PINNs for Electromagnetic Wave Propagation

Nilufer K. Bulut[0009−0005−1015−9570]

Izmir, Turkiye
niluferkbulut@outlook.com

Abstract. Physics-Informed Neural Networks (PINNs) solve physical systems by incorporating governing partial differential equations directly into neural network training. In electromagnetism, where well-established methodologies such as FDTD and FEM already exist, new methodologies are expected to provide clear advantages to be accepted. Despite their mesh-free nature and applicability to inverse problems, PINNs can exhibit deficiencies in accuracy and energy metrics compared to FDTD. This study demonstrates that hybrid training strategies can bring PINNs closer to FDTD-level accuracy and energy consistency.
A hybrid methodology addressing common challenges in wave propagation is presented. Causality collapse in time-dependent PINN training is addressed via time marching and causality-aware weighting. To mitigate discontinuities introduced by time marching, a two stage interface continuity loss is applied. To suppress cumulative energy drift in electromagnetic waves, a local Poynting-based regularizer is developed.
In the developed PINN model, high field accuracy is achieved with an average 0.09% NRMSE and 1.01% L^2 error over time. Energy conservation is achieved with only a 0.024% relative energy mismatch in the 2D PEC cavity scenario. Training is performed without labeled field data, using only physics-based residual losses; FDTD is used solely for post-training evaluation. The results demonstrate that PINNs can achieve competitive accuracy and energy results with FDTD in canonical electromagnetic examples, highlighting their fundamental advantages for mesh-free continuous modeling and inverse problem applications.

Keywords: Physics-Informed Neural Networks · Electromagnetic Wave Propagation · Machine Learning.

1 Introduction

In their most general form, Maxwell's equations form a coupled system of partial differential equations that describes the space-time dynamics of macroscopic electromagnetic fields. Different numerical techniques such as the finite difference time domain (FDTD) [1], finite element method (FEM) [2] and method of moments (MoM) [3] have been developed for solving these equations.

Physics-Informed Neural Networks (PINNs), introduced by Raissi et al. [4], incorporate governing partial differential equations directly into neural network training, reducing dependence on labeled data. However, applying PINNs to

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12

the time-dependent Maxwell equations remains challenging due to issues such as spectral bias, causality violations, and loss balancing. This study presents a hybrid methodology that addresses these difficulties.

Recent work has extended PINNs to Maxwell-type electromagnetic problems in several directions. For example, PINN-based formulations have been developed for transient and static electromagnetic problems in discontinuous media, with particular emphasis on first-order Maxwell systems and interface-aware representations [21]. PINN frameworks have also been explored for electromagnetic wave propagation problems, including TM/TE formulations, boundary and interface conditions, and non-dimensionalized Maxwell systems in canonical propagation settings [15]. Related studies have further considered inverse electromagnetic problems and nano-optics / metamaterial settings, indicating the broader relevance of PINNs across computational electromagnetics [19, 20].

At their core, PINNs are neural networks, and neural networks act as function approximators in the sense of function-space mapping [5]. However, for the time-dependent Maxwell system, the PDE residuals alone do not define a unique target function: without appropriate initial and boundary conditions, the problem is underdetermined and admits infinitely many solutions. In such a setting, a PINN may approximate a member of the solution family, but not necessarily the physically relevant one.

Therefore, before Maxwell's equations can be solved with a PINN, they must be cast into a well-posed initial-boundary value problem. According to Hadamard's classical framework, well-posedness requires existence, uniqueness, and continuous dependence on the data [6].

In this study, the 2D time-dependent Maxwell equations are considered in the TM_z polarization within a PEC cavity, so that the field evolution is described by the coupled components E_z, H_x, and H_y. A Gaussian pulse is used to excite the modal content. Under the source-free and lossless assumptions ($J = 0$, $\sigma = 0$), together with prescribed initial data and the PEC boundary condition $E_z = 0$ on the cavity walls, the resulting system defines a well-posed initial-boundary value problem:

$$\partial_x H_y - \partial_y H_x - \varepsilon \partial_t E_z = 0, \quad \partial_y E_z + \mu \partial_t H_x = 0, \quad \partial_x E_z - \mu \partial_t H_y = 0 \tag{1}$$

To improve numerical conditioning, a modal decomposition is adopted, as is common in PINN formulations for electromagnetic wave problems. For stable optimization, the governing equations are further cast into non-dimensional form so that the residual terms are evaluated on comparable scales. The resulting system is

$$\partial_{\bar{x}} \bar{H}_y - \partial_{\bar{y}} \bar{H}_x - \partial_{\bar{t}} \bar{E}_z = 0, \quad \partial_{\bar{y}} \bar{E}_z + \partial_{\bar{t}} \bar{H}_x = 0, \quad \partial_{\bar{x}} \bar{E}_z - \partial_{\bar{t}} \bar{H}_y = 0 \tag{2}$$

where, hereafter, the bar notation is dropped and all variables are understood to be non-dimensional.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12

1.1 Physics-Informed Loss Components

For electromagnetic cavity resonator modeling, the loss function consists of PDE, initial condition (IC), and boundary condition (BC) components, which enforce Maxwell residuals and constrain the solution at $t = 0$ and on the cavity boundaries. In this section, we develop these loss terms.

PDE Loss The non-dimensionalized system introduced in the previous section forms the basis of the PDE loss. By dropping the bar notation for simplicity, these equations are reformulated as residuals measuring how much the neural network predictions violate physical laws:

$$R_1 = \partial_x H_y - \partial_y H_x - \partial_t E_z, \quad R_2 = \partial_y E_z + \partial_t H_x, \quad R_3 = \partial_x E_z - \partial_t H_y \tag{3}$$

Specifically, R_1 enforces Ampere's law by relating the curl of the magnetic field to the time derivative of the electric field, while R_2 and R_3 enforce the components of Faraday's law by coupling the electric field gradients to the time derivatives of the magnetic field.

The PDE loss component is constructed by evaluating these residuals at collocation points distributed over the spatio-temporal domain. For aggregating residual violations, the mean squared error (MSE) metric is used:

$$L_{\text{PDE}}(\theta) = \lambda_1 \text{MSE}(R_1) + \lambda_2 \text{MSE}(R_2) + \lambda_3 \text{MSE}(R_3) \tag{4}$$

where $\text{MSE}(R_k) = \frac{1}{N_c} \sum_{i=1}^{N_c} |R_k(x_i, y_i, t_i)|^2$ for $k \in 1, 2, 3$.

Boundary Condition Loss Since a PEC cavity is considered in this study, the boundary condition requires the tangential component of the electric field to vanish on the conductor surface. For a rectangular cavity ($0 \leq X \leq a, 0 \leq Y \leq b$);

$x = X/a$, $y = Y/a$ and $\beta = b/a$ are defined. Thus, the cavity boundaries are drawn as ($0 \leq x \leq 1, 0 \leq y \leq \beta$). The coefficient β is included as an aspect ratio term so that, even though the cavity is defined as rectangular the equality $L = a$ defined during the non-dimensionalization step is preserved.

The cavity walls extend to infinity along the z-axis, therefore the z component is tangential along all walls and a single scalar condition for the electric field must be satisfied on all four walls. Thus, the BC loss can be written as

$$L_{BC}(\theta) = \frac{1}{N_{bc}} \sum_{k=1}^{N_{bc}} |E_z(x_k, y_k, t_k)|^2 \tag{5}$$

In the TM_z mode, imposing $E_z = 0$ on PEC walls provides the required electromagnetic boundary condition. The magnetic-field components are then determined implicitly through the PDE residuals and the enforced electric boundary; adding extra boundary conditions on H_x or H_y would overconstrain the system and is therefore avoided.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12

Initial Condition Loss In this study, a Gaussian pulse is selected as the excitation. The initial field configuration at $t = 0$ for the electric field distribution is expressed as

$$E_z(x, y, 0) = A \cdot \exp(-\frac{(x - x_0)^2 + (y - y_0)^2}{2\sigma^2}) \tag{6}$$

The magnetic field components are initialized to zero at $t = 0$. The pulse parameters are defined as centered in the cavity at $(x_0, y_0) = (0.5, 0.5\beta)$ with width $\sigma = 0.1$ and amplitude (A = 1.0).

The initial condition loss enforces the predicted field distribution at $t = 0$:

$$L_{IC} = \text{MSE}(E_z - E_{z_{tar}}) + \text{MSE}(H_x) + \text{MSE}(H_y)\big|_{t=0} \tag{7}$$

2 Implementation

In the theoretical formulation, expressing Maxwell's equations as a well-posed initial-boundary value problem reveals an elegant truth: when the equations are formulated correctly, the necessary physical properties emerge naturally. In the considered lossless PEC cavity setting, the PDE, IC and BC trio uniquely determine the solution. This is no coincidence, as Maxwell's equations automatically ensure energy conservation via Poynting's theorem, causality via the hyperbolic nature of the wave equations and interface continuity via the electromagnetic field boundary conditions. Therefore, at the **theoretical stage**, there is no need to define causality or energy loss; such additions would do little more than impose the existing laws again.

However, in the **implementation stage**, this changes. When training PINNs, the infinite dimensional solution space must be approximated using finite parameters and integral norms must be optimized using sampled estimators. In this discrete and local optimization based setting, scenarios that contradict physical reality are entirely plausible. For example, the optimization algorithm could violate energy conservation while minimising PDE residuals; the neural network could allow information to flow backwards in time; or field discontinuities could arise at material interfaces. These scenarios do not originate from a deficiency in the theoretical formulation, but from the intrinsic limitations of approximation and sampling based learning in neural networks. Consequently, the main goal at the implementation stage is to translate the physical guarantees implicit in theory into explicit constraints that guide the optimizer.

2.1 Training Pipeline

The core of the model is a deep fully connected neural network with 8 hidden layers, each consisting of 128 neurons. During training, to improve gradient flow and to overcome the optimization difficulties, a skip connection is applied after every second hidden layer inspired by the ResNet architecture [7]. To ensure

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12

robust performance, a set of problem specific engineering approaches has been adopted:

Input normalization. The tanh activation yields its strongest gradients near zero and saturates for distant inputs, causing gradient imbalance. To prevent this, spatial coordinates are dynamically normalized to $[-1, 1]$ depending on the cavity aspect ratio $\beta = b/a$ [16].

Sinusoidal time features. Electromagnetic wave propagation in a closed cavity inherently exhibits periodic behavior. To encode this structure, the temporal input is enriched with $\sin(2\pi t/T)$ and $\cos(2\pi t/T)$, providing the network with a natural basis for representing oscillatory field evolution [17].

Output scaling. During non-dimensionalization, the relationship $H_0 = E_0/Z_0$ was established. Since the network's randomly initialized output layer produces values of similar magnitudes for all components, the magnetic field outputs are scaled by 0.1 to ensure the starting point lies in a physically meaningful region [18].

Two-stage optimization. The optimization combines first-order and second-order methods. In the first stage, the global structure of the solution is learned for 1500 epochs using Adam with gradient clipping and a plateau-based LR scheduler. In the second stage, fine-tuning is performed with L-BFGS, which exploits local curvature information of the loss landscape.

Dynamic sampling. Collocation points are adapted according to the cavity aspect ratio, with increased density for narrow geometries. To preserve causality during time-window transitions, denser sampling is applied near the beginning of each window. Loss weights are also scheduled dynamically: $\mathcal{L}_{IC}$ is emphasized in early epochs, while $\mathcal{L}_{PDE}$ is gradually raised as training progresses.

2.2 Causality of PINNs

When the retarded solutions of Maxwell's equations are adopted, the field value at each point depends only on the initial and boundary data within that point's past light cone. This behavior is a natural consequence of causality in hyperbolic systems.

Within the PINN framework, however, the neural network treats time as a spatial dimension and minimizes the entire time axis in a single pass. During backpropagation, gradient signals from later times retroactively modify outputs at early times, violating causality.

Various approaches have been explored to address this issue, including causal training [8], time marching [9] and causal sweeping [10], XPINNs-type decompositions [11], and curriculum learning [12]. In this study, a time-domain decomposition strategy is adopted through the time-marching approach.

Time Marching The time-marching method partitions the temporal domain into sequential windows, training each separately. The temporal domain is divided into N_w contiguous windows:

$$\mathcal{T} = \{[t_0, t_1], [t_1, t_2], \ldots, [t_{N_w-1}, t_{N_w}]\} \tag{8}$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12

where $t_0 = 0$ and $t_{N_w} = T_{\max}$. For each window $\mathcal{W}_i = [t_i, t_{i+1}]$, a complete training procedure consisting of both Adam and L-BFGS stages is applied independently. With $T_{\max} = 2.0$ and $\Delta t_{\text{window}} = 0.10$, the domain is partitioned into 20 windows.

The main advantage is that the network handles limited temporal complexity within each window, focusing on a short interval where field evolution is more predictable. Based on experimental studies, $\Delta t_{\text{window}} = 0.10$ provides a good balance between accuracy and computational cost.

The sampling strategy is adapted for windowed training: fresh collocation points are generated uniformly over each window's domain, with 25% extra points concentrated within the first 10% of each window's temporal span. This denser sampling ensures stronger gradient signals for enforcing continuity at window interfaces.

Sequential Training Each window is trained conditioned on the completed solution of the previous window, enforcing causality and producing physically meaningful solutions.

In the first window $\mathcal{W}_0$, standard PINN training uses the analytic Gaussian pulse as the initial condition. For subsequent windows $\mathcal{W}_i$ $(i > 0)$, the trained model from $\mathcal{W}_{i-1}$ is evaluated at $t = t_i$ to obtain field values $[E_z, H_x, H_y]$, which serve as the initial condition. These predictions are computed without gradient tracking and stored as fixed targets. The $\mathcal{L}_{\text{IC}}$ term then penalizes deviations from these targets, enforcing field continuity across window boundaries.

Furthermore, network weights from the previous window initialize the current window's parameters. This transfer learning strategy provides a warm start that improves solution continuity.

Causality-Aware Weighting Even with time marching, a small PDE inconsistency at the beginning of a window can propagate forward through derivative terms. To address this, a causality-aware weighting mechanism is employed: for each window, time is normalized via $\tau = (t - t_{\text{start}})/(t_{\text{end}} - t_{\text{start}})$, and PDE residuals are weighted by $w_c(\tau) = \exp(-\gamma\tau)$.

With $\gamma = 2$, samples at the window start $(\tau = 0)$ receive weight 1, while those at the end $(\tau = 1)$ receive $e^{-2} \simeq 0.135$, creating a 7.4× ratio. This forces optimization to suppress early-time inconsistencies first, establishing an equation-consistent initial regime.

The within-window normalization is critical: it keeps the weighting's numerical meaning fixed per window, makes the gradient scale independent of window duration, and applies causality emphasis consistently across windows.

Interface Continuity Loss Although each time window is trained as an independent PINN, an additional constraint is needed to ensure smooth temporal evolution. A two-stage continuity mechanism is employed.

The first stage is the IC continuity loss at the start of each window. In the first window, the IC term enforces the analytic Gaussian pulse and zero magnetic

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12

field. In subsequent windows, it matches the trained solution from the previous window: E_z, H_x, and H_y at the beginning of window k are penalized against the fields at the terminal time of window $k-1$. The imported fields are detached during backpropagation to serve as fixed targets, keeping the training temporally causal.

The second stage is an interface continuity loss on a dedicated sampling set near the window boundary, comparing the current window's fields against the frozen outputs of the previous window at interface-centered points $(x_{\text{int}}, y_{\text{int}}, t_{\text{int}})$:

$$L_{\text{int}}(\theta) = \frac{1}{N_{\text{int}}} \sum_{j=1}^{N_{\text{int}}} \left| u_\theta(x_j, y_j, t_i) - u_{\text{prev}}(x_j, y_j, t_i) \right|_2^2 \tag{9}$$

Poynting Loss Maxwell's equations encode energy conservation, but a neural network does not inherently possess this structure. If not explicitly constrained, the optimizer may minimize L^2 loss while allowing cumulative energy drift. Therefore, Poynting's theorem is incorporated as a regularization term $\mathcal{L}_{\text{pyt}}$, implemented via two approaches.

Under the non-dimensionalization, the energy density is $u = \frac{1}{2}(E_z^2 + H_x^2 + H_y^2)$.

The **global approach** enforces the integral form: in an ideal PEC cavity with zero boundary flux, $\frac{d}{dt}\int_A u\, dA = 0$. Discretized via Gauss–Legendre quadrature:

$$\mathcal{L}_{\text{pyt-global}} = \left(\sum_{j=1}^{N_q} w_j \partial_t u_j \right)^2 \tag{10}$$

This formulation is practical but susceptible to cancellation of errors, where $\partial_t u > 0$ in one region compensates $\partial_t u < 0$ elsewhere.

The **local approach** enforces the differential form $\partial_t u + \nabla \cdot \mathbf{S} = 0$ at every collocation point:

$$\mathcal{L}_{\text{pyt-local}} = \frac{1}{N_c} \sum_{i=1}^{N_c} \left(\partial_t u_i + \nabla \cdot \mathbf{S}_i \right)^2 \tag{11}$$

The local approach provides pointwise control, making conservation violations visible where they occur. Comparative analysis is presented in the ablation.

Final Loss Function and Dynamic Weighting Combining all components, the total loss for any time window $\mathcal{W}_i$ is:

$$\mathcal{L}_{\text{total}} = w_{\text{pde}}\mathcal{L}_{\text{PDE}} + w_{\text{bc}}\mathcal{L}_{\text{BC}} + w_{\text{ic}}\mathcal{L}_{\text{IC}} + w_{\text{int}}\mathcal{L}_{\text{int}} + w_{\text{pyt}}\mathcal{L}_{\text{pyt-local}}$$

Note that $\mathcal{L}_{\text{PDE}}$ is already modified internally by the causality-aware weighting $w_c(\tau) = \exp(-\gamma\tau)$ from Section 2.2.

This is not a static sum. The weight vector $\mathbf{w}$ follows a dynamic schedule: during early Adam epochs, w_{ic} is increased so the model focuses on initial conditions, while w_{pyt} starts low for stability. As optimization progresses, weights shift to enforce tighter PDE and conservation compliance, guiding convergence toward physically consistent solutions.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12

3 Results and Ablation

A high-resolution FDTD simulation serves as the reference solution for evaluating the trained PINN model. The FDTD solver is not included in any stage of training; it is run only after training to generate reference data for quantitative comparison. The reference solution is generated on a Yee staggered grid with $dt = 0.5 \cdot \min(dx, dy)$ satisfying the CFL condition, using 41 equally spaced snapshots. Tables show selected time steps, but **Avg.** rows are computed over all 41 snapshots.

Two complementary field accuracy metrics are reported: NRMSE as the primary metric, selected for its robustness in wave problems where relative metrics can exhibit artificial spikes near nodal regions, and the relative L_2 error on the combined field vector $[E_z, H_x, H_y]$ as a secondary indicator. Energy behavior is evaluated through (i) instantaneous PINN–FDTD energy mismatch, (ii) each method's conservation relative to its initial value, and (iii) energy jumps at window interfaces. To ensure that the conclusions are not artifacts of the collocated evaluation grid, a native Yee-grid validation is also performed, where the discrete energy is computed using the standard staggered placement of the electric and magnetic field components.

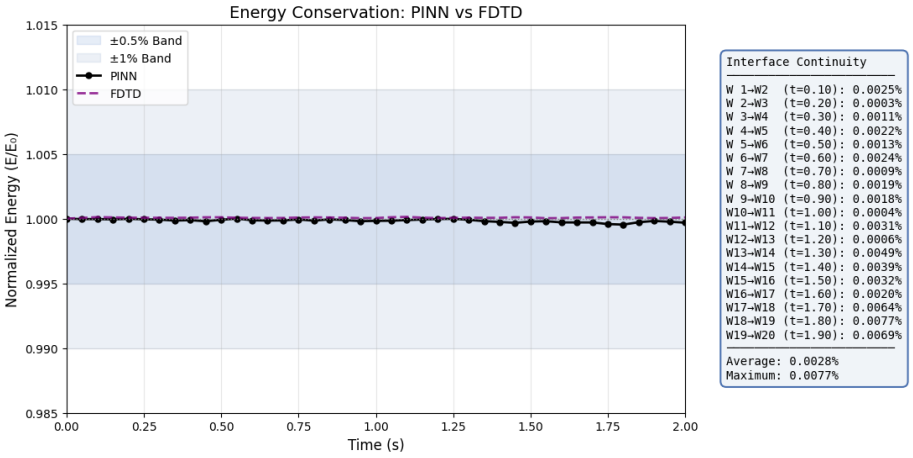

Fig. 1. Energy Conservation: PINN vs FDTD

Multiple versions of the PINN model developed in this study were trained under three primary scenarios and with different implementation approaches. The main version from which the other versions are derived will be referred to as **pinnA**.

One of the most fundamental challenges faced by time marching PINN models in the literature is the error accumulation problem [13]. In the continuous lossless PEC setting, total energy is constant; therefore deviations in discrete solvers like FDTD are treated as numerical error.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12

Table 1. pinnA: Energy and accuracy metrics.

	Energy Metrics				Accuracy Metrics	
Time	PINN.E	FDTD.E	Abs.Err	Rel.Err	NRMSE_{Tot}	$L_{2,Tot}$
t=0.0	0.015708	0.015708	1.482e-07	0.001%	0.01%	0.08%
t=0.5	0.015706	0.015710	3.210e-06	0.020%	0.10%	1.05%
t=1.0	0.015705	0.015709	3.535e-06	0.023%	0.08%	0.85%
t=1.5	0.015704	0.015709	5.139e-06	0.033%	0.10%	1.14%
t=2.0	0.015703	0.015709	6.503e-06	0.041%	0.12%	1.33%
Avg.	**0.015705**	**0.015709**	**3.759e-06**	**0.024%**	**0.09%**	**1.01%**

In neural network applications for electromagnetic problems, this error is often characterized as energy drift and most commonly manifests as decay. The underlying reason is the tendency of neural networks to converge toward trivial solutions. Rather than amplifying field magnitudes, the network tends to suppress them; this leads to amplitude damping and energy loss [14].

Although the time marching strategy is successful in preventing causality collapse, it causes energy drift to become cumulative. The terminal values of each time window are treated as the initial condition for the next window. This causes every small error formed in the previous stage to be inherited by the subsequent window. Therefore, energy drift becomes cumulative and deepens as time progresses.

The most distinctive property of **pinnA** is that it does not exhibit a cumulative energy drift pattern. The results show that when the model experiences an energy drop within a given window, it compensates for that drop in subsequent windows and brings the solution back toward the reference energy level. **pinnA** is a fully-connected neural network consisting of 8 hidden layers with 128 neurons per layer. The global time interval is partitioned into 20 consecutive windows of width $\Delta t_{\text{window}} = 0.10$, and each window is trained independently. The loss function is defined as a weighted sum of the PDE, boundary, initial condition, window-interface continuity and Poynting based energy conservation terms.

The FDTD solution is not incorporated into the training process in any form; therefore, the risk of overfitting to FDTD data is eliminated by construction. During the analysis stage, an FDTD solver is used purely as a reference to quantitatively evaluate the accuracy of **pinnA**.

Field-level accuracy of **pinnA** relative to FDTD is assessed using the Normalized Root Mean Square Error (NRMSE) and the relative L^2 norms defined in the previous section. The results for selected time steps are reported in Table 1. The reported average values are computed not only over the time steps shown in the table, but as an average over all time steps used in the analysis, $t = [0.00, 0.05, 0.10, 0.15, \ldots, 2.00]$.

As shown in the table, the NRMSE values for all three field components remain on the order of 0.1–0.2% throughout the full time interval. Even the relative L^2 error reaches only about 1.33% in the worst case and stays around 1.01% on average.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12

While accuracy metrics capture the convergence capability of the neural network, they are not sufficient on their own. To verify physical consistency, the energy metrics defined in the previous part are also employed. Since the values in the table and figure are computed on a collocated grid for the PINN solution, FDTD and PINN are additionally evaluated on a common Yee grid to confirm consistency. As shown in Table 2, the discrete energy functional of the PINN tracks the discrete FDTD energy across the entire spatio-temporal domain at a level well below 10^{-3} relative error.

Table 2. Discrete energy mismatch (Yee grid).

Metric	Value
Mean Absolute Difference	3.743911×10^{-6}
Max Absolute Difference	9.074444×10^{-6}
Mean Relative Difference (%)	0.0238
Max Relative Difference (%)	0.0578
PINN Energy Variation (%)	0.0472
FDTD Energy Variation (%)	0.0139

Four-Window Variant To examine the impact of time marching on **pinnA**, the training was repeated with $window_size = 0.50$, resulting in a four-window setup instead of 20 windows. Although reducing the number of windows increases the error level, the absolute performance of **pinnA** remains in a strong range.

Table 3. Four-Window Variant of pinnA

Time	**NRMSE**	**L2**	**FDTD.E**	**pinn.E**	**Rel.Enrgy.Err**
t=0.0	0.02%	0.27%	0.015708	0.015703	0.028%
t=0.5	0.47%	5.11%	0.015782	0.015662	0.765%
t=1.0	0.47%	5.24%	0.015603	0.015732	0.824%
t=1.5	0.52%	5.79%	0.015777	0.015715	0.395%
t=2.0	0.59%	6.59%	0.015658	0.015672	0.095%
Avg.	**0.42%**	**4.60%**	**0.015705**	**0.015697**	**0.421%**

The average $NRMSE_{\text{Total}}$ is measured as 0.42% and the average $L2_{\text{Total}}$ as 4.60%. On the energy side, agreement with FDTD is maintained, with the mean relative energy error remaining at 0.421%. Despite a relatively large initial energy mismatch of 0.028% compared to **pinnA**, the relative energy error decreases to 0.095% at $t = 2.0$. In this configuration, even with fewer windows, the energy jumps remain at 0.0187% on average and 0.0322% at worst. This indicates that the local Poynting constraint and the interface-based regularization, which are

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12

analyzed in detail in the ablation section, keep the average relative energy error below 0.5% even when wider time windows are used.

Lossy Scenario To test whether pinnA's behavior is preserved under lossy conditions, a lossy variant **pinnL** was trained with conductivity $\eta = \sigma/\varepsilon_0 = 0.5$. The conductivity term $-\eta E_z$ is added to the Ampere PDE residual, and the Poynting residual is extended to $R_{\text{pyt}} = \partial_t u + \nabla \cdot \mathbf{S} + \eta E_z^2$, where ηE_z^2 represents Joule heating.

Table 4. pinnL: Lossy scenario results.

Time	NRMSE	L_2	FDTD.E	pinnL.E	Rel.Enrgy.Err
$t = 0.0$	0.01%	0.08%	0.015708	0.015708	0.002%
$t = 0.5$	0.09%	1.04%	0.012601	0.012582	0.154%
$t = 1.0$	0.06%	0.91%	0.009353	0.009336	0.187%
$t = 1.5$	0.08%	1.26%	0.007390	0.007374	0.220%
$t = 2.0$	0.10%	1.80%	0.005858	0.005847	0.199%
Avg.	**0.08%**	**1.11%**	**0.009958**	**0.009941**	**0.176%**

Field accuracy remains at the same order as pinnA, with average NRMSE of 0.08% and L^2 error of 1.11%. The energy decreases from 0.0157 to 0.0059 ($\sim$63% reduction by $t = 2.0$), and the PINN tracks this dissipation curve with an average relative error below 0.2%. Window transition jumps remain negligible (avg. 0.0027%, max 0.0095%), confirming that the interface continuity mechanism generalizes to the lossy regime.

High Frequency Scenario To test the robustness of pinnA against spectral bias, a high frequency variant pinnHF was trained with $\sigma = 0.05$, increasing the maximum dominant frequency from 3.902 to 8.780 and the number of periods from approximately 7.8 to 17.6.

Table 5. Comparison of pinnA and pinnHF configurations.

Metric	pinnA	pinnHF
NRMSE (avg)	0.09%	0.10%
L^2 Error (avg)	1.01%	2.21%
Energy Rel. Error	0.02%	0.21%
Energy Mismatch (Yee Grid)	0.02%	0.21%
Max. Dominant Freq.	3.902	8.780
Max. Number of Periods	$\sim$7.8	$\sim$17.6

Despite the frequency content increasing by more than twofold, pinnHF demonstrated strong performance: average NRMSE of 0.10%, average L^2 error of 2.21%, and relative energy error of 0.21%. These results indicate that

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12

the methodology works effectively in the high frequency regime despite spectral bias, where neural networks tend to learn low frequency components faster. The fact that errors remain within acceptable limits suggests that time marching and Poynting-based regularization can partially compensate for spectral bias.

When the frequency band is increased further, model performance diminishes. For $\sigma = 0.03$ (approximately 29.9 periods), the average relative energy error reaches 6.862%. A large portion of this error occurs at the initial instant when the Poynting regularizer operates with low weight; with its active engagement, a decrease in energy drift was observed. It can be stated that $\sigma = 0.05$ represents a performance boundary for pinnA, beyond which stability is lost. However, the spectral bias problem can be substantially mitigated through improvements to the Poynting regularizer, which is targeted in future work.

3.1 Ablation

This ablation study isolates the cumulative energy drift behavior in time-windowed training while keeping all other components fixed. Three variants were trained with identical architecture, differing exclusively in how $\mathcal{L}_{\text{pyt}}$ is formulated:

pinnA: Local Poynting Residual. The differential form of Poynting's theorem is enforced at the same collocation points as the Maxwell residuals, penalizing $\partial_t u + \nabla \cdot \mathbf{S} = 0$ directly at every point. Since energy flux and density are locally constrained, each window establishes its own internal energy balance.

pinnGP: Global Poynting Regularizer. Instead of pointwise enforcement, pinnGP constrains the rate of change of total energy integrated over the domain, approximated via Gauss–Legendre quadrature. This produces a single scalar constraint rather than N_c independent pointwise constraints. Consequently, excess energy in one region can compensate for a deficit elsewhere through cancellation of errors, allowing local inconsistencies to persist while the global integral appears acceptable.

pinnWP: No Poynting Constraint. The energy term is omitted entirely ($\mathcal{L}_{\text{pyt}} = 0$), serving as a control to isolate the effect of energy regularization. All other loss components remain identical, so the differences in Table 6 can be attributed directly to this design choice.

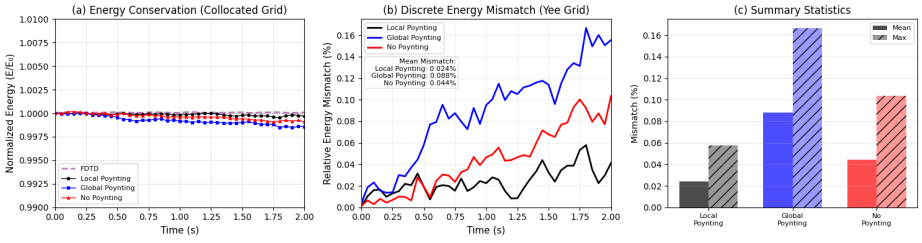

Fig. 2. Ablation: pinnA (local Poynting), pinnGP (global Poynting), and pinnWP (no Poynting constraint) compared against FDTD.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12

Table 6. Ablation study: Energy and accuracy metrics comparison.

Model	Time	Energy Metrics				Accuracy Metrics	
		PINN.E	FDTD.E	Abs.Err	Rel.Err	NRMSE_{Tot}	$L2_{Tot}$
pinnA	$t = 0.0$	0.015708	0.015708	1.482e-07	0.001%	0.01%	0.08%
	$t = 1.0$	0.015705	0.015709	3.535e-06	0.023%	0.08%	0.85%
	$t = 2.0$	0.015703	0.015709	6.503e-06	0.041%	0.12%	1.33%
	Avg.	**0.015705**	**0.015709**	**3.759e-06**	**0.024%**	**0.09%**	**1.01%**
pinnGP	$t = 0.0$	0.015707	0.015708	6.120e-07	0.004%	0.01%	0.07%
	$t = 1.0$	0.015694	0.015709	1.493e-05	0.095%	0.08%	0.90%
	$t = 2.0$	0.015685	0.015709	2.445e-05	0.156%	0.14%	1.51%
	Avg.	**0.015695**	**0.015709**	**1.385e-05**	**0.088%**	**0.10%**	**1.06%**
pinnWP	$t = 0.0$	0.015708	0.015708	2.020e-07	0.001%	0.01%	0.07%
	$t = 1.0$	0.015701	0.015709	7.296e-06	0.046%	0.08%	0.86%
	$t = 2.0$	0.015693	0.015709	1.627e-05	0.104%	0.11%	1.24%
	Avg.	**0.015702**	**0.015709**	**6.933e-06**	**0.044%**	**0.09%**	**0.99%**

As shown in Table 6, all three variants achieve comparable field accuracy against FDTD, with both NRMSE and L^2 errors remaining at the sub-percent level. This similarity is important: the differences in energy metrics cannot be explained by field accuracy, but are a direct consequence of the energy-regularization design.

In terms of energy, **pinnA** exhibits the most consistent behavior with FDTD, keeping normalized total energy within a narrow band and achieving the lowest conservation error. **pinnWP** shows the typical cumulative drift of time-windowed training: small per-window energy losses accumulate, causing monotonically decreasing energy and gradually diminishing field amplitudes. **pinnGP** only partially slows this drift and in fact attains the highest energy mismatch and variation, as its scalar constraint permits cancellation of errors across the domain. A clear hierarchy emerges: energy conservation is best for **pinnA**, intermediate for **pinnWP**, and worst for **pinnGP**.

4 Conclusion

Maxwell's equations have been successfully solved for decades using numerical methods. For a new methodology to gain acceptance in such a mature field, it must provide concrete advantages rather than merely offer an alternative. Although PINNs possess advantages such as a mesh-free structure and applicability to inverse problems, a systematic hybrid approach is required for them to approach classical solvers in accuracy and energy consistency. This study has presented such a methodology for TM_z mode wave propagation in a 2D PEC cavity and shown that hybrid PINNs can achieve competitive performance against FDTD on this canonical benchmark.

The model achieved an average NRMSE of 0.09% and relative L^2 error of 1.01% against the FDTD reference. In energy conservation, an average relative error of 0.024% and a maximum mismatch of 0.058% were attained, solely

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12

through physics-based residual losses without any labeled data. In the lossy scenario, the model tracked the expected dissipation curve with relative error below 0.2%, demonstrating generalizability across different physical regimes.

The main argument is that PINNs have requirements at both theoretical and implementation stages. At the theory level, Maxwell's equations guarantee energy conservation via Poynting's theorem, causality via their hyperbolic structure, and interface continuity via field continuity. However, these guarantees are not automatically preserved within the PINN framework; the network may converge to inconsistent solutions while minimizing PDE residuals. Therefore, incorporating these structures as explicit regularizers is essential, particularly for suppressing cumulative energy drift across sequential windows.

The proposed hybrid approach addresses this through three components: causality-aware weighting that prioritizes equation consistency at early times within each window, interface continuity loss that enforces smooth transitions across windows, and a local Poynting-based regularizer that suppresses cumulative energy drift. While the present framework is computationally more expensive to train than a standard FDTD run for the same forward problem, its value lies not in raw forward-simulation speed, but in providing a mesh-free continuous solution representation, rapid inference after training, and a flexible foundation for inverse problems. Taken together, the results show that these components contribute meaningfully and produce a strong combined effect for time-dependent electromagnetic wave propagation.

References

1. Yee, K.: Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. IEEE Transactions on Antennas and Propagation **14**(3), 302–307 (1966). https://doi.org/10.1109/TAP.1966.1138693
2. Clough, R.W.: Original formulation of the finite element method. Finite Elements in Analysis and Design **7**(2), 89–101 (1990). https://doi.org/10.1016/0168-874X(90)90001-U
3. Harrington, R.F.: Matrix methods for field problems. Proceedings of the IEEE **55**(2), 136–149 (1967). https://doi.org/10.1109/PROC.1967.5433
4. Raissi, M., Perdikaris, P., Karniadakis, G.E.: 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). https://doi.org/10.1016/j.jcp.2018.10.045
5. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Networks **2**(5), 359–366 (1989). https://doi.org/10.1016/0893-6080(89)90020-8
6. Hadamard, J.: Lectures on Cauchy's problem in linear partial differential equations. Courier Corporation (2014)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
8. Wang, S., Sankaran, S., Perdikaris, P.: Respecting causality is all you need for training physics-informed neural networks. arXiv preprint arXiv:2203.07404 (2022). https://doi.org/10.48550/arXiv.2203.07404

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12

9. Mattey, R., Ghosh, S.: A novel sequential method to train physics informed neural networks for Allen Cahn and Cahn Hilliard equations. Computer Methods in Applied Mechanics and Engineering **390**, 114474 (2022). https://doi.org/10.1016/j.cma.2021.114474
10. Penwarden, M., Jagtap, A.D., Zhe, S., Karniadakis, G.E., Kirby, R.M.: A unified scalable framework for causal sweeping strategies for physics-informed neural networks (PINNs) and their temporal decompositions. Journal of Computational Physics **493**, 112464 (2023). https://doi.org/10.1016/j.jcp.2023.112464
11. Jagtap, A.D., Karniadakis, G.E.: Extended physics-informed neural networks (XPINNs): A generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations. Communications in Computational Physics **28**(5), 2002–2041 (2020). https://doi.org/10.4208/cicp.OA-2020-0164
12. Bengio, Y., Louradour, J., Collobert, R., Weston, J.: Curriculum learning. In: ICML '09, pp. 41–48. ACM, New York (2009). https://doi.org/10.1145/1553374.1553380
13. Chen, Z., Lai, S.-K., Yang, Z.: AT-PINN: Advanced time-marching physics-informed neural network for structural vibration analysis. Thin-Walled Structures **196**, 111423 (2024). https://doi.org/10.1016/j.tws.2023.111423
14. Chang, C., Xin, Z., Zeng, T.: A conservative hybrid deep learning method for Maxwell–Ampère–Nernst–Planck equations. Journal of Computational Physics **501**, 112791 (2024). https://doi.org/10.1016/j.jcp.2024.112791
15. Pugin, A.: Simulation of electromagnetic waves propagation with Physics Informed Neural Networks. MSc thesis, Politecnico di Milano (2024)
16. LeCun, Y., Bottou, L., Orr, G.B., Müller, K.-R.: Efficient BackProp. In: Neural Networks: Tricks of the Trade, pp. 9–50. Springer (1998). https://doi.org/10.1007/3-540-49430-8_2
17. Tancik, M., Srinivasan, P.P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J.T., Ng, R.: Fourier features let networks learn high frequency functions in low dimensional domains. In: Advances in Neural Information Processing Systems **33**, pp. 7537–7547. Curran Associates, Inc. (2020). https://doi.org/10.48550/arXiv.2006.10739
18. Mandl, L., Mielke, A., Seyedpour, S.M., Ricken, T.: Affine transformations accelerate the training of physics-informed neural networks of a one-dimensional consolidation problem. Scientific Reports **13**, 15566 (2023). https://doi.org/10.1038/s41598-023-42141-x
19. Baldan, M., Di Barba, P., Lowther, D.A.: Physics-Informed Neural Networks for Inverse Electromagnetic Problems. IEEE Transactions on Magnetics 59(5), 1–5 (2023). https://doi.org/10.1109/TMAG.2023.3247023
20. Chen, Y., Lu, L., Karniadakis, G.E., Dal Negro, L.: Physics-informed neural networks for inverse problems in nano-optics and metamaterials. Optics Express 28(8), 11618–11633 (2020). https://doi.org/10.1364/OE.384875
21. Nohra, M., Dufour, S.: Approximating electromagnetic fields in discontinuous media using a single physics-informed neural network. arXiv preprint arXiv:2407.20833 (2024)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_12