H G_{s,i}\big) \odot \big(G_{r,i}^H G_{r,i}\big), \qquad A_R^{\top} A_R = \mathrm{Re}(A^H A), \tag{11}$$

where β_i includes the Born, frequency, and model-cell factors, and $\odot$ denotes elementwise multiplication. We use this Gram identity only for resolved diagnostic modes, since squaring the condition number can compromise the tiny spectral tail. At $n_r = 17$, direct TSQR/SVD agrees with these modes for both rules in all four cases, with maximum relative discrepancy 5.18×10^{-15}.

Interpretation. The observed orders support approximately first- versus second-order convergence in this finite regime. The high-band centered trapezoidal order is 1.813, so the evidence does not establish exact second order in every case. Shifting the array changes errors and passing

counts; the comparison does not isolate a unique symmetry mechanism. Source, frequency, and model discretizations remain fixed, so the experiment does not certify their convergence, continuum accuracy, or absence of aliasing. It is a new experiment on the GPU-study geometry, rather than a reproduction of the older $600 \times 360\,\mathrm{m}$ receiver study.

E.1 Historical receiver and aperture diagnostics

Figure 7 retains the earlier $600 \times 360\,\mathrm{m}$, $12\,\mathrm{m}$-grid study with 25 sources and 6 Hz-wide bands. Its rectangle/trapezoid fits were reported as $n_r^{-1.23}$ and $n_r^{-2.04}$ under the original discrepancy convention; the new experiment does not assign those curves to Equation (3). The aperture-extension panel supports the physical-coverage discussion in Section 5.2. The Hann receiver taper changes the target weighting and is excluded from comparisons of quadrature rules for a fixed operator.

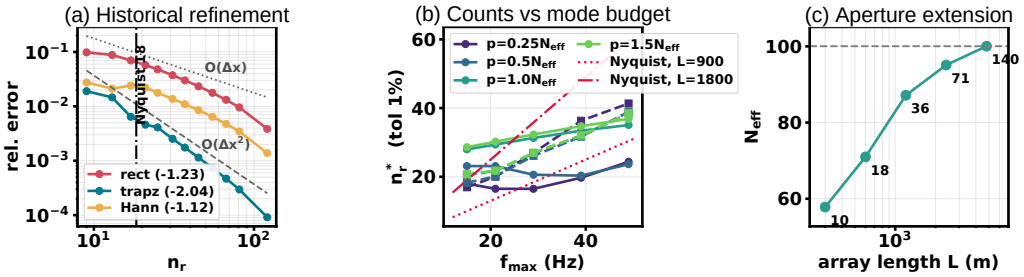

Figure 7: Historical diagnostics, retained with their original error convention. (a) Fixed-aperture receiver refinement; the Hann curve changes receiver weighting. (b) Counts satisfying the original 1% criterion; p indexes diagnostic modes. (c) Aperture extension changes physical coverage and rank; point labels show receiver counts. These curves are separate from the new reference-checked results in Figure 2.

F REAL-CHECKPOINT FROZEN-BASIS COEFFICIENT REFITTING

This experiment re-evaluates eight validation-selected checkpoints at 15–30 Hz: fixed and wavelength-matched ensembles, $p \in \{32, 64\}$, MLP and linear branches, and training seed 0. Each fit uses the original 800 training pairs, 256 ID validation pairs, and 512 samples per ID, rough-OOD, and anisotropic-OOD test set. Checkpoint and training-pool hashes are recorded; the two ensembles use identical OOD input and output arrays. Original-network inference runs on the RTX 5090 with TF32 disabled, while SVD and ridge fitting use CPU double precision. No neural model is retrained in this intervention.

Intervention and selection. We orthonormalize all columns of each learned output basis to obtain $Q \in \mathbb{R}^{n_y \times p}$, preserving its full span. We replace the original branch with a linear input-to-coefficient map; this is not merely a refit of the MLP's final layer. For row-wise training matrices X, Y, training-only RMS scale s_x, and $N = 800$, we solve

$$\widehat{W}_\alpha = \arg\min_W \frac{1}{N} \left\| \frac{X}{s_x} W - YQ \right\|_F^2 + \alpha \|W\|_F^2, \qquad \widehat{Y} = \frac{X}{s_x} \widehat{W}_\alpha Q^\top. \tag{12}$$

There is no intercept. The input SVD and scale use only the training prefix; validation selects α from $\{0, 10^{-14}, 10^{-12}, 10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}\}$. The recorded choices are 10^{-14}, except for the two matched-input MLP bases, which select 10^{-12}. Candidate errors near zero regularization differ only slightly; these choices do not identify a universal optimal regularization strength. Test and OOD data do not participate in fitting or selection.

Equal-data deployable baseline. Training POD uses a direct uncentered SVD of the same 800 training outputs, followed by Equation (12) and the same ID validation selection. It does not use the separate 2,048-output oracle pool from Appendix D, or true test coefficients at prediction time. The basis is deterministic for a fixed training set and is counted once, rather than as independent seed replicates.

Table 10: ID checkpoint/refit pairs. "Basis" identifies the branch used when the frozen basis was learned. Gains are relative reductions in total error. The unchanged span error agrees with the refitted total to the shown precision; nonzero coefficient residuals are reported separately.

Training	p	Basis	Original	Refit	Span	Refit coeff.	Gain (%)
Fixed	32	MLP	0.04153	0.02216	0.02216	2.82e-08	46.6
Fixed	32	Linear	0.01600	0.01449	0.01449	2.94e-08	9.4
Fixed	64	MLP	0.03360	0.01292	0.01292	3.70e-08	61.5
Fixed	64	Linear	0.00917	0.00706	0.00706	3.90e-08	23.0
Matched	32	MLP	0.13362	0.07566	0.07566	3.63e-08	43.4
Matched	32	Linear	0.02854	0.02685	0.02685	4.00e-08	5.9
Matched	64	MLP	0.11785	0.06348	0.06348	4.64e-08	46.1
Matched	64	Linear	0.02032	0.01589	0.01589	5.17e-08	21.8

ID results. All eight ID errors decrease, by 5.9%–61.5%. Refitted coefficient errors are $(2.82\text{–}5.17) \times 10^{-8}$, and the largest relative excess of total error above its unchanged span error is 1.53×10^{-11}. This is agreement to numerical precision on the recorded targets, not a claim of zero population error. At $p = 64$, the linear-basis refits yield 0.0070645 (fixed) and 0.015890 (matched), versus training-POD-plus-ridge errors of 9.851×10^{-7} and 0.0018843. Thus coefficient refitting alone does not close the basis gap, even when both predictors use the same sample budget.

OOD results. All 16 OOD pairs improve, by 8.3%–38.5% in total error, with at least 99.17% of the original coefficient squared error removed. Residual OOD coefficient errors remain 0.0314–0.0433 for fixed-input bases and 0.000166–0.00183 for matched-input bases. Hence reaching the ID span floor does not imply exact OOD coefficient recovery. At $p = 64$, matching still reduces linear-basis refit rough and anisotropic errors by 52.7% and 34.1%, respectively. Training-POD-plus-ridge remains the most accurate tested predictor on both OOD ensembles at both widths.

Table 11: Training-POD-plus-ridge prediction errors using the same 800 training pairs. ID sets differ between ensembles; OOD sets are shared.

Training	p	ID	Rough OOD	Anisotropic OOD
Fixed	32	0.000208208	0.224176	0.297887
Fixed	64	9.85134e-07	0.148502	0.226181
Matched	32	0.0157264	0.17008	0.251848
Matched	64	0.00188431	0.0871621	0.174586

Table 12: OOD checkpoint/refit pairs. The span is unchanged within each pair. All displayed errors are relative L_2 values; "Anis." denotes the anisotropic test ensemble.

Training	p	Basis	OOD	Original	Refit	Span	Refit coeff.
Fixed	32	MLP	Rough	0.61128	0.41495	0.41376	0.03142
Fixed	32	MLP	Anis.	0.63638	0.39887	0.39715	0.03704
Fixed	32	Linear	Rough	0.58934	0.40132	0.39999	0.03271
Fixed	32	Linear	Anis.	0.61041	0.39200	0.39019	0.03760
Fixed	64	MLP	Rough	0.60305	0.39842	0.39674	0.03653
Fixed	64	MLP	Anis.	0.63186	0.38919	0.38706	0.04064
Fixed	64	Linear	Rough	0.57908	0.35593	0.35367	0.04005
Fixed	64	Linear	Anis.	0.60887	0.38243	0.37998	0.04325
Matched	32	MLP	Rough	0.41085	0.29333	0.29333	0.00017
Matched	32	MLP	Anis.	0.48739	0.36042	0.36042	0.00109
Matched	32	Linear	Rough	0.22011	0.20174	0.20174	0.00036
Matched	32	Linear	Anis.	0.31536	0.28929	0.28929	0.00141
Matched	64	MLP	Rough	0.39416	0.27442	0.27442	0.00023
Matched	64	MLP	Anis.	0.47974	0.34027	0.34027	0.00136
Matched	64	Linear	Rough	0.19778	0.16850	0.16850	0.00046
Matched	64	Linear	Anis.	0.29355	0.25220	0.25219	0.00183

Consistency and scope. The archive contains 100 evaluation rows: 80 original/refit rows and 20 training-POD-plus-ridge rows across five splits. The maximum normalized Pythagorean defect is 4.09×10^{-11}, maximum prediction leakage is 1.47×10^{-7}, and paired span differences are zero in the records. Replayed original metrics agree with the previous high-band diagnostics within 3.40×10^{-9} absolute error. These checks establish consistency of the reported numerical diagnostics. The intervention establishes recoverable coefficient error in the tested linear Born setting. It does not isolate why joint training learns a less accurate basis, establish optimizer convergence, or resolve single-seed, two-band, fixed-grid, or trained-network full-wave generalization limitations. Appendix G separately assesses the physical Born targets without evaluating neural full-wave predictions.

G ENSEMBLE-SPECIFIC FULL-WAVE VALIDATION

This experiment checks the physical Born targets, independently of network training. It uses the geometry, quadrature, and input units in Appendix B: 25 sources and 25 receivers over $900\,\mathrm{m}$ apertures, and $14/21$ frequency nodes in the two bands. For low fixed, high fixed, and high matched conditions, we use cached sample IDs 0–7 from each ID, rough OOD, and anisotropic OOD split. Contrast multipliers $a \in \{0.25, 0.5, 1, 2\}$ give 288 reported cases. Each original $20 \times 20\,\mathrm{m}$ velocity pixel is subdivided without interpolation or regeneration. Refinements $r = 1, 2, 4, 8$ therefore compare the same medium, reaching a 320×192 grid with $2.5\,\mathrm{m}$ cells.

Physical equation and solver. With $c = c_0 + a\delta c$ and the $e^{-i\omega t}$ convention, we solve the free-space, constant-density equation

$$(\Delta + \omega^2/c^2)u = -\delta_{x_s}, \qquad u = u_{\mathrm{inc}} + G_0 * (qu), \qquad q = \omega^2\big[(c_0 + a\delta c)^{-2} - c_0^{-2}\big], \tag{13}$$

where $G_0 = iH_0^{(1)}(\omega|x|/c_0)/4$ and $u_{\mathrm{inc}} = G_0(x - x_s)$. The velocity-linear Born field substitutes u_{inc} for u and $q_{\mathrm{lin}} = -2\omega^2 a\delta c/c_0^3$ for q. Midpoint off-diagonal quadrature and an integrated rectangular self-cell define the volume discretization. The recorded runs use CPU complex128 FFT products with free-space padding and diagonally preconditioned GMRES (relative tolerance 10^{-10}, restart 40, at most 100 restart cycles). Self-cell angular quadrature uses 48 points on each of two smooth subintervals in a quadrant; the radial integral is analytic. Only scattered fields are compared, with the paper's frequency, source, and receiver weights applied once. Thus no singular incident field at collocated source and receiver positions enters the output metric.

Metrics and numerical gates. Let $y^F_{r,i}$ and $y^B_{r,i}$ be the weighted full-wave and velocity-Born outputs for sample i at refinement r. On a reporting group, define

$$D = \sum_i \|y^F_{8,i}\|_2^2, \qquad E_B^2 = D^{-1}\sum_i \|y^B_{8,i} - y^F_{8,i}\|_2^2, \tag{14}$$

$$E^2_{\mathrm{grid}} = D^{-1}\sum_i \|y^F_{4,i} - y^F_{8,i}\|_2^2, \qquad E^2_{\mathrm{paper}} = D^{-1}\sum_i \|y^{B,\mathrm{cache}}_{1,i} - y^F_{8,i}\|_2^2. \tag{15}$$

All errors use full-wave scattered-field energy, and pooling precedes the square root. Gates are applied *per sample*: the last grid change must be at most 0.005, at most 0.1 times that sample's Born/full-wave error, and smaller than the preceding change. These are empirical resolution checks, not rigorous continuum bounds. The maximum recorded true solver residual is 9.9996×10^{-11}; the independent coarse Born calculation matches the cached weighted target to 7.25×10^{-7} relative error. Our audit verifies all 288 recorded gates and reaggregates the supplied per-sample squared errors and energies; it is not an independent replay of the scattering solves.

Table 13: Cases passing all full-wave numerical gates. Each condition includes three splits and eight samples per amplitude. All cases pass the absolute 0.5% grid-change gate and have decreasing changes; remaining failures violate only the relative grid-to-model-error gate.

Condition	$a = 0.25$	$a = 0.5$	$a = 1$	$a = 2$	Total
Low fixed	22/24	24/24	24/24	24/24	94/96
High fixed	14/24	22/24	24/24	24/24	84/96
High matched	9/24	23/24	24/24	24/24	80/96
Total	45/72	69/72	72/72	72/72	258/288

G.1 Error magnitudes and interpretation

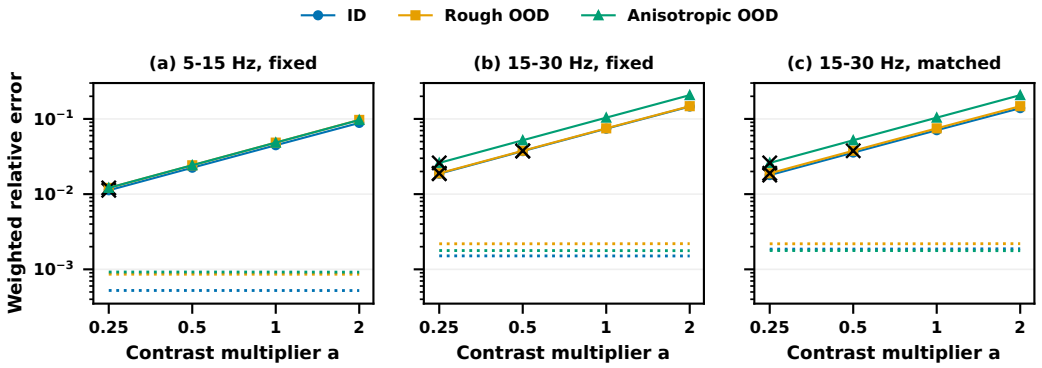

Figure 8: **Contrast dependence of the Born approximation.** Solid curves show E_B on the finest grid; dotted curves show E_{grid} from refinement 4 to 8. Each point pools eight inputs. A cross indicates that at least one constituent sample fails the prespecified numerical gate. Axes are shared across conditions. The two high-band conditions use identical OOD inputs and therefore have identical OOD physical results. No neural predictions appear in this figure.

Across individual cases, the last relative grid change is 0.0463%–0.3289% and decreases in every case. The observed order $\log_2(\|y_4^F - y_2^F\|_2/\|y_8^F - y_4^F\|_2)$ ranges from 2.004 to 2.029, consistent with second-order behavior over these grids. Failures occur only at reduced contrast (Table 13): the Born discrepancy decreases while relative grid changes remain nearly constant, making the relative gate stricter at smaller a. A cross indicates unresolved separation of errors, not GMRES failure or poorer Born accuracy at weak contrast.

Table 14: Nominal-contrast ($a = 1$) errors, in percent, pooled over eight samples. E_B compares Born and full-wave outputs on the same fine grid; E_{paper} instead uses the original cached Born target and also includes spatial-discretization effects. All samples pass both magnitude gates and the decreasing-change check.

Condition	Split	E_B (%)	E_{paper} (%)	E_{grid} (%)	Pass
Low fixed	ID	4.469	4.811	0.052	8/8
Low fixed	Rough	4.841	5.291	0.086	8/8
Low fixed	Anisotropic	4.843	5.304	0.092	8/8
High fixed	ID	7.416	8.572	0.151	8/8
High fixed	Rough	7.494	9.170	0.219	8/8
High fixed	Anisotropic	10.392	10.949	0.177	8/8
High matched	ID	7.076	8.630	0.186	8/8
High matched	Rough	7.494	9.170	0.219	8/8
High matched	Anisotropic	10.392	10.949	0.177	8/8

Passing the gate does not imply a small modeling error: high-band anisotropic error reaches 20.71% at $a = 2$. At $a = 1$, using the original Born targets gives high-band $E_{\text{paper}} = 8.57\%$–$10.95\%$ (Table 14). Parameter linearization, multiple scattering, and spatial-discretization differences are not orthogonal; their squared magnitudes cannot be added as in the learned-span decomposition.

Dependence and scope. High-band OOD hashes and metrics coincide across ensemble labels: 64 repeated reports leave 224 distinct band–medium–amplitude calculations, with 21 distinct failures. Amplitudes and bands are paired, not independent draws; all cases remain in group errors. These small-subset checks do not establish causal wavelength-matching benefits for physical OOD error, universal Born-error bounds, or neural full-wave accuracy.