forum acusticum 2026
8-12 September · Graz, Austria

PHYSICS-INFORMED NEURAL NETWORKS FOR ROOM-ACOUSTIC SIMULATIONS WITHIN BAYESIAN FRAMEWORK

William Zheng[1], Ning Xiang[1]

[1] *Graduate Program in Architectural Acoustics, Rensselaer Polytechnic Institute, Troy, NY, USA*

Abstract

Accurate prediction of sound energy decay in enclosed spaces is essential for acoustic design in performance halls, classrooms, and recording studios. Statistical room-acoustic theory, such as the Sabine reverberation formula, provides fast but spatially limited estimates, while wave-theoretical simulations are physically accurate but computationally expensive. This work presents a Physics-Informed Neural Network (PINN) trained to solve the acoustic diffusion equation, using finite-difference time-domain (FDTD) simulations for supervised learning. The network takes spatial position and time as inputs and predicts sound energy density without requiring re-simulation. The maximum likelihood enforces the governing partial differential equation, boundary conditions, and initial conditions alongside supervised data. Preliminary results show close agreement with FDTD ground truth across the full decay range, accurately reproducing the temporal decay process at the receiver positions. Full temporal coverage of the decay process provides stronger training than spatially dense but temporally sparse sampling. The current work also formulates the PINN within Bayesian inferential framework to highlight the learning process incorporating the prior.

Keywords: *physics-informed neural networks, acoustic diffusion equation, reverberation time, Bayesian inference, surrogate modeling*

1. Introduction

Prediction of sound energy decay in enclosed spaces is central to architectural acoustics. The acoustic diffusion equation (ADE) provides a physically grounded model for energy transport in rooms with diffusely reflecting surfaces [1]. Physics-informed neural networks (PINNs) have emerged as a promising surrogate modeling approach that embeds governing differential equations directly into the neural network training objective [2]. By penalizing violations of the PDE, boundary conditions, and initial conditions during training, PINNs can learn accurate solution operators from relatively small amounts of labeled data, with the physics serving as a regularizer. PINNs have since been applied to room-acoustic problems such as room impulse response reconstruction [3], and, more recently, to accelerating coupled normal-mode models for underwater acoustic propagation [4].

This work applies PINNs to room acoustic energy decay governed by the three-dimensional ADE, with particular attention to training efficiency. This work demonstrates that time series at a few receiver positions is sufficient to anchor the PINN training and achieve decay-curve accuracy within 3.5% of FDTD ground truth across four physically distinct room configurations, including rooms with a non-uniform sound absorber on the floor.

2. Physical Model

The acoustic diffusion equation governs energy transport throughout the enclosure, subject to boundary conditions at the walls and an initial condition at the source. The four room configurations described below provide the physical basis for training and evaluating the generalized PINN surrogate.

2.1. Governing Equations

The acoustic energy density $w(\mathbf{x}, t)$ in an enclosure satisfies the acoustic diffusion equation. For $t > 0$, after the initial energy pulse, the governing interior equation [1], [5] is

$$\frac{\partial w}{\partial t} = D\nabla^2 w - c\,m\,w, \tag{1}$$

where $D = c\,\ell/3$ is the diffusion coefficient, $\ell = 4V/S$ is the mean free path, V is the room volume, S is the

Corresponding author: .

total wall surface area, c is the speed of sound, and m is the air absorption coefficient. Wall absorption is enforced through the Jing & Xiang Robin boundary condition on all six surfaces [1]:

$$D\,\nabla w \cdot \hat{n} + \kappa\, w = 0, \quad \kappa = \frac{c\alpha}{4(1-\alpha/2)}, \qquad (2)$$

where α is the local absorption coefficient of the surface and $\hat{n}$ is the outward unit normal. For configurations with a sound absorber on the floor, α takes the absorber value α_{abs} within the absorber bounds and the background wall absorption value α_{wall} elsewhere, making κ spatially varying across the floor surfaces. The source excitation is encoded as a Gaussian initial condition centered at the source position $\mathbf{r}_s$:

$$w(\mathbf{x}, 0) = 2\exp\left(-\frac{|\mathbf{x}-\mathbf{r}_s|^2}{\sigma^2}\right), \quad \sigma = 3\,\Delta x, \quad (3)$$

where $\Delta x = \ell/6$ is the spatial grid spacing of the FDTD discretization. This replaces a continuous point-source term in the PDE, allowing the PINN to enforce the homogeneous diffusion equation Eq. (1) throughout the interior domain for all $t > 0$ while the initial energy distribution captures the source excitation at $t = 0$.

2.2. Room Configurations

Four room configurations are used to evaluate the generalized PINN surrogate, illustrated in Fig. 1. All configurations use a background wall absorption coefficient $\alpha_{\text{wall}} = 0.03$, air absorption coefficient $m = 0.0025\ \text{m}^{-1}$, and speed of sound $c = 343$ m/s. The total simulation time is $T = 2\,\text{s}$ for all configurations.

Room 1 is an $8.3 \times 6.3 \times 5.5\,\text{m}$ enclosure with uniform absorption $\alpha = 0.03$ on all surfaces, giving $D \approx 496\,\text{m}^2\,\text{s}^{-1}$. The source is located at $[8.0, 6.1, 5.3]$ m and the receiver at $[1.7, 1.9, 2.5]$ m.

Room 2 uses the same $8.3 \times 6.3 \times 5.5\,\text{m}$ geometry as Room 1 with the addition of a sound absorber on the floor ($\alpha_{\text{abs}} = 0.99$) covering the region $x \in [2.7, 7.7]$ m, $y \in [4.0, 6.0]$ m. All remaining surfaces retain $\alpha_{\text{wall}} = 0.03$.

Room 3 is a $9.6 \times 7.2 \times 7.2\,\text{m}$ enclosure with a sound absorber on the floor ($\alpha_{\text{abs}} = 0.99$) covering $x \in [2.7, 7.7]$ m, $y \in [4.0, 7.0]$ m, giving $D \approx 598\,\text{m}^2\,\text{s}^{-1}$. The source is located at $[8.0, 3.8, 3.0]$ m and the receiver at $[1.0, 1.8, 2.5]$ m.

Room 4 shares the $9.6 \times 7.2 \times 7.2\,\text{m}$ geometry of Room 3 with uniform absorption $\alpha = 0.03$ on all surfaces and no floor absorber, giving the same $D \approx 598\,\text{m}^2\,\text{s}^{-1}$, source, and receiver positions. Rooms 3 and 4 form a matched pair with Rooms 1 and 2, isolating the effect of the floor absorber across both geometries.

Together these four configurations expose the model to two distinct room geometries, uniform and non-uniform absorption, and varying absorber sizes and positions, providing a contrastive dataset for evaluating across room acoustic parameters.

3. Physics-Informed Neural Network

This section describes the PINN used to solve the physical model introduced in Section 2. We first detail the network architecture and its physics-informed loss function. We then describe the training strategy, which fits the model to receiver data alone.

3.1. Network Architecture

The generalized surrogate model is an MLP with 16-dimensional input

$$\mathbf{u} = [\tilde{\mathbf{r}}, \tilde{t},\ \mathbf{G},\ \tilde{\mathbf{r}}_s,\ \mathbf{a},\ \tilde{\mathbf{p}}] \qquad (4)$$

and scalar output $\hat{w}$. Space-time coordinates $[\tilde{\mathbf{r}}, \tilde{t}\,] = [\tilde{x}, \tilde{y}, \tilde{z}, \tilde{t}] \in [0,1]^4$ are normalized by room dimensions and total simulation time. Room dimensions $\mathbf{G} = [G_x, G_y, G_z]$ are passed in metres. Source position $\tilde{\mathbf{r}}_s = [\tilde{r}_{s,x}, \tilde{r}_{s,y}, \tilde{r}_{s,z}]$ and floor absorber bounds $\tilde{\mathbf{p}} = [\tilde{p}_x^{\min}, \tilde{p}_x^{\max}, \tilde{p}_y^{\min}, \tilde{p}_y^{\max}]$ are normalized by their respective room dimensions, and $\mathbf{a} = [\alpha_{\text{wall}}, \alpha_{\text{abs}}]$. For configurations without a floor absorber, absorber bounds are set to zero and $\alpha_{\text{abs}} = \alpha_{\text{wall}}$, reducing the absorber contribution to zero in the boundary condition. Each input is necessary for the model's ability to make predictions for various room configurations.

The architecture consists of one input layer ($16 \to 256$), six hidden layers ($256 \to 256$) with tanh activations and Dropout($p = 0.1$), and one output layer ($256 \to 1$) with Softplus activation to enforce $w \geq 0$, giving approximately 332,000 parameters. The diffusion coefficient D and Robin BC coefficient κ are computed per-batch from the room dimension and absorption columns of the input tensor, allowing a single forward pass to simultaneously process points from rooms with different geometries and absorption configurations.

Energy labels are normalized in log space,

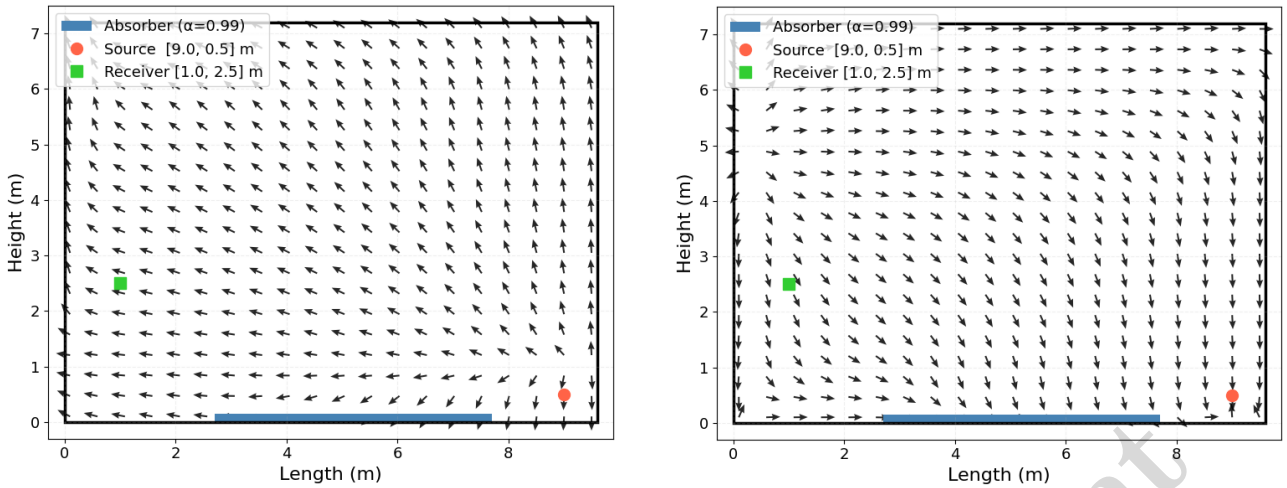
$$\tilde{w} = \frac{\ln(w+\varepsilon) - \ln_{\min}}{\ln_{\max} - \ln_{\min}}, \qquad (5)$$

where $\ln_{\min}$ and $\ln_{\max}$ are computed jointly across all room configurations. A physical noise floor $w_{\text{floor}} = 10^{-8}$ is applied to all receiver labels before normalization to prevent numerically insignificant tail values in high-absorption configurations from dominating the normalized range. Chain rule corrections are applied to all spatial and temporal derivatives to recover physical-space gradients from normalized autograd outputs.

3.2. Loss Function

The four residual error terms are defined as follows. The PDE interior residual ε_f, the Robin BC residual ε_b, the Gaussian IC residual ε_0, and the supervised receiver data residual ε_d are

$$\begin{aligned} \varepsilon_f &= \tfrac{\partial w}{\partial t} - D\nabla^2 w + cmw, \\ \varepsilon_b &= D\,\nabla w \cdot \hat{n} + \kappa\, w, \\ \varepsilon_0 &= w(\mathbf{x}, 0) - 2\exp\!\big(-|\mathbf{x}-\mathbf{r}_s|^2/\sigma^2\big), \\ \varepsilon_d &= \hat{w} - w_{\text{d}}, \end{aligned} \qquad (6)$$



(a) $t = 0$ s (b) $t = 1.5$ s

Figure 1: Cross-sectional side view (length–height plane at mid-width) of Room 3 ($9.6 \times 7.2 \times 7.2$ m) showing the acoustic energy flux field at (a) $t = 0$ s and (b) $t = 1.5$ s. Arrows indicate the direction and relative magnitude of energy flow. The sound absorber on the floor ($\alpha_{\text{abs}} = 0.99$, shown in blue) draws energy downward. The orange circle marks the source position, relocated here to illustrate the model's source-position generalization (see Section 2).

and the corresponding mean-squared errors are MSE_f, MSE_b, MSE_0, MSE_d. These are combined into the weighted training objective

$$\text{Loss}_{\text{total}} = \omega_f \,\text{MSE}_f + \omega_b \,\text{MSE}_b + \omega_0 \,\text{MSE}_0 + \omega_d \,\text{MSE}_d. \tag{7}$$

All four terms aggregate contributions from all room configurations simultaneously. Each epoch samples $K_f = 5{,}000$ interior points, $K_b = 5{,}000$ boundary points, and $K_0 = 3{,}000$ IC points independently per room, concatenated into a single batch before computing residuals. The supervised data term MSE_d uses the full combined receiver time series across all rooms ($N_{\text{total}} = \sum_r T_{\text{max}}^{(r)}$ points).

The PDE residual is divided by the per-batch D for numerical conditioning. The boundary condition residual computes κ per-point from the local absorption coefficient using differentiable element-wise multiplication rather than branching conditionals to preserve gradient flow. Loss weights are set as $\omega_f = 1.0$, $\omega_b = 0.0001$, $\omega_0 = 50.0$, $\omega_d = 20.0$. The boundary weight ω_b is substantially reduced relative to single-room training because the sound absorber ($\alpha_{\text{abs}} = 0.99$) produces boundary residuals two orders of magnitude larger than uniform-absorption walls.

3.3. Training Data Strategy

The model is trained exclusively on receiver time series data, the full FDTD energy decay curve at a single fixed receiver position per room configuration. No volumetric spatial labels are used. This corresponds to the minimal experimental scenario of one microphone measurement per room, requiring no spatial scanning or array measurements. Physics constraints (PDE, BC, IC residuals) sampled freshly each epoch provide the spatial regularization that labeled spatial data would otherwise supply.

The combined labeled dataset across all four room configurations contains N_{total} receiver points, compared to the millions of labeled space-time points that a full spatial sampling approach would require. This demonstrates that temporal coverage of the decay process, enforced through physics constraints, is a more data-efficient training strategy than spatial density.

4. Bayesian Framework

Bayesian learning for the estimation of neural network parameters $\boldsymbol{\Theta}$ applies Bayes' theorem to determine the posterior probability of $\boldsymbol{\Theta}$ given data $\mathbf{D}$, model $\mathbf{H}_{\text{P}}$, and background information I:

$$p(\boldsymbol{\Theta}|\mathbf{D}, \mathbf{H}_{\text{P}}, I) = \frac{p(\mathbf{D}|\boldsymbol{\Theta}, \mathbf{H}_{\text{P}}, I)\; p(\boldsymbol{\Theta}|\mathbf{H}_{\text{P}}, I)}{p(\mathbf{D}|\mathbf{H}_{\text{P}}, I)}, \tag{8}$$

where the model $\mathbf{H}_{\text{P}}$ collectively includes $[H_{\text{NN}}(\mathbf{r}, t, \boldsymbol{\Theta}),\ \nabla \cdot \mathbf{J},\ -\mathbf{J} \cdot \hat{n}]$, with the likelihood function $\mathcal{L} = p(\mathbf{D}|\boldsymbol{\Theta}, \mathbf{H}_{\text{P}}, I)$ and training data $\mathbf{D} = [\mathbf{D}_{\text{d}}, \mathbf{D}_0, \mathbf{D}_{\text{f}}, \mathbf{D}_{\text{b}}]$. The log-likelihood evaluated jointly across all room configurations is

$$10 \log_{10}(\mathcal{L}) = -5\Big[N_0 \log_{10}(\text{MSE}_0) + N_{\text{d}} \log_{10}(\text{MSE}_d) + N_f \log_{10}(\text{MSE}_f) + N_b \log_{10}(\text{MSE}_b)\Big], \tag{9}$$

where $N_0 = K_0 \cdot R$, $N_f = K_f \cdot R$, and $N_b = K_b \cdot R$ are the total IC, PDE, and boundary collocation points across all R room configurations per epoch. N_{d} is the total number of receiver timesteps pooled across all rooms. This quantity is recorded at each training improvement checkpoint and used for principled model comparison as model selection at the second

level of Bayesian inference. Unlike a two-term formulation applicable to parameter estimation alone, the four-term expression reflects the full physics-informed training objective, penalizing violations of the governing equation and boundary conditions alongside the supervised data fit.

Dropout is retained at inference time (Monte Carlo Dropout) to approximate a posterior distribution over network outputs. Running 50 stochastic forward passes yields a mean prediction and pointwise uncertainty estimate, with uncertainty highest at onset ($t < 0.1$ s) where the Gaussian pulse evolves rapidly, as illustrated in Fig. 2. T60 estimates from the MC Dropout mean are less sensitive to local prediction noise than single deterministic forward passes, due to the variance-reduction effect of averaging over stochastic samples.

5. Results

The trained PINN is evaluated against FDTD ground truth on two fronts: decay accuracy and reverberation time across all four rooms, and the extent to which a single model generalizes across distinct geometries and absorption regimes.

5.1. Energy Decay Accuracy

Fig. 3 shows the predicted receiver decay curves compared to FDTD ground truth for all four room configurations. The PINN tracks the decay range closely across all rooms, with visible divergence only near the physical noise floor where energy values fall below $w_{\text{floor}} = 10^{-8}$.

Results of reverberation time are summarized in Table 1. All four configurations achieve T_{60} errors below 3.5%, well within the ISO 3382 just-noticeable difference threshold of 5%. Rooms with a sound absorber on the floor (Rooms 2 and 3) show errors of 1.86% and 2.39% respectively, demonstrating that the model correctly distinguishes the accelerated decay introduced by high-absorption surfaces from the background uniform absorption. Room 4, the larger geometry without an absorber, achieves 3.37% error, reflecting the challenge of predicting a long reverberation time ($T_{60} = 4.408$ s) within a 2-second training window where the full decay is not directly observed.

5.2. Generalization Across Room Configurations

A single generalized PINN trained jointly on receiver time series from four physically distinct room configurations produces accurate decay curve predictions across all four without retraining. The four configurations span two room geometries, two absorption regimes (uniform α=0.03 and a sound absorber on the floor with α=0.99), and a T_{60} range from 1.522 s to 4.408 s. The model correctly predicts both faster decay due to high surface absorption and slower decay in the larger low-absorption room, demonstrating that the room geometry, absorption coefficient, and absorber geometry inputs are being meaningfully used to condition predictions.

Table 1: T_{60} (time to reach 60 dB) comparison across all four room configurations. FDTD (Finite-Difference Time-Domain) ground truth vs. generalized PINN (Physics-Informed Neural Network).

Configuration	T_{60}^{FDTD}	T_{60}^{PINN}	Err.
R1: uniform α=0.03 8.3×6.3×5.5 m	3.672 s	3.584 s	2.41%
R2: absorber α=0.99 8.3×6.3×5.5 m	1.522 s	1.494 s	1.86%
R3: absorber α=0.99 9.6×7.2×7.2 m	1.707 s	1.667 s	2.39%
R4: uniform α=0.03 9.6×7.2×7.2 m	4.408 s	4.260 s	3.37%

The receiver time series strategy requires only $N_{\text{total}} = \sum_r T_{\max}^{(r)}$ labeled points across all rooms combined, equivalent to one microphone measurement per configuration. Physics constraints sampled freshly each epoch from all four rooms provide spatial regularization in place of volumetric labeled data, making the approach practically deployable with minimal experimental effort.

6. Concluding Remarks

This research demonstrates that a single generalized PINN, trained jointly on receiver time series from four room configurations spanning two geometries and two absorption regimes, achieves T_{60} accuracy within 3.4% of FDTD ground truth across all configurations. The 16-dimensional input representation, encoding room geometry, absorption coefficients, source position, and floor absorber parameters, enables a single model to condition predictions on the specific physical parameters of each room without retraining. Crucially, the labeled training data requirement is limited to one receiver time series per room, with physics constraints providing the spatial regularization that would otherwise require volumetric measurements. Experimentally measured decay functions in reverberation chambers will be involved in training the PINN prior to Bayesian nested sampling.

Future work will extend the training dataset to a broader range of room geometries and absorption values to improve generalization. Bayesian nested sampling, guided by Monte Carlo Dropout uncertainty and PDE residual magnitude, will be used to identify optimal measurement locations for querying additional training data.

References

[1] Y. Jing and N. Xiang, "A modified diffusion equation for room-acoustic prediction," *Journal of the Acoustical Society of America*, vol. 121, no. 6, pp. 3284–3287, 2008.

[2] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learn-

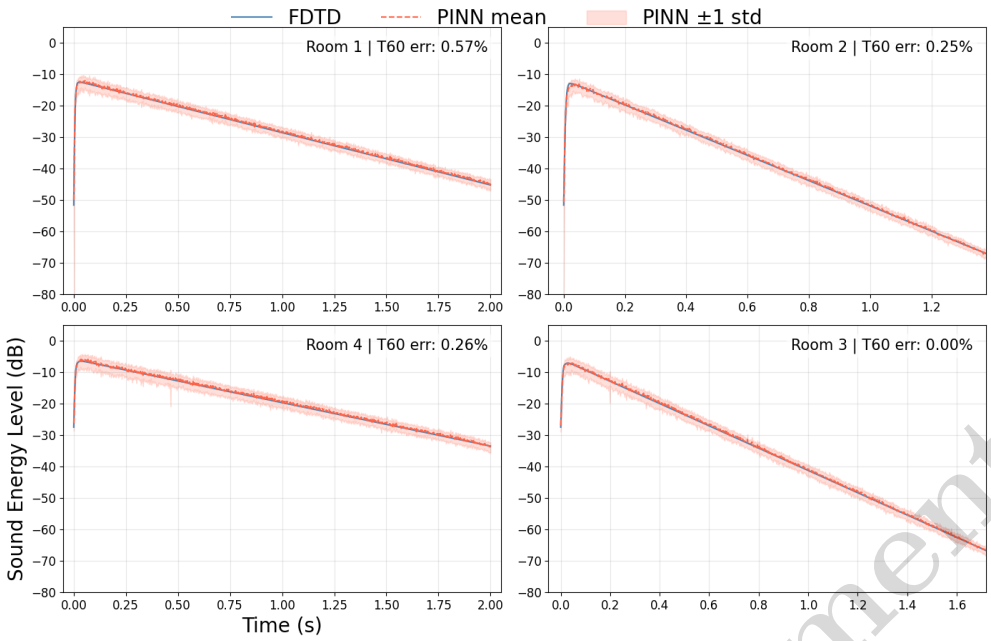


Figure 2: Monte Carlo Dropout uncertainty for all four room configurations. Dashed lines show PINN (Physics-Informed Neural Network) mean prediction across 50 stochastic forward passes. Shaded bands show ±1 standard deviation. Solid lines are FDTD (Finite-Difference Time-Domain) ground truth. Uncertainty is highest at onset and near the noise floor, and lowest in the physically meaningful decay region (−5 to −65 dB).

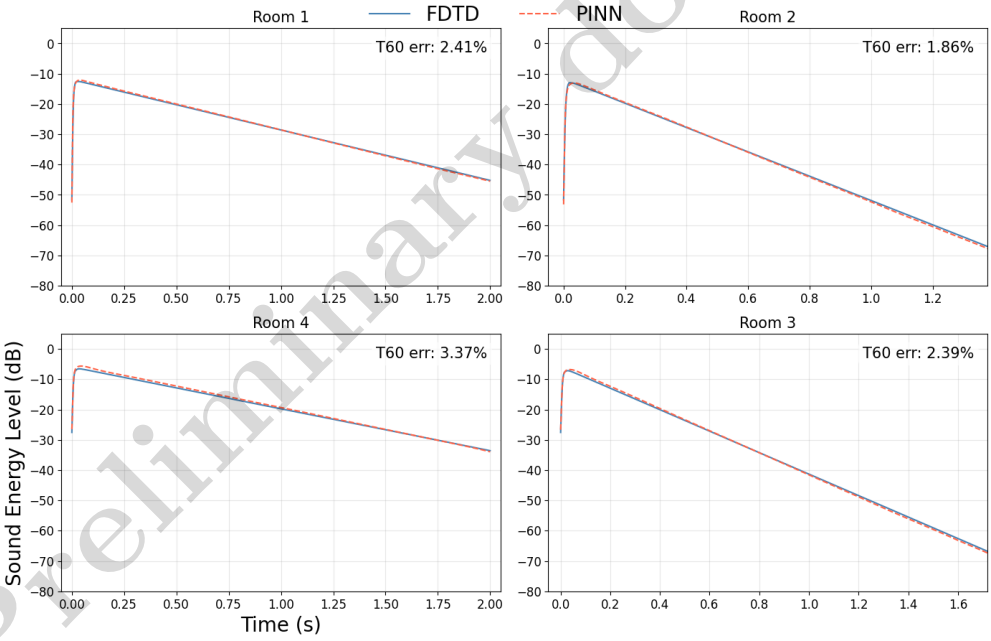


Figure 3: Comparison of energy decay curve (in level dB) at a receiver between the Physically Informed Neural Network (PINN) and the Finite-difference Time-domain (FDTD) simulation as the ground truth for all four room configurations. Solid lines are FDTD ground truth, dashed lines are PINN mean predictions. The x-axis ends when FDTD reaches −65 dB or at $T = 2\,\mathrm{s}$, whichever comes first.

ing framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.

[3] X. Karakonstantis, D. Caviedes-Nozal, A. Richard, and E. Fernandez-Grande, "Room impulse response reconstruction with physics-informed deep learning," *Journal of the Acousti-*

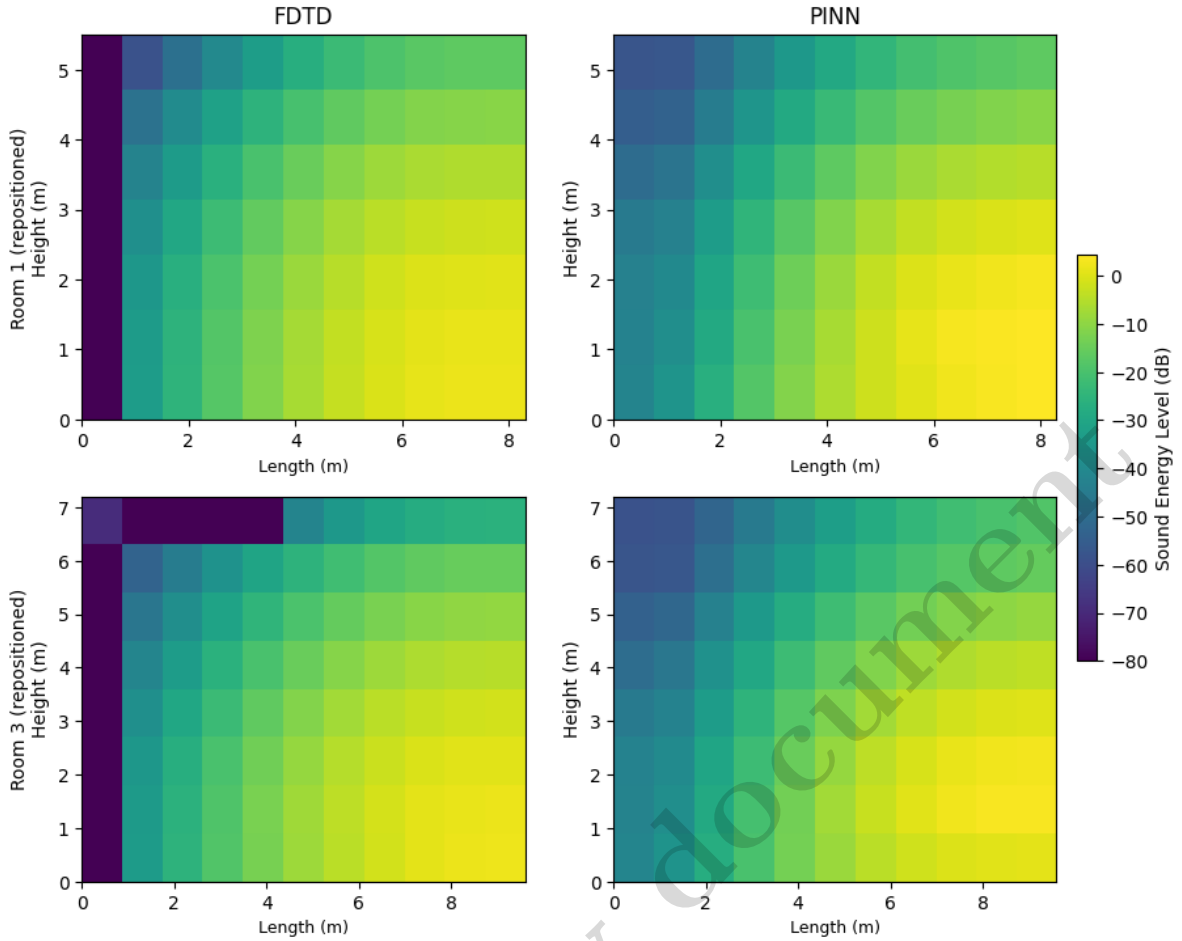


Figure 4: Spatial comparison of predicted sound energy level, the ground truth from the Finite-difference Time-domain (FDTD, left) and that from the Physically-informed Neural Network (PINN, right), for Room 1 (uniform $\alpha = 0.03$) and Room 3 (sound absorber on the floor, $\alpha_{\text{abs}} = 0.99$), at $t = 0.3$ s. The x-z cross-section is taken at mid-room width. The PINN reproduces the spatial energy distribution in both rooms, confirming that accuracy extends beyond the single receiver position used for training.

cal Society of America, vol. 155, no. 2, pp. 1048–1059, 2024.

[4] J. Chen, H. Wang, H. Xia, and J. Liu, "An accelerated coupled normal-mode model based on physics-informed neural networks," *Journal of the Acoustical Society of America*, vol. 159, no. 6, pp. 5105–5125, 2026.

[5] V. Valeau, J. Picaut, and M. Hodgson, "On the use of a diffusion equation for room-acoustic prediction," *Journal of the Acoustical Society of America*, vol. 119, no. 3, pp. 1504–1513, 2006.