STRUCTURE-FACTOR COORDINATES FOR FINITE-APERTURE OPERATOR LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning wave responses from aperture geometry requires resolving interference between spatially distributed elements. We introduce a raster-free physical interface that represents this interference through point-set Fourier intensity and multiscale angular moments of the structure factor. In paired acoustic controls, the interface reduces response MAE by 22.7% for Transolver-MS and 21.4% for Point-DeepONet. When the square-triangle family is entirely withheld, the gains persist across three seeds: 10.2–18.2% for Transolver-MS and 18.2–20.5% for Point-DeepONet across two training settings. SURFNO complements the same interface with a fitted kernel transform and learned residual, and performs competitively with spectral baselines. Antenna and scalar diffraction benchmarks reuse the Fourier-intensity construction under different observation measures, while solver-verified inverse design separates representation quality from feasible mask-space search. Equal-spectrum counterexamples delineate the information discarded by spectral compression. Together, the results establish structure-factor coordinates as a reusable physical interface across predictor architectures and an unseen aperture family; finite-window observations and geometry-dependent coupling require complementary geometric or source-state information.

1 INTRODUCTION

Many wave systems implement a cross-dimensional map: a two-dimensional source, element, or transmission pattern produces a one-dimensional response indexed by frequency, wavelength, or observation angle. This map recurs across acoustic transducer arrays, antenna apertures, and diffractive optical elements (Kinsler et al., 2000; Williams, 1999; Balanis, 2016; Goodman, 2017). Every observation combines contributions from spatial sites through phase interference, with additional effects from finite boundaries and coupling. Although raw-mask and point-cloud networks can learn this dependence, the pairwise interference information need not be learned entirely from data. We ask whether an explicit structure-factor representation can make that information reusable across predictor architectures and related wave-observation models.

Neural operators learn maps between functions or fields and have become effective surrogates for partial differential equations (Lu et al., 2021; Li et al., 2021; Kovachki et al., 2023; Azizzadenesheli et al., 2024). Geometry-aware variants extend this idea to irregular meshes and domains (Li et al., 2023a;b; Wu et al., 2024), while recent foundation models seek transfer across equation families (Herde et al., 2024; Holzschuh et al., 2025). For a finite aperture, however, the observable is often generated by a more specific law: spatial correlations encoded by a Fourier intensity or structure factor are sampled by a propagation-dependent kernel. A generic Fourier layer is not equivalent to exposing this law: grid Fourier coefficients depend on rasterization and do not identify which reciprocal coordinates contribute to each queried output.

Our contribution has three parts. **A physical input interface:** we organize classical directional structure factors into raster-free, multiscale coordinates that expose pairwise interference to neural predictors (Eqs. 1–3; Fig. 1). **Architecture-independent evidence:** paired input controls improve Point-DeepONet and Transolver-MS, and the gains persist across three seeds when an entire geometry family is withheld (Fig. 2; Table 4). **Scope and design implications:** known-law tasks, equal-spectrum counterexamples and common-solver inverse tests identify when the representation supports prediction and design and when complementary state is required (Table 5; Fig. 3). Reuse here

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.

concerns the physical construction and its utility across separately trained consumers. SURFNO is one such consumer, combining a fitted kernel transform with learned correction.

2 Related work

Neural and geometric operators. DeepONet, FNO, and subsequent neural-operator formulations learn maps between function spaces (Lu et al., 2021; Li et al., 2021; Kovachki et al., 2023). Geo-FNO and GINO accommodate irregular geometry through coordinate transforms or graph–Fourier latent grids (Li et al., 2023a;b); Transolver learns interactions over complex meshes (Wu et al., 2024). Our focus is complementary: supplying physically specified pairwise interference information to such predictors. The same spectral construction benefits the reported Transolver-MS and Point-DeepONet comparisons, without requiring either to use the fitted propagation core. Thus a stronger consumer can improve the predictor without invalidating the input representation. This differs from claiming a new universally superior attention or Fourier layer.

Physics-informed learning and inverse design. Physics-informed learning can impose equations, invariants, equivariances, or hybrid model components (Raissi et al., 2019; Karniadakis et al., 2021; Willard et al., 2022). Structure-informed operators have also projected predictions onto invariant manifolds (Obieke et al., 2026). Here physical knowledge enters as an input map, with an optional effective core and residual in SURFNO. Structure factors are classical in diffraction and ordered and disordered media (Goodman, 2017; Torquato, 2018); their definition is not our novelty. We evaluate their reuse as a multiscale, directional interface across learned consumers, and distinguish representation benefits from kernel and optimization effects. Their low-dimensional coordinates also motivate inverse design (Molesky et al., 2018; Pipi et al., 2025), but feasibility and response accuracy must be verified on generated masks.

3 Structure-factor coordinates for operator learning

Model names. Transolver-MS denotes our multiscale-input Transolver implementation, and Point-DeepONet denotes our point-set DeepONet implementation with the same multiscale inputs; MS means multiscale, not multiple seeds. SURFNO denotes the principal fitted-core-plus-residual acoustic model; CNN denotes the raster convolutional neural network baseline. Single-scale, zero-input and core ablations are qualified explicitly. These are distinct predictors, not aliases for one network. Appendix B maps the published names to archival identifiers.

3.1 Reciprocal-space representation

For active coordinates $\{\mathbf{r}_i\}_{i=1}^N$ and optional complex or real weights w_i, define the normalized directional structure factor

$$S_m(\mathbf{q}) = \frac{\left|\sum_{i=1}^N w_i \exp(-\mathrm{i}\mathbf{q}\cdot\mathbf{r}_i)\right|^2}{\left(\sum_{i=1}^N |w_i|\right)^2 + \epsilon}. \tag{1}$$

For approximately rotation-averaged observables, we use $S_m(q) = A^{-1}\sum_{a=1}^A S_m(q\hat{\mathbf{e}}_a)$. Directional Fourier moments retain angular information when the observation kernel is anisotropic. Because Eq. (1) is evaluated on the point set, the representation does not depend on raster resolution. A compatible point-set predictor receives these channels as a tensor of L scales $\times\ (1 + 2M)$ channels $\times\ n_q$ reciprocal samples—radial log-spectrum plus M complex angular moments per scale ($L = 7$, $M = 4$ in the acoustic benchmark)—encoded by a small shared convolution network per operator; nothing in the interface depends on a raster.

For prescribed source weights and a plane-wave observation law, the normalized angular intensity is $I_m(\boldsymbol{\theta}, \omega) = B(\boldsymbol{\theta}, \omega) S_m(\mathbf{q}(\boldsymbol{\theta}, \omega))$, where B contains the common element and propagation factors. Angular integration induces a measure $\mathrm{d}\nu_\omega$ on reciprocal space and hence

$$J_m(\omega) = \int S_m(\mathbf{q})\, \mathrm{d}\nu_\omega(\mathbf{q}). \tag{2}$$

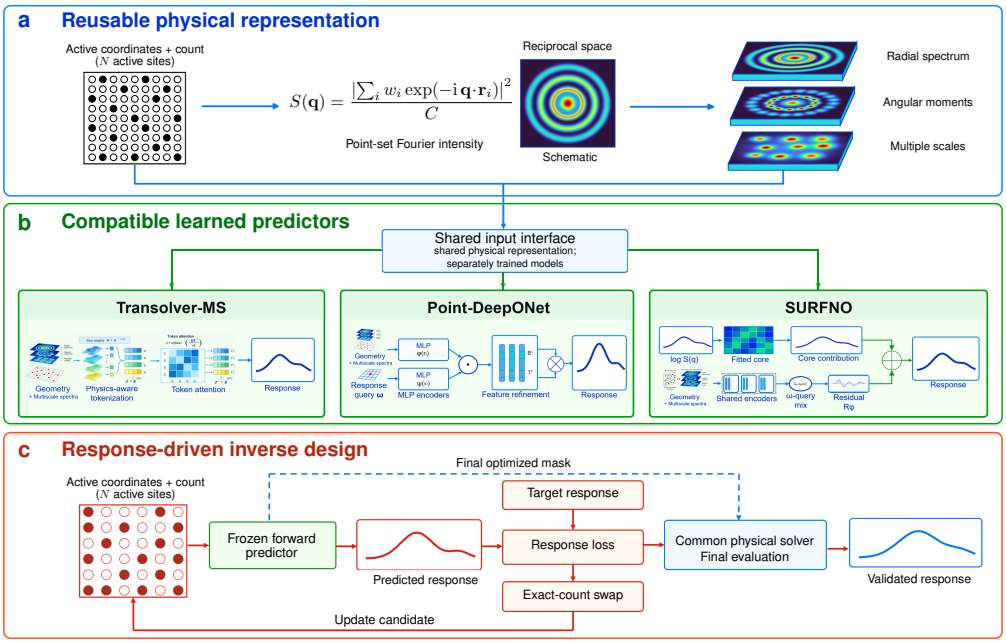

Figure 1: **Reusable structure-factor interface and response-driven inverse design.** (a) Active-site coordinates define a point-set Fourier intensity, from which radial, angular and multiscale features are constructed. C is the normalization C_m in Eq. (3); count, coordinates and local descriptors provide additional inputs. (b) Transolver-MS, Point-DeepONet and SURFNO share this input interface but are trained separately. Among these predictors, only SURFNO adds a fitted effective core to a learned residual; its acoustic core maps log-spectrum inputs to dB responses and is not an exact intensity kernel. (c) In the common-optimizer comparison, a frozen forward predictor scores candidate masks against a target response, and discrete swaps preserve the active count. CNN uses raster and descriptor inputs, whereas SURFNO uses spectral and geometric inputs. Only after optimization is the final mask passed to the common coupled analytical solver for independent response evaluation; the solver is outside the search loop. Masks, spectra, network icons and response curves are schematic.

This is a linear transform of the directional intensity, including the change-of-variable weights in ν_ω. A radial form is exact if the observation measure is rotationally invariant. A finite spatial observation window generally introduces off-diagonal reciprocal correlations; mean pressure magnitude and its logarithm are also different observables. Appendix A gives these distinctions.

What is reusable? Expanding the numerator of Eq. (1) makes the physical content explicit:

$$C_m S_m(\mathbf{q}) = \sum_i |w_i|^2 + 2\,\mathrm{Re}\sum_{i<j} w_i w_j^* e^{-i\mathbf{q}\cdot(\mathbf{r}_i-\mathbf{r}_j)}, \quad C_m = \left(\sum_i |w_i|\right)^2 + \epsilon. \tag{3}$$

The diagonal term is the self contribution and the remaining terms encode pairwise interference. Permuting sites leaves this representation unchanged. A common translation also leaves it unchanged when the weights are held fixed; this symmetry is appropriate for the plane-wave intensity law but can discard position information needed by a fixed observation window. The representation can be evaluated at physical reciprocal coordinates without first choosing an image grid.

Equation (2) separates aperture information in S_m from the measurement measure ν_ω. Under its assumptions, the same directional structure factor can be queried by different observation kernels. This is reuse of a physical representation, not evidence that a learned acoustic kernel transfers unchanged between systems. In practice, radial compression, retained harmonics and the q range determine which information is available. Coordinates, active count and local descriptors remain separate inputs where the compressed spectrum is insufficient.

3.2 KERNEL-TRANSFORM PREDICTOR (SURFNO)

SURFNO combines the spectral input with a fitted propagation-kernel transform and a learned residual. For prediction, let $u_m(q)$ denote the model's spectral input coordinate: linear intensity for the known-law tasks and $\log \max(S_m(q), \epsilon)$ for the acoustic radial core. The fitted structure-factor kernel transform is

$$\widehat{y}_m(\omega_k) = b(\omega_k) + \sum_{j=1}^{n_q} u_m(q_j) K_{\mathrm{eff}}(q_j, \omega_k) + \boldsymbol{\ell}_m^{\mathsf{T}} \mathbf{A}(\omega_k), \tag{4}$$

where $\boldsymbol{\ell}_m$ contains active fraction, vacancy, boundary, and local-pair descriptors. The acoustic output y_m is the dB transform of the mean pressure magnitude in the observation window. Thus K_{eff} is a regression coefficient in log-spectrum coordinates, not the analytical intensity kernel in Eq. (2). Ridge fitting uses the training partition only and gives the compact baseline, the structure-factor kernel transform (SFKT). Neural branches use training-set standardization. A separate continuous-core variant evaluates $\int u_m(q) K_\theta(q, \omega)\, \mathrm{d}q$ by quadrature on a supplied q grid; its discretization experiment is reported independently of the full residual model.

Radial averaging cannot fully encode anisotropy, local clustering, or strong collective coupling. SURFNO therefore predicts

$$\widehat{\mathbf{y}}_m = \mathbf{y}_{SFKT,m} + R_\phi\big(\{\mathbf{u}_{m,s}\}_{s=1}^{L}, \boldsymbol{\ell}_m, \{\mathbf{r}_i\}_{i=1}^{N}, \boldsymbol{\eta}_m\big)\,, \tag{5}$$

where $\mathbf{u}_{m,s}$ contains the standardized radial log spectrum and complex directional moments at scale s, and $\boldsymbol{\eta}_m$ contains system conditions. A response-frequency query selects a softmax mixture of scale embeddings before decoding. R_ϕ learns the combined error of the effective core, including effects of angular compression, boundaries, coupling, nonlinear observation preprocessing, and finite training data. These sources are not separately supervised or identified by the present decomposition. Directional-information controls test one component of the representation loss. Figure 1b places this architecture alongside other consumers of the shared representation; Appendix B gives implementation details.

3.3 SCOPE AND ASSUMPTIONS OF STRUCTURE-FACTOR TRANSFORMS

Table 1 distinguishes the directional angular law, its radial reduction and the separate whole-plane Parseval identity from the coupled acoustic observable.

Table 1: Scope of structure-factor intensity transforms and departures in coupled acoustics. Rotational symmetry is required for the radial reduction, not the directional law; whole-plane observation applies to the separate spatial Parseval identity. Model components are not independently identified physical corrections.

Condition	Departure in coupled acoustics	Model component
Known source weights	coupled amplitudes depend on geometry	residual / descriptors
Rotationally invariant angular measure	angular anisotropy	directional channels
Whole-plane spatial integration	finite 10×10 sample window, off-diagonal terms	residual
Squared-magnitude observable	mean$\lvert p\rvert$ + dB/log nonlinearity	effective core / residual

4 EXPERIMENTS

4.1 EVALUATION PROTOCOL

We ask whether spectral inputs improve different predictors, whether these gains persist on an unseen geometry family, and where the physical representation is insufficient. Known-law tasks distinguish analytical structure from empirical fitting; common-evaluator inverse tests assess usefulness in design. Reciprocal-grid and source-state diagnostics are reported separately in the appendix.

The acoustic benchmark contains 30,000 randomized arrays from square, triangle, and square-triangle templates, split into 21,000/4,500/4,500 train/validation/test samples. Targets are 23-point responses at one-sixth-octave center frequencies from 1 to 12.5 MHz. Spectral MLP, Conv1D, and FNO-1D receive identical seven-scale tensors. Point-DeepONet, Transolver-MS, and SURFNO also receive active coordinates, conditioning variables and local descriptors; only SURFNO receives the fitted core response. CNN uses a 50×50 raster, layout identity and local descriptors on a separate two-layout subset (14,011/2,986/3,003 samples). Its single-seed results are distinguished from optimization-seed averages. Checkpoints are selected on validation data; test data are excluded from fitting. Final inverse errors use independent physical evaluators: a calibrated coupled analytical solver for acoustics (ROI10 mean-$|p|$, $z = 50$ mm) and the exact antenna array factor. Appendix B details seeds, additional splits, preprocessing and budgets.

4.2 Acoustic prediction and input ablations

Paired interventions isolate spectral-input utility within each predictor (Fig. 2a), whereas complete-model comparisons retain differences in capacity and training (Fig. 2c).

At matched capacity (0.542M parameters), replacing the spectral input by its standardized mean increases Transolver-MS MAE from 0.1060 to 0.1371 dB: a 22.7% benefit (paired gain 95% CI 0.0282–0.0341 dB). The seed-0 Point-DeepONet comparison gives 0.1401 versus 0.1101 dB, a 21.4% reduction (0.0300 dB). Together, these paired controls show that the same physical input benefits two distinct consumers (Fig. 2a; Table 2). A capacity-matched information ladder reduces MAE from 0.1407 to 0.1044 dB when directional moments are added; seven scales then leave MAE similar but reduce RMSE from 0.2955 to 0.2503 dB. Thus angular information and scale serve different error regimes (Fig. 2b). The fitted kernel is assessed separately: Appendix Fig. 4b shows no from-scratch gain but a larger continuation effect, which measures dependence on a core already used by the checkpoint (Appendix B).

Count-stratified results expose a conditional advantage over Point-DeepONet (Fig. 2d). In a post hoc seed-0 analysis, restricting evaluation to $N \geq 30$ gives SURFNO/Point-DeepONet MAEs of 0.09559/0.10116 dB (4,447 samples); at $N \geq 100$, they are 0.08734/0.09237 dB (4,329 samples), a 5.4% reduction. The 53 cases with $N < 30$ instead favor Point-DeepONet (0.8608 dB versus 1.2371 dB for SURFNO), largely offsetting that benefit in the full seed-0 test set. Transolver-MS remains more accurate at $N \geq 100$ (0.07324 dB). These application-oriented subsets characterize the count regime behind the aggregate score; a universal minimum array size would require a dedicated multiseed study.

Core fine-tuning and descriptor conditioning provide small continuation gains, whereas an expanded angular basis gives no consistent capacity-matched benefit. Appendix C retains these refinements and negative controls; they do not alter the representation-first conclusion.

Sensitivity to the rasterization map. Re-rasterizing fixed physical geometries changes the frozen CNN's MAE from 0.1019 dB at its training grid to 0.94 dB at 56; a half-cell shift at 50 gives 0.132 dB. Point-set inputs remain unchanged. This experiment isolates sensitivity to the input map; generalization to new physical geometries is evaluated separately (Appendix C).

Table 2: Benefit of spectral inputs across two predictors on the three-layout test set (seed 0 in both arms). Each comparison retains the same architecture and parameter count; the no-spectrum control replaces the standardized spectral tensor by zero, rather than removing its encoder. Transolver-MS uses the formal stage and Point-DeepONet uses staged continuation. Lower MAE is better.

Predictor	Without spectral input	With spectral input	Relative MAE reduction
Transolver-MS	0.1371 dB	0.1060 dB	22.7%
Point-DeepONet	0.1401 dB	0.1101 dB	21.4%

Reduction is relative to the no-spectrum control; values are rounded. Capacity is unchanged within each pair (0.542M and 0.333M parameters, respectively).

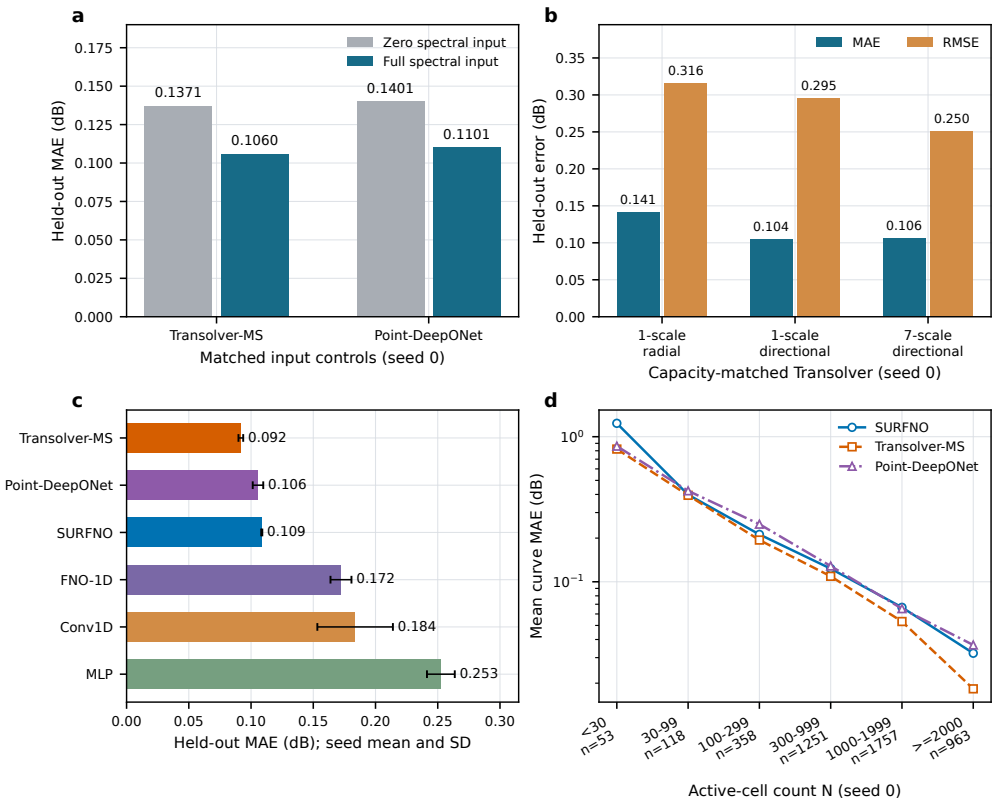

Figure 2: **Reusable spectral inputs and acoustic prediction.** (a) Seed-0 full/zero spectral-input controls at unchanged capacity within each predictor. Transolver-MS uses its formal stage; Point-DeepONet uses staged continuation. (b) MAE and RMSE for capacity-matched Transolver inputs with radial, directional and seven-scale directional information (seed 0). (c) Complete-model MAE in ascending order, with optimization-seed sample standard deviations; objectives and training stages differ (Appendix Table 7). (d) Seed-0 count-stratified errors of the same three point-set predictors; n denotes test geometries per bin, pooled over layouts. All panels use the three-layout split. CNN uses a separate two-layout subset (Table 3). Core and frequency diagnostics are retained in Appendix Fig. 4.

Native-raster comparison. At the actual 50×50 format, CNN is competitive (Table 3); the two-layout Transolver-MS retrain leads. Square has 2,500 sites, whereas square-triangle has 2,660 sites in the same 6 mm square aperture. Mapping the latter onto 2,500 pixels necessarily loses site identity and is one plausible contributor to the performance gap. Full per-layout and count results are retained in Appendix Fig. 6.

The continuous core obtains 1.129 dB MAE on its 32-coordinate reference grid and 1.170 and 1.171 dB at 64 and 128 coordinates without retraining (Appendix Fig. 4a). This experiment establishes reciprocal-grid quadrature transfer of the core; full-network resolution transfer and unseen physical frequencies require separate validation of the residual and queried response.

4.3 Transfer to an unseen template

We withhold the entire square-triangle family and train on square alone (7,000 samples) or square and triangle (14,000). Both settings use the same 9,400 unseen cases with $1 \leq N \leq 2500$, three optimization seeds and 120 epochs. Shared layout encoding avoids an untrained test-category embedding; normalization, ridge fitting and checkpoint selection use seen templates only. SURFNO is trained from scratch rather than resumed from its main-benchmark checkpoint. Appendix F gives the frozen splits, audits and all-count diagnostics.

Table 3: Two-layout forward evaluation (3,003 test samples); all errors are MAE in dB. CNN is trained on the native raster; Transolver-MS and Point-DeepONet are subset retrains (seed 0). Other rows retain archived three-layout training (four seeds for SURFNO, three for MLP/Conv1D/FNO-1D); $\pm$ denotes sample standard deviation. Rows should be interpreted within the indicated training protocols.

Model	MAE (dB)	square	square-triangle
Transolver-MS (subset retrain)	$\mathbf{0.0860}$	0.0816	0.0904
Point-DeepONet (subset retrain)	0.1009	0.0969	0.1049
CNN (native format)	0.1019	0.0944	0.1094
SURFNO	0.1070 ± 0.0005	0.1017	0.1124
FNO-1D	0.1646 ± 0.0076	0.1553	0.1738
Conv1D	0.1801 ± 0.0304	0.1694	0.1908
MLP	0.2381 ± 0.0098	0.2259	0.2503

Table 4: Transfer to unseen square-triangle masks (n = 9,400). MAE in dB, mean $\pm$ sample standard deviation over three seeds. Zero/full denote spectral input. All predictors use shared layout encoding and the same split within each column; capacity and elapsed time are not matched across architectures.

Predictor	Train: square	Train: square + triangle
Point-DeepONet, zero	0.7026 ± 0.0054	0.6954 ± 0.0023
Point-DeepONet, full	0.5585 ± 0.0018	0.5690 ± 0.0015
Transolver-MS, zero	0.7020 ± 0.0013	0.6543 ± 0.0196
Transolver-MS, full	0.5746 ± 0.0003	0.5873 ± 0.0049
SURFNO	0.5669 ± 0.0031	0.5718 ± 0.0025

Spectral inputs lower Point-DeepONet MAE by 20.5% and 18.2% in the two settings, and Transolver-MS MAE by 18.2% and 10.2% (Table 4). Every seed improves both absolute and offset-aligned shape MAE. SURFNO gives errors comparable to the spectral-input baselines. Its learned correction lowers core-only errors of 1.3961 and 1.4474 dB to 0.5669 and 0.5718 dB; this comparison measures correction relative to the fitted core. Representation utility therefore persists on the unseen family, although its errors remain above the roughly 0.09–0.11 dB seen-template range. Because adding triangle changes both training size and diversity, between-column differences combine these effects.

Table 5: Known-law forward MAE. Each task uses its own normalized intensity observable and separately fitted models; values are not dB errors.

Task	Kernel transform	Kernel + residual	Best generic control
Antenna	1.55×10^{-5}	1.30×10^{-5}	Conv1D: 5.36×10^{-5}
Scalar diffraction	8.63×10^{-6}	5.54×10^{-6}	FNO-1D: 3.19×10^{-5}

4.4 Known-law antenna and diffraction benchmarks

Antenna and scalar-diffraction tasks instantiate Eq. (2) under different wave-vector and angular-window parametrizations. They share a Fourier-intensity mechanism and use separately fitted models. The kernel-transform consumers have lower error than the generic controls on these matched observables (Table 5); example predictions and kernel plots appear in Appendix Fig. 7. Restricting antenna observation in azimuth increases radial-kernel MAE to 8.85×10^{-4}; directional inputs reduce it to 5.09×10^{-4}, consistent with the lost rotational symmetry.

The inverse-design comparison in Fig. 3a evaluates 72 held-out antenna targets after exact array-factor reevaluation. Spectrum-mediated paths achieve median MAEs of 3.44×10^{-5} and 3.41×10^{-5}, compared with 2.39×10^{-4} for the strongest direct-logit path. Panel (b) reports each method's own-forward score as a diagnostic, whereas panel (c) compares exact target and generated power. These results favor spectrum-mediated optimization in this known-law regime, although the paths use different optimizers and computational budgets. This advantage does not persist in the coupled-

acoustic setting below. The workflow, response example and generated masks appear in Appendix Fig. 8.

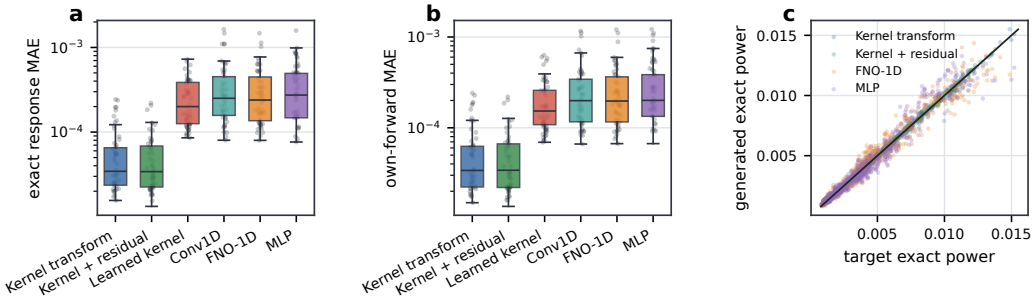

Figure 3: **Antenna inverse design: exact evaluation and surrogate diagnostics.** (a) Final target-response MAE over 72 held-out targets under exact array-factor reevaluation. (b) Each method's own-forward MAE serves as a within-method diagnostic. (c) Exact target versus generated power for selected methods. Performance conclusions rely on (a,c); adapted paths use different optimizers and budgets.

4.5 Information limits of structure factors

We construct 32 centered, noncongruent point-set pairs with identical ordered displacement multisets, hence identical directional structure factors before radial compression. At $z = 10\lambda_0$, fixed-window responses differ by a median 0.152 dB despite matched count, centroid and second moment. The construction demonstrates that spectrum alone can be insufficient; SURFNO therefore retains coordinates and local descriptors. Appendix E provides the construction, quadrature caveat and source-state diagnostics. Geometry-dependent source weights under coupling are another component not specified by the unweighted structure factor.

4.6 Solver-verified acoustic inverse design

We compare four inverse-design paths on 30 held-out targets (15 per layout, count-stratified, active count $N \geq 10$; the two $N = 2$ targets are excluded as a degenerate two-cell regime; the physical 50×50 format is the square layout's actual grid). Every final geometry is evaluated independently by the same calibrated coupled analytical solver. For the primary comparison, CNN and SURFNO drive the same *feasibility-preserving optimizer*: a discrete exact-count swap search with 300 steps $\times$ 16 single-swap proposals, response MSE only, and no spectrum penalty. The two swap arms therefore isolate surrogate choice more directly than the adapted paths. We additionally test a two-stage *spectrum-first* route (response $\rightarrow$ spectrum inversion with the linear core $\rightarrow$ spectrum-only synthesis) and a CNN *continuous-relaxation* route (gradient steps on sigmoid mask variables $\rightarrow$ top-N binarization), each with three shared initializations.

Table 6: Two-layout inverse design with a 50×50 CNN raster, solver-verified, $N \geq 10$ (30 targets, 15 per layout). Median final MAE (dB); "improved" = better than the shared random initial mask. CNN and SURFNO swap rows use the same exact-count discrete search and one shared initialization. Adapted paths report the median over three shared initializations per target, then the median across targets; they use different optimizers and computational budgets and are therefore not equal-cost comparisons. The pooled median is not the mean of layout medians.

Arm	square	square-triangle	merged	improved
CNN swap	0.0714	0.0786	0.0750	25/30
SURFNO swap	0.0806	0.0631	0.0771	26/30
Spectrum-first (3 inits)	0.2056	0.1881	0.2031	–
CNN continuous (3 inits)	0.3057	0.2367	0.2730	–

The main separation is between search paths rather than predictors. Under the shared swap search, CNN and SURFNO obtain similar median errors of 0.0750 and 0.0771 dB, a 16–14 target-wise split, and improvements on 25/30 and 26/30 targets. Spectrum-first and CNN continuous paths give larger errors (0.2031 and 0.2730 dB; Table 6). Realizing an intermediate spectrum with a discrete mask, evaluating fractional masks outside CNN's training distribution, and top-N binarization are plausible failure modes whose individual contributions remain unresolved. Small forward gains from refined SURFNO cores likewise produce no reproducible inverse gain (Appendix Fig. 6e). Feasibility-preserving search therefore matters more than the tested predictor refinements in this experiment.

A reference-free harmonic-proximity penalty gives no improvement at weights $\{0.01, 0.1, 1\}$. Across 320 masks from the archived 32-target cohort, the post hoc per-target median correlation between solver MAE and distance to the reference spectrum is $+0.718$. This is a diagnostic association: the reference spectrum is unavailable in response-and-count-only design, and spectral proximity does not ensure geometry recovery.

Reference-conditioning boundary. Earlier spectrum-inversion runs fixed descriptors at the reference mask, so they are not response-and-count-only design and are excluded from Table 6. Current swap arms recompute descriptors from every candidate. Reference spectra are used only for post hoc evaluation diagnostics, never as optimization inputs.

5 Discussion and limitations

Explicit structure factors supply pairwise interference information that would otherwise need to be learned from geometry. Their benefit across predictors and one unseen family supports separating the physical input construction from the choice of decoder. The strongest consumer need not be SURFNO.

The known-law tasks show the closest match between representation and observable: kernel-transform consumers improve forward prediction, and spectrum-mediated antenna optimization gives lower errors under the tested, unequal budgets. Coupled acoustics instead favors feasible mask-space search. Thus a useful forward representation need not be the best inverse parameterization. Finite-window counterexamples, coupling and nonlinear preprocessing explain why spectral inputs must be complemented by geometry or source-state information. The fitted acoustic core is not an analytical intensity kernel, and its residual does not separately identify these mechanisms.

Transfer remains limited to one unseen template with a substantial accuracy gap; cross-physics experiments fit separate models rather than transfer weights. The inverse comparisons likewise do not establish universal optimizer superiority. Future work could extend the interface with source-state and boundary information to account for geometry-dependent coupling and finite-window effects.

6 Conclusion

Explicit structure-factor inputs improve two different neural predictors, with benefits that persist on one unseen aperture family. Known-law experiments, acoustic counterexamples, and common-evaluator inverse tests jointly show both the utility and the limits of the interface: spectral coordinates capture reusable interference structure, but measurement and coupling effects can require additional state, and a strong forward representation need not be the best inverse parameterization. Our results therefore support a modular design principle: use structure-factor features as a reusable physics-informed interface, choose the learned predictor independently, and introduce regime-specific corrections only where the measurement or coupling physics requires them.

Reproducibility statement

Appendices A–F document the physical formulation, datasets, fixed splits, preprocessing, model inputs, optimization, metrics, inverse reevaluation, computational costs, counterexamples, and transfer controls. The anonymized material provides model components, toy-data generators, numerical

source tables, verification scripts, and one square-only SURFNO transfer checkpoint; it supports aggregate verification and selected replay, but not full end-to-end reproduction.

AI USE STATEMENT

Generative AI tools assisted with experiment planning, deterministic-code implementation and review, literature-search suggestions, manuscript organization, language editing, and preparation of the schematic in Fig. 1. All computations and research outputs were produced and checked by the authors, who take full responsibility for the submission; AI-generated schematic elements are illustrative only.

REFERENCES

Kamyar Azizzadenesheli, Nikola Kovachki, Zongyi Li, Miguel Liu-Schiaffini, Jean Kossaifi, and Anima Anandkumar. Neural operators for accelerating scientific simulations and design. *Nature Reviews Physics*, 6:320–328, 2024.

Constantine A. Balanis. *Antenna Theory: Analysis and Design*. Wiley, 4th edition, 2016.

Joseph W. Goodman. *Introduction to Fourier Optics*. W. H. Freeman, 4th edition, 2017.

Maximilian Herde, Bogdan Raonić, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel de Bézenac, and Siddhartha Mishra. Poseidon: Efficient foundation models for PDEs. In Amir Globerson, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37. Curran Associates, Inc., 2024.

Benjamin Holzschuh, Qiang Liu, Georg Kohl, and Nils Thuerey. PDE-transformer: Efficient and versatile transformers for physics simulations. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 23562–23602. PMLR, 2025.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3:422–440, 2021.

Lawrence E. Kinsler, Austin R. Frey, Alan B. Coppens, and James V. Sanders. *Fundamentals of Acoustics*. John Wiley & Sons, 4th edition, 2000.

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.

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.

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier Neural Operator with learned deformations for PDEs on general geometries. *Journal of Machine Learning Research*, 24(388):1–26, 2023a.

Zongyi Li, Nikola Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, and Anima Anandkumar. Geometry-informed neural operator for large-scale 3D PDEs. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36. Curran Associates, Inc., 2023b.

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.

Sean Molesky, Zin Lin, Alexander Y. Piggott, Weiliang Jin, Jelena Vucković, and Alejandro W. Rodriguez. Inverse design in nanophotonics. *Nature Photonics*, 12(11):659–670, 2018.

Victory C. Obieke, Christopher Chukwuemeka, and Emmanuel E. Oguadimma. Structure-informed neural operators for long-time prediction of parametric Hamiltonian PDEs. arXiv preprint arXiv:2606.14913, 2026.

Anastasia Pipi, Nivedha Gopinath, Valentin Duruisseaux, Myrl G. Marmarelis, Taylor L. Patti, Brucek Khailany, Prineha Narang, and Anima Anandkumar. Inverse design with Fourier Neural Operators for quantum system control. Machine Learning and the Physical Sciences Workshop at NeurIPS, 2025.

Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Salvatore Torquato. Hyperuniform states of matter. *Physics Reports*, 745:1–95, 2018.

Jared Willard, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*, 55(4):1–37, 2022.

Earl G. Williams. *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press, 1999.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for PDEs on general geometries. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 53681–53705. PMLR, 2024.

A Derivation and scope of the structure-factor law

Consider a finite aperture with source amplitudes $a_i(\omega)$ at positions $\mathbf{r}_i$. In a linear medium, an observed field can be written

$$p_m(\mathbf{R},\omega) = \sum_i m_i a_i(\omega) G(\mathbf{R},\mathbf{r}_i;\omega), \tag{6}$$

where G is a Green function or propagation factor. Its squared magnitude is

$$|p_m(\mathbf{R},\omega)|^2 = \sum_{i,j} m_i m_j a_i a_j^* G(\mathbf{R},\mathbf{r}_i;\omega) G^*(\mathbf{R},\mathbf{r}_j;\omega). \tag{7}$$

The Green-function product cannot generally be discarded. For prescribed weights $a_i = H(\omega) w_i$ and an angular observation with $G(\boldsymbol{\theta},\mathbf{r}_i;\omega) = g(\boldsymbol{\theta},\omega)\exp[-\mathrm{i}\mathbf{q}(\boldsymbol{\theta},\omega)\cdot\mathbf{r}_i]$, it factorizes into a common element/propagation magnitude and a weighted Fourier intensity. After normalization by $C_m = (\sum_i |m_i w_i|)^2$, integrating over the angular observation measure induces Eq. (2). The small numerical floor in Eq. (1) is omitted from this analytical identity. Source weights must be known; geometry-dependent coupled amplitudes cannot be replaced by equal weights without an approximation. Radial averaging preserves the measurement exactly only under the corresponding rotational symmetry, or with sufficient retained directional channels (Balanis, 2016; Goodman, 2017).

Finite spatial windows. To see why angular and finite-plane measurements differ, write a shift-invariant plane field as

$$p_m(\mathbf{x},z) = \int \widehat{a}_m(\mathbf{q}) T_z(\mathbf{q}) e^{\mathrm{i}\mathbf{q}\cdot\mathbf{x}} \frac{\mathrm{d}^2 q}{(2\pi)^2}. \tag{8}$$

For a real spatial window $W(\mathbf{x})$ and $\widehat{W}(\mathbf{k}) = \int W(\mathbf{x}) e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\,\mathrm{d}^2 x$, its integrated squared magnitude is

$$J_W = \iint \widehat{a}_m(\mathbf{q}) \widehat{a}_m^*(\mathbf{q}') T_z(\mathbf{q}) T_z^*(\mathbf{q}') \widehat{W}(\mathbf{q}'-\mathbf{q}) \frac{\mathrm{d}^2 q\,\mathrm{d}^2 q'}{(2\pi)^4}. \tag{9}$$

Under the usual integrability conditions, the whole-plane limit $W = 1$ makes the window transform a delta distribution and recovers a diagonal Parseval expression. A finite window generally retains $\mathbf{q} \neq \mathbf{q}'$ terms. Consequently, knowing only $|\widehat{a}_m(\mathbf{q})|^2$, even before radial averaging, need not determine a fixed-window intensity (Williams, 1999; Goodman, 2017).

Acoustic learning coordinates. The acoustic cache uses a discrete central 10×10 observation window at $z = 50$ mm and stores

$$y_m(\omega) = 20 \log_{10} \left[\frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} |p_m(\mathbf{x}, 50\,\mathrm{mm}, \omega)| + \epsilon_p \right], \qquad u_m(q) = \log \max(S_m(q), \epsilon_s), \quad (10)$$

with pressure expressed in the cache's numerical units. Mean magnitude is not mean squared magnitude, and neither the spatial magnitude average nor the logarithm commutes with the linear intensity transform. Equation (4) therefore fits u_m and local descriptors to y_m; it is not an exact consequence of Eq. (2). The core is fitted on unstandardized cached features using training samples only; its output and the residual-network inputs are then normalized with training statistics. The residual includes errors due to observation preprocessing, compression, model approximation and finite-data learning. Symbols such as Δ_{aniso} or $\Delta_{\mathrm{coupling}}$ name possible physical contributions, not independently inferred quantities.

B Model and evaluation protocols

Published names and archival identifiers. Point-DeepONet corresponds to `Point-MS`, `PointMS`, and `point-ms-deeponet` in saved results; Transolver-MS retains its name. The principal acoustic SURFNO corresponds to internal version P5. Earlier P3 controls remain explicitly labeled as the earlier-backbone control, not pooled with P5. The published refinements "SURFNO + trainable core" and "SURFNO + conditioned core" correspond to P5E and P5B, respectively. The known-law labels "Kernel transform", "Kernel + residual", and "Learned kernel" correspond to SFKT/SURFNO-KT, rSFNO/SURFNO-R, and eSFNO/SURFNO-E. SFKT remains the mathematical abbreviation in the equations. CNN denotes the native 50×50 raster model. These name mappings do not merge training stages, seeds, splits or model weights; archival filenames and checkpoint keys are unchanged.

Inputs and splits. The acoustic benchmark uses 10,000 masks per layout and a fixed 21,000/4,500/4,500 train/validation/test split. CNN uses only the square and square–triangle subset (14,011/2,986/3,003); its single-seed metrics are not pooled with the three-layout seed averages. Each response has 23 frequencies from 1 to 12.5 MHz. Spectra and responses use training-partition standardization. The seven reciprocal-scale factors are $1, 2, 4, 8, 16, 32, 64$, each with radial log intensity and real/imaginary moments of orders 2, 4, 6 and 12. The eleven local descriptors are active density, vacancy density, their product, edge–center imbalance, row, column and quadrant occupancy variation, active–active and vacancy–vacancy nearest-neighbor rates, and isolated active and vacancy fractions.

Architecture and optimization. A shared encoder embeds each scale; a frequency-query softmax mixes these embeddings. The decoder also receives active-coordinate features, layout identity and local descriptors. Only SURFNO receives the fitted core response. Transolver-MS uses four heads, 32 physics slices and 32-dimensional heads with padding masks. CNN uses a four-stage convolutional encoder plus layout identity and descriptors, without spectral inputs. Training uses validation-only selection. The SURFNO lineage fits its adapter, jointly optimizes, and continues with Huber loss ($\delta = 0.005$); active density below 0.05 receives a linearly increasing weight capped at three in the final continuation, with no residual-L_2 penalty. Table 7 states the other objectives.

Matched interventions. Spectrum removal keeps the architecture and parameter count unchanged and replaces spectral entries with their standardized training mean. The continuation core intervention starts from a full-core checkpoint and continues both arms for 25 epochs, with only the core input zeroed in one. It is not a from-scratch comparison. The corrected from-scratch control trains both arms for 120 fixed epochs with early stopping disabled and seeds 0–2, selecting on validation MAE.

Table 7: Optimization protocols for the acoustic models. Huber δ is stated where applicable. The table exposes architecture-dependent batch sizes and model-selection objectives rather than implying a single optimizer protocol. All models use AdamW with cosine learning-rate decay and validation-only checkpoint selection.

Model family	Training batch	Objective	Selection metric	Reported stage
Spectral MLP/Conv1D/FNO-1D	512	MSE	RMSE	single formal stage
CNN (two layouts)	128	Huber (0.01)	MAE	single formal stage
Point-DeepONet	128	Huber (0.01)	MAE	staged continuation
Transolver-MS and matched controls	24	Huber (0.01)	MAE	formal; continued where stated
SURFNO	128	Huber (0.005)	MAE	staged plus final continuation

Its with-core and zero-core test means are 0.0986 ± 0.0004 and 0.0974 ± 0.0008 dB, respectively. The earlier unequal-epoch exploratory comparison is superseded, not evidence for a core benefit.

In the three-layout continuation comparison, zero-core and full-core MAEs are 0.2056 and 0.1086 dB. Both arms start from a full-core checkpoint, so the 0.0970 dB difference measures dependence on the core during continuation, not its benefit under independent from-scratch training. The comparison in Fig. 4c uses the same spectral inputs for SURFNO and Transolver-MS on two layouts (0.1070 versus 0.0860 dB). Their capacities differ (0.513M versus 0.542M), so this is a model-level comparison, not an isolated architectural intervention.

Continuous-core and known-law tests. The separate continuous-core experiment evaluates 128 radial bins with 64 angles, resamples the 32-bin training inputs and checks a native 64-bin cache. A scalar training mean/standard deviation and trapezoidal quadrature preserve the meaning of the supplied q grid; this is not a discretization test of the full residual network. The antenna test uses a 32×32 grid at $0.5\lambda_0$ spacing, 24 frequencies and angular averaging. Scalar diffraction uses 1,024 sites, 64 radial bins and 24 wavelengths from $0.75\lambda_0$ to $1.35\lambda_0$. Each uses 2,100/450/450 train/validation/test samples and an epoch cap of 120. Physical wavevectors are interpolated onto radial bins and averaged over the observation angles to construct kernel proxies. The separate sector-symmetry test uses 2,400/600 train/test samples; its errors are not a matched leaderboard with the full-cone experiment. The lens-phase test uses prescribed phase weights and scalar-diffraction reevaluation, not a full electromagnetic solver.

Inverse evaluation. Current acoustic swaps use 30 count-stratified targets ($N \geq 10$, 15 per layout; two $N = 2$ targets excluded), 300 steps and 16 proposals per step, or 4,801 surrogate evaluations including initialization. Counts remain exact; descriptors and spectra are recomputed from every candidate. No nearest-training-mask projection or cached candidate response is used. The common coupled solver rebuilds mutual impedances and volume velocities at all 23 frequencies and averages $|p|$ over ROI10 at $z = 50$ mm. It agrees with the square-only solver to $< 10^{-3}$ dB; recomputed reference masks have median cache discrepancy 1.90×10^{-6} dB. The spectrum-first and continuous-relaxation controls use three shared initializations; the antenna inverse comparison uses 72 held-out targets and exact array-factor reevaluation. These adapted paths have different optimizers and are not equal-cost comparisons.

The current swap benchmark is reference-free. Earlier spectrum-mediated runs held eleven descriptors at reference-mask values and must not be interpreted as designer-only operation. Reference spectra in the separate 320-mask spatial-fidelity analysis are post hoc oracles, not optimization inputs. Near-zero printed distances do not establish spectral identity or recovery of the target geometry. Raster reconstruction uses float64 to preserve boundary-site assignments; Fourier bases retain complex values. Configurations, seeds, split indices, preprocessing and checkpoint lineages are required alongside the code to reproduce these comparisons.

C Acoustic accuracy, computational cost and ablations

Forward accuracy and cost. Tables 8 and 9 retain the complete three-layout forward metrics and measured resource costs. CNN is reported separately in the main text. Resource measurements use one NVIDIA A100-SXM4-80GB and exclude cache construction; they do not establish end-to-end savings after spectral preprocessing.

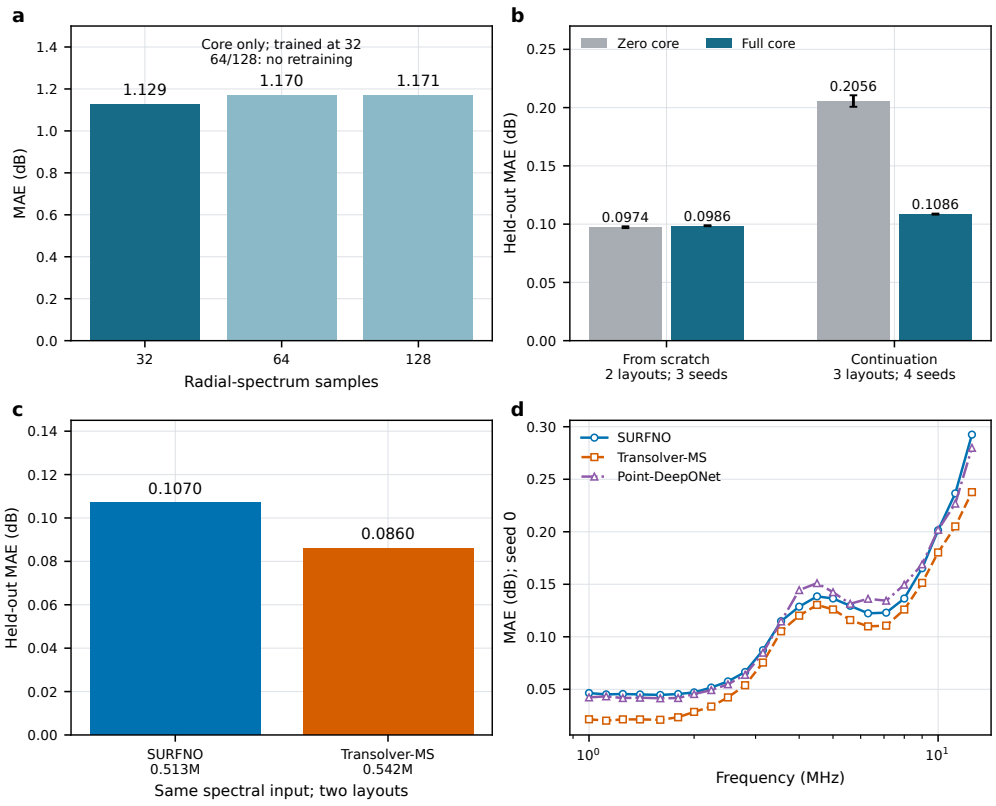

Figure 4: **Core, architecture and frequency diagnostics.** (a) Changing the reciprocal grid without retraining a separate continuous core; this is not full-network resolution transfer. (b) Core interventions under two distinct protocols: fixed 120-epoch training from scratch on two layouts (three seeds), and continuation from full-core checkpoints on three layouts (four seeds). Error bars are sample standard deviations. The from-scratch result does not support a core accuracy gain; continuation measures dependence on an already-used core. (c) Model-level comparison with the same spectral inputs on two layouts; unequal parameter counts preclude a capacity-matched architectural attribution. (d) Frequency-resolved MAE on the three-layout test set for seed-0 continued models, using the same predictions as Fig. 2d.

Table 8: Acoustic forward comparison on the primary test split. Values are mean ± sample standard deviation across optimization seeds. These three-layout results exclude CNN, which uses the two-layout subset reported in main-text Table 3.

Model	MAE (dB)	RMSE (dB)	R^2
MLP	0.2525 ± 0.0112	0.5027 ± 0.0301	0.99832 ± 0.00020
Conv1D	0.1836 ± 0.0304	0.3567 ± 0.0327	0.99915 ± 0.00015
FNO-1D	0.1723 ± 0.0084	0.3451 ± 0.0077	0.99921 ± 0.00004
Point-DeepONet	0.1056 ± 0.0042	0.2509 ± 0.0038	0.999582 ± 0.000013
Transolver-MS	0.0918 ± 0.0019	0.2415 ± 0.0115	0.999612 ± 0.000037
SURFNO(robust continuation)	0.1086 ± 0.0004	0.2722 ± 0.0021	0.999508 ± 0.000008

CNN input consistency. Figure 5 evaluates the same frozen CNN on the same 3,003 physical test geometries with different rasterizations. The model is not retrained. Dashed coordinate-model references have unchanged inputs; this is not evidence of transfer to new physical conditions. Only CNN is included, not a search over the best attainable CNN designs.

Core and angular refinements. The refinements below retain small or negative findings, with the full run history archived. SURFNO + trainable core (E) replaces the ridge core by a trainable linear head; SURFNO + conditioned core (B) continues E with a descriptor-conditioned head. Both

Table 9: Single-GPU resource profile on the acoustic benchmark. Batched inference uses the implementation's evaluation batch size and excludes cache construction. "Stage" denotes the final reported optimization stage; "lineage" sums checkpoint-parent stages and is reported to make staged optimization transparent rather than as a claim of lower training cost.

Model	Parameters	Stage (s)	Lineage (s)	Memory (GiB)	Infer. (ms/sample)
MLP	0.537M	62.7	62.7	0.06	0.035
FNO-1D	0.689M	44.0	44.0	0.14	0.034
Conv1D	0.586M	464.6	464.6	8.11	0.062
Point-DeepONet	0.333M	210.3	1051.2	11.40	0.171
Transolver geometry-only	0.366M	2354.4	2354.4	4.29	0.804
Transolver-MS	0.542M	1126.5	2429.5	11.65	0.735
SURFNO(robust seed 0)	0.511M	396.7	2981.1	9.93	0.161

initialize to the base prediction and use head learning rate 10^{-5} with no weight decay; the final-stage budget is 25 epochs with cosine backbone rate $5 \times 10^{-6} \to 5 \times 10^{-7}$. E improves MAE by 0.00073 ± 0.00011 dB (three layouts) and 0.00089 dB (two layouts), over four seeds; B adds 0.00125 and 0.00098 dB. The three-layout lineage changes from 0.10862 to 0.10664 dB. E core-output drift is 0.033–0.040 dB; B density-conditioned weight variation is approximately 5×10^{-4} relative Frobenius. Neither identifies a unique physical kernel. Under the same inverse swap protocol, E/B medians are 0.0775/0.0752 dB versus 0.0771 for the base, with no consistent advantage.

A expands the retained harmonics to orders 1–16 on 64 angles. The SURFNO paired changes on epoch-matched seeds are -0.0064 and -0.0049 dB; a third arm stopping at 45 epochs is excluded from that paired summary. Earlier-backbone (internal P3) changes are $+0.0048$ and $+0.0010$ dB on matched seeds; the 58/120-epoch pair is excluded. The capacity-matched earlier-backbone control makes the expanded basis $+0.0035$ dB worse. Transolver-MS continuation gives 0.08606 versus 0.08603 dB. Thus richer angular inputs show no consistent benefit under these controls; orders 1–16 are still a truncated basis.

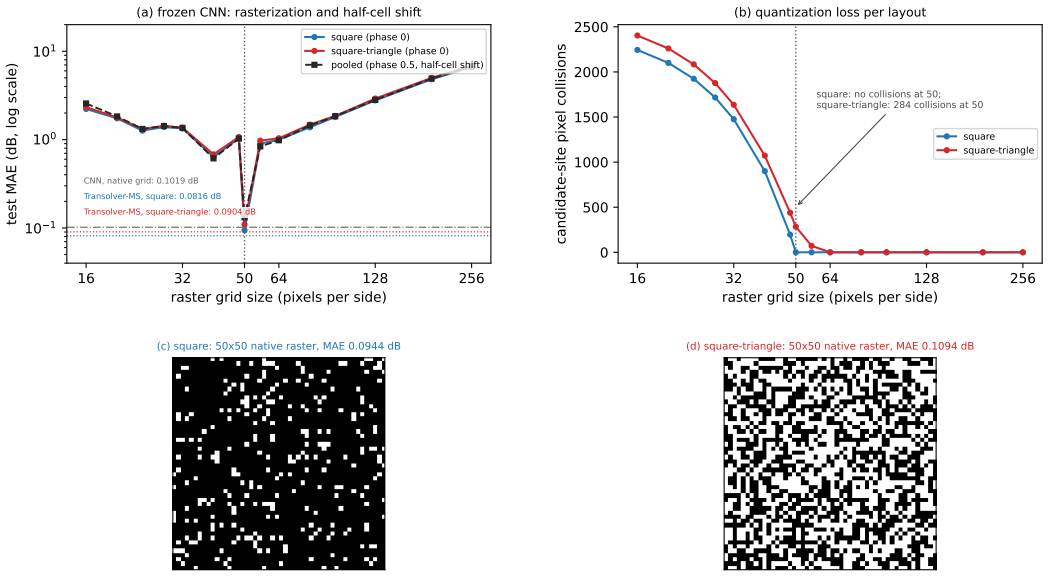

Figure 5: Frozen CNN rasterization probe. (a) Test MAE of the frozen CNN trained at the native 50×50 format, re-rasterized at grids from 16 to 256: per-layout curves at phase 0 plus the pooled half-cell shift (phase 0.5); dotted horizontal lines are the raster-free Transolver-MS references, unchanged across grids. (b) Candidate-site pixel collisions per layout; their causal contribution to prediction error is not isolated. (c,d) Example native 50×50 rasters; the displayed MAEs summarize each layout, not the individual masks. Geometries, descriptors and layout identities are fixed throughout; only the rasterization map changes, so the probe diagnoses input-map sensitivity rather than a universal convolutional limitation.

D KNOWN-LAW MODELS AND ADDITIONAL BENCHMARK RESULTS

The figures in this section provide additional visual evidence for three main-text comparisons. Figure 6 expands the native-grid forward and inverse results summarized in Tables 3 and 6, including per-layout and target-level variation. Figure 7 visualizes the known-law antenna and scalar-diffraction tests, which instantiate different measurement parametrizations of the same Fourier-intensity mechanism rather than independent cross-domain transfer tasks. Figure 8 complements the quantitative antenna inverse comparison in Fig. 3 with the workflow, a response example and generated masks.

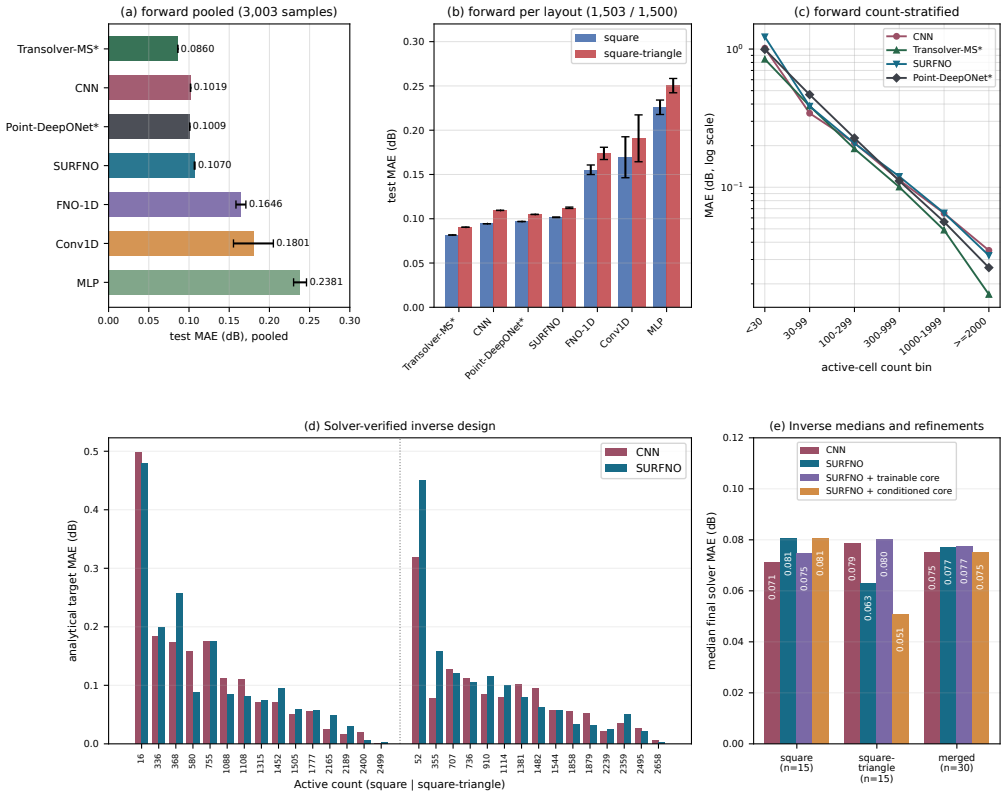

Figure 6: **Native-grid forward and inverse comparisons.** (a–c) Two-layout test subset (3,003 samples): pooled, per-layout and count-stratified MAE. Asterisks mark subset retrains (seed 0); other forward baselines retain archived training. (d) Common-solver errors for CNN and SURFNO swaps, sorted by count within each layout (30 targets, $N \geq 10$). (e) Per-layout and pooled medians including refinements E/B; only the pooled medians lie within 0.003 dB. The reported comparison quantifies performance on this cohort; raster collisions remain one plausible contributor.

Known-law forward models. Both benchmarks use coherent, equal-amplitude active sites in a planar aperture, with no mutual coupling. Let $m_i \in \{0,1\}$, $N = \sum_i m_i > 0$, $\tilde{\mathbf{r}}_i = \mathbf{r}_i/\lambda_0$, and $\mathbf{e}_\phi = (\cos\phi, \sin\phi)$. The normalized far-field amplitude and intensity are

$$\mathcal{A}_m(\tilde{\mathbf{q}}) = \frac{1}{N}\sum_i m_i e^{-i\tilde{\mathbf{q}}\cdot\tilde{\mathbf{r}}_i}, \qquad \mathcal{I}_m(\tilde{\mathbf{q}}) = |\mathcal{A}_m(\tilde{\mathbf{q}})|^2. \tag{11}$$

Here $\tilde{\mathbf{q}} = \lambda_0\mathbf{q}$ is dimensionless, and $\mathcal{I}_m$ is Eq. (1) with unit weights and no numerical floor. These are the array-factor and scalar Fraunhofer models, respectively, not full vector Maxwell simulations (Balanis, 2016; Goodman, 2017).

Antenna observable. With $\nu = f/f_0$ and $\lambda_0 = c/f_0$, the cached linear-power response is

$$P_m(\nu) = \sum_{t=1}^{T} \beta_t \frac{1}{A}\sum_{a=1}^{A} \mathcal{I}_m(2\pi\nu\sin\theta_t\,\mathbf{e}_{\phi_a}), \qquad \beta_t = \frac{\cos\theta_t}{\sum_{u=1}^{T}\cos\theta_u}. \tag{12}$$

The 3,000-mask cache uses a 32×32 square candidate grid with spacing $\lambda_0/2$, 24 uniformly spaced $\nu \in [0.6, 1.8]$, $T = 10$ uniformly spaced polar angles from 15° to 70°, and $A = 32$ azimuths $\phi_a = 2\pi(a-1)/A$. The cosine factor is a prescribed power weight. The sector control instead uses 32 equally spaced azimuths from -5° to 5°, including endpoints. Final antenna inverse-design evaluation uses Eq. (12) directly on each generated mask.

Scalar-diffraction observable. For normalized wavelength $\ell = \lambda/\lambda_0$, the flat-aperture cache uses

$$D_m(\ell) = \frac{1}{TA} \sum_{t=1}^{T} \sum_{a=1}^{A} \mathcal{I}_m\left(\frac{2\pi}{\ell} \sin\theta_t \, \mathbf{e}_{\phi_a}\right). \tag{13}$$

The 3,000 masks span square and square–triangle templates, each with 1,024 candidate sites in a $16\lambda_0 \times 16\lambda_0$ bounding square. There are 24 uniformly spaced $\ell \in [0.75, 1.35]$, ten uniformly spaced polar angles from 4° to 36°, and the same 32 full-circle azimuths. Active transmissions have unit amplitude and zero phase; the reported flat-aperture test does not include a focusing phase profile or material dispersion.

Both responses are dimensionless sampled angular averages, without a $\sin\theta$ solid-angle Jacobian; neither is total radiated power or a uniform-area detector integral. Their different weights and reciprocal-coordinate queries instantiate the measure in Eq. (2). The radial input uses 64 uniformly spaced reciprocal radii and 32 azimuths; the radius range extends from zero to 1.03 times the largest queried transverse wavenumber. The plotted physical-kernel proxy linearly interpolates these radial samples at the queried radii and applies the corresponding angular weights. It is a discretized proxy, distinct from the learned effective kernel. Forward labels are evaluated from the point sums, not from this interpolation.

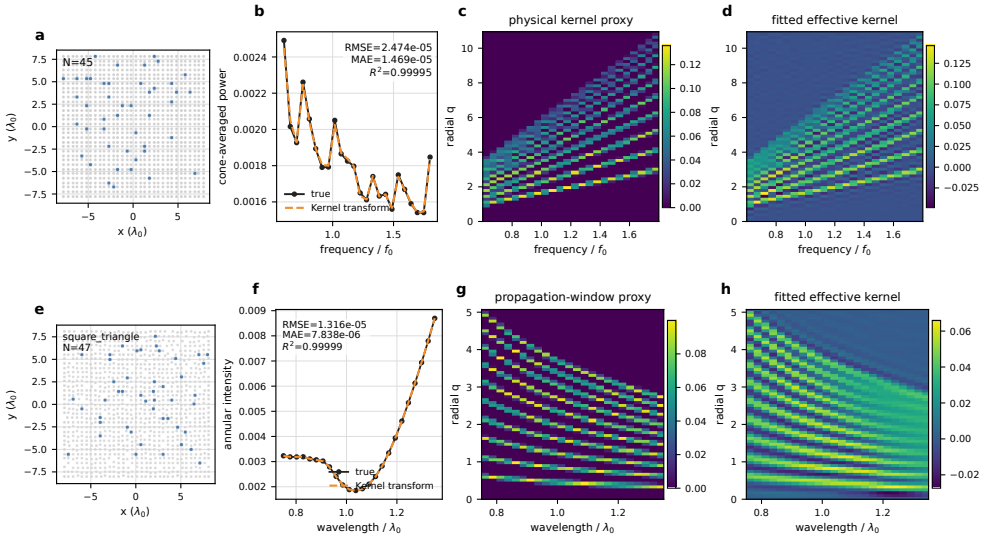

Figure 7: **Known-law verification: finite antenna (a–d) and scalar diffraction (e–h).** (a–d) Example aperture, held-out prediction, physical kernel proxy, and fitted effective kernel. (e–h) Square-triangle aperture, exact annular-intensity response, propagation-window proxy, and fitted effective kernel.

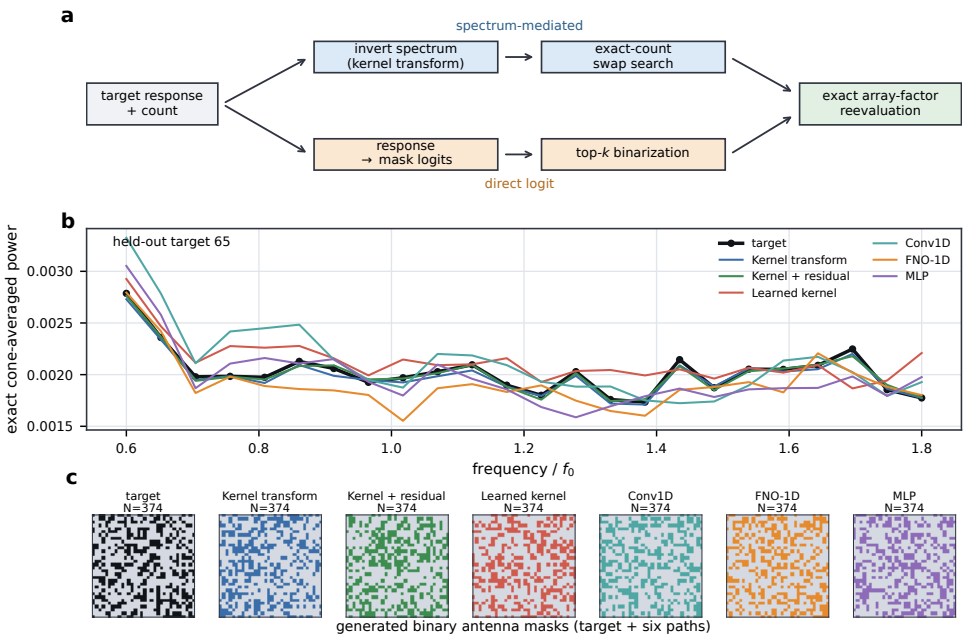

Figure 8: **Antenna inverse-design workflow and example.** (a) Spectrum-mediated and direct-logit paths, followed by exact array-factor reevaluation. (b) Target and generated exact cone-averaged power for held-out target 65. (c) Target and generated binary masks, all with 374 active sites. Aggregate exact-evaluator errors, own-forward diagnostics and power comparisons are shown in Fig. 3.

E Information boundaries and source-state diagnostics

Finite-window counterexamples. We first translate 32 prescribed coherent scalar-source patterns while holding the observation window fixed. Translation leaves their Fourier intensity unchanged but changes their response; translating the window with the sources restores the original response numerically. To exclude centroid as the sole explanation, we construct 32 noncongruent homometric pairs $A = V + W$ and $B = V - W$, rejecting collisions and separately centering each member. Their ordered displacement multisets coincide exactly; their multisets of per-vertex distance signatures differ, excluding rigid congruence. Counts are 12, 20, 30 and 42, eight pairs each. A common rescaling bounds coordinates by $2.75\lambda_0$. Count, centroid and second moment match, and sampled directional spectra agree within 10^{-15}. Local spacing descriptors are not matched.

Use prescribed weights $1/N$, Green function $\exp(\mathrm{i}kr)/(4\pi r)$, 23 frequencies $f/f_0 \in [0.75, 1.35]$, and an 11×11 central observation grid with half-width $1.2\lambda_0$. Median pairwise mean-magnitude response differences are 0.277, 0.152 and 0.021 dB at $z/\lambda_0 = 5, 10, 20$, respectively. On the balanced pair distribution at $z = 10\lambda_0$, half the *mean* pairwise absolute difference gives a 0.1058 dB lower bound for predictors receiving only the matched inputs. This bound does not apply to coordinate-augmented networks. Refining to a 21×21 grid leaves the median difference at 0.160 dB, but changes some individual curves by up to 0.286 dB. These discrete-observable diagnostics compare two numerical grids; continuum quadrature convergence and FEM validation are outside the scope of this test.

Correlation state and physical readout. For deterministic coherent source amplitudes, define $\mathbf{Q} = \mathbf{U}\mathbf{U}^\dagger$. It remains rank one even with strong coupling. With nonnegative observation weights $\mathbf{W}_\xi$, the quadratic readout is

$$J_\xi = \mathrm{tr}\left(\mathbf{G}_\xi^\dagger \mathbf{W}_\xi \mathbf{G}_\xi \mathbf{Q}\right). \tag{14}$$

Geometry is retained in $\mathbf{G}_\xi$; $\mathbf{Q}$ alone is not a geometry code. The mean-magnitude acoustic response instead averages $\sqrt{(\mathbf{G}_\xi \mathbf{Q} \mathbf{G}_\xi^\dagger)_{oo}}$ before taking the logarithm, as in Eq. (10). Direct and quadratic

readouts agree within 1.5×10^{-14} dB in the scalar tests. This checks an algebraic identity, not learned prediction performance. No separate covariance network is trained.

Complex-source feasibility pilot. The production chunks retain complex $N \times 23$ `Uvol` for each mask and $50 \times 50 \times 23$ `Ptot`. MATLAB column-major logical indexing aligns sources with active coordinates. Replaying six square cases using the original directivity and propagation at $z = 50$ mm gives relative complex-field error below 2.8×10^{-8} and maximum response discrepancy 1.2×10^{-5} dB. A matched source-network pilot selects 128 square cases with $30 \leq N \leq 300$, split 90/19/19. Both 24,334-parameter networks predict 23 complex amplitudes per source; one receives 32 log radial bins averaged over 24 directions, the other zeros in the same slots. Both use coordinates, count, identical initialization and a 200-epoch budget, selecting by validation source MSE. Central ROI MAE is 0.1212 versus 0.1215 dB: adding the spectrum does not improve this small-data experiment. Query shifts of ± 0.6 mm reuse the sources but test the supplied decoder, not learned query extrapolation.

Across those 19 cases, mean per-frequency source error is approximately 0.12, the full complex-field error is 0.037–0.038, and the impedance residual is 0.125. The true sources satisfy the reconstructed equation within 3.6×10^{-8}. The Spearman correlations between error and neighbor count within each geometry and frequency are 0.24 and 0.23. Larger correlations with the true coupling force do not constitute independent causal evidence, because the coupled-source equation, $\mathbf{CU} = \mathbf{b} - D\mathbf{U}$, already links that force algebraically to the source state. Here, D is the diagonal self-impedance matrix, $\mathbf{C}$ is the off-diagonal mutual radiation impedance matrix, and $\mathbf{b}$ is the uncoupled excitation vector. A fixed, unlearned four-step damped correction from independent sources gives 0.0163 dB ROI MAE for this inspected cohort. These small-cohort diagnostics do not establish generalization or computational savings from learned source corrections.

F TRANSFER TO AN UNSEEN TEMPLATE

Frozen protocols and fitting. We reserve all 10,000 square–triangle masks for testing. The first protocol trains on square only (7,000 training, 1,500 validation and 1,500 seen-test samples); the second adds the same numbers from triangle. Square indices and unseen-test indices are identical across protocols, with split seed 20260912. Point-DeepONet and Transolver-MS each have full- and zero-spectrum arms, paired in initialization and parameter count. SURFNO uses its frequency-conditioned multiscale-attention architecture. All five predictors are trained from scratch with seeds 0, 1 and 2, giving thirty fits across the two protocols. No all-template checkpoint is reused. Every template uses shared layout-embedding slot zero during training and testing, avoiding an untrained categorical embedding. These controls therefore differ from the categorical-layout main benchmark.

All fits use 120 epochs without early termination, AdamW, learning rate 10^{-3} decaying by cosine to 10^{-5}, weight decay 10^{-5}, Huber loss with $\delta = 0.01$, and validation-MAE checkpoint selection. Batch sizes are 128 for Point-DeepONet and SURFNO, and 24 for Transolver-MS. SURFNO uses ridge coefficient 10^{-3}, no sparse reweighting and no residual penalty; its fitted core remains frozen while the correction is learned. This is an epoch-matched comparison, not the historical staged SURFNO training recipe or a parameter-count/wall-time-matched ranking. Only training indices determine normalization and ridge fitting. The audit verifies all thirty fits, including disjoint splits, prediction ordering, paired baseline initializations, validation-selected epochs and independently refitted SURFNO cores. Template-centered, longest-span coordinate normalization is retained; absolute aperture-size transfer is not tested.

Evaluation. The pre-specified primary scope comprises 9,400 unseen cases with $1 \leq N \leq 2500$. All 10,000 predictions, count-bin metrics and RMSE are also archived. For frequency error $e_{ik} = \widehat{y}_{ik} - y_{ik}$, the offset-aligned diagnostic is

$$\mathrm{MAE}_{\mathrm{shape}} = \frac{1}{nF} \sum_{i=1}^{n} \sum_{k=1}^{F} \left| e_{ik} - \frac{1}{F} \sum_{j=1}^{F} e_{ij} \right|, \qquad F = 23. \tag{15}$$

It removes a constant dB error from each curve, supplementing rather than replacing absolute MAE. Reductions use ratios of seed-mean errors. Dispersion denotes sample standard deviation across three optimization seeds, not confidence intervals. The core-only fit is deterministic for each split.

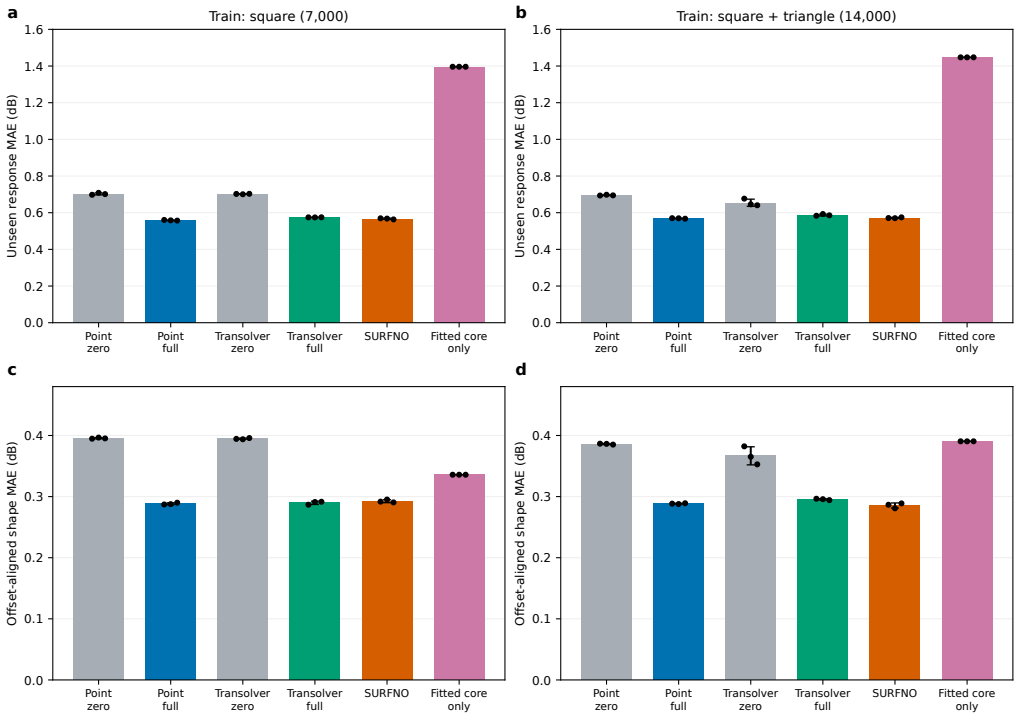

Figure 9: **Same-split transfer to an unseen template.** All panels evaluate the same 9,400 square–triangle masks without retraining. Columns distinguish square-only and square-plus-triangle training; rows report absolute MAE (a,b) and offset-aligned shape MAE (c,d). Point and Transolver abbreviate Point-DeepONet and Transolver-MS; zero/full indicate their spectral inputs. Bars show seed means, dots individual seeds and error bars sample standard deviations. Core-only dots coincide in value because the ridge fit is deterministic. Spectral-input controls improve in both splits; SURFNO improves over its fitted core but does not consistently lead the spectral-input baselines. The core is an effective ridge fit, not an exact analytical propagation kernel.

Main Table 4 reports absolute MAE. Figure 9 additionally shows shape errors and the deterministic core-only baseline; the complete seed-level MAE, RMSE and shape metrics remain in the accompanying results archive.

Results and limits. The shape diagnostic confirms that the input gains are not limited to a constant level correction. With square-only training, shape MAE falls by 27.1% for Point-DeepONet and 26.6% for Transolver-MS; with both templates, the reductions are 25.3% and 19.4%. SURFNO's shape MAEs are 0.2925 ± 0.0024 and 0.2855 ± 0.0041 dB, versus 0.3358 and 0.3905 dB for the deterministic core. Improvements over a core alone measure net learned correction, not the causal contributions of anisotropy, boundaries and coupling. They also do not replace the separate no-core architecture ablation. No architecture leads every metric in both splits.

Seen-template MAE for full-spectrum Point-DeepONet, Transolver-MS and SURFNO is 0.0955/0.0906/0.0988 dB with square-only training and 0.1032/0.1058/0.1065 dB with both templates. The substantial unseen-template gap remains. Adding triangle doubles the training set and changes the seen-test composition, so this is not a sample-size-matched diversity ablation; the results do not show that adding templates must improve transfer. Common count support does not match count distributions or physical extents. Coordinates and local descriptors remain available, and the zero-input controls retain their spectral encoders. This single-family holdout supports reuse of spectral-input utility, not universal geometric sufficiency, cross-physics weight transfer or a CNN transfer comparison.