ML-Assisted RF IC Design Enablement: the New Frontier of AI for EDA

Hyunsu Chae[1], Song Hang Chai[1], Taiyun Chi[2], Sensen Li[1], and David Z. Pan[1]
[1] ECE Department, The University of Texas at Austin, Austin, TX, USA
[2] ECE Department, Rice University, Houston, TX, USA
{hyunsu.chae, songhangc}@utexas.edu, taiyun.chi@rice.edu, {sensen.li@austin, dpan@ece}.utexas.edu

ABSTRACT

While AI for EDA has seen great success in digital IC design and some success in analog design, its potential for enabling RFIC design is yet to be fully explored. Due to its high-frequency nature, RFIC involves challenges such as parasitic effects, electromagnetic interference (EMI), signal integrity (SI), and other non-idealities. The modeling of passive networks and the associated computationally expensive EM simulations remain the major bottleneck in manual RFIC designs. This paper discusses the challenges and opportunities in ML-assisted RFIC design, covering topics from physics-augmented surrogate modeling to the inverse design of passive structures.

KEYWORDS

RFIC, EDA, Machine Learning

ACM Reference Format:
Hyunsu Chae[1], Song Hang Chai[1], Taiyun Chi[2], Sensen Li[1], and David Z. Pan[1]. 2025. ML-Assisted RF IC Design Enablement: the New Frontier of AI for EDA . In *30th Asia and South Pacific Design Automation Conference (ASPDAC '25), January 20–23, 2025, Tokyo, Japan.* ACM, Tokyo, Japan, 7 pages. https://doi.org/10.1145/3658617.3703637

1 INTRODUCTION AND BACKGROUND

Radio-frequency integrated circuits (RFICs) play an important role in numerous wireless systems, spanning from smartphones and Wi-Fi routers to satellite communication and radar systems. The evolution of RFIC design has traditionally leaned on intuition-based methods, which have demonstrated consistent success over time. These methods have produced a diverse collection of templates that are extensively utilized. However, conventional RFIC design methodologies are encountering limitations as the complexity of RF systems continues to grow, with increasing challenges in handling high-frequency signals with precision and reliability. The remarkable advancements in artificial intelligence (AI) and machine learning (ML) offer RFIC designers a new way to address these challenges. This breakthrough has the potential to liberate designers from the tedious cycle of iterative design and optimization, offering a more efficient path forward.

Conventional RFIC design involves many iterative steps (Figure 1, top) and requires careful balancing of trade-offs in specifications. Computationally intensive simulators are necessary for various design steps, such as transient simulations in circuit design, EM simulations in physical design, and Monte Carlo simulations in verification. These simulations require expert knowledge and are time- and resource-intensive, making RFIC design a complex process that demands significant expertise and effort to achieve optimal results.

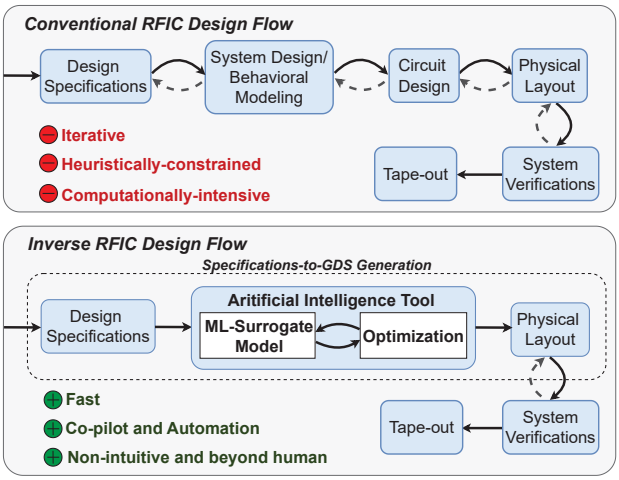


Figure 1: An illustration of the conventional design process compared to the ML-based inverse RFIC design framework.

RFIC designers must address various critical performance requirements while accounting for multiple electrical, magnetic, thermal, and physical constraints. Engineers often need several years of apprenticeship under experienced designers to gain proficiency. Automating repetitive tasks can accelerate the iterative design process, enabling engineers to explore a broader solution space with less manual intervention.

The fundamental elements of RFICs are active and passive components, each playing a distinct but complementary role in achieving the desired circuit performance [1–5]. Active components, such as transistors and the various active core circuits they form, are responsible for signal generation, amplification, frequency conversion, and other types of signal processing. These components are the heart of the signal chain, providing the essential power and gain. On the other hand, passive components–such as inductors, capacitors, transmission lines, and transformers–are crucial for tasks such as impedance matching, filtering, phase shifting, stabilization, and harmonic manipulation [6]. These components help define the frequency response of the RFIC and ensure the efficient operation of active stages. In RFICs, active and passive components do not function independently; their behavior and performance are closely interlinked due to the mutual dependence of their electrical characteristics and interactions [7–11]. These interactions influence key metrics such as gain, linearity, noise, efficiency, and bandwidth. Effective co-design of these components is crucial for optimizing performance. Together, active and passive elements form the backbone of RFICs, enabling compact, power-efficient, and high-performance RF communication systems.

The design of passive components is particularly challenging among RFIC design tasks. As essential building blocks, passives significantly impact chip area and performance. For instance, the passive component in an RF power amplifier (PA) typically occupies over 80% of the core chip area [3–5, 12, 13]. Passive design typically follows the selection of active components, with the design goals set according to

ASPDAC '25, January 20–23, 2025, Tokyo, Japan

ACM ISBN 979-8-4007-0635-6/25/01
https://doi.org/10.1145/3658617.3703637

the chosen actives. If the passive components fail to meet these goals, the active components may need to be reselected or adjusted accordingly. Poor passive design can result in inefficient power transfer, higher losses, or poor impedance matching, ultimately limiting the performance of active components. The design process of passives begins with selecting a well-defined topology, such as transformers or coupled transmission lines. Designers translate parameters such as inductance, coupling factor, and capacitance from behavioral models into physical structures constructed across multiple metal layers. The design is then optimized by iteratively adjusting the geometry to address discrepancies and meet specifications. This optimization process is limited by the chosen topology, with only a few adjustable parameters. As a result, the current design flow struggles to explore a larger design space and fully utilize the available area, constrained by the decisions made early in the process.

The necessity of AI/ML-driven tools in RFIC design is increasingly evident as these technologies could significantly enhance design productivity and lower entry barriers for new designers. Successes in various fields highlight the effectiveness of AI applications and suggest promising new directions for RFIC design. The use of ML techniques efficiently accelerates the entire design flow for Very Large Scale Integration (VLSI) [14–16] and Field-Programmable Gate Array (FPGA) [17]. Analog devices, which share significant similarities with RFIC design, have also seen substantial advancements through ML. The EDA community has successfully applied ML to tackle challenges in analog devices [18], including sizing problems [19, 20], topology selection [21–23] and layout generation [24–27]. Understanding the unique aspects of RFIC design allows us to effectively adapt ML methods from other areas. An ML-based inverse RFIC design framework, shown in Figure 1 (bottom), directly synthesizes the physical layout from target performance specifications. AI/ML models quickly predict circuit behavior and replace EM simulation for both passive and active components. Optimization methods enable a more comprehensive search of the design space. This ML-assisted framework can streamline the design process through automation, reduce the learning curve of RF design, and foster innovation.

The rest of this paper is organized as follows. In Section 2, we introduce various ML surrogate modeling techniques to replace conventional, time-consuming circuit simulations, with a focus on passive EM simulations. Section 3 presents advanced techniques for the inverse design of RF passive networks. Conclusions are provided in Section 4.

2 ML SURROGATE MODELING

Computationally expensive EM simulation is the major bottleneck in RF manual design. Conventional EM solver, such as Ansys High-Frequency Structure Simulator (HFSS) [28], employs numerical methods, including finite element methods (FEM) and finite difference time difference method (FDTD) to solve Maxwell's equations. HFSS constructs a 3D mesh of the entire simulation domain and divides it into small elements to approximate the field distribution. This approximation is achieved by solving matrix equations. A finer mesh is needed for a design with more complex geometry or higher frequency. This requirement substantially results in increased computation time. Additionally, HFSS iteratively refines the mesh based on field gradients until convergence criteria are met. As the complexity of the design grows, so do the memory and computation demands due to the iteration required for convergence. This highlights the need for an ML surrogate model that can serve as a fast EM solver while providing accurate predictions.

Table 1: Summary of Different ML Surrogate Models Proposed in RF Applications

Base ML Models	Input Data Type	Output Data	Applications
ANN	tabular	S-parameters	High-speed links [29], multilayer interconnects [30], three-pole H-plane filter [31]
TCNN	tabular	$L(f)$, $ESR(f)$	Integrated voltage regulator [32]
		S-parameters	Package interconnect structures [33]
GNN	graph	S-parameters	Resonator filter [34], chiplet interconnect [35]
GT	graph	S-parameters	PCB interconnects [36]
CNN	image	S-parameters	Planar microwave circuit [37], bandpass filter [38], impedance matching network [39]
UNet	image	S-parameters	Impedance transformation [40]

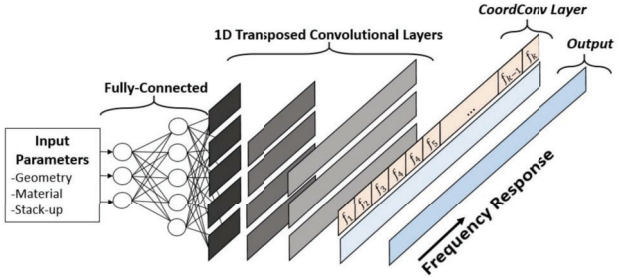


Figure 2: Architecture of S-TCNN for mapping input design parameters to predict frequency response [32].

In this section, we examine the current trend in ML surrogate modeling for RF passive design. Table 1 summarizes several recent works, each proposing different representations of RF passive designs and ML model architectures. The work [29–33] chooses tabular approach, while works [34–36] adopt a graph representation and works [37–40] use an image-based representation. All these methods aim to predict the frequency response of passive networks such as S-parameters. While the table provides a comprehensive overview, only a selection of these works will be discussed in detail in the following sections.

2.1 Tabular-based ML Modeling

The passive network can be represented as a tabular dataset, capturing its typical topological structure [29–32]. With a defined set of templates, each design is characterized by a unique set of parameters, such as geometry parameters and material properties.

Configuring the ML model input as a tabular dataset presents several challenges. The fully-connected (FC) layers are the most common choice for tabular inputs. However, they may struggle with capturing spatial hierarchies or local patterns in the data. The frequency response is smooth and continuous along the frequency axis. This characteristic must be captured by the ML model. 1D convolutional layers offer an alternative, as they incorporate the concept of spatial relationships. However, they focus on extracting spatial correlations in the input while we aim to predict them in the output. Additionally, convolutional layers are typically used for dimensionality reduction, whereas the goal is to expand the dimensionality. In RF applications, the frequency response output can expand significantly compared to the input dimension. This is especially true when observing behavior across a wide frequency range in multi-port networks.

Torun et al. [32] proposes a spectral transposed convolutional neural network (S-TCNN) to predict the frequency response of embedded inductors in integrated voltage regulators (IVR). Figure 2 shows the proposed S-TCNN architecture. The S-TCNN utilizes 1D

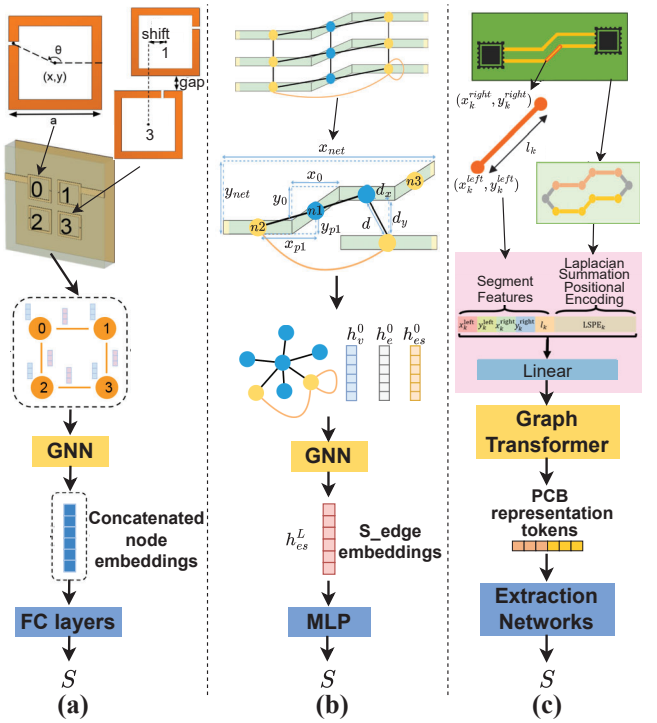


Figure 3: Graph representation and prediction flow for each selected work. (a) Circuit-GNN [34]. (b) GNN-SP [35]. (c) TraceFormer [36].

transposed convolutional layers and *CoordConv* layer [41]. The paper illustrates the intuition behind the transposed convolutional layer by deriving it from the standard 1D convolutional layer. A standard 1D convolutional layer outputs $\mathbf{y}$ from combining neighboring values of the input $\mathbf{x}$, weighted by a kernel $\mathbf{h}$ as shown in Eq. (1).

$$\mathbf{y} = f(\mathbf{h} * \mathbf{x}) = f(H\mathrm{x}), \tag{1}$$

where $*$ denotes the cross-correlation operation and $f(\cdot)$ denotes the non-linear activation function. This function allows to capture the spatial correlation of the $\mathbf{x}$. In the RF design problem, we aim to achieve spatial correlation in the output rather than the input. Thus, we can reverse the operation to obtain $\mathbf{x}$ as the output. Eq. (2) presents the reversed operation called transposed convolutional operation.

$$\hat{\mathbf{x}} = f(\hat{\mathbf{h}} *^{\mathrm{T}} \hat{\mathbf{y}}) = f(\hat{H}^{\mathrm{T}}\hat{\mathbf{y}}), \tag{2}$$

where $\hat{\mathbf{x}}$ is the final extracted frequency response when the input design parameters $\hat{\mathbf{y}}$ is given. Moreover, a *CoordConv* layer [41] is added as the last layer of the network. This layer serves to distinguish the location of two similar patterns in the frequency spectrum. This is done by hard-coding the location information to the output by concatenating a coordinate axis to its input.

2.2 Graph-based ML Modeling

The graph representation has been applied to various areas of circuit design, such as netlist-based digital circuit design, where nodes and edges can be mapped to logic gates and wires. A similar approach has been attempted in the RF applications, including distributed circuit design [34] and PCB interconnect design [35, 36]. They segment the metal trace structures in the design to form a graph.

Circuit-GNN [34] represents the N-resonator filter as a graph. Figure 3 (a) illustrates the overall flow of the model. In the graph, each node represents a square resonator, and each edge captures the electromagnetic interaction between resonators. Four design parameters define each resonator: center coordinates (x, y), side length a, and slit's angular position θ. The node feature is $[a, \theta]$, while edge feature between node i and j is $[\theta_i, \theta_j, x_i - x_j, y_i - y_j, g_{ij}, s_{ij}]$. The g_{ij} and s_{ij} represent the length of gap and shift between the two structures, respectively. An edge exists if two nodes are sufficiently close to interact electromagnetically. This graph is fed into the GNN to extract the node and edge embeddings. The node embeddings for the resonators connected to the input and output ports are concatenated to form a fixed-length global graph feature. This feature is then passed through multiple FC layers in the predictor to extract the S-parameters.

GNN-SP [35] employs a GNN to calculate the S-parameters in universal chiplet interconnect express (UCIe). The interconnects are segmented at points where they change direction. As shown in Figure 3 (b), each segment is represented as a node, categorized as either a port node (yellow) or a normal node (blue). There are two types of edges: normal edges and S_edges. S_edges are created to incorporate the ultimate goal of port-to-port S-parameter prediction. Therefore, it is generated in two cases: 1) between port nodes to account for insertion loss and crosstalk, and 2) a self-loop on a port node to account for return loss. Unlike traditional GNN models, GNN-SP embeds global information—such as the entire net length—into the node features to capture global coupling effects. Edge features are defined by the neighboring relationship of the two segments, such as the distances between them. GNN-SP performs edge-level regression, where S_edges captures essential global net information as it is connected between port nodes. This S_edges embedding is then passed into the multi-layer perceptron (MLP) decoding layer to generate S-parameter output.

TraceFormer [36] uses graph transformer-based model [42] to predict the S-parameter of PCB traces. As illustrated in Figure 3 (c), each PCB trace is converted into a graph by spitting the trace into segments. The branching points along each trace become nodes, and the segments between them form the graph's edges. To model all traces as one graph, dummy nodes are attached to the end of nodes on both the driver and receiver side. The graph information is then translated into tokens. Accurate S-parameter prediction requires processing the entire set of traces together, as they mutually influence each other. Therefore, each segment's position and graph topology are tokenized through a linear network. These tokens are then passed into the graph transformer, which consists of transformer encoder layers to extract PCB representations. Finally, four extraction networks predict each type of complex-valued S-parameters across the frequency range.

2.3 Image-based ML Modeling

The RF passive network can be mapped to layers of pixelated images [37–40]. This approach enables the application of various image-processing techniques in the AI field. Moreover, treating the layouts as images allows for leveraging data augmentation strategies such as rotation, scaling, and flipping. Data augmentation can help mitigate the limitations posed by small datasets in this domain.

The framework proposed by [39] utilizes Deep-CNN to predict S-parameters of the single-layer matching network. The Deep-CNN consists of twelve 2D convolutional layers followed by five FC layers, as shown in Figure 4 (a). The CNN serves as the base model, as it can capture the coupling of the neighboring pixels from the image input. The input to the model is an 18×18 matrix representing the EM structure. The output predicts the 2-port complex S-parameter across frequencies between 30 and 100 GHz, discretized into nine frequency points. The output size is, therefore, 54, as it only requires

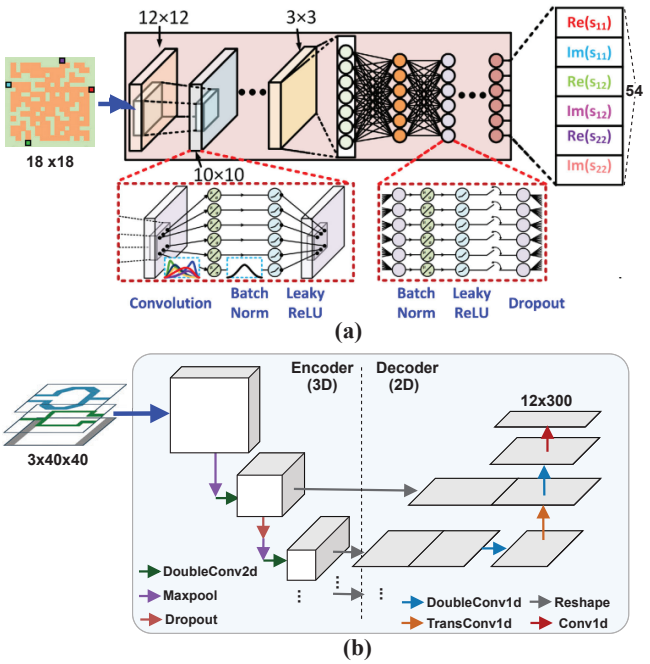


Figure 4: Image-based ML surrogate models. (a) CNN-based [39]. (b) UNet-based [40].

six S-parameter terms (real and imaginary components of S_{11}, S_{21}, and S_{22}) for each frequency point. The FC layers are used after CNN to transform 2D data into a 1D form for output.

PulseRF [40] framework utilizes UNet [43] as the backbone of the ML-surrogate model as illustrated in Figure 4 (b). The model predicts the S-parameters of a 4-port, multi-layer passive network. UNet is chosen primarily for its strengths in image-to-image transformation. The model's input space is 4800, for three layers of 40×40 images. The output size is 3600, representing 12 unique real and imaginary components of 4-port S-parameters for 300 frequency points (1-300GHz in 1GHz step). The similar sizes of input and output allow UNet to be effectively applied. However, since the input is 3D and the output is 2D, the UNet's decoder path is modified to produce a flattened output. Moreover, UNet has demonstrated its effectiveness in image segmentation tasks by effectively capturing local and contextual features. The underlying intuition of this work is to approach the RF passive prediction task as an image segmentation problem. In RF passive structures, specific signal path shapes generate particular EM fields. Accurately capturing and interpreting these local and global features is critical for the model's performance.

2.4 Challenges and Future Directions

While ML surrogate models are effective for accelerating the simulation process, they are fundamentally data-driven methods. Their performance relies heavily on the quality and quantity of the dataset used. Currently, there are no publicly available open-source datasets for RFIC design due to the non-disclosure agreement (NDA) constraint imposed by semiconductor foundries, especially in advanced technology nodes. As a result, each study generates its training dataset manually, relying on costly EM simulations that often take days or weeks to complete.

Researchers in different areas have proposed various methods to reduce the dependency on datasets. One approach is to integrate physical knowledge with ML [44–47]. By combining empirical data with established physical principles, they aim to improve performance on tasks involving physical laws and mathematical equations.

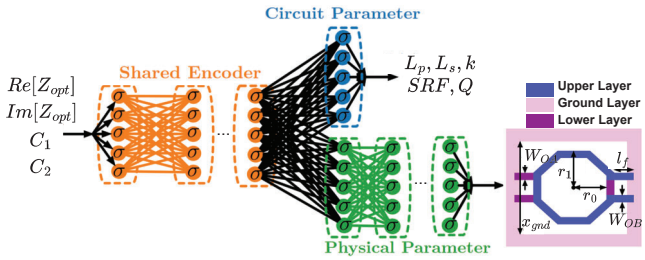


Figure 5: Design synthesis utilizing SE-NN [53].

Several strategies can be employed to achieve this integration. This includes embedding physical equations into the architecture of ML models and incorporating physical constraints into the training loss function.

Several studies have applied physics principles in the RF domain for surrogate modeling to enhance model accuracy. The work [33] introduces two key components: the causality enforcement layer (CEL) and the passivity enforcement layer (PEL). The CEL embeds the causal relationship between the real and imaginary part of s-parameters, which is derived as Kramers-Kronig (K-K) relations [48]. The PEL ensures the passivity of the RF passive structure, which means that the passive networks do not generate energy. Additionally, the work [40] introduces the K-K compensation layer (KKCL). This layer compensates for discrepancies that happen in high frequency between the imaginary components generated by the CEL and those from simulations.

Integrating physical principles into ML modeling presents a promising direction for RFIC design. Embedding CEL, PEL, and KKCL enhances the accuracy and reliability of surrogate models. Further refinement of these methods brings us closer to more efficient, less data-dependent ML models for RF design. Recent work has also demonstrated promising success in fast direct and iterative solvers with optimal (linear) complexity [49–51]. Moreover, the use of transfer learning [52] has shown potential for technology-agnostic applications. Combining these fast solver technologies with advanced data science and ML techniques could enable rapid, large-scale EM modeling and analysis tools. Additionally, the development of open-source datasets remains crucial for progress in this field.

3 ML-ASSISTED INVERSE DESIGN

In this section, we review recent studies on RF passive design automation. All approaches aim to generate a layout that meets the target RF performance. Er et al. [53] directly synthesize the design using an ML model, while Lee et al. [54], Karahan et al. [39], and Chae et al. [40] frame the task as an optimization problem. The optimization process requires frequent performance evaluations of different design parameters. Lee et al. [54] use EM simulators to assess the design, whereas Karahan et al. [39] and Chae et al. [40] utilize fast ML-based models for evaluation. All these methods differ in terms of design objectives, design space formulation, and optimization algorithms.

3.1 ML Synthesis Model-based Design

The work [53] designs an impedance-matching network using an ML synthesis model named SE-NN. While the ML models presented in Section 2 act as surrogates for EM simulations, this work's ML model performs the inverse of the EM simulation.

The proposed SE-NN model consists of a fully connected neural network with an embedded shared-encoder (SE). The overall model architecture is shown in Figure 5. Given the performance metrics such as input impedance Z_{opt} and load capacitors C_1 and C_2, the

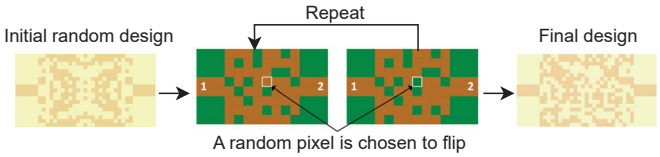


Figure 6: Random pixel flipping by PixelatedRF [54].

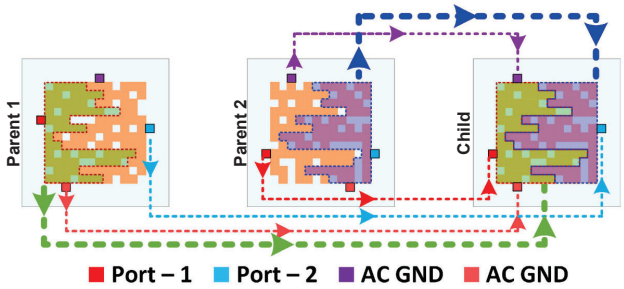


Figure 7: Example of population update in genetic algorithm-based approach [39].

model predicts six design parameters. These design parameters correspond to the physical dimensions that form a transformer topology-based design across three metal layers. Instead of solely relying on the model to learn the mapping between physical and performance parameters, the authors improve the process by incorporating the circuit parameters. The model is also trained to predict the circuit parameters: inductance of primary and secondary coil (L_p and L_s), coupling factor (k), self-resonant frequency (SRF), and quality factor (Q). These parameters are derived from the performance parameters, allowing SE to embed more circuit laws into the model weights. In this context, the circuit parameters serve as a bridge to link performance parameters and physical parameters, as they dictate the physical and circuit principles underlying the performance parameters.

This work demonstrates that the SE-NN model outperforms other ML architecture, including gradient boosting, linear regression, and naive-NN approach without SE. The model quickly identifies designs that meet target specifications without the need for a dedicated optimization algorithm to explore the design space. However, the approach has limitations: it only evaluates performance at a single frequency (30 GHz), and the design space is relatively small, mirroring conventional manual design constraints.

3.2 Random Search-based Optimization

PixelatedRF framework [54] proposes a direct binary search (DBS) method to design a passive filter.

In this framework, the design space is represented as a pixelated planar surface, where each pixel is a uniform unit of the top metal conductor. The layout is divided into a 25×19 grid of pixels, where each pixel can be filled with metal (conductor) or left unfilled (dielectric). The design flow begins by generating a random pixel map sample. An EM simulation is then performed on this initial layout to obtain S-parameters. The S-parameters are then translated to a figure of merit (FoM) metric based on the design specifications. In each iteration, five pixels are randomly selected to be flipped—either from metal to dielectric or vice versa, as shown in Figure 6. After that, a new EM simulation is conducted, and the updated FoM is evaluated. If the FoM improves, the modification is kept; otherwise, the pixel is reverted. This iterative process continues until the iteration budget is reached.

PixelatedRF framework successfully designs a dual-band filter. The prototyped and fabricated printed circuit board (PCB) design shows good alignment between measured and simulated results. However, the approach has limitations due to its reliance on a random-based optimization method. The authors report that this optimization method takes a few hundred to a few thousand iterations to converge. Each time there is a new design goal, the process requires repeating this intensive EM simulation. Additionally, the optimization's convergence highly depends on the initial random design. Starting with a suboptimal initial pixel layout may lead to slower convergence or cause the algorithm to become trapped in local optima, thereby reducing the effectiveness of the search process.

3.3 Genetic Algorithm-based Optimization

In this work [39], the authors proposed to synthesize an output matching network (OMN) based on the target specifications of a PA. In the inverse design flow, genetic algorithm-based optimization is employed.

The optimization design space is formulated as a pixelated map as shown in Figure 4 (a). The single metal layer is discretized into 16×16 pixels where each pixel can be filled with metal (1) or no metal (0), similar to the design space in PixelateRF [54]. Additionally, this work introduces flexible port locations. The ports can be placed anywhere along the boundary of the structure, providing greater flexibility in the design space with 16^4 possible pin location combinations.

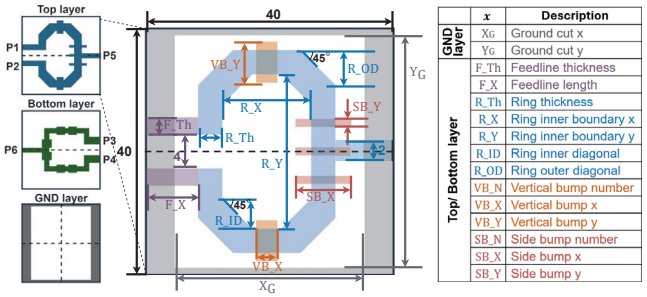
The authors utilized a genetic algorithm in this work to take care of the large searchable design space ($> 2^{16\times16}$). Genetic algorithms solve an optimization task by mimicking natural selection and evolving the solution over generations through selection, crossover, and mutation. Figure 7 illustrates the process of population update. Among the current population, the two best-performing parents are selected. Rows of the child matrix are reconstructed from each parent through the random cut line. The port locations are directly inherited from either parent. After this step, random mutation is applied by flipping metal and non-metal locations to enhance design diversity. Each design is evaluated through the cost metric that takes into account impedance mismatch and insertion loss. Following this optimization flow, the authors claim that the inverse-designed OMN meets performance specifications when paired with the PA's active core. It also achieves a broader matching bandwidth than conventional manually designed matching networks.

3.4 Bayesian Optimization

The framework PulseRF [40] is proposed to synthesize a generic multi-port and multi-layer passive layout. PulseRF maps the inverse design task of low-loss impedance transformation to a Bayesian optimization method.

PulseRF framework suggests a design space formulation that balances between the traditional template-based design and pixelated design. The description of physical design parameters is illustrated in Figure 8. Following the traditional transformer-topology-based design, the structure includes three layers: a ground layer and two metal ring layers. The proposed design space enables various ring shapes beyond the traditional symmetrical octagon shape. The shape of the ring structure's outer and inner boundary can be individually chosen from octagon, rectangle, hexagon, and rhombus. Moreover, they introduce localized deformations for random width variations in the ring. This enhanced design space offers more meaningful design candidates, building upon the traditional design space.

Bayesian optimization-based design synthesis is implemented to efficiently explore and optimize across a large design space ($>10^{16}$). Bayesian optimization selects new design samples through probabilistic decision-making. Due to its reduced dependency on the quantity



	x	Description
GND layer	X_G	Ground cut x
	Y_G	Ground cut y
Top/ Bottom layer	F_Th	Feedline thickness
	F_X	Feedline length
	R_Th	Ring thickness
	R_X	Ring inner boundary x
	R_Y	Ring inner boundary y
	R_ID	Ring inner diagonal
	R_OD	Ring outer diagonal
	VB_N	Vertical bump number
	VB_X	Vertical bump x
	VB_Y	Vertical bump y
	SB_N	Side bump number
	SB_X	Side bump x
	SB_Y	Side bump y

Figure 8: PulseRF design space [40].

		Manual	PulseRF
Design Target		Zp = Zn = 100 + j0 @ 140GHz	
Results	Design Time	a few hours	6 mins
	Layout		
	Synthesized Results	Zp=Zn=5.60+j35.92 $\eta = 0.65$	Zp=Zn=87.72+j0.20 $\eta = 0.89$

Figure 9: An example of PulseRF design and comparison with expert manual design [40].

and quality of the initial design samples, it is suitable for this expansive design space. This optimization framework targets ring structure parameters, treating deformation as noise to evaluate if slight width changes improve the design. The highlighted result of PulseRF is shown in Figure 9 for a common real-to-real impedance transformation. In manual design, no solution exists within the defined structures for the predefined transformer behavior model. The transformers inherently contain inductive components, thus requiring additional matching elements (e.g., capacitors) to facilitate real-to-real impedance transformation. The non-intuitive design generated by PulseRF can achieve a design close to the target. PulseRF presents results that bring speed up and achieve better performance by leveraging non-intuitive and incomprehensive structures.

3.5 Challenges and Future Directions

The works discussed in this section vary significantly in their final passive network designs. Er et al. [53] constrain the design space similarly to conventional manual design, resulting in a passive network that resembles a manually designed one. The design space formulation may be considered conservative, limiting the range of achievable layouts and restricting design flexibility. However, this approach provides a level of reliability, as it is based on a proven formulation that helps ensure performance remains close to design specifications. On the other end of the spectrum, Lee et al. [54] and Karahan et al. [39] explore the available design space as a pixelation of the entire available area. This approach results in a final design that differs significantly from conventional designs, offering a high degree of randomness. However, a completely pixelated implementation lacks explainability, making further fine-tuning and optimization nearly impossible—especially considering that AI-synthesized designs do not yet outperform expert designs at the current stage [1, 39, 55]. Furthermore, extending the work for multiple layers will significantly expand the design space, increasing the complexity of optimization exponentially.

Finding the balance between exploitation and exploration of the design space remains a challenge. Chae et al. [40] attempts to strike this balance. The design space is based on a template-driven passive network with introduced abnormalities. A template-based fundamental structure provides the designer with intuition and insight into the passive network's performance. For instance, the coupling between metal layers can be estimated by the overlap area of the two coil structures. The introduced abnormalities expand the design space, facilitating the search for optimal designs. Moving forward, incorporating RFIC domain knowledge into design space formulation is a promising direction, as it ensures a design space with meaningful performance outcomes. Additionally, introducing non-intuitive features could further enhance this approach, providing greater flexibility in the design process and exploring a bigger design space.

It is also essential to recognize that RFIC design is fundamentally a co-optimization of both active and passive components. Previous works have primarily focused on passive elements, as they are often the most time-consuming and challenging aspect of RF design. Typically, these works generate targets for passive components based on predetermined and fixed active components. However, the true performance of an RFIC can only be assessed through active-passive co-simulation. Integrating active components into the design process, i.e., an end-to-end AI-assisted design automation flow, is crucial to fully capture the complexity of RF design [56]. Adopting a co-optimization approach will not only provide a more holistic view but also unlock new possibilities for future RFIC design.

4 CONCLUSION

In this paper, we explored ML-based approaches for facilitating RFIC design automation, with a focus on passive networks. We first introduced conventional RFIC design, its challenges, and the need for ML models in RFIC design. We then delve into the discussion on the implementation of ML surrogate models, highlighting their potential to serve as a viable alternative to traditional EM simulation in RFIC design. Subsequently, we reviewed state-of-the-art inverse design examples for direct synthesis of RF passive networks based on target specifications. Potential future research directions include further incorporating physics into ML models to reduce dependency on expensive datasets. Additionally, integrating both active and passive elements into the automated design process could enable holistic co-optimization of RFIC performance.

ACKNOWLEDGMENTS

This work is supported in part by NSF CCF-2112665, MediaTek, and an equipment donation from Nvidia.

REFERENCES

[1] Tzu-Yuan Huang, Naga Sasikanth Mannem, Sensen Li, Doohwan Jung, Min-Yu Huang, and Hua Wang. A coupler balun load-modulated power amplifier with extremely wide bandwidth. *IEEE Transactions on Microwave Theory and Techniques*, 71(4):1573–1586, 2022.

[2] Sensen Li, Tzu-Yuan Huang, Yuqi Liu, Hyunjin Yoo, Yoosam Na, Youngsik Hur, and Hua Wang. A millimeter-wave LNA in 45nm CMOS SOI with over 23dB peak gain and sub-3dB NF for different 5G operating bands and improved dynamic range. In *IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pages 31–34, 2021.

[3] Xiaohan Zhang, Sensen Li, Daquan Huang, and Taiyun Chi. A millimeter-wave three-way doherty power amplifier for 5G NR OFDM. *IEEE Journal Solid-State Circuits (JSSC)*, 58(5):1256–1270, 2023.

[4] Xiaohan Zhang, Sensen Li, and Taiyun Chi. A millimeter-wave watt-level doherty power amplifier in silicon. *IEEE Transactions on Microwave Theory and Techniques*, 72(3):1674–1686, 2024.

[5] Tso-Wei Li, Sensen Li, Hossein Miri Lavasani, and Hua Wang. A highly linear and efficient v -band CMOS power amplifier using hybrid NMOS/PMOs for double nonlinearity cancellation with four-way distributed-active-transformer. *IEEE Journal Solid-State Circuits (JSSC)*, pages 1–14, 2024.

[6] Taiyun Chi, Yaolong Hu, Xiaohan Zhang, and Hao Guo. Pushing the performance boundaries of mmwave and sub-THz transceiver circuits through passive network design innovations. In *IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 759–763, 2024.

[7] Sensen Li, Min-Yu Huang, Doohwan Jung, Tzu-Yuan Huang, and Hua Wang. A mm-wave current-mode inverse outphasing transmitter front-end: A circuit duality of conventional voltage-mode outphasing. *IEEE Journal Solid-State Circuits (JSSC)*, 56(6):1732–1744, 2020.

[8] Sensen Li, Taiyun Chi, and Hua Wang. Multi-feed antenna and electronics co-design: An E-band antenna-LNA front end with on-antenna noise-canceling and G-boosting. *IEEE Journal Solid-State Circuits (JSSC)*, 55(12):3362–3375, 2020.

[9] Sensen Li, Taiyun Chi, Tzu-Yuan Huang, Min-Yu Huang, Doohwan Jung, and Hua Wang. A buffer-less wideband frequency doubler in 45-nm CMOS-SOI with transistor multiport waveform shaping achieving 25% drain efficiency and 46–89 GHz instantaneous bandwidth. *IEEE Solid-State Circuits Letters*, 2(4):25–28, 2019.

[10] Sensen Li, Taiyun Chi, Jong-Seok Park, Huy Thong Nguyen, and Hua Wang. A 28-GHz flip-chip packaged chireix transmitter with on-antenna outphasing active load modulation. *IEEE Journal Solid-State Circuits (JSSC)*, 54(5):1243–1253, 2019.

[11] Sensen Li, Taiyun Chi, Jong Seok Park, Wasif Tanveer Khan, Hua Wang, and John Papapolymerou. A fully packaged D-band MIMO transmitter using high-density flip-chip interconnects on LCP substrate. In *IEEE MTT-S International Microwave Symposium (IMS)*, pages 1–4. IEEE, 2016.

[12] Hua Wang et al. Power amplifiers performance survey 2000-present, 2023.

[13] Sensen Li, Min-Yu Huang, Doohwan Jung, Tzu-Yuan Huang, and Hua Wang. 24.3 a 28GHz current-mode inverse-outphasing transmitter achieving 40%/31% PA efficiency at Psat/6dB PBO and supporting 15Gbit/s 64-QAM for 5G communication. In *IEEE International Solid-State Circuits Conference (ISSCC)*, pages 366–368, 2020.

[14] Jinwook Jung, Andrew B. Kahng, Seungwon Kim, and Ravi Varadarajan. METRICS2.1 and flow tuning in the IEEE CEDA robust design flow and OpenROAD. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2021.

[15] Matthew M. Ziegler, Jihye Kwon, Hung-Yi Liu, and Luca P. Carloni. Online and offline machine learning for industrial design flow tuning. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2021.

[16] Zhiyao Xie, Guan-Qi Fang, Yu-Hung Huang, Haoxing Ren, Yanqing Zhang, Brucek Khailany, Shao-Yun Fang, Jiang Hu, Yiran Chen, and Erick Carvajal Barboza. FIST: A feature-importance sampling and tree-based method for automatic design flow parameter tuning. In *IEEE/ACM Asia and South Pacific Design Automation Conference (ASPDAC)*, 2020.

[17] Chang Xu, Gai Liu, Ritchie Zhao, Stephen Yang, Guojie Luo, and Zhiru Zhang. A parallel bandit-based approach for autotuning FPGA compilation. In *ACM Symposium on FPGAs*, 2017.

[18] Supriyo Maji, Ahmet F. Budak, Souradip Poddar, and David Z. Pan. Toward end-to-end analog design automation with ML and data-driven approaches. In *IEEE/ACM Asia and South Pacific Design Automation Conference (ASPDAC)*, pages 657–664, 2024.

[19] Hanrui Wang, Kuan Wang, Jiacheng Yang, Linxiao Shen, Nan Sun, Hae-Seung Lee, and Song Han. GCN-RL circuit designer: Transferable transistor sizing with graph neural networks and reinforcement learning. In *ACM/IEEE Design Automation Conference (DAC)*, 2020.

[20] Ahmet F. Budak, Shuhan Zhang, Mingjie Liu, Wei Shi, Keren Zhu, and David Z. Pan. Machine learning for analog circuit sizing. In Haoxing Ren and Jiang Hu, editors, *Machine Learning for Analog Circuit Sizing*, pages 307–335. Springer, 2022.

[21] Zehao Dong, Weidong Cao, Muhan Zhang, Dacheng Tao, Yixin Chen, and Xuan Zhang. CktGNN: Circuit graph neural network for electronic design automation. In *International Conference on Learning Representations (ICLR)*, 2023.

[22] Morteza Fayazi, Morteza Tavakoli Taba, Ehsan Afshari, and Ronald Dreslinski. Angel: Fully-automated analog circuit generator using a neural network assisted semi-supervised learning approach. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 70(11):4516–4529, 2023.

[23] Souradip Poddar, Ahmet Budak, Linran Zhao, Chen-Hao Hsu, Supriyo Maji, Keren Zhu, Yaoyao Jia, and David Z. Pan. A data-driven analog circuit synthesizer with automatic topology selection and sizing. In *IEEE/ACM Proceedings Design, Automation and Test in Eurpoe (DATE)*, pages 1–6, 2024.

[24] Hao Chen, Mingjie Liu, Xiyuan Tang, Keren Zhu, Abhishek Mukherjee, Nan Sun, and David Z. Pan. MAGICAL 1.0: An open-source fully-automated AMS layout synthesis framework verified with a 40-nm 1GS/s ΔΣ ADC. In *IEEE Custom Integrated Circuits Conference (CICC)*, pages 1–2, 2021.

[25] Keren Zhu, Hao Chen, Mingjie Liu, and David Z. Pan. Tutorial and perspectives on MAGICAL: A silicon-proven open-source analog IC layout system. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 70(2):715–720, 2023.

[26] Tonmoy Dhar, Kishor Kunal, Yaguang Li, Meghna Madhusudan, Jitesh Poojary, Arvind K. Sharma, Wenbin Xu, Steven M. Burns, Ramesh Harjani, Jiang Hu, Desmond A. Kirkpatrick, Parijat Mukherjee, Soner Yaldiz, and Sachin S. Sapatnekar. ALIGN: A system for automating analog layout. *IEEE Design & Test*, 38(2):8–18, 2021.

[27] Kishor Kunal, Meghna Madhusudan, Arvind K. Sharma, Wenbin Xu, Steven M. Burns, Ramesh Harjani, Jiang Hu, Desmond A. Kirkpatrick, and Sachin S. Sapatnekar. Invited: ALIGN – open-source analog layout automation from the ground up. In *ACM/IEEE Design Automation Conference (DAC)*, pages 1–4, 2019.

[28] AnsysHFSS. Ansys HFSS ver 2023.1.

[29] Katharina Scharff, Christian Morten Schierholz, Cheng Yang, and Christian Schuster. ANN performance for the prediction of high-speed digital interconnects over multiple PCBs. In *IEEE Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, pages 1–3. IEEE, 2020.

[30] Allan Sánchez-Masís, Renato Rimolo-Donadio, Kallol Roy, Modar Sulaiman, and Christian Schuster. FNNs models for regression of s-parameters in multilayer interconnects with different electrical lengths. In *IEEE MTT-S Latin America Microwave Conference (LAMC)*, pages 82–85. IEEE, 2023.

[31] Jianan Zhang, Jing Chen, Qianyi Guo, Wei Liu, Feng Feng, and Qi-Jun Zhang. Parameterized modeling incorporating MOR-based rational transfer functions with neural networks for microwave components. *IEEE Microwave and Wireless Components Letters*, 32(5):379–382, 2022.

[32] Hakki Mert Torun, Huan Yu, Nihar Dasari, Venkata Chaitanya Krishna Chekuri, Arvind Singh, Jinwoo Kim, Sung Kyu Lim, Saibal Mukhopadhyay, and Madhavan Swaminathan. A spectral convolutional net for co-optimization of integrated voltage regulators and embedded inductors. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 1–8. IEEE, 2019.

[33] Hakki Mert Torun, Ahmet Cemal Durgun, Kemal Aygün, and Madhavan Swaminathan. Causal and passive parameterization of s-parameters using neural networks. In *IEEE Transactions on Microwave Theory and Techniques*, volume 68, pages 4290–4304, 2020.

[34] Guo Zhang, Hao He, and Dina Katabi. Circuit-GNN: Graph neural networks for distributed circuit design. In *International Conference on Machine Learning (ICML)*, pages 7364–7373. PMLR, 2019.

[35] Lihao Liu, Yunhui Li, Beisi Lu, Li Shang, and Fan Yang. GNN-SP: Fast s-parameter estimation of chiplet interconnect via graph neural network. *IEEE Transactions on Components, Packaging and Manufacturing Technology (TCPMT)*, 2024.

[36] Doyun Kim, Jaemin Park, Youngmin Oh, and Bosun Hwang. TraceFormer: s-parameter prediction framework for PCB traces based on graph transformer. In *ACM/IEEE Design Automation Conference (DAC)*. IEEE, 2024.

[37] Shutong Qi and Costas D. Sarris. Deep neural networks for rapid simulation of planar microwave circuits based on their layouts. *IEEE Transactions on Microwave Theory and Techniques*, 70(11):4805–4815, 2022.

[38] Ren Shibata, Masataka Ohira, and Zhewang Ma. A novel convolutional-autoencoder based surrogate model for fast s-parameter calculation of planar BPFs. In *IEEE MTT-S International Microwave Symposium (IMS)*, pages 498–501, 2022.

[39] Emir Ali Karahan, Zheng Liu, and Kaushik Sengupta. Deep-learning-based inverse-designed millimeter-wave passives and power amplifiers. *IEEE Journal of Solid-State Circuits*, 58(11):3074–3088, 2023.

[40] Hyunsu Chae, Hao Yu, Sensen Li, and David Z. Pan. PulseRF: Physics-augmented ml modeling and synthesis for high-frequency RFIC design. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 1–9, 2024.

[41] Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the coordconv solution. *Advances in neural information processing systems*, 31, 2018.

[42] Jinwoo Kim, Dat Nguyen, Seonwoo Min, Sungjun Cho, Moontae Lee, Honglak Lee, and Seunghoon Hong. Pure transformers are powerful graph learners. *Advances in Neural Information Processing Systems*, 35:14582–14595, 2022.

[43] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer, 2015.

[44] Zijun Cui, Tian Gao, Kartik Talamadupula, and Qiang Ji. Knowledge-augmented deep learning and its applications: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–21, 2023.

[45] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

[46] Jared Willard, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. Integrating physics-based modeling with machine learning: A survey. *arXiv preprint arXiv:2003.04919*, 1(1):1–34, 2020.

[47] Valeriy Gavrishchaka, Olga Senyukova, and Mark Koepke. Synergy of physics-based reasoning and machine learning in biomedical applications: towards unlimited deep learning with limited data. *Advances in Physics: X*, 4(1):1582361, 2019.

[48] John S. Toll. Causality and the dispersion relation: Logical foundations. *Phys. Rev.*, 104:1760–1770, Dec 1956.

[49] Shuzhan Sun and Dan Jiao. Multiphysics modeling and simulation of 3-D Cu–graphene hybrid nanointerconnects. *IEEE Transactions on Microwave Theory and Techniques*, 68(2):490–500, 2020.

[50] Li Xue and Dan Jiao. Fast method for accelerating convergence of iterative partial differential equation solvers by changing system matrix to laplacian counterpart. *IEEE Transactions on Antennas and Propagation*, 70(2):1187–1197, 2021.

[51] Li Xue and Dan Jiao. Rapid modeling and simulation of integrated circuit layout in both frequency and time domains from the perspective of inverse. *IEEE TransaFctions on Microwave Theory and Techniques*, 68(4):1270–1283, 2020.

[52] Chenhao Chu, Yuhao Mao, and Hua Wang. Transfer learning assisted fast design migration over technology nodes: A study on transformer matching network. In *IEEE MTT-S International Microwave Symposium (IMS)*, pages 188–191, 2024.

[53] Siawpeng Er, Edward Liu, Minshuo Chen, Yan Li, Yuqi Liu, Tuo Zhao, and Hua Wang. Deep learning assisted end-to-end synthesis of mm-wave passive networks with 3d em structures: A study on a transformer-based matching network. In *IEEE MTT-S International Microwave Symposium (IMS)*, pages 66–69. IEEE, 2021.

[54] Jungmin Lee, Wei Jia, Berardi Sensale-Rodriguez, and Jeffrey S Walling. Pixelated RF: Random metasurface based electromagnetic filters. In *IEEE International NEWCAS Conference (NEWCAS)*, pages 1–5. IEEE, 2023.

[55] Tzu-Yuan Huang, Naga Sasikanth Mannem, Sensen Li, Doohwan Jung, Min-Yu Huang, and Hua Wang. A 26-to-60ghz continuous coupler-doherty linear power amplifier for over-an-octave back-off efficiency enhancement. In *IEEE International Solid-State Circuits Conference (ISSCC)*, volume 64, pages 354–356, 2021.

[56] Yaolong Hu, Xiaohan Zhang, Qiang Zhou, Fan Cai, Cindy Cui, and Taiyun Chi. Automated mmWave power amplifier design flow and a 28-GHz design example in 45-nm CMOS SOI. In *IEEE MTT-S International Microwave Symposium (IMS)*, pages 842–845, 2024.