HARMOCORE: FUNCTIONAL LATENT DIFFUSION FOR SPARSE RECONSTRUCTION OF OSCILLATORY WAVE FIELDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Reconstructing oscillatory wave fields from scattered sensors is a severely underdetermined inverse problem. Beyond the challenges of general physical-field reconstruction, wave responses are complex-valued, frequency-sensitive, and highly oscillatory, while costly simulation and sensing often leave only extreme-sparse observations. Existing low-rank, operator, and diffusion approaches are largely designed for real-valued, smoother fields; dense pixel-space diffusion is particularly inefficient for oscillatory complex fields and difficult to scale to 3D. We propose **HarmoCore**, which places a generative prior in a compact, continuous, and structured wave-field latent. HarmoCore represents joint real–imaginary channels with Functional Tucker cores over shared continuous spatial bases, learns a frequency-conditioned core diffusion prior, and performs Diffusion Posterior Sampling directly in core space. At fixed sensor coordinates, the multilinear decoder induces an explicit likelihood guidance operator, avoiding dense pixel-space correction. Optional target-equation residual guidance further promotes physical consistency. Experiments on 2D Helmholtz, 2D synthetic wave fields, and 3D Helmholtz show substantial gains under 1%–2% sensing while remaining practical in three dimensions.

1 INTRODUCTION

Reconstructing physical fields from sparse measurements is a fundamental inverse problem across science and engineering Arridge et al. (2019); Manohar et al. (2018). Wave-field reconstruction is particularly important in electromagnetic simulation Colton & Kress (2013), ocean acoustics Jensen et al. (2011), geophysical imaging Virieux & Operto (2009), and many other sensing and modeling problems. We focus on time-harmonic wave fields, which describe the steady-state response to excitation at a fixed frequency. In these applications, high-fidelity simulation or dense sensing can be prohibitively expensive, especially when responses must be acquired across many frequencies. The available training fields and sensor measurements are therefore often severely limited. Recovering a complete wave response from scattered sensors is thus both practically important and profoundly underdetermined.

Existing sparse-field reconstruction methods broadly include low-rank fitting, deep sparse-to-dense regression, neural operators, and diffusion-based generative reconstruction. Low-rank methods exploit compact spatial structure Dolgov et al. (2021); Luo et al. (2024); neural networks and operators learn direct field mappings Fukami et al. (2021); Li et al. (2021); Tran et al. (2023); Lu et al. (2021); Li et al. (2024; 2023); Kovachki et al. (2023); and diffusion priors can be combined with partial observations through posterior sampling Chung et al. (2023); Huang et al. (2024). These paradigms have shown strong results, but most have been developed for real-valued, relatively smooth spatial or spatiotemporal fields. Sparse reconstruction of frequency-sensitive, highly oscillatory complex-valued wave fields remains largely underexplored.

Oscillatory wave fields, however, differ sharply from the smoother data targeted by most existing methods. In the time-harmonic setting, the Helmholtz, frequency-domain Maxwell, and elastodynamic equations define a complex spatial response at each frequency. Its real and imaginary components jointly encode amplitude and phase, while sources, materials, and boundaries create nonlocal

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.

interference. Short wavelengths and sensitivity to frequency or medium parameters produce rapid spatial variation and can shift nodes and antinodes throughout the domain, leaving local sensors weakly informative about unobserved regions. Reconstruction is therefore not merely local interpolation: the model must infer a globally coherent phase pattern from sparse evidence, and small phase errors can alter interference across the domain. This structure mismatches common representations. Fixed low-rank models can suppress frequency-dependent modes; learned regressors and operators require broad training coverage to distinguish phase-sensitive responses; and pixel-space diffusion must model every rapidly varying complex value, making guidance costly and 3D scaling difficult. Generic visual latents remain grid-bound and are not designed for continuous-coordinate spatial queries or coupled complex channels. The central challenge is therefore to build a generative prior aligned with both the oscillatory field structure and its sparse observations.

To address these challenges, we propose **HarmoCore**, a functional latent diffusion framework for sparse complex wave-field reconstruction. We represent the real and imaginary components of each field as joint channels of a compact Functional Tucker core over shared continuous spatial bases. The bases capture continuous coordinate dependence, while the core retains compact, sample- and frequency-specific coefficients. We train a frequency-conditioned diffusion model on these cores and perform Diffusion Posterior Sampling directly in core space. We evaluate the shared bases at the sensor coordinates, so we can use scattered observations without rasterization. We exploit the decoder's multilinearity to write the sensor measurements as a linear operator on the core, which gives a closed-form observation-likelihood gradient. For optional governing-equation guidance, the same decoder provides a fixed core-to-field Jacobian that efficiently propagates the gradient of a possibly nonlinear residual. We thus keep posterior correction in the compact core space instead of the dense pixel space. Experiments on 2D Helmholtz, 2D synthetic wave fields, and 3D Helmholtz demonstrate substantial gains under 1%–2% sensing.

We summarize our contributions as follows:

- We propose a frequency-aware, compact representation of complex oscillatory wave fields based on Functional Tucker models, with joint real–imaginary channels and continuous-coordinate decoding.
- We train a frequency-conditioned diffusion prior in the core space and exploit the multilinear decoder to obtain a closed-form observation-likelihood gradient and efficient governing-equation residual guidance.
- We demonstrate substantial improvements under extreme sparsity across 2D Helmholtz, 2D synthetic wave fields, and 3D Helmholtz reconstruction settings.

2 Preliminaries and Problem Setup

2.1 Sparse Reconstruction of Time-Harmonic Wave Fields

A time-harmonic wave field describes the steady-state complex spatial response to excitation at angular frequency ω. The physical field $\tilde{u}_s(\mathbf{r}, \omega) \in \mathbb{C}^K$ encodes amplitude and phase jointly; real and imaginary parts together determine interference, node locations, and energy distribution throughout the domain. We represent the field through stacked real and imaginary channels and formalize the sparse observation model as

$$\begin{aligned} u_s(\mathbf{r}, \omega) &= \big[\mathrm{Re}(\tilde{u}_s),\, \mathrm{Im}(\tilde{u}_s)\big] \in \mathbb{R}^C, \quad C = 2K, \\ \mathbf{y}_m &= u_s(\mathbf{r}_m, \omega) + \varepsilon_m, \quad \varepsilon_m \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathrm{obs}}^2 I), \end{aligned} \tag{1}$$

where s indexes the field instance (source location, material parameters, boundary configuration), $\mathbf{r}_m \in \Omega$ is the m-th sensor coordinate, and $\mathbf{y}_m \in \mathbb{R}^C$ is the measured channel vector. For scalar wave fields $K = 1$ ($C = 2$, one real and one imaginary channel); in the experiments reported here all three benchmarks use $K = 1$. The sparse observation set $\{(\mathbf{r}_m, \mathbf{y}_m)\}_{m=1}^M$ may cover only a small fraction of the evaluation grid; in our extreme-sparse experiments, for example, the sensing ratio is as low as 1%–2%. The primary reconstruction target is the full continuous field $u_s(\cdot, \omega)$ over Ω, evaluated on both observed and unobserved regions, with the unobserved region being the main indicator of reconstruction quality.

The fields satisfy a governing equation $\mathcal{L}_{\xi_s,\omega}\,\tilde{u}_s = f_{\xi_s,\omega}$, where $\mathcal{L}$ is the differential operator (e.g., the Helmholtz operator $-\Delta - \omega^2/c(\mathbf{r})^2$), ξ_s collects the instance-specific medium parameters, and $f_{\xi_s,\omega}$ is the source term. Operator and source metadata available at test time are dataset-specific and used for optional equation guidance in Section 3.3.

2.2 Functional Tucker Representations

Tucker decomposition approximates a tensor $\mathcal{X} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ with a compact core $G \in \mathbb{R}^{r_1\times\cdots\times r_d}$ and factor matrices $U_k \in \mathbb{R}^{n_k\times r_k}$, $r_k \ll n_k$, reducing the parameter count from $\prod n_k$ to $\prod r_k + \sum n_k r_k$. Functional Tucker Dolgov et al. (2021); Fang et al. (2024) replaces the discrete row-lookup $U_k[\mathbf{i}]$ with a continuous coordinate-evaluable basis function $\phi_k : \mathbb{R} \to \mathbb{R}^{r_k}$:

$$u(\mathbf{r};\, G) = G \times_1 \phi_1(r_1) \times_2 \cdots \times_d \phi_d(r_d), \tag{2}$$

where $\times_k$ denotes mode-k contraction. At any fixed coordinate $\mathbf{r}$, the mapping $G \mapsto u(\mathbf{r};\, G)$ is multilinear—linear in each mode separately. Because ϕ_k can be evaluated at arbitrary real-valued inputs, the decoder supports queries at scattered off-grid sensor locations without rasterization. When a single set of basis functions is shared across a field family and only the core G varies per sample, the model separates common spatial structure (captured in the bases) from sample-specific coefficients (captured in the core).

2.3 Diffusion Priors and Posterior Sampling

A diffusion model places a generative prior $p_\theta(x_0)$ over an unknown variable x_0 by training a denoiser $\epsilon_\theta(x_t, t)$ to reverse a forward noising process $q(x_t \mid x_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\,x_0, (1-\bar{\alpha}_t)I)$. Given a noisy state x_t, the denoiser produces a clean estimate $\hat{x}_{0,t} = (x_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(x_t,t))/\sqrt{\bar{\alpha}_t}$. For an inverse problem with partial observations $\mathbf{y}$ related to x_0 through a forward measurement operator $\mathcal{A}$, the inference target is the posterior $p(x_0 \mid \mathbf{y}) \propto p_\theta(x_0)\, p(\mathbf{y} \mid x_0)$. Diffusion Posterior Sampling (DPS) Chung et al. (2023) approximates sampling from this posterior by correcting each unconditional reverse step with a measurement-consistency gradient evaluated on the clean estimate:

$$x_{t-1} = x'_{t-1} - \zeta_t \nabla_{x_t} \big\|\mathbf{y} - \mathcal{A}(\hat{x}_{0,t})\big\|_2^2, \tag{3}$$

where x'_{t-1} is the unconditional reverse sample obtained from $\hat{x}_{0,t}$ and $\epsilon_\theta(x_t, t)$, and ζ_t is a step-size schedule. The efficiency and accuracy of this guidance depend critically on how cheaply the measurement operator $\mathcal{A}$ and its gradient can be evaluated on the clean estimate $\hat{x}_{0,t}$.

3 Method

HarmoCore reconstructs sparse complex wave fields through two training stages and a guided inference stage. **Training stage 1:** shared continuous spatial basis networks (ϕ_x, ϕ_y) and per-field compact cores $G_{s,\omega}$ are learned jointly from sparse training observations, compressing the field family into a structured latent. **Training stage 2:** a frequency-conditioned diffusion model is trained on the normalized cores, capturing the distribution of valid latent coefficients. **Inference:** the frozen continuous basis networks are evaluated at the test-case sensor coordinates, yielding a precomputed observation operator $H_{\mathcal{O}}$; core-space posterior sampling guided by $H_{\mathcal{O}}$ and an optional governing-equation residual reconstructs the core, which is decoded to the continuous wave field. Figure 1 illustrates the overall pipeline.

3.1 Functional Latent Modeling of Complex Wave Fields

HarmoCore instantiates the Functional Tucker representation of Eq. (2) with a *joint real–imaginary channel core*. For each (sample, frequency) pair (s, ω), the field is encoded by a single core tensor and decoded as

$$\begin{aligned} G_{s,\omega} &\in \mathbb{R}^{R_x\times R_y\times C}, \quad C = 2K, \\ u_s(\mathbf{r};\, G_{s,\omega}) &= G_{s,\omega} \times_1 \phi_x(r_x) \times_2 \phi_y(r_y), \end{aligned} \tag{4}$$

where channels $1, \ldots, K$ are the real parts of the K field components and channels $K+1, \ldots, 2K$ are the imaginary parts. For scalar wave fields $K = 1$ ($C = 2$). In 3D a third basis network

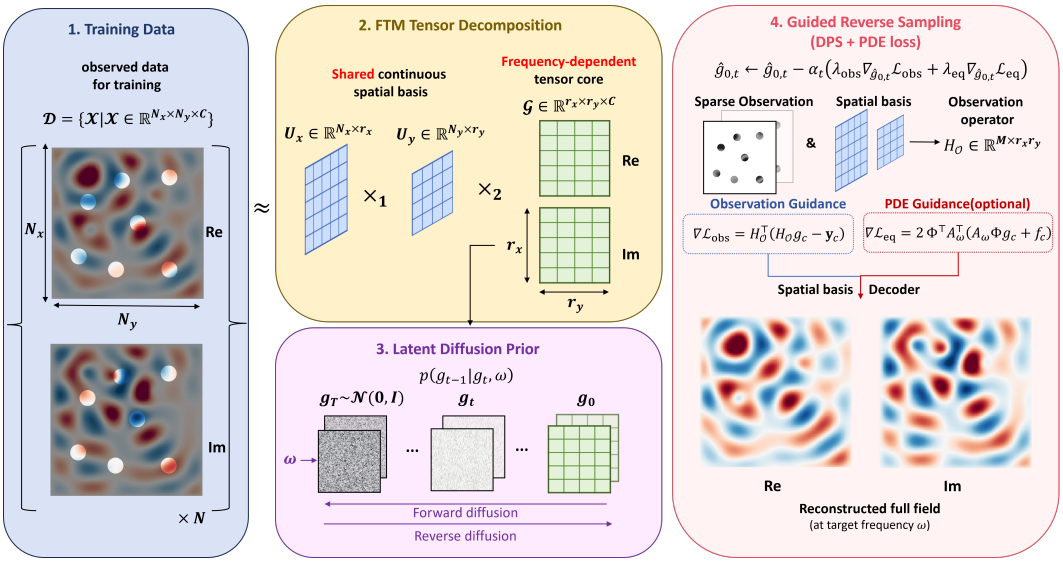


Figure 1: HarmoCore pipeline overview.(2d field as example) *Training* (left and center): sparse training observations are compressed into per-field joint real–imaginary channel cores via shared continuous basis networks; normalized cores are used to train a frequency-conditioned diffusion prior. *Inference* (right): basis evaluation at sensor coordinates yields the precomputed observation operator $H_{\mathcal{O}}$; core-space posterior sampling guided by $H_{\mathcal{O}}$ and an optional equation residual produces the reconstructed complex wave field.

ϕ_z is added and the core lives in $\mathbb{R}^{R_x \times R_y \times R_z \times C}$. A single pair of sine-activated MLP networks (ϕ_x, ϕ_y) Sitzmann et al. (2020) is *shared* across all samples, all frequencies, and all channels; only $G_{s,\omega}$ is sample- and frequency-specific. Each (sample, frequency) pair therefore yields a $C \times R_x \times R_y$ core ($R = R_x R_y$ coefficients per channel), which for the ranks used in our experiments (Section 5.1) is far smaller than a full-resolution field of size $C \times H \times W$.

Joint optimization of bases and cores. We learn (ϕ_x, ϕ_y) and all cores simultaneously by Adam optimization. Let $g_c = \text{vec}(G[\cdot, \cdot, c]) \in \mathbb{R}^R$ denote the per-channel vectorized core and $\Phi = [\phi_x(x_i) \otimes \phi_y(y_j)]_{ij} \in \mathbb{R}^{P \times R}$ ($P = H \times W$) the full-grid basis matrix. Let $\mathcal{O}$ index the observation coordinates used in the training objective below; restricting Φ to the rows indexed by $\mathcal{O}$ gives the observation operator $H_{\mathcal{O}} = \Phi[\mathcal{O}] \in \mathbb{R}^{M \times R}$, so that $H_{\mathcal{O}} g_c$ evaluates the decoded field at those positions against the corresponding measurements $\mathbf{y}_c$. The training objective is

$$\mathcal{L}_{\text{FTM}} = \frac{1}{2} \sum_{c=1}^{C} \mathbb{E}_{s,\omega} \left[\frac{\|H_{\mathcal{O}} g_c - \mathbf{y}_c\|}{\|\mathbf{y}_c\|} \right] + \lambda \mathcal{S}(G), \tag{5}$$

where $\mathcal{O}$ indexes the observed positions and $\mathcal{S}(G)$ is a frequency-weighted spatial-smoothness regularizer on the core matrices. After training, each core is normalized per channel using global mean and standard deviation computed over all (sample, frequency) pairs, yielding zero-mean, unit-variance inputs for the diffusion model.

3.2 Frequency-Conditioned Core Diffusion

The normalized cores are used to train a frequency-conditioned diffusion model. A conditional UNet ϵ_θ takes a noisy core image $g_t \in \mathbb{R}^{C \times R_x \times R_y}$, diffusion timestep t, and normalized frequency $\omega_{\text{norm}} = (\omega - \omega_{\min})/(\omega_{\max} - \omega_{\min}) \in [0, 1]$ injected via FiLM conditioning at each ResBlock. The model is trained with the standard noise-prediction objective Ho et al. (2020)

$$\begin{aligned} \mathcal{L}_{\text{diff}} &= \mathbb{E}_{g_0, \epsilon, t} \left[\|\epsilon - \epsilon_\theta(g_t, t, \omega)\|_2^2 \right], \\ g_t &= \sqrt{\bar{\alpha}_t}\, g_0 + \sqrt{1 - \bar{\alpha}_t}\, \epsilon, \end{aligned} \tag{6}$$

using a variance-preserving linear schedule ($T = 500$ steps, $\beta_1 = 10^{-4}$, $\beta_T = 2 \times 10^{-2}$). This yields the conditional prior $p_\theta(g_0 \mid \omega)$ over normalized joint-channel cores.

3.3 CORE-SPACE POSTERIOR RECONSTRUCTION

Explicit observation operator. At test time, the frozen basis networks are evaluated on the full evaluation grid to build $\Phi \in \mathbb{R}^{P\times R}$ once, and the observation operator $H_{\mathcal{O}} = \Phi[\mathcal{O}]$ is re-instantiated for the test-case sensor coordinates $\mathcal{O}$. The resulting observation loss and its core-space gradient are

$$\begin{aligned} \mathcal{L}_{\text{obs}}(g_c) &= \|H_{\mathcal{O}} g_c - \mathbf{y}_c\|_2^2, \\ \nabla_{g_c}\mathcal{L}_{\text{obs}} &= H_{\mathcal{O}}^{\top}(H_{\mathcal{O}} g_c - \mathbf{y}_c). \end{aligned} \tag{7}$$

Both $H_{\mathcal{O}}$ and the gradient require only matrix–vector products in $\mathbb{R}^R$, and $H_{\mathcal{O}}$ is reused across all reverse steps. The same spatial mask applies to all channels (real and imaginary), so a single $H_{\mathcal{O}}$ serves both.

Core-space sampling update. We run a DDPM-style reverse process with guidance applied directly to the clean estimate at each step:

$$\begin{aligned} \hat{g}_{0,t} &= \frac{g_t - \sqrt{1-\bar{\alpha}_t}\,\hat{\epsilon}_\theta}{\sqrt{\bar{\alpha}_t}}, \quad \hat{\epsilon}_\theta = \epsilon_\theta(g_t, t, \omega), \\ \hat{g}_{0,t} &\leftarrow \hat{g}_{0,t} - \alpha_t\big(\lambda_{\text{obs}}\nabla_{\hat{g}_{0,t}}\mathcal{L}_{\text{obs}} + \lambda_{\text{eq}}\nabla_{\hat{g}_{0,t}}\mathcal{L}_{\text{eq}}\big), \\ g_{t-1} &= \sqrt{\bar{\alpha}_{t-1}}\,\hat{g}_{0,t} + \sqrt{1-\bar{\alpha}_{t-1}}\,\hat{\epsilon}_\theta, \end{aligned} \tag{8}$$

where α_t is a step-specific guidance weight and the noise estimate $\hat{\epsilon}_\theta$ is not recomputed after the guidance correction.

Optional governing-equation guidance. For the Helmholtz benchmarks, the discretized operator A_ω (sparse finite-difference matrix) and source field f are available from dataset metadata at test time. The full-grid decoded field per channel is $\hat{u}_c = \Phi g_c$; the equation residual Raissi et al. (2019) is evaluated on the decoded dense grid via autograd. Because Φ is frozen after FTM training, the decoder Jacobian is constant across all reverse steps and is reused in each guidance evaluation without re-evaluation through the basis networks. For the linear Helmholtz residual $A_\omega\hat{u} + f$, the core-space gradient is $\nabla_{g_c}\mathcal{L}_{\text{eq}} = 2\,\Phi^{\top}A_\omega^{\top}(A_\omega\Phi g_c + f_c)$; in practice we compute this via autograd on the decoded output. The equation term is optional and supportive: ablations in Section 5.3 confirm that removing observation-guided posterior sampling causes a far larger accuracy collapse than removing the equation term alone.

4 RELATED WORK

Reconstructing dense fields from sparse sensors has been studied through sensor-to-dense networks, which pool irregular observations onto a surrogate regular input before prediction Fukami et al. (2021), and neural operators, which learn direct field mappings that generalize across PDE parameters Li et al. (2021); Tran et al. (2023); Lu et al. (2021); Li et al. (2024; 2023); Kovachki et al. (2023). Both perform well when training coverage captures the relevant response patterns, but under extreme sparsity (1%–2% sensing) the oscillatory structure and frequency sensitivity of time-harmonic wave fields make a single deterministic estimate depend heavily on whether training instances cover the specific frequency and boundary configuration. A generative model that captures the distribution of valid field configurations can supply the missing structural constraint when local sensors cannot resolve the global wave pattern alone.

Diffusion models Ho et al. (2020) supply such priors and can be combined with partial observations through posterior sampling Song et al. (2021b); Chung et al. (2023); Kawar et al. (2022); Song et al. (2022; 2023). building on efficient and controllable sampling techniques Song et al. (2021a); Ho & Salimans (2022); Diffusion Posterior Sampling Chung et al. (2023) steers a learned denoiser toward measurement-consistent states during the reverse process, and related work applies this directly in pixel space to PDE-governed field completion Huang et al. (2024). Modeling and guiding every value of the dense spatial field is costly for rapidly varying, channel-coupled complex fields, and grows more expensive as resolution or dimensionality increases—motivating a compact, continuous representation aligned with the real/imaginary channel coupling, in the spirit of latent-space diffusion Rombach et al. (2022).

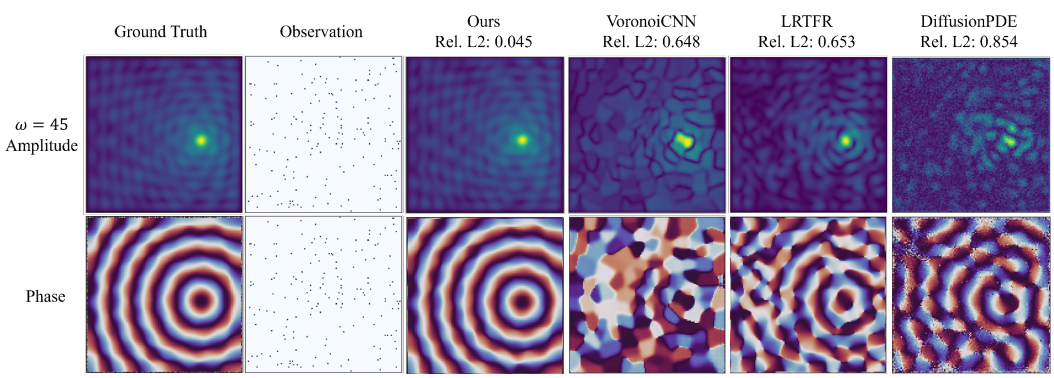


Figure 2: Qualitative comparison on 2D Helmholtz at extreme sparsity (1% sensing). HarmoCore recovers the global oscillatory structure more faithfully than competing baselines at this sensing density.

Tucker and tensor-train methods offer compact multi-mode decompositions of structured fields Kolda & Bader (2009); Oseledets (2011); Dolgov et al. (2021), and Functional Tucker models Fang et al. (2024) replace discrete factor matrices with continuous-coordinate basis functions Tancik et al. (2020); Mildenhall et al. (2020), separating shared spatial variation (the bases) from sample-specific coefficients (the core). Closest to our work, SDIFT Chen et al. (2025) combines Functional Tucker cores with diffusion for spatiotemporal reconstruction, using Gaussian-process noise, message-passing posterior guidance, and the linear observation model induced by fixed spatial bases. HarmoCore extends this functional core modeling approach to time-harmonic complex wave fields by conditioning the core prior on excitation frequency, jointly modeling real and imaginary components, and incorporating optional Helmholtz residual guidance. Our contribution is the resulting reconstruction framework and its empirical evaluation.

5 Experiments

5.1 Experimental Setup

We evaluate on three benchmarks, each isolating a different experimental question. **2D Helmholtz** is the primary benchmark and drives the main sparsity comparison across sensor densities. **2D Synthetic Wave Fields** test whether the same core-space formulation transfers to a physically different wave-field family generated from a closed-form ray model rather than a PDE solve: each field is a direct-plus-reflected superposition of per-source ray terms,

$$\tilde{u}_s(\mathbf{r},\omega) = \sum_{k=1}^{K_s} w_{s,k}\Big[a_1(\mathbf{r},\mathbf{r}_{s,k})\, e^{\,i2\pi f\,\tau_1(\mathbf{r},\mathbf{r}_{s,k})} + \beta\, a_2(\mathbf{r},\mathbf{r}'_{s,k})\, e^{\,i2\pi f\,\tau_2(\mathbf{r},\mathbf{r}'_{s,k})}\Big], \tag{9}$$

with $\omega = 2\pi f$, a_1, a_2 distance-dependent amplitude decays, τ_1, τ_2 the direct/reflected travel times, and $\mathbf{r}'_{s,k}$ each source's mirror-image reflector (full parameterization in Appendix A.1); having no governing equation, this benchmark receives no physics-residual metric or equation guidance (Section 3.3). **3D Helmholtz** reuses the same governing equation on a higher-dimensional domain to test whether the framework extends to 3D via the natural addition of a third shared basis network ϕ_z. All methods on a given benchmark are evaluated on the same held-out test set.

At each sensing ratio $r \in \{1\%, 2\%, 5\%, 10\%\}$, the observation mask is an i.i.d. Bernoulli(r) mask over grid points—each grid location is retained as an observed sensor independently with probability r—generated once per ratio and shared by every method on a given test case. We report 1%, 2%, and 5%, the extreme-to-moderate sparsity range that is the focus of our method.

We compare against three categories of baselines: a continuous low-rank tensor-function representation, LRTFR Luo et al. (2024); deterministic operator-regression networks trained on dense obser-

Table 1: Reconstruction Relative L2 Error and physics residual across three benchmarks (mean±std). Physics residual is a PDE residual under the Helmholtz operator for 2D/3D Helmholtz; the 2D Synthetic benchmark has no governing PDE (fields are generated by a closed-form ray model), so this column is "—" for all methods there.

	1%		2%		5%	
Method	Rel. L2	Phys. Res.	Rel. L2	Phys. Res.	Rel. L2	Phys. Res.
2D Helmholtz						
Ours	**0.130±0.069**	**0.032±0.001**	**0.068±0.039**	**0.032±0.001**	**0.037±0.023**	**0.032±0.001**
LRTFR	0.726±0.104	0.181±0.164	0.505±0.142	0.172±0.168	0.251±0.120	0.138±0.119
FNO	0.822±0.099	1.518±0.906	0.609±0.060	1.494±0.916	0.239±0.031	0.930±0.599
F-FNO	1.413±0.403	1.338±0.556	1.413±0.407	1.841±0.760	1.412±0.418	2.869±1.192
VoronoiCNN	0.638±0.298	0.201±0.048	0.446±0.227	0.186±0.038	0.222±0.123	0.153±0.024
DiffusionPDE	0.653±0.203	1.114±0.326	0.438±0.197	1.013±0.279	0.179±0.118	0.745±0.211
2D Synthetic Wave Fields						
Ours	**0.259±0.106**	—	**0.102±0.049**	—	**0.035±0.013**	—
LRTFR	0.503±0.038	—	0.224±0.017	—	0.170±0.070	—
FNO	1.101±0.034	—	1.091±0.033	—	1.058±0.032	—
F-FNO	1.075±0.049	—	1.070±0.050	—	1.056±0.051	—
VoronoiCNN	0.440±0.200	—	0.278±0.136	—	0.122±0.061	—
DiffusionPDE	0.434±0.169	—	0.194±0.120	—	0.038±0.021	—
3D Helmholtz						
Ours	**0.249±0.208**	**5.85±3.69**	**0.175±0.140**	**5.34±3.15**	**0.141±0.112**	**5.12±2.96**
LRTFR	3.139±2.293	603±1604	2.494±5.407	816±3032	5.575±3.221	710±1333
FNO	1.217±0.125	127±51	1.212±0.125	245±105	1.197±0.123	576±252
F-FNO	1.144±0.095	117±42	1.144±0.098	226±83	1.146±0.106	523±196
VoronoiCNN	0.440±0.162	14.8±8.0	0.291±0.109	11.7±6.5	0.161±0.060	8.47±4.92
DiffusionPDE	0.753±0.192	84.9±37.0	0.511±0.154	81.7±42.3	0.251±0.069	67.6±37.2

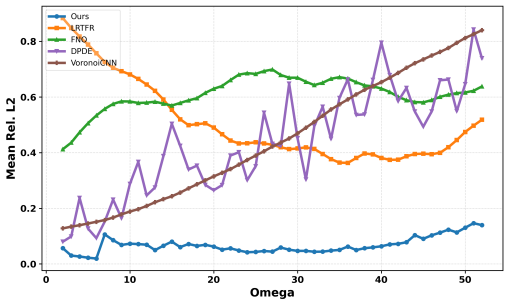
vations, FNO Li et al. (2021), F-FNO Tran et al. (2023), and VoronoiCNN Fukami et al. (2021); and a generative diffusion-based baseline, DiffusionPDE Huang et al. (2024).

The primary metric is the relative reconstruction error (relative ℓ_2 norm, mean±std over the test set), which we refer to throughout as *Relative L2 Error* (*Rel. L2* in tables); it is computed over the *full* evaluation grid including sensor locations, and the primary comparison focuses on 1%–2% sensing, where the inverse problem is most underdetermined. For the two Helmholtz benchmarks we additionally report a physics-residual metric under the discretized Helmholtz operator; the 2D Synthetic benchmark has no governing PDE, so this column is marked "—" for all methods there. Lower values are better for all metrics. Unless otherwise noted, the Functional Tucker core uses ranks $R_x = R_y = 24$ ($R_x = R_y = R_z = 24$ in 3D), yielding $R = 576$ (2D) or $R = 13{,}824$ (3D) coefficients per channel per (sample, frequency) pair (Section 3.1). Details of the experimental setup is given in Appendix A.

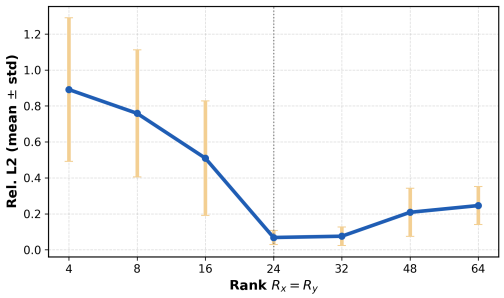
5.2 Main Results across Benchmarks

Table 1 reports reconstruction Rel. L2 error and physics residual across all three benchmarks. HarmoCore achieves the lowest Rel. L2 error and best physics consistency on every benchmark at all three sensing ratios (1%, 2%, and 5%), with the largest margin over the best baseline at 1%–2% sensing, where the inverse problem is least constrained. Figure 2 and Figure 4 show representative qualitative examples. Full qualitative examples are shown in Appendix A.4.

Figure 3a further breaks this down by frequency at 2% sensing: HarmoCore's error stays low and comparatively flat across the tested frequency range, while the baselines fluctuate irregularly with ω rather than following a stable trend, underscoring HarmoCore's comparative robustness across the frequency range.



(a) 2D Helmholtz mean reconstruction Relative L2 Error vs. frequency ω at 2% observation rate. HarmoCore (Ours) stays low and comparatively flat across the frequency range, while baseline error fluctuates irregularly with ω.



(b) Reconstruction Relative L2 Error (mean$\pm$std) vs. Functional Tucker rank on 2D Helmholtz at 2% sensing (102 test cases per rank), with the training configuration held fixed across ranks. The default rank $R_x = R_y = 24$ (dashed line) attains the lowest error under this fixed training budget.

Figure 3: Additional analyses of HarmoCore.

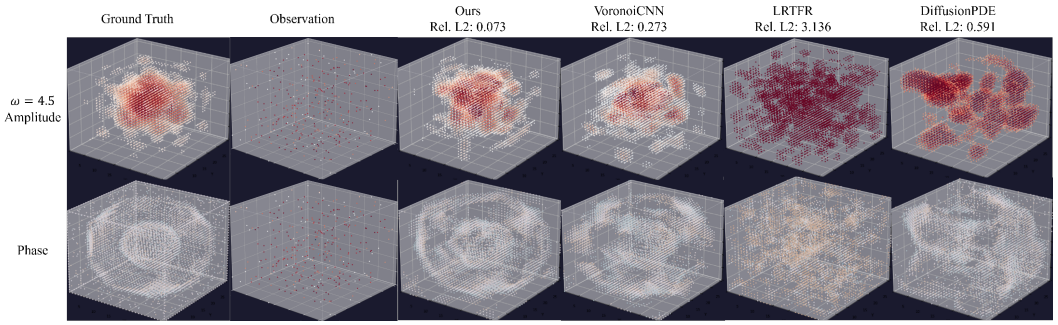


Figure 4: Qualitative 3D Helmholtz reconstructions at 1% sensing , comparing ground truth, HarmoCore, and the strongest baselines. HarmoCore recovers the correct interference and phase structure throughout the volume, while baselines flatten detail or drift in phase away from observed sensors.

A consistent pattern across all three benchmarks is that HarmoCore is most valuable in the truly underdetermined regime: when the sensor ratio is extremely low, sparse observations alone do not constrain the full field, and the learned core-space prior resolves this ambiguity, which is exactly where HarmoCore's margin over every baseline is largest.

5.3 Mechanism Analysis and Ablation

Table 2 reports ablations on 2D Helmholtz. Removing DPS guidance entirely—keeping only the PDE-residual term—causes a large accuracy collapse at both 1% and 2%, confirming that the learned prior and observation-guided posterior sampling are the primary source of reconstruction accuracy. Removing only the equation term produces a smaller but notable degradation, particularly at 1% sensing, which supports the interpretation that equation guidance acts as a supplementary physical regularizer rather than the primary reconstruction mechanism.

The "w/o DPS" variant achieves a lower PDE residual than the full method despite substantially worse reconstruction accuracy. This is expected: that variant optimizes directly toward equation consistency while receiving no constraint from observations. A low PDE residual alone is not sufficient for correct field recovery under extreme sparsity; physical consistency must be interpreted jointly with reconstruction error.

Together, these ablations indicate that the dominant ingredient is the combination of a compact Functional Tucker core with observation-guided DPS: the shared continuous basis reduces the dimensionality of the inverse problem while the diffusion prior and DPS inject the actual observations

Table 2: Ablation on 2D Helmholtz (lower is better).

	1%		2%	
Method	Rel. L2	Phys. Res.	Rel. L2	Phys. Res.
Ours	**0.130**	0.032	**0.068**	0.032
w/o DPS	0.960	**0.027**	0.921	**0.028**
w/o PDE	0.431	0.058	0.204	0.044

Table 3: Reconstruction Rel. L2 error on the 2D Synthetic OOD test set (shifted wave-speed range, no retraining).

Method	1%	2%	5%
Ours	**0.254**	**0.097**	**0.033**
LRTFR	0.496	0.206	0.174
FNO	1.119	1.110	1.083
F-FNO	1.087	1.083	1.069
VoronoiCNN	0.455	0.307	0.154
DiffusionPDE	0.542	0.326	0.140

into the posterior reconstruction, whereas the PDE-residual term acts mainly as a supporting constraint that improves physical consistency without being the main reason the method works.

5.4 Additional Analyses

Representation capacity. Figure 3b sweeps the Functional Tucker rank $R_x = R_y \in \{4, 8, 16, 24, 32, 48, 64\}$ on 2D Helmholtz at 2% sensing, retraining the diffusion prior at each rank with the same architecture and training budget. Reconstruction error is non-monotonic in rank: it falls sharply from R=4 (0.891 ± 0.400) to R=16 (0.509 ± 0.319), reaches its minimum at the default R=24 (0.068 ± 0.039, matching Table 1), stays close at R=32 (0.075 ± 0.052), and rises again at R=64 (0.246 ± 0.106). Since a larger core is at least as expressive as a smaller one, this reflects a fixed-budget diffusion prior becoming increasingly under-provisioned as the core grows, rather than an intrinsic capacity ceiling—we read R=24 as the best operating point under the current, rank-independent training budget.

Robustness to distribution shift. To test robustness to a shift in the underlying generative distribution at test time, we evaluate all methods, without retraining, on an out-of-distribution (OOD) variant of the 2D Synthetic benchmark that redraws the per-sample wave speed from a substantially wider range (full construction in Appendix A.2). Table 3 reports Rel. L2 error under this shift: HarmoCore's error is essentially unchanged from its in-distribution values and LRTFR is similarly robust, while the remaining baselines degrade substantially, most sharply DiffusionPDE.

6 Conclusion

We presented HarmoCore, a latent generative framework for sparse wave-field reconstruction representing complex fields as a compact Functional Tucker core over shared continuous spatial bases and performs diffusion posterior sampling directly in this core space rather than dense pixel space. Across 2D Helmholtz, 2D synthetic wave fields, and 3D Helmholtz, this formulation is most effective in the most underdetermined regimes, where deterministic reconstruction and prior-free low-rank fitting fail, while maintaining better physical consistency than dense operator and pixel-space generative baselines. The method relies on globally learned spatial parameterization and on first obtaining a sufficiently expressive FTM representation, and the paper offers stronger evidence for sparse reconstruction than for frequency extrapolation or uncertainty calibration; extending the core-space prior along these directions is left to future work.

AI USE STATEMENT

In this work, we used generative AI tools only as general-purpose writing and coding assistants, specifically to polish the exposition and to help scaffold and refactor experimental and plotting code. We did not use generative AI tools to generate research ideas, derive the proposed method, or produce experimental results. Other disclosure categories are not applicable to this work.

All AI-assisted text and code were reviewed, edited, and verified by the authors. In particular, AI-assisted code was checked and tested for correctness, and all AI-assisted writing was manually reviewed for technical accuracy and consistency with the authors' intended claims. We take full responsibility for the final content of this work, including all text, claims, code, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work studies physical field reconstruction using synthetic data generated from numerical PDE simulations. It does not involve human subjects, personally identifiable information, sensitive data, or user studies. The proposed method is intended for scientific computing and inverse problems, and we do not identify any immediate ethical concerns arising from the experiments presented in this work. The authors take full responsibility for the integrity and reporting of the results.

REPRODUCIBILITY STATEMENT

We provide the methodological details, model architecture, training and inference procedures, and experimental settings required to reproduce our results in the main paper and appendix. The appendix further specifies the PDE benchmarks, data generation procedures, observation settings, hyperparameters, and ablation configurations. Implementation details for the Functional Tucker representation, diffusion prior, and posterior-guided reconstruction are also documented. Source code and experiment scripts will be included in the supplementary material to facilitate reproduction of all reported results.

REFERENCES

Simon Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb. Solving inverse problems using data-driven models. *Acta Numerica*, 28:1–174, 2019. doi: 10.1017/S0962492919000059.

Panqi Chen, Yifan Sun, et al. Generating full-field evolution of physical dynamics from irregular sparse observations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. arXiv:2505.09284.

Hyungjin Chung, Jeongsol Kim, Michael T. Mccann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023.

David Colton and Rainer Kress. *Inverse Acoustic and Electromagnetic Scattering Theory*. Springer, 3rd edition, 2013.

Sergey Dolgov, Daniel Kressner, and Christoph Strössner. Functional Tucker approximation using Chebyshev interpolation. *SIAM Journal on Scientific Computing*, 43(3):A2190–A2210, 2021. doi: 10.1137/20M1356944. arXiv:2007.16126. .

Shikai Fang, Xin Yu, Zheng Wang, Shibo Li, Robert M. Kirby, and Shandian Zhe. Functional Bayesian Tucker decomposition for continuous-indexed tensor data. In *International Conference on Learning Representations (ICLR)*, 2024. arXiv:2311.04829.

Kai Fukami, Romit Maulik, Nesar Ramachandra, Koji Fukagata, and Kunihiko Taira. Global field reconstruction from sparse sensors with Voronoi tessellation-assisted deep learning. *Nature Machine Intelligence*, 3(11):945–951, 2021. doi: 10.1038/s42256-021-00402-2. arXiv:2101.00554; code: github.com/kfukami/Voronoi-CNN.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint*, 2022. arXiv:2207.12598.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.

Jiahe Huang, Guandao Yang, Zichen Wang, and Jeong Joon Park. DiffusionPDE: Generative pde-solving under partial observation. In *Advances in Neural Information Processing Systems*, 2024. arXiv:2406.17763.

Finn B. Jensen, William A. Kuperman, Michael B. Porter, and Henrik Schmidt. *Computational Ocean Acoustics*. Springer, 2nd edition, 2011.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, 2022. arXiv:2201.11793.

Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51 (3):455–500, 2009. doi: 10.1137/07070111X.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces. *Journal of Machine Learning Research*, 24, 2023. arXiv:2108.08481.

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, 2023. arXiv:2207.05209.

Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM/JMS Journal of Data Science*, 2024. doi: 10.1145/3648506. arXiv:2111.03794.

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(3):218–229, 2021. doi: 10.1038/s42256-021-00302-5.

Yisi Luo, Xile Zhao, Zhemin Li, Michael K. Ng, and Deyu Meng. Low-rank tensor function representation for multi-dimensional data recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5):3351–3369, 2024. doi: 10.1109/TPAMI.2023.3341688.

Krithika Manohar, Bingni W. Brunton, J. Nathan Kutz, and Steven L. Brunton. Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. *IEEE Control Systems Magazine*, 2018. doi: 10.1109/MCS.2018.2810460. arXiv:1701.07569.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision*, 2020.

Ivan V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011. doi: 10.1137/090752286.

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. doi: 10.1016/j.jcp.2018.10.045.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems*, 2020.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021b.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations*, 2022. arXiv:2111.08005.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, 2020.

Alasdair Tran, Alexander G. de G. Mathews, Lexing Xie, and Cheng Soon Ong. Factorized fourier neural operators. In *International Conference on Learning Representations*, 2023.

Jean Virieux and Stéphane Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26, 2009. doi: 10.1190/1.3238367.

A Implementation Details

This appendix collects implementation details that are omitted from the main paper for space reasons.

A.1 Datasets and Preprocessing

All three benchmarks (2D Helmholtz, 2D synthetic wave fields, 3D Helmholtz; Section 5.1) use scalar complex fields ($K = 1$, $C = 2$: one real and one imaginary channel per (sample, frequency) pair; Eq. equation 1).

All methods in Table 1 are evaluated on the same held-out test set per benchmark: 153 held-out fields for 2D Helmholtz, 170 held-out cases (10 samples $\times$ 17 frequencies) for 2D Synthetic, and 410 held-out cases (10 samples $\times$ 41 frequencies) for 3D Helmholtz.

Sensor ratios of 1%, 2%, and 5% of the evaluation grid are used in the main paper; the 10% ratio is reported in Appendix A.1.

After FTM training, each core is normalized per channel using the global mean and standard deviation computed over all (sample, frequency) pairs (Section 3.1), producing zero-mean, unit-variance inputs for diffusion training.

2D and 3D Helmholtz generation. Both Helmholtz benchmarks numerically solve the PML-damped Helmholtz equation

$$\Delta \tilde{u}_s(\mathbf{r}, \omega) + (\omega/c)^2 \, \tilde{u}_s(\mathbf{r}, \omega) = -f_s(\mathbf{r}, \omega), \tag{10}$$

with $L = 1$, $c = 1$, $\mathbf{r} \in \Omega = [0, L]^d$, $d \in \{2, 3\}$, discretized by second-order finite differences on a uniform 128^2 grid (2D) or 32^3 grid (3D). Δ is realized as a complex-coordinate-stretched Laplacian $\sum_{k=1}^{d} \partial_{r_k}[\, s_k(r_k)\, \partial_{r_k}\,]$ that implements a quadratic-profile PML absorbing layer of width ηL on every face ($\eta = 0.12$ in 2D, 0.15 in 3D):

$$\sigma(r_k) = \sigma_{\max}\Big(\tfrac{\eta L - \mathrm{dist}(r_k, \partial\Omega)}{\eta L}\Big)^2, \tag{11}$$

with $\sigma_{\max} = 50$ (2D) or 40 (3D); $\tilde{u}_s = 0$ is imposed on the outermost grid rows (Dirichlet). The source term is a sum of K_s random Gaussian point sources with unit-magnitude, random-phase complex amplitudes,

$$f_s(\mathbf{r}, \omega) = \sum_{k=1}^{K_s} e^{i\phi_{s,k}} \exp\!\big(-\|\mathbf{r} - \mathbf{r}_{s,k}\|^2 / 2\sigma_{\text{src}}^2\big), \tag{12}$$

with $\phi_{s,k} \sim \mathcal{U}(0, 2\pi)$, $K_s \in \{1, \ldots, 4\}$ (2D, $\sigma_{\text{src}} = 0.025$) or $K_s \in \{1, 2, 3\}$ (3D), source positions $\mathbf{r}_{s,k}$ drawn uniformly inside Ω away from the PML layer, and held fixed across every frequency of sample s (only ω changes the linear system, so $(\mathbf{r}_{s,1:K_s}, \phi_{s,1:K_s})$ is the per-sample instance descriptor ξ_s in $\mathcal{L}_{\xi_s,\omega}$, Section 2.1). Frequencies are sampled on a linear grid: 2D Helmholtz uses $\omega \in [2, 52]$ and 3D Helmholtz uses $\omega \in [2, 22]$. Each complex solution is split into real/imaginary channels and the whole dataset is divided by a single global scale (the maximum absolute value over all samples, frequencies, and grid points) before FTM fitting.

2D synthetic wave-field generation. The synthetic benchmark replaces the PDE solve with a closed-form direct-plus-reflected ray superposition on the same 128×128 grid over $\Omega = [0, 1]^2$. Sample s draws a fixed wave speed $v_s \sim \mathcal{U}(0.8, 1.2)$ and $K_s \in \{1, 2, 3\}$ source positions $\mathbf{r}_{s,k}$ (uniform in Ω), each held fixed across that sample's frequency sweep; every source has a mirror-image reflector $\mathbf{r}'_{s,k}$ across the boundary $r_2 = 0$ and a random weight $w_{s,k} \sim \mathcal{U}(0.8, 1.2)$. The field at frequency $\omega = 2\pi f$ is given by Eq. equation 9 (Section 5.1), where $\beta = 0.18$ scales the reflected term; the direct/reflected travel times are $\tau_1 = (r_1/v_s)\big(1 + \varepsilon\,\Phi_1(\mathbf{r})\big)$ and $\tau_2 = \big((r_2 + \delta)/v_s\big)\big(1 + \varepsilon\,\Phi_2(\mathbf{r})\big)$, with r_1, r_2 the distances from $\mathbf{r}$ to the source and its reflector, $\delta = 0.35$ a fixed delay bias, $\varepsilon = 0.08$ a phase-perturbation strength, and Φ_1, Φ_2 fixed smooth sinusoidal fields (in $\sin, \cos$ of r_1, r_2 over the domain extent) shared by every source in a sample; the amplitude decays are $a_1 = e^{-\alpha_1 r_1}/(r_1 + r_0)^p$ and $a_2 = e^{-\alpha_2 r_2}/\sqrt{r_2}$, with $\alpha_1 = 0$, $\alpha_2 = 0.12$, $r_0 = 0.4$, $p = 0.3$. In the unperturbed limit ($\varepsilon \to 0$) the direct-wave phase satisfies the eikonal relation $\|\nabla\theta\| = 2\pi f/v_s$; because the field is produced by this closed-form summation rather than a governing-equation solve, no PDE residual is available for this benchmark (Section 3.3). The evaluation frequency grid is 17 linearly spaced points in $[1, 5]$ Hz; data are globally rescaled by the maximum absolute value, as in the Helmholtz benchmarks.

A.2 Functional Tucker Representation

The shared spatial bases (ϕ_x, ϕ_y) (and ϕ_z in 3D) are sine-activated SIREN MLPs Sitzmann et al. (2020), one pair (triple in 3D) shared across all samples, frequencies, and channels (Section 3.1). Default ranks are $R_x = R_y = 24$, giving $R = R_x R_y = 576$ coefficients per channel per (sample, frequency) pair, versus a full-resolution field of size $C \times H \times W$ (Eq. equation 4). Bases and all per-(sample, frequency) cores are optimized jointly with Adam, using the objective in Eq. equation 5: a relative reconstruction loss on the observed positions plus a frequency-weighted spatial-smoothness regularizer $\mathcal{S}(G)$ (a spatial-gradient penalty weighted by normalized frequency). Both the basis networks and the cores use learning rate 10^{-4}, with a batch size of 64 samples over 25,000 training iterations; the smoothness regularizer weight is $\lambda = 10^5$ in Eq. equation 5. Each SIREN basis network has 4 hidden layers of width 512.

A.3 Latent Diffusion Model

The diffusion prior is a conditional UNet ϵ_θ operating on the normalized joint-channel core image $g_t \in \mathbb{R}^{C \times R_x \times R_y}$ (Section 3.2). Frequency conditioning uses the normalized scalar $\omega_{\text{norm}} \in [0, 1]$ injected via FiLM at each ResBlock. Training uses the noise-prediction objective (Eq. equation 6) under a variance-preserving linear schedule with $T = 500$ steps, $\beta_1 = 10^{-4}$, $\beta_T = 2 \times 10^{-2}$. Optimization uses AdamW with learning rate 10^{-4}, weight decay 10^{-6}, and a batch size of 32, trained for 500 epochs; no EMA of model weights is used. We additionally tried richer frequency encodings (Fourier features) during development; these did not yield consistent gains over the scalar-FiLM default and are treated as a negative result (Appendix A.3).

A.4 Posterior Sampling and Guidance

At inference, the frozen bases are evaluated once at the test-case sensor coordinates to build the observation operator $H_{\mathcal{O}}$ (Eq. equation 7), which is reused across all reverse steps. Guidance is applied directly to the clean estimate $\hat{g}_{0,t}$ at each reverse step (Eq. equation 8), combining the observation-likelihood gradient (weight λ_{obs}) and, optionally, a governing-equation residual gradient (weight λ_{eq}); the step-specific scale α_t multiplies the combined correction. Observations are treated as noiseless ($\sigma_{\text{obs}} = 0$) in all experiments, so λ_{obs} absorbs the likelihood scaling. For the Helmholtz benchmarks, the equation term uses the closed-form residual gradient given in Section 3.3. The 2D Synthetic benchmark has no governing PDE, so no equation-guidance term is applied there ($\lambda_{\text{eq}} = 0$); reconstruction on that benchmark uses the observation term only.

A.5 Evaluation Metrics

For a predicted field $\hat{u}$ and ground truth u (channel-stacked real/imaginary parts, size $C \times H \times W[\times D]$), the reconstruction error reported throughout the main tables is the relative ℓ_2 error over the full evaluation grid, which we refer to as *Relative L2 Error* (*Rel. L2*),

$$\text{Rel. L2} = \frac{\|\hat{u} - u\|_2}{\|u\|_2}, \tag{13}$$

computed jointly over all channels and grid points—including sensor locations—then aggregated as mean±std over the test set; this full-field definition is used consistently for HarmoCore and every baseline.

For the two Helmholtz benchmarks, the physics-residual metric evaluates the discretized governing operator A_ω (Eq. equation 10) against the predicted field and known source f on interior grid points $\mathcal{I}$ (the domain with the outermost boundary row/column excluded):

$$\text{PhysRes} = \sqrt{\frac{\text{mean}_{\mathcal{I}}\big(|A_\omega \hat{u} + f|^2\big)}{\text{mean}_{\mathcal{I}}\big(|f|^2\big)}}. \tag{14}$$

A.6 Baseline Settings

All learned baselines are trained on the same $80\%/20\%$ train/validation split of the dense training set for each benchmark (Appendix A.1) and evaluated on the same held-out test set as HarmoCore.

FNO Li et al. (2021): 4 Fourier layers, width 64, 12 Fourier modes per spatial dimension, trained for 200 epochs (Adam, learning rate 10^{-3}, batch size 64).

F-FNO Tran et al. (2023): 4 factorized Fourier layers, width 64, 12 modes per spatial dimension, trained for 200 epochs (Adam, learning rate 10^{-3}, batch size 64).

VoronoiCNN Fukami et al. (2021): convolutional encoder–decoder, base width 64, trained for 200 epochs (Adam, learning rate 10^{-3}, batch size 32).

DiffusionPDE Huang et al. (2024): conditional UNet (base width 64, conditioning dimension 256) trained with the same $T = 500$-step variance-preserving schedule as HarmoCore's diffusion prior (Appendix A.3), for 200 epochs (Adam, learning rate 10^{-4}, batch size 32); inference uses DPS-style guidance with weight $\zeta = 0.3$.

LRTFR Luo et al. (2024): at test time, per-case core coefficients are recovered by a linear least-squares fit of the observed positions against a fixed spatial basis, then decoded to the full grid.

A Additional Results and Ablations

This appendix collects supporting quantitative and qualitative results that complement the main paper.

Table 4: Reconstruction Relative L2 Error and physics residual at 10% sensing (mean±std), complementing Table 1.

Method	Rel. L2	Phys. Res.
2D Helmholtz		
Ours	**0.012±0.003**	**0.032±0.001**
LRTFR	0.177±0.104	0.116±0.108
FNO	0.062±0.035	0.161±0.035
F-FNO	1.408±0.437	4.012±1.680
VoronoiCNN	0.116±0.064	0.131±0.017
DiffusionPDE	0.101±0.066	0.594±0.170
2D Synthetic Wave Fields		
Ours	**0.025±0.010**	—
LRTFR	0.009±0.003	—
FNO	1.010±0.033	—
F-FNO	1.031±0.053	—
VoronoiCNN	0.064±0.030	—
DiffusionPDE	0.031±0.012	—
3D Helmholtz		
Ours	**0.131±0.105**	**5.001±2.854**
LRTFR	0.499±0.473	372±830
FNO	1.170±0.121	1080±480
F-FNO	1.147±0.118	968±364
VoronoiCNN	0.138±0.038	6.496±3.642
DiffusionPDE	0.161±0.033	60.56±33.80

Table 5: Reconstruction Rel. L2 error (mean±std) on the 2D Synthetic OOD test set (shifted wavespeed range, no retraining), complementing the condensed Table 3.

Method	1%	2%	5%	10%
Ours	**0.254±0.107**	**0.097±0.052**	**0.033±0.014**	**0.025±0.011**
LRTFR	0.496±0.034	0.206±0.024	0.174±0.096	0.112±0.037
FNO	1.119±0.040	1.110±0.039	1.083±0.037	1.043±0.035
F-FNO	1.087±0.062	1.083±0.062	1.069±0.064	1.045±0.067
VoronoiCNN	0.455±0.305	0.307±0.248	0.154±0.161	0.091±0.109
DiffusionPDE	0.542±0.262	0.326±0.290	0.140±0.262	0.093±0.215

A.1 Full Quantitative Tables

Table 4 extends Table 1 to 10% sensing, the ratio dropped from the main-paper table for width (Section 5.1). The 2D Synthetic benchmark has no governing PDE at any sensing ratio (Section 3.3), so its Phys. Res. column is fixed at "—" for all methods, matching Table 1.

A.2 Out-of-Distribution Robustness (2D Synthetic)

To probe sensitivity to a shift in the underlying generative distribution at test time, we build a second 2D Synthetic test set in which the per-sample wave speed v_s is redrawn from a substantially wider range than the in-distribution setting used everywhere else in the paper ($v_s \sim \mathcal{U}(0.8, 1.2)$, empirically $v_s \in [0.84, 1.19]$ across its held-out samples; Appendix A.1): the out-of-distribution (OOD) test set instead draws v_s from a shifted, wider range (empirically $v_s \in [0.44, 1.76]$ across its 10 samples), with every other generative factor—source count, source/reflector positions and weights, the reflection coefficient β, and the evaluation frequency grid—held fixed via the same random seed. No model is retrained: every method uses the same checkpoint evaluated in Table 1, applied unchanged to this shifted test set. Results are reported in Table 3 and discussed in Section 5.4.

Table 6: Reconstruction Rel. L2 error on 2D Helmholtz at 10% sensing under different frequency-conditioning encodings for the diffusion prior.

Frequency conditioning	Rel. L2
Fourier encoding (Eq. 15)	0.049
Fourier + ω_{norm} + ω_{norm}^2	0.044
ω_{norm} + ω_{norm}^2 (no Fourier)	0.044
ω_{norm} + ω_{norm}^2 + linear ($k\pi\omega_{\text{norm}}$)	0.052
ω_{norm} **only (paper default)**	**0.030**

Table 5 reports the same comparison in full (mean$\pm$std, all four sensing ratios), complementing the condensed version in Table 3, which omits standard deviations and 10% sensing for space. Every baseline with a dedicated OOD evaluation shows increased error under the shift, most sharply for DiffusionPDE (e.g. $0.038 \rightarrow 0.140$ at 5% sensing, more than a $3\times$ increase) and, to a lesser degree, VoronoiCNN, FNO, and F-FNO—consistent with these methods relying on a mapping fit to the in-distribution wave-speed range. Ours is flat to marginally lower than its in-distribution values at every sensing ratio (e.g. $0.035 \rightarrow 0.033$ at 5% sensing).

A.3 Frequency-Conditioning Encoding for the Diffusion Prior: A Negative Result

The diffusion prior conditions on frequency through the normalized scalar $\omega_{\text{norm}} \in [0, 1]$ injected via FiLM at each ResBlock (Appendix A.3). During development we also tried replacing this scalar with a richer Fourier-style encoding of ω_{norm},

$$\gamma(\omega_{\text{norm}}) = \left\{\sin(k\pi\,\omega_{\text{norm}}),\ \cos(k\pi\,\omega_{\text{norm}})\right\}_{k=0}^{K-1}, \tag{15}$$

with $K = 8$ frequency bands, concatenated and fed through the same FiLM conditioning path in place of the scalar default.

Table 6 compares this Fourier encoding against several variants—augmenting it with explicit low-order polynomial terms, dropping the Fourier features entirely in favor of the polynomial terms alone, further adding a linear spectral positional term, and the scalar-only default—on 2D Helmholtz at 10% sensing. None of the richer encodings improves over the plain scalar conditioning: the scalar-only variant attains the lowest error (0.030), while every richer variant is worse, with no consistent ordering among them—adding a linear spectral term on top of the polynomial encoding gives the worst result tested (0.052), Fourier encoding alone is only slightly better (0.049), and augmenting Fourier features with explicit polynomial terms or using the polynomial terms alone both land at an intermediate 0.044. We treat this as a negative result: for this benchmark, more elaborate frequency encodings for the diffusion prior's conditioning do not translate into improved reconstruction accuracy, consistent with the summary in Appendix A.3.

A.4 Additional Qualitative Comparisons

Qualitative 3D Helmholtz reconstructions at 1% sensing are shown in Figure 4 in Section 5.2; Figure 5 and Figure 6 below extend both Helmholtz benchmarks to additional test cases. Figure 7 shows qualitative 2D Synthetic wave-field reconstructions, complementing the quantitative results in Section 5.2; the main paper itself has no qualitative figure for this benchmark.

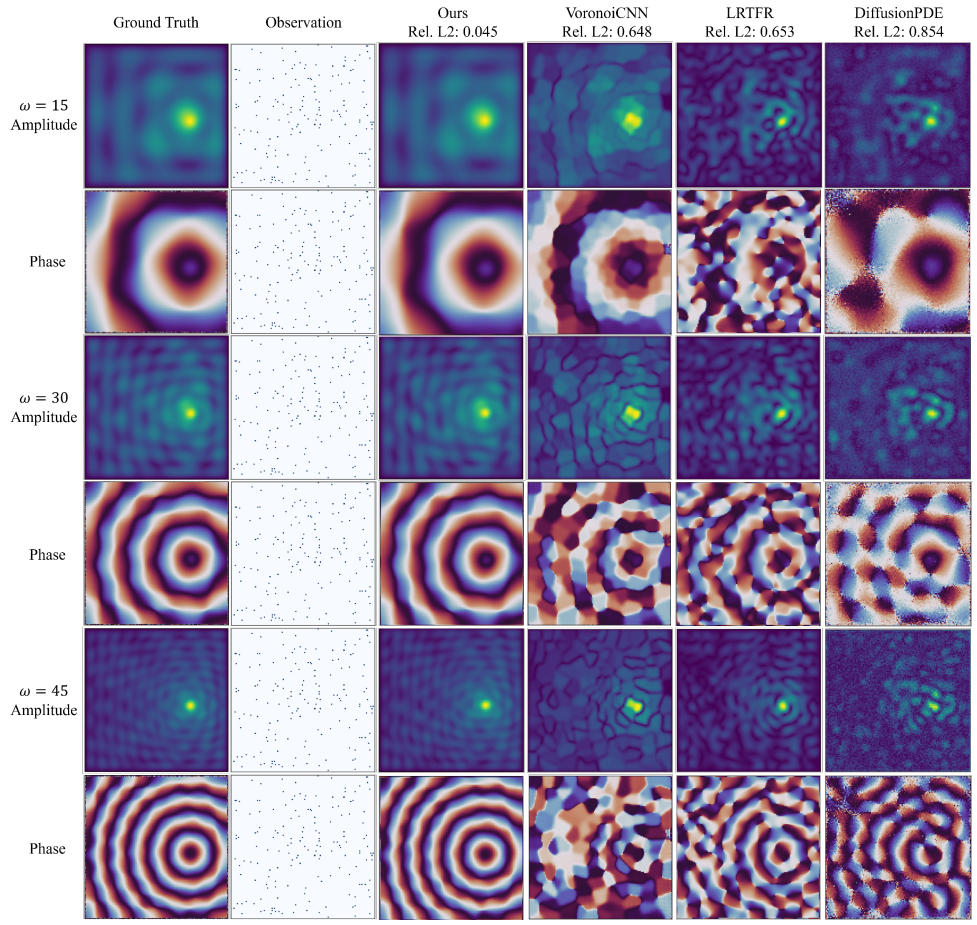


Figure 5: Additional qualitative 2D Helmholtz reconstructions across multiple test cases, with real and imaginary channels shown as separate rows per case, complementing the single-case comparison in Figure 2.

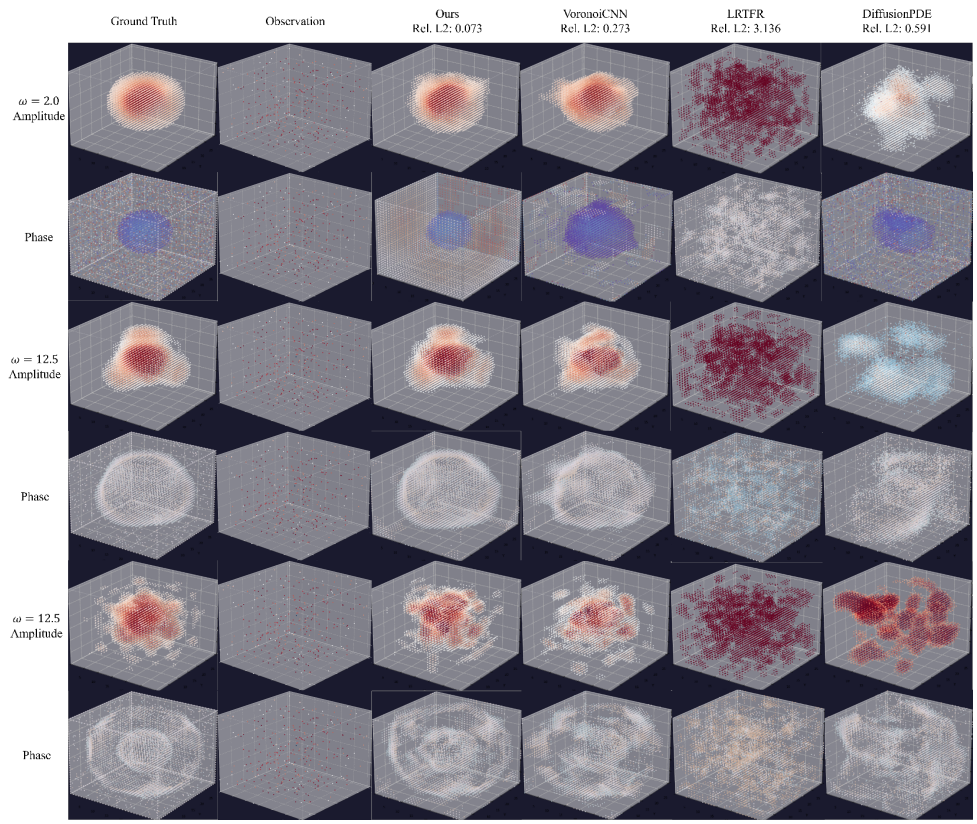


Figure 6: Additional qualitative 3D Helmholtz reconstructions across multiple test cases, with real and imaginary channels shown as separate rows per case, complementing the single-case comparison in Figure 4.

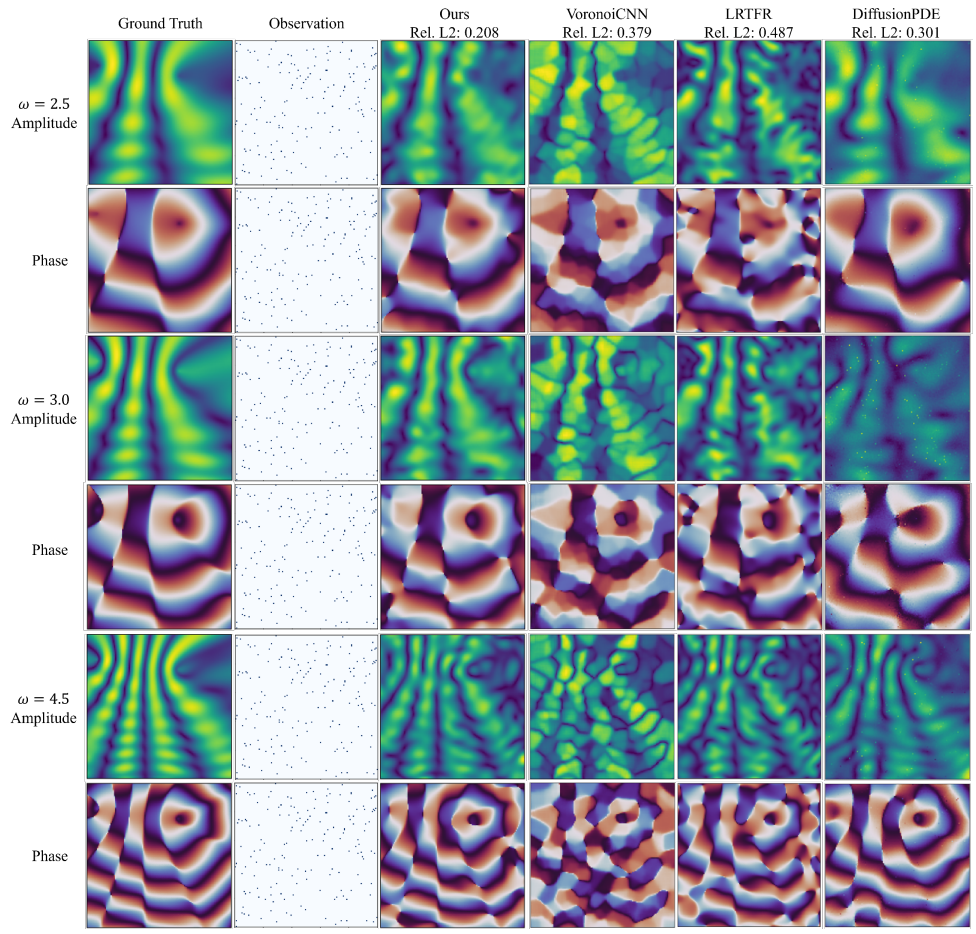


Figure 7: Qualitative 2D Synthetic wave-field reconstructions across three test cases, with real and imaginary channels shown as separate rows per case.