WAVE PHYSICS AS REPRESENTATION: TENSOR-STRUCTURED NEURAL RECONSTRUCTION FOR OCEAN ACOUSTIC FIELDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Reconstructing three-dimensional (3D) complex ocean acoustic fields from sparse measurements is challenging due to highly oscillatory wave structures and restricted sensing patterns/routes. Existing learning-based approaches generally fall into two categories. First, purely data-driven models can learn complex field patterns from data, but do not explicitly exploit the underlying wave physics that can potentially enhance interpretability and reduce training costs. Second, physics-informed methods impose a specified physical equation as a reconstruction constraint, but usually require the environment-dependent parameters to be known, which is difficult to acquire in practice. As a result, practical 3D acoustic field reconstruction that is both physically grounded and robust to uncertain environmental information remains largely unexplored. This work takes an alternative route: instead of solving a fully specified wave equation, it extracts a structured representation from the mathematical form of wave physics and learns its components from data. Starting from the Helmholtz equation and normal mode propagation, we derive a structured spatial–angular tensor representation, termed the S/c representation. Under a low-rank specialization, it takes the form of a rank-$(L, L, 1)$ block-term tensor decomposition, which connects structured sampling patterns/routes often used in ocean sensing practice to sufficient conditions for field recovery. Furthermore, building on this tensor representation, we propose TIDAL, a neural reconstructor that uses the S/c representation as an inductive bias for structured decoding and aligns information aggregation with the sampling geometry. The neural learner removes the need for known model parameters required by conventional physics-based recovery methods. Experiments across three structured sampling geometries show that TIDAL achieves the best reconstruction performance and fastest inference among baselines, with clear benefits from the physics-derived representation under limited data and model capacity.

1 INTRODUCTION

Accurate reconstruction of three-dimensional (3D) complex ocean acoustic fields from sparse measurements is important for underwater sensing and environmental characterization (Zhu et al., 2024). The reconstruction problem is particularly difficult in realistic ocean environments for two reasons. First, modal propagation and multipath effects produce highly oscillatory interference patterns, making the field fundamentally different from smooth signals (Jensen et al., 2011). Second, ocean observations cannot usually be placed arbitrarily, but are constrained by platform motion and array configurations. This leads to structured sampling patterns such as horizontal and vertical (H/V) transects (Wan et al., 2009; Tang et al., 2007), while many established inverse problem recoverability supports are based on random sampling patterns; see, e.g., Candes & Plan (2010). Recovering the full 3D field therefore requires an effective representation of oscillatory wave structure that is compatible with the geometry of sparse observations.

Existing learning-based approaches mainly follow two directions. Purely data-driven approaches can capture rich field patterns directly from data (Zhao et al., 2024; Santos et al., 2023; Huang et al., 2024). However, their representations are typically generic and do not explicitly exploit the governing wave physics. Physics-informed methods, such as physics-informed neural networks

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.

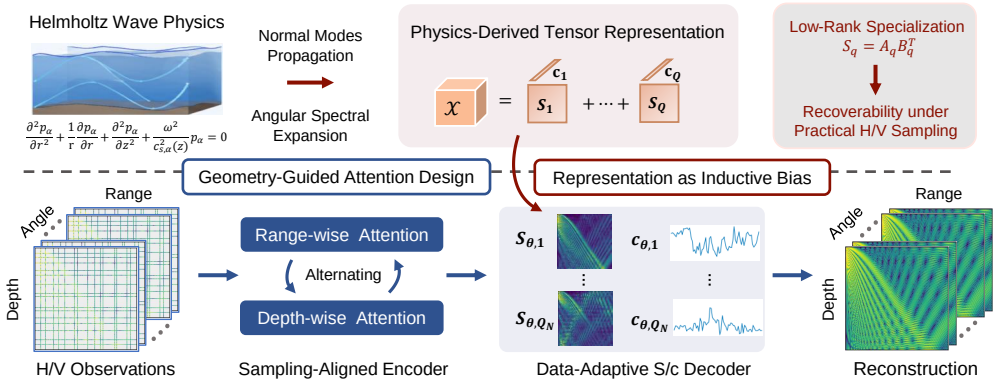


Figure 1: Overview of the proposed framework. Starting from the Helmholtz equation, we derive the physics-derived S/c tensor representation; under a low-rank specialization, it connects H/V observations to LL1 recoverability theory (top). Building on this representation, TIDAL uses the S/c structure as an inductive bias through data-adaptive decoding, while the H/V observation geometry guides information aggregation through sampling-aligned axial attention (bottom).

(PINNs) (Raissi et al., 2019), instead incorporate governing physical equations as constraints. This requires scene-dependent physical parameters to be known, which is often difficult in complex ocean environments. Consequently, practical reconstruction methods that can retain physical structure without relying on accurately specified scene-specific parameters remain largely unexplored.

This motivates a different route: using physics at the representation level. The central observation is that the mathematical structure of the governing wave equation already reveals how the physical field can be represented. Starting from the Helmholtz equation and the normal-mode representation of underwater propagation (Jensen et al., 2011), we organize the directional modal solutions along the periodic angular coordinate and apply an angular spectral expansion. This yields a structured representation of the field in terms of depth–range spatial components S modulated by angular factors c, termed the S/c representation. Under a low-rank specialization, the S/c representation connects the physics-derived field structure to rank-$(L, L, 1)$ block-term tensor decomposition (LL1) recoverability theory, yielding sufficient conditions for unique recovery from structured H/V observations.

Building on these insights, we propose TIDAL (Tensor-Inspired Decoding with Axial Learning), a neural reconstructor guided by both the physics-derived S/c representation and the structured H/V observation geometry. TIDAL retains the S/c representation as an inductive bias through structured decoding with data-adaptive components, while sampling-aligned axial attention aggregates information along depth and range following the H/V sensing geometry. In this way, wave physics determines how the field is represented, tensor structure explains when it is recoverable, and attention-based learning reconstructs the missing field from data. Fig. 1 summarizes the framework.

Experiments across equispaced, shifted, and jittered H/V sampling geometries show that TIDAL achieves the best reconstruction performance with efficient inference. Representation analysis shows that the learned spatial components exhibit pronounced singular value decay, while prescribed moderate ranks can limit representation capacity, motivating data-adaptive spatial modeling. Paired comparisons with direct-output decoding further support the S/c representation as an effective inductive bias, with particularly clear benefits under limited training data and model capacity.

Our main contributions are summarized as follows:

- We show that the mathematical form of wave physics directly reveals a structured S/c representation of 3D complex acoustic fields, without requiring exact physical parameters. Under a low-rank specialization, this representation takes the form of LL1, connecting structured H/V observations to classical tensor recoverability theory and yielding sufficient conditions for unique full-field recovery.
- Building on this representation, we propose TIDAL, which incorporates the physics-derived S/c structure as an inductive bias through structured decoding with data-adaptive

components. The H/V observation geometry further guides information aggregation through sampling-aligned axial attention along depth and range.

- Across three H/V sampling geometries, TIDAL achieves the best reconstruction performance and fastest inference among the baselines, while analysis further highlights the benefits of the physics-derived S/c representation under limited data and model capacity.

2 Related Work

Tensor and low-rank reconstruction. Low-rank tensor models, including Canonical polyadic decomposition (CPD) (Harshman et al., 1970), Tucker decomposition (Tucker, 1966), and block-term decomposition (BTD) (De Lathauwer, 2008), have been widely studied for representing and recovering high-dimensional data from incomplete observations. Coupled CP and LL1/BTD models have been studied for various applications; see, e.g., recoverability under structured degradation in hyperspectral super-resolution (Sørensen et al., 2015; Kanatsoulis et al., 2018; Ding et al., 2021). Most closely related, real-valued coupled LL1 recovery from structured slab and fiber observations has been studied for spectrum cartography (Zhang et al., 2020). In contrast, our S/c representation is derived from underwater wave physics, and its low-rank specialization connects complex acoustic field reconstruction to existing LL1 recoverability results under structured H/V observations.

Learning-based field reconstruction. Purely data-driven approaches have explored neural operators, attention-based models, and generative methods for reconstructing physical fields from incomplete observations (Zhao et al., 2024; Tiwari et al., 2025; Santos et al., 2023; Huang et al., 2024; Ma et al., 2025). Physics-informed methods instead incorporate governing PDEs into learning through residual-based objectives (Raissi et al., 2019), including Helmholtz-equation constraints for acoustic-field reconstruction (Chen et al., 2025). TIDAL differs by incorporating the S/c representation revealed by wave physics through structured decoding and using sampling-aligned attention to match the H/V observation geometry.

3 Problem Formulation

We consider reconstruction of a 3D complex ocean acoustic field from sparse measurements. At a fixed frequency, let $p(z, r, \alpha) \in \mathbb{C}$ denote the complex acoustic pressure at depth z, range r, and angular direction α. Such fields characterize acoustic propagation and interference across space and direction, providing useful information for underwater sensing and environmental characterization (Zhu et al., 2024). In practice, dense spatial sampling is often infeasible because measurements are constrained by platform trajectories and array configurations, naturally leading to structured acquisition patterns such as H/V transects (Wan et al., 2009; Tang et al., 2007).

Classical sparse-field reconstruction typically formulates this task as an underdetermined inverse problem, $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{x}$ is the vectorized field, $\mathbf{y}$ is the measurements, and $\mathbf{A}$ is the sensing matrix. Recovering $\mathbf{x}$ from $\mathbf{y}$ therefore relies on additional priors, including spatial or sparse representations (Sun et al., 2017; 2020), learned representations (Chen et al., 2026), and physics-based constraints or dictionaries (Chen et al., 2025; Damiano et al., 2025). However, the inverse formulation alone does not characterize recoverability under the structured H/V sampling geometry considered here. Such guarantees require the observation geometry to be analyzed jointly with a compatible field representation, as in coupled tensor recovery under systematic slab sampling (Zhang et al., 2020). We therefore formulate both the field and its structured H/V observations in tensor form. We discretize the field over I depths $\{z_i\}_{i=1}^{I}$, J ranges $\{r_j\}_{j=1}^{J}$, and K angular directions $\{\alpha_k\}_{k=1}^{K}$, yielding

$$\mathcal{X} \in \mathbb{C}^{I\times J\times K}, \qquad \mathcal{X}_{ijk} = p(z_i, r_j, \alpha_k). \tag{1}$$

Let $\mathcal{I}_H \subseteq [I]$ and $\mathcal{J}_V \subseteq [J]$ index the selected horizontal and vertical transects, respectively, where $[I] = \{1, ..., I\}$ and similarly for $[J]$ and $[K]$. The resulting observation set is

$$\Omega_{\mathrm{HV}} = \Big[(\mathcal{I}_H \times [J]) \cup ([I] \times \mathcal{J}_V)\Big] \times [K]. \tag{2}$$

The same depth–range sampling pattern is shared across angular directions. The observed field is therefore

$$\mathcal{X}_{\mathrm{obs}} = \mathcal{P}_{\Omega_{\mathrm{HV}}}(\mathcal{X}), \tag{3}$$

where $\mathcal{P}_{\Omega_{\mathrm{HV}}}$ denotes the sampling operator associated with Ω_{HV}.

Our objective is to recover the full complex tensor $\mathcal{X}$ from $\mathcal{X}_{\mathrm{obs}}$. We therefore view sparse H/V acoustic-field reconstruction as a structured tensor inverse problem. In Sec. 4, we derive a structured representation of $\mathcal{X}$ from the underlying wave physics and show how its low-rank specialization connects the H/V observation model to existing LL1 recoverability results.

4 From Wave Physics to Structured Tensor Representation

Following the physics-to-tensor pathway in Fig. 1, we first derive a structured tensor representation of the multi-angle acoustic field from the classical $N \times 2$D description of underwater propagation (Jensen et al., 2011). We then show that a low-rank specialization takes the form of an LL1 decomposition, which connects the structured H/V observation model to existing recoverability results.

4.1 From Modal Propagation to the S/c Representation

Under the classical $N \times 2$D approximation, a 3D acoustic field is represented by a collection of 2D depth–range fields, one for each angular direction α, with $p_\alpha(z, r) \equiv p(z, r, \alpha)$ (Jensen et al., 2011). For time-harmonic propagation in a range-independent waveguide, each directional field satisfies

$$\frac{\partial^2 p_\alpha}{\partial r^2} + \frac{1}{r}\frac{\partial p_\alpha}{\partial r} + \frac{\partial^2 p_\alpha}{\partial z^2} + \frac{\omega^2}{c_{\mathrm{s},\alpha}^2(z)} p_\alpha = 0, \tag{4}$$

where ω is the angular frequency and $c_{\mathrm{s},\alpha}(z)$ is the sound-speed profile along direction α.

To expose the depth–range structure of each directional field, we apply separation of variables, $p_\alpha(z, r) = \psi(z)\phi(r)$, reducing equation 4 to the standard depth eigenvalue problem and its associated range equation (Jensen et al., 2011). The resulting depth eigenfunctions define the propagation modes. Retaining L_m modes gives

$$p_\alpha(z, r) \approx \sum_{\ell=1}^{L_m} \psi_{\ell,\alpha}(z)\phi_{\ell,\alpha}(r), \tag{5}$$

where $\psi_{\ell,\alpha}$ is the depth eigenfunction of the ℓth mode and $\phi_{\ell,\alpha}$ is its corresponding range-dependent response. The separated modal equations and boundary conditions are provided in Appendix A.1.

We next organize these directional solutions along the angular coordinate. Since the angular coordinate $\alpha \in [0, 2\pi)$ is periodic, the multi-angle field admits the Fourier-series representation (Stein & Shakarchi, 2011)

$$p(z, r, \alpha) = \sum_{n\in\mathbb{Z}} \widehat{p}_n(z, r) e^{\mathrm{i}n\alpha}. \tag{6}$$

Here, n denotes the angular Fourier index, and the corresponding depth–range coefficient is

$$\widehat{p}_n(z, r) = \frac{1}{2\pi}\int_0^{2\pi} p(z, r, \alpha) e^{-\mathrm{i}n\alpha}\, d\alpha \approx \frac{1}{2\pi}\sum_{\ell=1}^{L_m}\int_0^{2\pi} \psi_{\ell,\alpha}(z)\phi_{\ell,\alpha}(r) e^{-\mathrm{i}n\alpha}\, d\alpha. \tag{7}$$

Thus, each angular Fourier coefficient aggregates the angle-dependent modal contributions into a depth–range component. Retaining angular orders $|n| \le N_\alpha$, with $Q = 2N_\alpha + 1$, gives the angularly truncated field approximation

$$p_{N_\alpha}(z, r, \alpha) = \sum_{|n|\le N_\alpha} \widehat{p}_n(z, r) e^{\mathrm{i}n\alpha} = \sum_{q=1}^{Q} S_q(z, r) c_q(\alpha), \tag{8}$$

where $n_q = -N_\alpha + q - 1$, $c_q(\alpha) \triangleq e^{in_q\alpha}$ is the qth retained angular Fourier harmonic, and $S_q(z, r) \triangleq \widehat{p}_{n_q}(z, r)$ is the corresponding depth–range coefficient field.

Proposition 1 (Finite-dimensional S/c representation)**.** *Under the L_m-mode approximation and angular truncation $|n| \le N_\alpha$, with $Q = 2N_\alpha + 1$, the depth–range components $S_q \in \mathbb{C}^{I\times J}$ and angular factors $c_q \in \mathbb{C}^K$ defined above yield the sampled field*

$$\mathcal{X} = \sum_{q=1}^{Q} S_q \circ c_q + \mathcal{E}, \tag{9}$$

where $\circ$ denotes the outer product. The residual $\mathcal{E}$ consists of the discarded modal contribution and the angular Fourier truncation error; if the angular dependence has s square-integrable derivatives, the latter decays as $\mathcal{O}(N_\alpha^{-s})$.

We refer to this form as the S/c representation. The detailed modal and angular truncation derivations are provided in Appendix A.1.

4.2 Low-Rank Specialization and H/V Recoverability

For theoretical characterization, we consider a low-rank specialization of the spatial components motivated by the normal-mode structure. For each sampled direction α_k, define $X_k = [p_{\alpha_k}(z_i, r_j)]_{i,j}$, together with the sampled depth-mode and range-response matrices $\Psi_k = [\psi_{\ell,\alpha_k}(z_i)]_{i,\ell}$, $\Phi_k = [\phi_{\ell,\alpha_k}(r_j)]_{j,\ell}$. equation 5 then gives $X_k \approx \Psi_k \Phi_k^\top$, so each directional slice is generated by at most L_m retained modal contributions. Across angle, these modal contributions are not independent: directional variations of the same propagation modes can remain strongly correlated. Consequently, the angular Fourier coefficient fields S_q in equation 9 can exhibit low effective spatial rank. Motivated by this structure, we consider the specialization

$$\operatorname{rank}(S_q) \le L, \qquad q = 1, \dots, Q, \tag{10}$$

where L denotes the prescribed spatial rank of each component. In the rank-L case, each spatial component admits a factorization $S_q = A_q B_q^\top$, where $A_q \in \mathbb{C}^{I \times L}$ and $B_q \in \mathbb{C}^{J \times L}$. For the theoretical analysis, we consider the exact representation case $\mathcal{E} = 0$. The tensor model becomes

$$\mathcal{X} = \sum_{q=1}^{Q} (A_q B_q^\top) \circ c_q, \tag{11}$$

which takes the form of an LL1 decomposition (De Lathauwer, 2008; Rontogiannis et al., 2021).

To relate this LL1 specialization to the H/V observation model, let P_H and Q_V denote the coordinate-selection matrices for the H/V transects defined in Sec. 3, with $I_H = |\mathcal{I}_H|$ and $J_V = |\mathcal{J}_V|$. Applying these coordinate selections preserves the LL1 structure:

$$\mathcal{Y}^H = \sum_{q=1}^{Q} (P_H S_q) \circ c_q, \qquad \mathcal{Y}^V = \sum_{q=1}^{Q} (S_q Q_V^\top) \circ c_q, \tag{12}$$

and form two coupled LL1 tensors sharing the angular factors $\{c_q\}_{q=1}^Q$. The H/V observation model therefore constitutes a coupled slab-sampling instance of the framework studied by Zhang et al. (2020), allowing its recoverability result to be specialized to the present setting.

Theorem 2. *(Zhang et al., 2020). Assume that the rank-$(L, L, 1)$ LL1 factors are generic and that $K \ge Q$. If*

$$I_H \ge 2L, \qquad J \ge LQ, \qquad \min\left(\left\lfloor \frac{I_H}{L} \right\rfloor, Q\right) + \min\left(\left\lfloor \frac{J}{L} \right\rfloor, Q\right) \ge Q + 2, \tag{13}$$

then the coupled slab observations identify the block terms $\{S_q, c_q\}_{q=1}^Q$ up to a common permutation and reciprocal nonzero scaling. The symmetric condition obtained by interchanging the depth and range modes is given in Appendix A.3.

These ambiguities leave each product $S_q \circ c_q$ unchanged and therefore do not affect the recomposed tensor. Consequently, the full tensor $\mathcal{X}$ is uniquely recoverable from the coupled H/V observations. Although the result of Zhang et al. (2020) is stated for real-valued LL1 factors, the required LL1 essential-uniqueness property also holds over $\mathbb{C}$ (De Lathauwer, 2008, Theorem 4.4), so the same coupled slab-sampling argument yields recoverability for the complex-valued acoustic field. Details are given in Appendix A.3.

This result is relevant to ocean sensing because measurements are often constrained to structured transects rather than freely placed spatial samples. The physics-derived S/c representation therefore converts H/V acquisition into a coupled slab-sampling problem, enabling existing LL1 theory to characterize sufficient conditions for full-field recovery.

5 TIDAL: PHYSICS-INSPIRED NEURAL RECONSTRUCTION

The S/c representation derived in equation 9 provides a physics-grounded structure of the field into depth–range spatial components and angular factors. Nonetheless, in addition to the low-rank structure of S_q that mostly models global "slow variation" of the field, in realistic propagation settings, acoustic fields can exhibit rich interference patterns and spatial variations. This motivates more detail-capturing, data-adaptive spatial modeling. We therefore propose TIDAL, with a decoder that retains the S/c representation while learning data-adaptive spatial and angular components. To infer these components from sparse H/V observations, TIDAL uses sampling-aligned axial attention to aggregate information along depth and range, following the observation geometry. Component-wise scale normalization further provides a consistent magnitude convention for the reciprocal scaling ambiguity of the S/c representation.

Given the sparse complex observation $\mathcal{X}_{\text{obs}} = \mathcal{P}_{\Omega_{\text{HV}}}(\mathcal{X})$, we separate its real and imaginary parts and concatenate them with the spatial observation mask and normalized depth and range coordinates:

$$Z_{\text{in}} = \text{Concat}\left[\Re(\mathcal{X}_{\text{obs}}), \Im(\mathcal{X}_{\text{obs}}), M, Z, R\right], \tag{14}$$

where M denotes the spatial sampling mask, and Z and R are the coordinate maps. The angular responses are represented along the channel dimension, while the full $I \times J$ depth–range grid is retained throughout the encoder. A pointwise lifting layer maps the input to an initial latent field $F^{(0)} \in \mathbb{R}^{I\times J\times D}$, where D denotes the latent feature dimension.

5.1 SAMPLING-ALIGNED AXIAL ATTENTION

The H/V sensing geometry is naturally aligned with the depth and range axes. We therefore factorize spatial attention along these two directions, alternating range- and depth-wise self-attention rather than attending globally over the full spatial grid (Ho et al., 2019). This geometry-aligned design enables long-range feature interaction while directly matching information propagation to the observation structure.

Given a latent feature field $F^{(\ell)}$, each axial block first applies self-attention independently along the range axis for each depth index,

$$\widetilde{F}^{(\ell)} = \mathcal{A}_r\left(F^{(\ell)}\right), \tag{15}$$

followed by attention along the depth axis for each range index,

$$F^{(\ell+1)} = \mathcal{A}_z\left(\widetilde{F}^{(\ell)}\right). \tag{16}$$

Here, $\mathcal{A}_r$ and $\mathcal{A}_z$ denote multi-head self-attention blocks with residual feed-forward updates along the corresponding axes. Alternating the two operations progressively propagates information from the observed transects across the depth–range field.

5.2 S/c DECODER

Let $F \in \mathbb{R}^{I\times J\times D}$ denote the output of the axial encoder, and Q_N the number of learned S/c components. The S/c representation in equation 9 organizes the field into spatial components and their corresponding angular factors. Rather than directly predicting the K complex field slices, the decoder parameterizes the output through separate spatial components and angular factors, which are recombined to reconstruct the full field.

Spatial components. The spatial branch directly predicts Q_N dense complex components,

$$\{S_{\theta,q}\}_{q=1}^{Q_N} = \mathcal{D}_S(F), \qquad S_{\theta,q} \in \mathbb{C}^{I\times J}, \tag{17}$$

where θ denotes the learnable model parameters, and $\mathcal{D}_S$ denotes the spatial decoder heads. The dense parameterization follows the spatial components structure in equation 9 while allowing them to adapt to richer field structure.

Angular factors. Following the Fourier parameterization in Sec. 4.1, we encode the angular coordinate with Fourier features and further learn data-adaptive angular factors to capture field-dependent directional structure. We first aggregate the encoder output by global average pooling (GAP) over the depth–range dimensions,

$$g = \mathrm{GAP}(F) \in \mathbb{R}^D, \tag{18}$$

and, for each angle α_k, predict

$$\{c_{\theta,q}(k)\}_{q=1}^{Q_N} = \mathcal{D}_c\left(g, \gamma(\alpha_k)\right), \tag{19}$$

where $\mathcal{D}_c$ denotes the angular decoder heads, and $\gamma(\alpha_k)$ is a Fourier encoding of the angular coordinate (Tancik et al., 2020). The pooled feature provides field-level context, while $\gamma(\alpha_k)$ encodes periodic angular dependence.

Scale normalization. The S/c recomposition is unchanged under the reciprocal rescaling $S_{\theta,q} \to \alpha_q S_{\theta,q}$ and $c_{\theta,q} \to \alpha_q^{-1} c_{\theta,q}$ (Kolda & Bader, 2009). This symmetry allows the same component contribution to be represented with arbitrary paired magnitudes in $S_{\theta,q}$ and $c_{\theta,q}$. We therefore normalize the spatial magnitude of each component before recomposition. Specifically, we define

$$\rho_q = \left(\frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} |S_{\theta,q}(i,j)|^2 \right)^{1/2}, \tag{20}$$

and apply the reciprocal rescaling $\overline{S}_{\theta,q} = \frac{S_{\theta,q}}{\rho_q}, \overline{c}_{\theta,q} = \rho_q c_{\theta,q}$. The reconstructed field is then

$$\widehat{\mathcal{X}}_{ijk} = \sum_{q=1}^{Q_N} \overline{S}_{\theta,q}(i,j)\overline{c}_{\theta,q}(k). \tag{21}$$

5.3 Training with Sampling-Geometry Augmentation

To reduce sensitivity to a particular transect layout, we resample the observation mask during training. The numbers of H/V transects are independently selected from a predefined set, and their locations are drawn from equispaced, shifted, and jittered sampling geometries.

The model is trained end-to-end with a complex-valued reconstruction objective combining full-field and region-specific reconstruction errors. A mild spatial smoothness regularizer is applied to the learned spatial components. Detailed loss definitions and optimization settings are provided in the Appendix B.

6 Experiments

6.1 Experimental Setup

Dataset. The 3D complex acoustic fields were obtained using KRAKEN (Porter, 1992), a physics-based propagation solver, under realistic ocean environments. The environments include temperature and salinity fields from finite volume coastal ocean model (FVCOM) reanalysis data (Chen et al., 2006), and bathymetry from the ETOPO1 Global Relief Model (Amante & Eakins, 2009), covering a region spanning approximately $17.2° - 17.6°$N and $110.1° - 110.5°$E. We select three source locations and sample 72 azimuthal directions at each location, yielding direction-dependent sound speed profiles and water depths. Combined with three acoustic frequencies (100, 150, 200 Hz) and four source depths (30, 50, 80, 120 m), these configurations span diverse source and propagation conditions. The resulting 660 fields are split into 500/80/80 for training, validation, and testing, with each sample represented on a $200 \times 200 \times 72$ depth–range–azimuth grid.

Baselines. We compare against representative learning-based reconstruction methods, including MxT (Chen et al., 2024), LaMO (Tiwari et al., 2025), DiffusionPDE (Huang et al., 2024), and SCPNet (Xia et al., 2023), using identical acoustic field splits and H/V observation protocols. Architecture-specific adaptations are detailed in Appendix B.5.

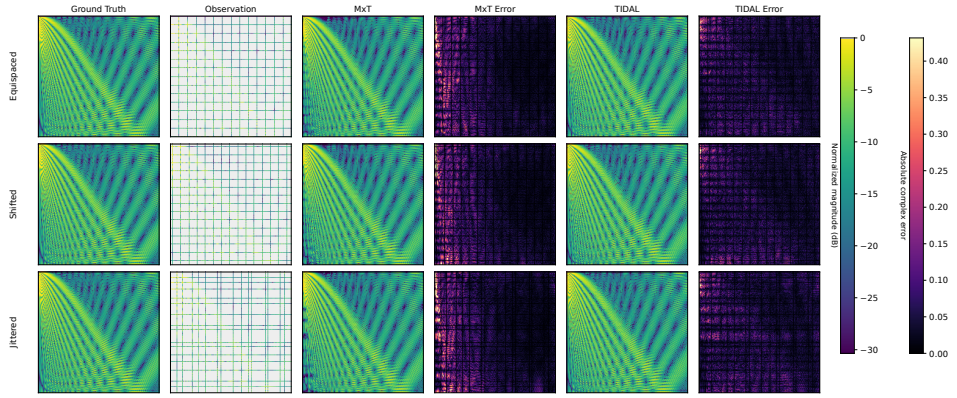


Figure 2: Qualitative reconstruction under three sampling geometries. Columns show the ground truth, sparse observations, and reconstructions and absolute complex errors of MxT and TIDAL for the same test field and angular slice. Field magnitudes and errors use shared scales across all cases.

Table 1: Reconstruction performance and inference efficiency under three sampling geometries. Results are reported as mean ± standard deviation over all test fields. "Time" denotes per-field inference time. Best results are shown in bold.

Method	Metric	Equispaced		Shifted		Jittered		Time (s) ↓
		Full	Unobs.	Full	Unobs.	Full	Unobs.	
LaMO	RMSE ↓	0.712 ± 0.068	0.749 ± 0.073	0.704 ± 0.068	0.743 ± 0.072	0.717 ± 0.067	0.757 ± 0.071	0.91
	MAC ↑	0.488 ± 0.098	0.434 ± 0.110	0.501 ± 0.097	0.444 ± 0.108	0.483 ± 0.096	0.423 ± 0.108	
SCPNet	RMSE ↓	0.504 ± 0.056	0.541 ± 0.060	0.486 ± 0.055	0.526 ± 0.059	0.542 ± 0.052	0.586 ± 0.056	1.71
	MAC ↑	0.743 ± 0.056	0.704 ± 0.065	0.761 ± 0.053	0.720 ± 0.062	0.704 ± 0.056	0.653 ± 0.066	
MxT	RMSE ↓	0.278 ± 0.073	0.298 ± 0.078	0.281 ± 0.070	0.303 ± 0.076	0.293 ± 0.072	0.317 ± 0.078	0.72
	MAC ↑	0.918 ± 0.042	0.905 ± 0.049	0.917 ± 0.041	0.903 ± 0.048	0.909 ± 0.044	0.894 ± 0.052	
DiffusionPDE	RMSE ↓	0.394 ± 0.131	0.423 ± 0.140	0.342 ± 0.137	0.370 ± 0.147	0.331 ± 0.121	0.358 ± 0.131	901.6
	MAC ↑	0.831 ± 0.108	0.806 ± 0.124	0.867 ± 0.103	0.845 ± 0.118	0.877 ± 0.093	0.857 ± 0.107	
TIDAL	RMSE ↓	$\mathbf{0.256 \pm 0.083}$	$\mathbf{0.269 \pm 0.085}$	$\mathbf{0.259 \pm 0.082}$	$\mathbf{0.274 \pm 0.085}$	$\mathbf{0.278 \pm 0.085}$	$\mathbf{0.295 \pm 0.089}$	**0.14**
	MAC ↑	$\mathbf{0.928 \pm 0.047}$	$\mathbf{0.920 \pm 0.051}$	$\mathbf{0.926 \pm 0.047}$	$\mathbf{0.918 \pm 0.052}$	$\mathbf{0.916 \pm 0.052}$	$\mathbf{0.905 \pm 0.057}$	

Evaluation metrics. We report relative root-mean-square error (RMSE) and modal assurance criterion (MAC) (Schmid et al., 2021) over the full tensor and over unobserved spatial locations. MAC provides a structure-aware measure of normalized similarity between complex acoustic fields, complementing the pointwise reconstruction error measured by RMSE. Unless otherwise specified, evaluation uses H15/V15 (15 horizontal and 15 vertical transects) observations under equispaced, shifted, and jittered geometries.

6.2 Main Reconstruction Results

Table 1 shows that TIDAL achieves the best reconstruction performance among the baselines. The larger gains on unobserved locations indicate improved recovery of missing field values rather than reproduction of measured transects. Although jittered sampling is more challenging than equispaced or shifted layouts, TIDAL maintains its advantage across all geometries. It also achieves the lowest inference time, yielding a favorable accuracy–efficiency trade-off.

Fig. 2 compares TIDAL with MxT, the strongest baseline in Table 1. While both recover the dominant propagation pattern, TIDAL better preserves the continuity and the oscillatory interference patterns in unobserved regions, with lower residual errors across all three sampling geometries.

6.3 Effect of Spatial-Rank Constraints

The H/V recoverability analysis in Sec. 4.2 considers a low-rank specialization of S_q. TIDAL directly learns dense spatial components within the same S/c representation. We therefore examine whether low-rank structure emerges in the learned components and the effect of prescribed spatial rank on representation capacity and end-to-end reconstruction.

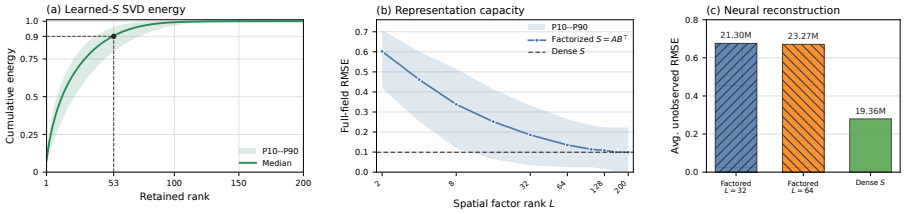


Figure 3: Effect of spatial-rank constraints. **(a)** Cumulative singular-value energy of the learned dense spatial components. **(b)** Representation error versus prescribed spatial rank L. **(c)** End-to-end reconstruction with factorized and dense-S decoders. Parameter counts are shown above each bar.

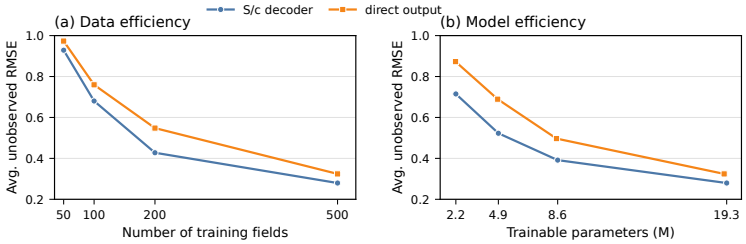


Figure 4: Data and model efficiency of S/c decoding. Unobserved-region RMSE versus **(a)** training-set size and **(b)** model size, comparing S/c and direct-output decoders.

Fig. 3(a) shows that the learned dense spatial components exhibit concentrated singular spectra, with the median cumulative energy reaching 90% at rank 53. This indicates that substantial low-rank structure emerges without being explicitly imposed. Despite this spectral concentration, Fig. 3(b) shows that oracle reconstruction error decreases with spatial rank L, but a clear gap to dense S remains at moderate ranks and closes only at high ranks. Thus, prescribing a moderate spatial rank can limit the representation capacity of the S/c model.

Fig. 3(c) further examines this effect in end-to-end neural reconstruction. Increasing the factorized rank from $L = 32$ to $L = 64$ yields only a marginal improvement, whereas dense spatial components substantially reduce the reconstruction error using fewer trainable parameters than either factorized variant. Together, these results support the use of data-adaptive spatial components in TIDAL while retaining the physics-derived S/c representation.

6.4 Data and Model Efficiency of the S/c Decoder

We next examine whether retaining the S/c representation provides a useful inductive bias under different training-set sizes and model capacities. As shown in Fig. 4, the S/c decoder consistently outperforms direct output, with larger gains in limited-data and lightweight-model regimes. Since paired models have comparable parameter counts, the gains are associated with structured output parameterization rather than model size, particularly under limited supervision or capacity.

7 Conclusion

This work demonstrates a representation-level perspective on incorporating physics into neural reconstruction: the mathematical structure of the governing wave equation can directly inform how a physical field is represented. For multi-angle ocean acoustic fields, the resulting S/c representation provides a structured organization of the field and, under a low-rank specialization, connects sparse H/V reconstruction to classical LL1 recoverability theory, enabling sufficient conditions for full-field recovery to be characterized. More broadly, these results illustrate how physics-derived representations can provide a principled interface between wave physics, structured observations, and data-driven reconstruction. Future work will investigate whether this representation-level use of physics can be extended to a broader range of physical fields.

AI USE STATEMENT

Generative AI tools were used to assist with writing and language polishing. All AI-assisted edits were reviewed and verified by the authors, who take full responsibility for the final content of this paper.

REFERENCES

Christopher Amante and Barry W Eakins. Etopo1 arc-minute global relief model: procedures, data sources and analysis. 2009.

Emmanuel J Candes and Yaniv Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98 (6):925–936, 2010.

Claudio Canuto, M Youssuff Hussaini, Alfio Quarteroni, and Thomas A Zang. *Spectral methods*, volume 285. Springer, 2006.

Changsheng Chen, Robert C Beardsley, and Geoffrey Cowles. Finite volume coastal ocean. *Oceanography*, 19(1):78, 2006.

Fangchao Chen, Youhong Xiao, Liang Yu, and Laixu Jiang. Frequency-domain physics-informed neural network for accurate reconstruction of 3d acoustic fields under sparse and multi-frequency measurements. *Neural Networks*, pp. 108476, 2025.

Fangchao Chen, Youhong Xiao, Liang Yu, and Laixu Jiang. Physics-aware graph learning for sound-field reconstruction from sparse measurements. *The Journal of the Acoustical Society of America*, 159(3):1973–1986, 2026.

Shuang Chen, Amir Atapour-Abarghouei, Haozheng Zhang, and Hubert P. H. Shum. Mxt: Mamba x transformer for image inpainting. In *35th British Machine Vision Conference 2024, BMVC 2024, Glasgow, UK, November 25-28, 2024*. BMVA, 2024. URL `https://papers.bmvc2024.org/0295.pdf`.

Stefano Damiano, Federico Miotello, Mirco Pezzoli, Alberto Bernardini, Fabio Antonacci, Augusto Sarti, and Toon Van Waterschoot. A zero-shot physics-informed dictionary learning approach for sound field reconstruction. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2025.

Lieven De Lathauwer. Decompositions of a higher-order tensor in block terms—part ii: Definitions and uniqueness. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1033–1066, 2008.

Meng Ding, Xiao Fu, Ting-Zhu Huang, Jun Wang, and Xi-Le Zhao. Hyperspectral super-resolution via interpretable block-term tensor modeling. *IEEE Journal of Selected Topics in Signal Processing*, 15(3):641–656, 2021.

Richard A Harshman et al. Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA working papers in phonetics*, 16(1):84, 1970.

Jonathan Ho, Nal Kalchbrenner, Dirk Weissenborn, and Tim Salimans. Axial attention in multidimensional transformers. *arXiv preprint arXiv:1912.12180*, 2019.

Jiahe Huang, Guandao Yang, Zichen Wang, and Jeong Joon Park. Diffusionpde: Generative pde-solving under partial observation. *Advances in Neural Information Processing Systems*, 37: 130291–130323, 2024.

Finn B Jensen, William A Kuperman, Michael B Porter, Henrik Schmidt, and Alexandra Tolstoy. *Computational ocean acoustics*, volume 2011. Springer, 2011.

Charilaos I. Kanatsoulis, Xiao Fu, Nicholas D. Sidiropoulos, and Wing-Kin Ma. Hyperspectral super-resolution: A coupled tensor factorization approach. *IEEE Transactions on Signal Processing*, 66(24):6503–6517, 2018.

Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3): 455–500, 2009.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Yuezhou Ma, Haixu Wu, Hang Zhou, Huikun Weng, Jianmin Wang, and Mingsheng Long. Physense: Sensor placement optimization for accurate physics sensing. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 35697–35725. Curran Associates, Inc., 2025. doi: 10.52202/085713-1200.

Gerlind Plonka, Daniel Potts, Gabriele Steidl, and Manfred Tasche. *Numerical Fourier Analysis*. Birkhäuser, Basel, 2018.

Michael B Porter. The kraken normal mode program. 1992.

Maziar Raissi, Paris Perdikaris, and George E 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.

Athanasios A Rontogiannis, Eleftherios Kofidis, and Paris V Giampouras. Block-term tensor decomposition: Model selection and computation. *IEEE Journal of Selected Topics in Signal Processing*, 15(3):464–475, 2021.

Javier E Santos, Zachary R Fox, Arvind Mohan, Daniel O'Malley, Hari Viswanathan, and Nicholas Lubbers. Development of the senseiver for efficient field reconstruction from sparse observations. *Nature Machine Intelligence*, 5(11):1317–1325, 2023.

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

Mikael Sørensen, Ignat Domanov, and Lieven De Lathauwer. Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank- $(l_r,n,l_r,n,1)$ terms—part ii: Algorithms. *SIAM Journal on Matrix Analysis and Applications*, 36(3):1015–1045, 2015. doi: 10.1137/140956865. URL https://doi.org/10.1137/140956865.

Elias M Stein and Rami Shakarchi. *Fourier analysis: an introduction*, volume 1. Princeton University Press, 2011.

Jie Sun, Aijun Song, Jiancheng Yu, Aiqun Zhang, and Fumin Zhang. Underwater acoustic field reconstruction using a compressive sensing approach. In *OCEANS 2017-Anchorage*, pp. 1–5. IEEE, 2017.

Jie Sun, Shijie Liu, Fumin Zhang, Aijun Song, Jiancheng Yu, and Aiqun Zhang. A kriged compressive sensing approach to reconstruct acoustic fields from measurements collected by underwater vehicles. *IEEE Journal of Oceanic Engineering*, 46(1):294–306, 2020.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33:7537–7547, 2020.

Dajun Tang, James N Moum, James F Lynch, Phil Abbot, Ross Chapman, Peter H Dahl, Timothy F Duda, Glen Gawarkiewicz, Scott Glenn, John A Goff, Hans Graber, John Kemp, Andrew Maffei, Jonathan D Nash, and Arthur Newhall. Shallow water '06: A joint acoustic propagation/nonlinear internal wave physics experiment. *Oceanography*, 20(4):156–167, 2007.

Karn Tiwari, Niladri Dutta, N M Anoop Krishnan, and Prathosh Ap. Latent mamba operator for partial differential equations. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 59639–59667. PMLR, 13–19 Jul 2025.

Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3): 279–311, 1966.

Lin Wan, Ji-Xun Zhou, Peter H Rogers, and David P Knobles. Spatial coherence measurements from two l-shape arrays in shallow water. *Acoustical Physics*, 55(3):383–392, 2009.

Zhaoyang Xia, Youquan Liu, Xin Li, Xinge Zhu, Yuexin Ma, Yikang Li, Yuenan Hou, and Yu Qiao. Scpnet: Semantic scene completion on point cloud. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17642–17651. IEEE, 2023.

Guoyong Zhang, Xiao Fu, Jun Wang, Xi-Le Zhao, and Mingyi Hong. Spectrum cartography via coupled block-term tensor decomposition. *IEEE Transactions on Signal Processing*, 68:3660–3675, 2020.

Xiaoyu Zhao, Xiaoqian Chen, Zhiqiang Gong, Weien Zhou, Wen Yao, and Yunyang Zhang. Recfno: A resolution-invariant flow and heat field reconstruction method from sparse observations via fourier neural operator. *International Journal of Thermal Sciences*, 195:108619, 2024.

Xiaohan Zhu, Ye Wang, Zeyu Fang, Lei Cheng, and Jianlong Li. Strategic deployment in the deep: Principled underwater sensor placement optimization with three-dimensional acoustic map. *The Journal of the Acoustical Society of America*, 156(4):2668–2685, 2024.

A Theory Details for the S/c Representation and Complex H/V Recoverability

This appendix reviews additional derivation details for the development in Sec. 4. We first introduce the preliminaries of the normal mode separation and finite-dimensional angular representation underlying Proposition 1, including the modal and angular truncation errors. We then summarize the classical complex LL1 essential-uniqueness result (De Lathauwer, 2008) used for the low-rank specialization and provide the complete symmetric conditions associated with the coupled slab-sampling result of Zhang et al. (2020) for structured H/V observations.

A.1 Details of the Finite-Dimensional S/c Representation

We first provide the normal-mode and angular approximation details underlying Proposition 1. For a fixed direction α, substituting the separated form $p_\alpha(z,r) = \psi(z)\phi(r)$ into equation 4 yields (Jensen et al., 2011)

$$\frac{\mathrm{d}^2\psi_{\ell,\alpha}}{\mathrm{d}z^2} + \left[\frac{\omega^2}{c_{\mathrm{s},\alpha}^2(z)} - k_{r,\ell,\alpha}^2\right]\psi_{\ell,\alpha} = 0, \tag{22}$$

and

$$\frac{\mathrm{d}^2\phi_{\ell,\alpha}}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}\phi_{\ell,\alpha}}{\mathrm{d}r} + k_{r,\ell,\alpha}^2\phi_{\ell,\alpha} = 0, \tag{23}$$

where $k_{r,\ell,\alpha}$ is the horizontal wavenumber associated with the ℓth mode. Together with the prescribed surface and bottom boundary conditions, equation 22 defines the depth eigenvalue problem, while equation 23 gives the corresponding range-dependent responses for each eigenmode (Jensen et al., 2011). The resulting eigenfunctions form the normal-mode basis, and linear superposition gives the modal expansion

$$p(z,r,\alpha) = \sum_{\ell\geq 1}\psi_{\ell,\alpha}(z)\phi_{\ell,\alpha}(r). \tag{24}$$

For a finite modal approximation, retaining L_m modes gives

$$p^{(L_m)}(z,r,\alpha) = \sum_{\ell=1}^{L_m}\psi_{\ell,\alpha}(z)\phi_{\ell,\alpha}(r), \tag{25}$$

with the discarded modal contribution $p(z,r,\alpha) - p^{(L_m)}(z,r,\alpha)$.

We next apply the angular expansion to $p^{(L_m)}$ (Stein & Shakarchi, 2011). Its Fourier coefficients are

$$\begin{aligned}\widehat{p}_n^{(L_m)}(z,r) &= \frac{1}{2\pi}\int_0^{2\pi} p^{(L_m)}(z,r,\alpha)e^{-\mathrm{i}n\alpha}\,\mathrm{d}\alpha \\ &= \frac{1}{2\pi}\sum_{\ell=1}^{L_m}\int_0^{2\pi}\psi_{\ell,\alpha}(z)\phi_{\ell,\alpha}(r)e^{-\mathrm{i}n\alpha}\,\mathrm{d}\alpha.\end{aligned} \tag{26}$$

Thus, each angular coefficient collects the directional variation of the retained modal contributions into a depth–range field. The superscript (L_m) is suppressed in Sec. 4.1 for notational simplicity.

Retaining the angular orders $|n| \le N_\alpha$ gives

$$p^{(L_m,N_\alpha)}(z,r,\alpha) = \sum_{|n|\le N_\alpha} \widehat{p}_n^{(L_m)}(z,r)e^{\mathrm{i}n\alpha}. \tag{27}$$

With $Q = 2N_\alpha + 1$, let $n_q = -N_\alpha + q - 1$ and define

$$S_q(z,r) \triangleq \widehat{p}_{n_q}^{(L_m)}(z,r), \qquad c_q(\alpha) \triangleq e^{\mathrm{i}n_q\alpha}, \qquad q = 1,\dots,Q. \tag{28}$$

equation 27 can be then written as

$$p^{(L_m,N_\alpha)}(z,r,\alpha) = \sum_{q=1}^{Q} S_q(z,r)c_q(\alpha). \tag{29}$$

Thus, S_q is the depth–range coefficient field associated with the retained angular Fourier harmonic c_q. Hence, the S/c form directly expresses the truncated angular Fourier expansion through paired depth–range coefficient fields and angular harmonics.

Sampling equation 29 at the depth, range, and angular grids defined in Sec. 3, with

$$[S_q]_{ij} = S_q(z_i,r_j), \qquad [c_q]_k = c_q(\alpha_k), \tag{30}$$

gives

$$\mathcal{X}^{(L_m,N_\alpha)} = \sum_{q=1}^{Q} S_q \circ c_q. \tag{31}$$

Here, Q is the number of retained angular Fourier harmonics, whereas K denotes the number of sampled angular directions.

The residual in Proposition 1 follows from the two successive truncations,

$$\begin{aligned} p(z,r,\alpha) - p^{(L_m,N_\alpha)}(z,r,\alpha) = &\left[p(z,r,\alpha) - p^{(L_m)}(z,r,\alpha)\right] \\ &+ \left[p^{(L_m)}(z,r,\alpha) - p^{(L_m,N_\alpha)}(z,r,\alpha)\right]. \end{aligned} \tag{32}$$

The first term is the discarded modal contribution. For the second term, Parseval's identity (Canuto et al., 2006) gives, for each fixed (z,r),

$$\frac{1}{2\pi}\int_0^{2\pi} \left|p^{(L_m)}(z,r,\alpha) - p^{(L_m,N_\alpha)}(z,r,\alpha)\right|^2 \mathrm{d}\alpha = \sum_{|n|>N_\alpha} \left|\widehat{p}_n^{(L_m)}(z,r)\right|^2. \tag{33}$$

If the angular dependence has s square-integrable derivatives, standard Fourier approximation further yields

$$\begin{aligned} &\left(\frac{1}{2\pi}\int_0^{2\pi} \left|p^{(L_m)}(z,r,\alpha) - p^{(L_m,N_\alpha)}(z,r,\alpha)\right|^2 \mathrm{d}\alpha\right)^{1/2} \\ &\qquad \le (N_\alpha+1)^{-s}\left(\frac{1}{2\pi}\int_0^{2\pi} \left|\partial_\alpha^s p^{(L_m)}(z,r,\alpha)\right|^2 \mathrm{d}\alpha\right)^{1/2}, \end{aligned} \tag{34}$$

which gives the $\mathcal{O}(N_\alpha^{-s})$ angular truncation rate stated in Proposition 1 (Canuto et al., 2006; Plonka et al., 2018). Sampling equation 32 on the spatial–angular grid produces the residual tensor $\mathcal{E}$ in equation 9.

A.2 Classical LL1 Results for the Low-Rank S/c Specialization

We next summarize the classical LL1 essential-uniqueness result used in the complex-valued recoverability analysis. Consider an LL1 model (De Lathauwer, 2008)

$$\mathcal{X} = \sum_{q=1}^{Q} \left(A_q B_q^\top\right) \circ c_q, \qquad A_q \in \mathbb{C}^{I\times L}, \quad B_q \in \mathbb{C}^{J\times L}, \quad c_q \in \mathbb{C}^K. \tag{35}$$

The corresponding spatial block is

$$S_q = A_q B_q^\top \in \mathbb{C}^{I \times J}. \tag{36}$$

Following De Lathauwer (2008), define the partitioned factor matrices

$$A = [A_1, \ldots, A_Q], \qquad B = [B_1, \ldots, B_Q], \qquad C = [c_1, \ldots, c_Q]. \tag{37}$$

For the partitioned matrix A, the block k-rank k'_A is the largest integer r such that every collection of r blocks contains rL linearly independent columns; k'_B is defined analogously. The ordinary Kruskal rank of C is denoted by k_C.

We assume that the real and imaginary parts of all entries of (A, B, C) are jointly absolutely continuous with respect to Lebesgue measure on the corresponding real Euclidean space. Under this standard genericity assumption, the relevant block and Kruskal ranks attain their maximal values almost surely (De Lathauwer, 2008). In particular,

$$k'_A = \min\left(\left\lfloor \frac{I}{L} \right\rfloor, Q\right), \qquad k'_B = \min\left(\left\lfloor \frac{J}{L} \right\rfloor, Q\right), \tag{38}$$

and, when $K \geq Q$,

$$k_C = Q. \tag{39}$$

Moreover, for $I \geq L$ and $J \geq L$, A_q and B_q have full column rank almost surely, and hence

$$\operatorname{rank}(S_q) = \operatorname{rank}(A_q B_q^\top) = L \tag{40}$$

for every q.

The classical LL1 essential-uniqueness result applies over both $\mathbb{R}$ and $\mathbb{C}$ (De Lathauwer, 2008, Theorem 4.4). For equation 35, a sufficient condition is

$$k_C = Q, \qquad k'_A + k'_B \geq Q + 2. \tag{41}$$

Therefore, for generic factors over $\mathbb{C}$, with

$$I \geq L, \qquad J \geq L, \qquad K \geq Q, \tag{42}$$

the condition

$$\min\left(\left\lfloor \frac{I}{L} \right\rfloor, Q\right) + \min\left(\left\lfloor \frac{J}{L} \right\rfloor, Q\right) \geq Q + 2 \tag{43}$$

is sufficient for essential uniqueness of the LL1 decomposition.

Under these conditions, the block terms $\{S_q, c_q\}_{q=1}^Q$ are identifiable up to a common permutation and reciprocal nonzero complex scaling. Equivalently, if another Q-term LL1 decomposition represents the same tensor, then there exist a permutation π of $\{1, \ldots, Q\}$ and nonzero complex scalars $\alpha_1, \ldots, \alpha_Q$ such that

$$\widetilde{S}_q = \alpha_q S_{\pi(q)}, \qquad \widetilde{c}_q = \alpha_q^{-1} c_{\pi(q)}. \tag{44}$$

The individual factors A_q and B_q are not uniquely determined. Indeed, for any nonsingular $F_q \in \mathbb{C}^{L \times L}$,

$$A_q B_q^\top = (A_q F_q)(B_q F_q^{-\top})^\top. \tag{45}$$

Thus, the identifiable spatial quantity is the block $S_q = A_q B_q^\top$ rather than its particular matrix factorization.

A.3 H/V Recoverability from Coupled Slab Sampling

We provide the complete specialization of Zhang et al. (2020, Theorem 3) to the H/V observation model in Sec. 4.2. Related recoverability results for coupled tensor models under structured observations have also been developed for hyperspectral super-resolution (Kanatsoulis et al., 2018; Ding et al., 2021).

Let $P_H \in \{0,1\}^{I_H \times I}$ and $Q_V \in \{0,1\}^{J_V \times J}$ denote the coordinate-selection matrices associated with the selected horizontal and vertical transects, respectively, where $I_H = |\mathcal{I}_H|$ and $J_V = |\mathcal{J}_V|$. For the exact low-rank model in equation 11, the corresponding observations are

$$\mathcal{Y}^H = \sum_{q=1}^{Q} \left((P_H A_q) B_q^\top\right) \circ c_q = \sum_{q=1}^{Q} (P_H S_q) \circ c_q, \tag{46}$$

$$\mathcal{Y}^V = \sum_{q=1}^{Q} \left(A_q (Q_V B_q)^\top\right) \circ c_q = \sum_{q=1}^{Q} (S_q Q_V^\top) \circ c_q. \tag{47}$$

Coordinate selection therefore preserves the LL1 form, and $\mathcal{Y}^H$ and $\mathcal{Y}^V$ form two coupled LL1 tensors with shared angular factors $\{c_q\}_{q=1}^{Q}$. Accordingly, the H/V observation model is a special case of the coupled slab-sampling framework analyzed by Zhang et al. (2020).

In Zhang et al. (2020, Theorem 3), a special case is the present H/V observation model, where the two coupled tensors contain Q block terms and retain all K samples along the shared angular mode. The corresponding third-mode overlap condition therefore reduces to

$$K \geq Q. \tag{48}$$

The remaining spatial sampling conditions yield two symmetric cases. The first requires

$$I_H \geq 2L, \qquad J \geq LQ, \tag{49}$$

together with

$$\min\left(\left\lfloor \frac{I_H}{L} \right\rfloor, Q\right) + \min\left(\left\lfloor \frac{J}{L} \right\rfloor, Q\right) \geq Q + 2. \tag{50}$$

Alternatively, the symmetric condition requires

$$J_V \geq 2L, \qquad I \geq LQ, \tag{51}$$

together with

$$\min\left(\left\lfloor \frac{J_V}{L} \right\rfloor, Q\right) + \min\left(\left\lfloor \frac{I}{L} \right\rfloor, Q\right) \geq Q + 2. \tag{52}$$

Together with equation 48, either set of conditions is sufficient for recoverability under the coupled slab-sampling model.

For complex-valued factors, genericity is understood with respect to the joint real and imaginary parts of the factors. Under this assumption, the LL1 essential-uniqueness condition used in the slab-sampling argument remains valid over $\mathbb{C}$ by De Lathauwer (2008, Theorem 4.4), as summarized in Appendix A.2. The subsequent coupling argument relies on full-rank linear relations and the alignment of the shared angular factors, which are unchanged over $\mathbb{C}$.

Under the above conditions, the coupled observations $(\mathcal{Y}^H, \mathcal{Y}^V)$ identify the block terms $\{S_q, c_q\}_{q=1}^{Q}$ up to a common permutation and reciprocal nonzero complex scaling. These ambiguities do not affect the recomposed tensor because

$$(\alpha_q S_q) \circ (\alpha_q^{-1} c_q) = S_q \circ c_q. \tag{53}$$

Consequently,

$$\mathcal{X} = \sum_{q=1}^{Q} S_q \circ c_q \tag{54}$$

is uniquely determined by the structured H/V observations.

B Implementation Details

B.1 Data Processing and Observation Masks

The dataset spans acoustic frequencies of 100, 150, and 200 Hz, source depths of 30, 50, 80, and 120 m, and three source locations, with approximately balanced coverage across these settings.

For preprocessing, each field is treated independently by compensating normal-mode geometric spreading along range and scaling the resulting complex field by its 99.5th-percentile magnitude. Real and imaginary parts are represented as two channels. During training, the numbers of horizontal and vertical transects are sampled independently from $\{10, 15, 20\}$, and masks are resampled at every epoch from equispaced, shifted, and jittered sampling geometries.

H/V sampling geometries. Let $\mathcal{I}_H \subseteq [I]$ and $\mathcal{J}_V \subseteq [J]$ denote the selected horizontal and vertical transects, respectively, with

$$\Omega_{\mathrm{HV}} = \big((\mathcal{I}_H \times [J]) \cup ([I] \times \mathcal{J}_V)\big) \times [K]. \tag{55}$$

The same spatial observation pattern is shared across all K angular slices. The sets $\mathcal{I}_H$ and $\mathcal{J}_V$ are generated independently using the same sampling rule along the depth and range axes.

For an axis of length n with s selected transects, using zero-based index $m = 0, \ldots, s-1$, we consider three sampling geometries.

Equispaced. The selected indices are uniformly distributed between the two endpoints:

$$\left\{ \mathrm{round}\left(\frac{m(n-1)}{s-1} \right) \right\}_{m=0}^{s-1}. \tag{56}$$

Shifted. An equally spaced grid is shifted by a random phase:

$$\left\{ \left\lfloor \frac{(m+\phi)n}{s} \right\rfloor \right\}_{m=0}^{s-1}, \qquad \phi \sim \mathcal{U}[0,1). \tag{57}$$

The phase shifts are sampled independently for the depth and range axes.

Jittered. The axis is partitioned into s approximately equal and non-overlapping intervals, and one transect is sampled uniformly from each interval:

$$\xi_m \sim \mathrm{Unif}\left(\left\{ \left\lfloor \frac{mn}{s} \right\rfloor, \ldots, \left\lfloor \frac{(m+1)n}{s} \right\rfloor - 1 \right\} \right), \qquad m = 0, \ldots, s-1. \tag{58}$$

The jittered selections are sampled independently along the depth and range axes.

B.2 Model and Objective

TIDAL uses $Q_N = 40$ components, latent width 384, and eight axial-attention blocks with eight heads each. Its input comprises the 72-angle real and imaginary observed fields, a shared spatial observation mask, and normalized depth and range coordinates (147 channels in total). The angular branch uses the normalized angle coordinate together with eight sine/cosine Fourier frequencies.

We optimize the sum of relative complex mean-squared-error terms over the full field, observed locations, unobserved locations, and the far half of the range axis, with weights 1, 1, 1, and 0.5, respectively. We additionally apply a finite-difference smoothness penalty to the learned spatial components with weight 10^{-4}.

B.3 Optimization and Model Selection

Models are trained for 150 epochs with batch size one using AdamW (Loshchilov & Hutter, 2019), an initial learning rate of 3×10^{-4}, and weight decay 10^{-5}. We use a StepLR scheduler that decays the learning rate by a factor of 0.5 every 40 epochs. The random seed is fixed to zero. We select the checkpoint with the lowest validation full-field relative RMSE and use this checkpoint for all test evaluations.

B.4 Evaluation Protocol

Unless stated otherwise, evaluation uses 15 horizontal and 15 vertical transects under the equispaced, shifted, and jittered sampling geometries. We report full-field and unobserved-region relative RMSE

and MAC. Let $\mathbf{x}$ and $\hat{\mathbf{x}}$ denote the vectorized reference and reconstructed fields, respectively. MAC is defined as

$$\mathrm{MAC}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{\left|\mathbf{x}^H \hat{\mathbf{x}}\right|^2}{(\mathbf{x}^H \mathbf{x})\,(\hat{\mathbf{x}}^H \hat{\mathbf{x}})}. \tag{59}$$

Both metrics are computed jointly over all 72 angles for each test field, and the mean and standard deviation are then reported over the 80 test fields. All methods are evaluated on the same held-out acoustic fields and observation masks. Inference time is measured on a single NVIDIA A800 GPU.

B.5 Baseline Implementation and Adaptation

All learning-based baselines are trained on the same 500 acoustic training fields and use the same field normalization, train/validation/test split, and H/V sampling protocol as TIDAL. We retain the main backbone of each baseline while adapting its input, output, and observation-conditioning interface to complex acoustic-field reconstruction.

MxT is adapted from image inpainting to complex field completion on individual depth–range slices. The real and imaginary parts of the sparse observations are provided together with the binary observation mask, and the output head predicts the corresponding dense complex field. The original Mamba–Transformer generator is retained.

LaMO is implemented as a regular-grid operator acting on depth–range slices. Each spatial location is represented by its real and imaginary observations together with the sampling mask, while normalized depth and range coordinates are provided as positional inputs. The model directly predicts the dense real and imaginary field components.

We adapt the completion backbone of SCPNet from semantic scene completion to continuous complex-field reconstruction. The semantic-specific prediction heads are removed, while the multi-scale 3D completion architecture is retained. The network takes the real and imaginary observations together with the occupancy mask and predicts the dense complex field over depth, range, and angle.

DiffusionPDE is adapted to model the real and imaginary components of individual depth–range acoustic slices using its original diffusion-based generative formulation. Sparse measurements are incorporated during iterative sampling through observation-gradient guidance rather than being directly concatenated with the diffusion-model input. The PDE-guidance term is disabled because the physical parameters required to instantiate the PDE constraint are unavailable in our reconstruction setting; reconstruction is therefore conditioned only on the observed acoustic measurements.

C Additional Results under Different H/V Sampling Budgets

To examine whether the main results depend on the default H15/V15 sampling budget, we further evaluate all methods under H10/V10 and H20/V20 observations. The training protocol is unchanged, and all models are evaluated on the same 80 test fields and observation masks under the equispaced, shifted, and jittered sampling geometries. Table 2 reports RMSE and MAC over the full field and unobserved spatial locations.

TIDAL consistently achieves the lowest RMSE under both sampling budgets and across all three geometries. The MAC results show the same overall trend, with TIDAL attaining the highest or closely comparable structural similarity across the evaluated settings. Together with the H15/V15 results in the main text, these results show that the advantage of the proposed reconstruction framework is maintained across different levels of H/V sampling sparsity.

D Complete Analysis of Spatial-Rank Constraints

This section provides additional details for the spatial-rank analysis in Sec. 6.3. We first characterize the singular-value structure of the learned dense spatial components, then examine the representation capacity under prescribed spatial ranks, and finally compare factorized spatial decoders with the dense-S decoder used in TIDAL.

Table 2: Reconstruction performance under H10/V10 and H20/V20 sampling. Results are reported as mean ± standard deviation over 80 test fields. Best results are shown in bold.

Budget	Method	Metric	Equispaced		Shifted		Jittered	
			Full	Unobs.	Full	Unobs.	Full	Unobs.
H10/V10	LaMO	RMSE ↓	0.792 ± 0.060	0.820 ± 0.063	0.770 ± 0.067	0.800 ± 0.069	0.781 ± 0.061	0.812 ± 0.064
		MAC ↑	0.370 ± 0.096	0.325 ± 0.104	0.404 ± 0.103	0.356 ± 0.111	0.387 ± 0.096	0.337 ± 0.104
	SCPNet	RMSE ↓	0.693 ± 0.044	0.724 ± 0.046	0.659 ± 0.047	0.694 ± 0.049	0.676 ± 0.045	0.712 ± 0.047
		MAC ↑	0.518 ± 0.061	0.474 ± 0.067	0.564 ± 0.062	0.517 ± 0.068	0.541 ± 0.060	0.491 ± 0.066
	MxT	RMSE ↓	0.329 ± 0.082	0.343 ± 0.086	0.329 ± 0.079	0.346 ± 0.084	0.343 ± 0.080	0.362 ± 0.084
		MAC ↑	0.886 ± 0.056	0.875 ± 0.061	0.886 ± 0.054	0.874 ± 0.060	$\mathbf{0.876 \pm 0.057}$	0.863 ± 0.063
	DiffusionPDE	RMSE ↓	0.622 ± 0.135	0.651 ± 0.141	0.535 ± 0.132	0.563 ± 0.139	0.538 ± 0.132	0.566 ± 0.138
		MAC ↑	0.605 ± 0.157	0.570 ± 0.170	0.701 ± 0.139	0.670 ± 0.152	0.699 ± 0.137	0.667 ± 0.151
	TIDAL	RMSE ↓	$\mathbf{0.317 \pm 0.095}$	$\mathbf{0.329 \pm 0.097}$	$\mathbf{0.311 \pm 0.091}$	$\mathbf{0.324 \pm 0.095}$	$\mathbf{0.343 \pm 0.093}$	$\mathbf{0.358 \pm 0.097}$
		MAC ↑	$\mathbf{0.891 \pm 0.065}$	$\mathbf{0.883 \pm 0.069}$	$\mathbf{0.895 \pm 0.061}$	$\mathbf{0.887 \pm 0.066}$	0.874 ± 0.068	$\mathbf{0.863 \pm 0.074}$
H20/V20	LaMO	RMSE ↓	0.671 ± 0.064	0.716 ± 0.071	0.667 ± 0.064	0.714 ± 0.070	0.678 ± 0.066	0.725 ± 0.072
		MAC ↑	0.547 ± 0.088	0.484 ± 0.102	0.552 ± 0.087	0.487 ± 0.101	0.538 ± 0.091	0.470 ± 0.105
	SCPNet	RMSE ↓	0.409 ± 0.052	0.451 ± 0.058	0.409 ± 0.053	0.455 ± 0.058	0.450 ± 0.054	0.501 ± 0.059
		MAC ↑	0.830 ± 0.043	0.793 ± 0.053	0.830 ± 0.043	0.790 ± 0.054	0.795 ± 0.049	0.746 ± 0.060
	MxT	RMSE ↓	0.250 ± 0.068	0.276 ± 0.075	0.253 ± 0.065	0.281 ± 0.072	0.263 ± 0.068	0.292 ± 0.075
		MAC ↑	0.933 ± 0.036	0.919 ± 0.043	0.932 ± 0.035	0.916 ± 0.043	0.927 ± 0.037	0.910 ± 0.046
	DiffusionPDE	RMSE ↓	0.321 ± 0.133	0.354 ± 0.147	0.278 ± 0.133	0.308 ± 0.148	0.271 ± 0.105	0.302 ± 0.117
		MAC ↑	0.881 ± 0.096	0.855 ± 0.115	0.906 ± 0.092	0.885 ± 0.112	0.916 ± 0.063	0.896 ± 0.077
	TIDAL	RMSE ↓	$\mathbf{0.237 \pm 0.079}$	$\mathbf{0.253 \pm 0.082}$	$\mathbf{0.241 \pm 0.079}$	$\mathbf{0.259 \pm 0.082}$	$\mathbf{0.254 \pm 0.080}$	$\mathbf{0.274 \pm 0.084}$
		MAC ↑	$\mathbf{0.938 \pm 0.042}$	$\mathbf{0.930 \pm 0.046}$	$\mathbf{0.936 \pm 0.043}$	$\mathbf{0.926 \pm 0.047}$	$\mathbf{0.929 \pm 0.045}$	$\mathbf{0.918 \pm 0.051}$

D.1 Learned Spatial-Component Spectra

We analyze the effective rank structure of the learned dense spatial components $S_{\theta,q} \in \mathbb{C}^{200\times 200}$. Across 80 test fields, $Q_N = 40$ components, and three sampling geometries, this yields 9,600 matrices. For each matrix, we compute the cumulative SVD energy

$$E(r) = \frac{\sum_{i=1}^{r} \sigma_i^2}{\sum_i \sigma_i^2}. \tag{60}$$

At each rank r, we report the pointwise median and P10–P90 range over all matrices, with all components and sampling geometries weighted equally.

D.2 Oracle Representation Capacity

We evaluate the approximation capacity of the low-rank spatial parameterization

$$S_q = A_q B_q^\top$$

for different prescribed spatial ranks L. We fix the number of components to $Q_N = 40$ and evaluate

$$L \in \{2, 4, 8, 16, 32, 64, 100, 128, 160, 200\}.$$

The experiment uses 20 fields uniformly selected from the test set. For each field and rank, we run three independent optimizations for 800 steps using Adam with a learning rate of 0.02, and keep the solution with the lowest reconstruction error. Statistics are then computed across the 20 fields. As a reference, we additionally report the corresponding dense-S representation with $Q_N = 40$.

Table 3 shows a clear reduction in representation error as the prescribed spatial rank increases. Low and moderate ranks retain a substantial approximation gap: the mean RMSE decreases from 0.3379 at $L = 8$ to 0.1855 at $L = 32$ and 0.1355 at $L = 64$, whereas the dense-S reference reaches 0.0992. The factorized representation approaches this reference only at very high ranks, with $L = 160$–200 required to nearly close the gap. These results indicate that imposing a moderate prescribed spatial rank can introduce a substantial representation bottleneck for the acoustic fields considered here.

D.3 Effect of Prescribed Spatial Rank in Neural Reconstruction

We next examine whether the same limitation appears in end-to-end reconstruction. We compare factorized spatial decoders with $L = 32$ and $L = 64$ against the dense-S decoder used in TIDAL. All models share the same backbone, training data, and optimization settings, differing only in the parameterization of the spatial components.

Table 3: Oracle representation error across different prescribed spatial ranks L. Results are full-field relative RMSE over 20 test fields.

Representation	L	Mean	Median	P10	P90
Factorized $S = AB^\top$	2	0.6020	0.6014	0.4201	0.7084
Factorized $S = AB^\top$	4	0.4616	0.4514	0.2524	0.6014
Factorized $S = AB^\top$	8	0.3379	0.3326	0.1188	0.5145
Factorized $S = AB^\top$	16	0.2518	0.2545	0.0647	0.4130
Factorized $S = AB^\top$	32	0.1855	0.1815	0.0344	0.3335
Factorized $S = AB^\top$	64	0.1355	0.1254	0.0256	0.2641
Factorized $S = AB^\top$	100	0.1140	0.1008	0.0246	0.2340
Factorized $S = AB^\top$	128	0.1086	0.0967	0.0162	0.2239
Factorized $S = AB^\top$	160	0.0996	0.0965	0.0032	0.2222
Factorized $S = AB^\top$	200	0.0992	0.0965	0.0032	0.2222
Dense S	–	**0.0992**	**0.0965**	**0.0032**	**0.2222**

Table 4: End-to-end reconstruction with factorized and dense spatial decoders under the H15/V15 protocol. Results are mean relative RMSE over 80 test fields. "Full" and "Unobs." denote full-field and unobserved-region relative RMSE, respectively.

		Equispaced		Shifted		Jittered	
Decoder	Params	Full ↓	Unobs. ↓	Full ↓	Unobs. ↓	Full ↓	Unobs. ↓
Factorized, $L = 32$	21.30M	0.6523	0.6668	0.6585	0.6752	0.6682	0.6855
Factorized, $L = 64$	23.27M	0.6452	0.6600	0.6535	0.6707	0.6643	0.6823
Dense S	**19.36M**	**0.2561**	**0.2692**	**0.2594**	**0.2744**	**0.2780**	**0.2950**

Increasing the prescribed spatial rank from $L = 32$ to $L = 64$ yields only a small improvement in neural reconstruction, reducing the average unobserved-region RMSE across the three geometries from 0.6758 to 0.6710. In contrast, the dense-S decoder achieves an average unobserved-region RMSE of 0.2795, with the same qualitative advantage under all three sampling geometries.

Notably, this improvement does not result from a larger model: the dense-S configuration contains 19.36M trainable parameters, compared with 21.30M and 23.27M for the $L = 32$ and $L = 64$ factorized decoders, respectively. Together with the oracle representation analysis, these results show that a moderate prescribed spatial rank can create a substantial bottleneck in both representation capacity and end-to-end reconstruction. This supports the use of data-adaptive spatial components in TIDAL while retaining the physics-derived S/c representation.

E Complete Data- and Model-Efficiency Results

This section provides the complete results underlying the data- and model-efficiency analysis in Sec. 6.4. We compare the S/c decoder with a direct-output decoder under different training-set sizes and model widths. Unless otherwise specified, all evaluations use H15/V15 observations and the same 80 test fields under the equispaced, shifted, and jittered sampling geometries.

E.1 Data Efficiency

We first vary the number of training fields while fixing the model width to 384. Both decoder variants use the same encoder, training protocol, normalization, and H/V sampling augmentation. Table 5 reports the complete results for $N \in \{50, 100, 200, 500\}$.

The structured S/c decoder achieves lower full-field and unobserved-region error for every tested training-set size and sampling geometry. The advantage is already present with only 50 training fields and becomes more pronounced with moderate supervision. For example, at $N = 200$, the average unobserved-region RMSE decreases from 0.548 with direct-output decoder to 0.428 with

Table 5: Reconstruction performance under different training-set sizes. Results are mean ± standard deviation over 80 test fields. "Full" and "Unobs." denote full-field and unobserved-region relative RMSE, respectively. "Avg. Unobs." is the average unobserved-region RMSE across the three sampling geometries. Lower is better.

		Equispaced		Shifted		Jittered		
Train fields	Decoder	Full ↓	Unobs. ↓	Full ↓	Unobs. ↓	Full ↓	Unobs. ↓	Avg. Unobs. ↓
50	Direct output	0.906 ± 0.021	0.973 ± 0.023	0.900 ± 0.021	0.972 ± 0.023	0.902 ± 0.020	0.975 ± 0.022	0.973
	S/c	$\mathbf{0.876 \pm 0.028}$	$\mathbf{0.928 \pm 0.026}$	$\mathbf{0.871 \pm 0.025}$	$\mathbf{0.928 \pm 0.022}$	$\mathbf{0.872 \pm 0.026}$	$\mathbf{0.930 \pm 0.024}$	**0.929**
100	Direct output	0.704 ± 0.076	0.754 ± 0.082	0.698 ± 0.074	0.754 ± 0.080	0.714 ± 0.076	0.771 ± 0.082	0.760
	S/c	$\mathbf{0.641 \pm 0.077}$	$\mathbf{0.677 \pm 0.078}$	$\mathbf{0.632 \pm 0.078}$	$\mathbf{0.671 \pm 0.080}$	$\mathbf{0.650 \pm 0.075}$	$\mathbf{0.692 \pm 0.076}$	**0.680**
200	Direct output	0.503 ± 0.096	0.539 ± 0.103	0.500 ± 0.094	0.539 ± 0.102	0.525 ± 0.094	0.566 ± 0.102	0.548
	S/c	$\mathbf{0.396 \pm 0.104}$	$\mathbf{0.418 \pm 0.108}$	$\mathbf{0.395 \pm 0.103}$	$\mathbf{0.419 \pm 0.107}$	$\mathbf{0.419 \pm 0.101}$	$\mathbf{0.446 \pm 0.105}$	**0.428**
500	Direct output	0.292 ± 0.088	0.313 ± 0.094	0.294 ± 0.086	0.317 ± 0.093	0.317 ± 0.089	0.343 ± 0.097	0.324
	S/c	$\mathbf{0.256 \pm 0.083}$	$\mathbf{0.269 \pm 0.085}$	$\mathbf{0.259 \pm 0.082}$	$\mathbf{0.274 \pm 0.085}$	$\mathbf{0.278 \pm 0.085}$	$\mathbf{0.295 \pm 0.089}$	**0.280**

Table 6: Reconstruction performance under different model capacities. Results are mean ± standard deviation over 80 test fields. "Params" denotes the total number of trainable parameters. Lower is better.

			Equispaced		Shifted		Jittered		
Width	Decoder	Params	Full ↓	Unobs. ↓	Full ↓	Unobs. ↓	Full ↓	Unobs. ↓	Avg. Unobs. ↓
128	Direct output	2.17M	0.809 ± 0.062	0.867 ± 0.067	0.808 ± 0.064	0.872 ± 0.069	0.814 ± 0.061	0.879 ± 0.066	0.873
	S/c	2.19M	$\mathbf{0.662 \pm 0.081}$	$\mathbf{0.706 \pm 0.085}$	$\mathbf{0.662 \pm 0.079}$	$\mathbf{0.710 \pm 0.084}$	$\mathbf{0.679 \pm 0.075}$	$\mathbf{0.729 \pm 0.080}$	**0.715**
192	Direct output	4.85M	0.634 ± 0.103	0.680 ± 0.111	0.633 ± 0.101	0.683 ± 0.110	0.652 ± 0.099	0.704 ± 0.107	0.689
	S/c	4.89M	$\mathbf{0.481 \pm 0.106}$	$\mathbf{0.511 \pm 0.111}$	$\mathbf{0.480 \pm 0.103}$	$\mathbf{0.512 \pm 0.110}$	$\mathbf{0.510 \pm 0.105}$	$\mathbf{0.545 \pm 0.111}$	**0.523**
256	Direct output	8.57M	0.453 ± 0.104	0.485 ± 0.112	0.453 ± 0.101	0.489 ± 0.110	0.477 ± 0.103	0.515 ± 0.111	0.497
	S/c	8.65M	$\mathbf{0.360 \pm 0.102}$	$\mathbf{0.380 \pm 0.106}$	$\mathbf{0.363 \pm 0.100}$	$\mathbf{0.386 \pm 0.104}$	$\mathbf{0.384 \pm 0.100}$	$\mathbf{0.409 \pm 0.105}$	**0.391**
384	Direct output	19.20M	0.292 ± 0.088	0.313 ± 0.094	0.294 ± 0.086	0.317 ± 0.093	0.317 ± 0.089	0.343 ± 0.097	0.324
	S/c	19.36M	$\mathbf{0.256 \pm 0.083}$	$\mathbf{0.269 \pm 0.085}$	$\mathbf{0.259 \pm 0.082}$	$\mathbf{0.274 \pm 0.085}$	$\mathbf{0.278 \pm 0.085}$	$\mathbf{0.295 \pm 0.089}$	**0.280**

structured decoding. The same trend is observed independently under equispaced, shifted, and jittered sampling, indicating that the benefit is not specific to a particular H/V geometry.

E.2 Model Efficiency

We next vary the encoder width while using all 500 training fields. The tested widths are $\{128, 192, 256, 384\}$. At each width, the S/c and direct-output models use the same backbone and training protocol, with only a small difference in total parameter count introduced by the corresponding output heads.

The S/c decoder also provides consistent gains across model capacities. The improvement is particularly clear for lightweight and intermediate models: at width 192, where both variants contain approximately 4.9M parameters, the average unobserved-region RMSE decreases from 0.689 to 0.523. Similar improvements are observed at widths 128 and 256, while the advantage remains at width 384. Since the paired models have comparable parameter counts at each width, these results indicate that the improvement is associated with the structured output parameterization rather than simply increased model size.

Together, the data- and model-efficiency results show that the S/c decoder remains beneficial across different supervision and capacity regimes, supporting its role as an effective inductive bias for sparse acoustic-field reconstruction.