POLITECNICO MILANO 1863

AI-based spatial audio

Research at Image and Sound Processing Lab – Politecnico di Milano

10-09-25 Fabio Antonacci

Motivations and outline

01

Motivations

- The Shannon-Nyquist spatial sampling theorem imposes an unfasible number of measurements, if we wish to adopt «standard» digital signal processing pipelines.
    1. Soundfield reconstruction in a 1m x 1m x 1m cube. Maximum frequency: 1000 Hz @ 343 m/s → λ=0.343 m → 7 measuring points along each dimension → >300 sampling points.
    2. Given a spherical microphone array of radius *r*, for a perfect spherical harmonic reconstruction up to a frequency *f*, all modes up to $N_{\max} = \left\lceil \frac{2\pi f r}{c} \right\rceil$ have to be captured, corresponding to $M_{\min} = (N_{\max} + 1)^2$ → quadratic increase of $M_{\min}$ with the maximum frequency and array size.
- Many inverse problems related to spatial audio are ill conditioned. Example: reconstructing the velocity field of a vibrating surface from the radiated soundfield. A linear operator (KH integral) can be used for this purpose. Unfortunately, this problem can be tackled only in simplistic cases.

Adopt ML to regularize the problem

Outline

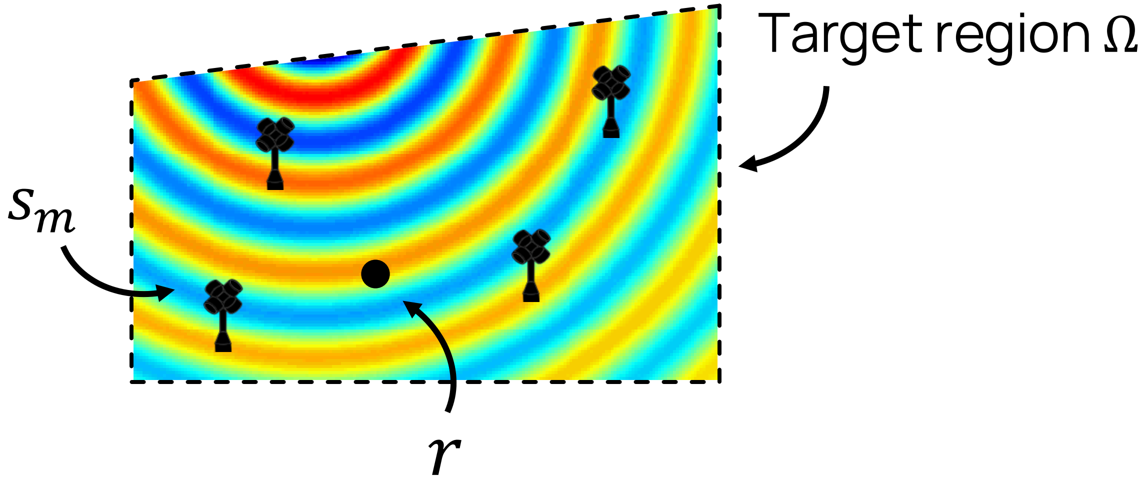
1. Soundfield reconstruction over an extended region
    - Complex-valued NNs
    - Physics-informed NNs
    - Diffusion-based NNs
2. Nearfield acoustic holography
    - Complex-valued physics-informed NNs
    - Physics-Informed Neural Network-driven Sparse Field Discretization method (PINN-SFD)
3. Upsampling of spherical microphone array measurements

Soundfield reconstruction over an extended region

02

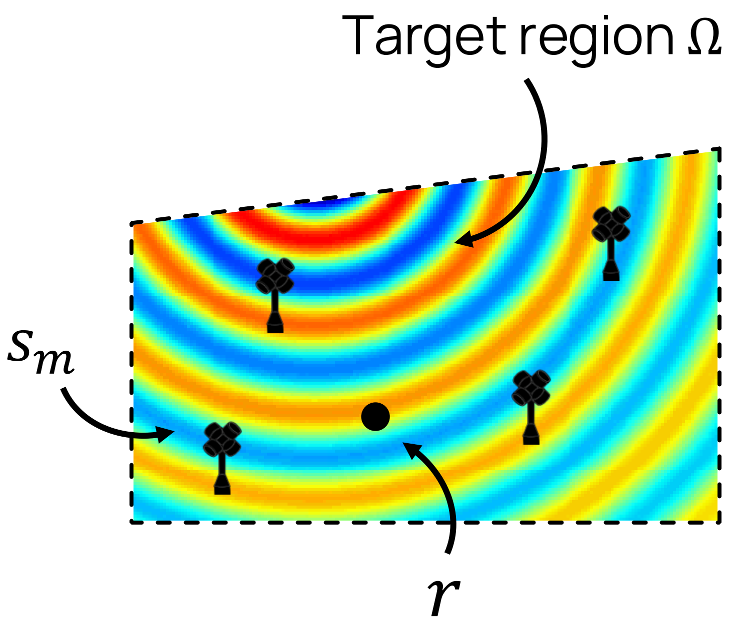
Sound field reconstruction over an extended area
Problem formulation

Estimate sound pressure distribution $u(\boldsymbol{r})(\boldsymbol{r} \in \Omega)$ from M microphone observations $\{s_m\}_{m=1}^{M}$



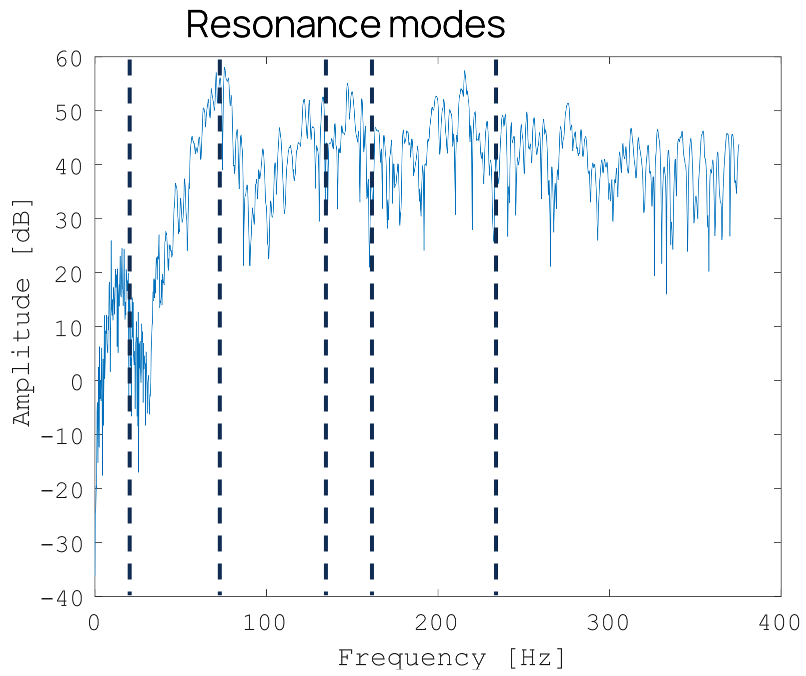
Soundfield reconstruction
Contextualization

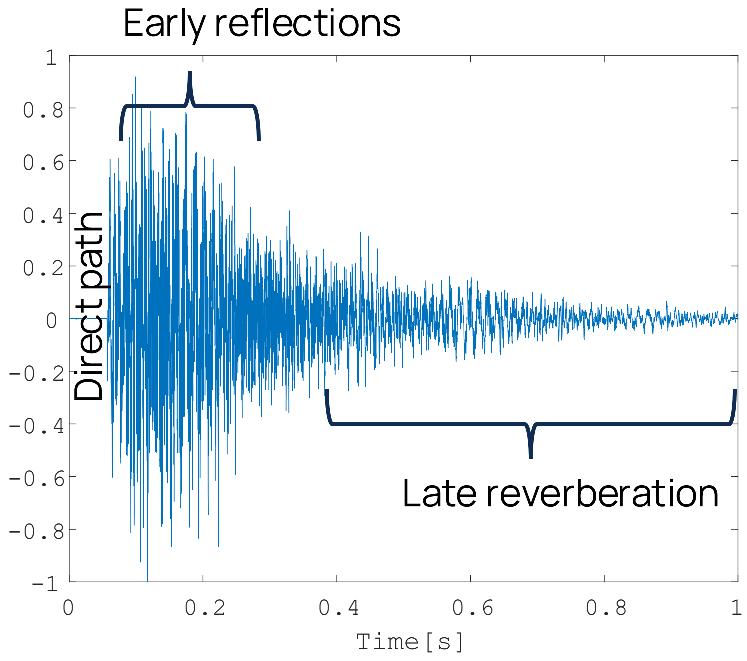
Goal: reconstruct the soundfield in a region from a sparse set of measurements (i.e. below Nyquist limit).



Representations of the acoustic response of a room:

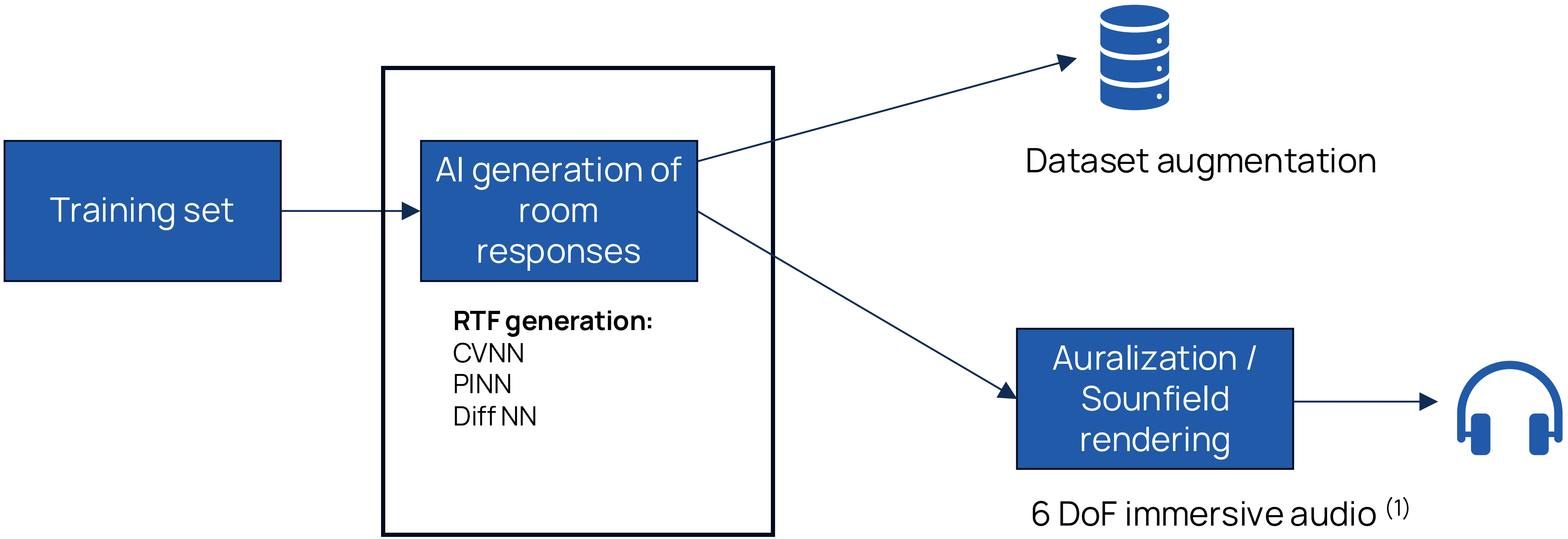
- **Room Transfer Function (RTF)**: frequency domain
- **Room Impulse Response (RIR)**: time domain





Different sampling requirements for different sensor types: HOMs require fewer sampling points than omni mics

Soundfield reconstruction
Contextualization



(1) Repertorium project, funded by EU in the RIA programme, GA number: 101095065 https://repertorium.eu/

Soundfield reconstruction
Complex-valued Neural Networks

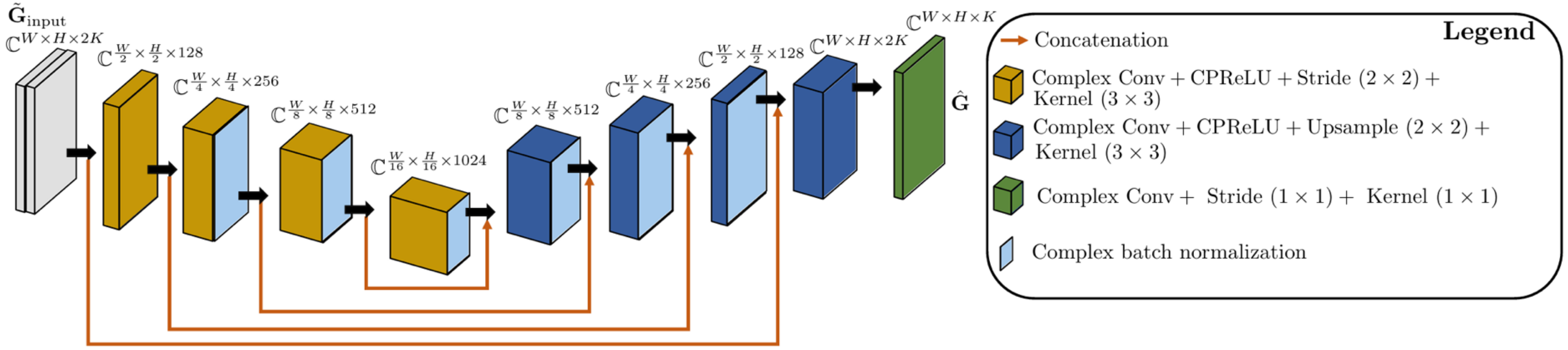
Reconstruction of the RTF over an extended area is typically approached as an image impainting approach, e.g. using encoder-decoder architectures. But… **RTF is complex valued**.

Two approaches:

1. Separate inputs for magnitude and phase: excellent reconstruction on the magnitude, bad on the phase
2. Separate inputs for real and imaginary parts: good reconstruction accuracy, but independent reconstructions on the two components may yield inaccurate phase if some countermeasures are not adopted in the loss function.

Solution: complex valued neural network

Conditioning of the learning on the points where measurements are available



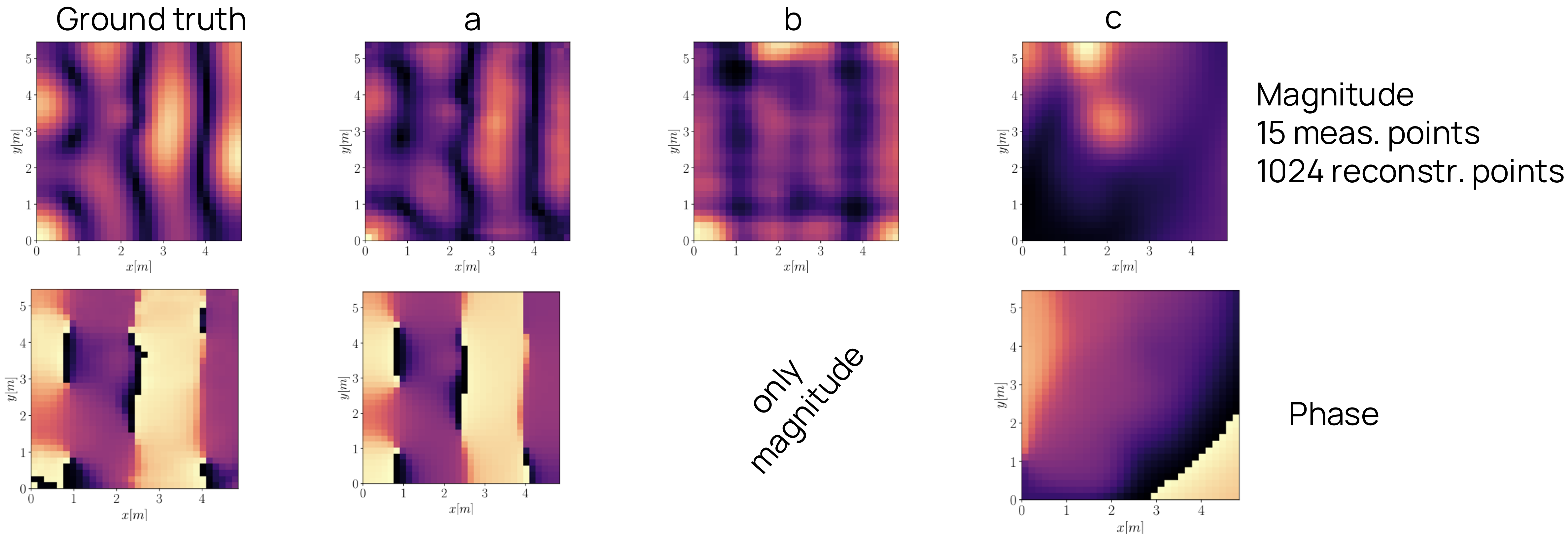
Setup:

- Eval: 15000 simulated rooms,
- Training: 5000 simulated rooms.
- 0.4s < T_{60} < 1.6s
- Room B of ISOBEL for real data evaluation
- [5,10,15,35,55] measurement points out of ~1000 virtual mic positions

F. Ronchini, L. Comanducci, M. Pezzoli, F. Antonacci, A. Sarti,. Room Transfer Function Reconstruction Using Complex-valued Neural Networks and Irregularly Distributed Microphones, EUSIPCO 2024

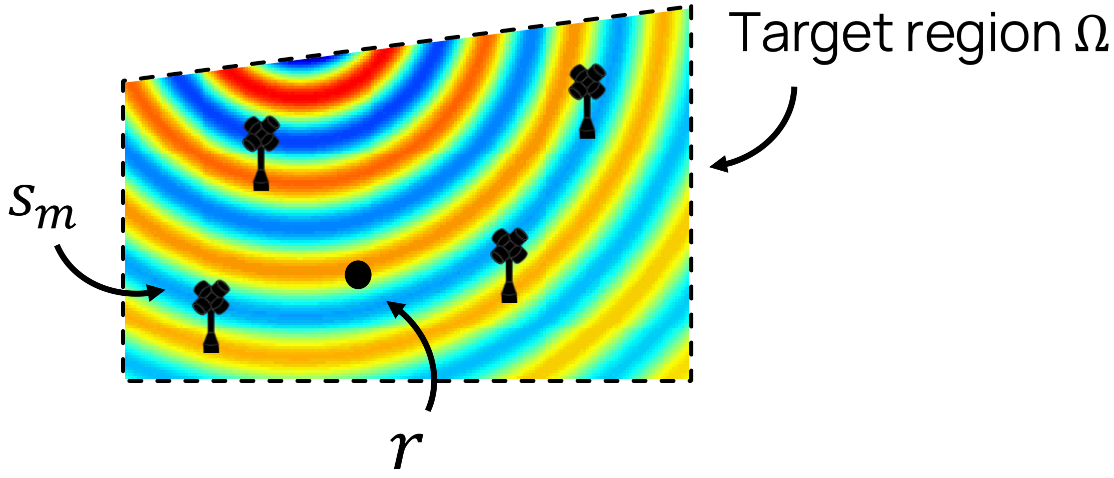
Soundfield reconstruction

Complex-valued Neural Networks



a) F. Ronchini, L. Comanducci, M. Pezzoli, F. Antonacci, A. Sarti,. Room Transfer Function Reconstruction Using Complex-valued Neural Networks and Irregularly Distributed Microphones, EUSIPCO 2024
b) Lluis, F., Martinez-Nuevo, P., Bo Møller, M., & Ewan Shepstone, S. (2020). Sound field reconstruction in rooms: Inpainting meets super-resolution. *The Journal of the Acoustical Society of America*, *148*(2), 649-659.
c) N. Ueno, S. Koyama, and H. Saruwatari, “Kernel ridge regression with constraint of Helmholtz equation for sound field interpolation,” in Int. Workshop Acoust. Signal Enhanc. IEEE, 2018

Sound field reconstruction
Physics constraints



General interpolation technique:

- Represent u using model parameters $\boldsymbol{\theta}$
- Regularization $\mathcal{R}(\boldsymbol{\theta})$
- Solve optimization problem

$$\underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathcal{L}(\{u(\boldsymbol{r}_m;\boldsymbol{\theta})\}_{m=1}^{M}, \boldsymbol{s}) + \mathcal{R}(\boldsymbol{\theta})$$

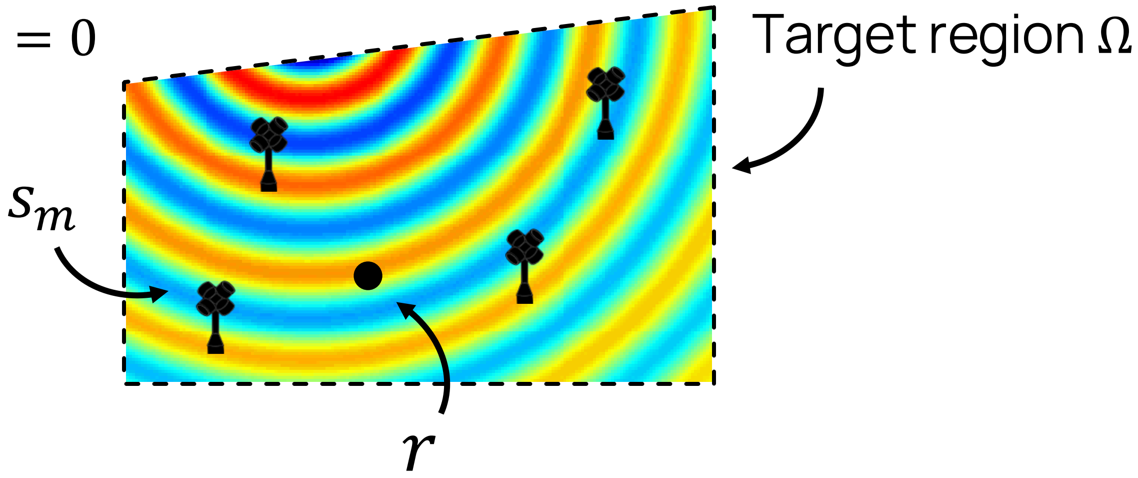
Prior knowledge: acoustics!

Sound field reconstruction
Physics constraints

Target function should satisfy

- Wave equation (time domain) $\left(\nabla_{\boldsymbol{r}}^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right) U(\boldsymbol{r}, t) = 0$

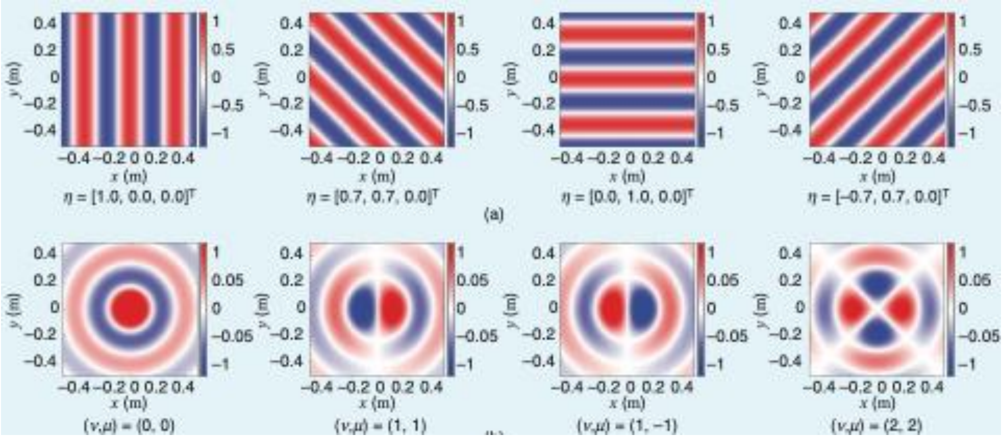
- Helmholtz equation (frequency domain) $(\nabla_{\boldsymbol{r}}^2 + k^2) u(\boldsymbol{r}, \omega) = 0$



Sound field reconstruction:
Common basis

Solutions of wave equation:

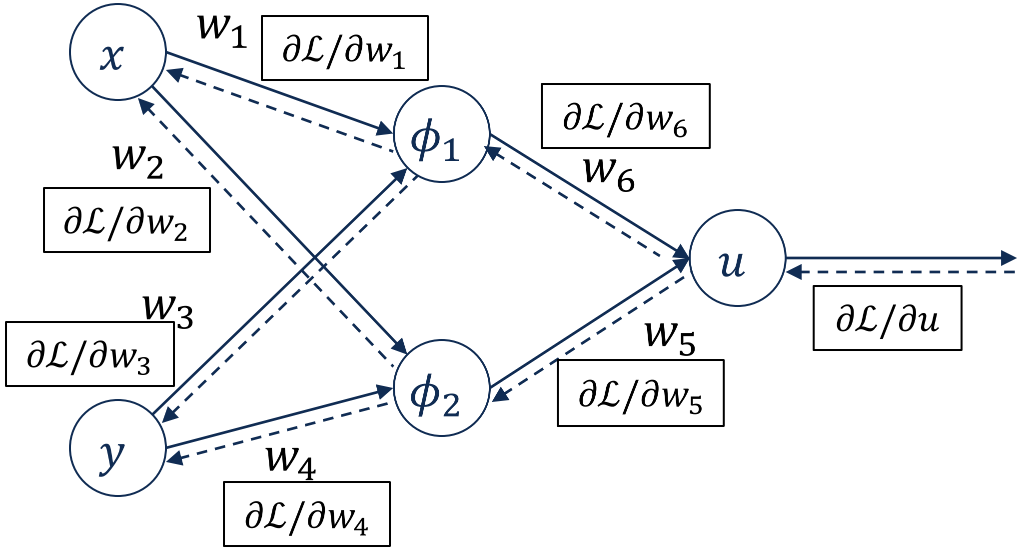
- Plane waves $u(\boldsymbol{r},\omega) = \int_{\mathbb{S}^2} \tilde{u}(\boldsymbol{\eta},\omega) e^{\iota k \langle \boldsymbol{\eta},\boldsymbol{r} \rangle} d\boldsymbol{\eta}$
- Spherical waves $u(\boldsymbol{r},\omega) = \sum_{\nu}\sum_{\mu} \dot{u}_{\nu,\mu}(\omega) j_{\nu}(k\|\boldsymbol{r}-\boldsymbol{r}_o\|) Y\left(\frac{\boldsymbol{r}-\boldsymbol{r}_o}{\|\boldsymbol{r}-\boldsymbol{r}_o\|}\right)$
- Equivalent sources $u(\boldsymbol{r},\omega) = \int_{\partial\Omega} \breve{u}(\boldsymbol{r}',\omega) \frac{e^{\iota k(\boldsymbol{r}-\boldsymbol{r}')}}{4\pi\|\boldsymbol{r}-\boldsymbol{r}'\|} d\boldsymbol{r}'$



Picture from: Koyama S, Ribeiro J., Nakamura T., Ueno N., Pezzoli M. ’’Physics-informed Machine Leraning for Sound Field Estimation"IEEE Signal Processing Magazine, vol. 41, no. 6, pp. 60-71, Nov. 2024

Physics-informed neural networks (PINNs)

Introduced in [1] in to solve PDE



PDE Loss

$$\mathcal{L}(u) = \begin{cases} \dfrac{\partial u}{\partial x} + \dfrac{\partial u}{\partial y} \cdots \\ \|u - \hat{u}\|^2 \end{cases}$$

Data loss

We can exploit the wave equation!

[1] Raissi, Maziar, et al. "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations." *Journal of Computational physics* (2019).

Sound field reconstruction: PINN

Use NN with physics-informed training

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(f(\boldsymbol{\theta}, \{r_m\}_{m=1}^{M}), \boldsymbol{s}) + \cancel{\mathcal{R}(\boldsymbol{\theta})}$$

NN parameters

NN

Regularization by

- Model structure
- PDE Loss

Sound field reconstruction:
PINN – PI Loss function

- Use physics-informed training

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(f(\boldsymbol{\theta}, \{r_m\}_{m=1}^{M}), \boldsymbol{s}) + \cancel{\mathcal{R}(\boldsymbol{\theta})}$$

NN parameters

NN

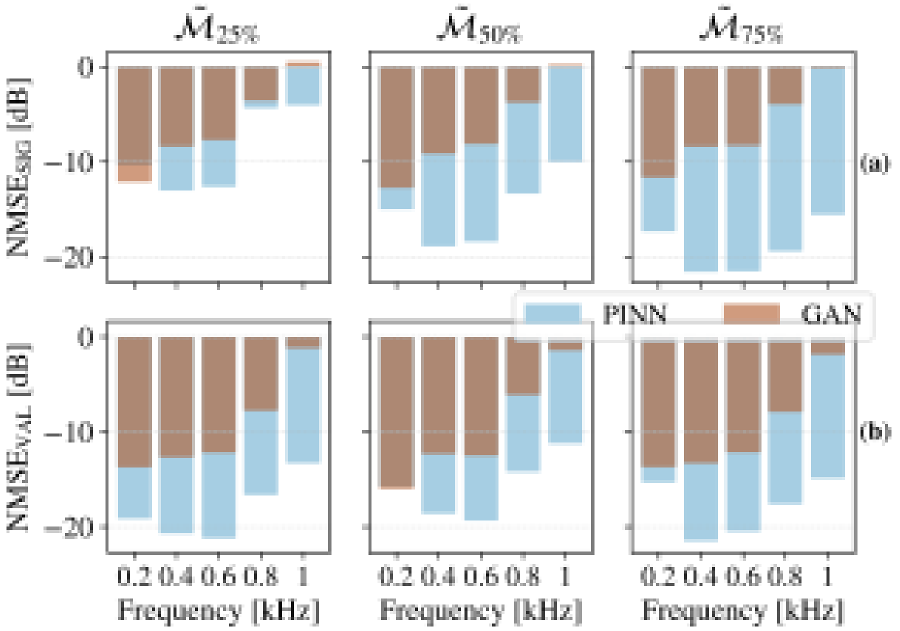
- PI-Loss function

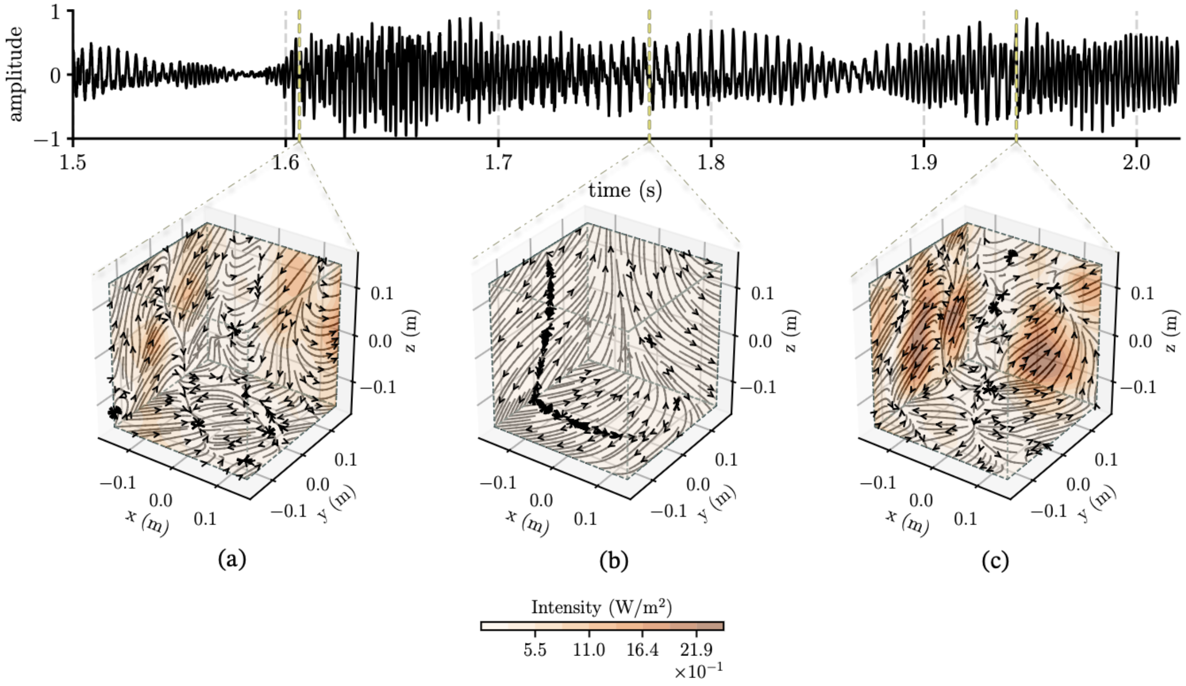
Estimate by NN

$$\mathcal{L} = \frac{1}{M}\sum_{m} \|\hat{u}_m - u_m\|_2^2 + \lambda \frac{1}{N}\sum_{n=1}^{N} \left\| \nabla^2 \hat{u}_n - \frac{1}{c^2}\frac{\partial^2 \hat{u}_n}{\partial t^2} \right\|_2^2$$

Sound field reconstruction:
PINN – Reconstruction of speech signals

- Reconstruction results on real arbitrary sound fields of speech signals
  - Comparison with GAN[x]





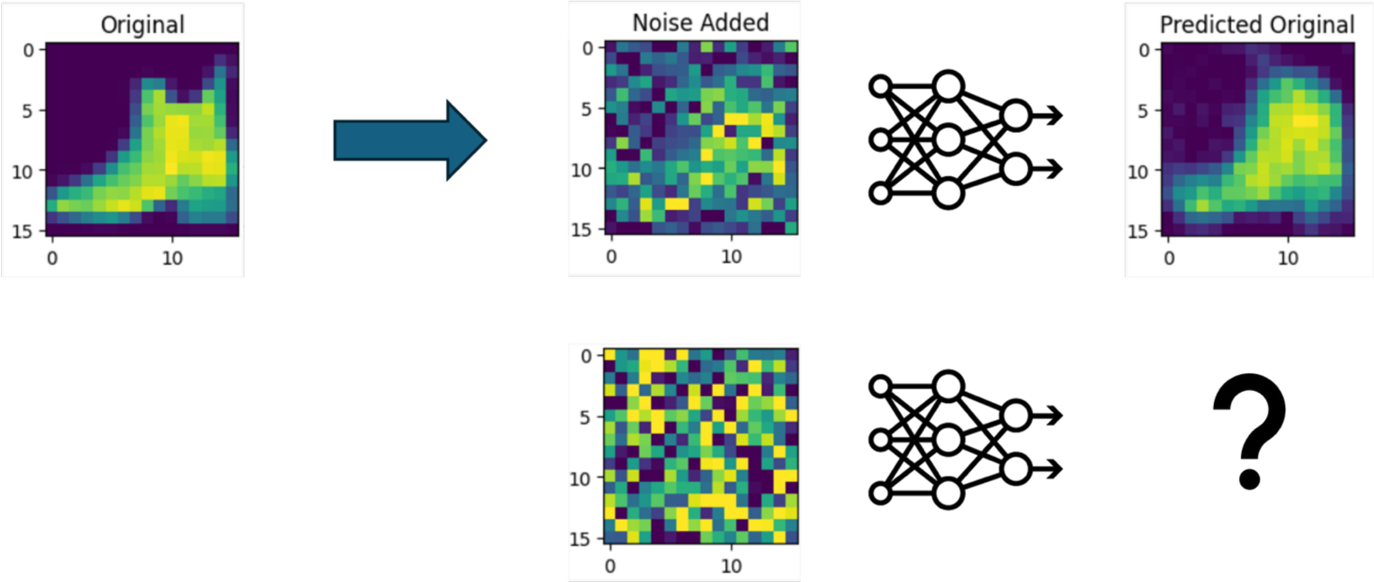
[1] X. Karakonstantis, E. Fernandez-Grande, Generative adversarial networks with physical sound field priors. J. Acoust. Soc. Am. 154(2), 1226–1238 (2023)

[2] M. Olivieri, X. Karakonstantis, M. Pezzoli, F. Antonacci, A. Sarti, and E. Fernandez-Grande, “Physics-informed neural network for volumetric sound field reconstruction of speech signals,” EURASIP J. Audio, Speech, Music Processing, vol. 2024, 2024, Art. no. 42.

Diffusion models

Basic idea: add noise to input data, and then use a NN (e.g. the **U-Net**) to separate the images from the noise (i.e., **denoising**).

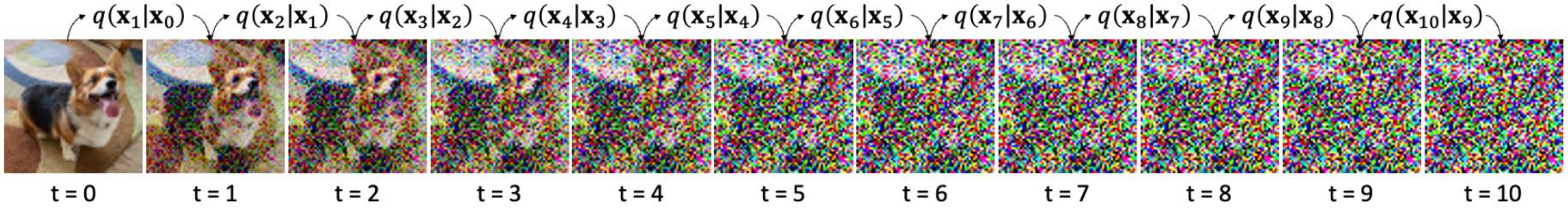
Could we then feed the model noise and create a relevant data?



NVIDIA Generative AI with Diffusion Models workshop - Prof. Dr. Andras Hajdu

Diffusion models – Forward process

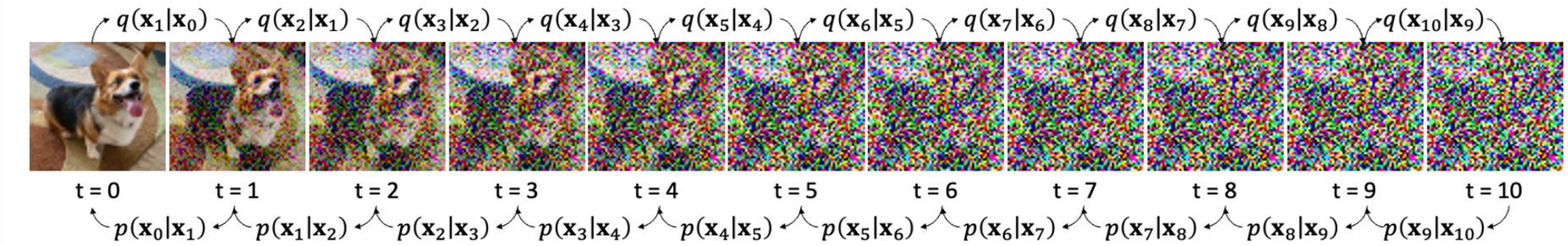
Solution: rather than adding noise to the data all at once, add a small amount of noise multiple times (*forward process*). Then use the neural network on a noisy image multiple times to generate new data (*reverse process*).



Given a data point sampled from a real data distribution, in the forward diffusion process we add small amount of Gaussian noise to the sample in steps (Markov chain), producing a sequence of noisy samples. The step sizes are controlled by a **variance schedule**.

Diffusion models – Reverse process

Solution: rather than adding noise to the data all at once, add a small amount of noise multiple times (*forward process*). Then use the neural network on a noisy image multiple times to generate new data (*reverse process*).

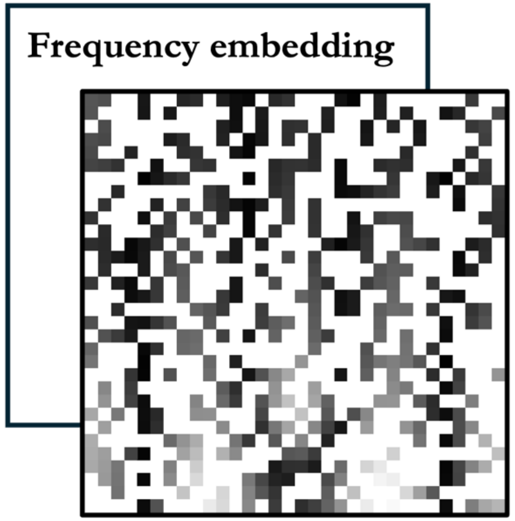


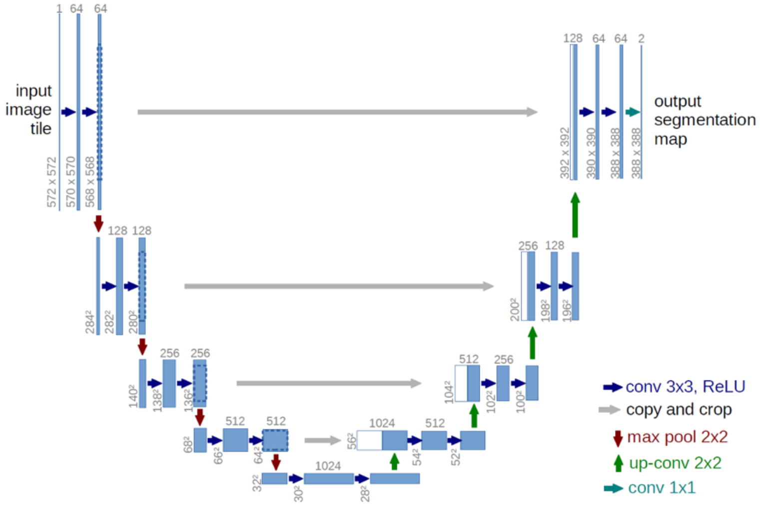
Now, we want to try to reverse the q distribution in order to remove noise, using distribution p. Since it is challenging to know the exact model, we define p as a Markov chain with learned Gaussian transitions starting at $p(\mathbf{x}_T) = N(\mathbf{x}_T; \mathbf{0}, \mathbf{I})$.

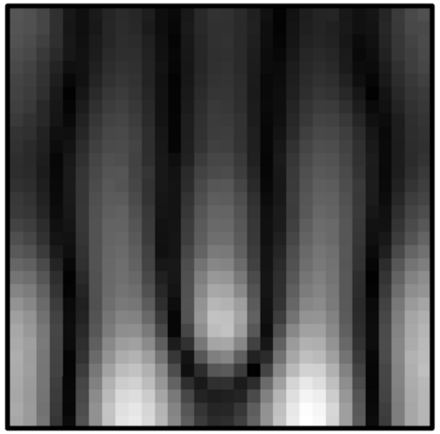
NVIDIA Generative AI with Diffusion Models workshop - Prof. Dr. Andras Hajdu

Sound field reconstruction: Diffusion models - Pipeline

- **Model**: Palette – diffusion denoising probabilistic model, designed for image-to-image translation tasks
- **Architecture**: U-Net – convolutional autoencoder initially designed for medical images processing
- **Input**: concatenation of sound field at measurement positions and frequency embedding, encoding a considered frequency (used for conditioning) - Noise injected at the unknown positions
- **Output**: reconstructed sound field
- **Loss**: Mean Squared Error



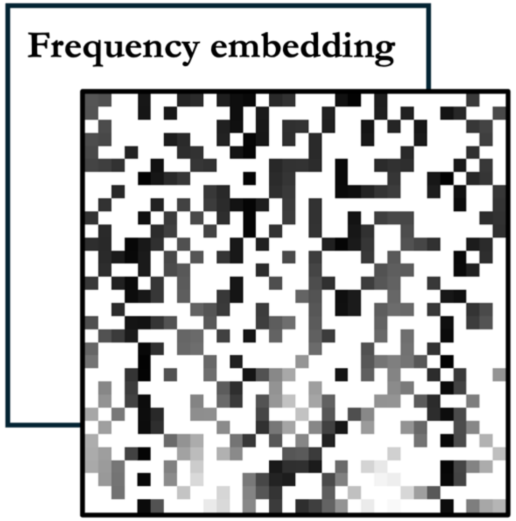


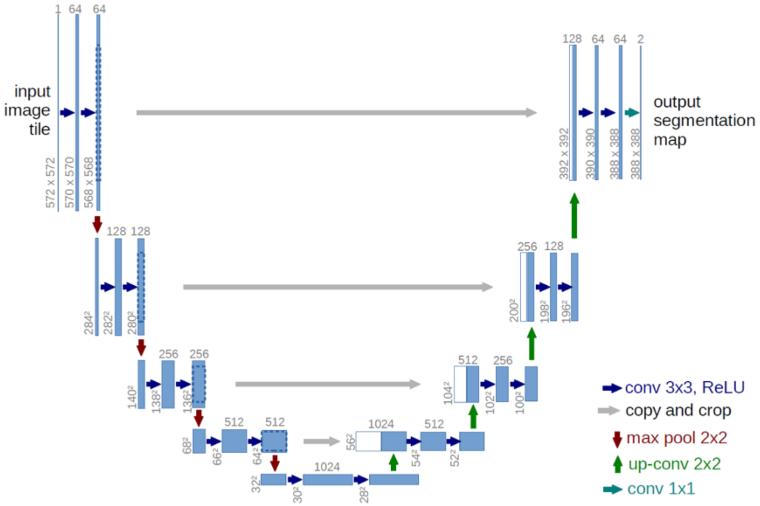


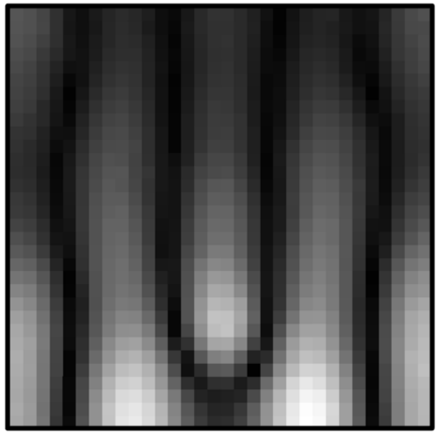
Miotello, Federico, et al. "Reconstruction of sound field through diffusion models." ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024.

Sound field reconstruction: Diffusion models - Setup

- Training carried out each frequency at a time (considered frequency range 30-300Hz) for 10k epochs
- Simulated **training data set**: frequency response (30-300Hz) of 10000 rooms
- Simulated **testing data set**: Frequency response (30-300Hz) of 250 rooms
- Room dimensions are random, with floor area 20-60m2
- T60 fixed to 0.6s
- Number of available microphones: **64 - 128 - 256 - 512**

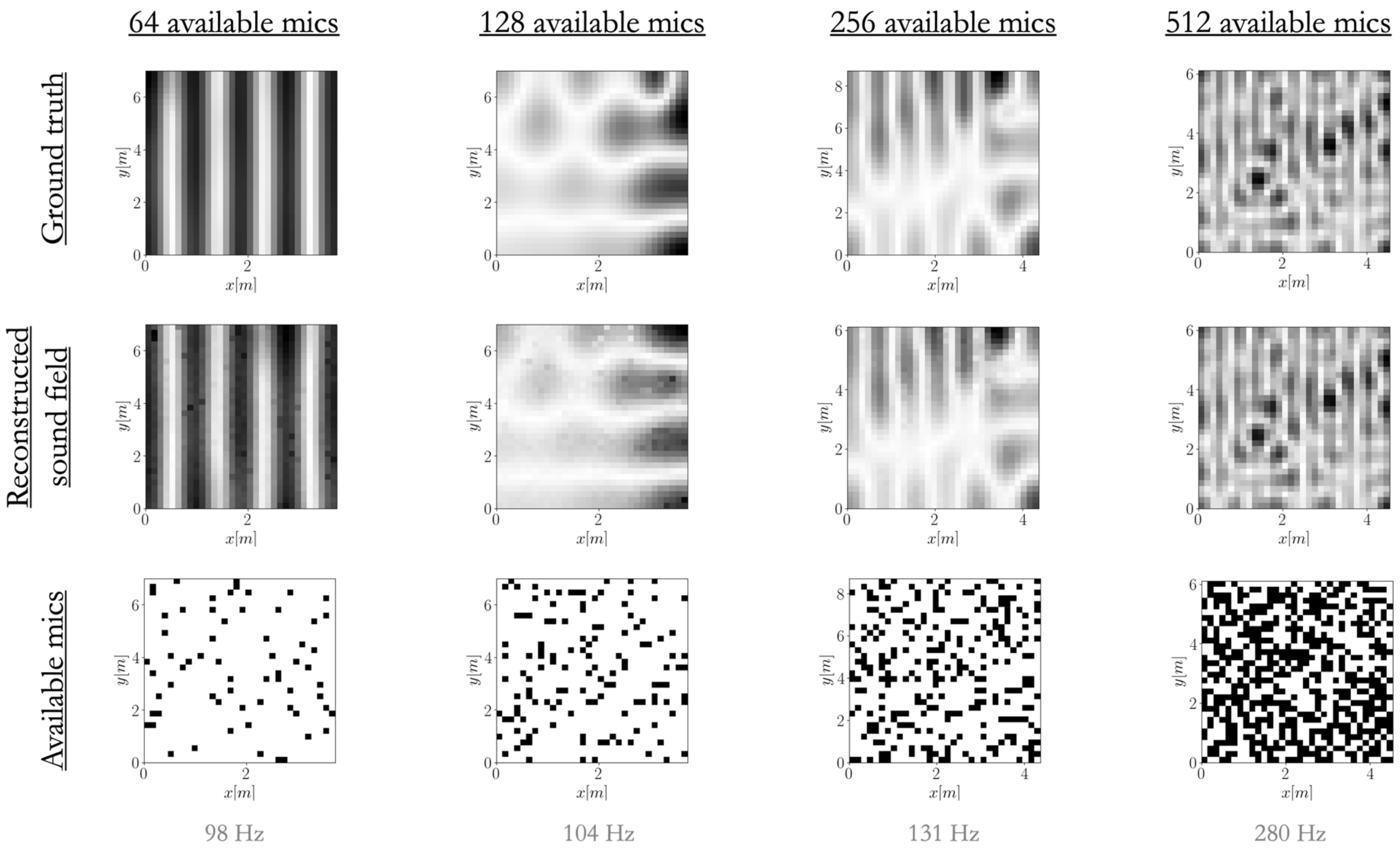






Miotello, Federico, et al. "Reconstruction of sound field through diffusion models." ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024.

Sound field reconstruction: Diffusion models - Results



Pros:
- Good results
- Generalization capabilities
- Easy to train (no complicated loss)

Cons:
- Limited to magnitude reconstruction
- Not using acoustic priors
- Training needs a lot of data
- One frequency at a time

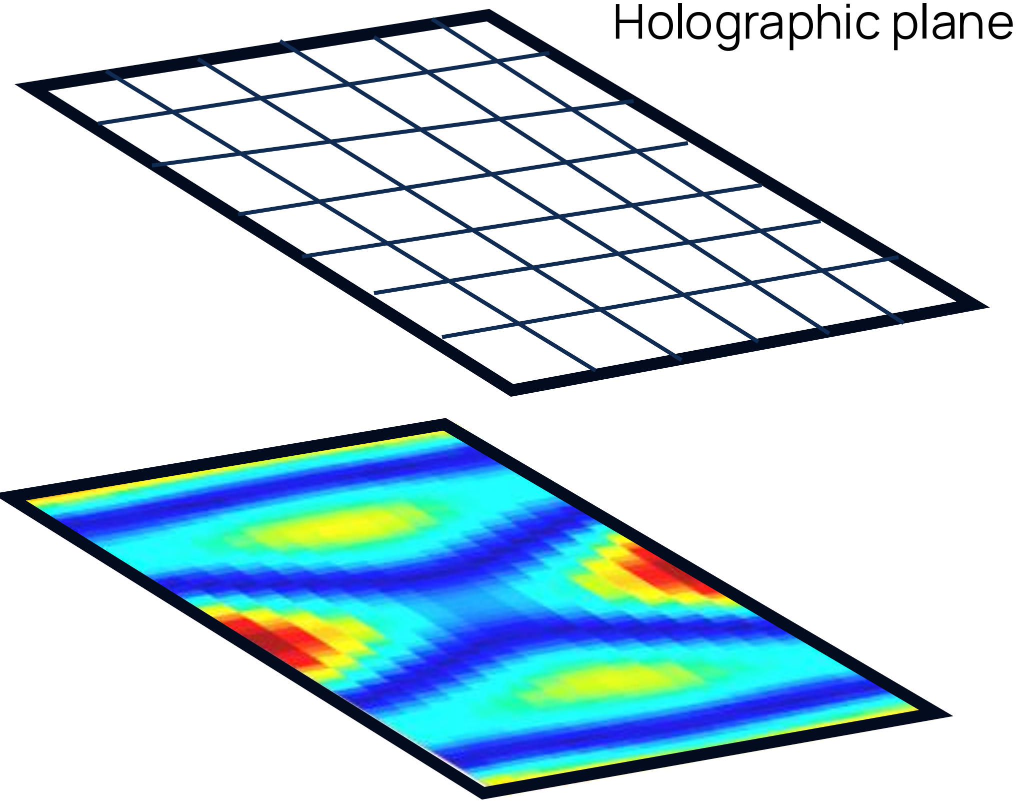
Miotello, Federico, et al. "Reconstruction of sound field through diffusion models." ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024.

Nearfield Acoustic Holography

03

Inverse Nearfield Acoustic Holography
Problem statement

Goal: reconstruct the velocity $v(x',y')$ on the vibrating surface starting from the pressure field $p(x,y)$ measured on the holographic plane



Direct NAH: Kirchhoff-Helmholtz integral

$$p(x,y,\omega) = \int_{\mathcal{S}} p(x',y')\frac{\partial}{\partial \boldsymbol{n}} g_\omega(x,y,z,x',y',z')d\boldsymbol{s}$$

$$-j\omega\rho_0 \int_{\mathcal{S}} v_{\boldsymbol{n}}(x',y')\, g_\omega(x,y,z,x',y',z')d\boldsymbol{s}$$

Inverse NAH:

$\hat{v}_{\boldsymbol{n}}(x',y') \approx \Gamma^{-1}[p(x,y)]$ discrete estimator: NN

Inverse Nearfield Acoustic Holography

Limits

Presence of evanescent components

↓

Highly ill-posed estimation problem

↓

Need of regularization

Nyquist sampling limit requires many mics

↓

Unfeasible if we aim at v above a few tens of Hz

↓

Need of superresolution algorithms

Limited number of mics on $\mathcal{H}$, high resolution on $\mathcal{S}$

Inverse Nearfield Acoustic Holography
Complex-valued Physics-Informed Neural Network for NAH

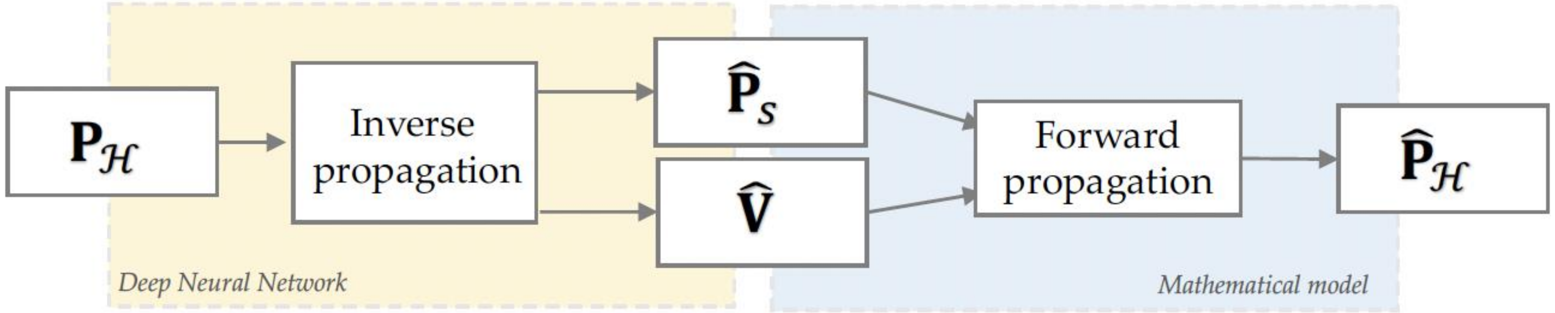
Problem: not possible to include the inverse problem physics directly into the network.

Idea: exploit the knowledge of the forward solution (KH equation) as an external element into the loss function

$$p(x,y,\omega) = \int_{\mathcal{S}} p(x',y') \frac{\partial}{\partial \boldsymbol{n}} g_\omega(x,y,z,x',y',z') d\boldsymbol{s} - j\omega\rho_0 \int_{\mathcal{S}} v_{\boldsymbol{n}}(x',y')\, g_\omega(x,y,z,x',y',z') d\boldsymbol{s}$$

Pressure field on $\mathcal{S}$

Velocity field on $\mathcal{S}$



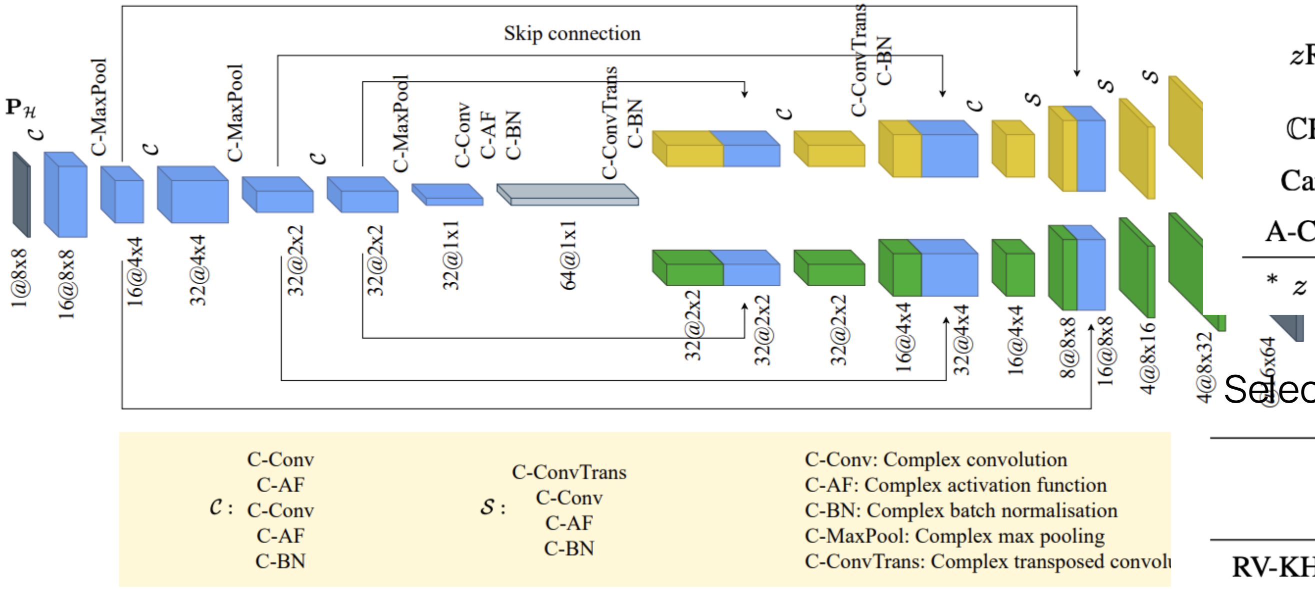
$\hat{\mathbf{P}}_{\mathcal{S}}$, $\mathbf{P}_{\mathcal{H}}$, $\hat{\mathbf{P}}_{\mathcal{H}}$ $\hat{\mathbf{V}}$ are complex

Use of CVNN

- Olivieri M, Pezzoli M, Antonacci F, Sarti A. A Physics-Informed Neural Network Approach for Nearfield Acoustic Holography. *Sensors*. 2021; 21(23):7834
- M. Olivieri, M. Pezzoli, F. Antonacci and A. Sarti, "Near field Acoustic Holography on arbitrary shapes using Convolutional Neural Network," 2021 29th European Signal Processing Conference (EUSIPCO)
- X. Luan, M. Olivieri, M. Pezzoli, F. Antonacci and A. Sarti, "Complex - Valued Physics-Informed Neural Network for Near-Field Acoustic Holography," 2024 32nd European Signal Processing Conference (EUSIPCO)

Inverse Nearfield Acoustic Holography

Complex-valued Physics-Informed Neural Network for NAH



Possible activation functions:

modReLU	$f(z) = \text{ReLU}(\lvert z\rvert + b)e^{i\theta_z}$
zReLU	$f(z) = \begin{cases} z & \text{if} \quad \theta_z \in [0, \pi/2] \\ 0 & \text{otherwise} \end{cases}$
$\mathbb{C}$ReLU	$f(z) = \text{ReLU}(\Re(z)) + i\text{ReLU}(\Im(z))$
Cardioid	$f(z) = \frac{1}{2}z(1 + \cos(\theta_z))$
A-Cardioid	$f(z) = \frac{1}{2}z(1 + \cos(\theta_z + \theta_b))$

* $z = \lvert z\rvert\, e^{i\theta_z}$ and b is the trainable bias.

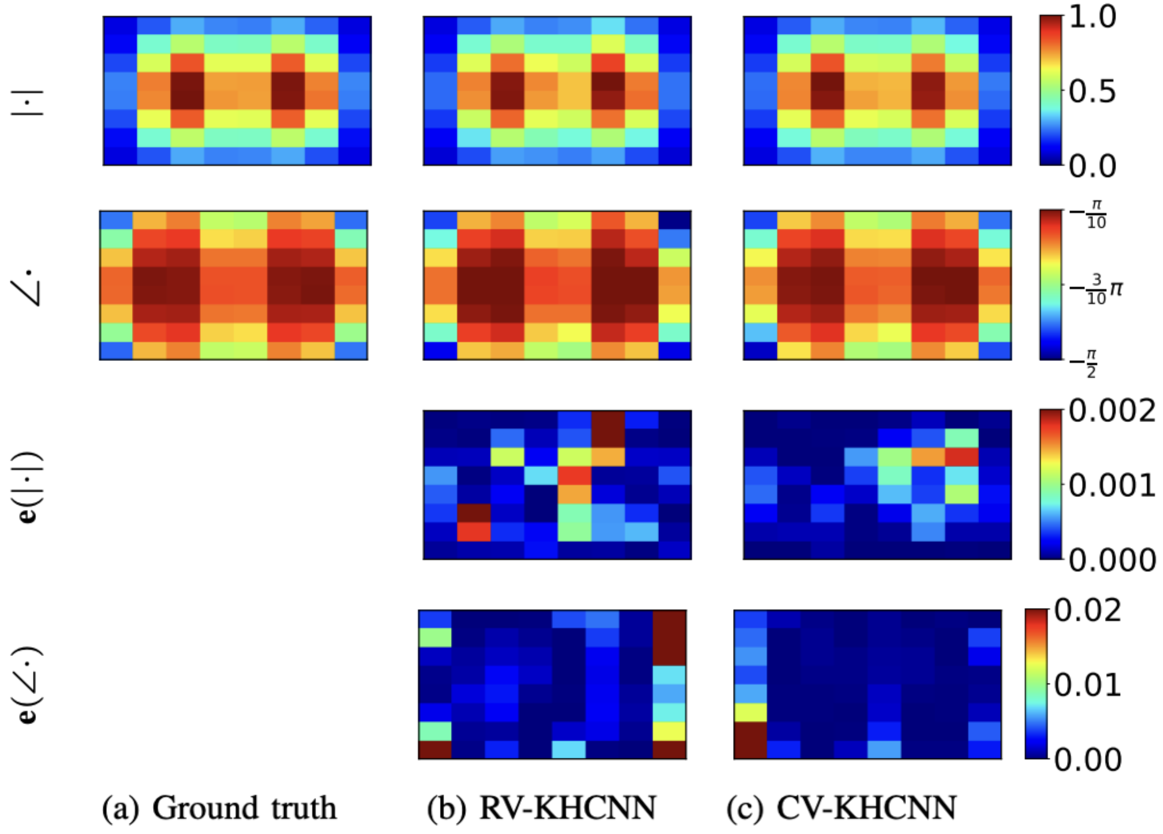
Selection of the activation function:

	$\hat{\mathbf{V}}$		$\hat{\mathbf{P}}_{\mathcal{H}}$	
	NMSE	**NCC**	**NMSE**	**NCC**
RV-KHCNN	-17.46	99.23%	-23.72	99.83%
modReLU	-13.55	98.28%	-24.27	99.87%
$\mathbb{C}$ReLU	-17.92	99.32%	-23.46	99.83%
Cardioid	**-18.99**	**99.48%**	**-26.30**	**99.91%**
A-Cardioid	-18.89	99.47%	-25.99	99.90%

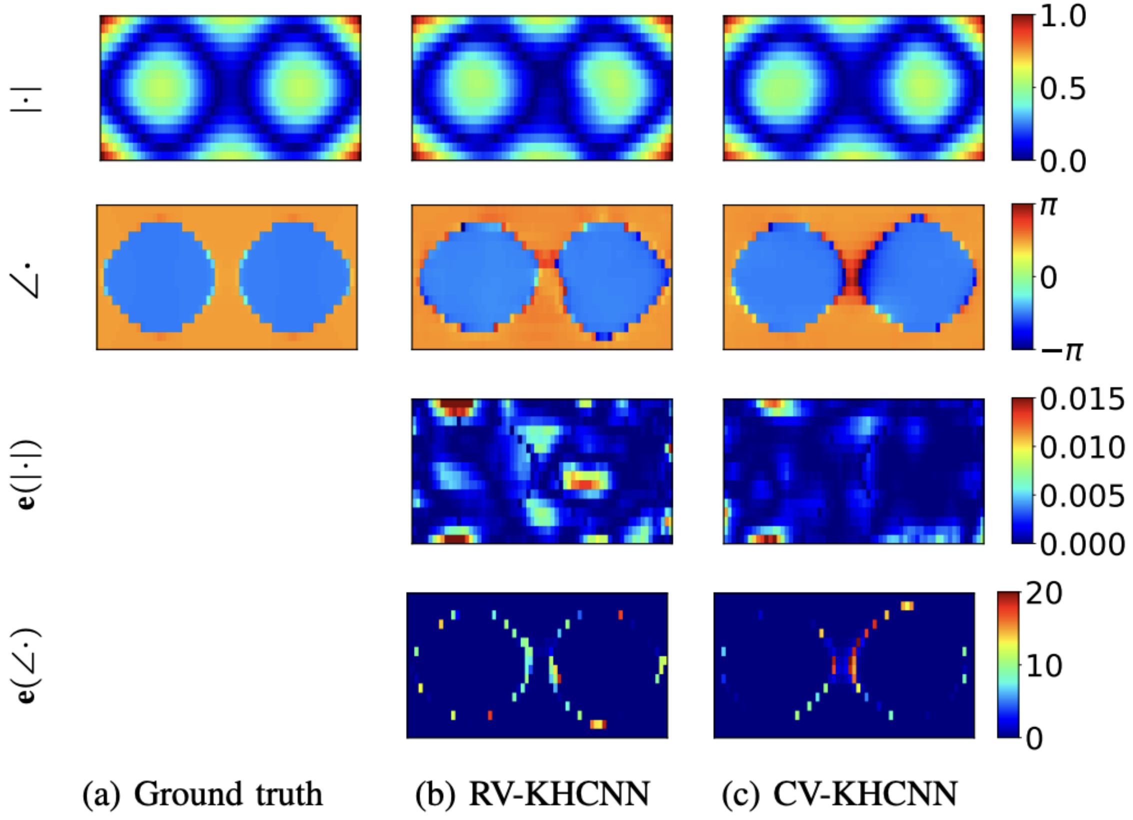
Inverse Nearfield Acoustic Holography

Complex-valued Physics-Informed Neural Network for NAH

Reconstruction example



$\mathbf{P}_{\mathcal{H}}$

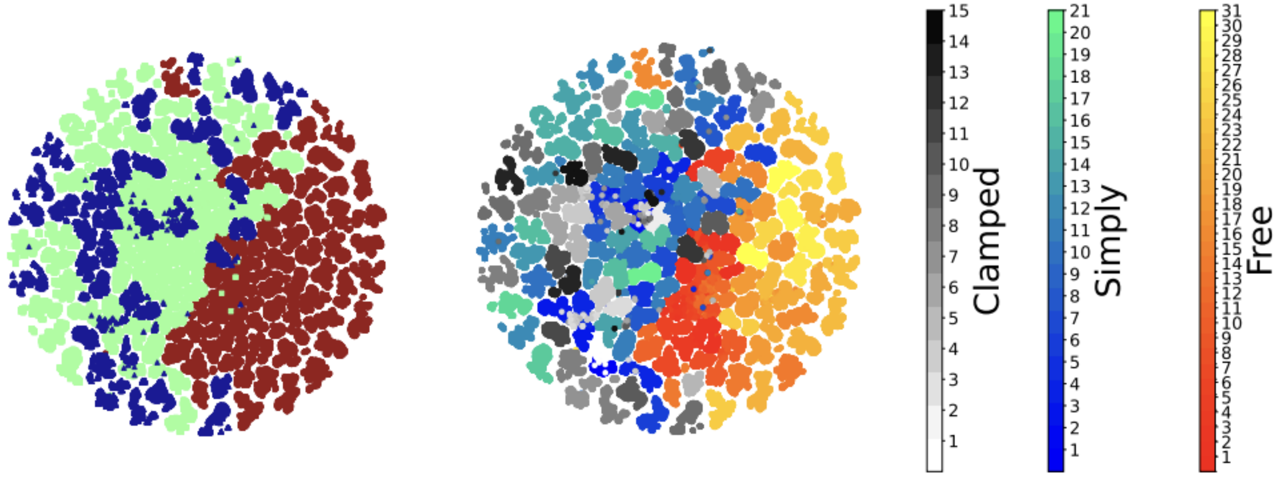


$\mathbf{V}$

Inverse Nearfield Acoustic Holography

Complex-valued Physics-Informed Neural Network for NAH

t-SNE visualization of the bottleneck



(a) (b)

(a): t-SNE of different boundary conditions. Red: free, blue: clamped, green: simply supported
(b): t-SNE for different boundary conditions and mode numbers

- H. Kafri, M. Olivieri, F. Antonacci, M. Moradi, A. Sarti and S. Gannot, "Grad-CAM-Inspired Interpretation of Nearfield Acoustic Holography using Physics-Informed Explainable Neural Network," *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*
- X. Luan, M. Olivieri, M. Pezzoli, F. Antonacci and A. Sarti, "Complex - Valued Physics-Informed Neural Network for Near-Field Acoustic Holography," 2024 32nd European Signal Processing Conference (EUSIPCO)

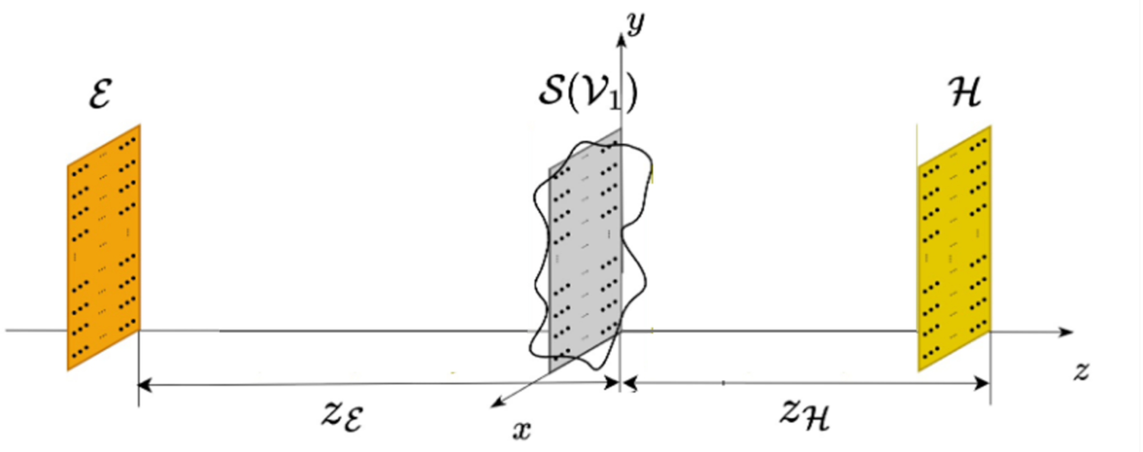
Inverse Nearfield Acoustic Holography

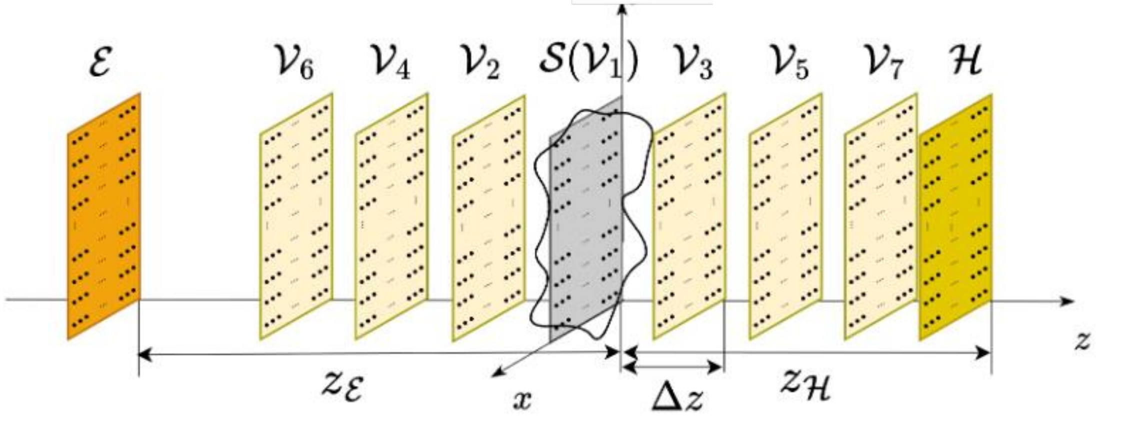
Physics-Informed Neural Network-driven Sparse Field Discretization method (PINN-SFD)

Inverse Equivalent Source method: model the pressure on the hologram plane as the propagation of equivalent sources from the plane ε to the hologram plane $\mathcal{H}$. The velocity field is obtained by propagating the sources on ε to the source plane $\mathcal{S}(\mathcal{V}_1)$.

PINN-SFD: introduce virtual planes (VPs) $\mathcal{V}_{1\cdots N_v}$ between ε and $\mathcal{H}$. The sound field is propagated from ε to the nearest VP and then between VPs up to $\mathcal{H}$. Pro: additional regularization constraints are imposed.

A one-shot self supervised learning strategy is adopted → no need of training datasets.

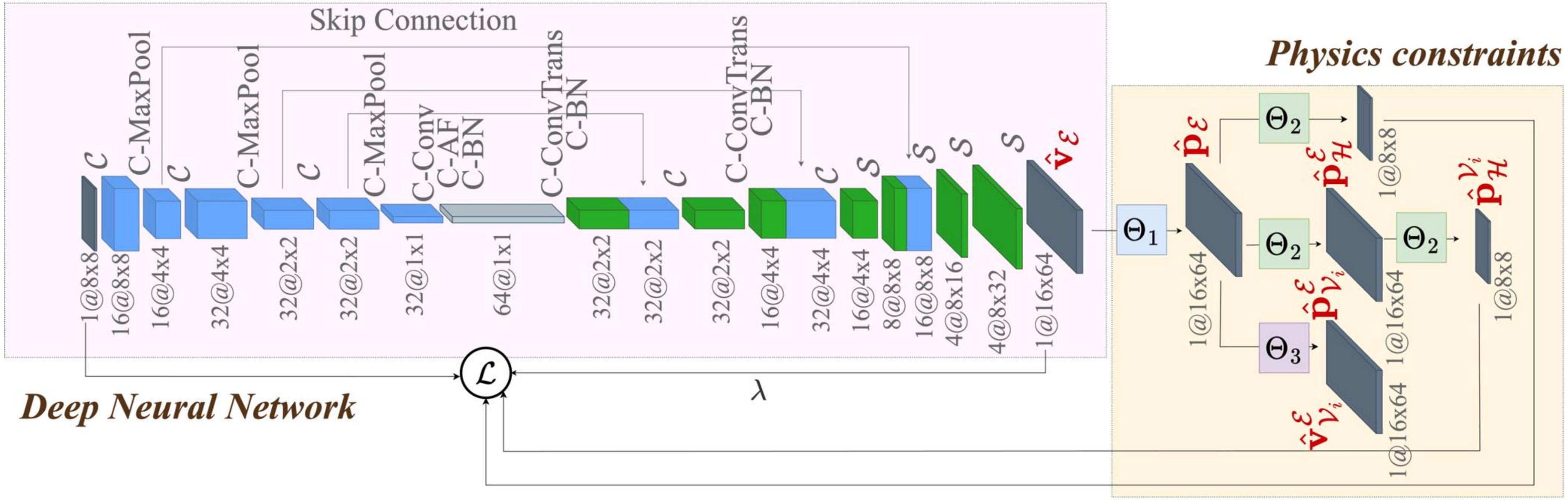




Xinmeng Luan, Mirco Pezzoli Fabio Antonacci, Augusto Sarti, Physics-Informed Neural Network-Driven Sparse Field Discretization Method for Near-Field Acoustic Holography, accepted for publication at IEEE/ACM Transactions on Audio, Speech and Language Processing

Inverse Nearfield Acoustic Holography

Physics-Informed Neural Network-driven Sparse Field Discretization method (PINN-SFD)



$\mathcal{C}$: C-Conv, C-AF, C-Conv, C-AF, C-BN

$\mathcal{S}$: C-ConvTrans, C-Conv, C-AF, C-BN

C-Conv: Complex convolution
C-AF: Complex activation function
C-BN: Complex batch normalisation
C-MaxPool: Complex max pooling
C-ConvTrans: Complex transposed convolution

$$\mathcal{L} = \frac{1}{M}\left(\|\mathbf{p}_{\mathcal{H}} - \hat{\mathbf{p}}_{\mathcal{H}}^{\mathcal{E}}\|_1 + \sum_{i=1}^{N_v} \|\mathbf{p}_{\mathcal{H}} - \hat{\mathbf{p}}_{\mathcal{H}}^{\mathcal{V}_i}\|_1 \right) + \lambda \|\hat{\mathbf{v}}_{\mathcal{E}}\|_1$$

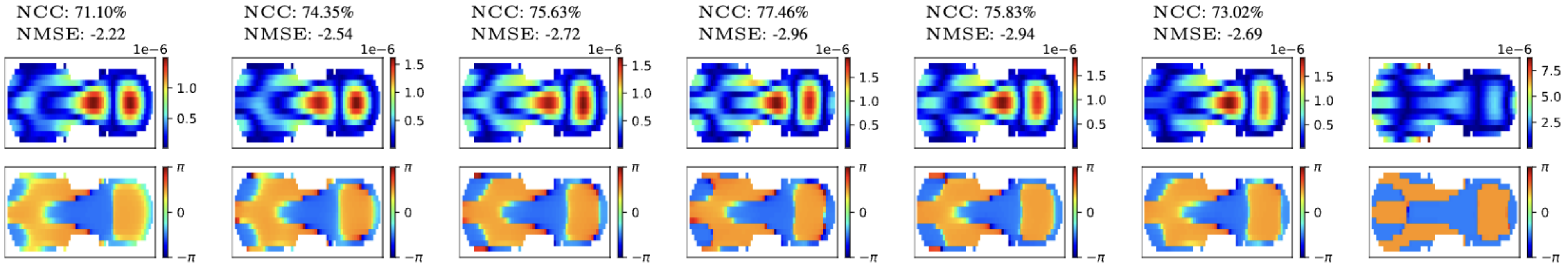
$\|\mathbf{p}_{\mathcal{H}} - \hat{\mathbf{p}}_{\mathcal{H}}^{\mathcal{E}}\|_a^a$ propagation from ε to $\mathcal{H}$ penalty term

$\|\mathbf{p}_{\mathcal{H}} - \hat{\mathbf{p}}_{\mathcal{H}}^{\mathcal{V}_i}\|_a^a$ propagation from $\mathcal{V}_i$ to $\mathcal{H}$ penalty term

Xinmeng Luan, Mirco Pezzoli Fabio Antonacci, Augusto Sarti, Physics-Informed Neural Network-Driven Sparse Field Discretization Method for Near-Field Acoustic Holography, accepted for publication at IEEE/ACM Transactions on Audio, Speech and Language Processing

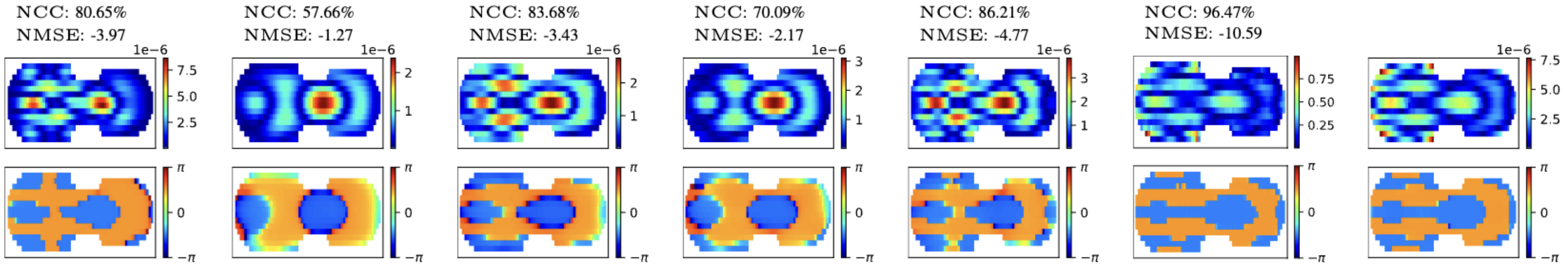
Inverse Nearfield Acoustic Holography

Physics-Informed Neural Network-driven Sparse Field Discretization method (PINN-SFD)



(a) $N_v = 0$ (b) $N_v = 1$ (c) $N_v = 2$ (d) $N_v = 3$ (e) $N_v = 5$ (f) $N_v = 7$ (g) GT

Mode 19, f = 800.3 Hz



(a) (b) (c) (d) (e) (f) (g)

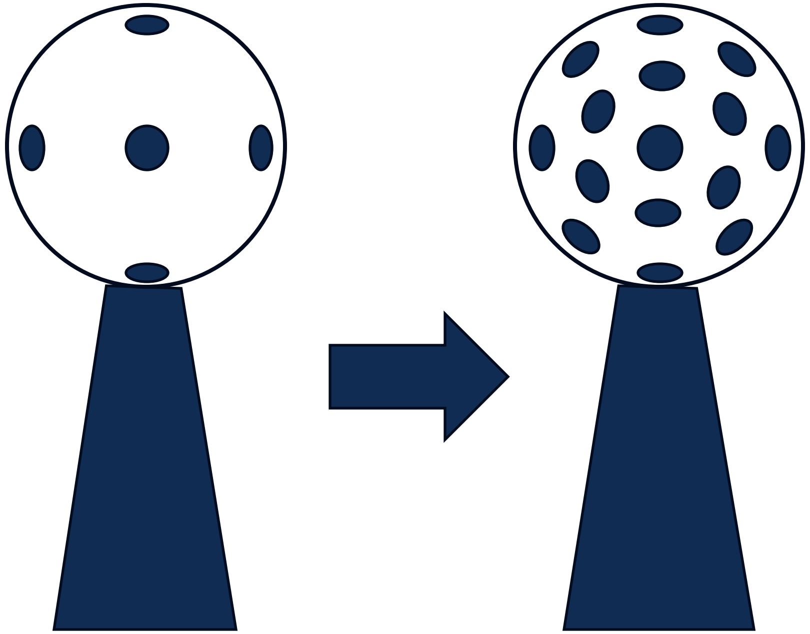
Mode 21, f = 863,7 Hz

Upsampling of spatial audio data

04

Upsampling spherical microphone array measurements
Problem statement

Goal: increase the spatial resolution of spherical microphone arrays (SMAs) for increasing the order of spherical harmonic decomposition,



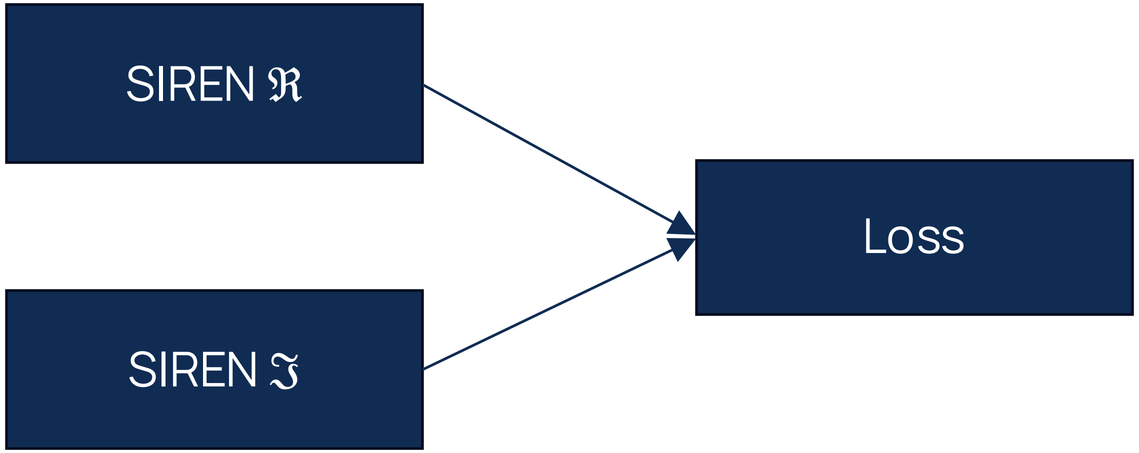
Needs:
- Mitigate the requirement of large datasets for training the network
- Include into the network some knowledge about the physics of the problem (e.g. rigid sphere)

Physics Informed Neural Network

- T. Lübeck, J. M. Arend, and C. Pörschmann, “Spatial upsampling of sparse spherical microphone array signals,” , IEEE Trans. Audio Speech Lang. Process., vol. 31, pp. 1163–1174, 2023
- F. Miotello, F. Terminiello, M. Pezzoli, A. Bernardini, F. Antonacci and A. Sarti, "A Physics-Informed Neural Network-Based Approach for the Spatial Upsampling of Spherical Microphone Arrays," IWAENC 2024, pp. 215-219

Upsampling spherical microphone array measurements
Model

$$\Lambda_i = \sigma_i(\boldsymbol{x}_i^T \boldsymbol{\theta}_i + \boldsymbol{b}_i)$$
$$\sigma_i(z) = \sin(\omega_0 z) + \sum_{w=1}^{W} n_w \sin(\alpha_w z)$$



Parameter	Value
Activation function	Rowdy
n_w	1
α_w	W
# layers	L=4

$$\mathcal{L} = \frac{1}{Q} \sum_{\mathbf{r}_q \in \mathcal{Q}} \underbrace{\|\hat{p}(\mathbf{r}_q, k) - \tilde{p}(\mathbf{r}_q, k)\|_2^2}_{\text{Data fidelity term}} + \lambda \frac{1}{S} \sum_{s=1}^{S} \underbrace{\|[\nabla^2 \hat{p}_{\Re}(\mathbf{r}_s, k) + i \nabla^2 \hat{p}_{\Im}(\mathbf{r}_s, k)] + k^2 \hat{p}(\mathbf{r}_s, k)\|_2^2}_{\text{Physics-based term (Helmholtz equation)}},$$

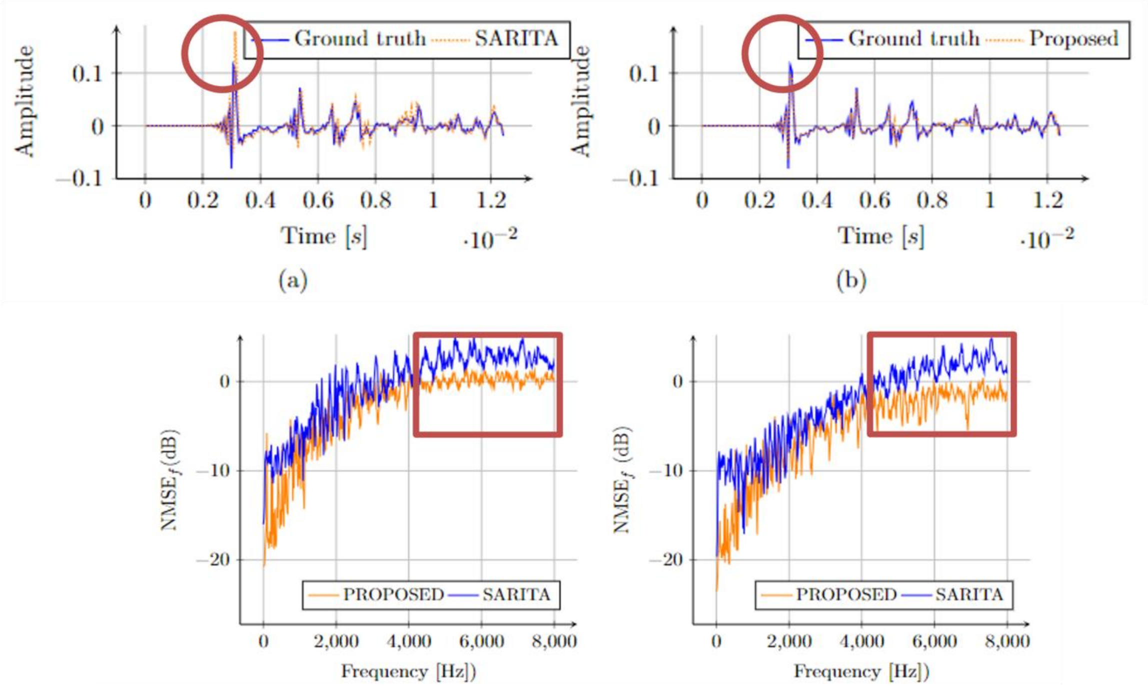
Upsampling spherical microphone array measurements PINN results

Upsampling of real measurements of a spherical microphone arrays

- Comparison with SARITA [1] - Signal processing method

Mean NMSE concerning the number of available channels in the SMA for EM32D dataset

	Mean NMSE			
Q	4	9	16	25
SARITA	-0.65	-2.6	-4.9	-5.57
SIREN	-1.17	-2.60	-5.76	-10.92
SIREN + PDE	-1.71	-4.97	-6.38	-11.13
Proposed	**-2.05**	**-5.40**	**-6.83**	**-12.44**



[1] T. Lübeck, J. M. Arend, and C. Pörschmann, “Spatial upsampling of sparse spherical microphone array signals,” Trans. Audio Speech Lang. Process., vol. 31, pp. 1163–1174, 2023.
[2] F. Miotello, F. Terminiello, M. Pezzoli, A. Bernardini, F. Antonacci and A. Sarti, "A Physics-Informed Neural Network-Based Approach for the Spatial Upsampling of Spherical Microphone Arrays," 2024 18th International Workshop on Acoustic Signal Enhancement (IWAENC), Aalborg, Denmark, 2024, pp. 215-219
[3] J. Xia and W. Zhang, “Upmix b-format Ambisonic room impulse responses using a generative model,” Applied Sciences, vol. 13, no. 21, p. 11810, 2023.

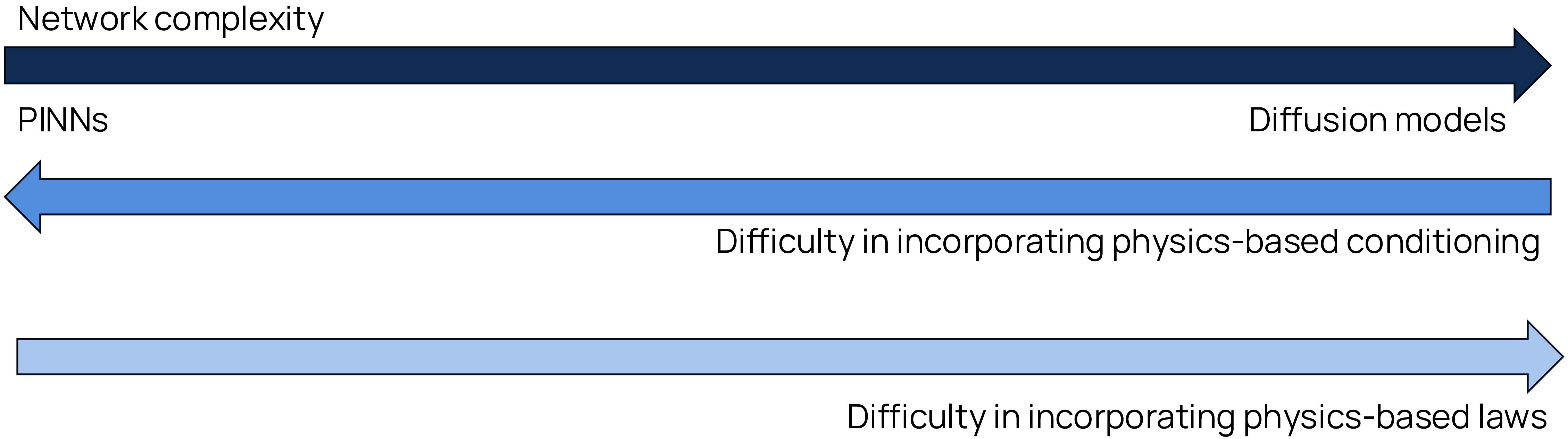
Considerations

Strict requirements if generated data are used for the development and training of space-time processing algorithms:

- Phase relationships (if we work with RTFs);
- Time delays (if we work with RIRs).

These constraints can be fulfilled through:

- Conditioning of the input (CVNN or diffusion models);
- Dedicated network architectures (PINNs).



Credits

Luca Comanducci
Federico Miotello
Luan Xinmeng
Mirco Pezzoli
Francesca Ronchini
Augusto Sarti
Ferdinando Terminiello

Thank you!