

PHYSICS-INFORMED NEURAL NETWORKS FOR ULTRASONIC LAMB WAVE RECONSTRUCTION IN NON-DESTRUCTIVE TESTING

Lukas Jilke[1, 2], Martin Schubert[2], Norbert Hosters[3], Florian Raddatz[4], Marek Behr[5] & Gerko Wende[6]

[1, 4, 6]German Aerospace Center (DLR), Institute of Maintenance, Repair and Overhaul, 21129 Hamburg, Germany
[2]German Aerospace Center (DLR), Facility Small Aircraft Technologies, 52146 Würselen, Germany
[3,5]RWTH Aachen University, Chair for Computational Analysis of Technical Systems, 52062 Aachen, Germany

Abstract

Ultrasonic guided wave (UGW)-based non-destructive testing (NDT) has emerged as an effective method for identifying damage in aerospace structures. Recently, the integration of machine learning models into NDT systems has shown the capability to accurately detect and localize structural damage. However, purely data-driven approaches require extensive datasets and lack physical interpretability. Furthermore, conventional Multi-Layer Perceptrons (MLPs) suffer from spectral bias (F-Principle), preferentially capturing low-frequency components and failing to resolve high-frequency wave propagation. By embedinng the governing partial differential equations of the underlying physical problem into the learning process, Physics-Informed Neural Networks (PINNs) have demonstrated the capability to ensure robust generalization even in data-sparse regimes. Utilizing periodic activation functions in Sinusoidal Representation Networks (SIRENs) facilitates the capture of high-frequency oscillations. This study proposes a Physics-Informed sinusoidal representation network (PI-SIREN) which embed elastodynamic wave equation under the Kirchhoff-Love and Mindlin-Reissner plate theory into the learning process, effectively mitigating spectral bias to enable ultrasonic wavefield reconstruction in data-sparse regimes. Numerical results demonstrate that the proposed PI-SIREN framework significantly outperforms purely data-driven baselines, achieving lower errors when trained on identical dataset sizes. In contrast, PINN configurations based on standard MLPs with hyperbolic tangent activation functions failed to converge, unable to resolve the high-frequency dynamics.

Keywords: Physics-Informed Neural Networks, Ultrasonic Guided Waves, Non-destructive Testing

1. Introduction

The aging nature of aircraft structures requires extensive maintenance efforts to prevent material degradation. Degradation poses a potential threat to aircraft reliability, heightening the risk of material failure [1]. In 2023, the airline industry allocated approximately $93.9 billion towards maintenance, repair and overhaul (MRO) activities, constituting roughly 11 percent of the total operating costs [2]. To address these challenges, diagnostic and prognostic methodologies in aircraft maintenance emerge as essential tools to enhance the reliability, availability and safety of operational aircraft, concurrently mitigating MRO costs [3, 4]. NDT is utilized to inspect the structural condition of an aircraft component. Conventional NDT relies heavily on ultrasonic inspection techniques, with UGW serving as the most widely utilized non-destructive evaluation (NDE) method for thin-walled structures [5, 6]. UGWs are elastic waves propagating within solid plates or layers constrained by free boundaries, where the wave behavior is intrinsically governed by the characteristics of the material [5, 6]. Due to their short wavelengths in the high-frequency region and low amplitude loss over distance, these waves are sensitive to minor perturbations in the waveguide, such as flaws or damages in the material [7]. Different approaches are utilized to increase the prediction accuracy of damage detection, localization, or characterisation in structural materials based on obtained UGW data. These methodologies

are classified as physics-based, data-driven, or hybrid models [8, 4]. Physics-based models, such as Finite Element Method (FEM) [9] and Boundary Element Method (BEM) [10], are based on physical principles that simulate UGW propagation in specific materials under pre-defined environmental conditions. These methods are associated with high modeling accuracies but have the major disadvantage of high computational costs and lack of generalization [11]. In contrast, data-driven models mostly rely on obtained and labeled UGW data. Numerous researchers have utilized these mostly machine learning-guided techniques [12, 13, 14, 15], but their effectiveness is heavily reliant on the quantity and quality of available training data. Purely data-driven machine learning approaches have not shown to be entirely feasible due to the lack of robustness, interpretability and limited training data [8]. The potential of hybrid models to incorporate the advantages of both domains is demonstrated by their integration of physics-based and data-driven methodologies [4]. A potential method is represented by physics informed machine learning models, where training of data-driven models is constrained by incorporating prior physical knowledge. The research of Raissi et al. [16] introduced these so-called PINNs, which integrate physical laws into deep learning architectures. These models possess robust generalization capabilities, allowing for reliable predictions even in data-sparse regimes [17]. Furthermore, PINNs are highly effective in solving complex inverse problems in a variety of application domains. Shukla et al. [18] introduce a PINN approach trained to identify and characterize surface-breaking cracks in aluminum plates using ultrasonic surface acoustic wave data. The model incorporates the acoustic wave equation and employs adaptive activation functions, demonstrating improved convergence and offering a promising model for addressing ill-posed inverse problems in various applications. Zhou and Xu [19] propose a method utilizes a PINN based on a classic MLP to reconstruct a "pseudo-pristine" wavefield from measured data by embedding the Kirchhoff-Love plate theory into the training process. Because the method lacks explicit boundary and initial conditions, it requires a dense dataset of sampled measurement points to accurately resolve the high-order derivatives and solve the ill-posed inverse problem for the plate's material parameters. Mehtaj and Banerjee [20] provide a comprehensive review, which classifies and analyzes four Scientific Machine Learning (SciML) architectures regarding their efficacy in modeling acoustic, elastic and guided wave propagation. The study concludes that while PINNs successfully address traditional limitations like discretization errors and the curse of dimensionality, they still face significant challenges regarding convergence, training time and difficulties in optimizing high-dimensional non-convex loss functions.
While conventional PINNs offer various advantages, they come with challenges, particularly in high-frequency regions. Spectral bias, also known as the F-Principle, describes the tendency of deep neural networks to prioritize the convergence of low-frequency components of a target function before high-frequency ones during gradient-based optimization [21, 22]. High-frequency oscillations in the PDE solution introduce rapid variations in the residual loss, which significantly increases the non-convexity of the objective function. This leads to optimization problems for standard PINNs where gradient-based solvers become trapped in sub-optimal local minima. Moreover PINNs struggle with numerical instability when dealing with partial differential equations (PDEs) in high-frequency regimes and convergence issues due to the multi objective optimization of the different loss terms [23]. To address spectral bias in PINN modelling, Moseley et al. [24] propose a domain decomposition approach for training PINNs, whereby the global domain is divided into multiple overlapping subdomains. Although these so-called Finite basis PINNs effectively mitigate spectral bias, they introduce architectural complexity that requires precise, problem-specific tuning of hyperparameters, including training schedules, subdomain interface alignment and window function configurations.
To address the challenges associated with UGW PINN modeling, this study proposes a PI-SIREN model architecture leveraging sinusoidal representation networks or SIRENs [25] for reconstructing UGW for isotropic plate-like structures in the time domain. A novel training methodology is proposed that integrates causal weighting [26] into a hybrid optimization framework. This strategy sequentially leverages first-order and quasi-second-order optimization methods, employing residual-based adaptive refinement (RAR) [27] between the training phases to dynamically increase collocation point sampling in high-error regions. The proposed architecture demonstrate superior accuracy in reconstructing the high-frequency wave-field in comparison to conventional PINN models.

2. Problem Formulation

Reliable damage identification in UGW inspection requires a physically consistent pristine baseline against which measured structural responses can be compared [14]. However, in practical deployments, acquiring high-fidelity experimental baselines across all operational conditions is rarely feasible. While accurate, finite element (FE) simulations are computationally expensive for repeated or real-time evaluation. Spatial measurements further constrain the available information to a sparse set of spatio-temporal observations, from which purely data-driven approaches cannot reliably reconstruct the full wave field without additional physical constraints. This work formulates the forward problem of lamb wave propagation as a physics-informed regression task to address these limitations. The structure under consideration is a homogeneous, isotropic, linear-elastic aluminum plate of dimensions $200 \times 200 \times 3$ mm. The material is characterised by a Young's modulus $E = 70$ GPa, Poisson's ratio $\nu = 0.33$ and density $\rho = 2700$ kg/m^3. The plate is excited by a five-cycle Hann-windowed sinusoidal burst at a center frequency of $f_0 = 50$ kHz. The excitation signal is modeled as a concentrated out-of-plane force of amplitude $F = 1$ N applied at the top surface of the plate at the center $(x, y) = (100, 100)$ mm. Reference displacement fields are obtained from transient dynamic FE simulations conducted in ABAQUS/Explicit. The plate is discretized using 8-node linear brick elements with reduced integration and hourglass control (C3D8R). The structured mesh ensures a minimum of 30 elements per shortest wavelength across all relevant lamb wave modes, as established through a preceding mesh convergence study. The spatial discretization comprises 8 elements through the thickness and 350 elements in both the x- and y-directions, yielding a total of 980,000 elements. Transient dynamics are solved using ABAQUS automatic time-increment scheme over a total simulation duration of $T_{\text{sim}} = 3 \times 10^{-4}$ s. For model training, only the structural response after excitation is considered. The observation window is therefore restricted to $t \in [t_0, T_{\text{sim}}] = [1.1 \times 10^{-4}, 3.0 \times 10^{-4}]$ s, corresponding to a training horizon of $T = 1.9 \times 10^{-4}$ s. Therefore, the governing PDE can be enforced in its homogeneous form without an explicit source term. This avoids the need to model the spatially and temporally localised forcing within the PINN residual. Full field displacement frames at a sampling rate of $\Delta t = 1 \times 10^{-6}$ s are extracted, yielding 191 temporal dataframes for training. From these frames, a sparse, fixed set of N_d spatio-temporal observation points is extracted to emulate the limited data coverage of a realistic inspection scenario and serves as the supervised training data for the PINN model.

The objective is to train a sinusoidal representation network $\mathscr{N}_{\boldsymbol{\theta}} : (\bar{x}, \bar{y}, \bar{t}) \mapsto \bar{\boldsymbol{u}}$ that simultaneously satisfies the governing plate equations in the temporal domain $\Omega \times [0, T]$, the free-edge boundary conditions, the initial conditions and data fidelity at the sparse observation points. The resulting model constitutes a continuous, differentiable surrogate of the pristine plate response, queryable at arbitrary spatio-temporal coordinates without repeated simulation. Two plate theories are evaluated within this PI-SIREN framework: Kirchhoff-Love (KL) and Mindlin-Reissner (MR), as detailed in the following section.

3. Governing Equations and Problem Formulation

The propagation of elastic waves in a continuous, homogeneous and isotropic solid can be described by the Navier-Lamé equations for elastodynamics, which represent Newton's second law applied to a deformable continuum. In the absence of body forces, the governing equations are:

$$\rho\, \ddot{\boldsymbol{u}} = (\lambda + \mu)\, \nabla(\nabla \cdot \boldsymbol{u}) + \mu\, \nabla^2 \boldsymbol{u}, \tag{1}$$

where $\boldsymbol{u}(\boldsymbol{x}, t) \in \mathbb{R}^3$ is the displacement vector field, ρ is the material density and λ and μ are the Lamé constants, related to the Young's modulus E and Poisson's ratio ν by:

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}, \qquad \mu = \frac{E}{2(1+\nu)}. \tag{2}$$

Equation (1) couples all three displacement components through the dilatational term $\nabla(\nabla \cdot \boldsymbol{u})$ and the shear term $\mu\, \nabla^2 \boldsymbol{u}$, admitting both compressional (P-wave) and shear (S-wave) solutions. In plate-like structures, the interaction of these bulk wave modes with the free top and bottom surfaces gives

rise to guided lamb waves, which propagate along the plate mid-plane as symmetric (S_n) and anti-symmetric (A_n) modes with frequency- and thickness-dependent phase velocities described by the Rayleigh-Lamb dispersion relations [28, 5]. A direct numerical treatment of Equation (1) in three spatial dimensions is computationally infeasible for the spatio-temporal scales relevant to ultrasonic NDT, particularly when embedded in the loss function of a PINN. For thin plates satisfying $h \ll L$, where h is the thickness and L is the lateral dimension, the full three-dimensional problem can be reduced to a set of two-dimensional equations governing the mid-plane displacement field by exploiting the separation of scales in the through-thickness direction. Specifically, in the low-frequency regime where the flexural wavelength λ_f exceeds the plate thickness ($h/\lambda_f \ll 1$), the dominant wave mode is the anti-symmetric mode A_0, whose kinematics are well captured by classical plate theories without recourse to the full three-dimensional elastodynamic solution [28, 29]. This dimensional reduction eliminates the z-dependence of the displacement field through kinematic assumptions on the through-thickness deformation, reducing the computational domain from three to two spatial dimensions. Although this simplification is beneficial in terms of dimensionality, it comes at the cost of increased derivative order in the classical KL theory, where the governing equation is fourth-order in space compared to the second-order Navier-Lamé equations. The MR theory avoids this increase by retaining the second-order spatial structure through the introduction of two independent rotation fields, θ_x and θ_y as auxiliary unknowns, at the expense of expanding the system from one to three coupled equations [30, 31]. The resulting formulations are presented in the following sections.

3.1 Kirchhoff-Love Plate Theory

The KL theory [32] is applicable to thin plates in which the thickness is significantly smaller than the characteristic lateral dimension. The theory rests on three kinematic assumptions: (i) straight lines initially normal to the mid-surface remain straight and normal to the deformed mid-surface after deformation (Kirchhoff normality hypothesis); (ii) the transverse normal strain ε_z is negligible; and (iii) the transverse normal stress σ_z is small compared to the in-plane stresses and is therefore neglected [33]. As a consequence, transverse shear deformation is entirely absent from the formulation, which renders the theory inadequate at higher frequencies or for thicker plates where shear effects become significant. Under these assumptions the displacement field through the plate thickness is:

$$u_1(x,y,z,t) = -z\frac{\partial w}{\partial x}, \qquad u_2(x,y,z,t) = -z\frac{\partial w}{\partial y}, \qquad u_3(x,y,z,t) = w(x,y,t), \tag{3}$$

where $w(x,y,t)$ is the transverse deflection of the mid-surface and $z \in [-h/2, h/2]$ is the through-thickness coordinate. The governing partial differential equation for vertical displacement w follows:

$$D\left(\frac{\partial^4 w}{\partial x^4} + 2\frac{\partial^4 w}{\partial x^2 \partial y^2} + \frac{\partial^4 w}{\partial y^4}\right) + \rho h \frac{\partial^2 w}{\partial t^2} = 0. \tag{4}$$

The flexural rigidity D characterizes the bending stiffness of the plate:

$$D = \frac{Eh^3}{12(1-\nu^2)}. \tag{5}$$

For a plate with free edges, the natural boundary conditions require that the bending moment M_{nn} and the Kirchhoff effective shear force V_n vanish at each edge, yielding:

$$M_{nn} = -D\left(\frac{\partial^2 w}{\partial n^2} + \nu\frac{\partial^2 w}{\partial s^2}\right) = 0, \qquad V_n = -D\left(\frac{\partial^3 w}{\partial n^3} + (2-\nu)\frac{\partial^3 w}{\partial n\,\partial s^2}\right) = 0, \tag{6}$$

where n and s denote the outward normal and tangential directions at the boundaries, respectively.

3.2 Mindlin-Reissner Plate Theory

The MR theory [30, 31] extends the KL formulation by relaxing the normality hypothesis, thereby admitting transverse shear deformation. Straight lines initially normal to the mid-surface remain straight after deformation but are no longer constrained to remain normal to the deformed mid-surface. This

introduces two independent rotation fields $\theta_x(x,y,t)$ and $\theta_y(x,y,t)$ as additional kinematic degrees of freedom and the displacement field becomes:

$$u_1(x,y,z,t) = z\,\theta_x(x,y,t), \qquad u_2(x,y,z,t) = z\,\theta_y(x,y,t), \qquad u_3(x,y,z,t) = w(x,y,t). \tag{7}$$

The theory is governed by a system of three coupled partial differential equations for the three unknowns (w, θ_x, θ_y). The equation of motion in the transverse direction is:

$$C_s\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial \theta_x}{\partial x} + \frac{\partial \theta_y}{\partial y}\right) - \rho h \frac{\partial^2 w}{\partial t^2} = 0, \tag{8}$$

and the two rotation equations are defined as:

$$D\left(\frac{\partial^2 \theta_x}{\partial x^2} + \frac{1-\nu}{2}\frac{\partial^2 \theta_x}{\partial y^2} + \frac{1+\nu}{2}\frac{\partial^2 \theta_y}{\partial x \partial y}\right) - C_s\left(\theta_x + \frac{\partial w}{\partial x}\right) - I_{\text{rot}}\frac{\partial^2 \theta_x}{\partial t^2} = 0, \tag{9}$$

$$D\left(\frac{\partial^2 \theta_y}{\partial y^2} + \frac{1-\nu}{2}\frac{\partial^2 \theta_y}{\partial x^2} + \frac{1+\nu}{2}\frac{\partial^2 \theta_x}{\partial x \partial y}\right) - C_s\left(\theta_y + \frac{\partial w}{\partial y}\right) - I_{\text{rot}}\frac{\partial^2 \theta_y}{\partial t^2} = 0. \tag{10}$$

Here $C_s = \kappa G h$ is the transverse shear stiffness, with $G = E/[2(1+\nu)]$ as the shear modulus and $\kappa = 5/6$ the shear correction factor accounting for the non-uniform distribution of shear stress through the thickness. The rotary inertia is $I_{\text{rot}} = \rho h^3/12$. For free-edge boundary conditions, the vanishing of the bending moment M_{nn}, the twisting moment M_{ns} and the transverse shear force Q_n at each edge is required:

$$M_{nn} = D\left(\frac{\partial \theta_n}{\partial n} + \nu\frac{\partial \theta_s}{\partial s}\right) = 0, \qquad M_{ns} = D\frac{1-\nu}{2}\left(\frac{\partial \theta_n}{\partial s} + \frac{\partial \theta_s}{\partial n}\right) = 0, \qquad Q_n = C_s\left(\theta_n + \frac{\partial w}{\partial n}\right) = 0, \tag{11}$$

where θ_n and θ_s are the rotation components in the normal and tangential directions at the boundary. The choice between the two plate theories is guided by the frequency-thickness product $f \cdot h$, which characterizes the relative importance of transverse shear effects. For the KL theory, the flexural wavelength λ_f must remain large compared to the plate thickness. Substituting the harmonic wave ansatz $w \propto e^{i(kx-\omega t)}$, with angular frequency $\omega = 2\pi f$, into the KL governing equation (4) yields the dispersion relation:

$$D k^4 = \rho h \omega^2, \tag{12}$$

from which the flexural wavelength follows as:

$$\lambda_f = \frac{2\pi}{k} = 2\pi\left(\frac{D}{\rho h}\right)^{1/4}\frac{1}{\sqrt{\omega}}. \tag{13}$$

As the frequency-thickness product $f \cdot h$ increases, transverse shear effects become non-negligible and the KL theory progressively overestimates the phase velocity of the antisymmetric lamb wave mode. The condition for the validity of thin plate theory is commonly stated as $h/\lambda_f \ll 1$, i.e. the plate thickness must remain small compared to the flexural wavelength [28]. In this regime the MR theory, which accounts for both shear deformation and rotary inertia, provides a more accurate representation of the wave kinematics. For the present study, with an excitation frequency of $f = 50\,\text{kHz}$ and plate thickness $h = 3\,\text{mm}$, the flexural wavelength is estimated from Equation (13) as $\lambda_f \approx 24\,\text{mm}$, yielding $h/\lambda_f \approx 0.12$. This value lies in a regime where transverse shear effects are no longer entirely negligible, motivating the comparative evaluation of both theories within the PINN framework [28].

4. Methodology

4.1 PI-SIREN Architecture

PINNs incorporate physical laws as residual penalty terms within the loss function to ensure predictions remain consistent with known physics. Unlike purely data-driven methods, PINNs integrate a physics-based term derived from the residual of the governing partial differential equations [16]. Consequently, model training is cast as a multi-objective optimization problem requiring the simultaneous

minimization of multiple loss components. To evaluate the PDE residual, PINNs leverage automatic differentiation (AD) to systematically compute the exact partial derivatives of the neural network outputs with respect to their spatio-temporal input variables [16, 18]. In Figure 1 the proposed PI-SIREN architecture is illustrated for forward solving the MR formulation in the normalized spatio-temporal domain $(\bar{x},\bar{y},\bar{t}) \in [-1,1]^3$. The neural network architecture employed in this work is SIREN [25], in which the activation function is defined by a periodic sinusoidal activation of the form $\phi(z) = \sin(\omega_0 z)$, where ω_0 is a frequency hyperparameter that governs the characteristic oscillation scale of each layer. The network takes the normalized spatio-temporal coordinates as inputs and maps them to the predicted displacement fields, with a single output $\bar{w}$ for the KL formulation and three outputs $(\bar{u},\bar{v},\bar{w})$ for the MR formulation. The architecture consists of five fully connected hidden layers with $N = 256$ neurons per layer. Following the initialization scheme of [25], the weights of the first layer are drawn from a uniform distribution $\mathscr{U}(-1/n_{\text{in}}, 1/n_{\text{in}})$, where n_{in} is the number of input features, while the weights of all subsequent hidden layers are initialized from $\mathscr{U}(-\sqrt{6/n_{\text{in}}}/\omega_0, \sqrt{6/n_{\text{in}}}/\omega_0)$, ensuring that the distribution of activations is preserved throughout the network depth. The frequency hyperparameters are set to $\omega_0 = 30$ for the first layer and $\omega_0 = 30$ for all hidden layers, values that are chosen to match the trade-off between stability and the center excitation frequency of $f_0 = 50\,\text{kHz}$ in the normalized domain.

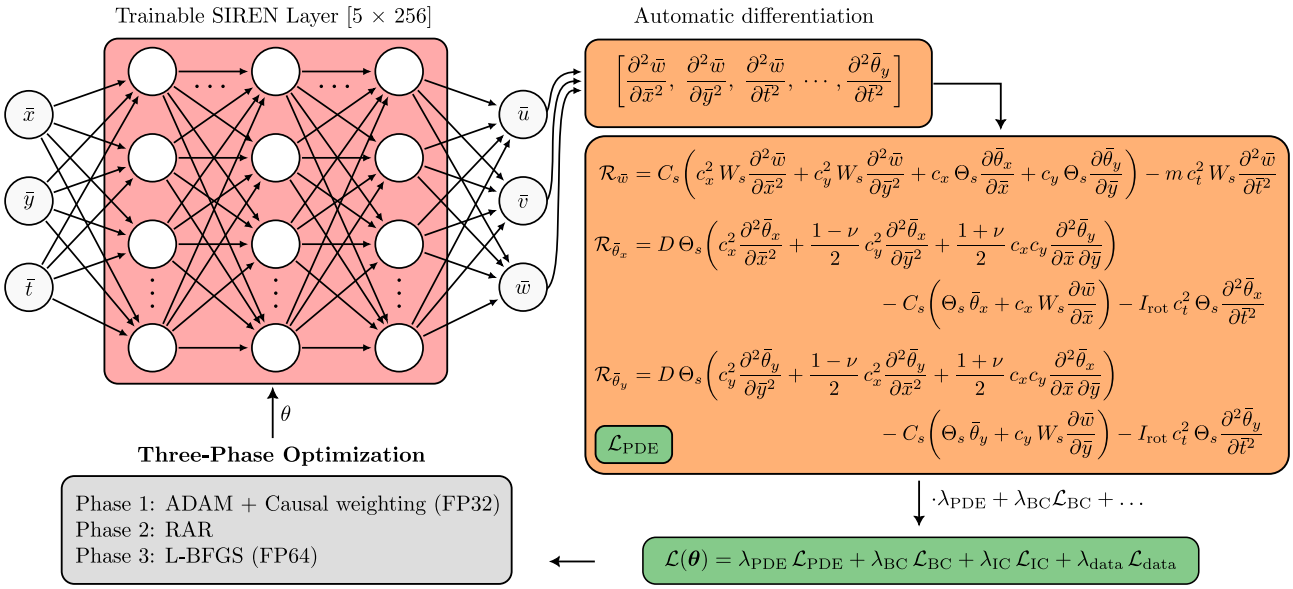


Figure 1 – Illustration of PI-SIREN model architecture for forward solving the MR formulation

Training of PINNs is significantly improved when the governing equations are expressed in a non-dimensional and normalized spatio-temporal domain, as this reduces the dynamic range of the network inputs and improves the conditioning of the loss function [34]. Normalized coordinates $(\bar{x},\bar{y},\bar{t}) \in [-1,1]^3$ are therefore introduced via the transformations:

$$\bar{x} = 2\left(\frac{x}{L}\right) - 1, \qquad \bar{y} = 2\left(\frac{y}{L}\right) - 1, \qquad \bar{t} = 2\left(\frac{t}{T}\right) - 1, \tag{14}$$

where L is the characteristic plate length and T is the total simulation duration. The corresponding inverse transformations yield the chain-rule scaling factors:

$$c_x = \frac{2}{L}, \qquad c_y = \frac{2}{L}, \qquad c_t = \frac{2}{T}. \tag{15}$$

Substituting these transformations into Equations (4)–(10) yields the normalized residual forms utilized in the loss function, as detailed in the following section 4.2.

4.2 Loss Function

The network parameters $\boldsymbol{\theta}$ are optimized by minimizing the composite loss function $\mathscr{L}(\boldsymbol{\theta})$. For the MR formulation the network predicts the surface displacement fields $\bar{u}(\bar{x},\bar{y},\bar{t})$, $\bar{v}(\bar{x},\bar{y},\bar{t})$ and $\bar{w}(\bar{x},\bar{y},\bar{t})$, corre-

sponding to the transverse and in-plane displacements at the top surface $z = h/2$, in the normalized domain $(\bar{x}, \bar{y}, \bar{t}) \in [-1, 1]^3$. The total loss is defined as:

$$\mathscr{L}(\boldsymbol{\theta}) = \lambda_{\text{data}}\mathscr{L}_{\text{data}} + \lambda_{\text{PDE}}\mathscr{L}_{\text{PDE}} + \lambda_{\text{BC}}\mathscr{L}_{\text{BC}} + \lambda_{\text{IC}}\mathscr{L}_{\text{IC}}, \tag{16}$$

where the λ parameters denote corresponding constant residual scaling terms. Each loss term of Equation (16) is described in the following. The data loss $\mathscr{L}_{\text{data}}$ enforces agreement between the predicted displacement fields and the reference measurements obtained from FE simulations at N_d spatio-temporal observation points $\{(\bar{x}_d^i, \bar{y}_d^i, \bar{t}_d^i)\}_{i=1}^{N_d}$:

$$\mathscr{L}_{\text{data}} = \frac{1}{N_d}\sum_{i=1}^{N_d}\left(\left|\bar{u}^i - \frac{u^{\text{ref},i}}{U_s}\right|^2 + \left|\bar{v}^i - \frac{v^{\text{ref},i}}{V_s}\right|^2 + \left|\bar{w}^i - \frac{w^{\text{ref},i}}{W_s}\right|^2\right), \tag{17}$$

where U_s, V_s and W_s are the characteristic displacement scales used to non-dimensionalize and normalize the reference data consistently with the network outputs. The characteristic values are determined from the maximum amplitudes of the reference data. The PDE residual loss $\mathscr{L}_{\text{PDE}}$ penalizes violations of the governing plate equations at N_f collocation points $\{(\bar{x}_f^j, \bar{y}_f^j, \bar{t}_f^j)\}_{j=1}^{N_f}$ sampled throughout the spatio-temporal domain. For the MR plate theory, three coupled residuals are enforced:

$$\mathscr{L}_{\text{PDE}} = \frac{1}{N_f}\sum_{j=1}^{N_f}\left(s_\perp^2\mathscr{R}_{\bar{w}}^2 + s_b^2\mathscr{R}_{\bar{\theta}_x}^2 + s_b^2\mathscr{R}_{\bar{\theta}_y}^2\right)_j, \tag{18}$$

where the individual residuals in the normalized domain are:

$$\mathscr{R}_{\bar{w}} = C_s\left(c_x^2 W_s\frac{\partial^2\bar{w}}{\partial\bar{x}^2} + c_y^2 W_s\frac{\partial^2\bar{w}}{\partial\bar{y}^2} + c_x\Theta_s\frac{\partial\bar{\theta}_x}{\partial\bar{x}} + c_y\Theta_s\frac{\partial\bar{\theta}_y}{\partial\bar{y}}\right) - m c_t^2 W_s\frac{\partial^2\bar{w}}{\partial\bar{t}^2}, \tag{19}$$

$$\mathscr{R}_{\bar{\theta}_x} = D\Theta_s\left(c_x^2\frac{\partial^2\bar{\theta}_x}{\partial\bar{x}^2} + \frac{1-\nu}{2}c_y^2\frac{\partial^2\bar{\theta}_x}{\partial\bar{y}^2} + \frac{1+\nu}{2}c_x c_y\frac{\partial^2\bar{\theta}_y}{\partial\bar{x}\,\partial\bar{y}}\right) - C_s\left(\Theta_s\,\bar{\theta}_x + c_x W_s\frac{\partial\bar{w}}{\partial\bar{x}}\right) - I_{\text{rot}}\,c_t^2\,\Theta_s\frac{\partial^2\bar{\theta}_x}{\partial\bar{t}^2}, \tag{20}$$

$$\mathscr{R}_{\bar{\theta}_y} = D\Theta_s\left(c_y^2\frac{\partial^2\bar{\theta}_y}{\partial\bar{y}^2} + \frac{1-\nu}{2}c_x^2\frac{\partial^2\bar{\theta}_y}{\partial\bar{x}^2} + \frac{1+\nu}{2}c_x c_y\frac{\partial^2\bar{\theta}_x}{\partial\bar{x}\,\partial\bar{y}}\right) - C_s\left(\Theta_s\,\bar{\theta}_y + c_y W_s\frac{\partial\bar{w}}{\partial\bar{y}}\right) - I_{\text{rot}}\,c_t^2\,\Theta_s\frac{\partial^2\bar{\theta}_y}{\partial\bar{t}^2}. \tag{21}$$

With $m = \rho h$ as the mass per unit area. The mid-plane rotation fields $\bar{\theta}_x$ and $\bar{\theta}_y$ are derived from the predicted surface displacements via the kinematic relation $\bar{\theta}_x = \Theta_s\,\bar{u}$, $\bar{\theta}_y = \Theta_s\,\bar{v}$, with $\Theta_s = 2U_s/h$. This allows the network to output surface displacements directly, since the rotation terms denoted by $\bar{\theta}_x$ and $\bar{\theta}_y$ are derived internally. To ensure that each residual component is of $\mathscr{O}(1)$ at initialization, yielding balanced gradient signals across all loss terms, each PDE equation is normalized by the magnitude of its dominant physical term. For the governing equations in (19)–(21) the loss residual is divided by the magnitude estimated from the maximum value of the characteristic scales:

$$\mathscr{M}_\perp = \max\left(C_s\,c_x^2 W_s,\ C_s\,c_x\,\Theta_s,\ m\,c_t^2\,W_s\right), \qquad \mathscr{M}_b = \max\left(D\,c_x^2\,\Theta_s,\ C_s\,\Theta_s,\ C_s\,c_x\,W_s,\ I_{\text{rot}}\,c_t^2\,\Theta_s\right), \tag{22}$$

The corresponding penalty weights applied in the loss function are given by the inverse of the calculated magnitudes: $s_\perp = \mathscr{M}_\perp^{-1}$, $s_b = \mathscr{M}_b^{-1}$.

The plate boundaries are simulated using free-plate boundary conditions, which strictly enforce zero normal bending moment, zero twisting moment and zero transverse shear along all unconstrained edges. These physical conditions are implemented as soft constraints at N_{BC} collocation points distributed along the normalized plate boundaries. Let $n \in \{x, y\}$ denote the coordinate normal to the current edge and $s \in \{y, x\}$ denote the tangential coordinate. The normalized boundary residuals evaluated at $\bar{n} = \pm 1$ are defined as:

$$\mathscr{R}_{M_{nn}} = D\Theta_s\left(c_n\frac{\partial\bar{\theta}_n}{\partial\bar{n}} + \nu\,c_s\frac{\partial\bar{\theta}_s}{\partial\bar{s}}\right), \tag{23}$$

$$\mathscr{R}_{M_{ns}} = D\Theta_s\frac{1-\nu}{2}\left(c_s\frac{\partial\bar{\theta}_n}{\partial\bar{s}} + c_n\frac{\partial\bar{\theta}_s}{\partial\bar{n}}\right), \tag{24}$$

$$\mathscr{R}_{Q_n} = C_s\left(\Theta_s\,\bar{\theta}_n + c_n W_s\frac{\partial\bar{w}}{\partial\bar{n}}\right). \tag{25}$$

According to $\mathscr{L}_{\mathsf{PDE}}$, the boundary constraints are normalized by the analytical estimates of their maximum expected magnitudes:

$$\mathscr{M}_{\mathsf{BC},M} = D\,c_x\,\Theta_s, \qquad \mathscr{M}_{\mathsf{BC},Q} = \max(C_s\,\Theta_s,\ C_s\,c_x\,W_s)\,. \tag{26}$$

The corresponding penalty weights applied in the loss function are given by the inverse of the calculated magnitudes: $s_{\mathsf{BC},M} = \mathscr{M}^{-1}_{\mathsf{BC},M}$ and $s_{\mathsf{BC},Q} = \mathscr{M}^{-1}_{\mathsf{BC},Q}$. The overall boundary condition loss is computed as the mean squared sum of the scaled residuals:

$$\mathscr{L}_{\mathsf{BC}} = \frac{1}{N_{\mathsf{BC}}} \sum_{j=1}^{N_{\mathsf{BC}}} \left(s^2_{\mathsf{BC},M}\,\mathscr{R}^2_{M_{nn}} + s^2_{\mathsf{BC},M}\,\mathscr{R}^2_{M_{ns}} + s^2_{\mathsf{BC},Q}\,\mathscr{R}^2_{Q_n} \right)_j . \tag{27}$$

The initial condition loss $\mathscr{L}_{\mathsf{IC}}$ imposes the measured displacement and velocity fields at the start of each time-marching window at N_{IC} spatial points:

$$\mathscr{L}_{\mathsf{IC}} = \frac{1}{N_{\mathsf{IC}}} \sum_{i=1}^{N_{\mathsf{IC}}} \left(\left|\bar{u}^i - \bar{u}^i_0\right|^2 + \left|\bar{v}^i - \bar{v}^i_0\right|^2 + \left|\bar{w}^i - \bar{w}^i_0\right|^2 + \left| c_t \frac{\partial \bar{u}}{\partial \bar{t}}\Big|^i - \dot{\bar{u}}^i_0 \right|^2 + \left| c_t \frac{\partial \bar{v}}{\partial \bar{t}}\Big|^i - \dot{\bar{v}}^i_0 \right|^2 + \left| c_t \frac{\partial \bar{w}}{\partial \bar{t}}\Big|^i - \dot{\bar{w}}^i_0 \right|^2 \right), \tag{28}$$

where $\bar{u}_0$, $\bar{v}_0$, $\bar{w}_0$ and $\dot{\bar{u}}_0$, $\dot{\bar{v}}_0$, $\dot{\bar{w}}_0$ denote the normalized displacement and velocity fields evaluated from the initial FE simulation data.

For the KL plate theory, the network predicts a single scalar output, the transverse displacement $\bar{w}(\bar{x},\bar{y},\bar{t})$. The residual of the governing biharmonic equation in the normalized domain is defined as:

$$\mathscr{R}_{\mathsf{PDE}} = D\,W_s \left(c_x^4 \frac{\partial^4 \bar{w}}{\partial \bar{x}^4} + 2\,c_x^2 c_y^2 \frac{\partial^4 \bar{w}}{\partial \bar{x}^2\,\partial \bar{y}^2} + c_y^4 \frac{\partial^4 \bar{w}}{\partial \bar{y}^4} \right) + m\,c_t^2\,W_s \frac{\partial^2 \bar{w}}{\partial \bar{t}^2}. \tag{29}$$

To balance gradient signals and maintain $\mathscr{O}(1)$ initialization, the residual is normalized by the maximum expected magnitude of its constituent terms: $\mathscr{M}_{\mathsf{PDE}} = \max\left(D\,c_x^4\,W_s,\ m\,c_t^2\,W_s\right)$. The final PDE loss is the mean squared sum of the scaled residuals evaluated over N_f collocation points:

$$\mathscr{L}_{\mathsf{PDE}} = \frac{1}{N_f} \sum_{j=1}^{N_f} \left(s_{\mathsf{PDE}}\,\mathscr{R}_{\mathsf{PDE}} \right)^2_j, \tag{30}$$

with $s_{\mathsf{PDE}} = \mathscr{M}^{-1}_{\mathsf{PDE}}$. For free edges, two macroscopic conditions must be satisfied: zero normal bending moment ($M_{nn} = 0$) and zero Kirchhoff effective shear force ($V_n = 0$). Utilizing the boundary-aligned coordinate system (n,s) where n is normal and s is tangential to the edge, the unscaled physical residuals evaluated at $\bar{n} = \pm 1$ are defined as:

$$\mathscr{R}_{M_{nn}} = D\,W_s \left(c_n^2 \frac{\partial^2 \bar{w}}{\partial \bar{n}^2} + \nu\,c_s^2 \frac{\partial^2 \bar{w}}{\partial \bar{s}^2} \right), \tag{31}$$

$$\mathscr{R}_{V_n} = D\,W_s \left(c_n^3 \frac{\partial^3 \bar{w}}{\partial \bar{n}^3} + (2-\nu)\,c_n c_s^2 \frac{\partial^3 \bar{w}}{\partial \bar{n}\,\partial \bar{s}^2} \right). \tag{32}$$

The boundary penalty weights are unified with the PDE scaling factor ($s_M = s_V = s_{\mathsf{PDE}}$). The boundary loss is applied as an unweighted global constraint over the N_{BC} edge points:

$$\mathscr{L}_{\mathsf{BC}} = \frac{1}{N_{\mathsf{BC}}} \sum_{j=1}^{N_{\mathsf{BC}}} \left(s^2_{\mathsf{PDE}}\,\mathscr{R}^2_{M_{nn}} + s^2_{\mathsf{PDE}}\,\mathscr{R}^2_{V_n} \right)_j . \tag{33}$$

Because the KL formulation yields a single partial differential equation of second order in time, it requires only the transverse displacement and its velocity as initial data:

$$\mathscr{L}_{\mathsf{IC}} = \frac{1}{N_{\mathsf{IC}}} \sum_{j=1}^{N_{\mathsf{IC}}} \left[\left(\bar{w} - \frac{w_0}{W_s} \right)^2 + \left(c_t \frac{\partial \bar{w}}{\partial \bar{t}} - \frac{\dot{w}_0}{W_s} \right)^2 \right]_j , \tag{34}$$

where w_0 and $\dot{w}_0$ are the measured reference displacement and velocity fields.

4.3 Optimization Strategy

For the training of the proposed PI-SIREN architecture, a hybrid three-phase optimization strategy is employed, which is illustrated in Algorithm 1. In the first phase, the Adam optimizer [35] is applied to minimize the composite loss function over randomly resampled collocation points at each gradient step. The first-order adaptive moment estimation of Adam is well-suited to the early stages of training. It prevents the loss function from stagnating in suboptimal local minima and is useful where large gradient magnitudes and the non-convex loss landscape make reliable curvature estimation infeasible for advanced quasi-Newton optimizer [36]. Random resampling of collocation points at each iteration reduces the risk of overfitting to a fixed grid and provides a broader coverage of the spatio-temporal domain. In the second phase, residual-based adaptive refinement (RAR) [37] is used to evaluate the residual error of the interior domain across a dense set of auxiliary collocation points. These sampled collocation points, which comprises the initial random samples and the hard points identified via RAR, is then used in the subsequent quasi-Newton optimisation phase, which is driven by the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) optimiser [38]. As demonstrated in the research of Xu et al. [39], the stiff high-order PDE gradients that arise in plate problems necessitate double-precision arithmetic (FP64) during the L-BFGS phase, since single-precision (FP32) arithmetic lacks sufficient resolution, causing convergence to sub-optimal minima [36]. Therefore all model parameters and arithmetic operations in the third phase are executed in FP64.

Standard PINNs attempt to optimise the entire spatio-temporal domain simultaneously, which can lead to violations of temporal causality when approximating future states before resolving initial conditions and early-time dynamics. Causal weighting introduced by Wang et al. [26] addresses this limitation by dynamically penalising the loss function at later times until the early-time dynamics are accurately captured. To enforce the physical direction of time during training, the causal weighting formulation introduces a dynamic penalty to the PDE residual at each discrete time step and is implemented during Phase 1 of Adam optimization. Specifically, the weight w_i assigned to the loss at time t_i decays exponentially based on the cumulative residual error of all preceding time steps, formulated as:

$$w_i = \exp\left(-\varepsilon \sum_{k=1}^{i-1} \left[\lambda_{\text{PDE}}\, \mathscr{L}_{\text{PDE}}^{(k)} + \lambda_{\text{BC}}\, \mathscr{L}_{\text{BC}}^{(k)}\right]\right), \quad i = 2, \ldots, N_T, \tag{35}$$

where $\varepsilon > 0$ is a hyperparameter defining the slope of the temporal weights. This mechanism ensures that the neural network resolves early-time dynamics before attempting to optimize later states, thereby preventing non-causal information propagation across the temporal domain. A multi-stage annealing schedule $\varepsilon_1 < \varepsilon_2 < \cdots < \varepsilon_J$ is employed, advancing to the next stage only when the minimum weight $\min_i w_i$ exceeds a prescribed tolerance δ, ensuring that each causal stage is adequately converged before tightening the causality constraint. The total loss minimized at each gradient step of Phase 1 is then defined as:

$$\mathscr{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N_T} w_i \left(\lambda_{\text{PDE}}\, \mathscr{L}_{\text{PDE}}^{(i)} + \lambda_{\text{BC}}\, \mathscr{L}_{\text{BC}}^{(i)}\right) + \lambda_{\text{IC}}\, \mathscr{L}_{\text{IC}} + \lambda_{\text{data}}\, \mathscr{L}_{\text{data}}, \tag{36}$$

where the IC and data terms are not causally gated since they provide global anchoring information. For a comprehensive breakdown of the implementation and gradient updates, it is referred to the work by Wang et al. [26].

The full time domain $[0, T]$ is partitioned into N_W non-overlapping windows $[t_{w-1}, t_w]$, each solved sequentially with a separate network parameter set $\boldsymbol{\theta}^{(w)}$. At the transition between windows, the network of the preceding window is evaluated pointwise to supply both the displacement field $\boldsymbol{u}_0 = \mathscr{N}_{\boldsymbol{\theta}^{(w-1)}}|_{t=t_{w-1}}$ and the velocity field $\dot{\boldsymbol{u}}_0 = \partial_t \mathscr{N}_{\boldsymbol{\theta}^{(w-1)}}|_{t=t_{w-1}}$ as initial conditions for the subsequent window. This formulation prevents the accumulation of temporal error that arises when a single network attempts to represent the solution over the full time and is particularly effective for the propagating wave patterns that arise in ultrasonic plate problems.

Algorithm 1 Hybrid Causal Time-Marching with Adaptive Collocation Refinement

Consider a physics-informed neural network $\mathscr{N}_{\boldsymbol{\theta}}(\boldsymbol{x},t)$ with free-edge boundary conditions enforced as soft constraints and the causally-weighted loss function

$$\mathscr{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N_T} w_i \left(\lambda_{\text{PDE}} \mathscr{L}_{\text{PDE}}^{(i)} + \lambda_{\text{BC}} \mathscr{L}_{\text{BC}}^{(i)} \right) + \lambda_{\text{IC}} \mathscr{L}_{\text{IC}} + \lambda_{\text{data}} \mathscr{L}_{\text{data}}.$$

The time domain $[0,T]$ is partitioned into N_W windows. Within each window the time axis is divided into N_T ordered slices. All weights w_i are initialized to 1.

for window $w = 1,\ldots,N_W$ **do**

Set initial condition. For $w = 1$: load measured $\boldsymbol{u}_0$, $\dot{\boldsymbol{u}}_0$. For $w > 1$: evaluate $\boldsymbol{u}_0 = \mathscr{N}_{\boldsymbol{\theta}^{(w-1)}}|_{t=t_{w-1}}$, $\dot{\boldsymbol{u}}_0 = \partial_t \mathscr{N}_{\boldsymbol{\theta}^{(w-1)}}|_{t=t_{w-1}}$.

Phase 1: Adam with causal epsilon annealing (FP32)

for $\varepsilon = \varepsilon_1,\ldots,\varepsilon_J$ **do**

for $n = 1,\ldots,S_{\max}$ **do**

(a) Compute and update causal weights:

$$w_i = \exp\left(-\varepsilon \sum_{k=1}^{i-1} \left[\lambda_{\text{PDE}} \mathscr{L}_{\text{PDE}}^{(k)} + \lambda_{\text{BC}} \mathscr{L}_{\text{BC}}^{(k)} \right] \right), \quad i = 2,\ldots,N_T, \qquad \mathbf{w} = \exp(-\varepsilon \mathbf{M}\mathbf{L}).$$

(b) Update parameters via Adam: $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \eta \nabla_{\boldsymbol{\theta}} \mathscr{L}(\boldsymbol{\theta}_n)$.

if $n \geq S_{\min}$ **and** $\min_i w_i > \delta$ **then**

break

end if

end for

end for

Phase 2: Hybrid Residual-Adaptive Refinement (RAR)

Evaluate residual magnitudes $r_i = \|\mathscr{R}(\boldsymbol{x}_i,t_i;\boldsymbol{\theta})\|^2$ on a set of dense candidate points N_{cand}. Select refined collocation set for L-BFGS:

$$\mathscr{T}_{\text{L-BFGS}}^{(w)} = \underbrace{\text{top}_{K_{\text{hard}}}(r_i)}_{\text{hard points}} \cup \underbrace{\text{rand}_{K_{\text{rand}}}(\mathscr{T}^{(w)})}_{\text{random points}}.$$

Phase 3: L-BFGS fine-tuning (FP64)

for $m = 1,\ldots,M_{\max}$ **until** $\|\nabla_{\boldsymbol{\theta}} \mathscr{L}\| < \tau$ **do**

$\boldsymbol{\theta}_{m+1} \leftarrow \text{L-BFGS}(\boldsymbol{\theta}_m, \mathscr{L}, \mathscr{T}_{\text{L-BFGS}}^{(w)})$

end for

Store $\boldsymbol{\theta}^{(w)} \leftarrow \boldsymbol{\theta}$.

end for

Hyper-parameters: $\delta = 0.99$, $\{\varepsilon_j\}_{j=1}^{J} = [10^{-3},10^{-2},10^{-1},10^{0}]$, $K_{\text{hard}} = K_{\text{rand}} = K/2$.

To capture sharp gradients and localized wave phenomena without over-saturating the entire domain with collocation points, a hybrid RAR strategy is employed prior the L-BFGS fine-tuning phase. Let $\mathscr{T}^{(w)}$ denote the collocation set used during the Adam phase for window w. At the transition to Phase 3, a dense candidate set $\mathscr{S}_{\text{cand}}$ of N_{cand} points is sampled uniformly across the spatial-temporal domain of the current window. The squared PDE residual magnitude:

$$r(\boldsymbol{x},t;\boldsymbol{\theta}) = \|\mathscr{R}(\boldsymbol{x},t;\boldsymbol{\theta})\|^2 \tag{37}$$

is evaluated at every point in $\mathscr{S}_{\text{cand}}$. A refined collocation set is then constructed as the union of two

complementary subsets:

$$\mathscr{T}^{(w)}_{\text{L-BFGS}} = \underbrace{\underset{K_{\text{hard}}}{\arg\operatorname{top}}\left\{r(\boldsymbol{x}_i,t_i;\boldsymbol{\theta})\right\}_{(\boldsymbol{x}_i,t_i)\in\mathscr{S}_{\text{cand}}}}_{\text{high-residual points}} \cup \underbrace{\text{rand}_{K_{\text{rand}}}\left(\mathscr{T}^{(w)}\right)}_{\text{random points}}, \tag{38}$$

where $K_{\text{hard}} + K_{\text{rand}} = K$ is the total L-BFGS collocation budget. The hard points concentrate resolution in regions where the physical constraints are most poorly satisfied, while the random component preserves global domain coverage and prevents over-concentration in a single spatial region. The set $\mathscr{T}^{(w)}_{\text{L-BFGS}}$ is used exclusively during Phase 3 and held fixed throughout the L-BFGS iterations, as the quasi-Newton method requires a stable objective function. The quasi-Newton iterations continue until the gradient norm falls below τ or the iteration budget $M_{\max}$ is exhausted. The original Adam collocation set $\mathscr{T}^{(w)}$ is not modified.

5. Results

This section presents the results of the proposed PI-SIREN framework for forward lamb wave simulation in the aluminum plate described in Section 2. The evaluation is structured around three axes of comparison. First, the capability of the trained models for forward wavefield reconstruction is analysed. Second, the trained models are tested as baseline models for damage identification and third, an elaboration of computational costs across the different trained models is presented. Six model variants are trained and evaluated to systematically assess the influence of plate theory, data supervision and training strategy on wavefield reconstruction accuracy. Table 1 summarises the configuration of each variant.

Table 1 – Overview of trained models - The asterisk (*) denotes ablation variants trained without causal time-marching in one window ($N_W = 1$, $\varepsilon = 0$).

Model	Plate theory	Data supervision	Physics losses	N_W
MR-full	Mindlin-Reissner	u, v, w	$\mathscr{L}_{\text{PDE}}$, $\mathscr{L}_{\text{BC}}$, $\mathscr{L}_{\text{IC}}$	12
MR-w	Mindlin-Reissner	w	$\mathscr{L}_{\text{PDE}}$, $\mathscr{L}_{\text{BC}}$, $\mathscr{L}_{\text{IC}}$	12
KL	Kirchhoff-Love	w	$\mathscr{L}_{\text{PDE}}$, $\mathscr{L}_{\text{BC}}$, $\mathscr{L}_{\text{IC}}$	12
MR-full*	Mindlin-Reissner	u, v, w	$\mathscr{L}_{\text{PDE}}$, $\mathscr{L}_{\text{BC}}$, $\mathscr{L}_{\text{IC}}$	1
KL*	Kirchhoff-Love	w	$\mathscr{L}_{\text{PDE}}$, $\mathscr{L}_{\text{BC}}$, $\mathscr{L}_{\text{IC}}$	1
Data-only	-	u, v, w	-	1

The MR models are trained on full data supervision (MR-full) as well as on transverse displacement w only (MR-w). This ensures a consistent basis for direct comparison with the KL formulation. To isolate the contribution of the proposed causal time-marching strategy, an ablation study is conducted in which the network is trained over the full time domain in a single window ($N_W = 1$). The Data-only baseline employs the same network architecture and training pipeline, but excludes all physics losses. This serves as a reference for quantifying the contribution of the physics constraints. The proposed PI-SIREN framework is implemented using JAX high performance array computing library [40]. All training pipelines and numerical simulations are executed on a single NVIDIA A100-SXM4 GPU with 80GB RAM at the Leibniz Supercomputing Centre (LRZ).

5.1 Wavefield reconstruction

This section outlines the results for forward wavefield reconstruction across the investigated PI-SIREN models from Table 1. All models are assessed against FE reference solutions on a held-out test set comprising $N_{\text{test}} = 5 \times 10^4$ spatio-temporal points per time frame. The primary accuracy metric used is the relative L^2 error:

$$e_{L^2} = \frac{\|\boldsymbol{u}_{\text{pred}} - \boldsymbol{u}_{\text{ref}}\|_2}{\|\boldsymbol{u}_{\text{ref}}\|_2} \tag{39}$$

According to section 2. for model training, only the structural response after excitation is considered. The observation window is therefore restricted to $t \in [t_0, T_{\text{sim}}] = [1.1 \times 10^{-4},\ 3.0 \times 10^{-4}]$ s, corresponding to the temporal domain of interest [19]. The training dataset employs a stratified sampling strategy with $N_d = 500$ fixed spatial positions distributed uniformly across all time frames. From each FEM frame, these nodes are selected by dividing the plate into $\lceil \sqrt{N_d} \rceil \times \lceil \sqrt{N_d} \rceil$ rectangular cells and drawing one node randomly from each occupied cell. The same fixed set of spatial indices is reused across all time frames, while training and test sets are drawn from different node pools to prevent data leakage.

All model variants share a fixed hyperparameter configuration, as detailed in Table 2. The scaling factors of the composite loss function from eq. 16 are initially chosen to ensure roughly equal loss magnitudes at the start of training. Due to the high-order spatial derivatives in the KL formulation, λ_{PDE} is drastically downscaled to 10^{-3} for time marching training. Conversely, λ_{data} and λ_{IC} are heavily weighted (10^4) to enforce convergence of the data and initial constraints [26]. The training protocol

Table 2 – Utilized training hyperparameters for each model variant

Parameter Description	Symbol	MR-full	MR-w	KL	MR-full*	KL*	Data-only
Physics Loss Weighting	λ_{PDE}	1	1	1×10^{-3}	1	1	1
Data Loss Weighting	λ_{data}	5000	5000	10000	5000	5000	5000
Boundary Loss Weighting	λ_{BC}	10	10	1000	10	10	10
Initial Loss Weighting	λ_{IC}	5000	5000	10000	5000	5000	5000
Number of Causal Windows	N_w	12	12	12	1	1	1
Time Slices per Window	N_T	40	40	25	50	50	-
Adam Learning Rate	η	2×10^{-5}					
Adam Collocation Points	N_f	500	500	100	500	400	-
Boundary Collocation Points	N_{BC}	400	200	50	400	200	-
Max. Adam Epochs	$S_{\max}$	20000	20000	40000	80000	30000	20000
RAR Candidate Points	N_{cand}	10000	10000	-	10000	2000	-
L-BFGS Optimization	-	✓	✓	—	✓	✓	✓
L-BFGS Collocation Points	N_f	200	200	100	200	200	-
Max. L-BFGS Iterations	$M_{\max}$	10000	10000	—	20000	50000	10000

leverages the hybrid causal time-marching algorithm (Algorithm 1), which distributes the time axis into N_T temporal slices. During the Adam optimization phase, collocation points are randomly resampled each epoch and the causal weighting scheme employs $\varepsilon \in [0.001, 0.01, 0.1, 1.0]$ with a convergence tolerance of $\delta = 0.99$. If this threshold is not met, training proceeds until a predefined maximum epoch limit. Prior to the L-BFGS phase, Residual-based Adaptive Refinement (RAR) evaluates the PDE residual on a dense grid of N_{cand} points per time slice. The algorithm ranks these points by error magnitude and selects the K_{hard} highest-residual points alongside an equal number of randomly sampled points (K_{rand}) to form the L-BFGS collocation set. This ensures a balanced distribution of fixed (hard) and random spatial collocation points within the domain. Boundary conditions are strictly enforced at N_{BC} points per temporal step. For the most complex configuration (MR-full), this results in a total of $500 \times 40 = 20{,}000$ $(N_f \times N_T)$ interior collocation points distributed across the temporal window.

The convergence behavior of the proposed models (MR-full, MR-w and KL) is analyzed in Figure 2. The left panel displays the total training loss $\mathscr{L}$, while the right panel shows the relative L^2 error over the training iterations for the initial (W1) and final causal window (W12), respectively. The results indicate that while the Adam optimizer with causal epsilon annealing effectively reduces the initial error, the transition to the L-BFGS optimizer has a heavy impact on the total training loss. As seen in the sharp decline following the vertical dashed lines, the L-BFGS phase drives the residuals towards their minimum. This semi second-order optimization is crucial for fine-tuning the model on the dense collocation sets generated by the RAR step. The selection of L-BFGS is motivated by the need to fit the large number of model parameters required to capture high-frequency wave content within GPU memory limits. The L-BFGS phase is omitted for the KL model because the fourth-order derivatives of the biharmonic operator restrict the collocation set to a sparse grid within the GPU memory limits.

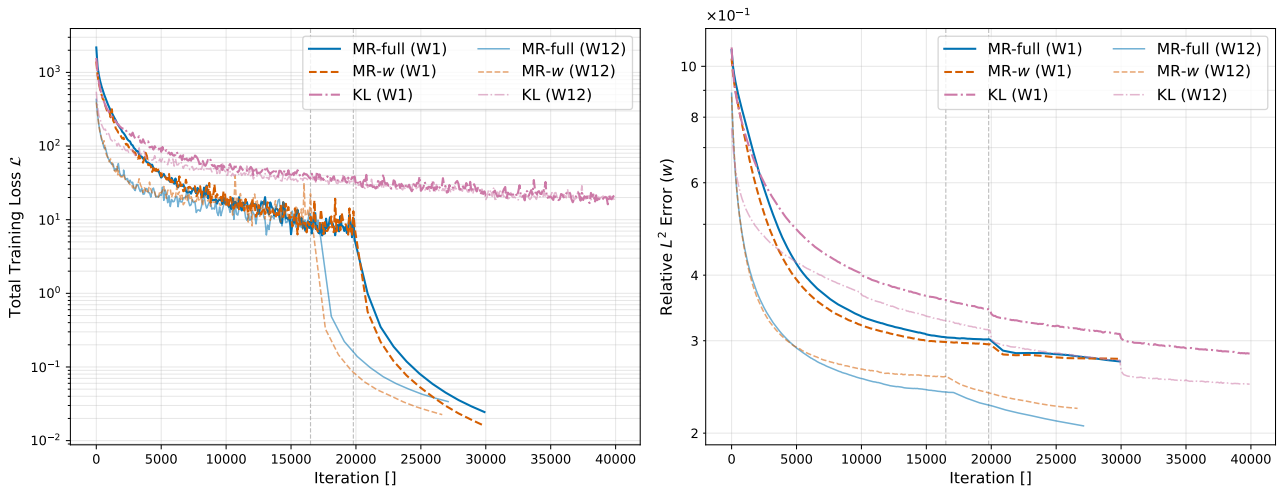


Figure 2 – Loss convergence and relative L^2-Error for the PI-SIREN models trained with Algorithm 1, displayed for the initial window (W1) and the final window (W12)

It is oberserved, that the KL model trained with L-BFGS and hybrid causal time-marching algorithm, rather tends to diverge on the sparse collocation grid than improve upon the solution obtained by Adam. Although higher-order optimizers can improve convergence, specific methods such as the advanced quasi-Newton optimizers Self-Scaled BFGS or Self-Scaled Broyden proposed in [36] are not utilized in this framework due to their prohibitive computational expense. Additionally, it should be noted that baseline MLPs with tanh activations failed to converge on the underlying problem with high-frequency data.

Figure 3 presents the relative L^2 error for the displacement components u, v and w across all investigated models. The MR-full configuration achieves the lowest error for all components. Notably,

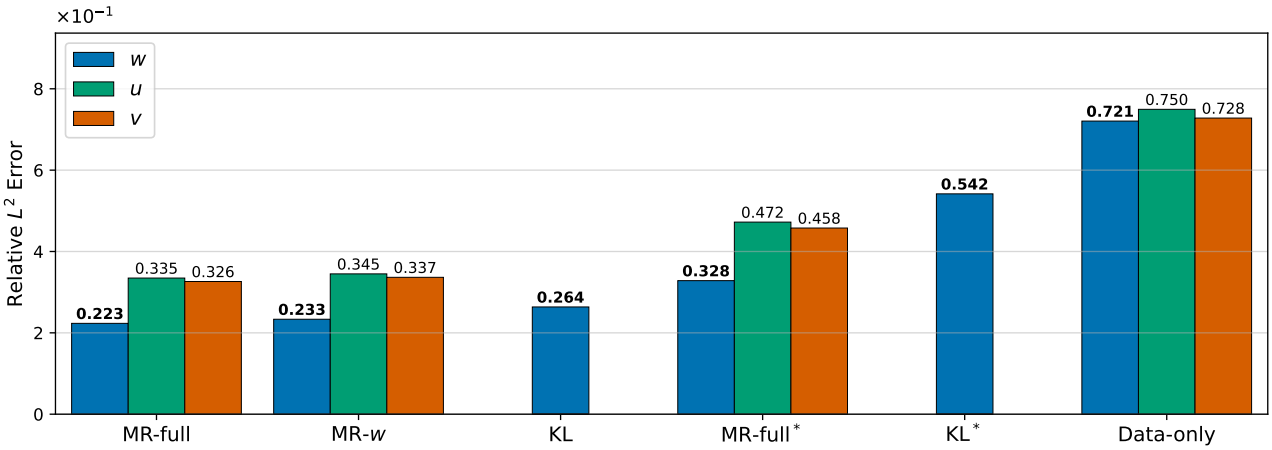


Figure 3 – Relative L^2 error for displacement components u, v, w across all model variants

the MR-w model, trained exclusively on transverse displacement data (w), yields comparable reconstruction accuracy for the in-plane displacements (u, v) as the fully supervised MR-full model. This indicates that the physics-informed constraints effectively guide the prediction of unobserved kinematic variables. The KL model reports errors only for w, consistent with its formulation. Both the ablation variants (*) and the Data-only baseline exhibit substantially higher errors, confirming the necessity of the causal time-marching strategy and physics-informed constraints.

Figure 4 presents a spatial comparison of the transverse displacement field w reconstructed by the MR-full model against the Data-only baseline, evaluated at $t = 1.5 \times 10^{-4}$ s. The top row displays the FEM reference solution, the MR-full prediction and the Data-only prediction, respectively. The bottom row shows the normalized absolute residual maps $|\Delta w|/w_{\max}$ for each model. The MR-full configuration captures the high-frequency wave interference patterns with high spatial fidelity, yielding

a relative L^2 error of $e_{L^2} = 0.207$. In contrast, the Data-only baseline exhibits pronounced localized deviations and a significantly higher error ($e_{L^2} = 0.699$).

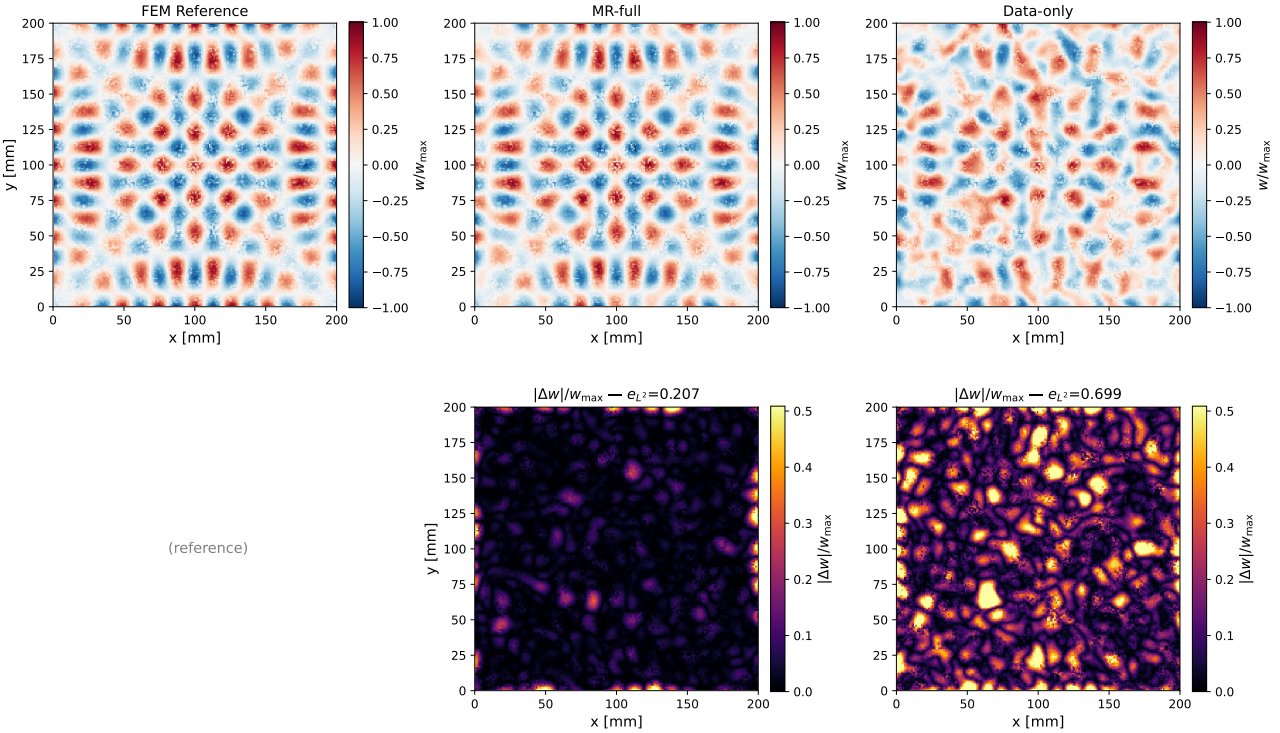


Figure 4 – Spatial reconstruction of the transverse displacement field w and corresponding residual maps for MR-full and Data-only baseline, compared against the FEM reference solution

Figure 5 illustrates the transverse displacement w over time at a specific spatial location (60.0, 140.0) mm, comparing the trained models against the FE ground truth. The physics-informed models, specifically MR-full and MR-w, demonstrate superior fidelity, closely tracking the FE wave propagation and accurately resolving the high-frequency oscillations throughout the entire temporal domain. Notably,

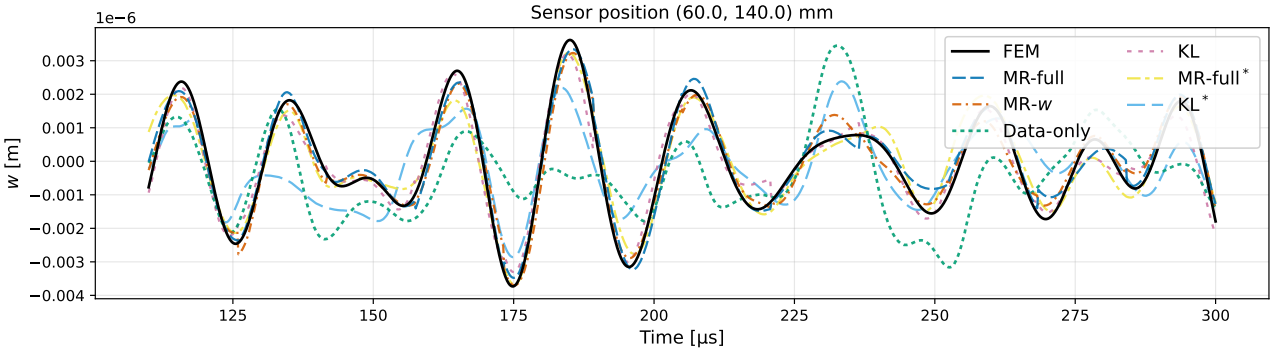


Figure 5 – Transverse displacement w over time at sensor position (60.0, 140.0) mm, comparing trained PI-SIREN models against FEM ground truth

despite being supervised on only temporal frames spanning the observation window, the trained models reconstruct the transient wavefield at arbitrary temporal resolution with high fidelity against the reference FE solution. Introducing the physics constraints yields efficient interpolation between the sparse data snapshots. While the ablation variants and the KL model capture the general wave behavior, they exhibit noticeable deviations in amplitude and phase compared to the full causal models. Conversely, the Data-only baseline fails to reconstruct the wavefield, exhibiting severe discrepancies in both amplitude and frequency response as the simulation progresses.

Table 3 quantifies the model performance averaged over $N_s = 80$ sensor positions distributed across

the spatial domain. The evaluation includes the relative L^2 error, Pearson correlation coefficient (ρ) and normalized root mean square error (NRMSE). As expected, the MR-full model achieves the lowest error bounds and highest correlation, indicating superior fidelity. The ablation variants and Data-only model demonstrate progressively higher errors, confirming the importance of the physics-informed constraints and the causal training strategy.

Table 3 – Sensor-level error metrics averaged over N_s sensor positions.

Model	e_{L^2}	ρ (Pearson)	NRMSE
MR-full	$\mathbf{0.108 \pm 0.059}$	$\mathbf{0.994 \pm 0.006}$	$\mathbf{0.025 \pm 0.013}$
MR-w	0.111 ± 0.063	0.993 ± 0.007	0.026 ± 0.015
KL	0.185 ± 0.050	0.984 ± 0.009	0.043 ± 0.011
MR-full*	0.194 ± 0.106	0.977 ± 0.025	0.045 ± 0.024
KL*	0.365 ± 0.202	0.910 ± 0.096	0.085 ± 0.045
Data-only	0.531 ± 0.268	0.808 ± 0.176	0.124 ± 0.060

5.2 Damage Identification using the PI-SIREN Baseline Model

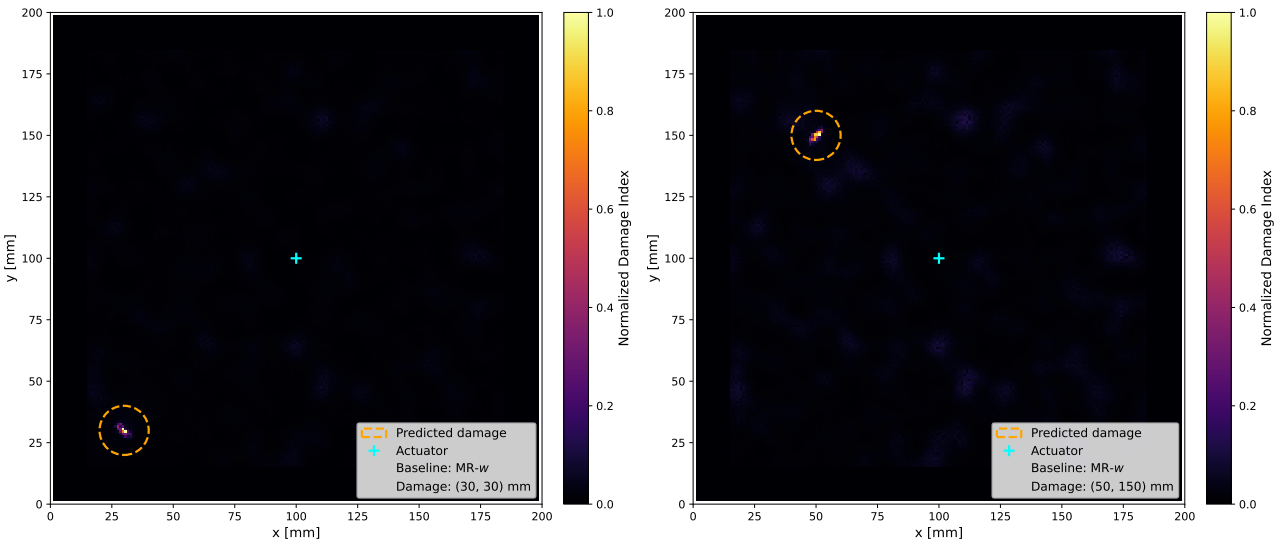
In this section, the trained PI-SIREN model is deployed as a baseline model for structural damage identification and localization. This capability is demonstrated by evaluating the predictions of the model against FE simulation data derived from a damaged plate. To simulate structural damage within the FE environment, a localized anomaly is introduced into the FE mesh. This is achieved by deleting nodes to create a physical discontinuity and defining contact interactions between the separated surfaces to capture nonlinear wave scattering and local friction effects. By subtracting the PINN predicted baseline wavefield $w_{\text{PINN}}(\boldsymbol{x},\boldsymbol{y},t_k)$ from the measured FE-simulated wavefield $w_{\text{meas}}(\boldsymbol{x},\boldsymbol{y},t_k)$, at each spatial snapshot t_k, the scattered wavefield is isolated according to $w_{\text{scat}}(\boldsymbol{x},\boldsymbol{y},t_k) = w_{\text{meas}}(\boldsymbol{x},\boldsymbol{y},t_k) - w_{\text{PINN}}(\boldsymbol{x},\boldsymbol{y},t_k)$. This scattered field isolates the damage-induced wave components. To amplify the weak signatures of the scattered wavefield, the two-dimensional discrete Teager Energy Operator (TEO) is applied to the scattered field at each time frame, following the methodology outlined in [19]:

$$\Psi[i,j] = 2w_{\text{scat}}[i,j]^2 - w_{\text{scat}}[i-1,j]w_{\text{scat}}[i+1,j] - w_{\text{scat}}[i,j-1]w_{\text{scat}}[i,j+1] \tag{40}$$

The TEO enhances regions of oscillatory scattering while simultaneously suppressing spatially distributed reconstruction errors inherent to the PINN model. The absolute TEO output is accumulated over all available time frames to formulate a cumulative damage index (DI) map $\text{DI}(\boldsymbol{x},\boldsymbol{y}) = \sum_k |\Psi(\boldsymbol{x},\boldsymbol{y},t_k)|$. Within this map, the damage location is represented as a peak. In Figure 6 the resulting damage index map for the MR-w model is shown. For demonstration purposes the accumulated Damage index for two spatially introduced damage positions is illustrated. Figure 6 presents the accumulated damage index $\text{DI}(x,y)$ of the MR-w model obtained for a partial-thickness crack introduced in the FE simulation at coordinates $(30, 30)$ mm and $(50, 150)$ mm, respectively. The defect is modelled as an out-of-plane contact crack with a length of 8.0 mm, oriented at $+/-45^\circ$ with respect to the x-axis and extending from a depth of 1.0 mm to 3.0 mm below the upper plate surface. It is shown that the true damage locations identified in the test data are accurately captured by the baseline MR-w model. For this model, an overall absolute error of 0.3 mm between the TEO-based $\text{DI}(x,y)$ and the centre of the true damage location is achieved, demonstrating the capability of the model serving as a baseline comparison model. In contrast, the purely data-driven model (Data-only) failed in damage location prediction with an absolute error of 57.8 mm. The error in prediction of the true damage location is strongly correlated with the wavefield reconstruction capabilities of the corresponding model.

5.3 Computational Costs and Efficiency Analysis

The final evaluation of the proposed methodology focuses on the computational efficiency and the resource overhead for model training. To ensure an equal assessment of training costs, all models



(a) True damage at $(30, 30)$ mm

(b) True damage at $(50, 150)$ mm

Figure 6 – Predicted TEO based normalized cumulative Damage Index on the spatial domain for baseline MR-w model

are trained until the relative L^2 error against the FEM test data no longer decreases. The training protocol strictly follows the convergence criteria outlined in Section 4.3 For the causal time-marching strategy, the optimization of each temporal window and epsilon value proceeds until the causal weight threshold $\delta = 0.99$ is satisfied or the predefined maximum number of Adam epochs is reached. Figure 7 provides a dual-perspective analysis of these computational expenses. The left plot in Figure 7

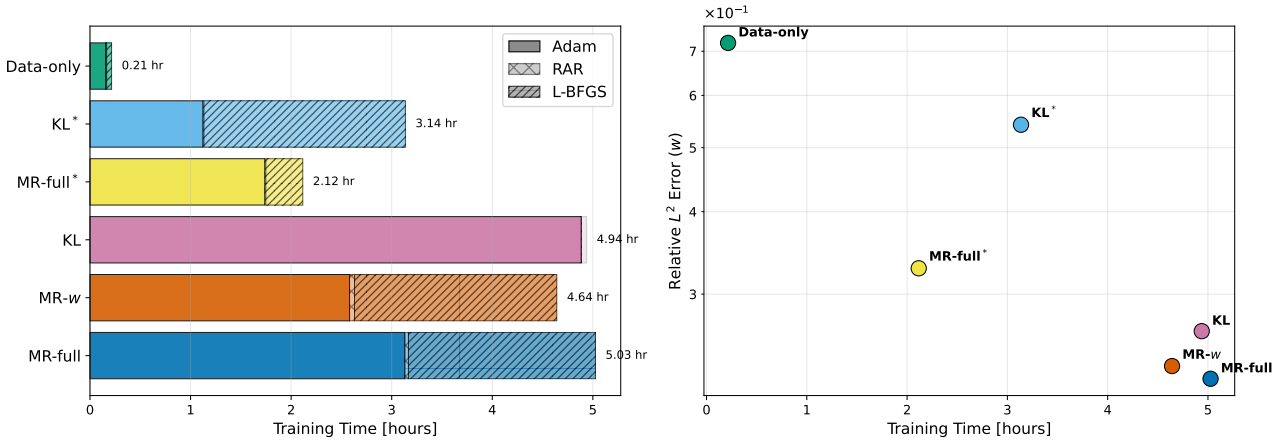


Figure 7 – Computational performance analysis, breakdown of total training time into optimization phases (Adam, RAR, L-BFGS) for each model configuration (left) and relative L^2 error as a function of training time (right)

decomposes the total training time into the contributions of the three optimization phases: the initial Adam optimization with causal epsilon annealing, the subsequent RAR and the final L-BFGS fine-tuning phase. The right plot shows the Relative L^2 error against the total training duration, illustrating the efficiency of each model variant under the defined stopping criteria. As expected, the data-driven baseline Data-only exhibits the lowest computational costs, requiring only 0.21 hours. However, this efficiency comes at the cost of physical adherence. Lacking the constraints necessary to enforce continuity, the model overfits to the discrete data samples and fails to generalize effectively across regions without supervision. This results in the highest relative L^2 error. Introducing physics-informed constraints significantly increases the computational costs. The ablation studies (*) isolate the cost of the causal time-marching strategy ($N_W = 12$). The comparison between MR-full* (2.12 hours) and MR-

full (5.03 hours) reveals that the multi-window causal propagation roughly doubles the computational cost. Similarly, the KL* model completes in 3.14 hours, whereas the full causal KL model requires 4.94 hours. Enforcing temporal consistency across the time domain via the windowed approach introduces a substantial, yet consistent, computational overhead. Despite the increased costs, the accuracy gains justify the computational expense. The MR-full model achieves a relative L^2 error of 2.23×10^{-1}, representing a $32\,\%$ reduction compared to the ablation variant MR-full* (3.28×10^{-1}). The cost of this performance improvement is an additional 3 hours of training time. The KL model follows a similar trend, achieving an error of roughly 2.64×10^{-1} at the cost of 4.94 hours, decreasing the L^2 error in comparison to the ablation variant KL* by approx. $50\,\%$. The proposed hybrid causal Time-Marching algorithm necessitates a higher computational budget compared to standard optimization strategies. The training process is computationally intensive due to the sequential nature of the causal windows. However, the trade-off is favorable: the method ensures solutions with increased accuracy and physical consistent error profiles that cannot be achieved by data-driven baselines or simplified ablation variants.

6. Conclusion

The proposed hybrid causal time-marching algorithm demonstrates superior performance in wave-field reconstruction within the high-frequency domain. By enforcing temporal consistency and causality, the method effectively guides the training process, although this comes at the cost of increased computational expense compared to standard optimization strategies. The results indicate that the PI-SIREN model MR-full, utilizing full data supervision (u, v, w), achieves the lowest relative L^2 error. Notably, the MR-w configuration, trained solely on transverse displacement data (w), yields comparable reconstruction results, suggesting that physics-informed constraints successfully infer unobserved kinematic variables such as in-plane displacements and rotations. When applied to damaged FE-simulation data, the models exhibit high fidelity in reconstructing damaged states. The integration of the TEO from [19] proves effective for damage identification, isolating damage zones and yielding low prediction errors of true damage locations. This confirms that the method is not only capable of forward wave propagation but also serves as a robust tool in the field of non-destructive testing. A primary limitation of this study is the exclusive use of clean, noise-free FEM data for training. Real-world applications often involve sensor noise, which is not accounted for in the present study. Furthermore, while the current investigation focuses on a single configuration, the primary benefit of PINNs emerges when models are parameterized to capture variations in material properties or plate geometries. While the offline training for such parameterized models entails high computational costs, the result is a highly efficient surrogate framework capable of real-time predictions. This offers a significant advantage over traditional forward FE simulations.

Future research will focus on validating the PI-SIREN framework on experimental datasets to test its robustness against noise. Furthermore, the investigation of the model in the spectral domain via Fourier transformation has been identified as a promising approach. This method has the potential to reduce computational costs by decreasing dimensionality and eliminating the necessity for causal time-marching. This would facilitate inverse damage prediction, despite the fact that transient wave behaviours would not be resolved.

7. Contact Author Email Address

Lukas Jilke (M.Sc.), German Aerospace Center (DLR)
ORCID: 0000-0001-6879-3279
mailto: Lukas.Jilke@DLR.de

8. Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

References

[1] N. Meyendorf, D. Eylon, G. S. Frankel, J. Hoffmann, M. Khobaib, H. Rösner, S. Sathish, and E. B. Shell. Degradation of aircraft structures. In Robert Hull, R. M. Osgood, Jürgen Parisi, Hans Warlimont, Norbert G. H. Meyendorf, Peter B. Nagy, and Stanislav I. Rokhlin, editors, *Nondestructive Materials Characterization*, volume 67 of *Springer Series in Materials Science*, pages 1–25. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[2] International Air Transport Association (IATA). Airline maintenance cost executive commentary, 2023.

[3] Juseong Lee and Mihaela Mitici. Multi-objective design of aircraft maintenance using gaussian process learning and adaptive sampling. *Reliability Engineering & System Safety*, 218, 2022.

[4] Lukas Jilke, Florian Raddatz, and Gerko Wende. Investigation of degradation modeling for aircraft structures: A systematic literature review. In *AIAA AVIATION 2023 Forum*, Reston, Virginia, 2023. American Institute of Aeronautics and Astronautics.

[5] C. Willberg, S. Duczek, J. M. Vivar-Perez, and Z. A. B. Ahmad. Simulation methods for guided wave-based structural health monitoring: A review. *Applied Mechanics Reviews*, 67(1), 2015.

[6] Muntazir Abbas and Mahmood Shafiee. Structural health monitoring (shm) and determination of surface defects in large metallic structures using ultrasonic guided waves. *Sensors (Basel, Switzerland)*, 18(11), 2018.

[7] A. HOUMAT. A sector fourier p -element applied to free vibration analysis of sectorial plates. *Journal of Sound and Vibration*, 243(2):269–282, 2001.

[8] Zhengyan Yang, Hongjuan Yang, Tong Tian, Deshuang Deng, Mutian Hu, Jitong Ma, Dongyue Gao, Jiaqi Zhang, Shuyi Ma, Lei Yang, Hao Xu, and Zhanjun Wu. A review on guided-ultrasonic-wave-based structural health monitoring: From fundamental theory to machine learning techniques. *Ultrasonics*, 133, 2023.

[9] Songling Huang, Shen Wang, Weibin Li, and Qing Wang. Finite element simulation of ultrasonic guided waves. In Songling Huang, Shen Wang, Weibin Li, and Qing Wang, editors, *Electromagnetic Ultrasonic Guided Waves*, Springer Series in Measurement Science and Technology, pages 237–270. Springer Singapore, Singapore, 2016.

[10] Jun Li, Zahra Sharif Khodaei, and M. H. Aliabadi. Boundary element modelling of ultrasonic lamb waves for structural health monitoring. *Smart Materials and Structures*, 29(10), 2020.

[11] Sebastian Uhlig, Ilkin Alkhasli, Frank Schubert, Constanze Tschöpe, and Matthias Wolff. A review of synthetic and augmented training data for machine learning in ultrasonic non-destructive evaluation. *Ultrasonics*, 134, 2023.

[12] Vivek Nerlikar, Olivier Mesnil, Roberto Miorelli, and Oscar D'Almeida. Damage detection with ultrasonic guided waves using machine learning and aggregated baselines. *Structural Health Monitoring*, 23(1):443–462, 2024.

[13] Mahindra Rautela and S. Gopalakrishnan. Ultrasonic guided wave based structural damage detection and localization using model assisted convolutional and recurrent neural networks. *Expert Systems with Applications*, 167, 2021.

[14] Mahindra Rautela, J. Senthilnath, Jochen Moll, and Srinivasan Gopalakrishnan. Combined two-level damage identification strategy using ultrasonic guided waves and physical knowledge assisted machine learning. *Ultrasonics*, 115, 2021.

[15] Lukas Jilke, Emy Arts, Florian Raddatz, Norbert Hosters, Marek Behr, and Gerko Wende. Federated two-stage machine learning model for ultrasonic guided wave structural health monitoring of composite structures. In *Digital Twins, AI, and NDE for Industry Applications and Energy Systems 2025*, volume 134380J. SPIE, 2025.

[16] 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*, 378:686–707, 2019.

[17] Veronika Trávníková, Eric von Lieres, and Marek Behr. Quantifying data needs in surrogate modeling for flow fields in two-dimensional stirred tanks with physics-informed neural networks. *Physics of Fluids*, 37(11), 2025.

[18] Khemraj Shukla, Patricio Clark Di Leoni, James Blackshire, Daniel Sparkman, and George Em Karniadakis. Physics-informed neural network for ultrasound nondestructive quantification of surface breaking cracks. *Journal of Nondestructive Evaluation*, 39(3), 2020.

[19] Wei Zhou and Y. F. Xu. Damage identification for plate structures using physics-informed neural networks. *Mechanical Systems and Signal Processing*, 209, 2024.

[20] Nafisa Mehtaj and Sourav Banerjee. Scientific machine learning for guided wave and surface acoustic

wave (saw) propagation: Pgnn, penn, pinn, and neural operator. *Sensors (Basel, Switzerland)*, 25(5), 2025.

[21] Ghazal Farhani, Alexander Kazachek, and Boyu Wang. Momentum diminishes the effect of spectral bias in physics-informed neural networks.

[22] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred A. Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. *International conference on machine learning*, pages 5301–5310, 2019.

[23] Shaikhah Alkhadhr and Mohamed Almekkawy. Wave equation modeling via physics-informed neural networks: Models of soft and hard constraints for initial and boundary conditions. *Sensors (Basel, Switzerland)*, 23(5), 2023.

[24] Ben Moseley, Andrew Markham, and Tarje Nissen-Meyer. Finite basis physics-informed neural networks (fbpinns): a scalable domain decomposition approach for solving differential equations. *Advances in Computational Mathematics*, 49(4), 2023.

[25] Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions.

[26] Sifan Wang, Shyam Sankaran, and Paris Perdikaris. Respecting causality for training physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 421, 2024.

[27] Lu Lu, Xuhui Meng, Zhiping Mao, and George Em Karniadakis. Deepxde: A deep learning library for solving differential equations. *SIAM Review*, 63(1):208–228, 2021.

[28] Joseph L. Rose. *Ultrasonic Guided Waves in Solid Media*. Cambridge University Press, 2014.

[29] U. Arasan, F. Marchetti, F. Chevillotte, G. Tanner, D. Chronopoulos, and E. Gourdon. On the accuracy limits of plate theories for vibro-acoustic predictions. *Journal of Sound and Vibration*, 493, 2021.

[30] Eric Reissner. The effect of transverse shear deformation on the bending of elastic plates. *Journal of Applied Mechanics*, 12(2):A69–A77, 1945.

[31] R. D. Mindlin. Influence of rotatory inertia and shear on flexural motions of isotropic, elastic plates. *Journal of Applied Mechanics*, 18(1):31–38, 1951.

[32] Augustus Edward Hough Love. Xvi. the small free vibrations and deformation of a thin elastic shell. *Philosophical Transactions of the Royal Society of London. (A.)*, (179):491–546, 1888.

[33] Valery V. Vasiliev and Evgeny V. Morozov. Laminated composite plates. In *Advanced Mechanics of Composite Materials and Structures*, pages 437–574. Elsevier, 2018.

[34] Shengfeng Xu, Yuanjun Dai, Chang Yan, Zhenxu Sun, Renfang Huang, Dilong Guo, and Guowei Yang. On the preprocessing of physics-informed neural networks: How to better utilize data in fluid mechanics. *Journal of Computational Physics*, 528, 2025.

[35] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *3rd International Conference for Learning Representations*, 2015.

[36] Elham Kiyani, Khemraj Shukla, Jorge F. Urbán, Jérôme Darbon, and George Em Karniadakis. Optimizing the optimizer for physics-informed neural networks and kolmogorov-arnold networks. *Computer Methods in Applied Mechanics and Engineering*, 446, 2025.

[37] Chenxi Wu, Min Zhu, Qinyang Tan, Yadhu Kartha, and Lu Lu. A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 403, 2023.

[38] Raghu Bollapragada, Dheevatsa Mudigere, Jorge Nocedal, Hao-Jun Michael Shi, and Ping Tak Peter Tang. A progressive batching l-bfgs method for machine learning.

[39] Chenhui Xu, Dancheng Liu, Amir Nassereldine, and Jinjun Xiong. Fp64 is all you need: Rethinking failure modes in physics-informed neural networks.

[40] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Yash Katariya, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. Jax: composable transformations of python+numpy programs, 2018.