

Lawrence Berkeley National Laboratory

LBL Publications

Title

Predicting Wave Reflection and Transmission in Heterogeneous Media via Fourier Operator-Based Transformer Modeling

Permalink

<https://escholarship.org/uc/item/4856c0qj>

Journal

International Conference on Artificial Intelligence Computer Data Sciences and Applications
Acdsa 2026, 00

Authors

Bai, Zhe
Johansen, Hans

Publication Date

2026-02-07

DOI

10.1109/acdsa67686.2026.11467920

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nd/4.0/>

Predicting Wave Reflection and Transmission in Heterogeneous Media via Fourier Operator-Based Transformer Modeling

Zhe Bai
AMCR Division
Lawrence Berkeley National Lab
Berkeley, USA
zhebai@lbl.gov

Hans Johansen
AMCR Division
Lawrence Berkeley National Lab
Berkeley, USA
hjohansen@lbl.gov

Abstract—We develop a machine learning (ML) surrogate model to approximate solutions to Maxwell’s equations in one dimension, focusing on scenarios involving a material interface that reflects and transmits electro-magnetic waves. Derived from high-fidelity Finite Volume (FV) simulations, our training data includes variations of the initial conditions, as well as variations in one material’s speed of light, allowing for the model to learn a range of wave-material interaction behaviors. The ML model autoregressively learns both the physical and frequency embeddings in a vision transformer-based framework. By incorporating Fourier transforms in the latent space, the wave number spectra of the solutions aligns closely with the simulation data. Prediction errors exhibit an approximately linear growth over time with a sharp increase at the material interface. Test results show that the ML solution has adequate relative errors below 10% in over 75 time step rollouts, despite the presence of the discontinuity and unknown material properties.

Keywords—*Electro-magnetics, wave propagation, vision trans-formers, Fourier transforms, predictive modeling*

I. INTRODUCTION

A. Overview

Electro-magnetic waves are ubiquitous in a wide range of scientific and engineering applications, including communications, photonics, material characterization and medical imaging. Accurately modeling how these waves propagate, reflect, and transmit across different media is critical for understanding and optimizing system performance. Maxwell’s equations [1] are often the starting point when simulating electro-magnetic (EM) devices, where combinations of complex geometries and material changes are designed to produce a desired effect. When simulating whole devices with nonlinear effects such as frequency-dependent materials or those with hysteresis, an inexpensive choice is the Finite Difference Time Domain method (FDTD) [2], [3]. However, FDTD can have several drawbacks, including: dispersion errors, errors introduced near boundaries and material interfaces, and very long serial time integration, among others [2], [5]. These errors and bottlenecks then become issues in the “inner loop” of EM design optimization problems, for example, wave guide multiplexers [4], [6]

and nonlinear antennas [7]. This work investigates a machine learning (ML) surrogate approach designed to predict the solution across *all times* from only a few initial times steps. The ML training data only includes the electric field, $E(x, t)$; the algorithm must infer the dynamics of $B(x, t)$ in Maxwell’s equations from just the space- and time-evolution of one variable. To evaluate the ML approach, we use training data from a 1D prototype of a transverse EM wave crossing a material boundary, and evaluate the accuracy of the ML solution. Because 1D Maxwell’s equations allows exact analytical solutions, we can assess the accuracy and spectral content at all times.

B. Related methods for Maxwell’s equations

Maxwell’s equations can be simulated using FDTD methods, and also finite volume (FV) [11] and finite element (FE) [10] approaches in the time domain. Spectral methods, including Fourier transforms-based approaches, can also be employed; however, nonlinear effects and material boundaries often limit their applicability except in special contexts. Similarly, reducing Maxwell’s equations to the wave equation can only be done in simplified circumstances. FDTD methods typically use lower-order discretizations in space (e.g., second-order differences on Yee-grids), and simple lower-order time integrators, such as the Leap Frog method [2]. For FD and FV methods, material boundaries are often constrained to be on regular grid lines, and implemented with lower-order or one-dimensional approximation [7]. FE approaches can represent sharp material changes exactly by having a mesh aligned with material boundaries, but explicit time step constraints (CFL condition) can be challenging with FE mesh refinement and higher-order discretizations. As a result, FE methods typically lead to solvers in the frequency domain or use implicit time stepping [10], which can be expensive or scale poorly. Recent improvements in FV methods allow for sharp interfaces and higher-order accurate methods that do not have severe conditioning issues, which can otherwise impose restrictive

time step constraints when material boundaries are not aligned with the computational grid [12], [13].

C. Related ML methods for PDE solutions

Data-driven approaches provide a new paradigm to solve PDEs using simulation or experimental measurements that are amenable to leveraging physical constraints, such as physics-informed neural networks (PINNs) [14] that embeds governing laws into neural networks. However, PINNs may struggle in stiff loss landscapes with sharp gradients or discontinuities (e.g., material interfaces), and is limited to the geometry and boundary conditions that may not generalize to others. Neural operator methods [15]–[17] build upon the universal approximation theorem, and effectively learn a mapping from the input functions to output functions, allowing them to learn PDE solutions with improved generalization. Recently, transformer-based models that leverage self-attention mechanism have shown more promise in modeling long-dependencies between input sequence elements and support parallel processing of sequence as compared to conventional types of recurrent networks [18]–[20].

In this work, we develop a Fourier operator-based surrogate model to emulate the evolution of electric field derived from Maxwell’s equations using a dual-path neural architecture that incorporates both embeddings with self-attention from frequency and physical space, which enables the model to capture inherent dynamics for long-range predictions of wave propagation behaviors.

II. METHOD

A. Maxwell’s equations and discretization

Maxwell’s equations govern the time evolution of electric field ($\mathbf{E}$) and magnetic flux ($\mathbf{B}$), traveling at the speed of light in each material. The simplest form normalizes against the speed of light in a vacuum, so that $c_m \leq 1$ is a constant relative speed of light in each material m :

$$\partial_t \begin{bmatrix} \mathbf{E} \\ \mathbf{B} \end{bmatrix} = \begin{bmatrix} c_m^2 \nabla \times \mathbf{B} & 0 \\ 0 & -\nabla \times \mathbf{E} \end{bmatrix}. \quad (1)$$

In the case of fields *transverse* to the x dimension this can be simplified to just one component, $\mathbf{E} = E\hat{y}$ and $\mathbf{B} = B\hat{z}$:

$$\partial_t \begin{bmatrix} E \\ B \end{bmatrix} = - \begin{bmatrix} 0 & c_m^2 \\ 1 & 0 \end{bmatrix} \partial_x \begin{bmatrix} E \\ B \end{bmatrix} \quad (2)$$

$$= -\partial_x \begin{bmatrix} c_m^2 B \\ E \end{bmatrix} \quad (3)$$

where this has been written in *characteristic* (2) and *finite volume* (3) forms, respectively. Using the *method of characteristics* [9] in 1D, the solution evolves along characteristics, $x_0 = x \pm c_m t$, with $E(x_0, t = 0)$ and $B(x_0, t = 0)$ contributing to the exact solution at all times. This allows us to trace back from the current time, through or reflecting off the material interface, to a value from the initial condition.

In finite volume form, we can use finite volume approximations that respect the transverse wave interface jump conditions, $[E] = 0$ and $[B] = 0$. As is typical in FV methods, the

E and B fields are transformed into left- and right-moving characteristic variables; we use a fifth-order upwind finite volume scheme that enforces jumps exactly at the interface. A Riemann problem is solved on each finite volume cell interface, and at the material boundary the transmitted and reflected waves are calculated. A flux-based update to E and B is used to evolve (3) with an RK4 time integrator. More details on this FVTD method can be found in [8] where it is applied to nonlinear compressible gas dynamics.

B. Autoregressive modeling

To model the long-range evolution of the physical process, we consider an autoregressive formulation as illustrated in Fig. 1. We define a vector $\tilde{s}_t$ that represents the previous snapshots of the variable at the timestamps from $t - m + 1$ to t , in a total of m ($m \ll T$) snapshots of $\tilde{s}_t = [s_t, s_{t-1}, \dots, s_{t-m+1}]$. The one-step forward neural operator is defined as

$$s_{t+1} = \mathcal{F}_{\theta_d}(\tilde{s}_t), \quad (4)$$

where the operator $\mathcal{F}$ is composed of a dual-path Fourier Transformer that learns both the embeddings from the original physical space and the Fourier transformed space as in shown in Fig. 2.

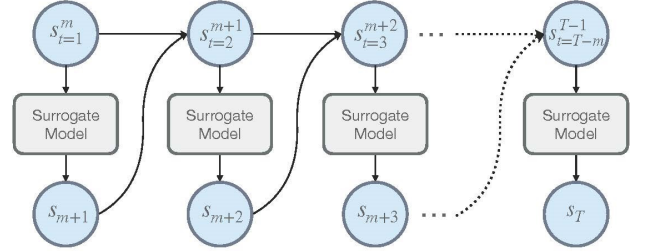


Fig. 1. Autoregressive modeling for forecasting long-range dynamics with a series of time snapshots of $\tilde{s}_t = [s_t, s_{t-1}, \dots, s_{t-m+1}]$ at each step rolling out for prediction of s_{t+1} .

C. Fourier-based vision transformers

The model learns spatial scales through a hierarchy of Fourier transformer block that each composition block is designed to handle a specific scale. At each level, two transformation paths are performed: one is in the spatio-temporal domain, which operates self-attention on spatio-temporal embeddings; the other path is in the frequency domain, which operates on its transformed Fourier components.

- *Frequency embeddings*: the model first applies a Fourier transform to the input tokens, retaining the low-frequency components. These components are then passed through a Fourier Transformed module to generate the frequency embeddings f_t . These embeddings are subsequently processed by a series of transformer layers that perform feature mixing and nonlinear transformations in the frequency domain. Finally, an inverse Fourier transform is applied, followed by a linear projection that maps the frequency-domain representations back to the spatial domain.

- *Spatio-temporal embeddings*: the same set of tokens first pass through a spatio-temporal embedder, after which the spatio-temporal embeddings e_t are processed by multiple standard transformer layers for mixing correlations and nonlinear transformation. Finally a separate linear projection is applied to get predictions for each patch of s_{t+1} .

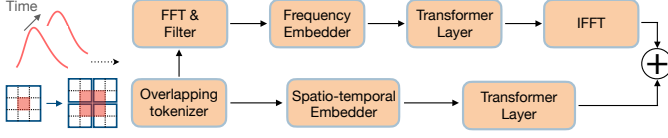


Fig. 2. Dual-path architecture: Fourier transformer. Overlapping tokenizer is applied prior to the transformations into embeddings of f_t and e_t . After the two embeddings are merged, an overlapping detokenizer is employed to reconstruct the spatial field, mapping the latent representation back to the original physical domain.

In addition, each Fourier transformer block first tokenizes the discretized functions $\tilde{s}_t$ along the spatial dimensions with an overlapping tokenizer (OLT), which allows adjacent patches to share boundaries through overlapping regions and enhance the spatial continuity. For instance, as shown in Fig. 2, a 3×3 input generates four 2×2 patches with a 1×1 overlap, where the overlapping areas (gray) are shared between patches. During detokenization, overlapping regions are reconstructed by averaging the corresponding values from neighboring patches in overlapping detokenizer (OLDT). This step mitigates discontinuity issues at the patch boundaries, improving the smoothness of continuous target functions incorporating shared boundaries into patch embeddings for sharp transitions.

III. RESULTS

A. Maxwell's Equations Test Setup

The domain $x \in [0, 1]$ is split into two material regions, with speed of light $c_1 = 1$ for $x \leq x_j = 0.5$, and c_2 for the right region, $x \geq x_j$. The initial condition ϕ is entirely contained in the left half domain, and is a right-moving wave packet with $E_0 = \phi$, $B_0 = \phi/c_1$.

Test Case 1. We use two uniform random parameters $r_1, r_2 \in [0, 1]$ to generate 200 test cases varying only the initial conditions:

$$\begin{aligned} \phi(x) &= \begin{cases} g(x), & \text{if } r_1 < 0.15 \\ g(x) \sin(2\pi k(x - x_s)), & \text{otherwise.} \end{cases} \quad (5) \\ g(x) &= \exp(-(x - x_g)^2/\sigma^2), \text{ with } x_g = 0.25, \\ \sigma &= 0.25/(5 - 2r_1) \in [1/20, 1/12], \\ x_s &= 0.35 + 0.2r_1 \in [0.35, 0.55], \\ k &= 3 + 3r_2 \in [3, 6]. \end{aligned}$$

For this test case, we fix $c_2 = 1/3$ in the second domain. To generate accurate training data, we use the FV method described in §II-A, with $N = 256$ grid points, resulting in a pointwise relative error $< 0.1\%$ for all tests. We run for

200 time steps, and train the model across the time domain $[0, 200]$.

Test Case 2. Here we add a third uniform random variable, r_3 , that specifies the speed of light in the second medium:

$$c_2 = 1/(1 + 2r_3) \in [1/3, 1] \quad (6)$$

so that $c_2 \leq c_1 = 1$. This is an important variation to test if the ML approach can learn the speed of light directly from training data, and then *infer* it from only the first few time steps. Because c_2 has no influence before the wave packet has crossed x_j , we first run for 100 time steps, then train the model on time steps $[100, 200]$, so that the effect of c_2 on the PDE solution can be observed as the wave packet crosses the material interface, x_j . In both cases, the curated data is partitioned into 160 samples for training, 20 samples for validation and 20 for test.

B. Wave propagation prediction

We investigate $E(x, t)$ over the entire time domain in 2D that is lifted from the 1D domain, in which values of one dimension are constant as exemplified in Fig. 3. Comparisons with the simulated results indicate that the ML model accurately captures the dynamics, with relative errors remaining small at earlier times ($t = 50$) but increasing at later times ($t = 100$), reflecting the temporal accumulation of approximation error. The ML model hyperparameters, including the network depth, hidden dimension, patch size are searched for optimized validation performance, as detailed in Table I. Experiment results further indicate that incorporating the overlapping tokenizer alleviates sharp discontinuities at wave peaks introduced by parchification, leading to a notable improvement with relative error reductions of 20%.

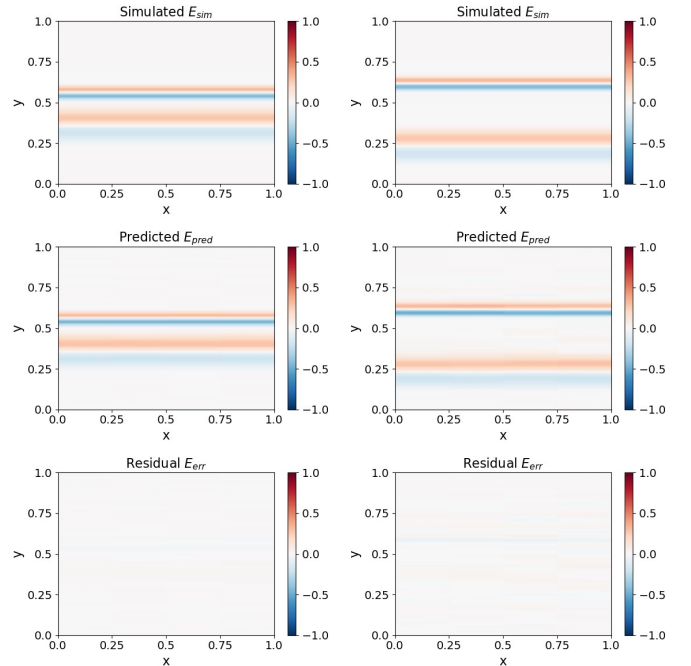


Fig. 3. Comparison between the simulated (top), ML predicted (middle), and residual (bottom) results for the electric field in Case 2, at $t = 50$ (left column), and $t = 100$ (right column).

TABLE I.
HYPERPARAMETER SEARCH RANGE

Hidden dimension	[256, 512]
Patch size	[33, 66, 132]
Network depth	[6, 10, 12]
Overlap tokenizer	[0, 1]

C. Error analysis - Test Case 1

For Test Case 1, we fix the speed of light in the right half of the domain, $c_2 = 1/3$, and evaluate 200 timesteps of the solution. The initial wave packet begins in the left half of the domain, and crosses the material interface around step $t = 100$. By that point, the wave packet has started to reflect (invert) and travel left, as well as transmit into the right half of the domain, where it narrows substantially and continues traveling but at the slower speed c_2 . By $t = 200$ both the reflected and transmitted waves have mostly moved away towards the boundaries.

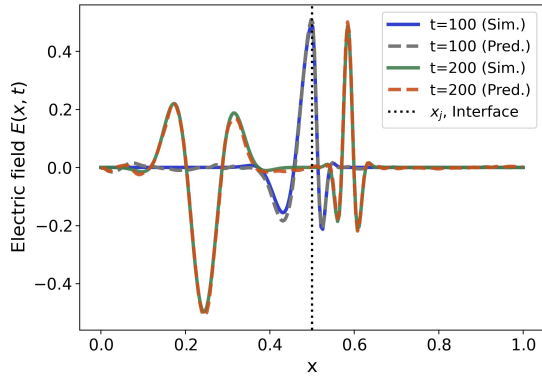


Fig. 4. Simulated and predicted solutions for Case 1, at two time steps $t = 100$ and $t = 200$. The dotted line at $x_j = 0.5$ is the material interface where the speed of light changes.

Fig. 4 shows the simulation training data and the predicted ML solution at time steps 100 and 200. Note that the predicted wave packets are of nearly the same profile and location, with only a small amount of visible amplitude error. The final solution appears to match the transmitted and reflected waves as well. There is also a small amount of higher wave-number error in the prediction throughout the domain.

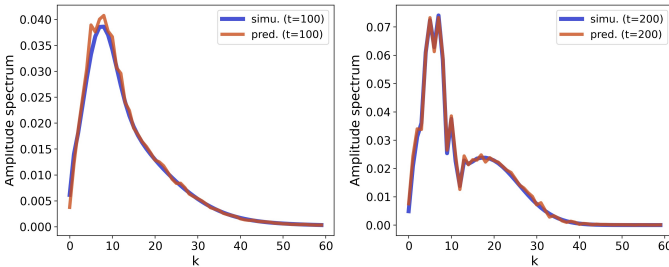


Fig. 5. Fourier domain content of the predicted solution for Case 1, at step $t = 100$ (left), when the wave packet is crossing the material boundary, and $t = 200$ (right) when it has moved away after reflecting/transmitting.

Fig. 5 shows the Fourier spectrum (wave mode content), which supports these conclusions as well. At timestep 100, the predicted solution has largest relative errors at the peak of the Fourier spectrum, but is overall consistent with the FVTD spectrum. At timestep 200, the spectrum matches better at the peak, even though it contains high wave number features from the narrower transmitted wave. However, there is some noise throughout the spectrum in both cases, indicating that spatial errors are distributed relatively uniformly, everywhere.

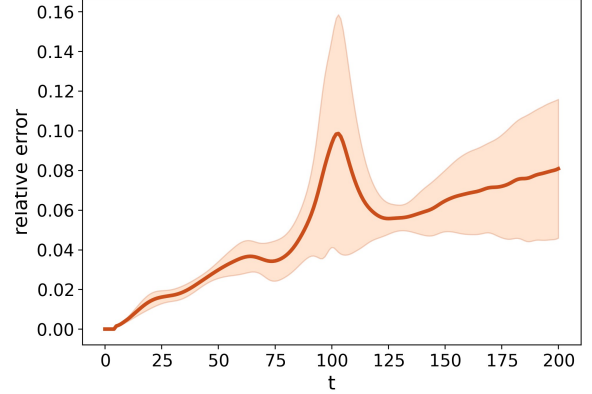


Fig. 6. Case 1 predicted solution relative error over time steps $t = [0, 200]$, with the mean value (dark red line) and standard deviation (shaded region) across all test cases. Time step $t = 100$ is approximately when the wave pulse is crossing the material boundary.

Fig. 6 shows the relative error of the predicted solution over the whole time interval, $t = [0, 200]$. The mean appears to grow approximately linearly in time, except around step $t = 100$, where the error approximately doubles, and there is a much larger error distribution across the test data, reaching a relative error of around $\sim 10\% \pm 5\%$. This may not be surprising, as the wave packet develops a discontinuity as it moves through the material interface, which may affect its Fourier representation and require more modes than have been retained. This is also the point where the reflected and transmitted wave packets are superimposed across the discontinuity, making it more difficult to separate the smooth spatial patterns that are clearer at earlier and later times.

For times after $t \approx 125$, the mean error reverts to the previous trend, but continues to grow, albeit with a much smaller variation across the test data. In general, we see that the variation in error continues to grow more substantially over time, so that a significant portion of the tests are again around $\sim 10\%$ relative error by the end of the time range. Referring to Figs. 4-5 for comparison, the E field amplitudes and spectral content may still match relatively well across the time domain.

D. Error analysis - Test Case 2

For this test, the speed of light c_2 in the right half of the domain varies randomly across individual test cases. This affects both the amplitude of the transmitted and reflected waves, as well as the speed that the transmitted wave propagates. For the algorithm to reliably infer c_2 from the time-series data, it must observe the dynamics over several time steps. Accordingly, we

start the training data at time step $t = 100$, which corresponds to the onset of observable transmitted and reflected waves. The time variable is then reindexed starting at this point $t = 0$, and continued for an additional 100 time steps subsequently.

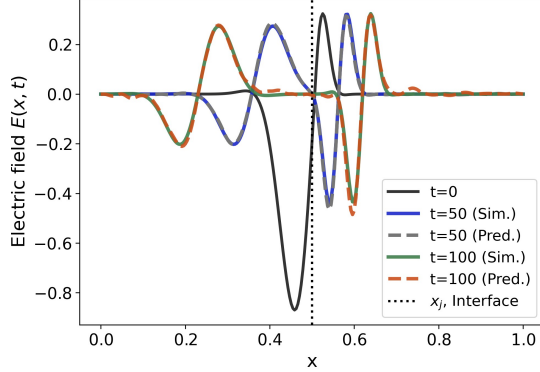


Fig. 7. Simulated and predicted solutions for Case 2, with the initial condition and two time steps $t = 50$ and $t = 100$. The material to the right of the interface (dotted line) has a random speed of light that must be learned from training data.

Referring to Fig. 3, we see a color plot of the 2D solution at the initial condition (time-shifted, $t = 0$) and later steps $t = 50$ and $t = 100$. The wave packet location and amplitude are well-predicted, and the error at later times is larger, as before. As illustrated in Fig. 7, a similar trend is observed: the ML model accurately reproduces the phase evolution of the wave packet, but the peak amplitude is less precise. Note that the initial condition already contains some structure in the right half of the domain, and because the first few time steps are used for inference, the model is able to determine an approximate c_2 . However, the spectral content is not as clearly discernible from the initial condition due to the superimposition of incident and reflected waves in the left domain. As a result, the model must infer the propagation during training, yet it predicts the transmitted and reflected waves relatively well at later times, for example at $t = 100$. The Fourier content, seen in Fig. 8, matches well in both wave number and amplitude at both times, as previously observed in Case 1.

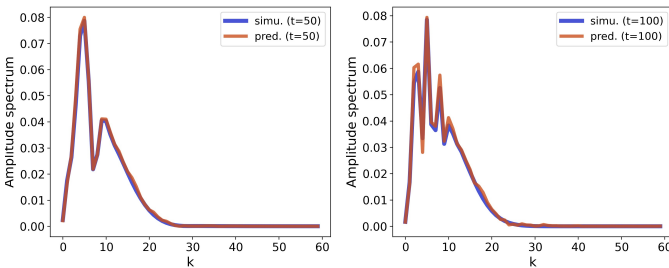


Fig. 8. Fourier domain content of the predicted solution for Case 2, at step $t = 50$ (left) and $t = 100$ (right), as it is moving away from the interface after reflecting/transmitting.

A more substantial difference from Case 1 is seen in the time evolution of the relative error, in Fig. 9. The error and its variation in time are more consistent, but grow at approximately $2\times$ the rate of Case 1, effectively $4\times$ higher

with half the number of steps. We suspect it is more challenging to start the training data when the wave pulse is discontinuous crossing the interface, especially with unknown material (c_2 parameterized) in the second medium. So although the randomly varying speeds of light are immediately evident to the model, it does exhibit greater errors overall. Still, the short-term errors are small and the error growth is more steady than in Case 1.

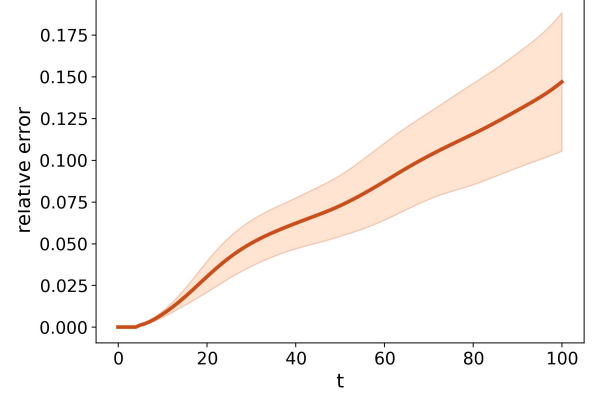


Fig. 9. Case 2 predicted solution relative error for $t \in [0, 100]$, mean value (dark red line) and standard deviation (shaded region) across all test data. The error grows approximately linearly in time.

E. Ablation study

1) *The Frequency Path*: To evaluate the effectiveness of the frequency path embeddings, we conduct an ablation study removing the frequency path in each hierarchical layer. The experimental results are summarized in Table III, with all experiments trained for the same number of epochs. The L_2 relative error drops to 0.0506 from 0.3945 in Case 1, and for Case 2 with a half test time window, the relative error drops from 0.1819 to 0.0827.

TABLE II.
EFFECT OF THE FREQUENCY PATH.

Problem	Model(s)	L_2 error
Case 1	$\hat{f}_t + e_t$	0.0506
	e_t	0.3945
Case 2	$\hat{f}_t + e_t$	0.0827
	e_t	0.1819

TABLE III.
ABLATION RESULTS ON THE FOURIER MODE TRUNCATION.

Problem	Setting(s)	L_2 error
Case 1	$r = 4$	0.0825
	$r = 8$	0.0506
	$r = 16$	0.4795
Case 2	$r = 4$	0.1191
	$r = 8$	0.0827
	$r = 16$	0.3099

2) *The Fourier modes*: We also investigate the number of truncated Fourier modes in the Frequency path $\hat{f}_t$. It is observed that $r = 8$ yields the best overall performance. When fewer modes are retained, the model fails to adequately capture the dominant peak frequencies (see Figs. 5 and 8).

In contrast, retaining a larger number of frequency modes does not further improve accuracy, as the expanded frequency range exhibits reduced effectiveness in resolving the fine-scale structures compared to the spatio-temporal path.

IV. CONCLUSION

We have presented a Fourier operator-based Transformer method as a surrogate model for predicting the time series solution to 1D Maxwell's equations with a material boundary. With the results of two test cases, it is shown that the ML approach is able to predict solution to a relative error of $\sim 10\%$ across most of the time domain. The errors are much smaller early in the simulation, and grow approximately linearly through later time steps, with a growth in variation of errors across the test data as well. More importantly, the Fourier domain representation of the solution is also accurate, representing the dominant wave numbers well and maintaining the solution's form throughout.

The ablation study results demonstrate the effectiveness of the frequency path in capturing the dominant spectral components of the system. In particular, the optimal number of retained modes in the frequency path corresponds closely to the peak frequencies in the spectrum. For the first test case, where the wave packet crosses the material boundary half way through the simulation, the ML prediction error increases briefly, then decreases again in later time steps. We suspect that this is due to the material discontinuity introducing high wave-number content that is not in the truncated Fourier representation. The second test case demonstrates that, if the model is trained on data that includes solution values in both material domains, it greatly reduces this error at the interface. The model is then able to recover long-term accuracy, even when the material property is random distributed within the test data, implying that it effectively infers the hidden variable or characteristics of the media.

In future work, building on our understanding of the limitations and sources of error in the present approach, we will extend it to higher dimensions and nonlinear materials, as seen in waveguide multiplexers [4]. An additional consideration is the assessment of computational gains offered by the ML model over traditional low-accuracy finite-volume models. From a scaling standpoint, the ability to predict the entire time evolution at once would vastly improve parallelism over traditional sequential time stepping; this could open the door for doing parallel in time algorithms [21], as well. Finally, we expect that these time series predictions may lead to better initial guesses in solver preconditioners and inverse design problems of these complex systems.

ACKNOWLEDGMENT

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research under contract number DE-AC02-05CH11231. Z.B. acknowledges support from the U.S. Department of Energy, Office of Science, SciDAC/Advanced Scientific Computing Research under Award Number DE-AC02-

05CH11231. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy User Facility.

REFERENCES

- [1] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [2] J. B. Schneider, "Understanding the finite-difference time-domain method," School of electrical engineering and computer science, Washington State University, 2025. Downloaded from <http://eeecs.wsu.edu/~schneidj/ufdtd/ufdtd.jpeg>.
- [3] Y. Wenhua, "Electromagnetic simulation techniques based on the FDTD method," John Wiley & Sons, 2009. ISBN: 978-0-470-50203-7.
- [4] J. Lu, and J. Vučković, "Nanophotonic computational design," *Opt. Express*, 21, 2013. DOI: 10.1364/OE.21.013351.
- [5] W. Jin, T. Hughes, Z. Yu, and S. Fan, "FDTD 101," Downloaded Oct 2025. Online course at <https://www.flexcompute.com/fdtd101>.
- [6] Y. Mahlau, et al, "A flexible framework for large-scale FDTD simulations: open-source inverse design for 3D nanostructures," *Photonics and Phononic Properties of Engineered Nanostructures XV*, vol. 13377, pp. 40-52, Mar 2025. DOI: 10.1117/12.3052639.
- [7] A. Oskooi, D. Roundy, M. Ibanescu, P. Bermel, J.D. Joannopoulos, and S.G. Johnson, "MEEP: A flexible free-software package for electromagnetic simulations by the FDTD method," *Computer Physics Communications*, vol. 181, pp. 687-702, 2010. DOI: 10.1016/j.cpc.2009.11.008.
- [8] P. McCorquodale, M.R. Dorr, J.A.F. Hittinger, and P. Colella. "High-order finite-volume methods for hyperbolic conservation laws on mapped multiblock grids," *Journal of Computational Physics*, 288, 2015. DOI: 10.1016/j.jcp.2015.01.006.
- [9] L. C. Evans, *Partial Differential Equations*. Providence: American Mathematical Society, 1998. ISBN 0-8218-0772-2.
- [10] S. Couture, X. Wang, A. Goncharov, and V. Lomakin, "A coupled micromagnetic-Maxwell equations solver based on the finite element method," *Journal of Magnetism and Magnetic Materials*, vol. 493, 2020. DOI: 10.1016/j.jmmm.2019.165672.
- [11] K. S. Yee, and J. S. Chen, "The finite-difference time-domain (FDTD) and the finite-volume time-domain (FVTD) methods in solving Maxwell's equations," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 3, 1997. DOI: 10.1109/8.558651.
- [12] N. Overton-Katz, X. Gao, S. Guzik, O. Antepara, D. T. Graves, and H. Johansen, "A Fourth-Order Embedded Boundary Finite Volume Method for the Unsteady Stokes Equations with Complex Geometries," *SIAM Journal on Scientific Computing*, 45, 5. 2023. DOI: 10.1137/22M1532019.
- [13] H. Johansen, N. Overton-Katz, R. Chowdhury, and S. Guzik, "Accurate Higher-order Embedded Boundary Methods without Small Cell Issues," *APS Division of Fluid Dynamics*, 2024. Bibcode: 2024APS..DFDA13001J.
- [14] M. Raissi, P. Perdikaris, and G. Karniadakis. "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, pp. 686–707, 2019. DOI: 10.1016/j.jcp.2018.10.045.
- [15] Z. Li, D. Z. Huang, B. Liu, A. Anandkumar, "Fourier neural operator with learned deformations for pdes on general geometries," *Journal of Machine Learning Research*, vol. 24, pp. 1–26, 2023.
- [16] L. Lu, P. Jin, G. Pang, Z. Zhang, G. Karniadakis, "Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators," *Nature machine intelligence*, vol. 3, pp. 218–229, 2021.
- [17] Z. Li, K. Meidani, A.B. Farimani, "Transformer for partial differential equations operator learning," *arXiv preprint, arXiv:2205.13671*, 2022.
- [18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, E. Kaiser, I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [19] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint, arXiv:2010.11929*, 2020.
- [20] S. Khan, M. Naseer, M. Hayat, S.W. Zamir, F.S. Khan, M. Shah, "Transformers in vision: A survey," *ACM computing surveys (CSUR)*, vol. 54, no. 10s, pp.1-41, 2022.
- [21] A. Q. Ibrahim, S. Götschel, D. Ruprecht, "Space-Time Parallel Scaling of Parareal with a Physics-Informed Fourier Neural Operator Coarse Propagator Applied to the Black-Scholes Equation," *Proceedings of PASC '25, ACM*, pp. 1–11, 2025. DOI: 10.1145/3732775.3733574.