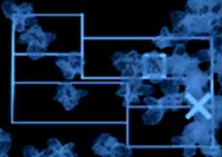
NEURAL ACOUSTIC MULTIPOLE SPLATTING

FOR ROOM IMPULSE RESPONSE SYNTHESIS



Geonwoo Baek & Jung-Woo Choi

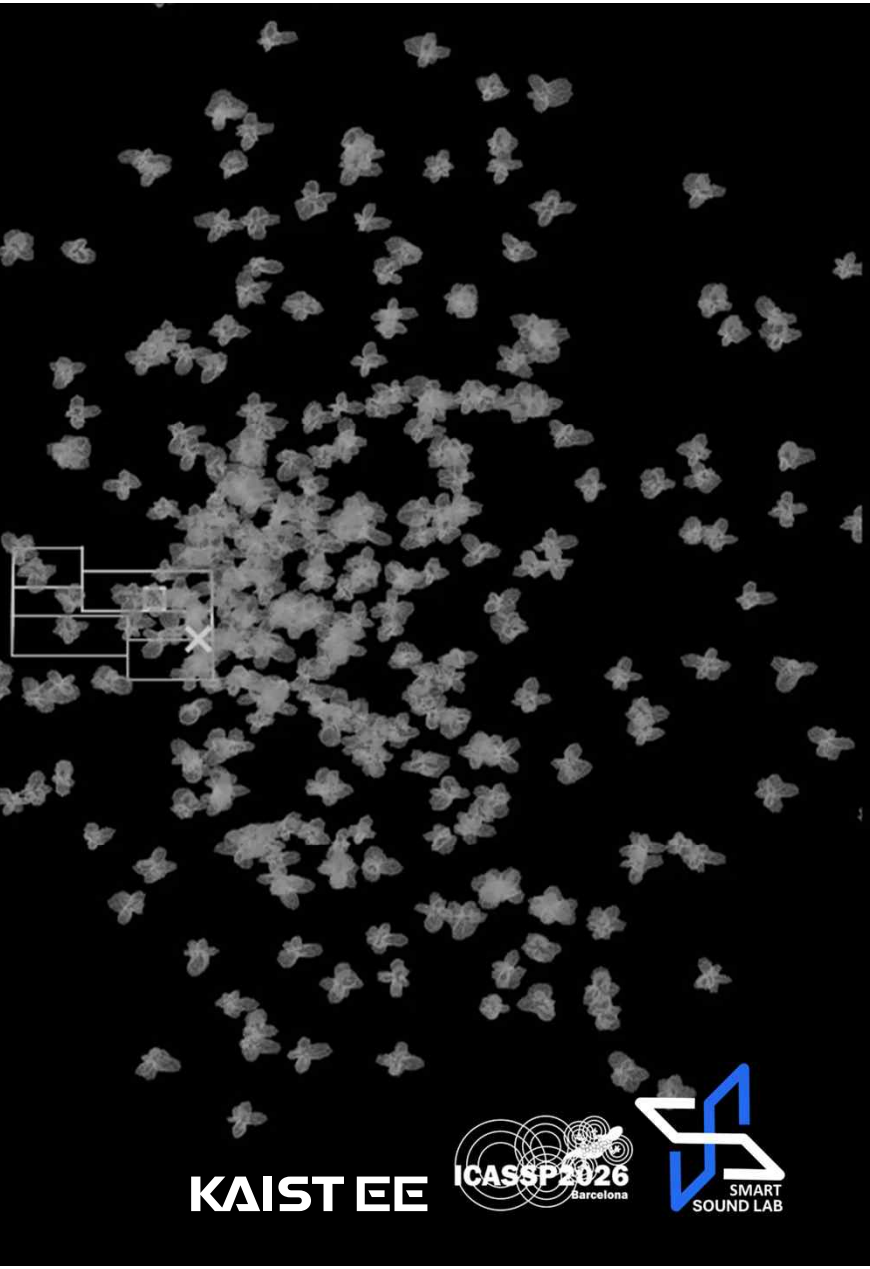
School of Electrical Engineering
Korea Advanced Institute of Science & Technology

SS-L15 Neural Spatial Audio Processing, Room 111, Fri 15:00

KAIST EE ICASSP2026 Barcelona SMART SOUND LAB

Outline

1. **Introduction**
2. **Proposed Method**
3. **Experiment**
4. **Conclusion**



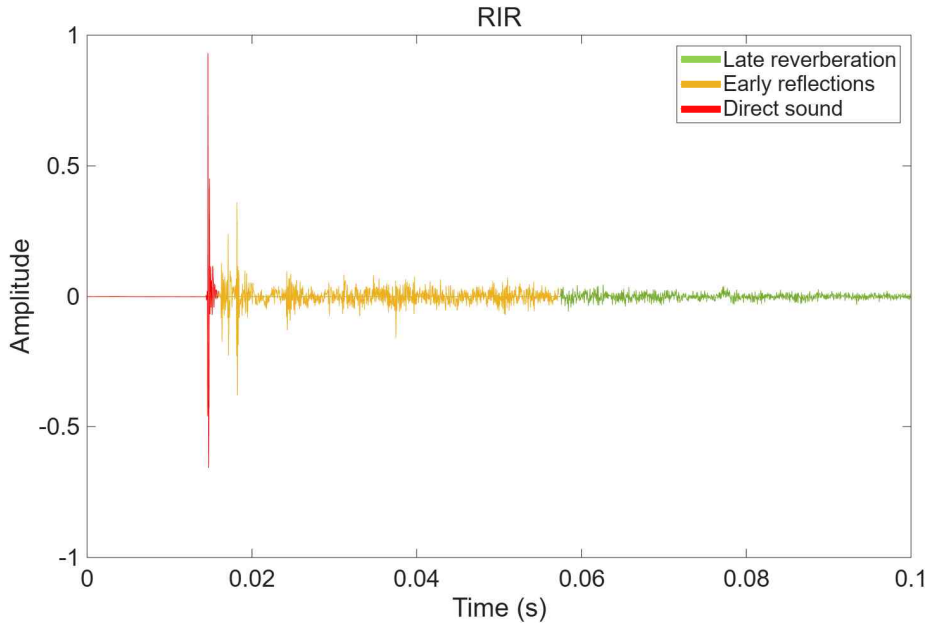
KAIST EE ICASSP2026 Barcelona SMART SOUND LAB

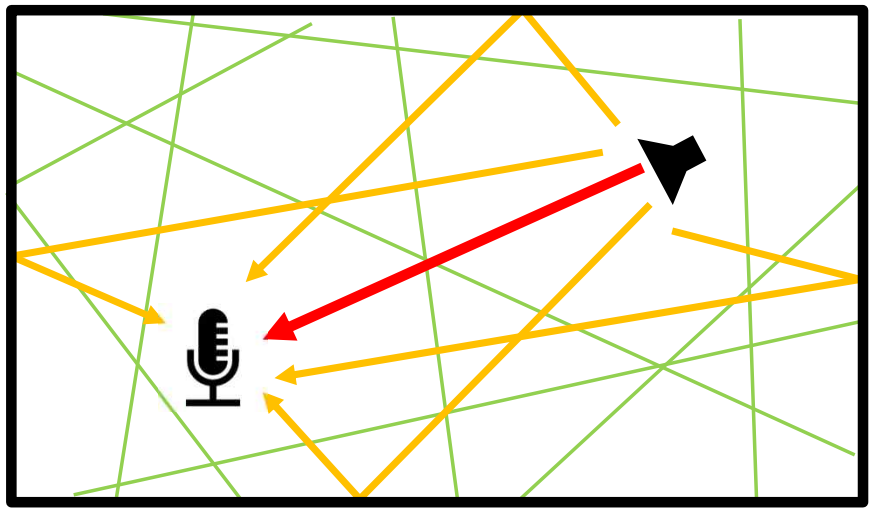
Introduction

NEURAL ACOUSTIC MULTIPOLE SPLATTING
FOR ROOM IMPULSE RESPONSE SYNTHESIS

Room Impulse Response (RIR)

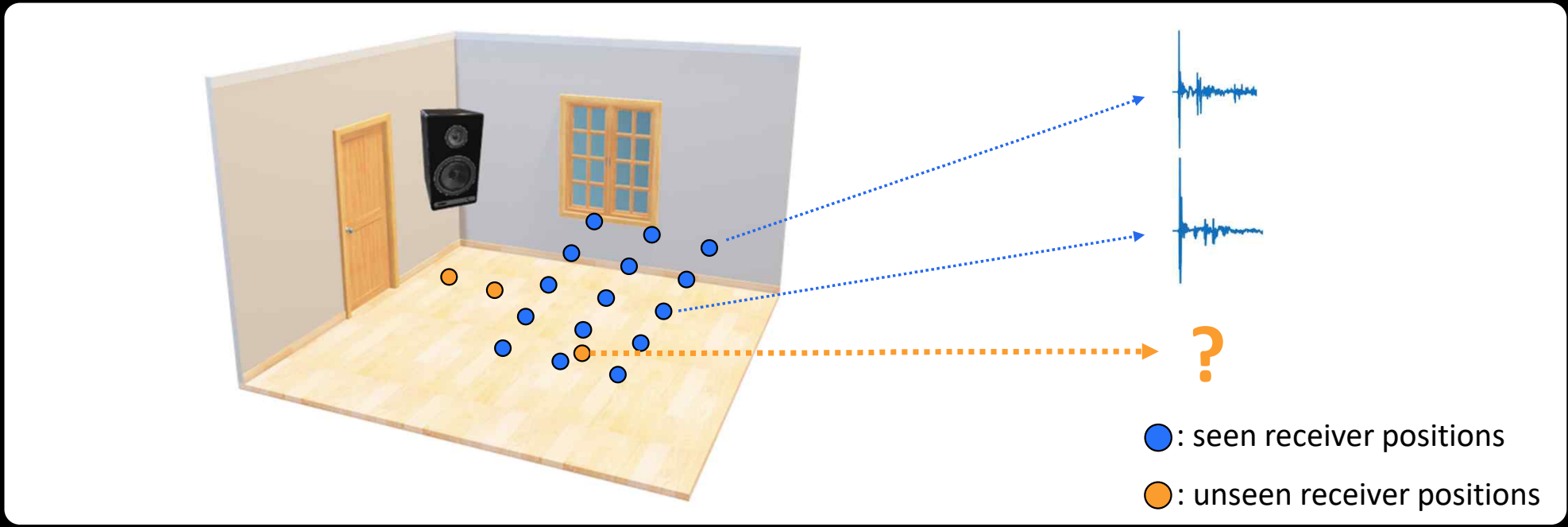
- Describes how sound propagates from a **source** to a **receiver**
- Includes direct sound, early reflections, and late reverberation





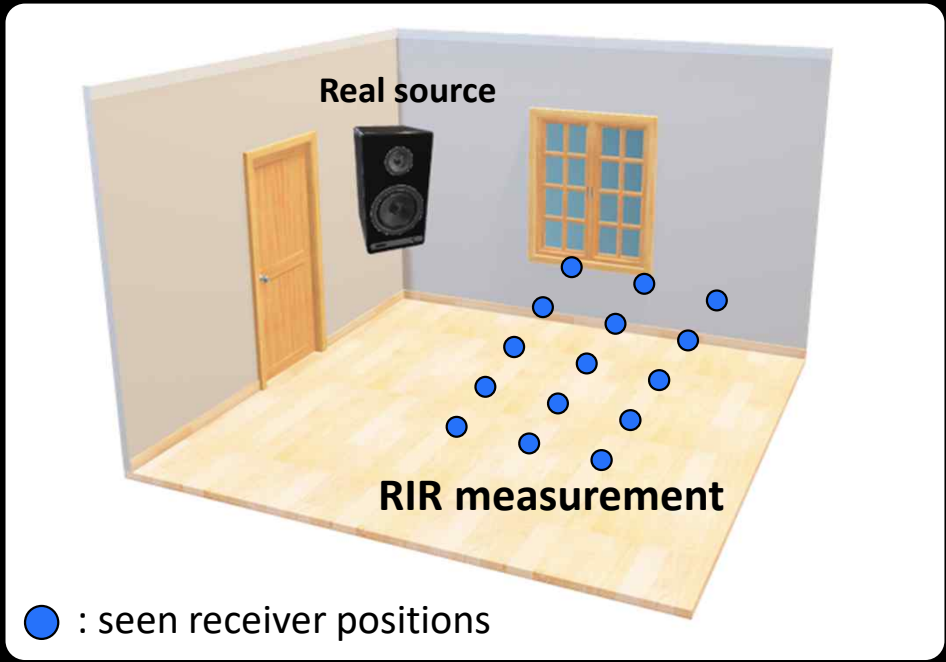
RIR synthesis

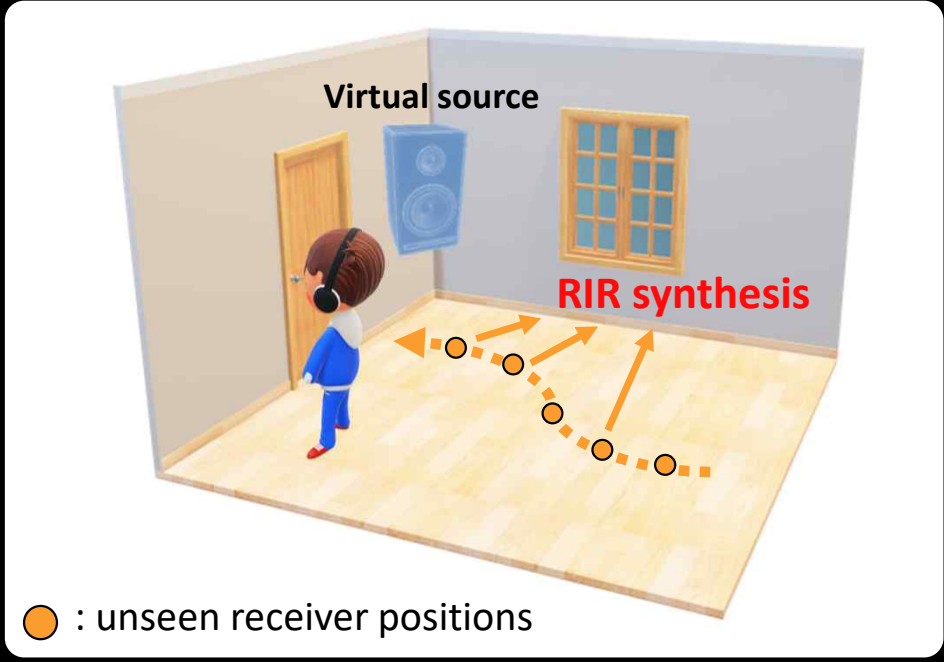
- Predict RIRs at arbitrary receiver positions using limited RIR measurements
- Generalization to **unseen receiver positions** is important



Application | AR audio rendering

- For AR applications, RIRs should be rendered at all receiver positions
- Measuring or simulating RIRs at all receiver positions is infeasible



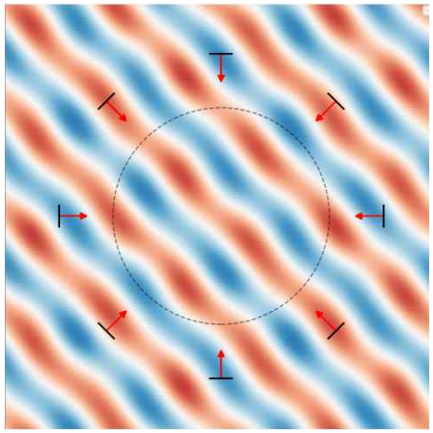


Approaches | Basis Expansion

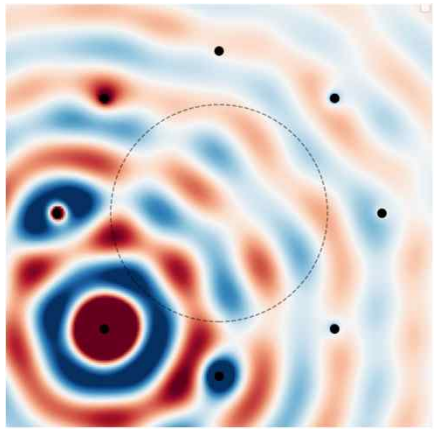
- **Model the given sound field as a linear combination of basis functions**
  - Finite number of fixed basis functions

    → Struggle to describe complex sound fields (spatial aliasing)

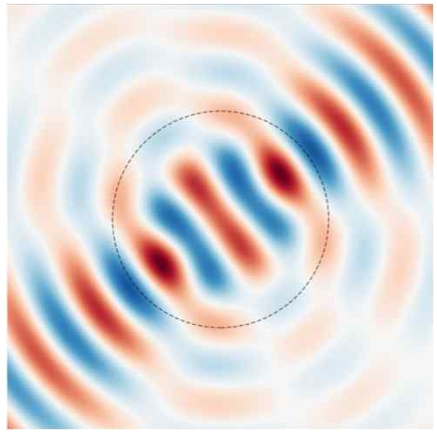
Plane Wave Expansion (PWE)



Equivalent Source Method (ESM)



Spherical Harmonic Expansion



Approaches | Parameterization

- **Room acoustic parameter interpolation** [Raghuvanshi 2018, Chakravarty 2020]
  - **Direct sound**: Initial delay, Initial direction, Initial loudness
  - **Reflections:** Reflections delay, Decay time, Energy flux …
    → Requires dense sampling for smooth interpolation

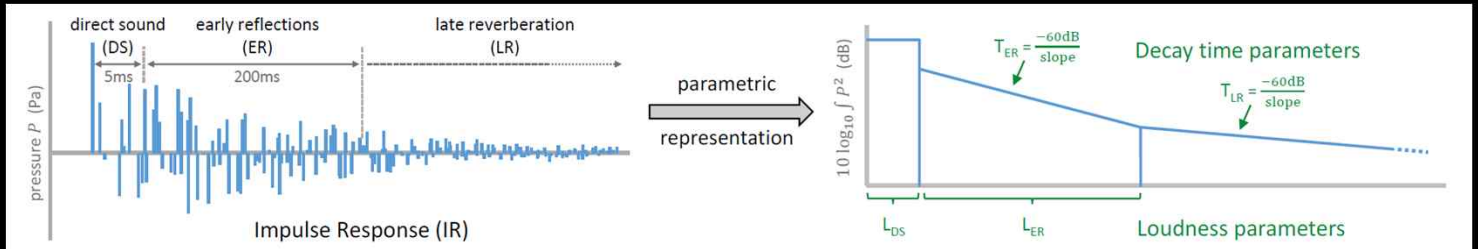
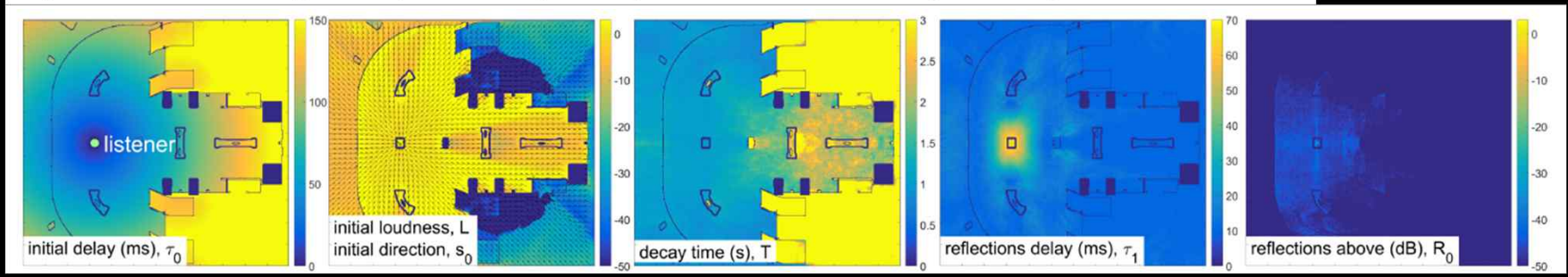


Figure 3: Parametric IR encoding schematic (time not to scale). The four parameters we extract are shown in green on the right for an IR shown on the left.

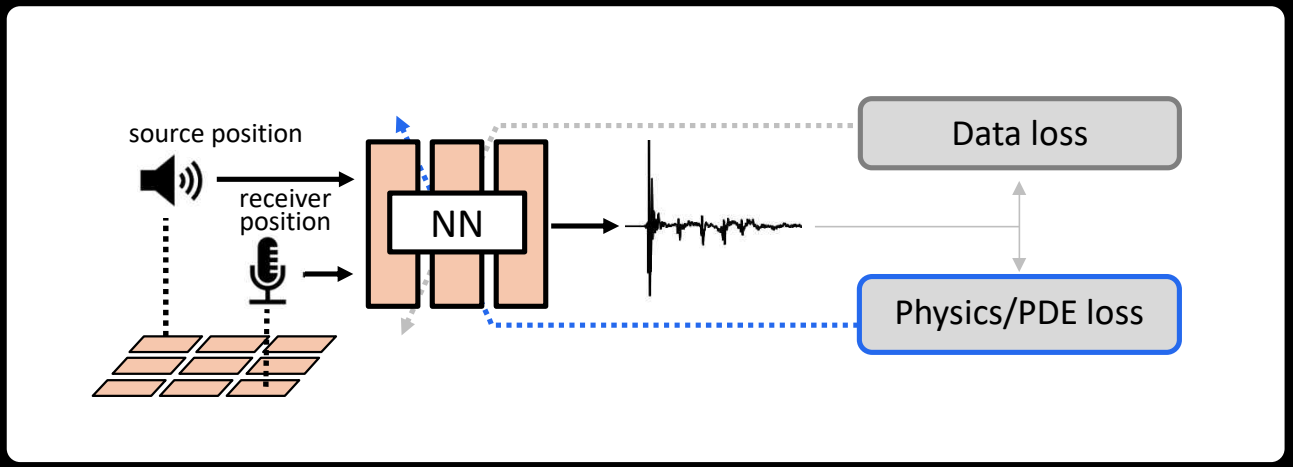


Images from Raghuvanshi et al., ACM Trans. Graph., Vol. 37, No. 4, Article 108, 2018
Chakravarty et al., ACM Trans. Graph., Vol. 39, No. 4, Article 44, 2020

Approaches | Neural Net (Direct)

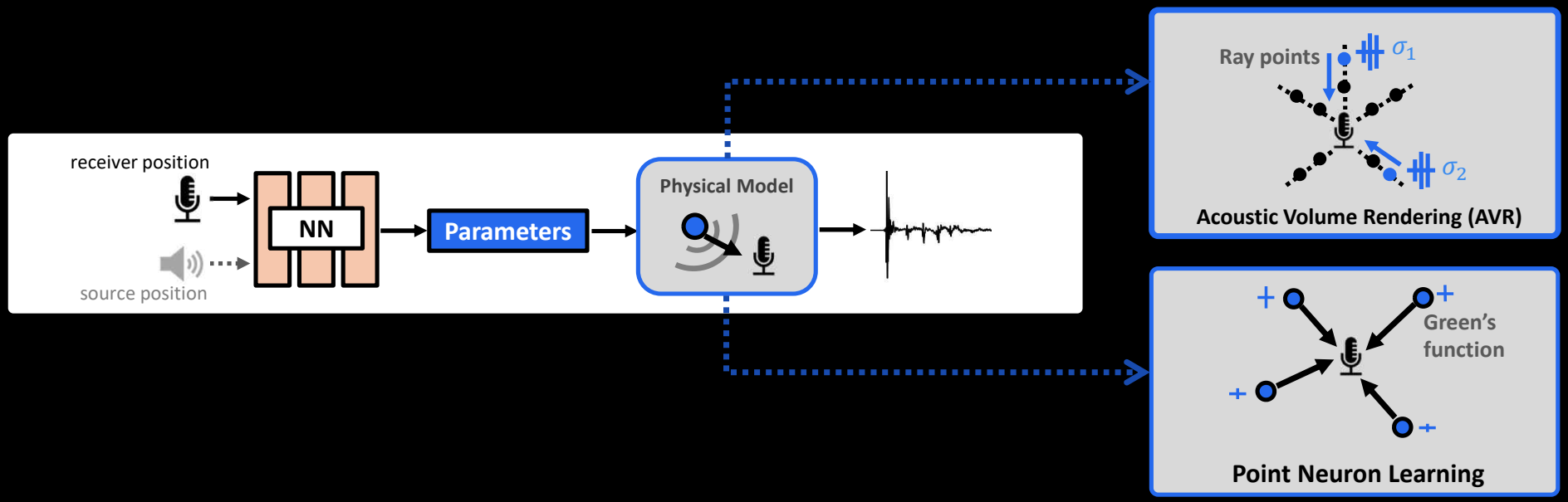
- **Direct RIR synthesis through nonlinear mapping**
  - Power of nonlinear mapping: **Rich representation** with limited parameters
  - Neural Acoustic Field [Luo 2022] : Train NN to generate an RIR for given source & receiver positions
- **Physically-informed loss function**
  - Learns a continuous field under Physical constraint

  → Struggle to extrapolate beyond observation regions



Approaches | Neural Net + physical model

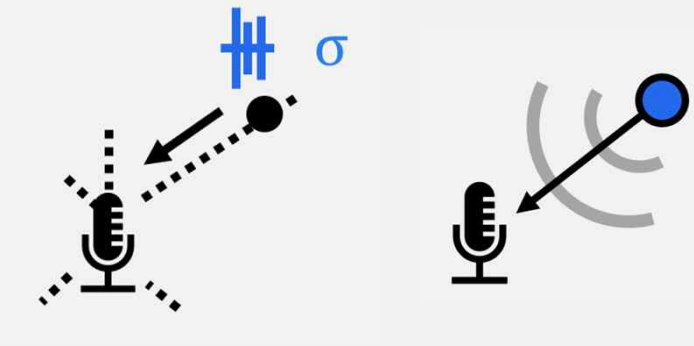
- **NN learns parameters for a given physical model**
  - Acoustic Volume Rendering (AVR) [Lan 2024] : ray-based acoustic rendering
  - Point Neuron Learning [Bi 2024] : Green's-function-based point sources
  - (+) Better connection to physics



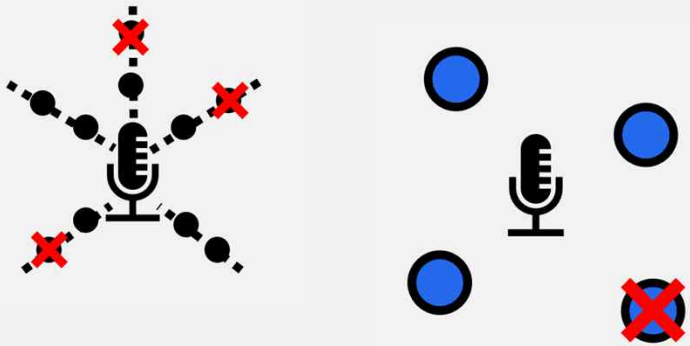
Challenges

- **Expression capability & Efficiency of Basis or Models**
  - More expressive & physically plausible basis than monopole or ray
  - Efficiency: finding a minimal set of basis functions

Highly expressive basis?

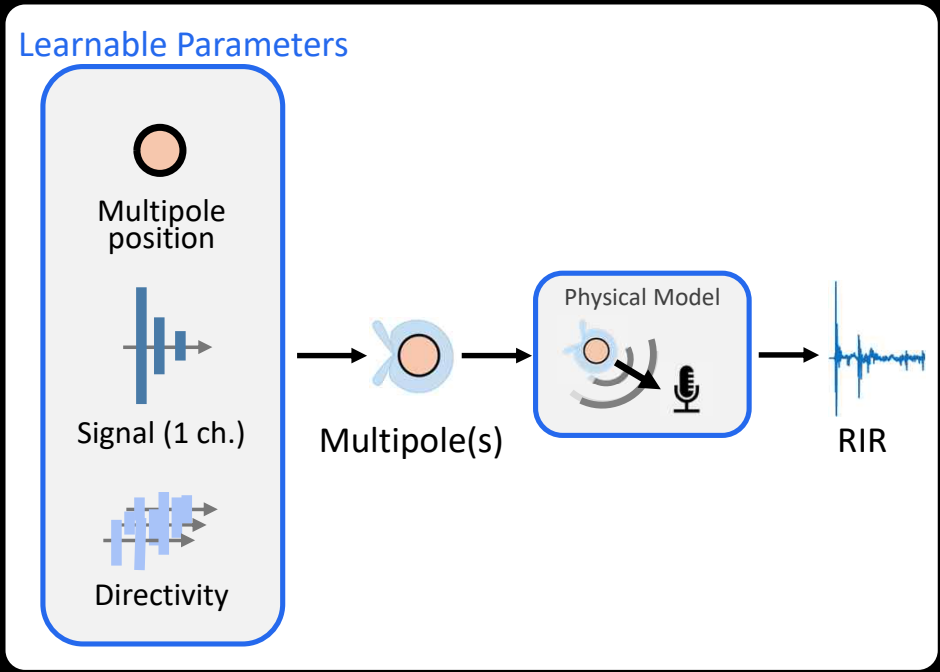


Minimizing the necessary basis?

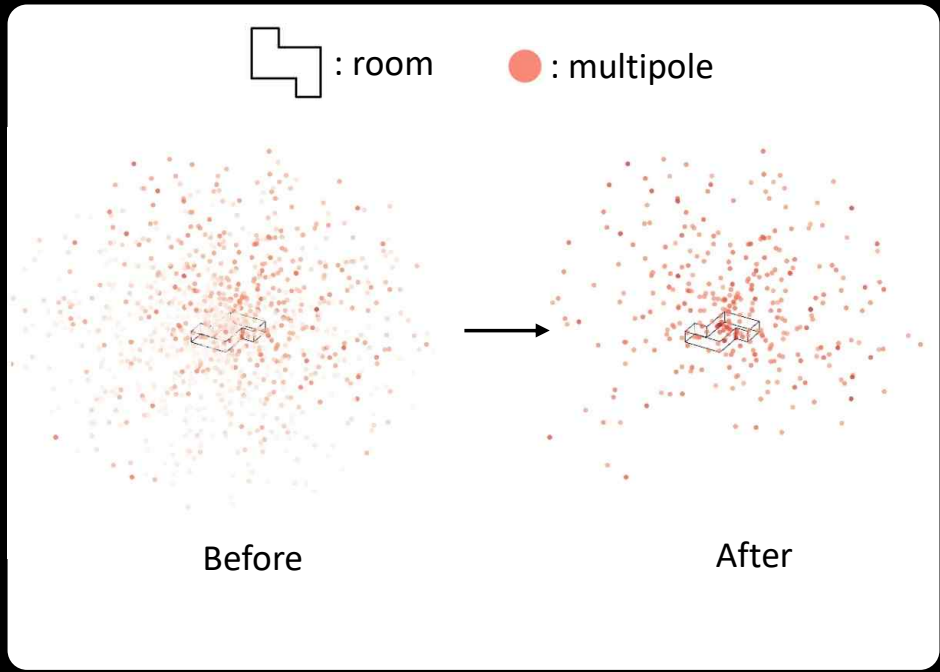


Our approach

- **Neural Acoustic Multipole**
  - Highly expressive basis
  - Synthesis RIR by superposing multipoles



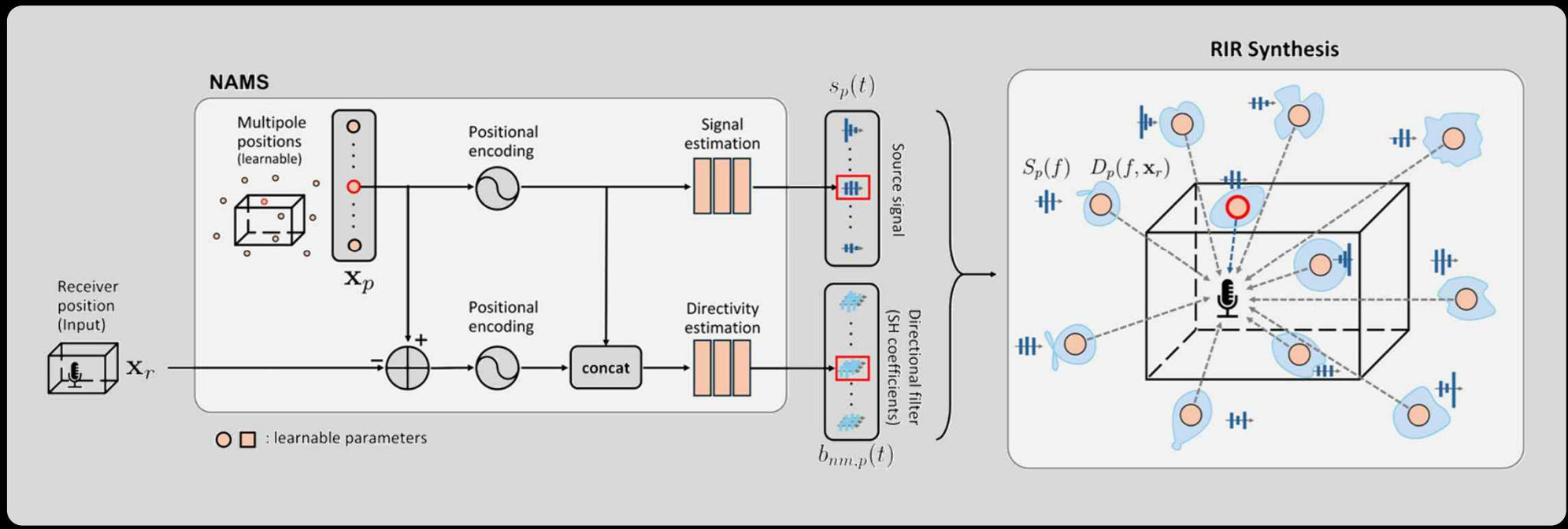
- **Multipole Pruning**
  - Selection of significant multipoles
  - For computational complexity reduction



Proposed method

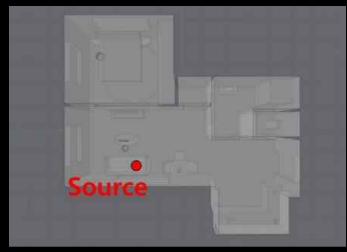
Neural Acoustic Multipole Splatting (NAMS)

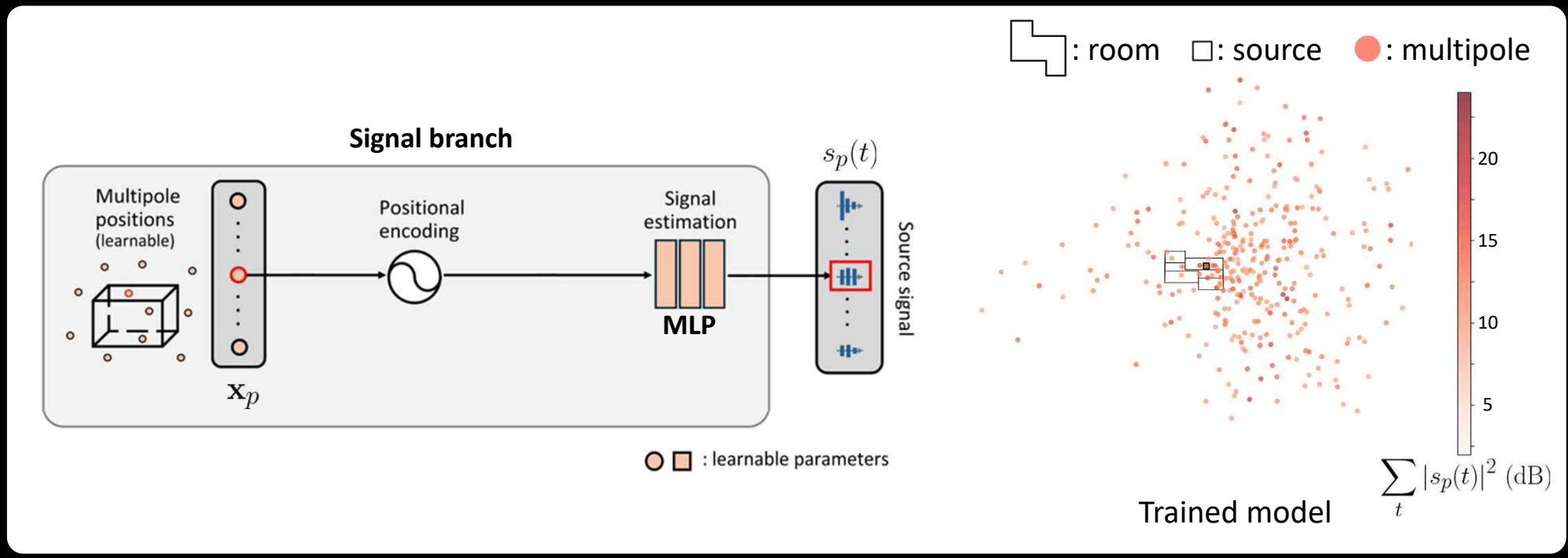
- NAMS learns **multipole positions** and predicts **signals and directivities**
- RIRs are synthesized through a differentiable process using multipoles



1. Signal branch

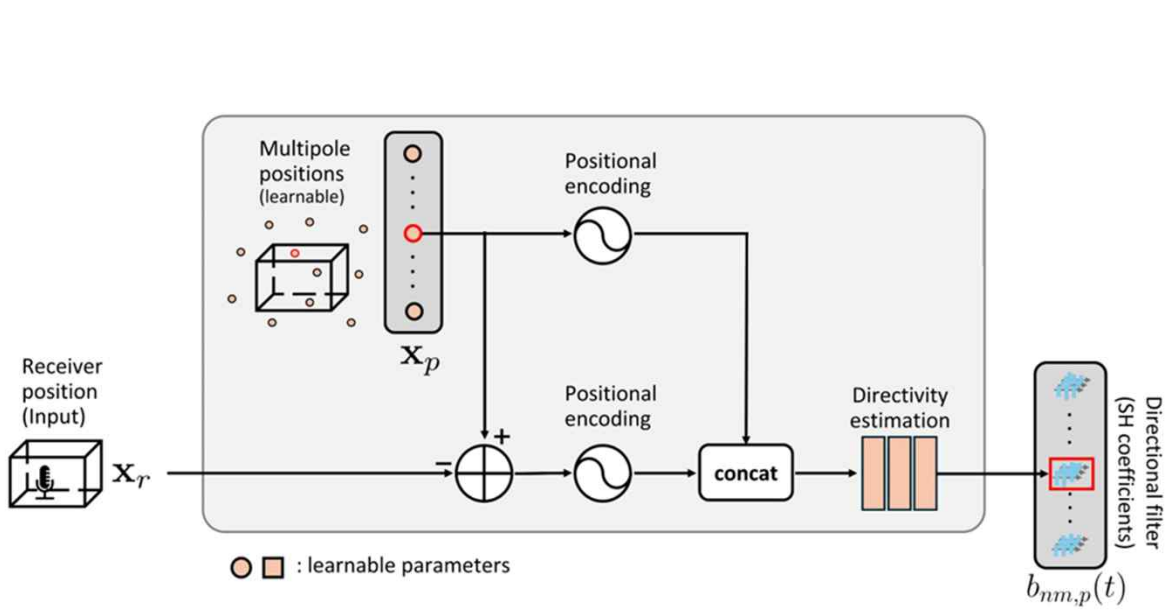
- **Multipole signal: derived solely from its position**
  - Signal branch ensures that each multipole signal is receiver-independent
  - Signal estimation MLP generates 3 ms signals to reduce temporal redundancy

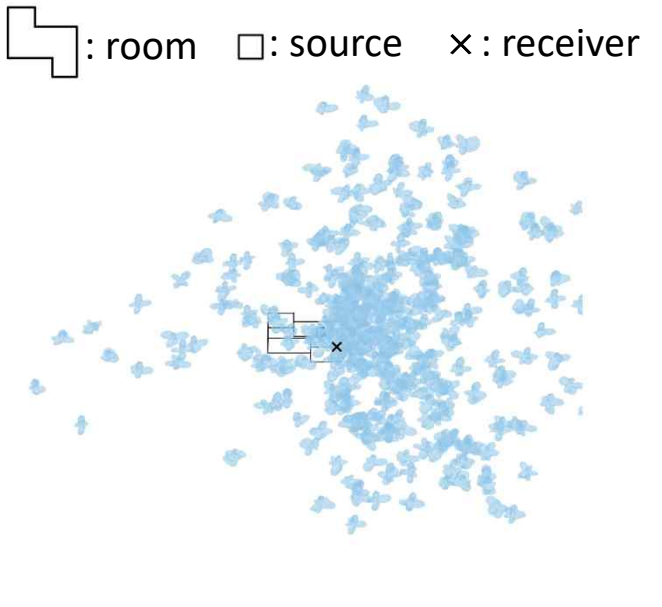




2. Directivity branch

- **Directivity depends on both multipole and receiver positions**
  - Directivity estimation MLP generates receiver-dependent variable directivity
  - Directivity is represented by **3rd-order SH coefficients of 3 ms duration**

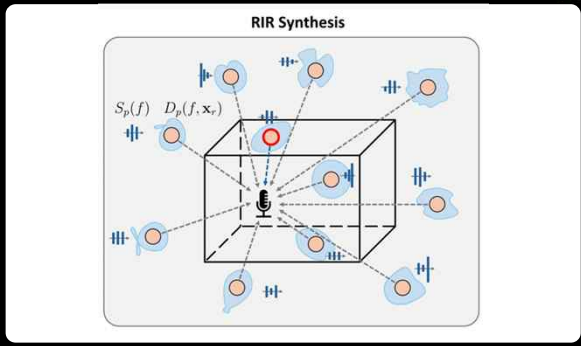




Trained model (directivities at 1kHz)

3. RIR synthesis

- **RIR = superposed multipoles**
  - Normalize directivity to ensure consistent energy scaling
  - Frequency-domain synthesis with propagation delay and attenuation



Directivity

$$D_p(f, \mathbf{x}_r) = \sum_{n=0}^{N} \sum_{m=-n}^{n} B_{nm,p}(f)\, Y_n^m(\boldsymbol{\Omega}_p(\mathbf{x}_r))$$

Normalization

$$\tilde{D}_p(f, \mathbf{x}_r) = \frac{D_p(f, \mathbf{x}_r)}{\left\| D_p(\cdot, \mathbf{x}_r) \right\|}$$

RIR Synthesis

$$H(f, \mathbf{x}_r) = \sum_{p=1}^{P} S_p(f) \frac{e^{-j2\pi f r_p(\mathbf{x}_r)/c}}{r_p(\mathbf{x}_r)} \tilde{D}_p(f, \mathbf{x}_r)$$

signal | decay & delay | directivity

p : multipole index

$\mathbf{x}_p$: position of p-th multipole

$\mathbf{x}_r$: position of receiver

$\boldsymbol{\Omega}_p(\mathbf{x}_r)$: relative direction of p-th multipole

$S_p(f)$: signal of p-th multipole

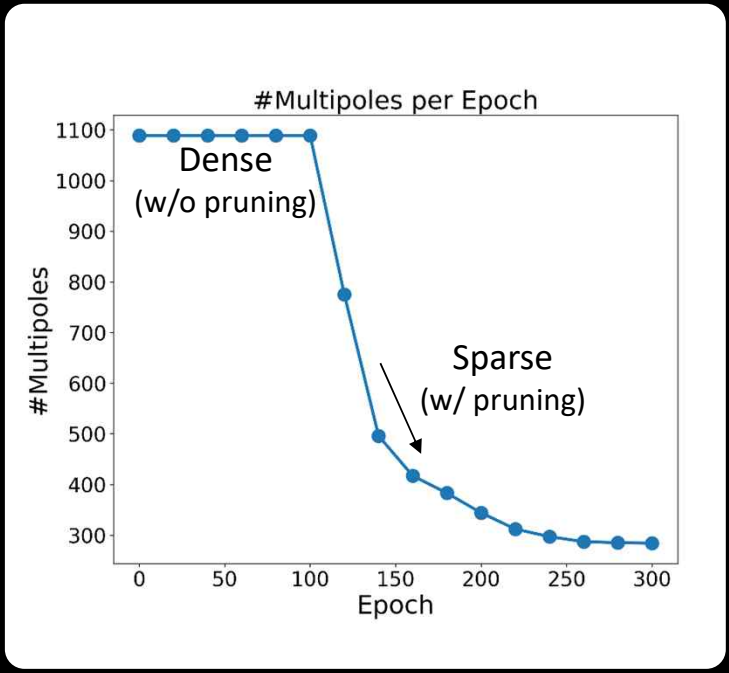
$B_{nm,p}(f)$: SH coefficients of p-th multipole

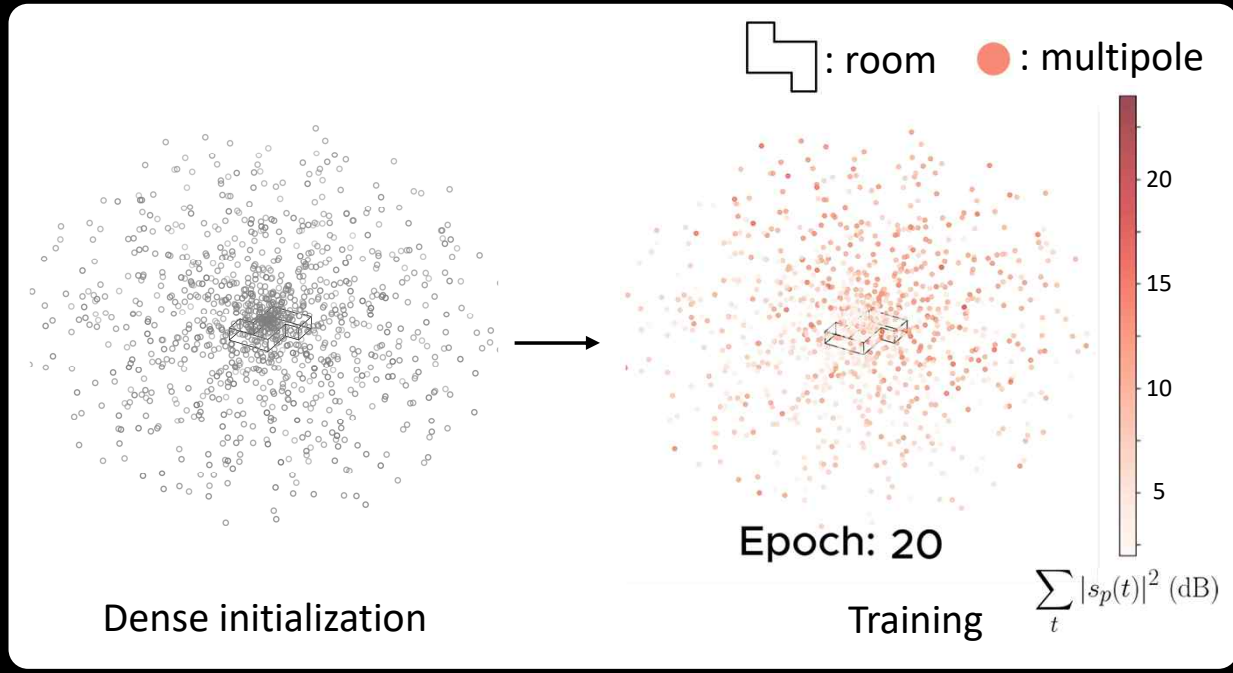
Y_n^m : spherical harmonic basis

$r_p(\mathbf{x}_r) = \|\mathbf{x}_r - \mathbf{x}_p\|$

4. Multipole pruning

- Initialize multipoles densely in space
- Iteratively prune insignificant multipoles during training
  - Every 20 epochs after the first 100 epochs

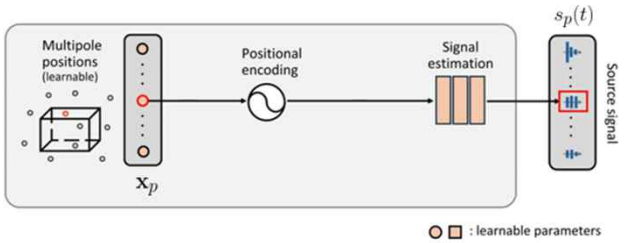




Signal energy-based pruning

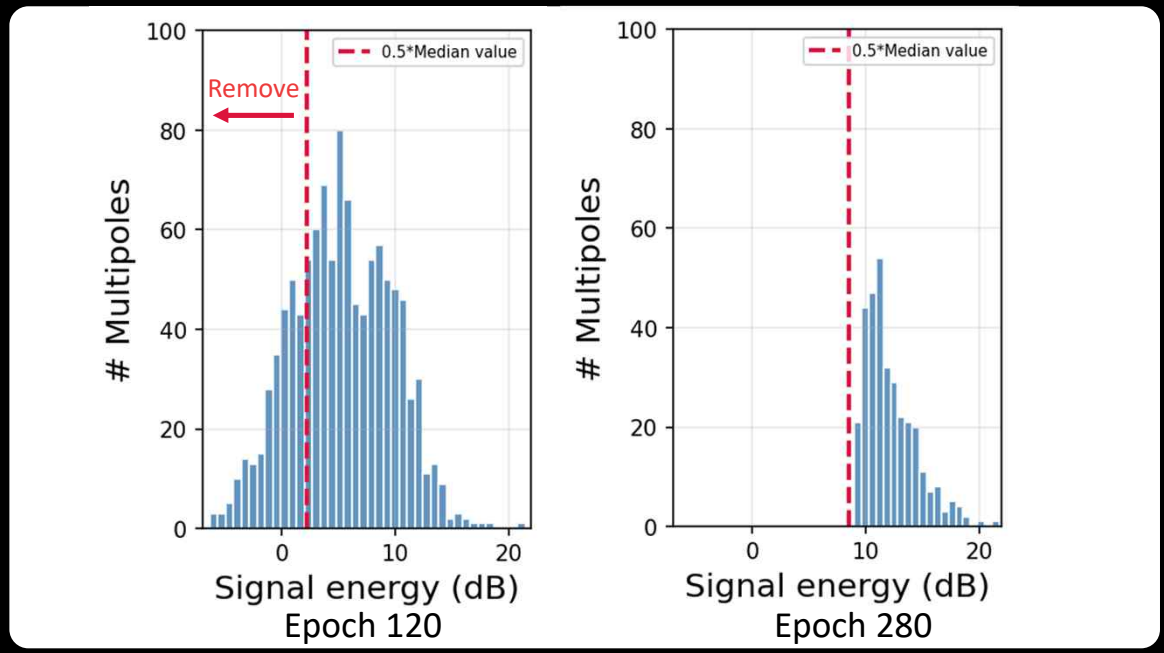
- Signal energy: computed without receiver position
  - Signal branch is receiver-independent & directivity is normalized
- Remove multipoles with signal energy below 50% of the global median

Receiver-independent signal branch



Normalized directivity

$$\tilde{D}_p(f, \mathbf{x}_r) = \frac{D_p(f, \mathbf{x}_r)}{\|D_p(\cdot, \mathbf{x}_r)\|}$$

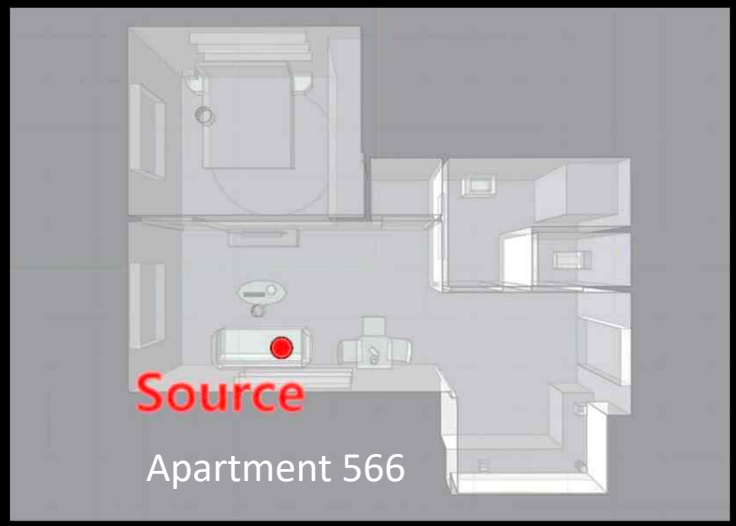


Experiment

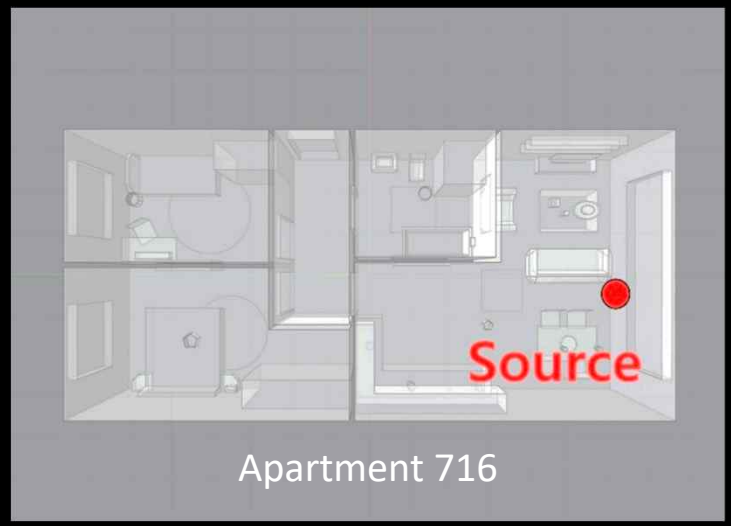
Datasets

- **Scene information**

Scene	Scene complexity	Volume (m^3)	RT60 (s)	Type
MeshRIR[Koyama et al., 2021]	Simple	121	0.38	Real
Apartment 566	Complex	105	0.48	Simulated
Apartment 716	Complex	183	1.8	Simulated



Apartment 566

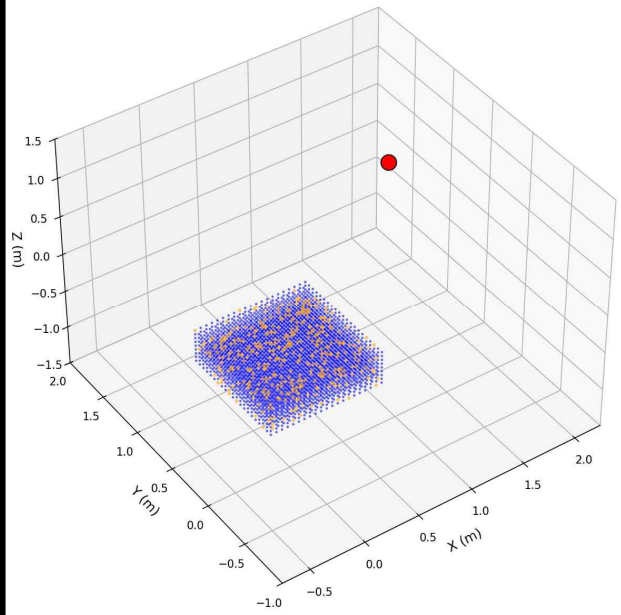


Apartment 716

Datasets

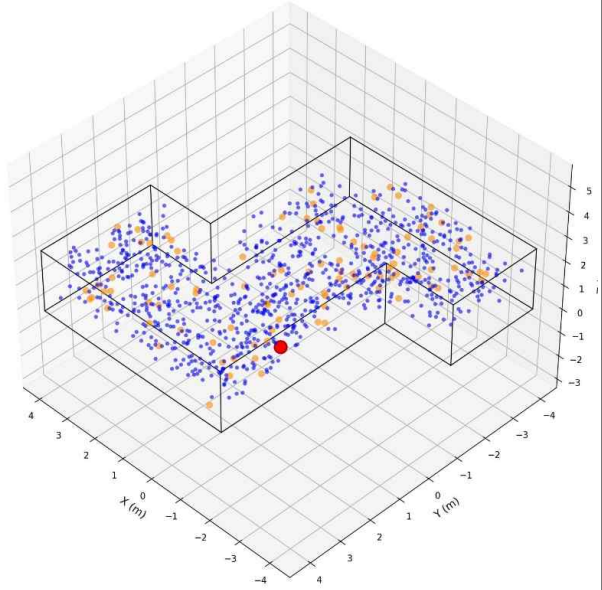
: source : train data : test data

MeshRIR



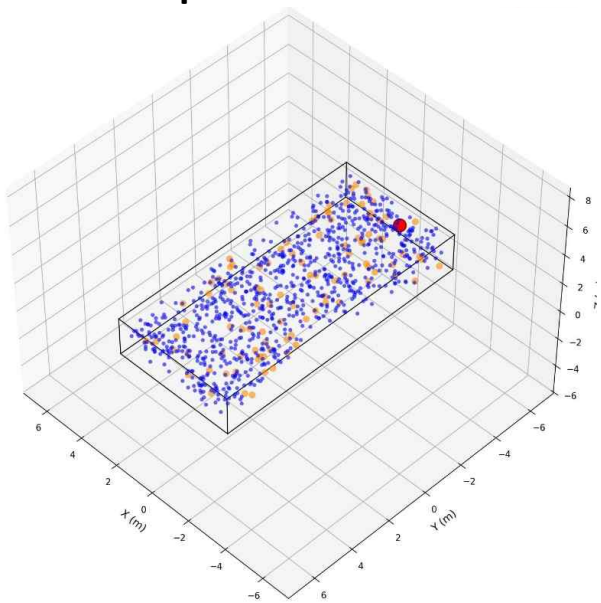
Train : Test	9:1
RIR length	0.1 s (24 kHz)
#RIRs	3969
#RIR(train) / m³	8930.3

Apartment 566



Train : Test	9:1
RIR length	0.1 s (24 kHz)
#RIRs	1000
#RIR(train) / m³	8.6

Apartment 716



Train : Test	9:1
RIR length	0.1 s (24 kHz)
#RIRs	1000
#RIR(train) / m³	4.9

Losses and Metrics

- **Losses and evaluation metrics:** following AVR[Lan et al., 2024]
- **Loss**
  - Spectral loss, Amplitude loss, Phase loss, Waveform loss, Multi-resolution STFT loss, Energy Decay loss
- **Metric**
  - Amplitude error (Amp.)
  - Envelope error (Env., %)
  - T60 error (T60, %)
  - C50 error (C50, dB)
  - EDT error (EDT, ms)

$h^*[n]$: ground-truth impulse response

$h[n]$: predicted impulse response

$H^*[f]$: ground-truth frequency response

$H[f]$: predicted frequency response

$$\text{Amplitude error} = \frac{1}{F}\sum_f \frac{\left|\widehat{|H^*[f]|} - \widehat{|H[f]|}\right|}{\widehat{|H^*[f]|}}$$

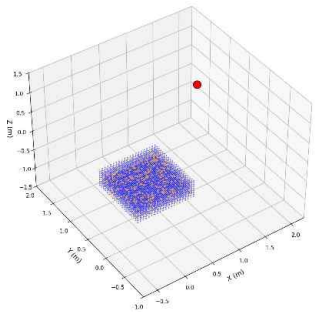
$\widehat{\cdot}$: moving avg.smoothing

$$\text{Envelope error} = \frac{100}{N}\sum_n \frac{|\text{Env}^*[n] - \text{Env}[n]|}{\max_n \text{Env}^*[n]} \qquad \text{Env}[n] = |\text{Hilbert}(h[n])|$$

Performance comparison with existing models

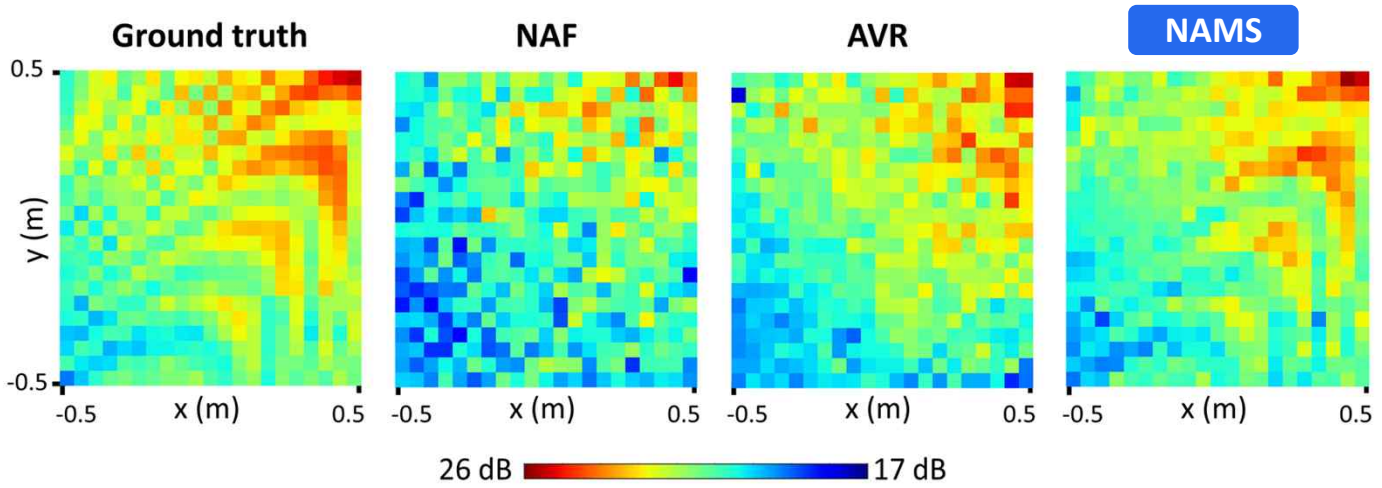
- **[MeshRIR] dataset**
  - NAMS demonstrates superior performance over existing models

MeshRIR	Amp. ↓	Env. ↓	T60 ↓	C50 ↓	EDT ↓	Param. ↓	#points ↓	$T_{Inference}$ (ms) ↓
NAF[Luo et al., 2022]	0.57	1.98	3.5	0.88	31.0	2.7M	-	1.9
AVR[Lan et al., 2024]	0.28	1.44	2.9	0.66	19.4	57.2M	205k	62.5
NAMS	**0.11**	**1.21**	**2.0**	**0.34**	**9.8**	1.8M	225	2.2



MeshRIR

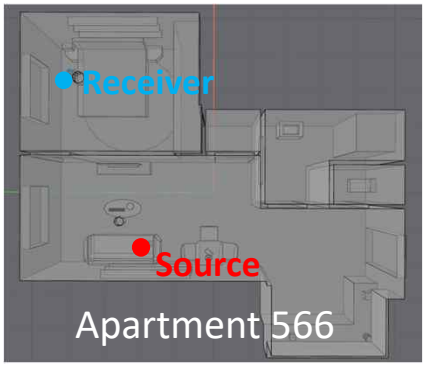
RIR magnitudes:
z=-0.15 m plane
at 4kHz (1/3 octave band)

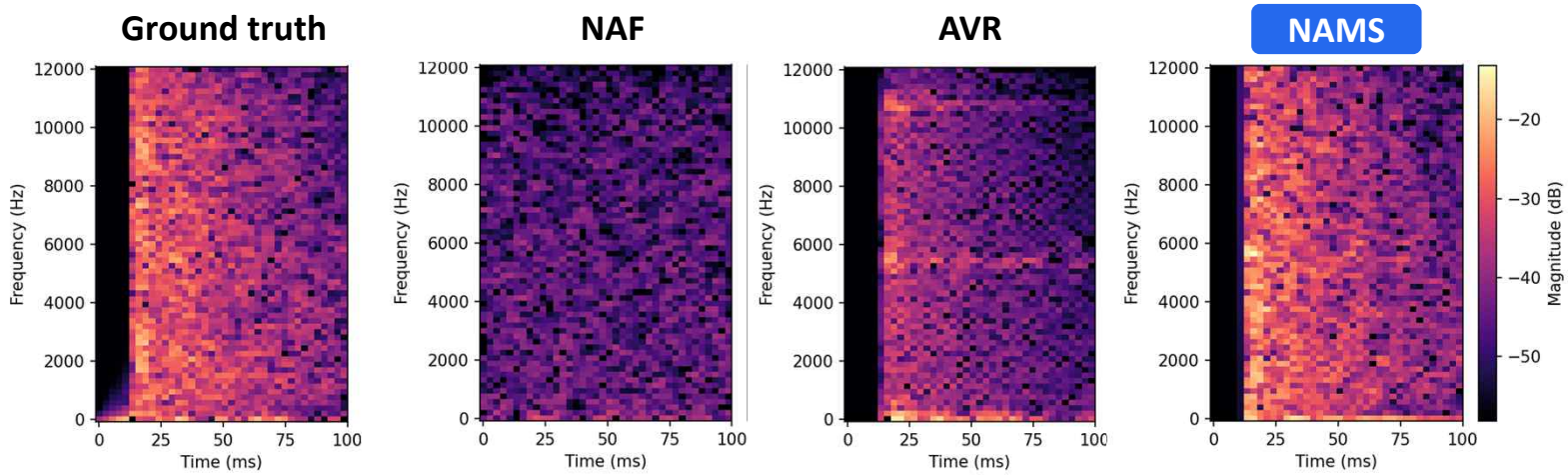


Performance comparison with existing models

- **Apartment 566 & 716 datasets**
  - NAMS performs better in complex environments with sparsely sampled receiver-position RIRs

	Apartment 566					Apartment 716				
	Amp. ↓	Env. ↓	T60 ↓	C50 ↓	EDT ↓	Amp. ↓	Env. ↓	T60 ↓	C50 ↓	EDT ↓
NAF	0.77	5.05	15.2	8.66	22.0	0.76	6.56	21.4	7.35	18.8
AVR	0.46	**4.20**	5.0	1.20	29.8	0.54	**5.79**	9.0	2.46	24.2
NAMS	**0.22**	4.46	**3.4**	**0.48**	**12.0**	**0.30**	6.70	**6.3**	**0.82**	**13.7**

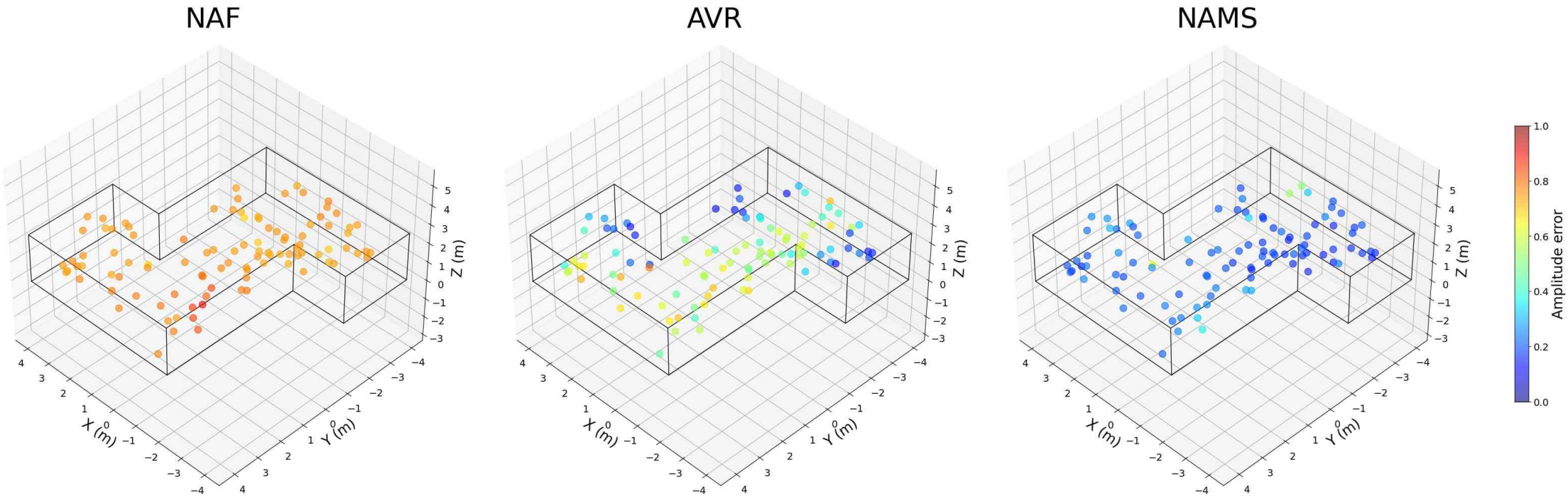




Amplitude errors – Apartment 566

- Amplitude errors at test positions (↓)

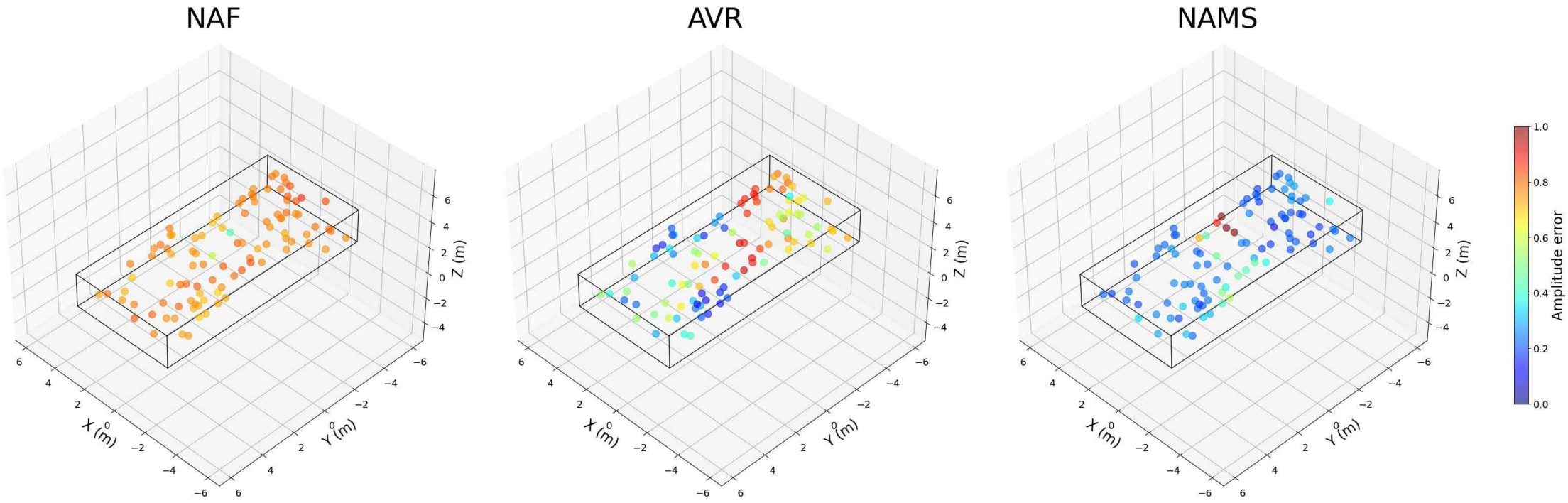
Proposed



Amplitude errors – Apartment 716

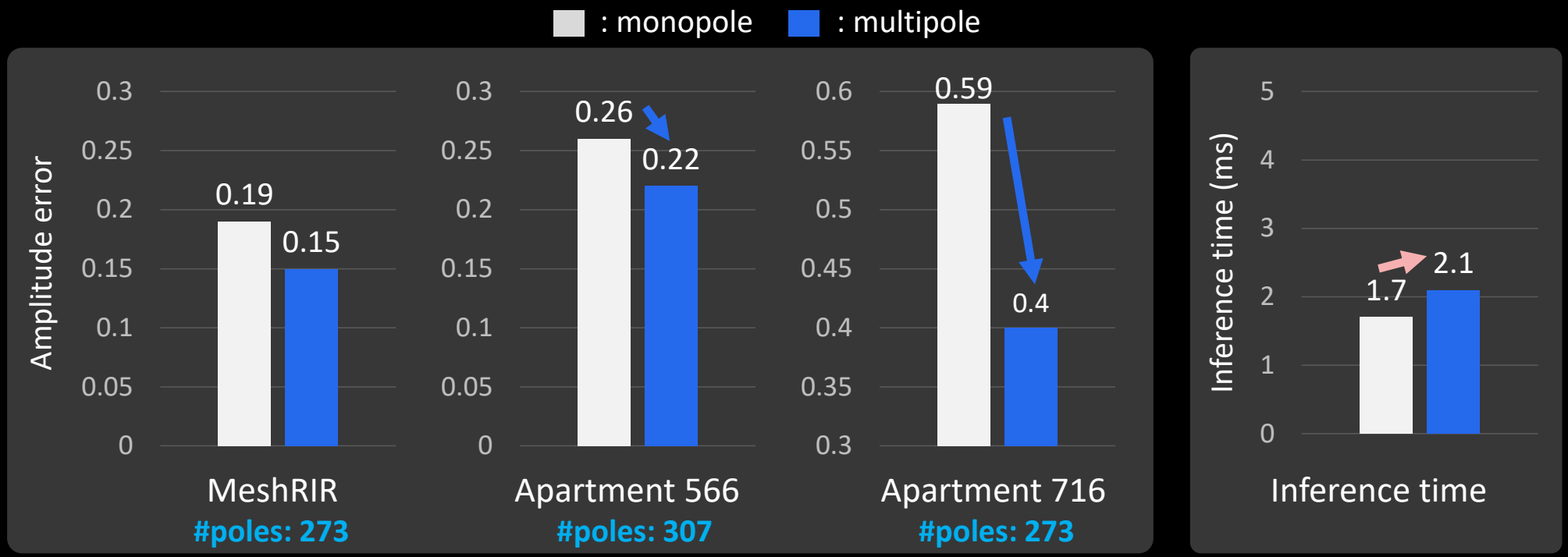
- Amplitude errors at test positions (↓)

Proposed



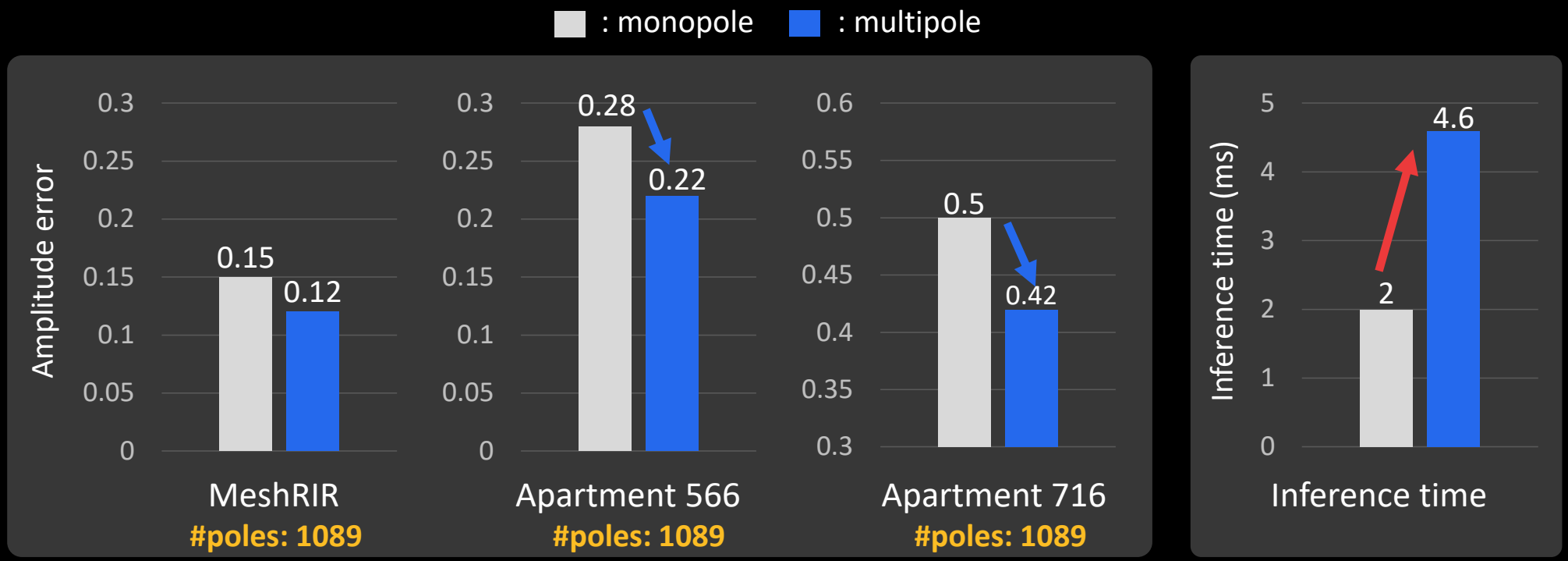
Monopole vs. Multipole

- Comparison under the same number of poles (sparse pole setting)
  - Multipole representation outperforms monopole representation in amplitude error



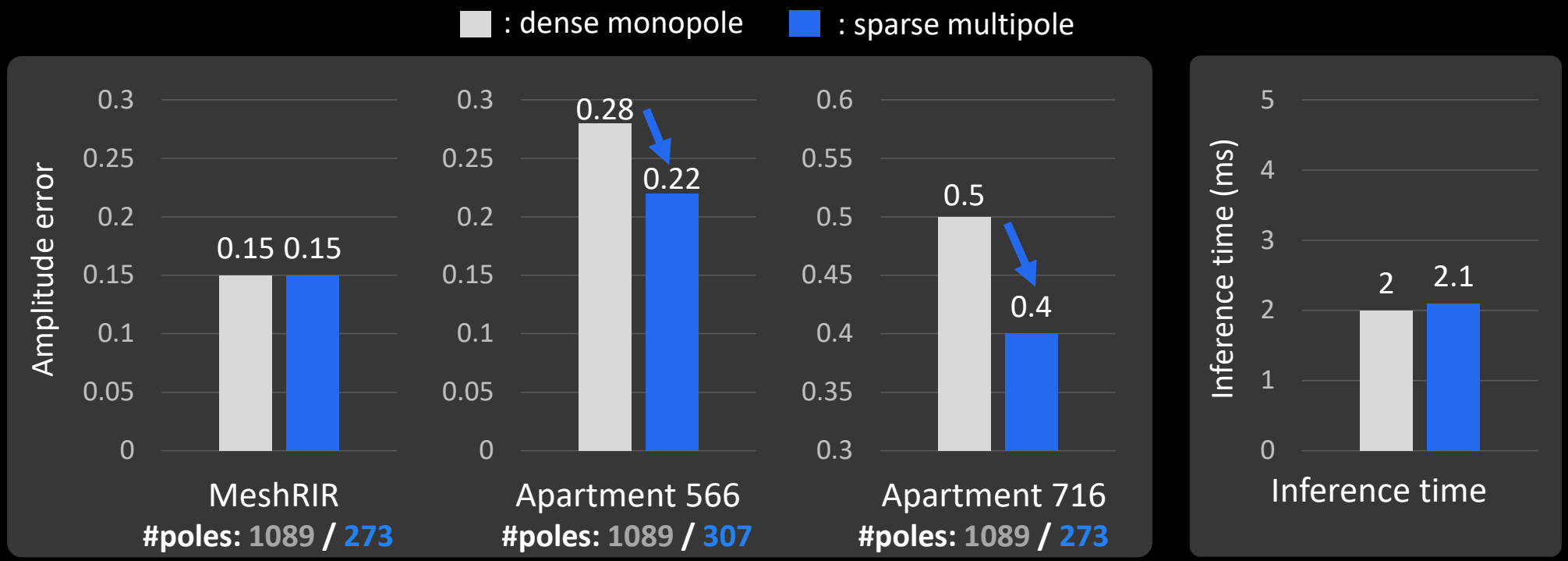
Monopole vs. Multipole

- Comparison under the same number of poles (dense pole setting)
  - Multipole representation provides lower amplitude error, but computational cost increases



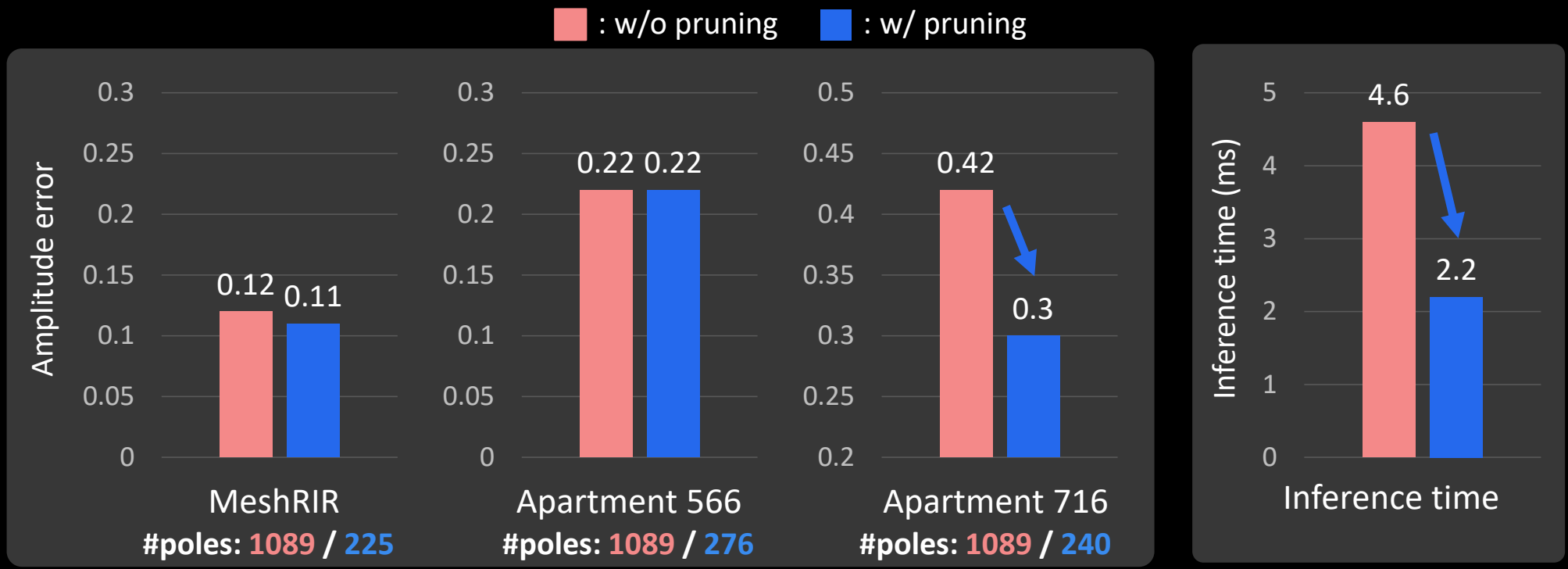
Dense Monopole vs. Sparse Multipole

- **Sparse multipole representation** performs better in complex scenes (Apartment 566, 716)



Effect of pruning

- Comparison with/without pruning under the dense multipole initialization
  - Pruning removes unnecessary multipoles while preserving performance



Conclusion & Limitations

- **Conclusion**
  1. NAMS outperforms existing RIR synthesis models
  2. Multipoles provide **more efficient & rich representations** than monopoles in complex scenes
  3. **Pruning** further reduces inference time while maintaining similar performance
- **Limitations & Future directions**
  1. Synthesis for **arbitrary source positions**
  2. Still requires **dense** RIR measurements

Thank you

Geonwoo Baek & Jung-Woo Choi

Paper

Demo

Github

KAIST EE

ICASSP2026 Barcelona

SMART SOUND LAB

Reference

- N. Raghuvanshi and J. Snyder, “Parametric directional coding for precomputed sound propagation,” ACM Transactions on Graphics (TOG), vol. 37, no. 4, pp. 1–14, 2018.
- A. Luo, Y. Du, M. Tarr, J. Tenenbaum, A. Torralba, and C. Gan, “Learning neural acoustic fields,” in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2022, vol. 35, pp. 3165–3177.
- Z. Lan, C. Zheng, Z. Zheng, and M. Zhao, “Acoustic volume rendering for neural impulse response fields,” in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2024, vol. 37, pp. 44600–44623.
- W. Jin and W. Kleijn, “Theory and design of multizone sound field reproduction using sparse methods,” IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP), vol. 23, no. 12, pp. 2343–2355, 2015.
- Antonello, N., De Sena, E., Moonen, M., Naylor, P. A., and Van Waterschoot, T, “Room impulse response interpolation using a sparse spatio-temporal representation of the sound field,” IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP), vol. 25, no. 10, pp. 1929 1941, 2017.
- X. Karakonstantis, D. Caviedes-Nozal, A. Richard, and E. Fernandez-Grande, “Room impulse response reconstruction with physics-informed deep learning,” The Journal of the Acoustical Society of America (JASA), vol. 155, no. 2, pp. 1048–1059, 2024.
- H. Bi and T. Abhayapala, “Point neuron learning: a new physics-informed neural network architecture,” EURASIP Journal on Audio, Speech, and Music Processing, vol. 2024, no. 1, pp. 56, 2024.
- S. Koyama, T. Nishida, K. Kimura, T. Abe, N. Ueno, and J. Brunnstrom, “MeshRIR: A dataset of room impulse responses on meshed grid points for evaluating sound field analysis and synthesis methods,” in Proc. IEEE workshop on applications of signal processing to audio and acoustics (WASPAA). IEEE, 2021, pp. 1–5.