A FOUNDATION MODEL FOR TIME-HARMONIC ELECTROMAGNETIC BEHAVIOR

Anonymous authors
Paper under double-blind review

ABSTRACT

3D electromagnetic (EM) systems underpin semiconductor packaging, conformal antennas, and consumer devices such as phones and wearables. Full-wave simulations are necessary to model these systems but remain costly across large design and frequency spaces, while existing surrogates typically specialize in a narrow design family and predict either fields or aggregate responses. We introduce the first EM foundation model for complex 3D systems, predicting both electromagnetic fields and scattering parameters for unseen devices. Our contribution is threefold. First, we generate a corpus of approximately 250k simulations spanning 25 template families of 3D printed-circuit-board and package structures. Second, we train a 340M-parameter model that predicts complex-valued electric and magnetic fields at arbitrary locations and S-parameters across frequencies and port pairs. Finally, we introduce `EMVal`, the first public evaluation suite for EM foundation models covering both data modalities. We find that joint field and S-parameter supervision preserves in-distribution S-parameter accuracy while reducing out-of-distribution error relative to S-parameter-only training, suggesting that field supervision encourages representations that generalize more effectively across geometry families. Also, ablations identify impedance-aware target scaling, a robust field objective, and relative spatial encoding as critical to accurate complex-field prediction. Ultimately, our model's full EM vector-field predictions provide a rich physical description of unseen devices that enables design-critical downstream calculations and visual debugging, opening a path toward general-purpose learned models of EM systems.

1 INTRODUCTION

Modern electronic systems often rely on structures such as interconnects, filters, and antennas to control how electromagnetic (EM) energy propagates, couples, and radiates. Designing these structures typically requires repeated full-wave simulations using numerical techniques such as the finite-element method (FEM). Given a three-dimensional geometry, material distribution, port configuration, and excitation frequency, a commercial FEM solver computes the complex-valued fields $\mathbf{E}$ and $\mathbf{H}$ and the matrix of complex-valued scattering parameters (S-parameters, which describe how energy is transmitted and reflected between ports). These solvers are accurate, but costly across design iterations and frequency sweeps. Learned surrogates amortize this cost, but existing models generally specialize to a narrow design family or output modality.

Progress is limited by the absence of a large, geometrically diverse corpus of field and S-parameter data, a model that predicts both outputs across three-dimensional structures, and a common evaluation protocol.

In this paper, we make the following contributions:

- **A procedural dataset-generation pipeline.** We describe a flexible pipeline that creates geometrically diverse 3D design and simulation data, including both traditional and unconventional design families, along with their simulated near-field and S-parameter values.
- **A foundation model for time-harmonic electromagnetic behavior.** For an input with arbitrary 3D geometric structure, dielectric and metallic material properties, and a collection of ports, our model predicts (1) the complex 3D electric and magnetic field values at an arbitrary query location

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

and excitation frequency, and (2) the complex S-parameter values given an excited port, receiving port, and excitation frequency.

- **An evaluation suite, `EMVal`.** We introduce a collection of metrics and datasets for standardized evaluation of time-harmonic EM foundation models.

2 Related work

Learned EM models divide broadly into response surrogates and field solvers. RF and microwave response surrogates range from component-specific low-dimensional parameterizations (Feng et al., 2017; Torun et al., 2020) to image- and frequency-conditioned emulators developed for inverse design over broader but still two-dimensional, fabrication-constrained design spaces (Karahan et al., 2024; Bi et al., 2025). Full-field prediction includes frequency-domain Maxwell solvers for planar photonic devices (Gu et al., 2022; Chen et al., 2022) and, more recently, 3D scattering models for canonical metallic targets, periodic metasurfaces, and surface meshes (Jiang et al., 2026a; Furat et al., 2025; Jiang et al., 2026b). Learned subdomain operators have also been used as preconditioners for conventional two-dimensional transverse-electric solvers (Mao & Fan, 2026). We are not aware of prior work that predicts both complex vector $\mathbf{E}$ and $\mathbf{H}$ over diverse multilayer PCB and package geometries, conditioned jointly on arbitrary position, frequency, and excited port.

In fluid dynamics, UPT provides a common query interface for meshes, grids, and particles, while AB-UPT scales it to automotive CFD meshes with up to 150 million cells (Alkin et al., 2024; 2025). Poseidon instead pretrains across diverse operators and adapts to 15 downstream PDE tasks, including physics absent from pretraining (Herde et al., 2024). Collectively, these works move beyond surrogates fitted to a single parameterized design family through broad pretraining, adaptation across tasks, or discretization-independent interfaces. We find no analogous pretrained model for time-harmonic electromagnetics.

Architecturally, our system combines the occupied-volume patch encoder of Volt (Yilmaz et al., 2026) with the encode–compress–query interface of UPT (Alkin et al., 2024). Unlike Volt, which applies 3D rotary position embeddings to discrete patch indices, we compute rotations from continuous physical coordinates, making attention depend on metric displacement. Unlike UPT, we do not evolve a latent state: time-harmonic response is stationary at a specified frequency, so frequency and port excitation enter as query conditions and the encoded geometry is reused directly across queries.

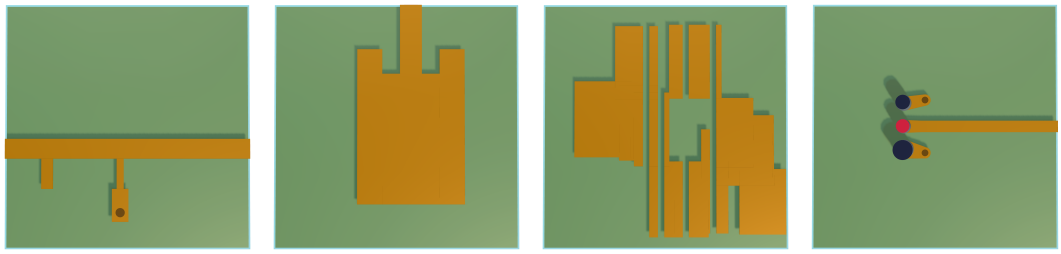
3 Methods

3.1 Dataset construction

We construct a training dataset of $\sim$250k time-harmonic EM simulations of excited three-dimensional printed circuit board (PCB) structures. Each design in the corpus is generated by first selecting one of 25 design families and then sampling values for the parameters of that family's template. For in-distribution validation, we randomly hold out $\sim$6k designs from the training set. For out-of-distribution evaluation, we generate a dataset of $\sim$2k designs from two additional design families that are fully excluded from the training set.

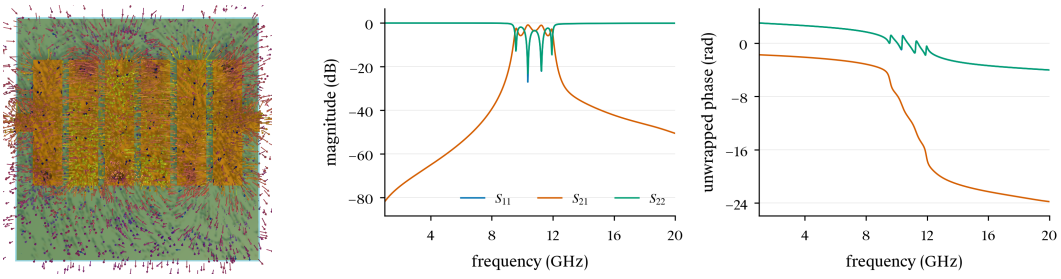
The PCB design families include a mix of common devices in RF and microwave engineering that exhibit useful electromagnetic behavior (e.g. transmission-line structures, filters, resonators, matching networks, antennas, and ball grid array (BGA) packages), as well as unconventional randomly-generated designs intended to increase the geometric diversity of the dataset. Each design family includes configurable properties like trace widths and lengths, gap and coupling distances, resonator counts, stub positions, turn counts and pitches, feed placements, via positions, etc. Figure 1 shows a collection of representative designs.

All boards have an 8×8 mm square footprint and comprise alternating layers of copper and dielectric. Planar boards are modeled as a patterned signal layer over a ground plane, with optional cylindrical vias connecting metal layers. BGAs have additional 3D complexity from rounded solder balls mounted to the board's upper surface. Boards contain one or two ports from which the system can be excited or measured, each of which is terminated in a standard $50\,\Omega$ load. Dielectrics are randomly sampled from materials including FR4 and Rogers RO4003C.



(a) (b) (c) (d)

Figure 1: **Example designs.** Four boards drawn from distinct design families: (a) a π matching network, (b) a patch antenna, (c) a free-form rectilinear layout, and (d) a BGA package with solder balls and an escape trace. Copper is shown in gold, ground plane in green, dielectric is transparent, vias are brown, port signal pins are red, and port ground pins are black.



(a) (b)

Figure 2: **Example simulation results.** A combline bandpass filter with via-grounded resonators. (a) Simulated complex $\mathbf{E}$ vectors at the 10k probe points, viewed from above at 11.26 GHz with the first port excited, colored on the field's 5th–95th percentile magnitude. (b) The complex S-parameters across the 1–20 GHz frequency range, as dB-scaled magnitude and unwrapped phase. On this symmetric filter S_{11} lies almost exactly under S_{22}.

3.2 Simulation

We simulate each board using *Palace*, an open-source FEM solver (AWS Center for Quantum Computing, 2025), with an adaptive sweep capped at 20 full solves per board within the 1 to 20 GHz frequency range. From those solves, we obtain interpolated solves at 101 uniformly spaced frequencies from 1 to 20 GHz. At each frequency, we store the complex S-parameter matrix for all ports and the near-field values given each excited port at 10,000 probe locations as prediction targets. Instead of sampling probe locations on a uniform grid, we dynamically sample a point cloud that is denser closer to metal regions of the board where the fields tend to be stronger, with a small share of points drawn through the conductor volumes. Figure 2 shows the near-field and S-parameter targets for one simulated board.

3.3 Model architecture

Our model is a 340M-parameter transformer-based neural operator with an encoder-decoder architecture, detailed in Figure 3.

System encoder. The board is first voxelized at 256^3 resolution ($\sim 31\mu$m per voxel) so that each voxel is represented by a feature vector of values derived from that voxel's fractional occupancy and material properties (metal occupancy, overall occupancy, average relative permittivity, average loss tangent, whether it intersects a port, etc.). The domain is divided into 16^3 patches, and patches containing only air are discarded, following Volt's occupancy-sparse tokenization (Yilmaz et al., 2026). We replace Volt's linear patch projection with a small convolutional stem that transforms each remaining patch into a token embedding. Typed port tokens representing the excited and passive ports

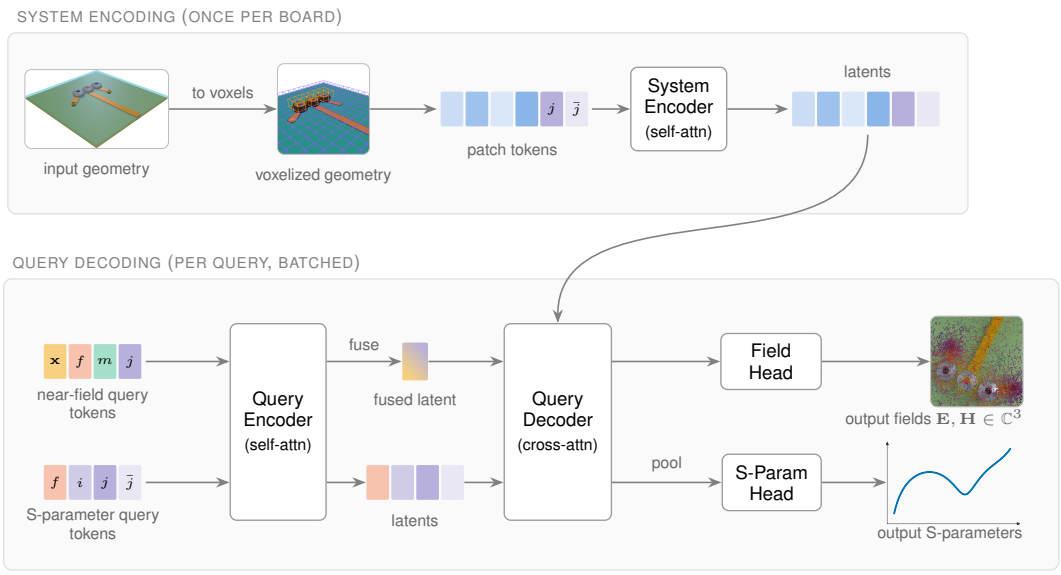


Figure 3: **Model architecture.** A sparse voxel-patch encoder represents the system geometry and material composition; a shared query encoder and cross-attention decoder condition field or S-parameter predictions on that representation before task-specific output heads. Token labels: $\mathbf{x}$ query position, f frequency, m local material, i receiving port, j excited port, $\bar{j}$ passive ports.

are concatenated to the patches, and the whole sequence is passed through a transformer encoder with global self-attention to produce a sequence of latents.

Query encoder. The model can be queried by passing a sequence of typed tokens to a query encoder, where each type has a corresponding learned embedding. A *near-field query* contains tokens for the probe location, excitation frequency, probe location material properties, and excited port identity. Note that including the material properties at the probe location allows us to specify the material at a higher granularity than is captured by the coarse voxelized system representation. An *S-parameter query* contains tokens for the excitation frequency, the excited port, receiving port, and any passive ports not already captured by the excited and receiving ports. Both query sequences are passed through the same transformer encoder to produce a sequence of latents. The near-field latents are then collapsed into a single fused latent by a learned projection, while the S-parameter latents are kept separate to maintain the positional information of the individual port tokens.

Query decoder. Query latents then cross-attend to system latents using a transformer decoder, following the query-based interface of UPT (Alkin et al., 2024). Resulting S-parameter query latents are mean-pooled into a single latent. While the query encoder and decoder modules are shared across query types, queries do not interact with each other within a model forward pass. Instead, queries are parallelized along a batch dimension to allow for efficient inference of multiple simultaneous queries given a single cached system encoding.

Output heads. To produce final model outputs, the decoder output flows through either a field head or an S-parameter head, depending on the query type. The field head maps a decoded near-field latent to the 12 real components of $\mathbf{E}, \mathbf{H} \in \mathbb{C}^3$ in physical units, together with a per-query log-scale parameter ℓ consumed by the loss. The S-parameter head applies a small MLP with a *tanh* final activation function to produce the real and imaginary components of $S_{ij}(f)$, each bounded in $[-1, 1]$.

Positions. All attention in both the encoders and the decoder uses relative 3D rotary position encoding (RoPE) (Su et al., 2024; Heo et al., 2024) computed from continuous physical coordinates in meters, making the attention operations depend on relative metric displacement rather than absolute position.

3.4 TRAINING

3.4.1 LOSS FUNCTION

We train our model using a multi-task loss function L that combines a near-field term L_{NF} and S-parameter term L_{SP},

$$L = L_{\text{NF}} + L_{\text{SP}}, \tag{1}$$

each component of which activates only for the task corresponding to the input query type. The S-parameter term is the complex MSE loss on the S-matrix column driven by the excited port j, averaged over the N_i receiving ports and N_f sampled frequencies of the board:

$$L_{\text{SP}} = \frac{1}{N_i N_f} \sum_{i=1}^{N_i} \sum_{f=1}^{N_f} \left| \hat{S}_{ij}(f) - S_{ij}(f) \right|^2. \tag{2}$$

The near-field term is a heteroscedastic robust objective based on a multivariate Student-t negative log-density, with impedance-normalized field targets:

$$\begin{aligned} L_{\text{NF}} &= \text{sg}\big(e^{\beta \ell}\big) \left[\frac{\nu + d}{2} \, \log\!\left(1 + \frac{\|\mathbf{r}\|^2 e^{-\ell}}{\nu}\right) + \frac{d}{2}\,\ell \right], \qquad d = 12, \qquad \beta = 0.5, \\ \mathbf{r} &= \frac{1}{E_{\text{ref}}} \left[\hat{\mathbf{E}} - \mathbf{E},\ \eta_0(\hat{\mathbf{H}} - \mathbf{H}) \right] \in \mathbb{R}^{12}, \end{aligned} \tag{3}$$

where $\mathbf{r}$ concatenates the real and imaginary components of the complex 3D electric and magnetic fields. The remaining terms are constructed as follows:

- **The bracketed expression** contains the residual and scale-matrix terms of an isotropic multivariate Student-t negative log-density over d dimensions, with scale matrix $e^{\ell}\mathbf{I}$ and tail parameter ν.
- **The learned** ℓ parameterizes the scale-matrix diagonal in log space and is inspired by the heteroscedastic-regression formulation of Kendall & Gal (2017). Instead of fitting ℓ globally, we make it a per-input prediction of the network.
- **The** $\text{sg}(e^{\beta\ell})$ **prefactor**, where $\text{sg}(\cdot)$ denotes stop-gradient, follows the β-NLL weighting of Seitzer et al. (2022). We apply this weighting to our robust objective to prevent scale estimation from degrading mean estimation, and adopt their recommended value of $\beta = 0.5$.
- **The learned** ν is parameterized as $\nu = 2 + \text{softplus}(\tilde{\nu})$, which enforces $\nu > 2$ and hence finite variance; $\tilde{\nu}$ is a learned global scalar.
- **The scaling of r by** E_{ref} **and** η_0 is standard physics-based non-dimensionalization. We use the free-space wave impedance $\eta_0 = 376.73\,\Omega$ to put $\mathbf{E}$ and $\mathbf{H}$ on a common footing through $|\mathbf{E}| = \eta_0|\mathbf{H}|$, and $E_{\text{ref}} = 125$ V/m to represent a 1 V excitation over the 8 mm board scale and yield unitless residuals near 1.

3.4.2 OPTIMIZATION

We train with AdamW (Loshchilov & Hutter, 2019) at a base learning rate of 5×10^{-4} under μP width scaling (Yang et al., 2021), with no weight decay, gradient clipping at 1.0, and bf16 mixed precision. We apply a warmup-stable-decay schedule (Hägele et al., 2024): 3,000 warmup steps, a stable phase at the peak rate, and a decay to zero over the final 10% of the run. Each optimizer step processes a global batch of 128 boards, sampling 512 field query positions and 8 frequencies per board. Training runs for 130,000 steps, about 72 epochs over the training dataset. To avoid exact-instance repeats across multiple epochs, we apply 90-degree in-plane rotations, reflections, and sub-patch translations pre-voxelization to every training example as a form of geometric augmentation.

3.5 EVALUATION SUITE: EMVAL

We introduce `EMVal`, a benchmark suite to evaluate the performance of foundation models for time-harmonic electromagnetic behavior. The suite consists of two tracks: `EMVal-NF`, for evaluating near-field predictions (the primary target of an electromagnetic foundation model), and `EMVal-SP`, for evaluating S-parameter predictions (a downstream task). Each track consists of several metrics defined at the level of a board, so we average each board-level metric across all boards to compute a

dataset-level metric. We estimate 95% confidence intervals by resampling boards with replacement 1,000 times and taking the 2.5th and 97.5th percentiles of the bootstrap distribution as the lower and upper bounds. To assess generalization performance, we split the evaluation dataset by geometry into two subsets, *in-distribution (ID)* and *out-of-distribution (OOD)*, where OOD examples are generated from design families completely excluded from the training set. In all cases, a lower metric value is better, with 0 being the best possible score on each metric.

EMVal-NF. Given a board excited at port j with frequency f, let $\mathbf{E}_{jfp} \in \mathbb{C}^3$ be the true electric field vector at probe p. Likewise, let $\hat{\mathbf{E}}_{jfp}$ be the predicted electric field vector (with equivalent $\mathbf{H}$ and $\hat{\mathbf{H}}$ for the magnetic field). We define the board's *global relative L_2 electric field error* over all excited ports, frequencies, and probe locations as:

$$\text{rel-}L_2(\mathbf{E}) = \frac{\sqrt{\sum_{jfp} \|\hat{\mathbf{E}}_{jfp} - \mathbf{E}_{jfp}\|_2^2}}{\sqrt{\sum_{jfp} \|\mathbf{E}_{jfp}\|_2^2}}. \tag{4}$$

Because the numerator sums squared absolute errors, this metric is dominated by regions where the true field is large. We also define the board's *local-q_{50} electric field error* by normalizing each probe's prediction error by the magnitude of its true field, then taking the median over probes,

$$\text{local-}q_{50}(\mathbf{E}) = \text{median}_{jfp}(e_{jfp}), \qquad e_{jfp} = \frac{\|\hat{\mathbf{E}}_{jfp} - \mathbf{E}_{jfp}\|_2}{\|\mathbf{E}_{jfp}\|_2}. \tag{5}$$

By normalizing per probe, this metric prevents outlier probes from dominating the score, allowing low-magnitude fields to matter just as much as high-magnitude fields.

EMVal-SP. Given a board, let $S_{ijf} \in \mathbb{C}$ be the true complex S-parameter for receiving port i, excited port j, and frequency f. Let $\hat{S}_{ijf} \in \mathbb{C}$ be the prediction and N be the number of (i, j, f) entries. We define the board's *S-parameter real-and-imaginary RMSE* as:

$$\text{riRMSE} = \sqrt{\frac{1}{2N} \sum_{ijf} \Big[\big(\text{Re}(\hat{S}_{ijf} - S_{ijf})\big)^2 + \big(\text{Im}(\hat{S}_{ijf} - S_{ijf})\big)^2 \Big]}, \tag{6}$$

where the $2N$ counts the real and imaginary parts of each complex entry as two separate dimensions. Writing magnitude in dB as $M = 20 \log_{10} |S|$ (and prediction $\hat{M} = 20 \log_{10} |\hat{S}|$), we also define the board's *S-parameter magnitude weighted-MAE* as:

$$\text{mag-wMAE} = \frac{\sum_{ijf} w_{ijf} \, |\hat{M}_{ijf} - M_{ijf}|}{\sum_{ijf} w_{ijf}}, \qquad w_{ijf} = \sigma\left(\frac{M_{ijf} - N_0}{\tau}\right), \tag{7}$$

where σ is the logistic sigmoid, $N_0 = -20$ dB is its midpoint, and $\tau = 2$ dB controls the transition width. This metric prioritizes aspects of S-parameter curves that engineers care most about by measuring magnitude error on a logarithmic scale and largely ignoring errors where the magnitude is very small.

3.6 Experimental comparisons

To assess whether 3D field supervision improves downstream task performance compared to supervising on that task alone, we compare our main jointly-trained model to an equivalent model trained on only the S-parameter targets of the full training set, all other aspects of the model architecture and optimization held equal.

We also run an ablation study to assess the impact of each of our main modeling decisions. We train a collection of ∼52M-parameter models on a reduced dataset of 10k boards, running three training seeds per variant. A 2×2 loss-and-scaling grid compares the Student-t objective with MSE and impedance-based scaling with component-wise z-score scaling. The position ablation replaces continuous relative RoPE with absolute multiband Fourier position features. The backbone ablation replaces the sparse transformer geometry encoder with a residual 3D CNN backbone matched to within 10% in parameter count.

Table 1: **Main results.** Joint field and S-parameter supervision versus S-parameter-only supervision for 340M-parameter models trained on 250k boards. Values are means over boards; parentheses give 95% bootstrap confidence intervals. Statistically significant OOD improvements are bold, and the jointly supervised model is shaded.

	EMVal-NF				EMVal-SP	
Model	rel-L_2 $\mathbf{E}\downarrow$	rel-L_2 $\mathbf{H}\downarrow$	q_{50} $\mathbf{E}\downarrow$	q_{50} $\mathbf{H}\downarrow$	riRMSE $\downarrow$	mag-wMAE (dB) $\downarrow$
ID						
Ours, field + S	0.206 (0.202, 0.210)	0.181 (0.177, 0.186)	0.234 (0.230, 0.238)	0.260 (0.255, 0.264)	0.061 (0.059, 0.063)	0.457 (0.441, 0.474)
Ours, S only	–	–	–	–	0.060 (0.057, 0.062)	0.437 (0.420, 0.455)
OOD						
Ours, field + S	0.329 (0.321, 0.336)	0.316 (0.307, 0.324)	0.289 (0.285, 0.294)	0.319 (0.315, 0.324)	**0.102 (0.098, 0.105)**	**0.489 (0.468, 0.512)**
Ours, S only	–	–	–	–	0.153 (0.148, 0.157)	0.940 (0.899, 0.979)

Additionally, we perform an architecture comparison using the same 10k-board field-only dataset, which compares a scaled-down version of our model against GINO (Li et al., 2023) and Transolver (Wu et al., 2024), all at ∼28M parameters each. GINO maps the mesh-vertex point cloud to a regular latent grid with an input graph operator, processes the grid with a 3D Fourier neural operator, and maps the result to the requested field-query coordinates with an output graph operator. Transolver instead predicts on its input point set, so for each sampled frequency we concatenate the mesh vertices and field-query points into a single point cloud, condition the Physics-Attention layers on frequency, and apply the field head only at the query points. Both baselines receive material and excited-port information; Appendix B gives the full adaptations and configurations.

3.7 Computational resources

The jointly supervised and S-parameter-only models were trained on 32 NVIDIA L40S GPUs (48 GB each) for 72 and 53 hours, respectively, totaling ∼4,000 GPU-hours. The component ablations and architecture comparison used four NVIDIA A10G GPUs (24 GB each) per run and required ∼1,100 and ∼500 GPU-hours, respectively. Generating the ∼250k-example simulation corpus on single-A10G workers required an additional ∼25,000 GPU-hours.

4 Results

4.1 Fields improve S-parameter generalization

Table 1 reports the main 340M-parameter comparison. On `EMVal-NF`, the jointly supervised model obtains in-distribution relative L_2 errors of 0.206 for $\mathbf{E}$ and 0.181 for $\mathbf{H}$, which increase to 0.329 and 0.316 on held-out design families; Appendix A presents one example held-out prediction. On `EMVal-SP`, field supervision leaves in-distribution S-parameter performance nearly unchanged, while reducing out-of-distribution riRMSE from 0.153 to 0.102 and mag-wMAE from 0.940 to 0.489 dB.

4.2 Component ablations

Table 2 isolates the effects of the field objective, position encoding, and geometry backbone in our small-scale ablation study.

Loss and target scaling. The naïve combination of MSE and per-component z-scoring performs poorly across all in-distribution metrics, with global relative errors near 1 and local-q_{50} errors above 4. Impedance scaling provides the dominant improvement: holding MSE fixed, it reduces global relative

Table 2: **Component ablations.** Results for 52M-parameter models trained on 10k boards; mean ± standard deviation over three seeds. Best values are bold, second-best values are underlined, and the default configuration is shaded.

	EMVal-NF				EMVal-SP	
Variant	rel-L_2 $\mathbf{E}$ ↓	rel-L_2 $\mathbf{H}$ ↓	q_{50} $\mathbf{E}$ ↓	q_{50} $\mathbf{H}$ ↓	riRMSE ↓	mag-wMAE (dB) ↓
ID						
default	0.274 ± 0.002	0.252 ± 0.002	**0.406** ± 0.009	**0.509** ± 0.006	**0.065** ± 0.003	0.468 ± 0.008
MSE, impedance scale	**0.264** ± 0.013	**0.234** ± 0.012	0.769 ± 0.015	0.971 ± 0.036	0.068 ± 0.022	**0.420** ± 0.068
MSE, z-score	0.993 ± 0.001	1.003 ± 0.001	5.787 ± 0.301	4.186 ± 0.604	0.466 ± 0.009	7.183 ± 0.352
Student-t, z-score	0.959 ± 0.055	1.000 ± 0.000	1.002 ± 0.028	1.014 ± 0.004	0.501 ± 0.005	8.192 ± 0.429
absolute positions	0.642 ± 0.023	0.598 ± 0.010	0.732 ± 0.019	0.913 ± 0.031	0.206 ± 0.017	2.523 ± 0.433
CNN backbone	0.278 ± 0.011	0.267 ± 0.015	0.423 ± 0.010	0.536 ± 0.012	0.082 ± 0.007	0.528 ± 0.020
OOD						
default	0.401 ± 0.006	0.411 ± 0.005	**0.529** ± 0.003	**0.655** ± 0.005	0.133 ± 0.004	0.817 ± 0.055
MSE, impedance scale	**0.371** ± 0.015	**0.374** ± 0.015	0.781 ± 0.015	0.940 ± 0.010	**0.128** ± 0.014	**0.762** ± 0.071
MSE, z-score	0.989 ± 0.005	1.000 ± 0.001	3.146 ± 0.404	2.347 ± 0.179	0.392 ± 0.002	4.697 ± 0.113
Student-t, z-score	0.984 ± 0.018	1.000 ± 0.000	0.999 ± 0.002	1.001 ± 0.000	0.398 ± 0.002	4.900 ± 0.116
absolute positions	0.761 ± 0.021	0.791 ± 0.014	0.787 ± 0.020	0.895 ± 0.008	0.303 ± 0.016	2.947 ± 0.222
CNN backbone	0.473 ± 0.049	0.478 ± 0.048	0.596 ± 0.027	0.735 ± 0.038	0.165 ± 0.017	1.139 ± 0.078

error to 0.264 for $\mathbf{E}$ and 0.234 for $\mathbf{H}$ and substantially improves both S-parameter metrics. However, it does not fully resolve local field prediction; the corresponding q_{50} errors remain 0.769 and 0.971, with $\mathbf{H}$ particularly close to the zero-prediction baseline. Replacing MSE with the heteroscedastic Student-t objective recovers much of this remaining local accuracy, reducing the q_{50} errors to 0.406 and 0.509 while retaining comparable global and downstream performance. The tradeoff persists on held-out families: impedance-scaled MSE achieves lower global relative errors and slightly better S-parameter metrics, whereas the Student-t objective substantially reduces the q_{50} errors, from 0.781 and 0.940 to 0.529 and 0.655. We therefore use the Student-t objective by default: although MSE performs better on global relative error, Student-t gives substantially lower error at the typical probe, which better reflects the model's arbitrary-location query setting.

Position encoding. Replacing relative 3D RoPE with absolute positions degrades every metric in both evaluation tiers. In distribution, for example, electric-field relative error increases from 0.274 to 0.642 and S-parameter riRMSE increases from 0.065 to 0.206.

Backbone. The matched CNN and transformer backbones perform similarly in distribution, while the transformer is better on every metric in the OOD evaluation.

4.3 Architecture comparison

Table 3 compares field-only models under a common training and evaluation protocol. Our model substantially outperforms both GINO and Transolver on every metric in both evaluation tiers. Transolver's global errors remain close to the zero-prediction baseline and its local errors exceed 1, indicating that it struggles to learn the field mapping under this protocol.

5 Discussion

Our studies show that 3D field supervision can improve S-parameter prediction even though S-parameters have a dedicated output head. The effect is small in distribution but substantial on held-out design families, suggesting that joint field prediction encourages the shared system encoder to capture electromagnetic behavior that transfers across geometry families rather than fitting only the S-parameter curves of the training families.

The ablations show that learning the challenging distributions of electric and magnetic fields depends critically on both the training objective and the representation of the physical problem. Per-component z-scoring does not preserve the impedance-based relationship between the field targets and fails

Table 3: **Architecture comparison.** Field-only results for our model, GINO, and Transolver at 28M parameters and 10k training boards; mean ± standard deviation over three seeds. Best values are bold, second-best values are underlined, and our model is shaded.

Model	rel-L_2 $\mathbf{E}\downarrow$	rel-L_2 $\mathbf{H}\downarrow$	q_{50} $\mathbf{E}\downarrow$	q_{50} $\mathbf{H}\downarrow$
ID				
Ours, field-only	**0.398** ± 0.027	**0.397** ± 0.027	**0.855** ± 0.017	**0.951** ± 0.025
GINO	0.517 ± 0.002	0.521 ± 0.001	1.049 ± 0.026	1.056 ± 0.007
Transolver	0.989 ± 0.012	0.949 ± 0.046	1.656 ± 0.231	1.570 ± 0.124
OOD				
Ours, field-only	**0.575** ± 0.035	**0.581** ± 0.036	**0.892** ± 0.021	**1.000** ± 0.015
GINO	0.680 ± 0.004	0.684 ± 0.001	1.043 ± 0.036	1.031 ± 0.009
Transolver	0.975 ± 0.019	0.959 ± 0.047	1.213 ± 0.129	1.209 ± 0.060

under both objectives, whereas impedance scaling preserves that relationship and enables useful field prediction. With impedance scaling, the contrast between global relative error and local-q_{50} shows that MSE and Student-t emphasize different spatial regimes: MSE favors high-magnitude regions, while the robust objective improves the typical probe. Relative coordinates provide a complementary inductive bias by making attention depend on relative displacement.

6 Limitations

Our OOD evaluation varies geometry family only within a bounded domain of PCB and package structures; it does not establish generalization to new material systems, substantially different electrical scales, or higher frequencies.

`EMVal` currently evaluates near fields and S-parameters, but not other quantities of practical interest such as far-field radiation or task-specific design objectives. Future work should expand both the data distribution and `EMVal` along these axes.

7 Conclusion

We introduced, to our knowledge, the first foundation model for complex 3D time-harmonic electromagnetic systems that jointly predicts complex electric and magnetic fields and S-parameters, together with a large procedurally generated simulation corpus and `EMVal`, a benchmark suite for both output modalities. Field supervision improves S-parameter prediction most strongly on held-out geometry families, while our ablations identify impedance-aware target scaling, a robust field objective, and relative spatial encoding as important ingredients for accurate complex-field prediction. Together with gains over established neural-operator baselines, these results support a path toward reusable learned models of electromagnetic behavior.

AI use statement

We used Claude Code and Codex to assist with experimental design, code implementation, training-run management, and preliminary analysis of experimental results. The authors reviewed all AI-assisted designs, code, and analyses for correctness and performed the final analyses themselves. We also used Codex to survey related literature and edit the manuscript for clarity and style. The authors wrote most of the manuscript by hand and take responsibility for its final content.

Reproducibility statement

Section 3.1 describes dataset construction. Section 3 describes the model and training objectives. The experimental results report evaluation metrics and ablation settings. The appendices provide additional implementation and training details.

REFERENCES

Benedikt Alkin, Andreas Fürst, Simon Schmid, Lukas Gruber, Markus Holzleitner, and Johannes Brandstetter. Universal physics transformers: A framework for efficiently scaling neural operators. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Benedikt Alkin, Maurits Bleeker, Richard Kurle, Tobias Kronlachner, Reinhard Sonnleitner, Matthias Dorfer, and Johannes Brandstetter. AB-UPT: Scaling neural CFD surrogates for high-fidelity automotive aerodynamics simulations via anchored-branched universal physics transformers. *Transactions on Machine Learning Research*, 2025. arXiv:2502.09692.

AWS Center for Quantum Computing. Palace: 3D finite element solver for computational electromagnetics, 2025. URL `https://github.com/awslabs/palace`. Open-source parallel finite element code for full-wave 3D electromagnetic simulation, built on MFEM.

Jingyun Bi, Xinyu Zhou, Jing Xia, Shichang Chen, and Wing Shing Chan. Frequency-query enhanced electromagnetic surrogate modeling with edge anti-aliasing pixelation for bandpass filter inverse design. In *IEEE International Microwave Symposium (IMS)*, pp. 614–617, 2025. doi: 10.1109/IMS40360.2025.11103901.

Mingkun Chen, Robert Lupoiu, Chenkai Mao, Der-Han Huang, Jiaqi Jiang, Philippe Lalanne, and Jonathan A. Fan. High speed simulation and freeform optimization of nanophotonic devices with physics-augmented deep learning. *ACS Photonics*, 9(9):3110–3123, 2022. doi: 10.1021/acsphotonics.2c00876.

Feng Feng, Venu-Madhav-Reddy Gongal-Reddy, Chao Zhang, Jianguo Ma, and Qi-Jun Zhang. Parametric modeling of microwave components using adjoint neural networks and pole-residue transfer functions with EM sensitivity analysis. *IEEE Transactions on Microwave Theory and Techniques*, 65(6):1955–1975, 2017. doi: 10.1109/TMTT.2017.2650904.

Orkun Furat, Vinay Chakravarthi Gogineni, Henrik Bindslev, and Esmaeil S. Nadimi. Physics-informed neural operators for predicting 3D electromagnetic fields transformed by metasurfaces. *arXiv preprint arXiv:2512.15694*, 2025.

Jiaqi Gu, Zhengqi Gao, Chenghao Feng, Hanqing Zhu, Ray T. Chen, Duane S. Boning, and David Z. Pan. NeurOLight: A physics-agnostic neural operator enabling parametric photonic device simulation. In *Advances in Neural Information Processing Systems*, 2022.

Alexander Hägele, Elie Bakouch, Atli Kosson, Loubna Ben Allal, Leandro von Werra, and Martin Jaggi. Scaling laws and compute-optimal training beyond fixed training durations. In *Advances in Neural Information Processing Systems*, 2024.

Byeongho Heo, Song Park, Dongyoon Han, and Sangdoo Yun. Rotary position embedding for vision transformer. In *European Conference on Computer Vision (ECCV)*, 2024.

Maximilian Herde, Bogdan Raonić, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel de Bézenac, and Siddhartha Mishra. Poseidon: Efficient foundation models for PDEs. In *Advances in Neural Information Processing Systems*, 2024.

Qile Jiang, Marc Salvadori, Dale Ota, Vijaya Shankar, and Khemraj Shukla. Complex valued deep operator network (DeepONet) $[\mathcal{G}]$ for three dimensional Maxwell's equations: $\mathcal{G} \in \mathbb{C}^{m \times n}$. *Journal of Computational Physics*, 562:114993, 2026a. doi: 10.1016/j.jcp.2026.114993.

Xiaotian Jiang, Huan Huan Zhang, Kaitai Guo, Yang Zheng, and Jimin Liang. Mesh-based deep learning surrogate model for rapid wideband 3-D electromagnetic scattering computation. *IEEE Transactions on Antennas and Propagation*, 74(5):4546–4561, 2026b. doi: 10.1109/TAP.2026.3661600.

Emir Ali Karahan, Zheng Liu, Aggraj Gupta, Zijian Shao, Jonathan Zhou, Uday Khankhoje, and Kaushik Sengupta. Deep-learning enabled generalized inverse design of multi-port radio-frequency and sub-terahertz passives and integrated circuits. *Nature Communications*, 15:10734, 2024. doi: 10.1038/s41467-024-54178-1.

Alex Kendall and Yarin Gal. What uncertainties do we need in Bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems*, 2017.

Zongyi Li, Nikola Borislavov Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Prakash Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, and Anima Anandkumar. Geometry-informed neural operator for large-scale 3D PDEs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Chenkai Mao and Jonathan A. Fan. Accurate and scalable deep Maxwell solvers. *Proceedings of the National Academy of Sciences*, 123(18):e2530330123, 2026. doi: 10.1073/pnas.2530330123.

Maximilian Seitzer, Arash Tavakoli, Dimitrije Antić, and Georg Martius. On the pitfalls of heteroscedastic uncertainty estimation with probabilistic neural networks. In *International Conference on Learning Representations*, 2022.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. RoFormer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024. doi: 10.1016/j.neucom.2023.127063.

Hakki Mert Torun, Ahmet Cemal Durgun, Kemal Aygün, and Madhavan Swaminathan. Causal and passive parameterization of S-parameters using neural networks. *IEEE Transactions on Microwave Theory and Techniques*, 68(10):4290–4304, 2020. doi: 10.1109/TMTT.2020.3011449.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for PDEs on general geometries. In *International Conference on Machine Learning (ICML)*, 2024.

Greg Yang, Edward J. Hu, Igor Babuschkin, Szymon Sidor, Xiaodong Liu, David Farhi, Nick Ryder, Jakub Pachocki, Weizhu Chen, and Jianfeng Gao. Tuning large neural networks via zero-shot hyperparameter transfer. In *Advances in Neural Information Processing Systems*, 2021.

Kadir Yilmaz, Adrian Kruse, Tristan Höfer, Daan de Geus, and Bastian Leibe. Volume Transformer: Revisiting Vanilla Transformers for 3D Scene Understanding. In *European Conference on Computer Vision (ECCV)*, 2026.

A EXAMPLE MODEL PREDICTION

Figure 4 shows one field prediction on a held-out hairpin filter, together with the corresponding per-probe field calibration and S-parameters.

B NEURAL-OPERATOR BASELINE DETAILS

All three models in the architecture comparison use the same training and evaluation data, sampled field-query workload, augmentation, physics-scaled MSE objective, optimizer family, 80,000-step schedule, and checkpoint-evaluation protocol; the learning rate is calibrated separately for each architecture.

GINO. Our GINO implementation uses mesh vertices as its input point cloud. Each vertex is represented by its material properties, interface status, port and excited-port indicators, pin type, and distances to the nearest metal face and excited-port centroid. A radius graph operator maps these features to a 24^3 latent grid over a fixed 16 mm cube. The grid is processed by four width-66 3D FNO blocks with 11 Fourier modes per dimension, after which a second radius graph operator evaluates the latent representation at the exact field-query coordinates. Query material and distance features are concatenated before the output head, which predicts the 12 real-valued components of the complex $\mathbf{E}$ and $\mathbf{H}$ vectors. Frequency and excited-port embeddings condition the input graph operator, FNO blocks, and query readout. We use input and output graph radii of 2.0 and 1.5 mm, respectively, with at most 64 neighbors. Relative offsets in each graph operator are encoded with a fixed log-spaced Fourier basis and used to predict normalized neighbor weights and channel-wise directional gates. This learned graph kernel and the use of feature-wise linear modulation for conditioning are adaptations of the reference GINO architecture (Li et al., 2023).

Transolver. Our Transolver implementation follows the Physics-Attention sequence of learned slicing, self-attention between slice tokens, and deslicing to the input points (Wu et al., 2024). Because the resulting representation is defined on the input point set, we group field queries by frequency and concatenate the queries for each group with a copy of the geometry mesh. Mesh vertices carry coordinate, material, interface, port, excited-port, and pin-type features. Query points carry their coordinates and local material properties and are assigned the default non-interface, non-port, and no-pin values; no separate query-type embedding is used. Frequency conditions each Physics-Attention layer through feature-wise linear modulation. We use six layers of width 512, eight attention heads, and 64 slice tokens. The final field head is applied only to the desliced query-point representations, and predictions at the mesh vertices are discarded.

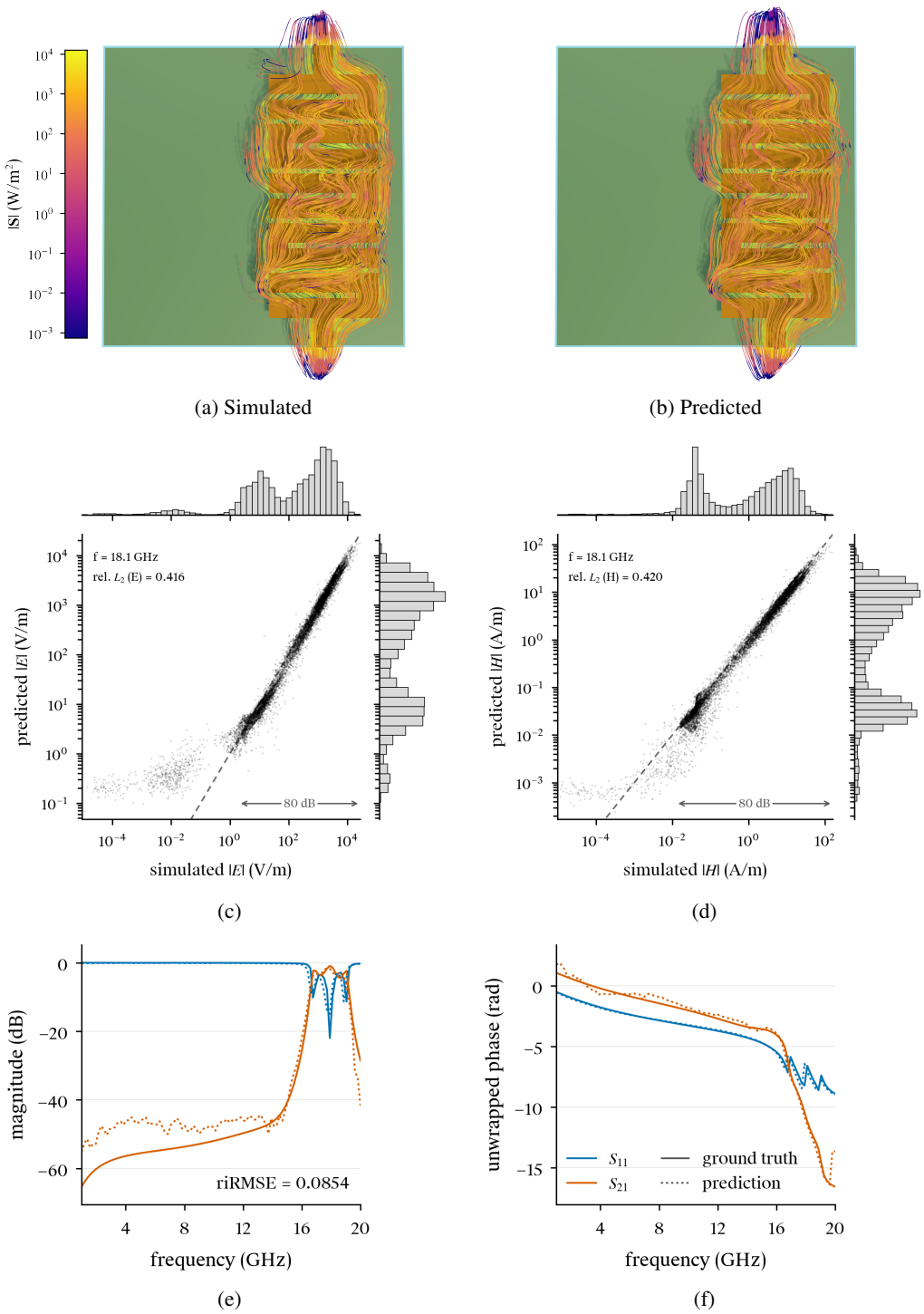


Figure 4: **Example model prediction on a held-out hairpin filter.** All panels show the same board and excited port. (a, b) Simulated and predicted time-averaged Poynting vector $\mathbf{S} = \frac{1}{2}\,\mathrm{Re}(\mathbf{E} \times \mathbf{H}^*)$ at 18.1 GHz, near the center of the filter's passband, drawn as streamlines for the reference simulation and for our model at the same 10k probe coordinates. (c, d) Predicted against simulated field magnitude at that frequency, one point per probe, for $|\mathbf{E}|$ and $|\mathbf{H}|$. Both calibrations follow the dashed line at $y = x$ closely over a dynamic range of 80 dB, which, when measured from the peak, captures 91% of $|\mathbf{E}|$ values and 97% of $|\mathbf{H}|$ values in the distribution. (e, f) The predicted S-parameter response against simulation, decomposed into magnitude and phase.