Cross-Topology Transfer Learning Using Bayesian Optimization for Scalable Surrogate Modeling of 3D Packaging Structures

Md Sultan Mahmud
Pennsylvania State University
University Park, PA, USA
mqm7099@psu.edu

Xianbo Yang
IBM Corporation
Austin, TX, USA
xianboyang@ibm.com

Priyank Kashyap
Hewlett Packard Enterprise
Durham, NC, USA
priyank.kashyap@hpe.com

Madhavan Swaminathan*
Pennsylvania State University
University Park, PA, USA
mvs7249@psu.edu

***Abstract*—Accurate simulation of electromagnetic structures is critical for semiconductor packaging, yet full-wave solvers impose prohibitive computational costs. This paper proposes a machine learning (ML) surrogate framework integrated with Bayesian optimization (BO) to accelerate signal integrity (SI) analysis. Unlike traditional topology-specific models, this approach utilizes transfer learning to adapt pretrained behavioral models to new geometries, preserving transferable features while isolating topology-specific variations. The framework is validated on a vertical via transition and a high-speed connector interface. Experimental results demonstrate that the proposed transfer learning framework significantly outperforms a baseline model utilizing the identical architecture but trained without knowledge transfer. For the via model, transfer learning achieves greater than 90% test accuracy with only 100 training samples, whereas the baseline yields 73% on the same subset and requires a 1000-sample dataset to reach comparable convergence when predicting insertion loss (S21) over the target frequency range. In the connector analysis, the framework achieves a fourfold reduction in required training data for both insertion loss (S21) and time domain reflectometry (TDR) predictions. Specifically, TDR prediction test accuracy for a 50-sample dataset improves from 58% to 83%, highlighting the efficacy of the proposed method in capturing time-domain impedance discontinuities with minimal high-fidelity simulation overhead. The end-to-end training requires approximately 1 hour, enabling rapid and scalable design closure for next-generation interconnects.**

***Keywords*—3D packaging, heterogeneous integration, signal integrity (SI), electronic design automation (EDA), surrogate modeling, machine learning (ML), transfer learning (TL), Bayesian optimization (BO), Gaussian processes (GP), simulation acceleration, design space exploration (DSE).**

I. Introduction

THE relentless scaling of data transmission rates in high-performance computing (HPC) systems has necessitated a paradigm shift toward significantly more complex packaging structures. Driven by the exponential growth in computing performance required by artificial intelligence (AI) and machine learning (ML) workloads, the industry is increasingly adopting heterogeneous integration and advanced 2.5D and 3D packaging technologies to overcome the physical scaling limits of monolithic designs [1]. Within these architectures, high-bandwidth interconnects, such as connectors, vias, and transmission lines, must support unprecedented data rates while navigating the stringent constraints of bandwidth density and energy efficiency. However, this increased structural complexity introduces severe signal integrity (SI) challenges, as higher data transmission rates make the system more susceptible to parasitic effects, crosstalk, and impedance discontinuities [2].

To ensure robust performance in these high-speed links, accurate electromagnetic (EM) characterization is essential. Traditionally, this relies on full-wave 3D EM solvers to capture intricate field interactions and frequency-dependent behaviors. While these physics-based simulations provide the necessary high fidelity for capturing complex parasitics, they impose prohibitive computational costs, often requiring hours or days to simulate complex vertical via transitions or connector interfaces [3], [4]. This latency creates a critical bottleneck in the design cycle, rendering comprehensive design space exploration and iterative optimization practically infeasible for modern time-to-market windows.

ML surrogate models have emerged as a powerful alternative to address this efficiency gap by approximating the mapping between design parameters and EM responses [2]–[9]. However, existing ML frameworks are constrained by their lack of topological generalizability, which severely limits their applicability to new circuit topologies. Training a high-accuracy surrogate for a new interconnect geometry usually requires generating a massive dataset of full-wave simulations from scratch, which paradoxically reintroduces the very computational overhead the model aims to eliminate. Existing methods often fail to leverage the physical knowledge embedded in previously simulated, correlated topologies, treating each new design task as an isolated learning problem. While transfer learning has been introduced to mitigate this bottleneck by repurposing pretrained models, existing implementations typically rely on full-model fine-tuning [10]. This process overwrites the source model parameters, eliminating their reusability and resulting in a fragmented collection of isolated surrogates rather than a cohesive library. Furthermore,

This research was supported by the National Science Foundation under Grant No. CNS 2137288, 2137283 and 2137259 – Center for Advanced Electronics through Machine Learning (CAEML) and its industry members.
*Corresponding author.

these approaches frequently depend on handcrafted feature extraction and heuristic hyperparameter tuning, which increases development time and impedes the automated generation of optimal behavioral models.

To strictly address these limitations, this work proposes a novel transfer learning framework integrated with Bayesian optimization (BO), designed to achieve three central objectives. First, to eliminate the reliance on manual feature extraction and trial-and-error tuning, the framework employs BO guided by Gaussian processes (GP) to autonomously explore the architectural search space, ensuring optimal model convergence without human intervention. Second, to facilitate efficient cross-topology adaptation, we introduce a transfer learning architecture utilizing lightweight adapters. By aligning the shared design space between source and target topologies, these adapters preserve transferable latent features while isolating and learning only the variations unique to the new geometry. Third, the framework preserves source model integrity by decoupling transferable components from adapters assigned to specific tasks. This separation prevents model fragmentation and enables the construction of a scalable library of reusable packaging surrogates.

The remainder of this paper is organized as follows. Section II details the proposed methodology. Section III presents the experimental validation on a vertical via transition, demonstrating the efficacy of the framework in scaling layer counts. Section IV extends this validation to a high-speed connector interface, analyzing the transferability between distinct interface standards. Finally, Section V concludes the paper.

II. Proposed Framework

The primary objective of this work is to develop a generalized and data efficient surrogate modeling framework capable of predicting the electromagnetic behavior of diverse high speed interconnect topologies with minimal simulation overhead. We consider two distinct circuit topologies, denoted as the source topology (α) and the target topology (β), where f_α and f_β represent their corresponding behavioral models.

Let the design space of the source topology be denoted by $\mathsf{D}_\alpha \subset \mathbb{R}^{d_\alpha}$ where d_α represents the dimensionality of its design parameters including geometric dimensions, material properties, and stackup configurations. Similarly, let $\mathsf{D}_\beta \subset \mathbb{R}^{d_\beta}$ represent the design space of the target topology. The corresponding electromagnetic responses such as S-parameters or TDR waveforms are denoted by $\mathbf{Y}_\alpha \in \mathbb{C}^N$ and $\mathbf{Y}_\beta \in \mathbb{C}^N$ where N is the number of frequency or time points sampled.

In a standard supervised learning paradigm a surrogate model $f_\beta : \mathsf{D}_\beta \rightarrow \mathbf{Y}_\beta$ is trained by minimizing a loss function L over a dataset $\mathsf{S}_\beta = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{M_\beta}$ where $\mathbf{x}_i \in \mathsf{D}_\beta$ and $\mathbf{y}_i \in \mathbf{Y}_\beta$ are obtained via full wave electromagnetic simulations. However generating a sufficiently large target dataset size M_β to ensure generalization is computationally prohibitive for complex 3D EM structures.

To circumvent this data scarcity, we propose a transfer learning approach that leverages a pretrained source model f_α trained on a comprehensive dataset S_α where $|\mathsf{S}_\alpha| \gg |\mathsf{S}_\beta|$.

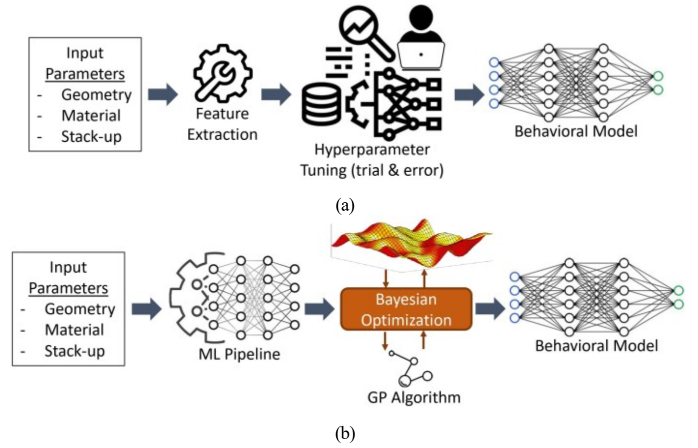


Fig. 1: Comparison of surrogate modeling workflows. (a) Traditional approach relying on manual feature extraction and trial-and-error hyperparameter tuning. (b) Proposed automated framework using Bayesian Optimization (BO). The BO process utilizes a GP surrogate to model the validation loss of the neural network, iteratively selecting the optimal architectural and training hyperparameters ($\boldsymbol{\lambda}$) to ensure convergence to a robust model.

We postulate that α and β share underlying physical characteristics despite their differing geometric parametrizations. Consequently, the design spaces can be conceptually decomposed into shared and topology specific subspaces as

$$\mathsf{D}_\beta = \mathsf{D}_{\text{shared}} \cup \mathsf{D}_{\text{unique}}^{\beta}, \tag{1}$$

where $\mathsf{D}_{\text{shared}}$ encompasses parameters physically analogous between topologies and $\mathsf{D}_{\text{unique}}^{\beta}$ contains parameters exclusive to the target structure.

The optimization problem is thus reformulated as learning a transfer function $\mathsf{F}_{\text{trans}}$ that adapts the pretrained knowledge of f_α to the target domain using a minimal set of target samples M_β. This is achieved by optimizing a set of adapter parameters θ_{adapt} while keeping the pretrained weights θ_{shared} of the source model frozen as

$$\min_{\theta_{\text{adapt}}} \sum_{j=1}^{M_\beta} \mathsf{L}\left(\mathsf{F}_{\text{trans}}(\mathbf{x}_j; \theta_{\text{shared}}, \theta_{\text{adapt}}), \mathbf{y}_j\right), \tag{2}$$

subject to the constraint that M_β is a small fraction of the data typically required for training from scratch. To realize this framework effectively we address three fundamental research questions:

- **RQ1:** Can we build fast and accurate behavioral models while eliminating the inefficiencies of manual hyperparameter tuning and feature extraction?
- **RQ2:** Can we leverage the behavioral model for the common design space shared across related topologies while effectively isolating and discarding non transferable design parameters?
- **RQ3:** Can we preserve the reusability of pretrained behavioral models when adapting to new design tasks without incurring the computational cost of retraining or the risk of catastrophic forgetting?

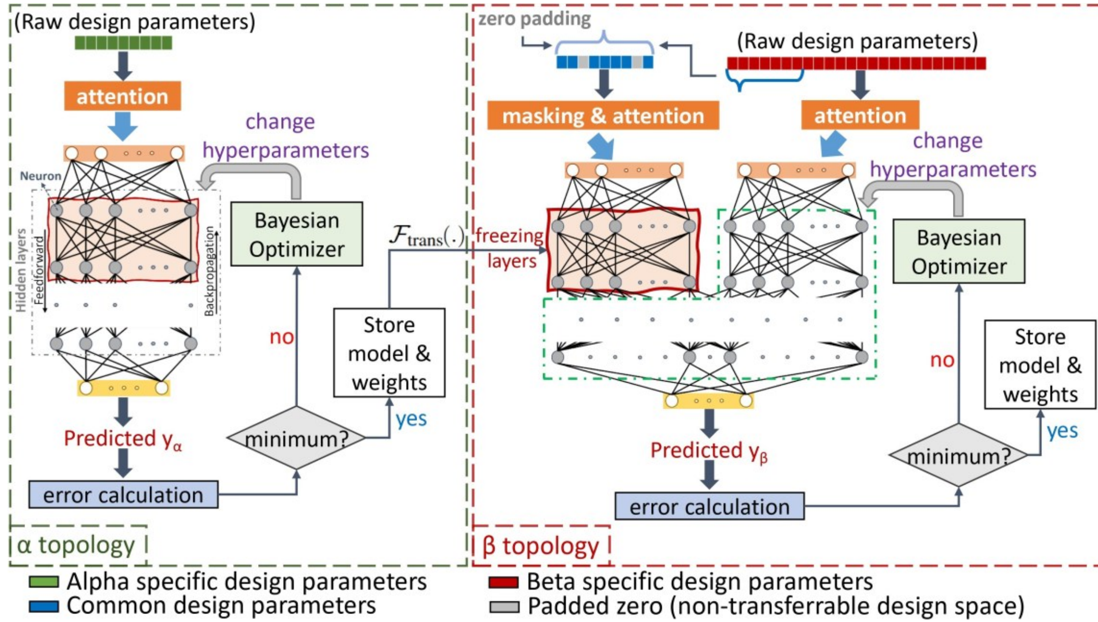


Fig. 2: Proposed transfer learning architecture with input unification and lightweight adapters. The design spaces of source (α) and target (β) topologies are aligned via masking and zero-padding, creating a unified input vector $\mathbf{x}_{unified}$. The backbone layers (θ_{shared}) pretrained on the source are frozen to preserve latent physical features. Trainable adapter modules (θ_{adapt}) are inserted to map topology-specific geometric variations to the shared feature space, minimizing the training data required for the target task.

A. Bayesian Optimization

As illustrated in Fig. 1a, the development of accurate surrogate models is traditionally hindered by the need for manual feature extraction and iterative hyperparameter tuning. This heuristic process is computationally inefficient, prone to human bias, and often fails to locate the global optimum. To strictly address **RQ1** and eliminate these inefficiencies, this framework integrates a Bayesian optimization (BO) strategy, as depicted in Fig. 1b. As detailed in [8], BO provides a sample-efficient mechanism for optimizing black-box objective functions that are expensive to evaluate, such as the validation loss of a neural network.

We define the hyperparameter search space Λ, which encompasses architectural parameters (e.g., number of hidden layers L, neurons per layer N_l) and training hyperparameters (e.g., batch size B, optimizer type). The objective is to identify the configuration $\boldsymbol{\lambda}^* \in \Lambda$ that minimizes the validation loss L_{val} of the neural network f trained on dataset D:

$$\boldsymbol{\lambda}^* = \underset{\boldsymbol{\lambda} \in \Lambda}{\operatorname{argmin}} \, \mathsf{L}_{val}(f(\cdot\,; \boldsymbol{\lambda}), \mathsf{D}). \tag{3}$$

Since the objective function $g(\boldsymbol{\lambda}) = \mathsf{L}_{val}$ has no closed-form expression, we employ a Gaussian Process (GP) as a probabilistic surrogate. The GP defines a prior distribution over functions specified by a mean function $m(\boldsymbol{\lambda})$ and a covariance kernel $k(\boldsymbol{\lambda}, \boldsymbol{\lambda}')$:

$$g(\boldsymbol{\lambda}) \sim \mathsf{GP}(m(\boldsymbol{\lambda}), k(\boldsymbol{\lambda}, \boldsymbol{\lambda}')). \tag{4}$$

Given observed evaluations D_t, the posterior distribution at a new point $\boldsymbol{\lambda}$ is Gaussian with mean $\mu_t(\boldsymbol{\lambda})$ and variance $\sigma_t^2(\boldsymbol{\lambda})$. To balance exploration and exploitation, we utilize the Expected Improvement (EI) acquisition function:

$$EI(\boldsymbol{\lambda}) = (g_{best} - \mu_t(\boldsymbol{\lambda}))\Phi(Z) + \sigma_t(\boldsymbol{\lambda})\phi(Z), \tag{5}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function (CDF) and probability density function (PDF) of the standard normal distribution, respectively. The term $Z = \frac{g_{best} - \mu_t(\boldsymbol{\lambda})}{\sigma_t(\boldsymbol{\lambda})}$ represents the standardized improvement, quantifying the potential gain relative to the uncertainty. The optimization proceeds iteratively by maximizing EI to select the next configuration until convergence, ensuring the automated generation of an optimal behavioral model.

B. Adapter-Based Transfer Learning Architecture

To strictly address **RQ2** and **RQ3**, we propose a neural network architecture augmented with lightweight adapters. As demonstrated in recent advancements in computer vision and natural language processing (NLP) [11], [12], adapter-based transfer learning has emerged as a highly parameter-efficient alternative to full-model fine-tuning. In these domains, large pretrained models are adapted to downstream tasks by inserting small, trainable modules while keeping the vast majority of the pretrained parameters frozen. This strategy not only drastically reduces the computational cost of training but also prevents the catastrophic forgetting of the source domain knowledge.

To prevent catastrophic forgetting in our context, the network must retain access to learned source features while accommodating new inputs. We achieve this via a dual strategy of parameter freezing and input space unification.

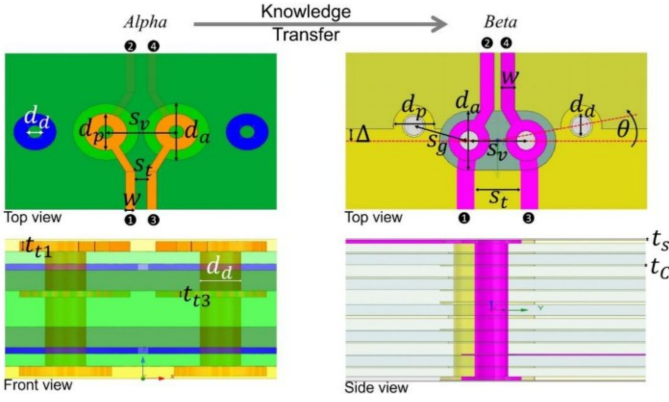


Fig. 3: 3D Electromagnetic models of the vertical via transitions. *Alpha*: 6-layer via. *Beta*: 12-layer via.

TABLE I: Design Parameters for 6-Layer Via. **Bolded parameters** denote unique geometric features specific to the topology (D_{unique}), while unbolded entries represent common design variables shared across α and β.

Parameter	Range
i/p ports Trace spacing ($s_{t,in}$)	3.5 to 20 mil
i/p ports Trace thicknes ($t_{t,in}$)	0.6 to 1.55 mil
i/p ports Trace width (w_{in})	2.5 to 12 mil
o/p ports Trace spacing ($s_{t,out}$)	3.5 to 20 mil
o/p ports Trace thicknes ($t_{t,out}$)	0.6 to 1.55 mil
o/p ports Trace width (w_{out})	2.5 to 12 mil
Anti-pad diameter (d_a)	22 to 28 mil
Drill diameter (d_d)	8 to 10 mil
Pad diameter (d_p)	16 to 20 mil

The fundamental challenge in this physical domain lies in the dimensional mismatch between the source design space D_α and the target design space D_β. As depicted in Fig. 2, we employ a masking and zero-padding strategy to unify the input spaces. We construct a unified input vector $\mathbf{x}_{unified}$ where inputs corresponding to unique source parameters are masked (set to zero) and inputs corresponding to unique target parameters are padded. This aligns the shared design space D_{shared} while isolating D_{unique}:

$$\mathbf{x}_{input} = \text{Concat}(\mathbf{x}_{shared}, \mathbf{0}_\alpha, \mathbf{x}^\beta_{unique}). \quad (6)$$

Here, $\mathbf{0}_\alpha$ represents a zero vector with dimensionality equal to the number of unique source parameters (i.e., $\mathbf{0}_\alpha \in \mathbb{R}^{\dim(D^\alpha_{unique})}$). This explicit zero-padding ensures that the input neurons in the frozen backbone corresponding to source-specific geometric features are multiplied by zero, effectively deactivating them during inference on the target topology.

The network processes this unified input through a composition of frozen backbone layers and trainable adapter layers. The backbone weights θ_{shared} are transferred from the source model and frozen to preserve prior knowledge (satisfying **RQ3**). The adapter layers θ_{adapt} are inserted between backbone blocks to capture topology-specific variations. The forward pass for layer l is given by:

$$\mathbf{h}^{(l)} = \mathbf{W}^{(l)}_{adapt}\sigma\left(\mathbf{W}^{(l)}_{frozen}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}_{frozen}\right) + \mathbf{b}^{(l)}_{adapt}. \quad (7)$$

Training minimizes the loss solely with respect to θ_{adapt}, allowing the model to adapt to β using only a small fraction of the training data typically required.

TABLE II: Design Parameters for 12-Layer Via. **Bolded parameters** denote unique geometric features specific to the topology (D_{unique}), while unbolded entries represent common design variables shared across α and β.

Parameter	Range
1st layer Copper thickness (t_{Cu1})	0.6 to 1.55 mil
2nd layer Copper thickness (t_{Cu2})	0.6 to 1.55 mil
3rd layer Copper thickness (t_{Cu3})	0.6 to 1.55 mil
4th layer Copper thickness (t_{Cu4})	0.6 to 1.55 mil
5th layer Copper thickness (t_{Cu5})	0.6 to 1.55 mil
6th layer Copper thickness (t_{Cu6})	0.6 to 1.55 mil
2nd layer Core thickness (t_{c2})	1.9 to 8 mil
4th layer Core thickness (t_{c4})	1.9 to 8 mil
6th layer Core thickness (t_{c6})	1.9 to 8 mil
1st layer Prepreg thickness (t_{p1})	1.9 to 8 mil
3rd layer Prepreg thickness (t_{p3})	1.9 to 8 mil
5th layer Prepreg thickness (t_{p5})	1.9 to 8 mil
Solder mask thickness (t_s)	0.4 to 0.6 mil
GND via angle (θ)	-45 to 45 degree
GND via spacing (s_g)	19.5 to 64 mil
GND via offset (Δ)	-10 to 40 mil
i/p ports Trace spacing ($s_{t,in}$)	3.5 - 20 mil
o/p ports Trace spacing ($s_{t,out}$)	3.5 - 20 mil
i/p ports Trace width (w_{in})	2.5 - 12 mil
o/p ports Trace width (w_{out})	2.5 - 12 mil
Anti-pad diameter (d_a)	22 - 36 mil
Drill diameter (d_d)	8 - 10 mil
Pad diameter (d_p)	16 - 20 mil
Via spacing (s_v)	19.5 - 64 mil

III. Case Study I: PTH Vias-in-package

To validate the proposed transfer learning framework, we consider a family of Plated Through-Hole (PTH) vertical via transitions, a fundamental discontinuity in 3D packaging. The source topology (α), illustrated in Fig. 3, consists of a 6-layer stack-up with a Ground-Signal-Signal-Ground (GSSG) configuration. The signal path transitions from a differential microstrip on the top layer to differential striplines on the third metal layer. This structure is parameterized by 9 design variables. On the other hand, the target topology (β), shown in Fig. 3, represents a 12-layer stack-up. Here, the signal transitions from differential microstrips on the top layer to striplines on the 10th layer and finally to microstrips on the 12th layer. This increased vertical depth and layer count expands the design space to 24 parameters. The design parameters for both topologies, including their respective ranges, are detailed in Table I and Table II.

A. Data Generation

Ground truth data for both topologies was generated using the ANSYS HFSS 3D full-wave electromagnetic solver [13]. To ensure a comprehensive coverage of the design space, Latin Hypercube Sampling (LHS) was employed to generate the training and testing samples. The simulations were configured with a center frequency of 10 GHz, and each full-wave simulation required approximately 8 minutes of computational time.

For α, the scattering parameters (*S*-parameters) were simulated over a frequency range of 0.02 GHz to 20 GHz, discretized into 1001 frequency points. A total of 1858 samples were generated to train and test the robust backbone model.

For β, simulations were conducted over a range of 0.05 GHz to 20 GHz. A total of 1000 samples were generated to

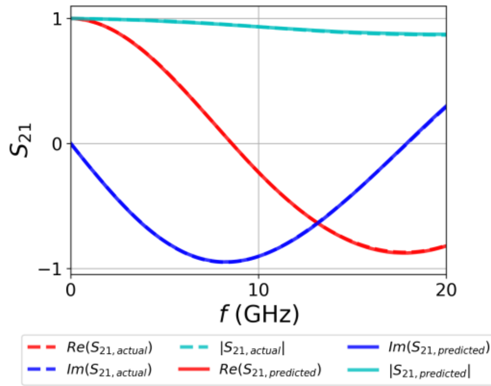


Fig. 4: Comparison of HFSS simulated (solid) and ML predicted (dashed) S-parameters for the source 6-layer via topology.

constitute the full training dataset (100%). To comprehensively evaluate the efficiency of the proposed method, we created data subsets representing 10%, 25%, 50%, and 100% of the total samples. Both the transfer learning framework and the baseline model (trained from scratch) were trained on these identical subsets to provide a fair comparison of convergence behavior and data efficiency across the entire spectrum of data availability.

B. Model Setup and Training

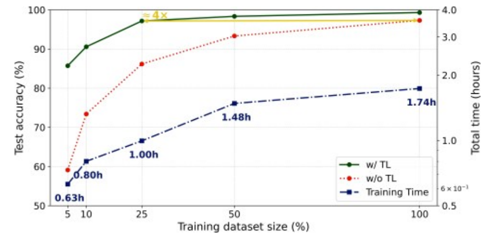
The machine learning framework was implemented in PyTorch [14] and trained on Windows with an Intel Xeon w5-3433 CPU and an NVIDIA RTX 4000 Ada Generation GPU.

The training process proceeded in two stages. First, the source model was optimized using the BO framework described in Section II, identifying the optimal ML architecture that minimized validation loss on the source topology. Second, for the target model, the weights of this pretrained backbone were frozen. The 24-dimensional input of the target topology was mapped to the shared feature space using the zero-padding strategy, and only the lightweight adapter layers were optimized using the BO framework.

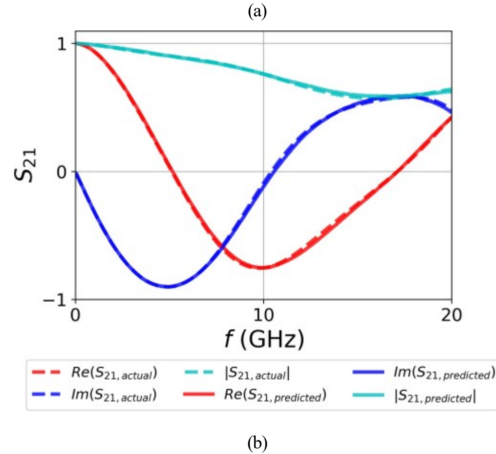
C. Results and Analysis

The performance of the proposed transfer learning (TL) framework was benchmarked against the baseline model (without pretraining) across the varying training dataset sizes.

The source model (f_a) achieved a baseline accuracy of 99.97% on its test set, confirming the ability to capture the underlying physics of vertical via transitions. This high-fidelity prediction is illustrated in Fig. 4, where the predicted S-parameters for a randomly selected test sample show near-perfect agreement with the HFSS simulation. When adapted to the target topology, the TL framework demonstrated superior initialization and rapid convergence, as shown in Fig. 5a. In the data-scarce regime (10% data or 100 samples), the TL model achieved over 90% accuracy, whereas the baseline model struggled to generalize, reaching only 73%. This performance gap of approximately 17% highlights the effectiveness of the transferred physical features. As the dataset size increased



(a)



(b)

Fig. 5: (a) Test accuracy comparison between the proposed Transfer Learning (TL) method and the baseline model (trained from scratch) as a function of training data size for the 12-layer via, including total training time for the target model. The ≈ 4× annotation highlights the fourfold reduction in required training data for the TL model to achieve parity with the maximum accuracy of the baseline model. (b) Comparison of HFSS simulated (solid) and TL predicted (dashed) S_{21} parameters (Real, Imaginary, and Magnitude) for the target 12-layer via topology, generated using the model trained on 25% of the dataset (250 samples).

to 25%, the TL model reached 97% accuracy compared to 87% for the baseline. While both models eventually converged toward high accuracy as the dataset size approached 100%, the TL framework reached acceptable fidelity thresholds (e.g., $> 95\%$ accuracy) with significantly fewer samples.

Fig. 5b presents the comparison between the HFSS-simulated (solid lines) and ML-predicted (dashed lines) S-parameters (S_{21}) for the 12-layer target via, generated using the model trained on only 25% of the dataset (250 samples). The results indicate excellent agreement across the entire frequency band, validating that the optimized adapters successfully mapped the complex 24-parameter space of the 12-layer via to the learned features of the 6-layer counterpart.

A key contribution of this framework is its ability to generalize across fundamentally different stackup configurations. In Case Study I, the model successfully transfers physical intuition from a 6-layer source to a 12-layer target, despite a nearly 3x expansion in the design parameter space (from 9 to 24 variables). This demonstrates that the lightweight adapter architecture does not merely fine-tune weights but effectively maps a more complex, high-dimensional vertical interconnect

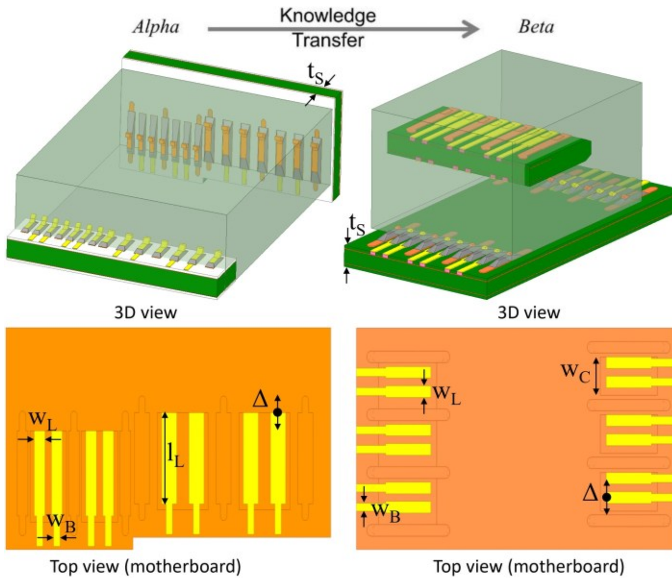


Fig. 6: 3D Electromagnetic models of the high-speed connector interfaces. *Alpha*: SATA footprint on Rogers substrate. *Beta*: PCIe footprint on Megatron6 substrate.

TABLE III: Design Parameters for SATA. **Bolded entries** indicate topology-specific variations, whereas unbolded parameters represent transferable physical dimensions.

Parameter	Range
Landing pad displacement (Δ)	-0.5 to 1 mm
Landing pad width (w_L)	0.3 to 0.7 mm
Landing pad length (l_L)	2 to 3.2 mm
Breakout trace width (w_B)	0.2 to 0.4 mm
Rogers substrate thickness (t_S)	0.2 to 0.4 mm

topology onto a lower-dimensional latent subspace of learned electromagnetic behaviors. This capability allows for the rapid characterization of advanced 3D packaging structures without the linear increase in simulation overhead typically associated with increased layer counts.

IV. Case Study II: Connector Interface

To demonstrate the versatility of the proposed framework across distinct interface standards we investigate the transfer of knowledge between two high speed connector footprints comprising Serial ATA (SATA) and Peripheral Component Interconnect Express (PCIe).

This case study presents a significantly more challenging scenario than the previous example due to substantial differences in material properties and stackup configurations alongside stringent signal integrity requirements which rigorously tests the framework adaptability.

The source topology (α) corresponds to a SATA connector interface implemented on a Rogers substrate. The design focuses on optimizing the transition from the connector landing pads to the breakout traces to minimize reflection. It is parameterized by five key geometric variables.

The target topology (β) is a PCIe Gen 5 connector interface designed on a Megatron6 high speed laminate. This structure introduces stricter impedance control requirements and different geometric constraints compared to SATA including a cutout region to manage capacitance.

TABLE IV: Design Parameters for PCIe. **Bolded entries** indicate topology-specific variations, whereas unbolded parameters represent transferable physical dimensions.

Parameter	Range
Landing pad displacement (Δ)	-0.1 to 0.1 mm
Landing pad width (w_L)	0.2 to 0.4 mm
Cutout size width (w_C)	1 to 1.4 mm
Breakout trace width (w_B)	0.11 to 0.31 mm
Megatron6 Substrate thickness (t_S)	0.6 to 0.8 mm

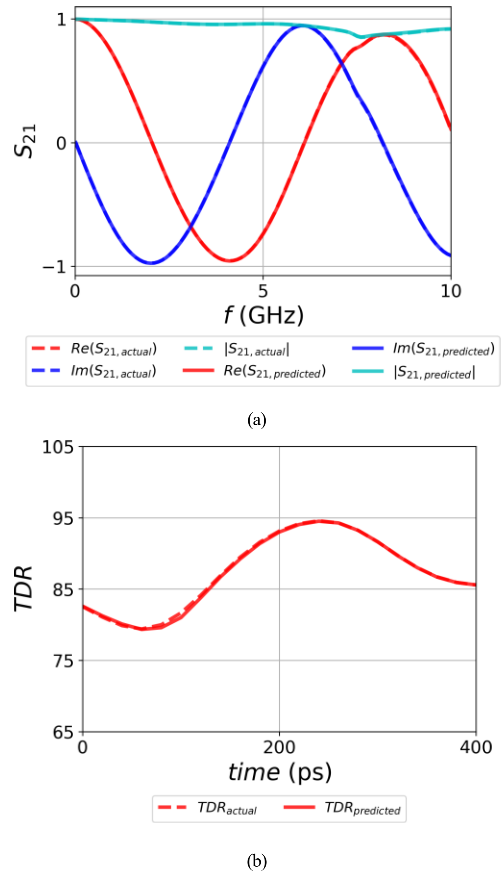


Fig. 7: Baseline performance of the source SATA model: Comparison of HFSS simulated (solid) and ML predicted (dashed) results for (a) S_{21} and (b) TDR impedance profile.

The design parameters for both topologies are listed in Table III and Table IV. Fig. 6 depicts the 3D electromagnetic models for both interfaces. It is pertinent to mention that specific internal structural details of these commercial connector models are encrypted to preserve vendor intellectual property. However this encryption does not affect the ultimate objective of this work as the port definitions remain accessible for generating the necessary electromagnetic response data via full wave simulation.

A. Data Generation

Consistent with the methodology detailed in Section III-A, ground truth data for both connector topologies was generated using the ANSYS HFSS 3D full-wave electromagnetic

solver [13]. Latin Hypercube Sampling (LHS) was similarly employed to ensure a statistically representative coverage of the design space defined in Tables III and IV. Given the increased structural complexity, these simulations were configured with a center frequency of 5 GHz, and each full-wave extraction required approximately 50 minutes of computation time.

For α, the SATA interface was simulated up to 10 GHz. A comprehensive dataset comprising 1580 samples was generated to train and test a robust backbone model capable of capturing the fundamental signal propagation characteristics of the connector footprint.

For β, the PCIe interface simulations were conducted over the same frequency range up to 10 GHz. A total of 1000 samples were generated to constitute the full training dataset (100%). To comprehensively evaluate the efficiency of the proposed method, we created data subsets representing 10%, 25%, 50%, and 100% of the total samples. Both the transfer learning framework and the baseline model (trained from scratch) were trained on these identical subsets to provide a fair comparison of convergence behavior and data efficiency across the entire spectrum of data availability. A separate dataset of 120 samples was generated solely for testing purposes to ensure rigorous validation of the adapted models.

B. Model Setup and Training

The machine learning framework was implemented in PyTorch [14] utilizing the identical hardware configuration described in Section III-B.

The training protocol strictly followed the two stage transfer learning strategy established in the previous case study. First the SATA source model architecture (f_α) was optimized using the BO framework to minimize validation loss. Second for the PCIe target model (f_β) the weights of this pretrained backbone were frozen. The design parameters of the PCIe interface were mapped to the shared feature space using the zero padding strategy to account for geometric differences. Finally only the lightweight adapter layers were optimized using the BO framework to align the predicted response with the target ground truth following the same optimization constraints and hyperparameters.

C. Results and Analysis

The performance of the proposed transfer learning (TL) framework was benchmarked against the baseline model (without pretraining) across varying training dataset sizes on the held-out test set.

The source model (f_α) achieved a baseline accuracy of 99.92% for the broadband frequency response (S_{21}) and 94.14% for the transient impedance characteristics (TDR) on its test set. This confirms the ability to capture the fundamental signal propagation physics across both domains. These high-fidelity predictions are illustrated in Fig. 7, where the predicted responses for the SATA interface show strong agreement with HFSS simulations.

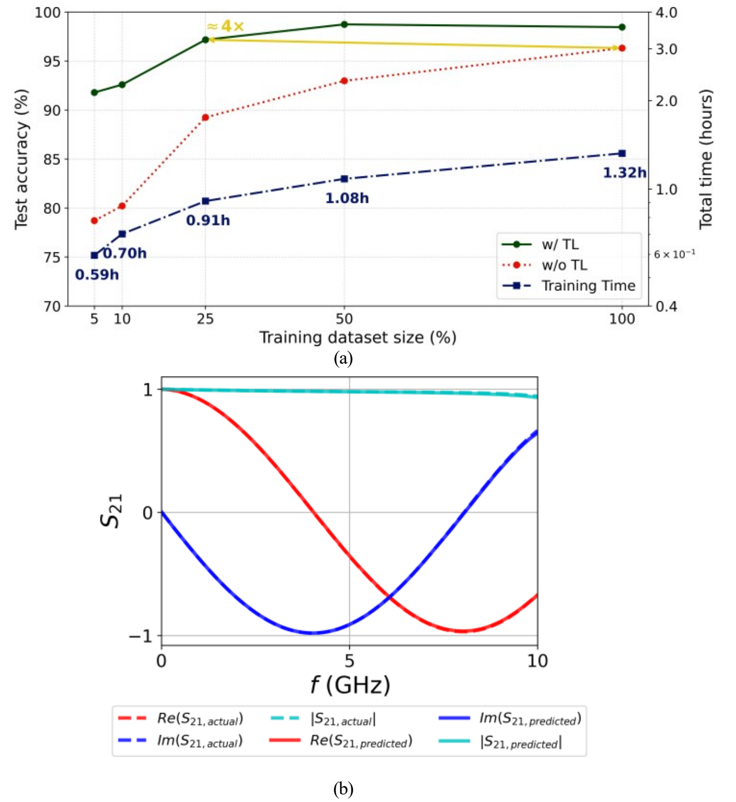


Fig. 8: Frequency Domain (S_{21}) Results for PCIe: (a) Test accuracy comparison vs. training size along with the total training time for the target model. The ≈ 4× annotation indicates a fourfold reduction in the training data required by the TL framework to reach the peak accuracy of the baseline model. (b) Comparison of HFSS simulated (solid) and TL predicted (dashed) S_{21} parameters, generated using the model trained on 25% of the dataset (250 samples).

1) Frequency Domain (S_{21}) Adaptation: When adapted to the target PCIe topology for frequency-domain prediction, the TL framework demonstrated superior initialization and rapid convergence, as shown in Fig. 8a. In the data-scarce regime (10% data), the TL model achieved over 95% accuracy, whereas the baseline model struggled to generalize. This performance gap highlights the effectiveness of the transferred spectral features. Fig. 8b presents the comparison between HFSS-simulated (solid) and ML-predicted (dashed) S_{21} parameters, generated using the model trained on only 25% of the dataset (250 samples). The results indicate excellent agreement up to 10 GHz, validating that the adapters successfully mapped the complex resonance dips and roll-off of the PCIe footprint.

2) Time Domain (TDR) Adaptation: Similarly, for the TDR task, the TL framework exhibited significant efficiency gains, as illustrated in Fig. 9a. The pre-trained backbone accelerated the learning of impedance discontinuities, reducing the training time by approximately 4× compared to the baseline. Fig. 9b shows the TDR prediction for the PCIe target, generated using the model trained on only 25% of the dataset (250 samples). The model accurately captures the time-domain responses at 5GHz, confirming that the framework effectively adapts

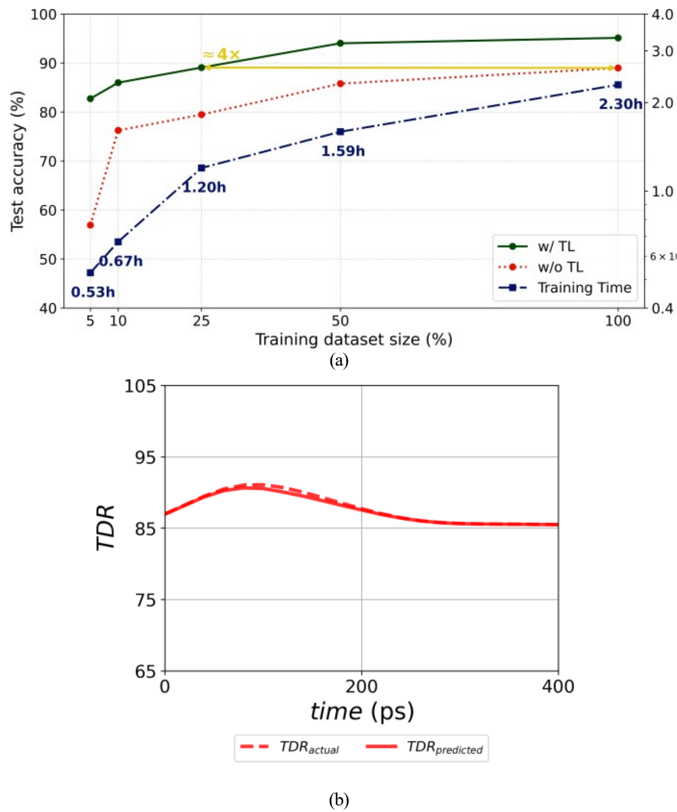


Fig. 9: Time Domain (TDR) Results for PCIe: (a) Test accuracy comparison vs. training size along with the total training time for the target model. The $\approx 4\times$ annotation indicates a fourfold reduction in the training data required by the TL framework to reach the peak accuracy of the baseline model. (b) Comparison of HFSS simulated (solid) and TL predicted (dashed) TDR impedance profiles at 5 GHz, generated using the model trained on 25% of the dataset (250 samples).

static geometric features to dynamic transient responses with minimal data.

V. Conclusion

This paper presented a novel Bayesian Optimization-guided transfer learning framework for the scalable surrogate modeling of complex 3D packaging structures. By decoupling topology-specific variations from shared physical features through a lightweight adapter architecture, the proposed method effectively overcomes the data scarcity bottleneck inherent in full-wave electromagnetic simulations. Experimental validation on vertical via transitions and high-speed connector interfaces demonstrated that the framework achieves superior generalization in data-scarce systems, delivering over 90% accuracy with only 10% of the training samples and providing a fourfold reduction in training time for time-domain tasks. These results establish the framework as a robust solution for accelerating design space exploration and enabling the creation of reusable, cross-topology surrogate libraries for next-generation signal integrity analysis.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. CNS 2137288, 2137283 and 2137259 – Center for Advanced Electronics through Machine Learning (CAEML) and its industry members. The authors thank Amphenol Corporation for providing the commercial connector models utilized in this study.

References

[1] S. Li, M.-S. Lin, W.-C. Chen, and C.-C. Tsai, "High-bandwidth chiplet interconnects for advanced packaging technologies in ai/ml applications: Challenges and solutions," *IEEE Open Journal of the Solid-State Circuits Society*, 2024.

[2] O. Akinwande, S. Erdogan, R. Kumar, and M. Swaminathan, "Surrogate modeling with complex-valued neural nets for signal integrity applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 72, no. 1, pp. 478–489, 2023.

[3] J. Konduru, O. Mikulchenko, L. Y. Foo, and J. E. Schutt-Aine´, "Signal integrity analysis and design optimization using neural networks," in *2024 IEEE 74th Electronic Components and Technology Conference (ECTC)*. IEEE, 2024, pp. 924–928.

[4] Y. Guo, Y. Wang, J. Corsello, and M. Swaminathan, "Deep kernel based hyperparameter adaptive learning and frequency response predictions using transposed convolutional neural network," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 2025.

[5] C. M. Schierholz, K. Scharff, and C. Schuster, "Evaluation of neural networks to predict target impedance violations of power delivery networks," in *2019 IEEE 28th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*. IEEE, 2019, pp. 1–3.

[6] T. Lu, J. Sun, K. Wu, and Z. Yang, "High-speed channel modeling with machine learning methods for signal integrity analysis," *IEEE Transactions on Electromagnetic Compatibility*, vol. 60, no. 6, pp. 1957–1964, 2018.

[7] S. Chen, J. Chen, T. Zhang, and S. Wei, "Semi-supervised learning based on hybrid neural network for the signal integrity analysis," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 10, pp. 1934–1938, 2019.

[8] M. Swaminathan, O. W. Bhatti, Y. Guo, E. Huang, and O. Akinwande, "Bayesian learning for uncertainty quantification, optimization, and inverse design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 11, pp. 4620–4634, 2022.

[9] H. M. Torun, A. C. Durgun, K. Aygu¨n, and M. Swaminathan, "Causal and passive parameterization of s-parameters using neural networks," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 10, pp. 4290–4304, 2020.

[10] O. Akinwande, X. Jia, A. Deroo, Y. Lu, H. Lin, B.-C. Tseng, and M. Swaminathan, "Transfer learning framework for 3-d electromagnetic applications," *IEEE Transactions on Microwave Theory and Techniques*, 2025.

[11] Y.-L. Sung, J. Cho, and M. Bansal, "Vl-adapter: Parameter-efficient transfer learning for vision-and-language tasks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5227–5237.

[12] T. Lei, J. Bai, S. Brahma, J. Ainslie, K. Lee, Y. Zhou, N. Du, V. Zhao, Y. Wu, B. Li *et al.*, "Conditional adapters: Parameter-efficient transfer learning with fast inference," *Advances in Neural Information Processing Systems*, vol. 36, pp. 8152–8172, 2023.

[13] Ansys Inc. (2022) Ansys HFSS. [Online]. Available: https://www.ansys.com/

[14] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.