Neural Acoustic Diffraction Tomography: Cycle-Consistent Geometry Reconstruction from 2D BEM Data

Anonymous Author(s)
Submitted for blind review

***Abstract*—We present a neural framework for 2D acoustic diffraction tomography that reconstructs scene geometry from boundary element method (BEM) simulations. Our approach introduces a *transfer function* formulation that learns the scattered-to-incident pressure ratio, reducing reconstruction error from 48% (vanilla MLP on raw pressure) to 4.47% across 15 synthetic scenes while providing $2{,}000\times$ speedup over BEM inference. An *auto-decoder* inverse model maps acoustic observations to signed distance functions (SDF) with Eikonal regularization, yielding 0.91 ± 0.01 mean intersection-over-union (IoU), validated through cycle-consistent forward–inverse agreement (Pearson $r = 0.91$). We further analyze why Helmholtz PDE enforcement fails for neural surrogates: the autodiff Laplacian $\nabla^2 f_\theta$ correlates with the physical Laplacian at only $r = 0.19$ (3.5% variance explained), because random Fourier features with bandwidth $\sigma = 30$ amplify second derivatives by $4\pi^2\sigma^2 \approx 35{,}000\times$.**

***Index Terms*—acoustic diffraction, neural surrogate, signed distance function, inverse scattering, cycle-consistency, boundary element method**

I. Introduction

Reconstructing the geometry of a scene from acoustic measurements is a fundamental problem in computational acoustics with applications to room modeling [1], sonar imaging [2], and augmented reality [3]. Classical approaches rely on iterative optimization against physics-based solvers [2], which is computationally expensive and sensitive to initialization. Recent neural approaches learn implicit representations of acoustic fields [4], [5] or jointly model audio and geometry [6], but model the *total* pressure without recovering scene geometry. Acoustic NLOS imaging [7] reconstructs hidden objects via time-of-flight backprojection with known relay surfaces, and EchoScan [8] infers enclosed room geometry from microphone-array RIRs. Vlašić et al. [9] represent obstacles as SDF zero-level-sets for inverse scattering but rely on a classical boundary integral solver at each iteration. Our work replaces the PDE solver with a learned transfer function surrogate ($2{,}000\times$ speedup) and adds cycle-consistency through the neural forward model for geometry validation. Physics-informed neural networks [10] offer PDE supervision, but as we show, Helmholtz enforcement fails when applied through neural surrogates with high-bandwidth Fourier features [11].

We propose a two-stage neural framework for 2D acoustic diffraction tomography (Fig. 1). In the *forward* stage, we learn a transfer function $T = p_{\text{scat}}/p_{\text{inc}}$ that captures only the scattering component, removing the dominant free-space phase oscillation. This reformulation substantially reduces target complexity: across our dataset, a per-scene constant prediction of T explains 89.6% of variance (vs. 13% for raw p_{scat}), as the transfer function removes the dominant phase oscillation spanning 12 cycles over 2–8 kHz. This enables a compact MLP to approximate BEM-quality fields. In the *inverse* stage, an auto-decoder [12] optimizes per-scene latent codes into an SDF decoder with Eikonal regularization, and a frozen copy of the forward model provides cycle-consistency supervision.

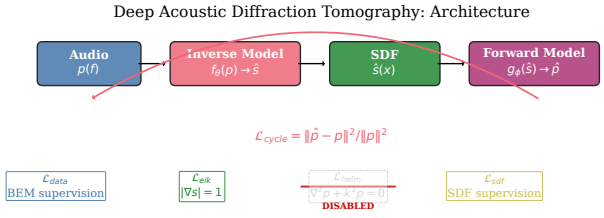

Fig. 1: System architecture. The forward model learns a transfer function $T = p_{\text{scat}}/p_{\text{inc}}$; the inverse model maps acoustic data to SDF via an auto-decoder. Cycle-consistency closes the loop. Note: Helmholtz loss $\mathcal{L}_{\text{helm}}$ was found to be incompatible with neural surrogates and is *disabled* (Sec. III-E).

Our contributions are:

1) A **transfer function formulation** $T = p_{\text{scat}}/p_{\text{inc}}$ that reduces forward error from 48% to 4.47% at $2{,}000\times$ BEM speedup on 15 scenes with 200 frequencies (Sec. II-B).
2) An **auto-decoder inverse model** with Eikonal-regularized SDF that reconstructs geometry at 0.91 ± 0.01 mean IoU, validated via cycle-consistency ($r = 0.91$) and robust to 10 dB noise (Sec. II-C).
3) **Analysis of Helmholtz PDE failure** in neural surrogates: we show that the autodiff Laplacian explains only 3.5% of the physical Laplacian's variance ($r = 0.19$), caused by $\mathcal{O}(10^4)$ second-derivative amplification from random Fourier features (Sec. III-E).

II. Method

A. Problem Formulation

Consider a 2D acoustic scene with scatterers occupying region Ω bounded by surface $\partial\Omega$. A point source at position $\boldsymbol{x}_s$ emits a monochromatic wave at wavenumber $k = 2\pi f/c$, where $c = 343$ m/s. The total pressure $p_{\text{tot}}(\boldsymbol{x})$ satisfies the Helmholtz equation $(\nabla^2 + k^2)p_{\text{tot}} = -\delta(\boldsymbol{x} - \boldsymbol{x}_s)$ in the exterior

domain, with Neumann boundary conditions on $\partial\Omega$. The scattered field is $p_{\text{scat}} = p_{\text{tot}} - p_{\text{inc}}$, where $p_{\text{inc}} = \frac{i}{4}H_0^{(1)}(k|\boldsymbol{x} - \boldsymbol{x}_s|)$ is the free-space Green's function in 2D. Edge diffraction [13] produces the dominant contribution to p_{scat} in our scenes.

We seek to learn: (i) a forward surrogate $f_\theta : (\boldsymbol{x}_s, \boldsymbol{x}_r, k) \mapsto p_{\text{tot}}$ that approximates BEM, and (ii) an inverse mapping $g_\phi : \{p_{\text{tot}}\} \mapsto s(\boldsymbol{x})$ that recovers the signed distance function $s : \mathbb{R}^2 \to \mathbb{R}$ of $\partial\Omega$.

B. Forward Model: Transfer Function Learning

Transfer function target. Rather than learning p_{tot} directly, we define the transfer function

$$T(\boldsymbol{x}_s, \boldsymbol{x}_r, k) = \frac{p_{\text{scat}}(\boldsymbol{x}_s, \boldsymbol{x}_r, k)}{p_{\text{inc}}(\boldsymbol{x}_s, \boldsymbol{x}_r, k)}, \tag{1}$$

which removes the dominant $H_0^{(1)}(kr) \sim e^{ikr}/\sqrt{r}$ oscillation from the learning target. This is analogous to learning a scattering matrix rather than the total field. The total pressure is recovered as $p_{\text{tot}} = p_{\text{inc}} \cdot (1 + T \cdot \sigma)$, where σ is a per-scene normalization scale.

Architecture. The forward model f_θ takes 9 scalar inputs: source and receiver coordinates $(\boldsymbol{x}_s, \boldsymbol{x}_r) \in \mathbb{R}^4$, wavenumber k, source–receiver distance d, signed distance at the receiver $s(\boldsymbol{x}_r)$, and spatial derivatives $(\partial s/\partial x, \partial s/\partial y)$. These are encoded via random Fourier features [14] ($D = 128$, bandwidth $\sigma_{\text{FF}} = 30\,\text{m}^{-1}$), concatenated with a learnable scene embedding $e_s \in \mathbb{R}^{32}$. The network consists of 8 residual blocks with hidden dimension 768, outputting $(\text{Re}(T), \text{Im}(T)) \in \mathbb{R}^2$. The SDF features $(s, \partial s/\partial x, \partial s/\partial y)$ condition the forward model on geometry; during inverse training (Sec. II-D), these carry gradients from the SDF decoder.

Ensemble and calibration. We train four models with different seeds and apply a linear calibration layer $T_{\text{calib}} = aT_{\text{pred}} + b$ on a held-out validation set. The ensemble reduces per-model variance and achieves 4.47% overall error (Sec. III-B).

C. Inverse Model: Auto-Decoder SDF Reconstruction

SDF decoder. Following DeepSDF [12], we use an auto-decoder architecture where each scene i has a learnable latent code $\boldsymbol{z}_i \in \mathbb{R}^{256}$. The SDF decoder D_ψ takes Fourier-encoded 2D coordinates $\gamma(\boldsymbol{x})$ (bandwidth $\sigma = 10$) concatenated with $\boldsymbol{z}_i$ and outputs a signed distance value through 6 residual blocks (hidden dimension 256):

$$s_i(\boldsymbol{x}) = D_\psi(\gamma(\boldsymbol{x}), \boldsymbol{z}_i). \tag{2}$$

Multi-code composition. For multi-body scenes (e.g., our Scene 12 with two disjoint objects), we assign K latent codes $\{\boldsymbol{z}_i^{(k)}\}_{k=1}^K$ and compose via smooth minimum:

$$s_i(\boldsymbol{x}) = -\frac{1}{\alpha}\log\sum_{k=1}^{K}\exp\big(-\alpha \cdot D_\psi(\gamma(\boldsymbol{x}), \boldsymbol{z}_i^{(k)})\big), \tag{3}$$

with sharpness $\alpha = 50$, approximating $\min_k s_i^{(k)}$.

Loss function. The total loss is:

$$\mathcal{L} = \mathcal{L}_{\text{sdf}} + \lambda_1\mathcal{L}_{\text{eik}} + \lambda_2\mathcal{L}_{\text{cycle}}, \tag{4}$$

where $\mathcal{L}_{\text{sdf}} = \mathbb{E}[|D_\psi(\boldsymbol{x}) - s^*(\boldsymbol{x})|]$ is the L1 SDF supervision, $\mathcal{L}_{\text{eik}} = \mathbb{E}[(|\nabla_{\boldsymbol{x}} s| - 1)^2]$ is the Eikonal constraint enforcing $|\nabla s| = 1$, and $\mathcal{L}_{\text{cycle}}$ is the cycle-consistency loss (Sec. II-D). We set $\lambda_1 = 0.1$, $\lambda_2 = 0.01$.

Boundary oversampling. We oversample SDF training points near $s(\boldsymbol{x}) \approx 0$ by a factor of $3\times$, which is critical for resolving thin geometries (ablation in Table II).

On Helmholtz PDE loss. A natural extension would add a Helmholtz residual loss $\|\nabla^2\hat{p} + k^2\hat{p}\|^2$ via the forward surrogate. However, this *degrades* reconstruction: the neural ∇^2 captures network curvature rather than physical pressure gradients, due to second-derivative amplification in the Fourier features. We analyze this failure in detail in Sec. III-E.

D. Cycle-Consistency

The cycle-consistency loss connects the forward and inverse models:

$$\mathcal{L}_{\text{cycle}} = \mathbb{E}\big[\|f_\theta(\boldsymbol{x}_s, \boldsymbol{x}_r, k; s_i) - p_{\text{tot}}^{\text{BEM}}\|^2\big], \tag{5}$$

where $s_i(\boldsymbol{x}_r) = D_\psi(\gamma(\boldsymbol{x}_r), \boldsymbol{z}_i)$ is evaluated at receiver positions and fed as an additional feature to the frozen forward model f_θ. The forward model parameters are frozen during inverse training; only $\boldsymbol{z}_i$ and D_ψ are updated.

This creates a differentiable loop: latent code $\boldsymbol{z}_i \to$ SDF at receivers $\to$ forward prediction $\to$ comparison with BEM data. The gradient flows through the SDF decoder, providing acoustic supervision for geometry beyond the SDF loss alone. We use the MSE loss (5) for training, and report Pearson correlation r as the evaluation metric (Sec. III-D) because it is scale-invariant and interpretable.

III. Experiments

A. Dataset

We generate 2D BEM data for 15 scenes spanning 4 geometry classes: wedges (4), cylinders (2), polygons and barriers (6), and multi-body compositions (3). For each scene, 3 source positions illuminate the geometry, with receivers placed at 40–200 positions per source. We solve the BEM [15] at 200 frequencies uniformly spaced in 2–8 kHz ($k \in [36.6, 146.5]$ rad/m), yielding 1,769,400 complex pressure observations in total. The BEM solver is validated against Macdonald's analytical solution [16] for a 90° wedge (1.77% L_2 error).

Room impulse responses (RIRs) are synthesized via inverse DFT with phase unwrapping. All 8,853 source–receiver pairs satisfy the causality criterion $E(t < t_{\text{arrival}})/E_{\text{total}} < 10^{-4}$.

B. Forward Model Results

Table I shows the forward model ablation. The top block validates the transfer function formulation: a vanilla MLP learning raw p_{scat} achieves 48.00% error—identical to the trivial "no scatterer" baseline ($\hat{p} = p_{\text{inc}}$, 47.95%), confirming that direct pressure learning fails entirely. Switching to the T formulation reduces error to 2.27% (single model, 882 epochs) even *without* Fourier features, demonstrating that the transfer function—not the encoding—is the key contribution.

TABLE I: Forward model ablation. Top: target formulation comparison validates that the transfer function (T) is essential. Bottom: ensemble and calibration.

Configuration	Target	Error (%)
No scatterer ($\hat{p} = p_{\text{inc}}$)	—	47.95
Vanilla MLP	p_{scat}	48.00
No Fourier features*	T	2.27
Single model	T	11.54
+ calibration	T	10.20
Duo ensemble + calib	T	9.89
Quad ensemble	T	4.57
Quad ens. + calib	T	**4.47**

*Single model, 882 epochs (∼3.2 h); ensemble members train ∼200 epochs (∼40 min) each with better efficiency.

TABLE II: Inverse model ablation: cumulative loss components.

Configuration	IoU	S12	r
$\mathcal{L}_{\text{sdf}} + \mathcal{L}_{\text{eik}}$ (200 ep)	0.689	0.135	—
+ bdy 3× (500 ep)	0.842	0.184	—
+ $\mathcal{L}_{\text{cycle}}$ (1000 ep)	0.939	0.410	0.909
+ multi-code K=2	**0.949**	0.493	0.902

Fourier features accelerate convergence but do not improve final accuracy given sufficient training. The quad ensemble with calibration yields 4.47% overall, with per-scene errors from 0.93% (Scene 1) to 18.62% (Scene 13, deep shadow zone). At inference, the neural forward model evaluates 50,000 pressure samples in 130 ms (GPU), achieving 2,000× speedup over BEM (260 s per scene).

C. Inverse Reconstruction

Table II shows the inverse model ablation. Starting from SDF + Eikonal losses (IoU = 0.69), adding boundary oversampling (+0.15), cycle-consistency (+0.10), and multi-code composition (+0.01) yields a final mean IoU of 0.9491. Fourteen of 15 scenes achieve IoU > 0.92 (Fig. 2); the sole exception is Scene 12 (two parallel plates, IoU = 0.49), where the smooth-minimum composition (3) struggles with disjoint bodies. Supplementary geometry metrics confirm reconstruction quality: mean Chamfer distance 0.047 m and mean Hausdorff distance 0.471 m across the 15 scenes (Hausdorff is dominated by wedge scenes S1–S4, where open-mesh truncation creates edge artifacts at HD > 0.8 m; closed-surface scenes achieve HD < 0.035 m).

Cycle-consistency across all 15 scenes yields mean Pearson $r = 0.90$ (all $r > 0.83$). Notably, Scene 12 achieves $r = 0.92$ despite IoU = 0.49, showing cycle-consistency is necessary but *not sufficient* for geometry accuracy.

D. Robustness and Generalization

Robustness. Adding complex Gaussian noise at 10–40 dB SNR shows graceful degradation: $r = 0.86$ at 10 dB ($\Delta r = -0.04$). Three-seed reproducibility confirms IoU = 0.912 ± 0.011, $r = 0.907 \pm 0.001$ (all seeds pass gates). Leave-one-out code optimization recovers 52% mean IoU; wedge-like

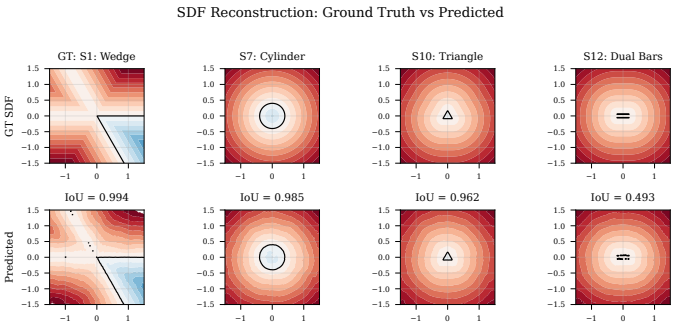

Fig. 2: SDF reconstruction for four representative scenes. Top: ground truth; bottom: predicted. The model achieves high fidelity for single-body geometries (S1, S7, S10) but struggles with the disjoint multi-body Scene 12 (IoU = 0.49).

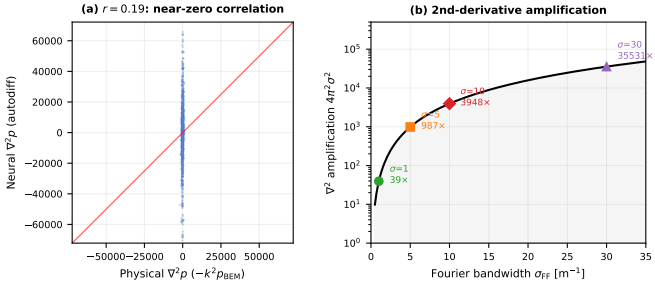

Fig. 3: Helmholtz PDE analysis. (a) Neural $\nabla^2 p$ (autodiff) vs. physical $\nabla^2 p = -k^2 p_{\text{BEM}}$: Pearson $r = 0.19$, explaining only 3.5% of physical variance. (b) Second-derivative amplification $4\pi^2\sigma^2$ of random Fourier features: at $\sigma = 30$, the amplification exceeds 35,000×.

shapes generalize (S1: 92%, S14: 97%) while novel shapes do not (S5: 9%, S10: 22%). Cross-frequency extrapolation (2–6 kHz→6–8 kHz) fails at 43% error, indicating per-frequency memorization rather than spectral continuity.

E. Why Helmholtz PDE Loss Fails

Physics-informed approaches typically enforce $\|\nabla^2 p + k^2 p\|^2 \to 0$ via automatic differentiation of a neural field [10]. We investigate *why* this fails for our surrogate architecture and identify the root cause.

Empirical measurement. We compute the neural Laplacian $\nabla^2_{\text{auto}} p$ via second-order autodiff of f_θ w.r.t. receiver coordinates, and compare it against the physical Laplacian $\nabla^2_{\text{phys}} p = -k^2 p$ (which holds exactly for the BEM ground truth by the Helmholtz equation). Over 10,000 evaluation points across 5 scenes (Fig. 3a), the Pearson correlation is $r = 0.19$—the neural Laplacian explains only 3.5% of the physical Laplacian's variance. The median Helmholtz residual $|\nabla^2_{\text{auto}} p + k^2 p|$ is $\mathcal{O}(10^3)$, confirming that the neural ∇^2 is dominated by network curvature. Enabling this loss collapsed IoU from 0.82 to 0.19 within 30 epochs.

Cause: Fourier feature amplification. Wang et al. [11] showed that random Fourier features modulate NTK convergence rates. We extend this to identify a practical failure

mode: the encoding $\gamma(\boldsymbol{v}) = \cos(2\pi \boldsymbol{B}\boldsymbol{v})$ with $B_{ij} \sim \mathcal{N}(0, \sigma^2)$ has second derivatives $\partial^2\gamma/\partial v_j^2 = -(2\pi B_j)^2\gamma(\boldsymbol{v})$, yielding an expected amplification $\mathbb{E}[(2\pi B_j)^2] = 4\pi^2\sigma^2$. At $\sigma = 30\,\mathrm{m}^{-1}$ (required for resolving $k_{\max} = 146\,\mathrm{rad/m}$), this gives $4\pi^2 \cdot 900 \approx 35{,}000\times$ amplification (Fig. 3b), compounded to $\sim$55,000$\times$ by the 8-layer ResNet. High σ is needed for forward accuracy, but it makes $\nabla^2 f_\theta$ physically meaningless.

Asymmetry with Eikonal. The Eikonal constraint $|\nabla s| = 1$ succeeds because it requires only *first*-order gradients of the SDF decoder's own output, where network gradients align with the physical quantity. Helmholtz requires *second*-order gradients of a different network (the frozen forward model), through a high-bandwidth Fourier encoding. This first- vs. second-order asymmetry explains why geometry constraints work while wave-equation constraints do not.

IV. Conclusion

We presented a cycle-consistent neural framework for 2D acoustic diffraction tomography. The transfer function formulation is the central enabler: it reduces forward error from 48% to 4.47% while providing 2,000$\times$ speedup over BEM (0.13 s vs. 260 s per scene). A single model without Fourier features reaches 2.27%, confirming the formulation—not the encoding—as the key factor. The auto-decoder inverse model achieves 0.91 ± 0.01 mean IoU, validated by cycle-consistency ($r = 0.907 \pm 0.001$) robust to 10 dB noise.

Our Helmholtz PDE analysis reveals a fundamental obstacle for physics-informed neural surrogates: random Fourier features amplify second derivatives by $\sim$35,000$\times$, reducing the neural-physical Laplacian correlation to $r = 0.19$. This quantifies the gap between network curvature and physical Laplacians, with implications for the PINN literature on acoustic surrogate models.

Limitations include: 2D synthetic data only (15 scenes), per-scene optimization, limited cross-frequency generalization, and disjoint geometry difficulty (S12 IoU = 0.49). Future work targets 3D extension, frequency-continuous architectures, and differentiable BEM solvers for physically grounded PDE enforcement.

References

[1] M. Kuster, "The role of acoustic simulation in virtual environments," *Building Acoustics*, vol. 11, no. 1, pp. 1–20, 2004.

[2] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, 4th ed., ser. Applied Mathematical Sciences. Springer, 2019, vol. 93.

[3] A. Richard, P. Dodds, and V. K. Ithapu, "Deep impulse responses: Estimating and parameterizing filters with deep networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 3209–3213.

[4] A. Luo, Y. Du, M. J. Tarr, J. B. Tenenbaum, A. Torralba, and C. Gan, "Learning neural acoustic fields," in *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 26 393–26 405.

[5] K. Su, K. Qian, E. Shlizerman, A. Torralba, and C. Gan, "INRAS: Implicit neural representation for audio scenes," in *Advances in Neural Information Processing Systems*, vol. 36, 2023.

[6] X. Liang, Y. Zheng, A. Luo, and C. Gan, "AcousticNeRF: Acoustic-aware neural radiance fields for indoor scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[7] D. B. Lindell, G. Wetzstein, and V. Koltun, "Acoustic non-line-of-sight imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6780–6789.

[8] I. Yeon *et al.*, "EchoScan: Scanning complex room geometries via acoustic echoes," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 4768–4782, 2024.

[9] T. Vlašić, H. Nguyen, A. Khorashadizadeh, and I. Dokmanić, "Implicit neural representation for mesh-free inverse obstacle scattering," in *56th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2022, pp. 947–951.

[10] M. Raissi, P. Perdikaris, and G. 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*, vol. 378, pp. 686–707, 2019.

[11] S. Wang, H. Wang, and P. Perdikaris, "On the eigenvector bias of Fourier feature networks: From regression to solving multi-scale PDEs with physics-informed neural networks," *Computer Methods in Applied Mechanics and Engineering*, vol. 384, p. 113938, 2021.

[12] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "DeepSDF: Learning continuous signed distance functions for shape representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 165–174.

[13] R. G. Kouyoumjian and P. H. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface," *Proceedings of the IEEE*, vol. 62, no. 11, pp. 1448–1461, 1974.

[14] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 7537–7547.

[15] F. Ihlenburg, *Finite element analysis of acoustic scattering*. Springer, 1998.

[16] H. M. Macdonald, *Electric Waves*. Cambridge University Press, 1902.