Machine learning-assisted design and analysis of electromagnetic structures

Zhou, Zhao

DOI (link to publication from Publisher):
10.54337/aau695980301

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

MACHINE LEARNING-ASSISTED DESIGN AND ANALYSIS OF ELECTROMAGNETIC STRUCTURES

BY
ZHAO ZHOU

DISSERTATION SUBMITTED 2023

AALBORG UNIVERSITY
DENMARK

Machine Learning-Assisted Design and Analysis of Electromagnetic Structures

Ph.D. Dissertation

Zhao Zhou

Dissertation submitted December, 2023

Dissertation submitted: December, 2023

PhD supervisor: Associate Professor Ming Shen
Aalborg University

PhD committee: Associate Professor John-Josef Leth (chairman)
Aalborg University, Denmark

Professor Roberto Gomez-Garcia
University of Alcalá, Spain

Associae Professor Fei Ding
University of Southern Denmark, Denmark

PhD Series: Technical Faculty of IT and Design, Aalborg University

Department: Department of Electronic Systems

ISSN (online): 2446-1628
ISBN (online): 978-87-7573-587-7

Published by:
Aalborg University Press
Kroghstræde 3
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Stibo Complete, 2023

Curriculum Vitae

Zhao Zhou

Zhao Zhou was born in Shaanxi, China. He received the B.Sc. and M.Eng. degrees in electronic engineering from Xidian University, Xi'an, China, in 2017 and 2020, respectively. In 2020, he became a Ph.D. researcher in the Antennas, Propagation, and Millimeter-Wave Systems Section, Department of Electronic Systems, Aalborg University, Aalborg, Denmark. In 2023, he visited the University of Alcalá, Spain for 3 months.

Curriculum Vitae

Abstract

Electromagnetic structures play a crucial role in diverse applications, yet conventional design and analysis methods suffer from reliance on experience, labor-intensive processes, and inefficiencies due to the absence of closed-form formulations that accurately describe their behaviors. Existing machine learning-based methods accelerate this process but require heavy simulation needs. This thesis addresses these challenges by developing efficient machine learning to enhance electromagnetic structure design and analysis.

A two-order machine learning approach is proposed to alleviate the simulation requirements for array radiation synthesis. Traditional calculation-based synthesis is incomplete and inaccurate due to its omission of the coupling effect; the simulation-aided synthesis yields precise results at a high computation and time cost. Existing machine learning-based synthesis methods approximate the projection from synthesized radiations to array excitation states purely through simulation data. To reduce the simulation need, the proposed method decomposes the learning task into two orders. The first-order learns fundamental superposition from low-cost calculation-based syntheses, and only a reduced simulation dataset is needed to impart the coupling effect to the second-order. Comparison results suggest a 50 % reduction in simulation needs.

A backward machine learning method is developed to achieve fast array calibration, which aims to determine the complex array excitation that leads to the expected array radiation. A novel feature extraction scheme is proposed first for compressing the array radiation into compact features, as its redundant information might complicate the projection and increase the data requirement. This minimizes the measurement times needed for calibrating an array. The requirement for simulation data is further reduced through knowledge-based transfer learning. The radiation superposition theorem is exploited to simplify the training task, thus saving the simulation data needed to establish the backward projection. Experimental results show comparable calibration performance with fewer measurement times compared with typical power-only calibration methods. Besides, it imposes no demands for phase shifters, involves no equation manipulations, and executes fast.

A fully automated inverse design framework is proposed to accelerate the design of frequency-selective surfaces. Typical machine learning-based inverse design methods rely on human engineers to select the topology first according to their experience, and then collect many simulation data to train surrogate models. They are not technically fully automated and suffer from a heavy simulation need. The proposed framework utilizes representation learning to realize the fully automated and low-cost inverse design of frequency-selective surfaces. Given electromagnetic constraints, the auto-selection module decides a topology based on classification; the auto-evolution module generates an optimal design through the combination of neural networks and particle swarm optimization. Four structures are designed, fabricated, and tested to validate the framework.

To universally reduce simulation needs, a high-quality data acquisition strategy is proposed. It can be applied to various design tasks and integrated with various machine learning methods to reduce the simulation need. Conventional methods acquire data by sweeping key geometrical parameters of electromagnetic structures on a fine, uniform grid of the parameter space. The qualities of collected data are unbalanced and unstable due to the variability in electromagnetic responses. To solve this, an acquisition algorithm is developed to recognize high-quality data before simulating their electromagnetic responses. It leverages computational resources to simulate high-quality data. Four implementations are carried out to compare the simulation requirements of uniform sampling and the proposed high-quality data acquisition method, suggesting an average reduction of 40 % in simulation requirements for the same model accuracy.

Finally, an efficient Bayesian-inspired sampling-based machine learning method is introduced. Unlike typical machine learning-assisted methods that collect all the training data first and then train the machine learning model, this approach employs a Bayesian-inspired algorithm to dynamically evaluate and generate the training data during the training procedure. The process begins with the training of the machine learning model using a limited initial dataset. Following the Bayesian-based expressions, the training loss and training data distribution are continuously assessed; new sets of high-quality data are iteratively generated and incrementally incorporated into the training dataset. Three design cases are conducted to evaluate the proposed method's effectiveness. The experimental results demonstrate a significant reduction in the simulation need and a noteworthy enhancement in the overall efficiency compared to conventional machine learning methods.

Resumé

Elektromagnetiske strukturer spiller en afgørende rolle i forskellige anvendelser, men konventionelle metoder til design og analyse lider under afhængighed af erfaring, arbejdskrævende processer og ineffektivitet på grund af manglen på lukkede formler, der præcist beskriver deres adfærd. Eksisterende metoder baseret på maskinlæring fremskynder denne proces, men kræver omfattende simuleringer. Denne afhandling adresserer disse udfordringer ved at udvikle effektiv maskinlæring til at forbedre design og analyse af elektromagnetiske strukturer.

En totrins maskinlæringstilgang foreslås for at lette simuleringen af array-radiationssyntese. Traditionel beregningsbaseret syntese er ufuldstændig og unøjagtig på grund af manglende koplingseffekt; simuleringssupporteret syntese giver præcise resultater, men kræver høj beregning og tidsomkostning. Eksisterende maskinlæringsbaserede syntesemetoder tilnærmer projektionen fra syntetiserede strålinger til array-ekscitationsstatistikker udelukkende gennem simuleringer. For at mindske simuleringen opdeles den foreslåede metode opgaven i to trin. Første trin lærer grundlæggende overlejringsprincipper fra omkostningsberegnet syntese, og kun en reduceret simuleringsdatasæt er nødvendigt for at overføre koplingseffekten til andet trin. Sammenligningsresultater antyder en reduktion på 50 % i simuleringens behov.

En metode til bagudrettet maskinlæring er udviklet for at opnå hurtig kalibrering af arrays, der sigter mod at bestemme den komplekse array-ekscitation, der fører til den forventede array-stråling. En ny featureudvindingsmetode foreslås først til komprimering af array-stråling til kompakte funktioner, da redundant information kan komplicere projektionen og øge datakravet. Dette minimerer måletiderne, der er nødvendige for kalibrering af et array. Kravet til simuleringsdata reduceres yderligere gennem vidensbaseret transfer learning. Strålingsoverlejringsteoremet udnyttes til at forenkle træningsopgaven og derved spare på de simuleringsdata, der er nødvendige for at etablere bagudrettet projektion. Eksperimentelle resultater viser sammenlignelig kalibreringspræstation med færre måletider sammenlignet med typiske metoder baseret på effektalgoritmer. Derudover stiller den ingen krav til faseskiftere, involverer ingen ligningsmanipulationer og udføres hurtigt.

Et fuldautomatisk inverse designframework foreslås for at accelerere designet af frekvensselektive overflader. Typiske metoder til inverse design baseret på maskinlæring er afhængige af, at menneskelige ingeniører først vælger topologien i overensstemmelse med deres erfaring og derefter indsamler mange simuleringsdata for at træne surrogatmodeller. De er ikke teknisk fuldt automatiserede og lider under et tungt behov for simulering. Det foreslåede framework udnytter repræsentationsindlæring til at realisere fuldautomatiseret og omkostningseffektivt inverse design af frekvensselektive overflader. Givet elektromagnetiske begrænsninger beslutter auto-selektionsmodulet en topologi baseret på klassifikation; auto-evolutionsmodulet genererer en optimal design gennem kombinationen af neurale netværk og partikelsværmeoptimering. Fire strukturer designes, fabrikeres og testes for at validere frameworket.

For at universelt reducere behovet for simulering foreslås en strategi for høj kvalitet til dataindsamling. Den kan anvendes til forskellige designopgaver og integreres med forskellige metoder inden for maskinlæring for at reducere behovet for simulering. Konventionelle metoder indsamler data ved at gennemgå centrale geometriske parametre for bekymrede elektromagnetiske strukturer på en fin, uniform gitter af parameterområdet. Kvaliteten af indsamlede data er ubalanceret og ustabil på grund af variabiliteten i elektromagnetiske reaktioner. For at løse dette udvikles en algoritme til at genkende høj kvalitet data, før simulering af deres elektromagnetiske reaktioner. Den udnytter beregningsressourcer til simulering af høj kvalitet data. Fire implementeringer udføres for at sammenligne kravene til simulering af uniform prøvetagning og den foreslåede metode til dataindsamling af høj kvalitet, hvilket antyder en gennemsnitlig reduktion på 40 % i kravene til simulering for samme modelnøjagtighed.

Endelig introduceres en effektiv Bayesian-inspireret prøvetagingsbaseret maskinlæringsmetode. I modsætning til typiske maskinlæringsbaserede metoder, der først indsamler al træningsdata og derefter træner maskinlæringsmodellen, anvender denne tilgang en Bayesian-inspireret algoritme til dynamisk at evaluere og generere træningsdata under træningsproceduren. Processen begynder med træningen af maskinlæringsmodellen ved hjælp af et begrænset indledende datasæt. Efter de bayesiske udtryk vurderes træningstab og fordeling af træningsdata kontinuerligt; nye sæt af høj kvalitet data genereres iterativt og inkorporeres gradvist i træningsdatasættet. Tre designcases udføres for at evaluere metodens effektivitet. De eksperimentelle resultater demonstrerer en betydelig reduktion i behovet for simulering og en markant forbedring i samlet effektivitet sammenlignet med konventionelle metoder inden for maskinlæring.

Contents

Thesis Details

Thesis Title:	Machine Learning-Assisted Design and Analysis of Electromagnetic Structures
Ph.D. Candidate:	Zhao Zhou
Supervisor:	Assoc. Prof. Ming Shen, Aalborg University

This thesis is submitted as partial fulfillment of the requirements for the degree of Doctor of Philosophy (Ph.D.) from Aalborg University, Denmark. The thesis is structured as a collection of papers with the main part of the thesis being scientific papers published in, or submitted to, peer-reviewed journals and conferences. The thesis is a result of three years of research, as a PhD researcher in the section of Antennas, Propagation and Millimetre-wave Systems (APMS), Department of Electronic Systems, Aalborg University, Denmark.

This work was supported in part by the China Scholarship Council (CSC) under grant 202006960011.

The main body of this thesis consists of the following papers:

[A] Z. Zhou, Z. Wei, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Two-Order Deep Learning for Generalized Synthesis of Radiation Patterns for Antenna Arrays," in *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 5, pp. 1359–1368, 2023.

[B] Z. Zhou, Z. Wei, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Transfer-Learning-Assisted Multielement Calibration for Active Phased Antenna Arrays," in *IEEE Transactions on Antennas and Propagation*, vol. 71, no. 2, pp. 1982–1987, 2023.

[C] Z. Zhou, Z. Wei, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Representation Learning-Driven Fully Automated Framework for the Inverse Design of Frequency-Selective Surfaces," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 6, pp. 2409–2421, 2023.

[D] Z. Zhou, Z. Wei, A. Tahir, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "A High-Quality Data Acquisition Method for Machine-Learning-Based Design and Analysis of Electromagnetic Structures," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 10, pp. 4295–4306, 2023.

[E] Z. Zhou, Z. Wei, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Bayesian-Inspired Sampling for Efficient Machine-Learning-Assisted Microwave Component Design," in *IEEE Transactions on Microwave Theory and Techniques*, Early Access, 2023.

According to the Ministerial Order no. 1039 of August 27, 2013, regarding the PhD Degree § 12, article 4, statements from each co-author about the PhD students contribution to the above-listed papers have been provided to the PhD school for approval prior to the submission of this thesis. These co-author statements have also been presented to the PhD committee and included as a part of their assessment.

In addition to the listed papers as the main content of this thesis, the following papers were also either authored or co-authored during the PhD studies. As these papers are not a part of the main body of this thesis they have not been included in print. The reader is therefore kindly referred to their respective publishing channels as listed hereafter.

[1] Z. Wei, Z. Zhou, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Equivalent Circuit Theory-Assisted Deep Learning for Accelerated Generative Design of Metasurfaces," in *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 7, pp.5120–5129, 2022.

[2] Z. Wei, Z. Zhou, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Fully Automated Design Method Based on Reinforcement Learning and Surrogate Modeling for Antenna Array Decoupling," in *IEEE Transactions on Antennas and Propagation*, vol. 71, no. 1, pp. 660–671, 2023.

[3] Z. Wei, Z. Zhou, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Automated Antenna Design via Domain Knowledge-Informed Reinforcement Learning and Imitation Learning," in *IEEE Transactions on Antennas and Propagation*, vol. 71, no. 7, pp. 5549–5557, 2023.

[4] Z. Zhou, Z. Wei, Y. Zhang, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Training of deep neural networks in electromagnetic problems: a case study of antenna array pattern synthesis," in *IEEE MTT-S International Wireless Symposium*, Nanjing, China, 2021, pp. 1-3.

[5] Z. Zhou, Z. Wei, J. Ren, N. Sun, J. Kang, Y. Yin, M. Shen, "A study on machine learning assisted accelerated design of microwave structures," in

Photonics & Electromagnetics Research Symposium, Prague, Czech Republic, 2023, pp. 1189-1192.

[6] Z. Wei, Z. Zhou, Y. Zhang, P. Li, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Domain knowledge assisted training dataset generation for metasurface designs," in *IEEE MTT-S International Wireless Symposium*, Nanjing, China, 2021, pp. 1-3.

[7] Z. Wei, Z. Zhou, M. Shen, "Application and challenges of machine learning-assisted antenna design," in *Photonics & Electromagnetics Research Symposium*, Prague, Czech Republic, 2023.

Preface

Reflecting on the past three years, I am humbled to admit that it is an unparalleled dream for me. Born and growing up in a village in Shaanxi Province in China, I never envisioned myself embarking on such a path. I hesitated every time on whether to pursue a master's degree and further a Ph.D., let alone in a foreign land. A series of uncertainties have driven and guided me throughout this three-year Ph.D. research journey in Denmark. Words fail to express the depth of my gratitude to those remarkable individuals who have made irreplaceable contributions to these uncertainties. Still, I would like to take this opportunity to thank them for bringing me here.

First and foremost, I wish to express my sincerest thanks to my supervisor, Assoc. Prof. Ming Shen, and our esteemed section leader, Prof. Gert Frølund Pedersen, for offering me this opportunity and for your unwavering guidance throughout my journey as a Ph.D. researcher in the Antenna Propagation and Millimeter-wave Systems (APMS) section. To my colleagues in the APMS section, thank you for your support during my study in Denmark.

Beyond uncertainties, one thing remains indisputable—I am incredibly fortunate to have the best family and friends. My parents have steadfastly supported every decision I have made, regardless of their own limited exposure to higher education or their roots firmly grounded in the village. Their unwavering faith in me has been a constant source of strength. Additionally, I have been consistently recharged and rejuvenated through the connections and interactions with my family and friends. I hope that I can reciprocate this support and always be a pillar for them in return.

To all of you who supported me and to the readers, this thesis is based on the collection of my research results on machine learning-assisted design and analysis of electromagnetic structures during the past three years. I sincerely hope that I have come close to meeting your expectations.

Zhao Zhou
Aalborg University, 28th November, 2023

Part I

Introduction

Introduction

1 Motivation

1.1 Electromagnetic Structures

Electromagnetic structures, such as antennas [1], metasurfaces [2], and filters [3], are deliberately designed to manipulate electromagnetic signals, finding applications in diverse fields including wireless communication, radar detection, and medical imaging. The characterization of their effects on electromagnetic signals within a specific frequency range is described by their electromagnetic responses. There are various electromagnetic responses, such as scattering coefficients (transmission or reflection), gain, beam width, axial ratio, etc. Achieving desired electromagnetic responses involves selecting the topologies and tuning the geometries of these electromagnetic structures to meet the specific constraints of the intended scenarios.

The performance of electromagnetic structures is affected by many factors, including geometric configurations, material properties, and environmental conditions. These factors intricately interact with one another, resulting in nonlinear and non-uniform influences on the characteristics of electromagnetic structures. Consequently, deriving precise closed-form equations that account for all possible variations and interactions presents a considerable challenge. As a result, the design and analysis of electromagnetic structures heavily rely on knowledge and experience.

The dependency on knowledge and experience manifests in two distinct aspects, corresponding to two stages of the design process: topology selection and geometry adjustment. Given a particular application scenario, designers rely on their historical knowledge and experience to initially select a suitable topology from their familiar options. Subsequently, they iteratively fine-tune the geometrical parameters of the chosen topology to achieve the desired electromagnetic behavior that complies with the specific electromagnetic constraints.

The past decades have witnessed the development of electromagnetic designs. Before the widespread use of electromagnetic simulators, engineers

relied on theoretical analysis, experimentation, and manual calculation to design electromagnetic structures. Theoretical analysis involves the calculation of Maxwell's equations and their variations to gain insights into the fundamental working principle and determine the dimensions of electromagnetic structures. For example, the theoretical length of a dipole wire antenna is half of the wavelength at the operating frequency, as expressed in equation 1,

$$L = \frac{\lambda}{2}, \tag{1}$$

where L represents the length of the dipole, λ represents one wavelength at the operating frequency. Following the theoretical analysis, prototypes of designed structures are fabricated and tested to measure their performance. For example, using network analyzers to measure the impedance, gain, etc. The measured data is taken as a reference for further refining the structures to get close to the desired performance. The refinement is estimated through manual calculations based on the measured data. Engineers use charts and graphs, for example, the Smith chart, to assist the estimation process. They repeat the testing of performance and the adjusting of physical structures iteratively until the prototypes generate the desired performance. However, the calculation accuracy was low, leading to a tedious and time-consuming design process. The availability was limited to only the simple and standard types of geometries like dipole wire antennas. As the study progressed, empirical formulas were derived from accumulated experimental observations. Engineers improved the accuracy and extended the availability by combining theoretical analysis and empirical formulas. They added some empirical factors to the theoretical equations to compensate for the difference between theoretical and experimental results. As in the dipole design case, its length is modified to an empirical expression from 0.47λ to 0.48λ [1]. The length of thinner dipole wires is closer to 0.48λ; The length should be further reduced for thicker wires. The reduction of the length is caused by the capacitive end effect. Engineers established approximate design guidelines for these simple and standard electromagnetic structures. The design and analysis of electromagnetic structures were rather limited in scope during this period.

The design and analysis of electromagnetic structures were revolutionized by the widespread use of full-wave electromagnetic simulators since the 1990s when typical full-wave simulators were released, including Computer Simulation Technology (CST), High-Frequency Structure Simulator (HFSS), FEld & KOhn (FEKO), etc. The core procedure of full-wave simulation is to mesh the computational domain into discrete elements or cells and to solve Maxwell's equations within these elements or cells. The three simulators utilize various numerical methods to solve Maxwell's equations. The main solvers of CST include the finite integration technique and time domain solver. HFSS uses finite element analysis as its primary solver. FEKO employs the Method

of Moments, Multilevel Fast Multipole Method, and Physical Optics. These simulators provide engineers with powerful tools to accurately predict and optimize the responses of the electromagnetic structures. They allow engineers to design and analyze modified and intricate electromagnetic structures. The utilization of these simulators has become the standard practice in the design and analysis process of electromagnetic structures. The main limitation is that they require specific computational resources. Larger or more intricate structures require more meshing, hence leading to a number of elements or cells and a heavy calculation burden, which is more computationally expensive. Currently, both the complexity and scale of electromagnetic structures continue to increase to meet the more restrictive requirements of new communication eras, posing a growing computational burden on simulation-based design methods. While the advancement of computing capabilities facilitates the design and analysis of increasingly intricate structures, integrating them with more intelligent solving techniques can enhance the overall impact.

Machine learning offers a promising chance to further improve the design and analysis of electromagnetic structures. As mentioned above, the rise in computational demands challenges the efficiency of simulation-based design methods. To address this, there is a growing interest in integrating machine learning techniques into the design and analysis process. Instead of relying solely on solving Maxwell's equations within the computational domain, machine learning models can learn and capture patterns and relationships from historical data, hence allowing for more efficient and accurate predictions of the behaviors of electromagnetic structures. Furthermore, machine learning has the potential to uncover novel insights and optimizations in higher dimensional space that can be difficult to observe through manual calculation or conventional simulation-based methods. The captured knowledge can significantly improve the design and analysis of electromagnetic structures. However, current machine-learning methods require sufficient simulation data to train the surrogate models. The challenge lies in striking a balance between the utilization of electromagnetic simulators and the development of machine-learning models. A better integration of these two approaches could pave the way for more comprehensive and efficient design and analysis methods of increasingly complex electromagnetic structures demanded in modern communication eras.

1.2 Conventional Design and Analysis Methods

1.2.1 Antennas

Antennas are designed to receive or transmit electromagnetic wireless signals [1]. Key indicators to evaluate the performance of an antenna include operating frequency—at which frequency the antenna works—, bandwidth—the

maximum frequency range where the antenna remains functional—, gain—the radiation intensity in certain directions—, polarization—the orientation of the electric field vector as electromagnetic signals radiate: linear, circular, elliptical—, beamwidth—the angular width of the main lobe that the antenna radiates the maximum power—, size, etc.

Antenna engineers design antennas according to specific requirements with respect to the abovementioned indicators [2]. First, they need to select a proper antenna type from their familiar antenna types. Typical antenna types include microstrip patch antenna, dipole antenna, horn antenna, Yagi-Uda antenna, log-periodic antenna, slot antenna, etc. They vary a lot with respect to the abovementioned indicators and have different limitations in real application scenarios. For example, microstrip patch antennas have a low profile but a narrow bandwidth; horn antennas have a wide bandwidth but a bulky structure. After a proper antenna type is selected, designers need to model a draft structure. Its geometric parameters should be roughly estimated and initialized based on domain knowledge and experience about the selected antenna type. For example, a microstrip patch antenna should have an approximate half-wavelength patch resonator. Then designers need to fine-tune the geometry until it fulfills the electromagnetic requirements. The fine-tuning process involves multiple iterations of sweeping geometric individual parameters or several parameters jointly, simulating the swept combinations of geometric parameters using full-wave simulators, reasoning the simulation result, and tuning the geometric parameters or modifying the geometry if needed.

1.2.2 Metasurfaces

Metasurfaces are periodic sub-wavelength structures that act as spatial filters to reflect or transmit wireless signals at certain frequencies [4]—also known as frequency selective surfaces—, absorbers [5], polarization converters [6], radar cross-section reducers [7], etc. Key indicators to evaluate a metasurface include reflection/transmission coefficient, polarization, phase, size, etc. The performance of a metasurface is decided by the structures of its unit cells. Depending on whether all the unit cells are identical, there are uniform metasurfaces and multi-bit metasurfaces. Uniform metasurfaces have identical unit cells. Multi-bit metasurfaces have several types of modified unit cells, for example, an n-bit metasurface has 2^n types of modified unit cells.

According to the desired electromagnetic function that a metasurface is expected to perform, designers need to choose a suitable topology for the unit cell from their familiar options first, such as patches, particles, Jerusalem cross, etc. Afterward, they initialize one/multiple unit cell(s) and array the unit cells based on domain knowledge and experience. Finally, they fine-tune the unit cell until the formed metasurface performs the expected functionality.

1.2.3 Filters

Filters reflect or transmit wired signals at certain frequencies. Key indicators of filters include operating frequency, bandwidth, reflection/transmission coefficient, insertion loss—transmission loss within the operating frequency band—, roll-off rate—how fast signals attenuate as moving from passband to stopband—, group delay—phase delay across the passband—, size, etc.

Filter design requires a good understanding of microwave circuit theory and electromagnetic field theory. Given desired filtering responses, designers determine the filter topology first. Common filter topologies include Butterworth, Chebyshev, Elliptic, etc. Afterward, designers build an equivalent circuit that generates the desired filtering performance. Radiofrequency simulators can support modeling and tuning of the circuit, for example, the Advanced Design System (ADS). Based on the equivalent circuit, proper microwave components are selected to replace all the circuit elements. The projection between microwave structures and circuit components can be guided through microwave circuit and electromagnetic field knowledge, but it can hardly be formulated. The formed microwave filter structure requires further fine-tuning to fulfill the filtering requirements.

1.2.4 Limitations

Conventional approaches in the design and analysis of electromagnetic structures heavily depend on the expertise and experience of human engineers, necessitating a manual and iterative process, because the precise characterization of the intricate relationship between geometries and electromagnetic responses remains a challenging task.

The conventional design process involves the careful selection of suitable topologies and the subsequent fine-tuning of their geometrical parameters. However, the choice of topologies is often limited by the designers' familiarity with a specific set of candidate options. Consequently, the process of selecting an optimal topology becomes constrained and less flexible. The fine-tuning stage further adds to the conventional approach's labor-intensive and computationally demanding nature. Designers rely on iterative trial and error methods to adjust the geometries until the desired electromagnetic responses are achieved within the specified constraints. This manual process lacks efficiency and is prone to inefficiencies and suboptimal solutions.

During the fine-tuning process, designers base their adjustment decisions on their reasoning and understanding of historical simulation results. Full-wave simulators such as CST, HFSS, or FEKO provide the necessary support for these simulations. Thus, the choices made by individual designers can vary significantly, influenced by their different understanding and reasoning abilities, leading to unstable design efficiency. Due to the complexity involved,

designers normally tune only one or several parameters simultaneously and keep most of the parameters unchanged to ensure reasonability. However, the effects of the tuning parameters on the performance heavily depend on the current values of other parameters. The tuning parameters can just arrive at a suboptimal combination with other parameters being fixed at the current values. Designers have to retune these parameters after other parameters are changed in the following tuning process. Therefore, the tuning efficiency is rather low. The comprehension of the design space becomes increasingly challenging as the number of jointly changed parameters increases. This limitation will be particularly significant in future scenarios where large and sophisticated electromagnetic structures are required to meet restrictive constraints in complex environments.

The limitations of conventional methods are mainly imposed by their reliance on human expertise and heavy computation. By exploring machine learning, it is possible to address these limitations to improve the design and analysis efficiency of electromagnetic structures.

1.3 Machine Learning-Assisted Design and Analysis Methods

Machine learning can significantly improve the design and analysis efficiency of electromagnetic structures, due to its incredible ability to summarize the regularity of a complex numerical space from historical data samples. Designers can use machine learning to develop surrogate models to approximate the complex characteristics of electromagnetic structures. There are roughly three types of surrogate models: forward models [8], inverse models [9], and generative models [10], which promote three different applications.

1.3.1 Forward Model-Based Method

The primary objective of forward models is to establish a mapping or projection from the space of geometrical parameters to the space of electromagnetic responses. By training these forward models, it becomes possible to estimate the electromagnetic responses of a structure for unseen combinations of geometrical parameters, without relying on computationally expensive full-wave simulators.

The process begins by acquiring a dataset that consists of known geometric parameters and their corresponding electromagnetic responses through simulation. The acquired dataset is then utilized to train the forward model. The geometric parameter settings are taken as the input and the corresponding electromagnetic responses are taken as the output. The well-trained forward model can replace the full-wave simulator for estimating the electromagnetic responses for new combinations of geometric parameters. This enables the use of forward models in conjunction with optimization algorithms (e.g., ge-

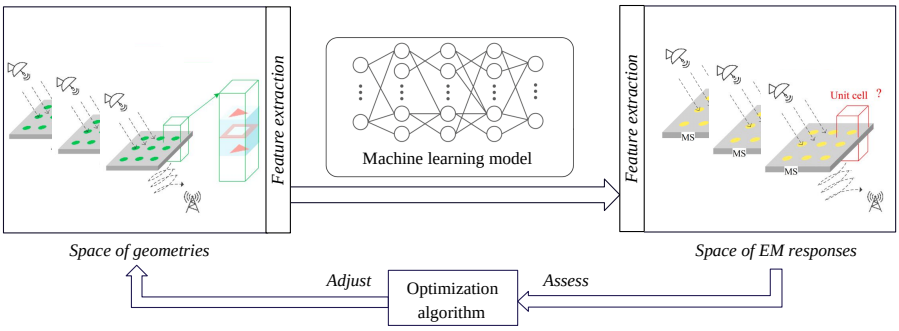


Fig. 1: Forward model-based method.

netic algorithm [11] or particle swarm optimization [12]) to accelerate the design of electromagnetic structures, as demonstrated in Fig. 1. These algorithms work iteratively, starting with a set of randomly initialized geometries. Each geometry is analyzed using the forward model to obtain its predicted electromagnetic response. The optimization algorithm then analyzes the responses and identifies potential improvements. Based on the analysis, the algorithm suggests modifications to the geometries by perturbing their parameters, such as shape, size, or material properties. The modified geometries are subsequently analyzed using the forward model to evaluate their new electromagnetic responses. This iterative process continues, with the algorithm iteratively refining the geometries based on the predicted responses until an optimal geometry is obtained.

Compared to traditional methods that rely on full-wave simulators for simulation, the integration of forward model-based optimization significantly improves the efficiency of the design process. By leveraging the learned knowledge encoded in the forward model, the optimization algorithm can explore the parameter space more efficiently, leading to faster convergence towards optimal designs.

Overall, the utilization of machine learning-based forward models in the design and analysis of electromagnetic structures provides an efficient and effective approach for exploring the geometrical parameter space, estimating electromagnetic responses for unseen configurations, and accelerating the design process through integration with optimization algorithms.

1.3.2 Inverse Model-Based Method

Unlike forward models, which focus on predicting electromagnetic responses based on given geometrical parameters, inverse models work in the opposite direction. They develop a reverse projection from the electromagnetic response space to the geometrical parameter space, enabling the selection of a topology

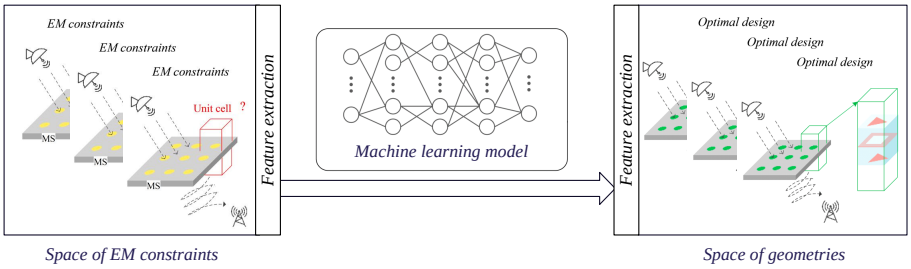


Fig. 2: Forward model-based method.

or determination of an optimal geometry that is likely to generate desired electromagnetic responses.

Likewise, it starts with acquiring a dataset that consists of known geometries or known geometric parameter settings and their corresponding electromagnetic responses through simulation. The difference is that the inverse model is trained using known geometric parameter settings as the output and their corresponding electromagnetic responses as the input.

The usage of the inverse model is more straightforward compared to the forward model. This is because the objective of the inverse model aligns perfectly with the ultimate objective, which is the projection from electromagnetic constraints to satisfying geometries, as demonstrated in Fig. 2. By directly selecting the topology or determining the geometrical parameters based on desired electromagnetic constraints, the inverse model provides a streamlined pathway toward achieving the desired electromagnetic characteristics.

It is important to note that inverse models may encounter a challenge known as the "one-input-multi-output" problem, which can hinder model convergence. This problem arises when one setting of electromagnetic responses corresponds to multiple combinations of geometric parameters. One solution to solve this problem is to utilize more deterministic electromagnetic responses as the input [13]. For example, a reflection or transmission coefficient curve within a frequency range contains more deterministic information than only a bandwidth value. A more deterministic setting of electromagnetic responses has more restrictive constraints and is less likely to correspond to multiple combinations of geometric parameters. It enhances the inverse model's accuracy of projection from each setting of electromagnetic responses to its correct combination of geometric parameters. However, more deterministic electromagnetic responses can involve redundant features, hence increasing the dimensionality of the input. It might increase the complexity of the projection or fail to find a suitable combination of geometric parameters.

Inverse model-based methods provide a powerful tool for the inverse design of electromagnetic structures. Inverse models can be utilized to automatically select a proper topology or determine an optimal combination of ge-

ometric parameters that can generate the desired electromagnetic responses. Unlike forward model-based methods, they do not need to be integrated with optimization algorithms. The input of inverse models should be carefully considered. On one hand, deterministic electromagnetic responses should be selected as input to solve the one-input-multi-output problem. On the other hand, we should avoid involving redundant information to minimize the complexity of the projection.

1.3.3 Generative Model-Based Method

Generative models (e.g., generative adversarial network [14], variational autoencoder [15]) aim to generate new structures that resemble existing structures designed by human engineers. Unlike forward or inverse models that approximate the projection between geometric parameters and electromagnetic responses, generative models focus on capturing the characteristics of well-behaved electromagnetic structures. These models are trained on a diverse set of existing electromagnetic geometries to learn the underlying patterns and characteristics. For example, a variational autoencoder consists of an encoder and a decoder. The encoder takes each electromagnetic structure as an input and compresses it to a reduced-dimension vector in a latent space; the decoder takes the reduced-dimension vector in the latent space as an input and reconstructs the electromagnetic structure. The encoder and decoder are trained together. After training, the encoder can project the space of electromagnetic structures to a low-dimension latent space; the decoder can reconstruct electromagnetic structures from low-dimension vectors in the latent space.

Generative models are integrated with forward models and optimization algorithms for the design of electromagnetic structures. Generative models project between the space of electromagnetic structures and the reduced-dimension latent space. Forward models predict the corresponding electromagnetic responses. Optimization algorithms initialize and tune the structures iteratively. Compared with typical optimization algorithm-based methods, the main difference is that generative models compress the solving space by representing electromagnetic structures into reduced-dimension vectors. A typical representation of electromagnetic structures is to pixelate the solving space. Each electromagnetic structure is represented as an array. The size of the array depends on the number of pixels. More pixels lead to higher complexity and design freedom. Each element value of an array represents each pixel being metal or non-metal. Typical optimization algorithm-based methods directly optimize the structures in the pixelated space, hence resulting in high optimization complexity. Generative model-based methods project the pixelated space to a reduced-dimension latent space. This approach has two major advantages: the optimization complexity is significantly reduced, as the dimension of the solving space is decreased; an improved initial population of electromagnetic

structures that resemble human designs can be acquired, instead of randomly initialized in optimization algorithm-based methods. Therefore, generative model-based methods enhance the design efficiency of electromagnetic structures.

In summary, generative model-based methods accelerate the design of electromagnetic by representing the space of electromagnetic structures into a reduced-dimension latent space. Generative models can generate new structures that resemble well-behaved structures designed by human engineers. It provides promising initialized structures and reduces the dimension of the solving space for the optimization process. The integration of generative models with optimization algorithms and forward models can achieve an efficient electromagnetic design and analysis.

1.3.4 Challenges

Machine learning-based methods for the design and analysis of electromagnetic structures have their challenges, primarily stemming from computation resources and data accessibility. The potential contributions one can make in this field are contingent upon their computational abilities and access to data.

Giant technology companies and institutions with powerful computational resources and vast datasets have the opportunity to develop large surrogate models with impressive capabilities for solving complex tasks. For instance, OpenAI, a prominent example, has pioneered the development of a large language model known as ChatGPT, revolutionizing various text-processing-related industries.

Individual research groups and researchers can still contribute to machine learning-assisted electromagnetic solutions. This is because certain challenges persist regardless of the complexity of the task or the size of the model. We can contribute to addressing the remaining common challenges by investigating small-scale tasks using machine learning techniques.

One of the most significant challenges encountered by both large and small models is the interpretability of machine learning-based electromagnetic solutions. While machine learning models have shown impressive performance in various domains, their internal mechanisms lack transparency, making it difficult for human engineers to understand the reasoning behind their predictions or recommendations. In the context of electromagnetic structures, interpretability is crucial for engineers and researchers to trust and effectively utilize machine learning models. One approach to address this challenge is to incorporate electromagnetic domain knowledge into the development of machine learning models. By enhancing interpretability, researchers can enable better understanding and validation of machine learning-assisted electromagnetic solutions, thereby fostering their wider adoption in practical applications.

Another shared challenge for both large and small models is data acquisi-

tion. Obtaining relevant and high-quality data for training machine learning models in the domain of electromagnetic structures is a complex task. Electromagnetic data is often scarce, expensive to obtain, and of unstable, unpredictable quality. Additionally, acquiring electromagnetic data requires expert knowledge and is difficult to outsource. There is barely any publicly accessible dataset providing electromagnetic data. Researchers usually acquire data by simulating the target electromagnetic structures and sweeping the geometrical parameters using full-wave simulators. It is computationally expensive and the obtained simulation data suffer from unstable quality. These data acquisition challenges hinder the progress of machine learning-assisted electromagnetic solutions and represent a common obstacle for both giant institutions and individual researchers.

Although machine learning-based methods for the design and analysis of electromagnetic structures have difficulties in terms of computation resources and data accessibility, these challenges are not insurmountable. While tech giants work on large models, small research groups and individual researchers can contribute by tackling the aforementioned challenges. By improving interpretability and data acquisition, machine learning-assisted electromagnetic design and analysis can continue to advance, benefiting a wide range of applications and industries.

1.4 Key Machine Learning Techniques

1.4.1 Neural Networks

Neural networks [16] aim to approximate relationships within data, recognize patterns, and make data-driven predictions. They are inspired by biological neural structures and consist of interconnected artificial neurons organized into layers. The architecture of a neural network can be observed in Fig. 3. Depending on its location within the network, each layer is categorized as the input layer, the output layer, or the hidden layer. Each neuron acts as a processing unit that takes input and generates output via connections. The relationships between input and output are determined by the weight value (w) assigned to each connection and the bias value (b) assigned to each neuron. The weight decides the strength of a connection and the bias provides additional tuning flexibility. Activation functions are inserted between every two layers to introduce non-linearity into the network to enable the approximation of complex functions. Typical activation functions include linear function, binary step function, sigmoid function, hyperbolic tangent ($tanh$), rectified linear unit (ReLU) function, and variants like Leaky ReLU function, etc., as demonstrated in Fig. 4. Neural networks have many variations: deep neural networks (DNNs) have multiple hidden layers and have been instrumental in solving complex problems in various domains; convolutional neural networks

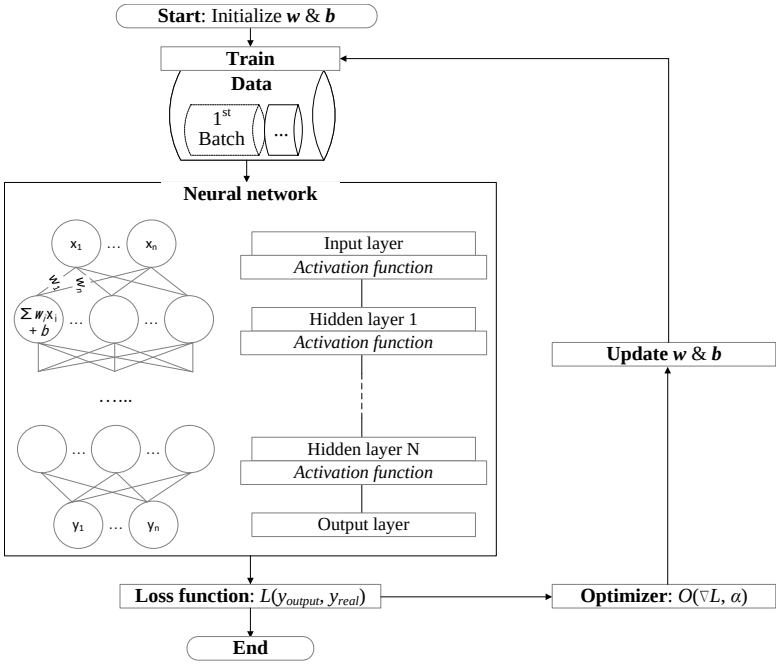


Fig. 3: Demonstration of the architecture and the training process of a neural network.

(CNNs) use convolutional neuron layers and are widely used in image recognition, object detection, etc.; recurrent neural networks include loops in their architectures and are suitable for time series prediction like natural language processing, speech recognition, etc.

As human beings learn from experience, neural networks are trained with experienced data consisting of pairs of input and output. The training process of the neural networks is controlled by epoch, learning rate (α), batch size, loss function (L), and optimizer. Epoch is the total number of training iterations. The learning rate decides the step at which the weights and biases update during training. A large learning rate leads to fast convergence but low stability, and a small learning rate results in high stability but slow convergence. The batch size determines the update frequency of weights and biases as they update once after a batch of data is processed. A larger batch size leads to more stable updates but a slower convergence; a smaller batch size adds more noise but converges faster. The updates are controlled by the loss function and the optimizer. The loss function evaluates the approximation performance of the neural network equipped with its current weights and biases by calculating the difference between the network's output and the real output. Common loss functions include mean squared error, cross-entropy, and various custom loss functions tailored to specific tasks. The optimizer determines the next update

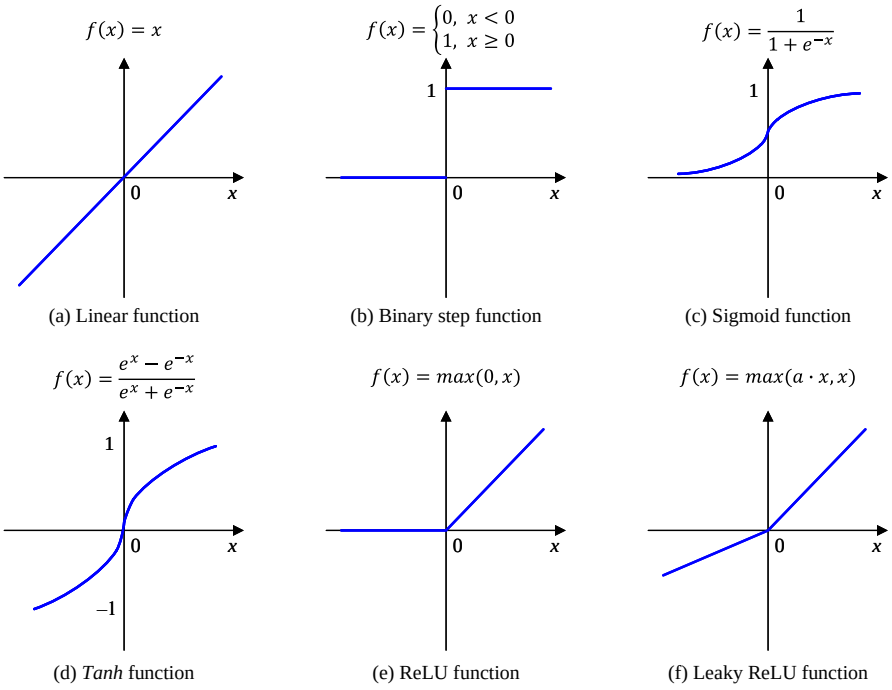


Fig. 4: Various activation functions. (a) Linear function. (b) Binary step function. (c) Sigmoid function. (d) $Tanh$ function. (e) ReLU function. (f) Leaky ReLU function.

for each weight and bias based on the current loss. Typical optimizers include stochastic gradient descent (SGD) and its variants (e.g., Adam, RMSprop).

Neural networks' training process can be described as a series of forward propagation and backpropagation processes. Forward propagation is referred to as the process of passing data from the input layer, through each hidden layer, to the output layer. Each layer takes the output of the last layer as input and calculates its output based on the input, weights, biases, and activation functions. The final output of the output layer is used to calculate the loss by being compared to the real output. Backpropagation calculates the gradients of the loss with respect to weights and biases in the direction from the output layer to the input layer. The gradients are processed using the optimizer to determine the next update steps of the weights and biases. By updating weights and biases, the neural network learns to minimize the loss and improve its approximation accuracy. The neural network continues to update its weights and biases iteratively until the maximum epoch is reached or a desired approximation accuracy is achieved.

1.4.2 Transfer Learning

Transfer learning is a machine learning technique that adapts or fine-tunes a pre-trained model on a source task to perform a different but related task. It aims to leverage the knowledge, information, and features learned from a source domain to help solve a new but related problem. The learned knowledge, information, and features are contained in the neuron layers of the pre-trained model. Depending on the application scenario, one can decide whether to keep specific neuron layers unchanged during the fine-tuning; certain neuron layers can be removed from the pre-trained neural network, or new neuron layers can be added to it.

Transfer learning offers three practical advantages, including improving model performance, reducing training time, and enhancing generalization ability. There are many scenarios where only limited data is available and collecting extra labeled data for the target task is expensive, time-consuming, or impractical. Transfer learning leads to substantial improvements in model performance for these conditions. Since less data is required for training, it can also significantly reduce the training time and computational resources required for the new task by starting with a pre-trained model. Moreover, transfer learning enhances the generalization ability and robustness of the machine learning model against various tasks. Researchers strive to develop universal models that generalize well across various scenarios, which requires large volumes of data and computation resources. Transfer learning addresses this challenge by recognizing the knowledge and information acquired in former tasks that can be leveraged to improve performance on other related tasks. It offers a high generalization ability across various scenarios in a more efficient manner.

1.4.3 Data Acquisition

Data acquisition for machine learning-assisted design and analysis of electromagnetic structures is a unique challenge, compared to other fields [17], such as text processing, image recognition, etc. It is easy to acquire texts and images online, from sharing archives, or by crowd-outsourcing, because the production and formation of texts and images do not require much domain knowledge. In contrast, the generation of electromagnetic data requires necessary electromagnetic knowledge, and it is difficult to format electromagnetic data as there are enormous categories of electromagnetic geometries and electromagnetic responses with various data types and shapes. Consequently, there are barely any online archives that share electromagnetic data, and it is impossible to outsource electromagnetic data acquisition. Most researchers acquire electromagnetic data by performing the full-wave simulation on their own. The simulation process is computationally expensive and time-consuming, es-

pecially as the scale and complexity of electromagnetic structures increase. Therefore, reducing the simulation data requirement while maintaining the model accuracy is meaningful for machine learning-assisted design and analysis of electromagnetic structures.

Typical methods in existing literature initially define all parameter settings through either uniform sampling [18] or Latin hypercube sampling [19]. Subsequently, individual simulations corresponding to these parameter settings are conducted. In conventional uniform sampling, parameters are swept uniformly at a fine step across the entire parameter range. Latin hypercube sampling adjusts the sampling of parameter settings according to the specified parameter space. However, the variation of electromagnetic responses is unstable and non-uniform as the parameter settings change. Uniform sampling and Latin hypercube sampling only consider the distribution of the parameter range, disregarding the distribution of electromagnetic responses. Their collected data have unstable quality, which means some data offer limited contributions to the model convergence, hence involving redundant simulations. Data that have limited contributions to the model performance are classified as low-quality data; data that significantly enhance the model accuracy are referred to as high-quality data. A more efficient data acquisition method is needed to reduce redundant simulations by leveraging computational resources to obtain high-quality data.

There have been many data acquisition methods [20, 21], but they are unsuitable for electromagnetic-related machine learning applications. Also, many works [22, 23] have focused on recognizing the most possible parameter areas for fine-tuning the electromagnetic structures. While they have substantially improved the global optimization of electromagnetic structures, their emphasis is on the fast convergence of the optimization process rather than generating a better dataset for improving the training of machine learning models.

2 Research Objectives

Five research objectives are involved in this Ph.D. project: generalized array radiation synthesis, multi-element phased array calibration, inverse design of frequency-selective surfaces, high-quality data acquisition, and efficient machine learning. These five objectives were selected for several reasons.

1. The selected objectives address significant issues in the field of electromagnetic research.

2. Machine learning-assisted electromagnetic design and analysis represent a relatively uncharted territory with numerous unsolved problems to explore.

3. There was no authoritative guidance, and no clear hierarchy of the significance of key research topics existed.

4. It is worthwhile to investigate any unsolved problems in this field.

Consequently, three initial research objectives were chosen from the familiar topics: radiation synthesis, array calibration, and frequency-selective surface design. As the study progressed, it became evident that data acquisition and efficiency pose the main challenges for electromagnetic-related machine learning applications. Therefore, high-quality data acquisition and efficient machine learning were selected as the final objectives.

2.1 Generalized Array Radiation Synthesis

The communication eras beyond the fifth (B5G) and sixth (6G) generations need large-scale antenna arrays with high gain and good directivity. The array radiation pattern is one of the essential responses that needs to be evaluated. The synthesis of array radiation could be a severe problem as the size increases.

The array radiation can be calculated on the basis of the radiation superposition theorem. The radiation superposition theorem suggests that the array radiation approximately equals the composition of all the antenna elements' radiation. The tricky thing is that the radiation of each antenna element distorts due to the coupling effects caused by surrounding antenna elements. The distorted radiation of an antenna element is called its active radiation, and its original radiation is called the ideal radiation. The active radiation of each antenna element varies and depends on its relative position within the array. The coupling effects happen in every untraceable path and are difficult to formulate. The typical radiation superposition theorem omits the coupling effects and considers only the ideal element radiation. The ideal radiation and relative position of each element decides the synthesized array radiation. It is at a low computational cost but lacks accuracy due to the omission of the coupling effects. A radiation pattern example of a 1×4 antenna array under a specific excitation state is calculated by using the radiation superposition theorem, and it is compared to the real radiation pattern measured in an anechoic chamber in Fig. 5. A big difference, caused by the omission of the coupling effects, can be observed between them.

The accurate array radiation can be acquired by full-wave simulation using a full-wave simulator (e.g., CST, HFSS, or FEKO). The coupling effects are fully considered during the simulation, as the simulation considers the array as a whole single structure and meshes and executes enormous calculations for the whole solving space. We acquire a simulated radiation pattern example of the same 1×4 antenna array under the same excitation state using CST and compare it with the measurement result in Fig. 5. A good agreement level can be observed. However, the simulation process is time-consuming,

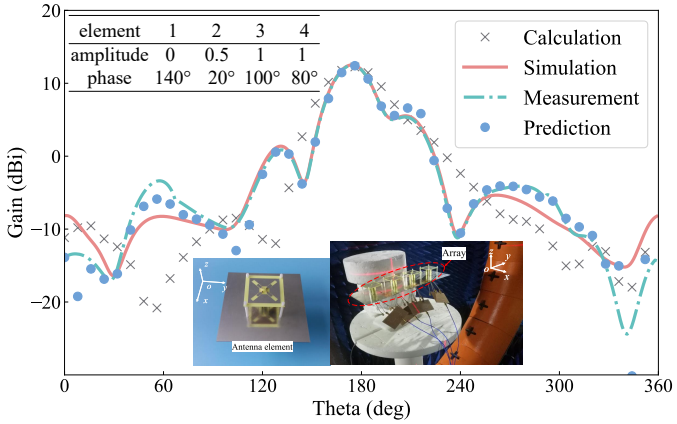


Fig. 5: Comparison of radiation patterns of a 1×4 antenna array that generated through four different ways: calculation based on radiation superposition theorem; simulation using CST; measurement in an anechoic chamber; prediction using an machine learning surrogate model. This figure is taken from paper A, and detailed descriptions can be found there.

computationally expensive, and inefficient, and its cost rises proportionally to the size of the array.

Machine learning can achieve a fast and accurate synthesis of array radiation by developing a surrogate model to mimic the projection from the array excitation state to the array radiation pattern. A radiation pattern example of the same 1×4 antenna array under the same excitation state is generated by using a machine learning surrogate model and is exhibited in Fig. 5. The generated radiation pattern matches well with the simulation result and the real measurement result, which proves its accuracy. Compared to the typical machine learning method that relies on massive simulation data to learn the projection, we propose a two-order neural network to reduce the requirement for simulation data. The first order learns the superposition by training with massive calculation-based synthesis data, and the second order captures the coupling effects by learning from reduced simulation data. A significant reduction of simulation data is expected, compared to the typical machine learning-based synthesis method, while maintaining the same synthesis performance. The corresponding work is presented in Paper A.

2.2 Multi-Element Phased Array Calibration

We mentioned in Section 2.1 that the large-scaled antenna array plays a significant role in the upcoming B5G and 6G communication eras. In addition to radiation synthesis as a challenge during its design stage, array calibration remains another challenge during its operation stage. Unpredictable environ-

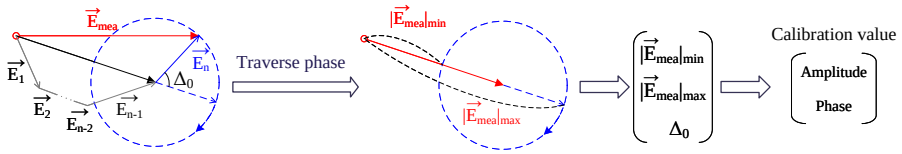


Fig. 6: Rotating element electric field vector method. More details can be found in [24].

mental variations (e.g., temperature, moisture, pressure) and aging effects can severely distort the excitation path and deteriorate the array radiation performance eventually. Engineers ought to calibrate the array periodically to correct the excitation distortions. The main objective of the array calibration is to determine the relative amplitude and phase of each antenna element.

There have been many array calibration methods. They can be categorized into complex methods and power-only methods according to whether they require both amplitude and phase measurement or just amplitude measurement. Power-only methods are more prevalent compared to complex methods, as restrictive phase measurement is not available in some scenarios. The rotating element electric field vector (REV) is a typical power-only calibration method. Depending on whether it allows simultaneous calibration of multiple antenna elements, there is the single-element REV [24] and the multi-element REV [25]. As demonstrated in Fig. 6, the single element REV traverses the phase of each element from 0° to 360°, with the rest of the elements being matched. The maximum and minimum electric power, along with the phase difference between these two states, are recorded for each element. By executing the well-established formulas which can be found in detail in [24], the relative amplitude and phase of each antenna element can be determined. The single-element REV can be extended into the multi-element REV, providing phase shifters with a certain resolution level. The working principle is that the composite measurement results of multiple antenna elements can be decomposed into independent measurement results for each antenna element by keeping their phase shifters in independent phase states. The next step follows the single-element method. To ensure decomposition ability, the maximum number of antenna elements that must be calibrated simultaneously is limited by the resolution of phase shifters. Other than combining the calibration of multiple elements, H.-J. Yoon *et al.* improved REV by reducing the required hardware operations [26]. With 4-bit phase shifters, they proposed a simple and accurate algorithm to select the correct calibration formula and to determine the calibration value. It requires only a small magnitude change in the antenna elements and avoids a complete disconnection of the antenna elements.

We propose a transfer learning-assisted multielement array calibration method. Compared to traditional power-only calibration methods, the pro-

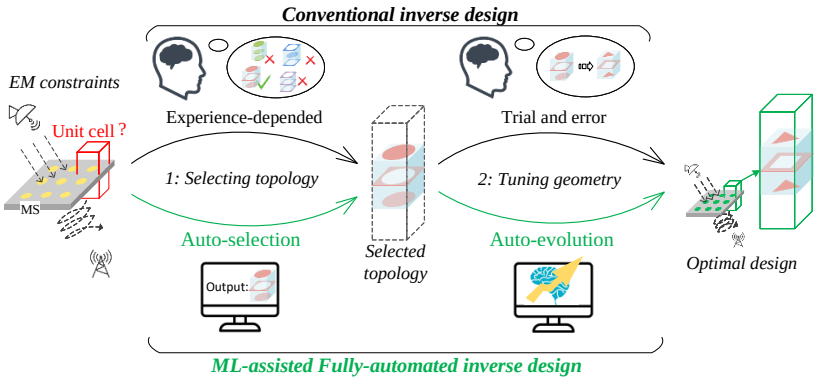


Fig. 7: Comparison of conventional inverse design and proposed machine learning-assisted fully automated inverse design. This figure is taken from paper C, and detailed descriptions can be found there.

posed method imposes no limitation on phase shifters and requires fewer measurement times for calibrating a same-scaled array. Unlike existing machine learning-assisted calibration methods [27–29] that determine only amplitude or phase, our method decides the complex excitation, and we reduce the required simulation data by applying knowledge-transfer learning. The learning burden is partially transferred to the training on theoretical data that takes a lower computation cost to acquire. Therefore, it is expected to reduce the need for simulation significantly. The corresponding work is presented in Paper B.

2.3 Inverse Design of Frequency Selective Surfaces

The inverse design of frequency-selective surfaces is another attractive inverse machine-learning problem in the electromagnetic field. A frequency-selective surface is a planar electromagnetic structure and is composed of many periodical sub-wavelength units. It is widely applied in wireless communication scenarios to manipulate the transmission or reflection characteristics of incident electromagnetic signals. The inverse design of a frequency-selective surface can be summarized mathematically as projecting known transmission or reflection constraints to the geometry of the frequency-selective surface unit. It is usually carried on in two stages, selection of topology and adjustment of geometry. Similar to other electromagnetic structures, the traditional inverse design of frequency-selective surfaces relies on iterative trial and error procedures due to the lack of closed-form formulations for guiding the design choices.

Many machine learning-assisted inverse design methods for frequency-selective surfaces have been proposed. Most of them [30] contribute to the second stage, geometry selection. In their methodologies, a topology is fixed

and a geometry is initialized by human designers according to the electromagnetic constraints. The common objective of their work is to establish a projection from electromagnetic constraints to parameters of the fixed geometry. It is not technically a fully automated inverse design process. Some methods [31] achieved fully automated inverse design by expanding the projection space to cover all representative topologies. The surrogate model is responsible for both selecting the topology and tuning the geometry. It requires an exponentially increasing number of simulation data for training the model, as the expanded space severely complicates the projection.

We design a representation learning-driven framework for the fully automated inverse design of frequency-selective surfaces. A more efficient complete projection is established, as shown in Fig. 7. The topology is auto-selected with the support of principal component analysis and support vector machine classifier, and the geometry is auto-evolved into an optimal design guided by cross-evolution using neural networks and particle swarm optimization. For four classical application scenarios, four frequency-selective surfaces are generated, manufactured, and measured to validate their effectiveness. The reduction of required simulation data proves superior efficiency compared to existing cases. The corresponding work is presented in Paper C.

2.4 High-Quality Data Acquisition

As clarified in Section 1.4.3, data acquisition is a unique problem and the bottleneck for machine learning-assisted electromagnetic methods. The former objectives, in Section 2.1, 2.2, and 2.3, aim to reduce the required simulation data for the design and analysis of specific electromagnetic structures. We mitigate the burden of approximating the projection by utilizing correlated electromagnetic knowledge, such as the radiation superposition theorem. The methodology is not universally compatible but is especially dedicated to specific cases. It could happen in some scenarios where no electromagnetic knowledge is available to simplify the projection. A universal data acquisition methodology is in great demand.

Most machine learning-assisted methods acquire simulation data through uniform sampling or Latin hypercube sampling [32]. Uniform sampling randomly selects geometric parameter settings on a fine and uniform grid within the parameter space, as shown in Fig. 8 (a). The Latin hypercube sampling distributes all the geometrical parameter settings evenly along each dimension of the parameter space. Their sampling results only depend on the parameter space and ignore the electromagnetic responses. Although the geometric parameter settings defined are distributed uniformly or evenly, the distribution of the corresponding electromagnetic responses could be unstable and unbalanced. Electromagnetic responses may change incrementally in one parameter range, yet strongly in another. Data in the incrementally changed parameter

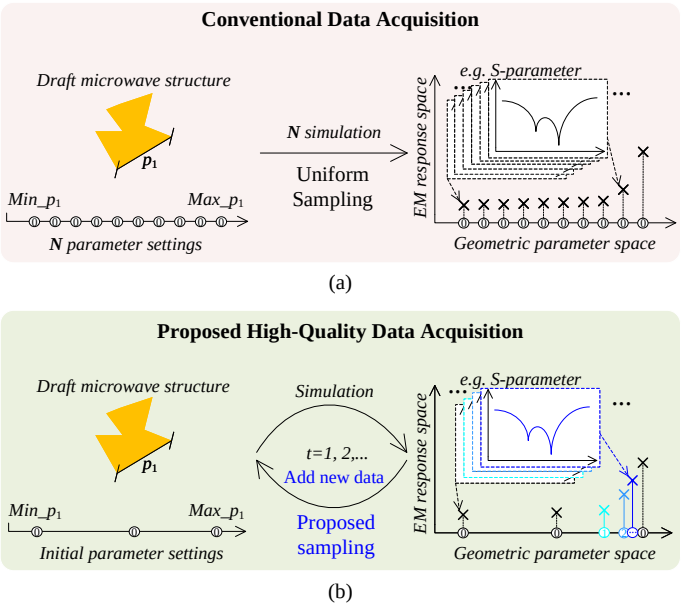


Fig. 8: (a) Conventional data acquisition method; (b) Proposed high-quality data acquisition method.

area contribute little to the model convergence, which is called low-quality data; data in the heavily varied parameter area contribute a lot to the model convergence, hence being called high-quality data. The data acquired using uniform sampling or Latin hypercube sampling are usually of unstable and unbalanced quality. In other words, they involve redundant simulation for training the model.

We propose a high-quality data acquisition method, as demonstrated in Fig. 8 (b). It aims to recognize high-quality data before simulation and leverage computation resources to simulate these high-quality data. We recognize and generate high-quality data iteratively based on the joint analysis of the distributions of both parameter settings and electromagnetic responses. The distribution of parameter settings generated adapts dynamically to that of correlated electromagnetic responses. By maximizing the quality of each simulation data sample, the simulation data required to reach the same model performance is significantly reduced. The corresponding work is presented in Paper D.

2.5 Efficient Machine Learning

There have been many works that managed to improve the efficiency of machine learning methods. Most of these works exploited correlated electromag-

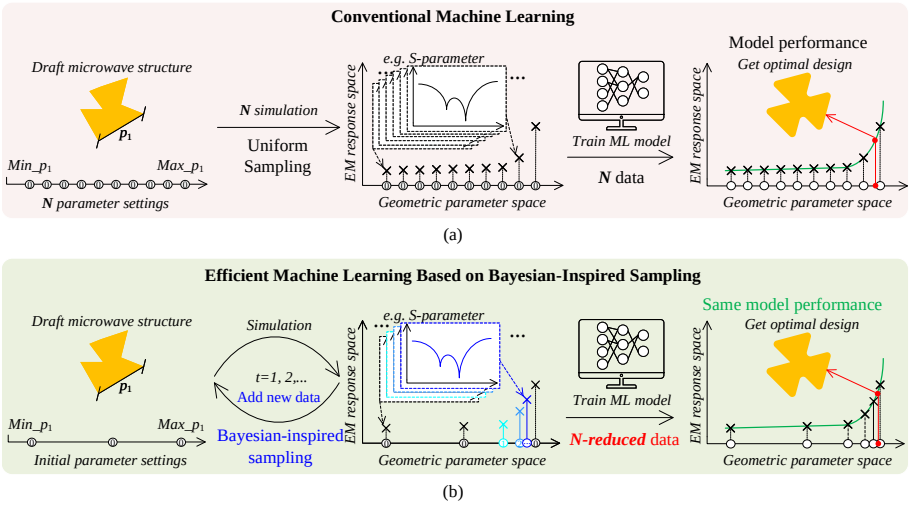


Fig. 9: (a) Conventional machine learning; (b) Efficient machine learning based on Bayesian-inspired sampling.

netic domain knowledge to simplify the complexity of the projection to be approximated, as we mentioned in Section 2.1, 2.2, and 2.3, which are called physical-driven machine learning techniques. For example, J. Zhang *et al.* improved the efficiency of machine learning-assisted design of metasurfaces by incorporating temporal coupled mode theory [33]. However, as we discussed in Section 2.4, these techniques are not universally compatible with various electromagnetic structures since correlated electromagnetic knowledge varies or may not exist. A generally compatible efficient machine learning-assisted method that is suitable for the design and analysis of various electromagnetic structures is needed.

Machine learning methods usually take two major stages: data collection and model training. In Section 2.4, we aim to develop a universal high-quality data acquisition method that improves the efficiency of the data collection. Here, the objective is to improve the overall efficiency of machine learning-assisted methods for satisfying the design and analysis of various electromagnetic structures. Inspired by the Bayesian theorem, we develop an algorithm to improve the overall efficiency by dynamically interacting between the data collection stage and the model training stage. The corresponding work is presented in Paper E.

3 Contributions

In alignment with the five objectives presented in Section 2, the main contributions of this thesis are summarized in five parts and presented in five papers as listed below.

1. Generalized radiation synthesis of antenna arrays is achieved using a two-order deep learning approach. The two-order architecture leverages the theoretical synthesis data to reduce the simulation need for reaching the same synthesis accuracy. Detailed content can be found in Paper A.

2. Fast calibration of phased antenna arrays is realized with a transfer learning-based method. It accelerates the determination of the real excitation amplitude and phase with simplified measurement requirements. Detailed content can be found in Paper B.

3. A fully automated inverse design framework for frequency selective surfaces is proposed based on representation learning. It allows designers without electromagnetic domain knowledge to generate frequency-selective surfaces that fulfill their requirements. Detailed content can be found in Paper C.

4. A high-quality data acquisition method is presented to significantly reduce the simulation need for achieving the same model accuracy. It can be integrated with many machine learning models in various electromagnetic scenarios. Detailed content can be found in Paper D.

5. An efficient machine learning framework is developed with Bayesian-inpsired sampling. It significantly enhances the overall efficiency of machine learning-based electromagnetic design and analysis without compromising the performance. Detailed content can be found in Paper E.

This section briefly introduces the motivation, content, and results of each paper. The complete papers are included in Part II.

3.1 Paper A

Two-Order Deep Learning for Generalized Synthesis of Radiation Patterns for Antenna Arrays
Zhao Zhou, Zhaohui Wei, Jian Ren, Yingzeng Yin, Gert Frølund Pedersen, and Ming Shen
Published in the *IEEE Transactions of Artificial Intelligence*, vol. 4, no. 5, pp. 1359–1368, 2023.

3.1.1 Motivation

Antenna arrays are widely applied in radars, base stations, etc. An antenna array is composed of a number of antenna elements for obtaining high gain and good directivity. Synthesis of the array radiation pattern remains a challenge, as simulating the whole array is computationally expensive and time-consuming, and the rising demand for larger arrays outruns the development of computation resources. In paper A, we develop a two-order deep learning model for fast synthesis of array radiation patterns.

3.1.2 Paper Content

The basic idea is to approximate the projection from the complex array excitation state to the array radiation pattern. By taking advantage of the radiation superposition theorem, we divide the projection into two stages and propose a two-order deep learning model for reducing the simulation requirement and extending the generalization ability. Several typical arrays are selected as training objects. Based on the radiation superposition theorem, enormous theoretically synthesized radiation data of these arrays are collected and used to train a first-order model. Afterward, the first-order model is extended into a two-order model. The two-order model is trained on a reduced number of simulation data. The well-trained two-order model can synthesize the radiation pattern of a wide range of arrays that have a similar topology to the training objects.

3.1.3 Main Results

The proposed two-order model is trained on five 1 by 4 coaxial-fed patch antenna arrays operating at five frequencies. After training, it is tested on an extended range of antenna arrays: coaxial-fed patch antenna arrays, coupling-fed patch antenna arrays, and stacked patch antenna arrays operating at new frequencies. Experimental results validate its high generalization ability and a 50 % reduction of simulation requirement compared with a conventional model.

3.2 Paper B

Transfer-Learning-Assisted Multielement Calibration for Active Phased Antenna Arrays
Zhao Zhou, Zhaohui Wei, Jian Ren, Yingzeng Yin, Gert Frølund Pedersen, and Ming Shen
Published in the *IEEE Transactions on Antennas and Propagation*, vol. 71, no. 2, pp. 1982–1987, 2023.

3.2.1 Motivation

Phased antenna arrays feature flexible adjustment of the radiation direction by precisely controlling the excitation state of each antenna element. Periodic calibration is essential to minimize the distortions of excitation resulting from aging effects, environmental deviations, and other factors. Normally, calibration methods focus on revealing the distortions of the complex array excitation. Among the conventional calibration methods, power-only methods are preferred as they avoid restrictive measurements of complex gains. Typical power-only methods, such as the rotating element electric vector (REV) method, repetitively measure the gains while tuning the phase shifters sequentially. Therefore, it requires either a large number of measurement times or high-resolution phase shifters, since high-resolution phase shifters may allow multiple antenna elements to be calibrated simultaneously. Several machine learning-based methods have been proposed but deal with only amplitude or phase. We propose a multi-element fast calibration method assisted by transfer learning in paper B, which reduces the measurement times and has no requirement for phase shifters.

3.2.2 Paper Content

We develop a surrogate model for determining the complex array excitation state according to the array radiation pattern by integrating well-established antenna theory with machine learning techniques. A novel feature extraction scheme is proposed in paper B to compress the raw array radiation pattern into a compact vector. By reducing the dimension, the amount of required data for reaching the convergence threshold decreases accordingly. The simulation data requirement is further reduced by transferring part of the training responsibility from simulation data to theoretical data. Theoretical data can be easily acquired by formula operations at a significantly lower computation cost. The well-trained surrogate model can reveal the excitation distortions within milliseconds by taking a significantly reduced number of measurement times.

3.2.3 Main Results

The proposed method is implemented on a 1 by 8 linear patch antenna array for validation. The target array is pre-distorted arbitrarily within a certain range for collecting sufficient simulation data for training. The well-trained surrogate model is then tested on a set of unseen distorted simulation data. Experimental results suggest that the proposed method arrives at a comparable calibration accuracy by utilizing a significantly reduced number of measurement times compared with typical calibration methods. Additionally, no limitation is

imposed on the phase shifters, no phase shifting is involved, and no complex equation manipulation is needed.

3.3 Paper C

Representation Learning-Driven Fully Automated Framework for the Inverse Design of Frequency-Selective Surfaces
Zhao Zhou, Zhaohui Wei, Jian Ren, Yingzeng Yin, Gert Frølund Pedersen, and Ming Shen
Published in the *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 6, pp. 2409–2421, 2023.

3.3.1 Motivation

Frequency-selective surfaces are deliberately designed planar structures that have specific transmission or reflection characteristics within a concerned frequency range. Different topologies correlate to different functionalities and fit in different scenarios. The conventional design of frequency-selective surfaces relies on experienced human designers to determine the topology and tune the geometry manually, hence it is labor-intensive and inefficient. There have been many works that automate the tuning process by using machine learning or optimization-based methods. However, most of them still need experienced human designers to determine the topology first, and they require enormous simulation data for developing surrogate models. We propose a representation learning-driven fully-automated inverse design framework. For given electromagnetic constraints, it determines the topology and evolves the geometry into an optimal design automatically.

3.3.2 Paper Content

Driven by representation learning, we establish a machine-friendly projection from known electromagnetic constraints to an optimal frequency-selective surface design. With reference to the design process of human designers, the proposed projection takes two major steps: auto-selection and auto-evolution. Auto-selection starts from electromagnetic constraints, going through discrete sampling, dimensionality reduction, and support vector machine (SVM)-aided classification step by step, and arrives at a proper topology. Auto-evolution initializes the topology as a fan-shaped geometry and evolves it into an optimal design. In order to maximize the efficiency, the evolution is guided and controlled through a novel cross-evolution system supported by the integration of the particle swarm optimization (PSO) and the neural network. The feasibility and efficiency of the proposed framework are validated on several design examples.

3.3.3 Main Results

We validate the proposed framework in four scenarios. Given four sets of pre-defined electromagnetic constraints, a band-pass frequency-selective surface, a dual-band-pass frequency-selective surface, a high-pass frequency-selective surface, and a polarizer frequency-selective surface are generated, fabricated, and measured. Measurement results meet the pre-defined electromagnetic constraints. Compared with several typical automated inverse design methods, the proposed method achieves full automation of the whole inverse design process and requires fewer simulation data. The adjustment of the number of fans extends the availability by allowing complicated designs at low frequencies and compact designs at high frequencies. The proposed framework allows non-experienced users to realize basic functionalities and frees experienced designers up to focus on more meaningful innovation tasks.

3.4 Paper D

A High-Quality Data Acquisition Method for Machine-Learning-Based Design and Analysis of Electromagnetic Structures
Zhao Zhou, Zhaohui Wei, Abdullah Tahir, Jian Ren, Yingzeng Yin, Gert Frølund Pedersen, and Ming Shen
Published in the *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 10, pp. 4295–4306, 2023.

3.4.1 Motivation

Machine learning-assisted electromagnetic methods usually rely on simulation data for developing surrogate models. They are often criticized for taking too much simulation data. The collection of simulation data is time-consuming and computationally expensive, hence deteriorating the efficiency of machine learning-based methods. Conventional data acquisition methods in electromagnetic research typically involve sweeping geometric parameters of an electromagnetic structure on a fine, uniform grid within a specified parameter space. However, a challenge arises due to the nonuniform and unbalanced nature of the variance in electromagnetic responses. Certain areas within the parameter space yield high-quality data, which significantly contribute to model convergence, while other areas produce low-quality data with less impact. Conventional data acquisition methods often generate data of unstable quality, leading to low training efficiency. We propose a high-quality data acquisition method for improving the efficiency of machine learning-assisted electromagnetic methods.

3.4.2 Paper Content

The idea is to estimate the quality of potential simulation data before simulation and allocate computation resources to the generation of high-quality simulation data. For a specific machine learning-based electromagnetic task, we initialize a small set of simulation data by sweeping its key geometrical parameters on a sparse uniform grid within a pre-defined parameter space. The qualities of existing simulation data are evaluated by measuring the variances of their electromagnetic responses. A pair of adjacent simulation data of the highest quality is picked out for determining the parameter setting of a potential high-quality new data sample, and it is simulated then to get its electromagnetic responses. This new data sample is then added to the existing simulation dataset. The evaluation and generation process runs iteratively until sufficient simulation data has been acquired. The final dataset has improved quality and is used to train the surrogate model. Compared with conventional data acquisition methods, the proposed data acquisition method takes fewer simulation data and leads to better model performance.

3.4.3 Main Results

The advancement is validated by comparing the results of using the proposed data acquisition method and the conventional data acquisition method in four implementations. The distribution of generated data is optimized adaptively by the proposed data acquisition method. On average, a 40 % reduction of simulation data requirement is realized for reaching the same model accuracy. The proposed method may be integrated with most machine learning-assisted electromagnetic solutions to reduce the burden of simulation, and it may be utilized for revealing the sensitivity variance of geometrical parameters and provide guidance for understanding the working principle of correlated electromagnetic structures.

3.5 Paper E

Bayesian-Inspired Sampling for Efficient Machine-Learning-Assisted Microwave Component Design
Zhao Zhou, Zhaohui Wei, Jian Ren, Yingzeng Yin, Gert Frølund Pedersen, and Ming Shen
Published in the *IEEE Transactions on Microwave Theory and Techniques*, Early Access, 2023.

3.5.1 Motivation

The efficiency of existing machine learning-assisted electromagnetic methods is severely compromised due to their heavy simulation needs for obtaining

accurate machine learning models. They take two separate steps: data collection and model development. The collected data should provide sufficient informativeness and representativeness of the solving space for training the model. To ensure the informativeness and representativeness of generated data, they often extensively explore the solving space by sweeping over the whole parameter space in a high-resolution grid. However, their simulation cost exceeded the real necessary requirement because they spent many computation resources on redundant simulation data. We propose an efficient machine learning-assisted electromagnetic design method that minimizes the simulation need.

3.5.2 Paper Content

Instead of collecting data and training the model separately, we generate data and train the model in the meantime. Inspired by the Bayesian theorem, the generation of data is dynamically tuned based on analyzing the distribution of the existing dataset and current model accuracy. We develop a Bayesian-inspired algorithm to analyze the distribution of the existing dataset and evaluate the probability of the outcome of adding new data. New data that maximize the probability of improving the model accuracy are generated and added to the training dataset. While the dataset updates, it is used to train the model at the same time. The data generation process stops once the model accuracy meets the minimum need to avoid redundant simulation costs. To validate its effectiveness, we implement the proposed method in three design cases. Experimental results suggest that the proposed method shows superior efficiency compared to existing methods.

3.5.3 Main Results

We propose an efficient machine learning-assisted method based on Bayesian-inspired sampling integrated with neural networks and apply the proposed method in three real design cases to evaluate its efficiency. For comparison, the three design examples are reimplemented using different combinations of several typical machine learning surrogate models (Neural Networks, Support Vector Regression Machine, Generalized Regression Neural Network, and 3-order Polynomial Design Response Network) combined with different sampling techniques (Proposed Sampling, Uniform Sampling, and Latin Hypercube Sampling), respectively. We also investigate the performance of the proposed method when the dimensionality of the design case increases. The experimental and comparative results demonstrate an improved efficiency of the proposed method compared to existing methods.

4 Conclusion

In conclusion, this Ph.D. thesis makes contributions to the field of electromagnetic structures by leveraging machine learning techniques for the design and analysis of such structures.

In paper A, a two-order deep learning strategy is proposed for machine learning-assisted array radiation synthesis. This approach reduces the reliance on simulation data by decomposing the function between array excitation states and synthesized radiations into two orders. The first-order captures fundamental superposition from low-cost calculation-based syntheses, while the second-order incorporates the coupling effect with a reduced simulation dataset. The results demonstrate a significant reduction of approximately 50 % in simulation burden.

Paper B develops a multielement fast calibration method for phased antenna arrays. A novel feature extraction scheme is introduced for organizing and compressing array radiation into compact features during phased array calibration. This reduces redundancy and measurement times. Additionally, knowledge-based transfer learning is employed to minimize the need for extensive simulation data. By leveraging the radiation superposition theorem, the training task is simplified, resulting in reduced simulation data for establishing the backward projection. Experimental results show comparable calibration performance to other methods, with fewer measurement times, no demands for phase shifters, no equation manipulations, and fast execution.

For the inverse design of frequency-selective surfaces, a representation learning-driven framework is developed to enable fully automated design in paper C. This framework includes an auto-selection module that determines the frequency-selective surface topology based on classification and an auto-evolution module that generates optimal designs through cross-evolution using neural networks and particle swarm optimization. The framework is validated through the design, fabrication, and measurement of four frequency-selective surfaces in different scenarios.

In paper D, a generally compatible high-quality data acquisition strategy is proposed to reduce reliance on simulation data. The strategy optimizes the distribution of simulation data to be acquired to maximize model performance. The proposed algorithm recognizes high-quality data before their electromagnetic responses are known and utilizes computational resources to simulate such data. A comparison with the common data acquisition method demonstrates an average reduction of 40 % in the simulation data required to achieve the same model accuracy.

Finally, we develop an efficient machine learning that is universally applicable for the design and analysis of various electromagnetic structures. A Bayesian-inspired algorithm is proposed to minimize the simulation need and

computation cost for the development of machine learning models. A significant improvement of the overall efficiency is achieved compared to existing machine learning techniques incorporating typical sampling strategies.

Overall, the Ph.D. thesis successfully applies machine learning techniques to address various objectives in the design and analysis of electromagnetic structures. The contributions made in three specific electromagnetic scenarios and two universally compatible approaches significantly enhance the efficiency of machine learning-assisted design and analysis of electromagnetic structures.

References

[1] C. A. Balanis, *Antenna theory: analysis and design*. John wiley & sons, 2016.

[2] K. Achouri and C. Caloz, *Electromagnetic metasurfaces: Theory and applications*. John Wiley & Sons, 2021.

[3] R. Levy, R. V. Snyder, and G. Matthaei, "Design of microwave filters," *IEEE Transactions on Microwave Theory and techniques*, vol. 50, no. 3, pp. 783–793, 2002.

[4] J. Zhang, S. Zhang, and G. F. Pedersen, "Dual-band structure reused antenna based on quasi-elliptic bandpass frequency selective surface for 5G application," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 11, pp. 7612–7617, 2020.

[5] Y. Wei, J. Duan, H. Jing, Z. Lyu, J. Hao, Z. Qu, J. Wang, and B. Zhang, "A multiband, polarization-controlled metasurface absorber for electromagnetic energy harvesting and wireless power transfer," *IEEE Transactions Microwave Theory and Techniques*, vol. 70, no. 5, pp. 2861–2871, 2022.

[6] W. Tang, S. Mercader-Pellicer, G. Goussetis, H. Legay, and N. J. Fonseca, "Low-profile compact dual-band unit cell for polarizing surfaces operating in orthogonal polarizations," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 3, pp. 1472–1477, 2017.

[7] H. Shan, K. Jiang, J. Xing, and T. Jiang, "BPSO and staggered triangle layout optimization for wideband rcs reduction of pixelate checkerboard metasurface," *IEEE Transactions Microwave Theory and Techniques*, vol. 70, no. 7, pp. 3406–3414, 2022.

[8] C. C. Nadell, B. Huang, J. M. Malof, and W. J. Padilla, "Deep learning for accelerated all-dielectric metasurface design," *Optics Express*, vol. 27, no. 20, pp. 27 523–27 535, 2019.

[9] L.-Y. Xiao, F.-L. Jin, B.-Z. Wang, Q. H. Liu, and W. Shao, "Efficient inverse extreme learning machine for parametric design of metasurfaces," *IEEE Antennas and Wireless Propagation Letters*, vol. 19, no. 6, pp. 992–996, 2020.

[10] J. A. Hodge, K. V. Mishra, and A. I. Zaghloul, "RF metasurface array design using deep convolutional generative adversarial networks," in *IEEE International Symposium on Phased Array Systems and Technology*, pp. 1–6, 2019.

[11] M. Ohira, H. Deguchi, M. Tsuji, and H. Shigesawa, "Multiband single-layer frequency selective surface optimized by genetic algorithm with geometry-refinement technique," in *IEEE Antennas and Propagation Society International Symposium*, pp. 833–836, 2003.

[12] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *IEEE International Conference on Evolutionary Computation Proceedings*, pp. 69–73, 1998.

[13] Z. Gu, D. Li, Y. Wu, Y. Fan, C. Yu, H. Chen, and E. Li, "A solution to the dilemma for FSS inverse design using generative models," *IEEE Transactions on Antennas and Propagation*, early access, 2023.

[14] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.

[15] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proceedings of International Conference on Learning Representations*, 2014.

[16] C. M. Bishop, *Neural networks for pattern recognition*. Oxford university press, 1995.

[17] Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: A big data-AI integration perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1328–1347, 2019.

[18] R. Zhu, J. Wang, Y. Han, S. Sui, T. Qiu, Y. Jia, M. Feng, X. Wang, L. Zheng, and S. Qu, "Design of aperture-multiplexing metasurfaces via back-propagation neural network: Independent control of orthogonally-polarized waves," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 6, pp. 4569–4575, 2022.

[19] J. C. Helton and F. J. Davis, "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," *Reliability Engineering & System Safety*, vol. 81, no. 1, pp. 23–69, 2003.

[20] J. Bouvette, H.-F. Liu, X. Du, Y. Zhou, A. P. Sikkema, J. da Fonseca Rezende e Mello, B. P. Klemm, R. Huang, R. M. Schaaper, M. J. Borgnia *et al.*, "Beam image-shift accelerated data acquisition for near-atomic resolution single-particle cryo-electron tomography," *Nature Communications*, vol. 12, no. 1, pp. 1–11, 2021.

[21] K. Jeong, M. Babović, V. Gorshkov, J. Kim, O. N. Jensen, and O. Kohlbacher, "FLASHIda enables intelligent data acquisition for top–down proteomics to boost proteoform identification counts," *Nature Communications*, vol. 13, no. 1, pp. 1–12, 2022.

[22] A. Pietrenko-Dabrowska, S. Koziel, and U. Ullah, "Reduced-cost two-level surrogate antenna modeling using domain confinement and response features," *Scientific Reports*, vol. 12, no. 1, Art. no. 1, 2022.

[23] S. Koziel and A. Pietrenko-Dabrowska, "Expedited variable-resolution surrogate modeling of miniaturized microwave passives in confined domains," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 11, pp. 4740–4750, 2022.

[24] M. Tanaka, Y. Matsumoto, S. Kozono, K. Suzuki, S. Yamamoto, and N. Yoshimura, "On-orbit measurement of phased arrays in satellites by rotating element electric field vector method," *Electronics and Communications in Japan*, vol. 81, no. 1, pp. 1–13, 1998.

[25] T. Takahashi, Y. Konishi, S. Makino, H. Ohmine, and H. Nakaguro, "Fast measurement technique for phased array calibration," *IEEE Transactions on Antennas and Propagation*, vol, 56, no. 7, pp. 1888–1899, 2008.

[26] H.-J. Yoon and B.-W. Min, "Improved rotating-element electric-field vector method for fast far-field phased array calibration," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 11, pp. 8021–8026, 2021.

[27] R. G. Ayestaran, F. Las-Heras, and L. F. Herran, "High-accuracy neural-network-based array synthesis including element coupling," *IEEE Antennas and Wireless Propagation Letters*, vol. 5, pp. 45–48, 2006.

[28] R. Lovato and X. Gong, "Phased antenna array beamforming using convolutional neural networks," in *IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting* IEEE, 2019, pp. 1247–1248.

[29] J. H. Kim and S. W. Choi, "A deep learning-based approach for radiation pattern synthesis of an array antenna," *IEEE Access*, vol. 8, pp. 226 059–226 063, 2020.

[30] M. Abdullah and S. Koziel, "Supervised-learning-based development of multibit RCS-reduced coding metasurfaces," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 1, pp. 264–274, 2022.

[31] P. Naseri and S. V. Hum, "A generative machine learning-based approach for inverse design of multilayer metasurfaces," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 9, pp. 5725–5739, 2021.

[32] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239–245, 1979.

[33] J. Zhang, J. W. You, F. Feng, W. Na, Z. C. Lou, Q.-J. Zhang, and T. J. Cui, "Physics-driven machine-learning approach incorporating temporal coupled mode theory for intelligent design of metasurfaces," *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 7, pp. 2875–2887, 2023.

Part II

Papers

Paper A

Two-Order Deep Learning for Generalized Synthesis of Radiation Patterns for Antenna Arrays

Zhao Zhou, Zhaohui Wei, Jian Ren, Yingzeng Yin, Gert Frølund Pedersen, and Ming Shen

The paper has been published in the
IEEE Transactions of Artificial Intelligence Vol. 4(5), pp. 1359–1368, 2023.

The layout has been revised.

Abstract

This letter tackles the generalized synthesis of antenna arrays using two-order deep learning. Existing deep learning-assisted antenna synthesis approaches mainly rely on model training using electromagnetic (EM) simulation data and hence feature limited generalization ability and the need for a huge amount of EM simulations. The proposed two-order deep learning method uses the first-order model to learn the generic features of radiation patterns from the data efficiently generated by applying conventional array factors. After that, the second-order model learns from EM simulations to capture the detailed pattern variations due to concrete coupling effects in the case of a specific array arrangement with different operating frequencies, radiation structures, and feeding schemes. Therefore the two-order DL model can predict the radiation patterns of a series of antenna arrays while reducing the needed amount of EM simulation data. Implementation was carried out on a series of patch antenna arrays to verify the feasibility and robustness of the proposed approach. The validation includes conditions for the array operating at arbitrary new frequencies, with modified radiation structures or new feeding schemes. The results show that the proposed two-order model provides a prediction accuracy of about 84% for a series of 1 by 4 antenna arrays, clearly outperforming the existing regular one-order DL model, which obtains around 65% with the same EM simulation data. The proposed method reveals a promising direction for applying deep learning to assist the design and analysis of antenna arrays.

1 Introduction

Deep learning (DL) has shown the potential to solve complex, especially non-analytic problems, such as face or speech recognition and image classification. The DL-based beyond-5G/6G (beyond-fifth/sixth communication generation) solutions have significantly inspired both academia and industry, including smart wave management, imperfect channel estimation [1–10], direction-of-arrival estimation [11–13], inverse scattering [14–20], inverse design of microwave components [21–23], and array radiation synthesis [24–28].

Wireless communication systems have been widely applied in many scenarios, such as self-driving vehicles, remote surgery, home automation, indoor localization, etc. [29]. Wireless communication systems rely on a well-behaved antenna array. In the design and analysis of antenna arrays, the array radiation is a core indicator. To constantly monitor this indicator, the whole analysis process will witness repetitive procedures of array radiation synthesis. Thus, a convenient generalized array synthesis method is required. A traditional way is to calculate radiation with the element radiation pattern and excitation condition based on the array factor. It is convenient and computationally low cost, yet inaccurate to describe non-ideal array characteristics such as coupling effects, which can hardly be generalized with closed-loop expressions. Thus,

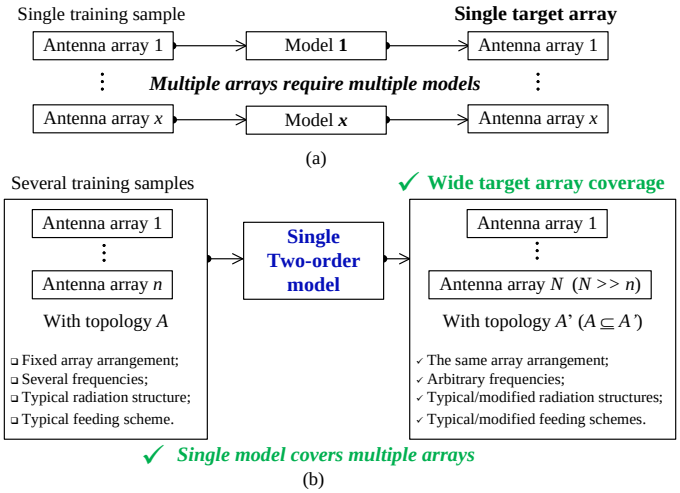


Fig. A.1: Comparison of (a) published regular DL-based method and (b) the proposed DL-based two-order method for antenna array radiation synthesis.

the oriented array must be modeled and simulated via EM simulation software to synthesize its radiation pattern. EM simulation provides high accuracy by paying a high computation cost. On the one hand, the computation complexity might be tolerable for small-scale antenna arrays but is sometimes tricky to handle with ordinary hardware when it comes to large-scale antenna arrays.

On the other hand, its computation is non-reusable because any tiny modification to the structure could void the existing results and require a new simulation. In general, it could take many changes to reach an optimal design, and thus often, many simulation iterations are needed. In this way, deep learning may help significantly accelerate the whole design process by re-using the former simulation results. After being trained with the former simulation results, deep neural networks can predict the future array's radiation patterns at a low computation cost.

Radial basis function neural networks were employed in [24] to estimate the directivity for linear dipole arrays. Similarly, the authors in [25] used a multibranch ANN to solve the inverse problem of the antenna array directivity. However, in most cases, we need more detailed information on the whole array radiation pattern to evaluate beamwidth, sidelobe suppression, etc.

The authors of [26–28] managed to apply deep learning for radiation synthesis, where [26] presented a neural network-based (NN) method capable of establishing a relation between input voltages for all elements and radiation patterns of the target array. The NN they used as a black box considered the interference between elements and possible obstacles positioning within the radiating area. The solution published in [27] focused on the mutual coupling

between antenna elements. NN was proven able to exploit the hidden principle of radiation to predict the voltages needed to generate given radiation patterns. Instead of NN, [28] investigated the possibility of applying a deep neural network (DNN) for antenna array radiation synthesis. All the three methods from [26–28] can imitate the mapping between the excitation conditions and the antenna arrays' radiation patterns. However, as illustrated in Fig. A.1(a), they are single-array-oriented methods; the well-trained models can only exclusively serve one specific antenna array. That means every array calls for an exclusive model, and we need to perform the whole process of model generation for each array. For the sake of higher efficiency, generalized synthesis methods are in great demand, which is precisely the motivation for us to develop a wide-array-coverage-oriented generalized synthesis method.

This letter proposes a wide-array-coverage-oriented deep learning solution for the generalized antenna array radiation synthesis. We have two main contributions: 1) compared to the regular model only serving for one exclusive array, the proposed two-order model inspired by electromagnetic theory can offer generalized radiation synthesis for a series of antenna arrays once being trained with several array samples, as demonstrated in Fig. A.1; 2) unlike most existing deep learning methods heavily relying on simulation, we significantly alleviate the burden on simulation by taking advantage of well-established array factors [30]. As the name suggests, the two-order model incorporates a two-order training process. In advance, we train a conventional fully-connected deep neural network to grab the generic radiation patterns of a group of target antenna arrays, with massive theoretical data handily collected via array factors. The model is then elevated into a two-order model and trained using EM simulation data to improve the accuracy for customized array designs. Thus, the first-order model can learn the basic features of radiation synthesis. The second-order model can then capture the detailed radiation variations due to concrete coupling effects in the customized arrays: specific array arrangements with different operating frequencies, radiation structures, and feeding schemes. We verified the proposed method's feasibility, performance, and robustness by implementing a series of patch antenna arrays. The experiment took five 1 by 4 coaxial-fed patch antenna arrays with elements matched to different levels at different frequencies as source samples. The output model achieves an accuracy of around 84 % defined in section 3.2, while the regular model reaches only 65 %. Also, compared to the regular model only serving for one single target array, the two-order model can cover a series of arbitrary 1 by 4 antenna arrays with similar topology: other coaxial-fed patch antenna arrays at new frequencies, coaxial-fed stacked patch antenna arrays, or coupling-fed patch antenna arrays at arbitrary frequencies. The proposed method can be a competitive alternative tool for designing and analyzing antenna arrays due to its merits of high generalized synthesis capability and low computation cost.

The content is arranged as follows. Section 2 demonstrates the principle and the workflow of the presented method in detail, followed by instruction and discussion of the implementation on patch antenna arrays in section 3. The conclusion is drawn in section 4.

2 Two-order DL-based Method

2.1 Workflow

The workflow of the regular DL-based method and the proposed two-order method are illustrated and compared in Fig. A.2. As seen in Fig. A.2(a), the conventional DL methods often target one single array and feature at direct rough training with the array excitation conditions and the simulated array radiation of the target sample. Consequently, a demanding simulation workload is obliged to yield the regular models; the standard models are poorly generalized and only applicable to a single target array.

The major improvement of our method is to include several arrays as training samples and propose a two-order training method to arrive at a highly generalized two-order model. Several typical arrays instead of a single target array as training samples ensure a supply of more informative, representative, and diverse array synthesis knowledge. This richer knowledge, of course, upgrades the complexity of learning. For this reason, the two-order learning strategy is proposed as shown in Fig. A.2(b).

Inspired and theoretically supported by the conventional EM knowledge, we degrade the complicated learning object to two unmixed tasks: grasping the generic characteristics of radiation synthesis; capturing the detailed radiation variations. The general rules of radiation synthesis hide behind the array factor. Accordingly, to start with, we collect sufficient theoretical data by applying the array factor to train a first-order model. After that, we simulate several customized arrays to gather the complete variated radiation. A two-order model is elevated based on the well-trained first-order model to focus on learning the radiation variances during the second order. Under the first-order model, the two-order model can quickly capture the personalized radiation variations caused by substantial coupling effects in the customized arrays.

Overall, the two-order learning strategy significantly alleviates the burden on simulation and renders the final two-order model with higher generalization ability. As illustrated in Fig. A.2(b), its workflow contains three main procedures: data processing, first-order training, and second-order training, which will be demonstrated in detail below.

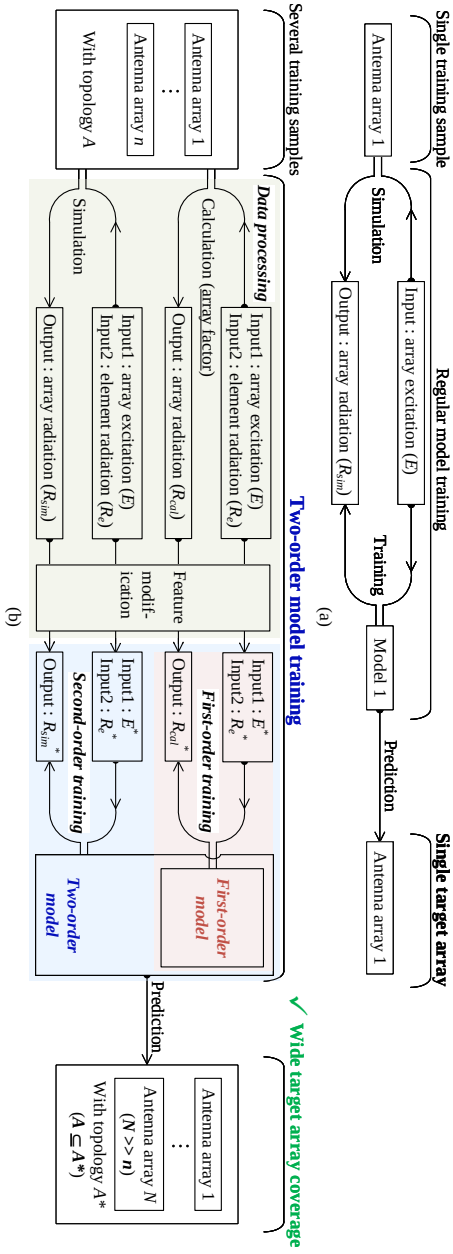


Fig. A.2: The principle diagram of (a) regular method and (b) the proposed two-order method.

2.2 Data Processing

For radiation synthesis of a series of antenna arrays with similar topology, only a limited number of different antenna arrays need to be taken along with

their elements as source samples for training. These source samples, utilizing a specific array arrangement, radiation structure, and feeding scheme, work at different frequencies and are matched to different levels to represent most of the target arrays adequately. These source samples then need to go through a calculation based on well-established array factors [30] to collect data for the first-order training and EM simulation for the second-order model.

The data sets for training either model consist of given excitation signals and element radiation patterns as input, and corresponding array radiation patterns as the output label. The only difference is that the calculation carries only the array factor, as seen in Fig. A.2(b). The data for training the first-order model contains only generic features related to array factors. In contrast, the data for the second order considers the detailed radiation variations due to concrete coupling effects in the case of a specific array arrangement with different operating frequencies, radiation structures, and feeding schemes. The data from calculation or simulation need to be modified to highlight the features, facilitating the following training. The modification of the excitation signals and radiation patterns differs since they have different physical meanings and data characteristics.

For each set of excitation signals (E), a standard definition is n amplitudes (I) and phases (α):

$$E = [I_1, ..., I_n, \alpha_1, ..., \alpha_n], \tag{A.1}$$

where n is the number of ports, I could be any non-negative value, and α may range from 0° to 360°. So, E can have large values and a large positive mean. However, neural networks prefer small values and a zero mean to ensure advanced stability. Thus, E should better be transformed into the complex format by applying Euler's formula, as expressed in:

$$\begin{aligned} E^* &= [Re(Ie^{j\alpha}), Im(Ie^{j\alpha})] \\ &= [[I_1 \cos\alpha_1, I_2 \cos\alpha_2, ..., I_n \cos\alpha_n] \\ &\quad [I_1 \sin\alpha_1, I_2 \sin\alpha_2, ..., I_n \sin\alpha_n]]. \end{aligned} \tag{A.2}$$

After transformation, the mean value for excitation (E^*) would be close to zero, and then, after normalization, its magnitude value would be between 0 and 1.

We only take the radiation pattern in one elevation plane where the elements are aligned since it dominates in the linear array. More elevational or azimuthal radiations can be added in case detailed performance matters. The radiation values are usually in decibel format (dB). The decibel format visually shows more precise details within the whole range than the linear format since the proportion of maximum and minimum values is smaller in the decibel format. However, in practice, the values within the main lobe have higher priority than others. Thus, it is better to use linear data to train the model to achieve higher performance within the main lobe. Then, after normalization,

the values out of the main lobe would significantly approach zero, and the rest be small positives.

After these modifications, the data sets emphasized their features and are now ready for training. If not explicitly noted, all the following data sets from now on refer to the modified data sets.

2.3 First-order Training

During training, the first step is using the calculation data to learn the generic features of radiation synthesis. As mentioned above, the data sets for the first-order model are array radiation patterns calculated from the array factor along with given excitation signals and element radiation patterns. Noted that modified array factor or other available formulas can also be applied if adaptive or customized performance is required. For each data set, a set of excitation signals and the element radiation patterns are taken as input; the corresponding calculated array radiation pattern is as output label. Numerically, each input contains two 1 by n (n is the number of elements) vectors for real and imaginary parts of excitation and another vector for element radiation pattern. At the same time, each output is a similar vector for the array radiation patterns. The vector for either element or array radiation pattern can be one-dimensional in the case of the linear array and two-dimensional for the planar array.

With sufficient processed calculation data, a fully-connected DNN, as shown in Fig. A.3, is employed to learn the generic features of radiation synthesis. The basic architecture of the first-order model consists of an input layer, an output layer, and several Dense layers as the hidden layers. For optimal performance, possible adjustments include the number of layers, each layer's size and activation function, the optimizer, and other available techniques. Here, the adaptive moment estimate (Adam) is recommended as an optimizer. Adam proposed in [31] combines the advantages of AdaGrad [32] and RMSProp [33] and is well-suited to optimization problems with high-dimensional parameter spaces:

$$
\begin{aligned}
Initialize &: \beta_1,\ \beta_2,\ \alpha,\ \epsilon,\ g_0,\ G_0,\ G_0^2,\\
Gradient &: g_i = \nabla Y_i = \nabla F(w_{i-1}^T \cdot X + b_{i-1}),\\
1_{st} moment &: G_i = [\beta_1 \cdot G_{i-1} + (1-\beta_1)\cdot g_i]/(1-\beta_1^i),\\
2_{nd} moment &: G_i^2 = [\beta_2 \cdot G_{i-1}^2 + (1-\beta_2)\cdot g_i^2]/(1-\beta_2^i),\\
Update &: w_i = w_{i-1} - \alpha \cdot G_i/(\sqrt{G_i^2} + \epsilon).
\end{aligned}
$$

Rectified Linear Unit (ReLU) [34] is taken as activation for each layer, except Linear for the last layer. We choose ReLU because our expected output, the

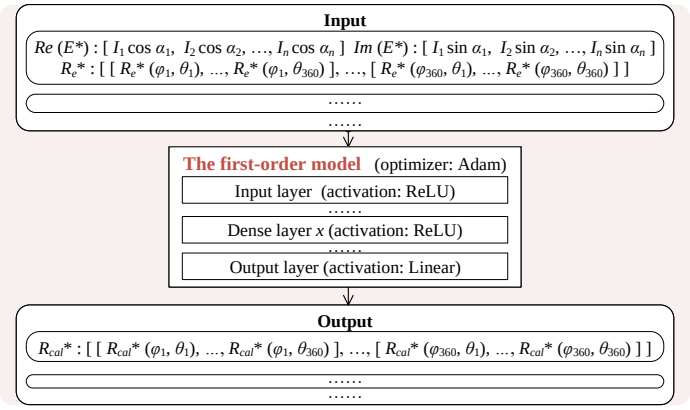


Fig. A.3: The principle diagram of the first-order model.

radiation strength, should be positive and continuous float values. At the same time, ReLU is exactly defined as the positive part of its argument:

$$F(X) = Max(0,\ w^T \cdot X + b). \tag{A.3}$$

Also, ReLU can alleviate vanishing gradient problems. After training using appropriate architecture and techniques, the well-trained first-order model holds the generic radiation synthesis features.

2.4 Second-order Training

With the well-trained first-order model as a basis, a two-order model is built to further capture the detailed radiation variations due to coupling effects in the case of specific array arrangements with different operating frequencies, radiation structures, and feeding schemes. To summarize, its rough principle is to extract the coupling effects, then compress and integrate them into the original excitation signals, marked as coupled excitation signals, and, at last, feed them along with the element radiation patterns into the well-trained first-order model. Since the well-trained first-order model has already acquired the generic features of radiation synthesis, the final two-order model can better capture the detailed radiation variations due to substantial coupling effects. The absolute architecture of the two-order model is exhibited in Fig. A.4, where sharp rectangles represent structures of the model, and rounded ones hold the input, intermediate, or output data.

In practice, as seen in Fig. A.4, the extraction and compression of the coupling effects are accomplished by using two separate DNNs with similar architectures: an input layer, an output layer, and several Dense layers as the hidden layers. As for the optimizer and activation function, the consideration is

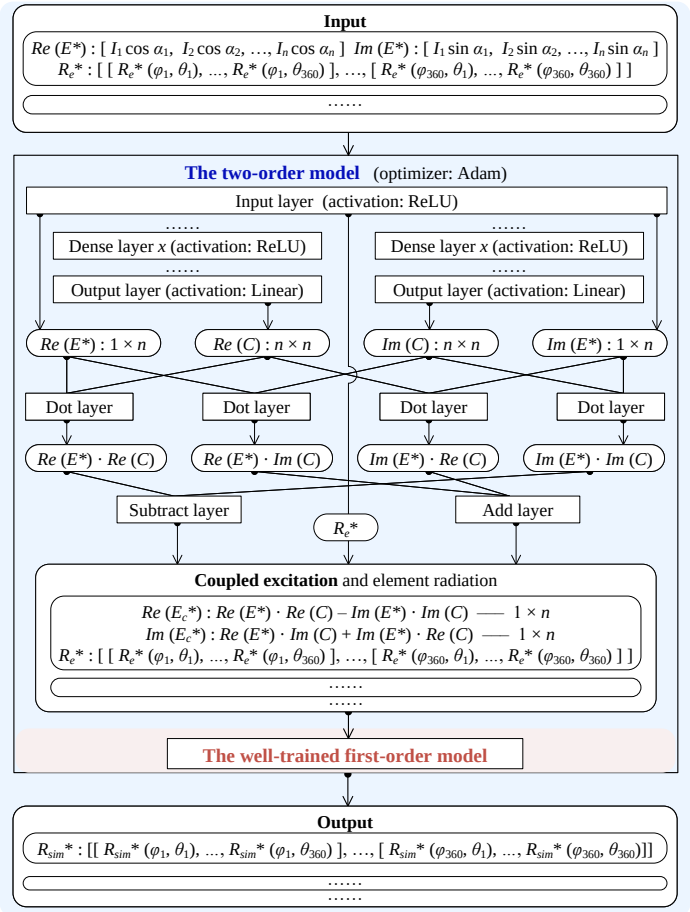


Fig. A.4: The principle diagram of the final two-order model.

the same as what we discussed for the first-order model since these two-order models share the same final goal: Adam is again utilized as the optimizer; ReLU is taken as the activation function.

These two DNNs manage to reveal and compress the coupling information behind the input−−the excitation signals (E^*) and the element radiation patterns (R_e^*)−−into two n by n vectors, standing for the real ($Re(C)$) and imaginary ($Im(C)$) parts of an assumed complex coupling factor (C), respectively. The size of the coupling factor being confined as n by n is supported by a reasonable assumption that, given any certain excitation condition, the coupling effect between every two elements can be approximately transferred and generalized into a complex port-to-port coefficient. For example, the two values at row i and column j of $Re(C)$ and $Im(C)$ together evaluate the coupling influence of

element i on element j and vice versa.

The two DNNs are then followed with Dot layer, Add layer, and Subtract layer as shown in Fig. A.4. These operation layers are employed to execute the complex multiplication of E^* and C. This step is to integrate the extracted coupling factor into the original excitation signals, forming so-called coupled excitation signals. The actual multiplication here is divided into several separate operations on their real and imaginary vectors ($Re(E^*)$, $Im(E^*)$, $Re(C)$, $Im(C)$).

Finally, the well-trained first-order model is attached to the posterior end of the whole architecture. The coupled excitation signals yielded from the afore-mentioned operation layers, along with the element radiation patterns, are fed into the well-trained first-order model for the second-order training. The actual radiation patterns from simulation or measurement results are used as label data. For simulation, we set up the array in the virtual EM environment supported by Computer Simulation Technology (CST) Studio Suite® and then proceed with its frequency domain solver to yield the data. Since the well-trained first-order model has mastered the generic features of radiation synthesis in advance, it is easier for the second-order training to capture the detailed radiation variations, probably leading to high generalization ability, performance, and robustness.

3 Implementation on Patch Antenna Arrays

We validated the proposed two-order method's feasibility, high generalization ability, and performance on a series of patch antenna arrays. Expressly, five 1 by 4 coaxial-fed patch antenna arrays at different frequencies, along with their elements, are set as training samples. (We use a 1 by 4 patch antenna array as the test vehicle for no particular reason, consider it an example to testify our method.) The final two-order model was proven superior to the regular deep learning model and capable of generalized synthesis of radiation patterns of a series of 1 by 4 arrays, including coaxial-fed patch antennas at other different frequencies and stacked patch antennas and coupling-fed patch antennas at arbitrary frequencies.

Next, we will demonstrate the training process, including data processing, the first and second-order training, a comparison with the regular deep learning model, and the verification of three cases.

3.1 Data Processing

We simulated five coaxial-fed patch antennas well-matched at different frequencies via CST to collect their element radiation patterns. And they made up five 1 by 4 arrays at around half-wavelength distance. Fig. A.5 shows the

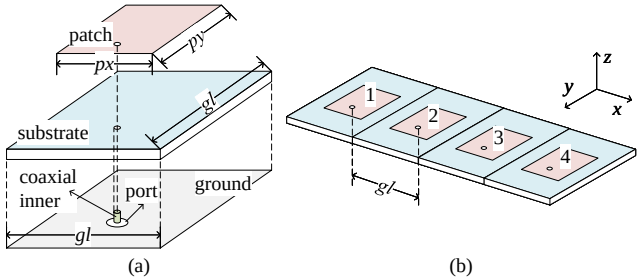


Fig. A.5: The structures of the source samples. (a) Element. (b) Array.
(*Note: The substrate used FR-4 with an ϵ_r of 4.4 and a thickness of 1 mm.)

structure of the element and array. The parameters and characteristics in all five cases are listed in Table A.1. Then, each array went through pattern multiplication based on the array factors to generate 10000 sets of array radiation patterns for the first-order model, with various excitation conditions−−the amplitudes varied from 0 to 1.5 at a step of 0.5 and the phases from 0° to 180° at a step of 20° given. Likewise, it underwent simulation via CST to collect 5000 sets for training the second model. There were 50000 sets for the first order and 25000 sets for the second order. It is worth mentioning that the calculation data outnumbers the simulation data because the calculation takes much less time. It took just 1 minute to collect the 50000 sets of calculation data and roughly 108 hours and 20 minutes to collect 25000 sets of simulation data for the second-order model. The amounts of data sets for the first and second orders are negotiable. In general, more data sets probably lead to higher accuracy. Thus, the portions here mainly depend on the tolerance range of accuracy. Then, as demonstrated in section 2.2, we modified the collected raw data sets to highlight the features before training. The data set for each excitation condition was two 1 by 4 vectors, and each radiation pattern of either the element or array was 1 by 360. Thus, each input was two 1 by 4 vectors along with one 1 by 360 vector, and the output was one 1 by 360 vector for either the first or the second-order model training.

3.2 First-order Training

Within the 50000 sets of calculated array radiation patterns, we took 70 %, 35000 sets as output labels, their corresponding excitation conditions and element radiation patterns as input for the first-order training, and the rest 30 % data for testing. After careful adjustment, the first-order model was fixed, as illustrated in Fig. A.6. In practice, this model had four hidden layers, holding 750, 550, 1550, and 1350 neurons, respectively. Fig. A.8(a) shows the accuracy during training and testing. The accuracy reached 90.20 % during training and 89.61 % for testing.

Table A.1: The Parameters and Characteristics of Source Sample Elements

Sample No.	Freq. (GHz)	Parameters (mm)				S_{11} (dB)	Gain (dBi)	HPBW (deg)
		gl	*px*	*py*	*df*			
1	0.95	150	83.9	75.2	12	−14.8	6.75	94.3
2	1.94	75	40.8	36.5	7.5	−22.6	6.68	95.3
3	2.96	50	24.1	24	4	−15.9	6.6	96.3
4	4	37.5	17.8	17.7	2.7	−14.2	6.53	97.1
5	4.94	30	17.5	14.1	1.9	−34.9	6.62	93.8

(*Note: Freq. means the operating frequency;
HPBW means half-power beamwidth;
S_{11}s, Gains, and HPBWs are at the operating frequency.)

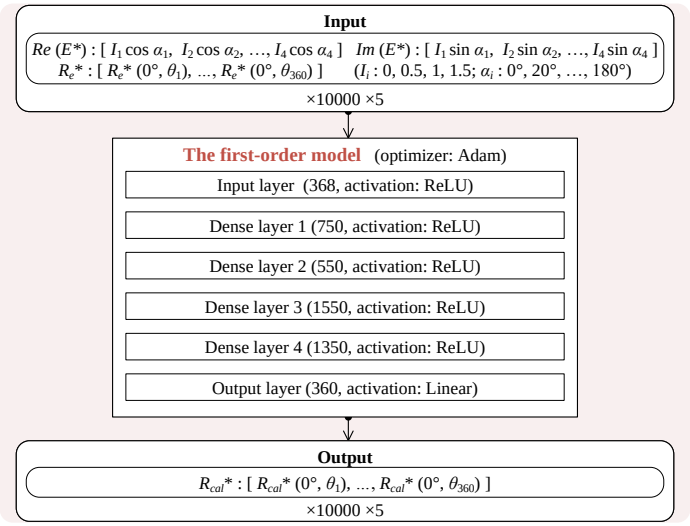


Fig. A.6: The instance structure of the first-order model in this case.

Note that the accuracy mentioned in this article is the proportion of the radiation pattern values where the difference between prediction and simulation meets the predefined criteria, no more than 0.5 dB for positive values and 1 dB negative ones. Thus, although the accuracy may not seem that high, a good agreement has been achieved, verified, and observed in the following comparison between simulation and prediction results.

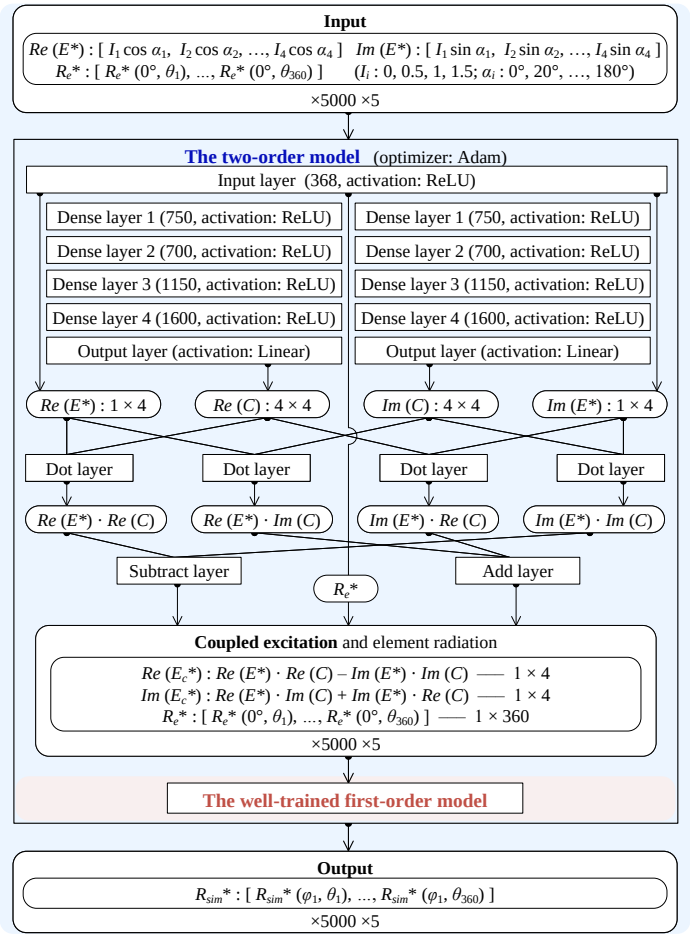


Fig. A.7: The instance structure of the second-order model in this case.

3.3 Second-order Training

Similarly, we took 70 %, 17500 sets of the simulated array radiation patterns as output labels, along with their related excitation conditions and element radiation patterns as input for the second-order training and the rest for testing. With the well-trained first-order model from section 3.2 as a basis, the final two-order model was built and shown in Fig. A.7. In the final model, the sub-models to predict the real and imaginary parts of the assumed coupling factor share the same architecture, also four hidden layers with 750, 700, 1150, and 1600 neurons.

The accuracy of the final two-order model is shown in Fig. A.8(b), which shares the exact definition from section 3.2. The final model showed its op-

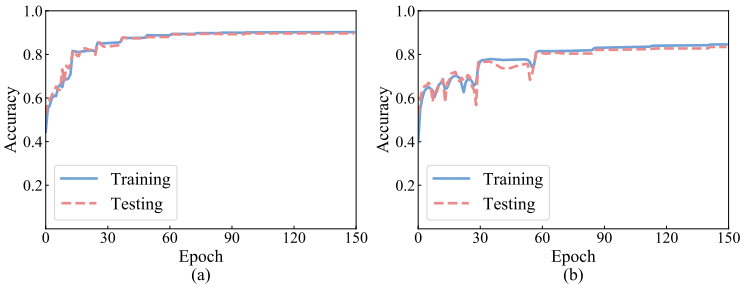


Fig. A.8: The accuracy of (a) the first-order and (b) the second-order model.

timal performance with an accuracy of 84.57 % during training and 83.52 % for testing. One might notice that the accuracy of the final model is lower than that of the first-order model by around 12 %, which is reasonable as the second-order training use much fewer data sets than the first order. But again, the achieved accuracy is considered high according to its definition.

3.4 Verification

The two-order model for radiation synthesis of a series of 1 by 4 patch antenna arrays has been obtained. It was then applied to other 1 by 4 patch antenna arrays than the source samples to verify its performance and robustness. The cases for verification take into account the most typical variations in array designs, like operating frequency, radiator structure, and feeding scheme. In practice, we arbitrarily chose a 1 by 4 coaxial-fed patch antenna array at 3.48 GHz, a 1 by 4 stacked patch antenna array at around 1.5 GHz, and a 1 by 4 coupling-fed patch antenna array at 2.41 GHz for the validation. Also, we validated our model with a published microstrip antenna from [35]; for experimental validation, the proposed method was testified on an antenna array prototype from our published work in [36].

3.4.1 Case I

The first case is a 1 by 4 coaxial-fed patch antenna array working at 3.48 GHz. It has the same structure as the source samples shown in Fig. A.5. The fixed parameters and basic characteristics of its elements are listed in Table A.2. With its element radiation pattern and an arbitrary excitation condition fed into the model, the model was proven to give the predicted array radiation pattern that matches perfectly with the simulation result. Fig. A.9 shows the calculated, simulated, and predicted array radiation patterns given one certain random excitation condition listed beside.

Note that the predicted radiation pattern output from the two-order model

Table A.2: The Parameters and Characteristics of Element in case I

Freq.	Parameters (mm)				S_{11}	Gain	HPBW
(GHz)	*gl*	*px*	*py*	*df*	(dB)	(dBi)	(deg)
3.48	42.8	20.5	20.4	3.2	−14.4	6.56	96.3

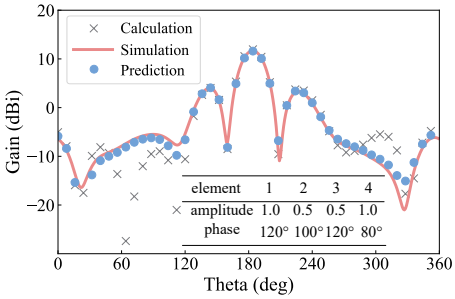


Fig. A.9: A comparison between the calculated, simulated, and predicted array radiation patterns given a certain excitation condition in case I.

is initially in the linear format as we used linear data for training. The prediction is transformed into decibel format to compare with simulation and calculation. A noticeable difference between the simulation and calculation can be observed. The prediction result can ideally correct this difference and agree with the simulation.

3.4.2 Case II

The second case is a 1 by 4 stacked patch antenna array working around 1.5 GHz to test its robustness in terms of radiator structure. It uses a two-layer patch as its radiator, which is widely applied in patch antenna designs to expand its bandwidth, and it is exhibited in Fig. A.10. Since the two-layer patch antenna can resonate at two frequencies, Table A.3 shows the element's parameters and characteristics at their two operating frequencies, 1.47 GHz and 1.53 GHz. And the results for radiation synthesis at both frequencies are presented in Fig. A.11. At both frequencies, the predictions match pretty well with the simulations when excited by the random excitation signals and outperform the calculations.

3.4.3 Case III

In this case, we took a 1 by 4 coupling-fed patch antenna array working at 2.41 GHz to prove its compatibility with feeding techniques. Here, the patch is coupling-fed by an L-shaped strip instead of directly fed by a coaxial cable. This

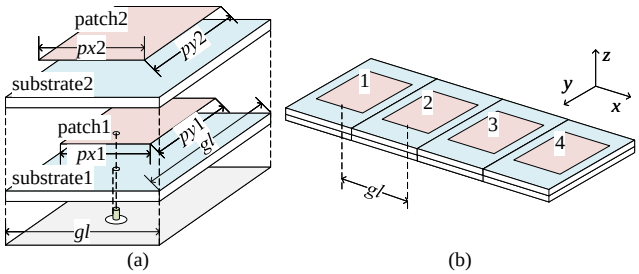


Fig. A.10: The structure of the stacked patch antenna for verification in case II. (a) Radiation element. (b) 1 by 4 array.
(*Note: The substrate1&2 used FR-4 with an ϵ_r of 4.4 and a thickness of 1 mm.)

Table A.3: The Parameters and Characteristics of Element in Case II

Freq.	Parameters (mm)						S_{11}	Gain	HPBW
(GHz)	*gl*	*px*1	*py*1	*px*2	*py*2	*df*	(dB)	(dBi)	(deg)
1.47	100	49	46	50	48	1.6	−11.3	6.56	95.3
1.54	100	49	46	50	48	1.6	−22.4	6.68	92.8

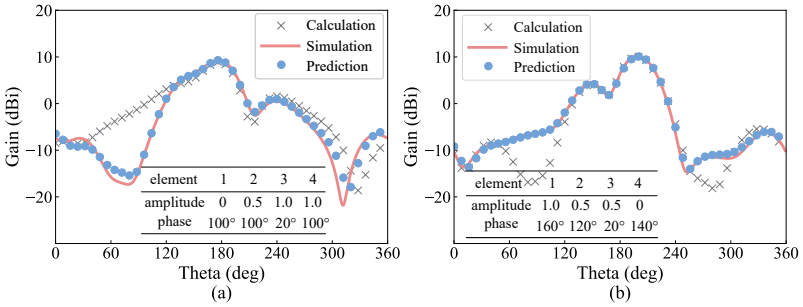


Fig. A.11: A comparison between the calculated, simulated, and predicted stacked patch array radiation patterns given a certain excitation conditions in case II at (a) 1.47 GHz and (b) 1.54 GHz.

coupling-fed approach is a common alternative feeding way for bandwidth enhancement. Its structure, parameters, and characteristics can be found in Fig. A.12 and Table A.4, respectively. And also, Fig. A.13 shows one of the verification results. Generally, a high matching level between the prediction and simulation can be observed, proving that the model works well even with modified feeding techniques.

3.4.4 Case IV

We validated our two-order model on a published microstrip antenna from [35]. It is a classical coaxial-fed microstrip antenna covered with an extra

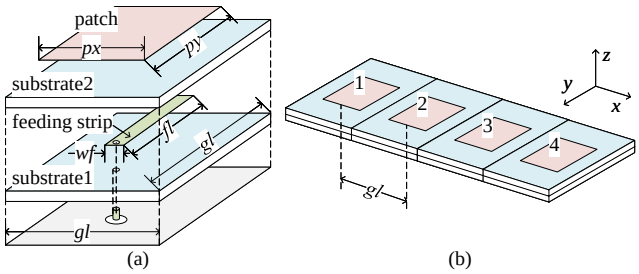


Fig. A.12: The structure of the coupling-fed patch antenna in case III. (a) Element. (b) Array. (*Note: The substrate1&2 used FR-4 with an ϵ_r of 4.4 and a thickness of 1 mm.)

Table A.4: The Parameters and Characteristics of Element in Case III

Freq.	Parameters (mm)						S_{11}	Gain	HPBW
(GHz)	*gl*	*px*	*py*	*wf*	*lf*	*df*	(dB)	(dBi)	(deg)
2.41	60	28.8	28.8	1.5	14	1	−18.3	6.43	98.4

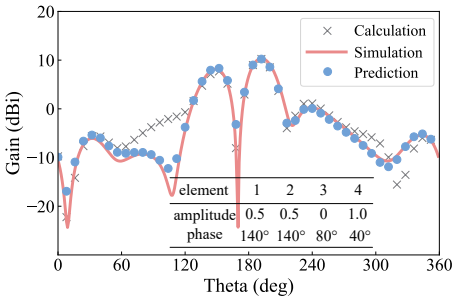


Fig. A.13: A comparison between the calculated, simulated, and predicted coupling-fed array radiation patterns given certain excitation conditions in case III.

dielectric layer to prevent environmental influences, as seen in Fig. A.14. The dielectric layer can also affect the antenna's EM performance by lowering the resonate frequency. According to the coefficients provided in [35], we set up the antenna element and the corresponding array in CST to collect the required simulation data. Then we utilized our model to predict its array radiation. The comparison result is exhibited in Fig. A.15. We can observe that the predicted radiation agrees well with the simulation radiation.

3.4.5 Experimental validation

For experimental validation, we proceeded with the proposed method on an antenna array prototype from our previous work [36]. We have the array prototype measured using a SATIMO multi/probe spherical near-field system.

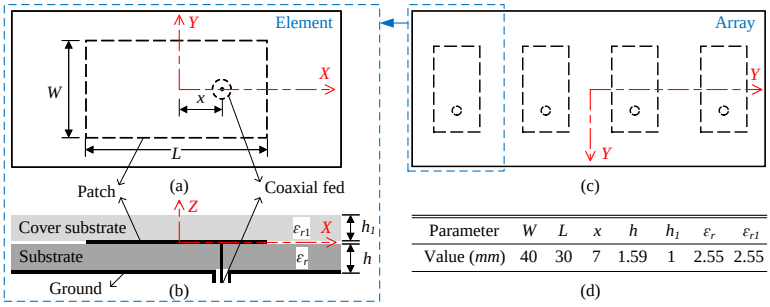


Fig. A.14: The structure of the dielectric-covered patch antenna in case IV: (a)top view and (b) side view of the Element; (c) top view of the array; (d) coefficients table.

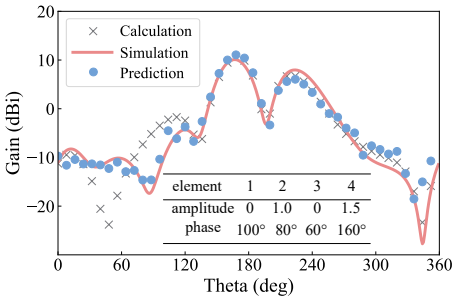


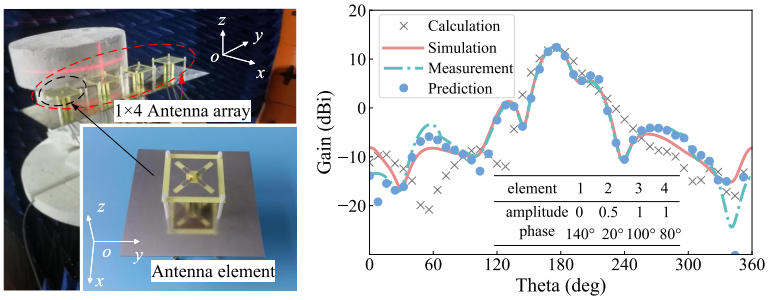
Fig. A.15: A comparison between the calculated, simulated, and predicted coupling-fed array radiation patterns given certain excitation conditions in case IV.

The antenna prototype was initially designed as a slant dual-polarization corresponding to two ports. During the experiment, we fed identical signals to each pair of ports to make it operate in the patch mode. Fig. A.16 compares its calculation, simulation, measurement, and prediction array radiation on an arbitrary excitation state. The prediction here is based on measured element radiation. It is safe to say that the prediction radiation approximately equals the experimental result despite a bit of variance of less than 3 dB outside the main lobe.

We have tested the well-trained two-order model in five cases and verified its high generalization ability, performance, and robustness.

3.5 Comparison and discussion

To quantify our improvement over the existing regular deep learning synthesis methods [26–28], we compare the proposed model to the typical model, as shown in Fig. A.17, regarding the accuracy and the total time consumption. Fig. A.18 exhibits their accuracy when being trained with one array sample;



(a) (b)

Fig. A.16: (a) Experimental setup. (b) A comparison between the calculated, simulated, measured, and predicted coupling-fed array radiation patterns given certain excitation conditions in experimental validation.

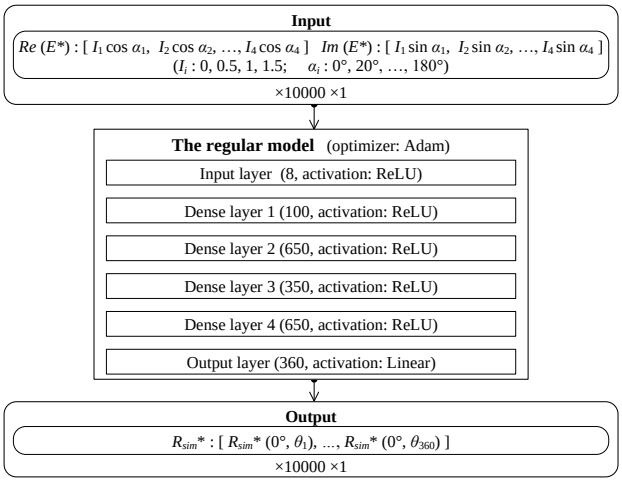


Fig. A.17: The architecture of regular model.

Table A.5 lists the detailed time consumption over their training process.

Our quantitative advantages can be summed up in three aspects. For one thing, we cut down nearly half of the simulation needed to tackle each training sample. It turns out that, with 5000 sets of simulation data per sample, the traditional method can only reach an accuracy up to 65 %. In contrast, the proposed two-order method realizes about 84 %. The standard method requires double the amount (10000 sets) of simulation to achieve the same level of accuracy (around 83 %). Therefore, our method saves nearly half of simulation time for each training sample, which means a lot since the simulation occupies over 99 % of time for the regular model. For the second point, we significantly reduce the number of training samples needed for covering the exact scaled N target arrays, thereby freeing a huge amount (Nn) of simulation workload.

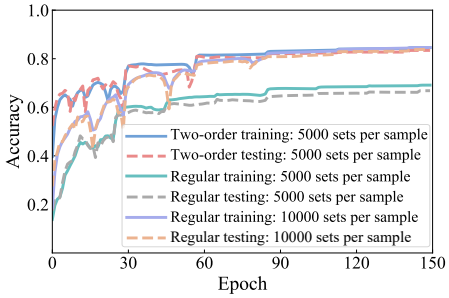


Fig. A.18: The accuracy of regular model and the proposed two-order model.

Table A.5: The comparison of time consumption of regular method and the proposed two-order method

	Target arrays	Training samples	Time			
			Cal. data	Sim. data	Training	In total
Pro.	N	n (= 5)	1m×5	21h40m×5	(18m+16m)×5	21h31m×5
Reg.	N	N	——	43h20m×N	7m×N	43h27m×N

(*Note: Usually $N \gg n$;
Pro. represents the proposed two-order method;
Reg. represents the regular method.)

The last merit is that after a once two-order training process, the two-order model of high generalization ability can solve N target arrays' synthesis problem once and for all. In contrast, the traditional methods have to repetitively produce N models for different target arrays. Here, N depends on the training samples' informativeness, representativeness, and diversity. N can hardly be quantified because the generalized two-order model covers a continuous target array space rather than one single target array like the traditional methods do. Thus, the proposed two-order approach takes much less time to provide a more generalized synthesis solution than the existing regular methods.

This framework can also be expected to outperform traditional methods in the case of the planar array. The main difference is that the planar array must consider radiations over at least two elevation planes instead of one for the linear array. Therefore, we must adopt an advanced feature extraction scheme to compress the complex data. Accordingly, the size of data and the number of hidden layers and neurons should increase.

4 Conclusion

This letter proposed a two-order deep learning method to generalize radiation synthesis of a series of antenna arrays with similar topology. This method uses the well-established array factor to learn the generic features of radiation synthesis using the first order. Then the second order focuses on further capturing the detailed radiation variations caused by concrete coupling effects in the case of a specific array arrangement with different operating frequencies, radiation structures, and feeding schemes. Compared with most published deep learning methods, this method has higher generalization ability, performance, and robustness, which has been testified with an implementation on a series of patch antenna arrays. The two-order model can synthesize a series of 1 by 4 patch antenna arrays with different operating frequencies, radiation structures, and feeding schemes at an accuracy of around 84 %. In beyond-5G/6G scenarios where multiple antenna arrays co-exist in each single base station, we can utilize the proposed model to conveniently monitor all the arrays' radiation patterns.

References

[1] Y. Jin, J. Zhang, S. Jin, and B. Ai, "Channel estimation for cell-free mmwave massive mimo through deep learning," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 10 325–10 329, 2019.

[2] A. S. Alwakeel and A. H. Mehana, "Data-aided channel estimation for multiple-antenna users in massive mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10 752–10 760, 2019.

[3] Y. Liao, Y. Hua, and Y. Cai, "Deep learning based channel estimation algorithm for fast time-varying mimo-ofdm systems," *IEEE Communications Letters*, vol. 24, no. 3, pp. 572–576, 2019.

[4] C. Lu, W. Xu, S. Jin, and K. Wang, "Bit-level optimized neural network for multi-antenna channel quantization," *IEEE Wireless Communications Letters*, vol. 9, no. 1, pp. 87–90, 2019.

[5] X. Wei, C. Hu, and L. Dai, "Deep learning for beamspace channel estimation in millimeter-wave massive mimo systems," *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 182–193, 2021.

[6] E. Balevi, A. Doshi, and J. G. Andrews, "Massive mimo channel estimation with an untrained deep neural network," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2079–2090, 2020.

[7] Y. Zhang, Y. Mu, Y. Liu, T. Zhang, and Y. Qian, "Deep learning-based beamspace channel estimation in mmwave massive mimo systems," *IEEE Wireless Communications Letters*, vol. 9, no. 12, pp. 2212–2215, 2020.

[8] Y. Sui, Y. He, T. Cheng, Y. Huang, and S. Ning, "Broad echo state network for channel prediction in mimo-ofdm systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13 383–13 399, 2020.

[9] M. E. Morocho-Cayamcela, M. Maier, and W. Lim, "Breaking wireless propagation environmental uncertainty with deep learning," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5075–5087, 2020.

[10] W. Ma, C. Qi, Z. Zhang, and J. Cheng, "Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive mimo," *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 2838–2849, 2020.

[11] Z.-M. Liu, C. Zhang, and S. Y. Philip, "Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections," *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 12, pp. 7315–7327, 2018.

[12] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and doa estimation based massive mimo system," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8549–8560, 2018.

[13] Y. Cao, T. Lv, Z. Lin, P. Huang, and F. Lin, "Complex resnet aided doa estimation for near-field mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 11 139–11 151, 2020.

[14] L. Li, L. G. Wang, F. L. Teixeira, C. Liu, A. Nehorai, and T. J. Cui, "Deepnis: Deep neural network for nonlinear electromagnetic inverse scattering," *IEEE Transactions on Antennas and Propagation*, vol. 67, no. 3, pp. 1819–1825, 2018.

[15] H. M. Yao, E. Wei, and L. Jiang, "Two-step enhanced deep learning approach for electromagnetic inverse scattering problems," *IEEE Antennas and Wireless Propagation Letters*, vol. 18, no. 11, pp. 2254–2258, 2019.

[16] Y. Zhou, Y. Zhong, Z. Wei, T. Yin, and X. Chen, "An improved deep learning scheme for solving 2d and 3d inverse scattering problems," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 5, pp. 2853–2863, 2021.

[17] H. Zhou, T. Ouyang, Y. Li, J. Liu, and Q. Liu, "Linear-model-inspired neural network for electromagnetic inverse scattering," *IEEE Antennas and Wireless Propagation Letters*, vol. 19, no. 9, pp. 1536–1540, 2020.

[18] K. Xu, L. Wu, X. Ye, and X. Chen, "Deep learning-based inversion methods for solving inverse scattering problems with phaseless data," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 11, pp. 7457–7470, 2020.

[19] Z. Ma, K. Xu, R. Song, C.-F. Wang, and X. Chen, "Learning-based fast electromagnetic scattering solver through generative adversarial network," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 4, pp. 2194–2208, 2021.

[20] Z. Wei and X. Chen, "Uncertainty quantification in inverse scattering problems with bayesian convolutional neural networks," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 6, pp. 3409–3418, 2021.

[21] L.-Y. Xiao, W. Shao, X. Ding, Q. H. Liu, and W. T. Joines, "Multigrade artificial neural network for the design of finite periodic arrays," *IEEE Transactions on Antennas and Propagation*, vol. 67, no. 5, pp. 3109–3116, 2019.

[22] J. Jiang and J. A. Fan, "Simulator-based training of generative neural networks for the inverse design of metasurfaces," *Nanophotonics*, vol. 9, no. 5, pp. 1059–1069, 2020.

[23] P. Naseri and S. V. Hum, "A generative machine learning-based approach for inverse design of multilayer metasurfaces," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 9, pp. 5725–5739, 2021.

[24] S. Mishra, R. N. Yadav, and R. P. Singh, "Directivity estimations for short dipole antenna arrays using radial basis function neural networks," *IEEE Antennas and Wireless Propagation Letters*, vol. 14, pp. 1219–1222, 2015.

[25] L. Yuan, X.-S. Yang, C. Wang, and B.-Z. Wang, "Multibranch artificial neural network modeling for inverse estimation of antenna array directivity," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 6, pp. 4417–4427, 2020.

[26] R. G. Ayestaran and F. Las-Heras, "Obstacle modeling in array synthesis using neural networks," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 8, pp. 2420–2424, 2006.

[27] R. G. Ayestaran, F. Las-Heras, and L. F. Herrán, "Neural modeling of mutual coupling for antenna array synthesis," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 3, pp. 832–840, 2007.

[28] J. H. Kim and S. W. Choi, "A deep learning-based approach for radiation pattern synthesis of an array antenna," *IEEE Access*, vol. 8, pp. 226 059–226 063, 2020.

[29] A. Jagannath, J. Jagannath, and T. Melodia, "Redefining wireless communication for 6g: Signal processing meets deep learning with deep unfolding," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 6, pp. 528–536, 2021.

[30] C. A. Balanis, *Antenna theory: analysis and design*. John wiley & sons, 2015.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, Y. Bengio and Y. LeCun, Eds., 2015, pp. 1–15.

[32] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization." *Journal of machine learning research*, vol. 12, no. 7, 2011.

[33] T. Tieleman, G. Hinton *et al.*, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, pp. 26–31, 2012.

[34] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *International Conference on Machine Learning*, 2010, pp. 1–8.

[35] G. Kumar and K. P. Ray, *Broadband microstrip antennas*. Artech house, 2003.

[36] Z. Zhou, Z. Wei, Z. Tang, Y. Yin, and J. Ren, "Compact and wideband differentially fed dual-polarized antenna with high common-mode suppression," *IEEE Access*, vol. 7, pp. 108 818–108 826, 2019.

Paper B

Transfer-Learning-Assisted Multielement Calibration for Active Phased Antenna Arrays

Zhao Zhou, Zhaohui Wei, Jian Ren, Yingzeng Yin, Gert Frølund Pedersen, and Ming Shen

The paper has been published in the
IEEE Transactions on Antennas and Propagation Vol. 71(2), pp. 1982–1987, 2023.

The layout has been revised.

Abstract

A transfer learning-based method for accelerating power-only calibration of phased array antennas by combining conventional array theory with deep learning is presented in this paper. Existing power-only calibration methods either require a significant number of measurement cycles or have restrictive phase shifter resolution requirements. The proposed array calibration method uses a surrogate model to calibrate all array elements in one pass without restricting phase resolution requirements. We developed a novel feature extraction scheme (FES) that picks out the most important power features resulting in reduced measurement cycles. The burden of data acquisition for model training is further reduced by relational knowledge transfer learning. The surrogate model acquires its general calibration capability from massive theoretical data, which is easily collected by the radiation multiplication theorem, and captures the detailed non-ideal response from a small number of simulations. The proposed methodology has been demonstrated and tested on several arrays for validation. The effectiveness and performance of the method have been verified, hence it can serve as a complementary tool to accelerate the calibration process of phased antenna arrays.

1 Introduction

Phased antenna arrays are becoming increasingly important and are widely used in wireless communication systems. They rely on precise control of array excitation. In practice, tolerances in component fabrication and aging effects can cause non-negligible distortions. The actual array excitation may deviate from the expected values and depreciate the performance. Therefore, regular calibration must be performed to maintain the performance.

The calibration methods can be divided into two groups: single-element and multi-element methods. The typical single-element methods calibrate only one element in one pass [1–7], and the extended multi-element methods calibrate multiple elements simultaneously during one calibration cycle [8–17]. Moreover, depending on the methodology, they can be further categorized as the mutual coupling-based method [1], the phase toggling method [2, 13], the rotating element electric field vector (REV) method [3–5, 14, 15], and the orthogonal code-based method, etc.

Since the typical single-element calibration methods [1–7] calibrate each element individually, an array antenna with N elements requires at least N calibration cycles. During each calibration cycle, a target element to be calibrated is activated and measured with different phase settings. Meanwhile, the remaining elements are isolated as they are turned off or terminated. Different calibration methods require measurement at different sets of phase settings.

The mutual coupling-based method [1] technically makes a phase adjustment during each measurement cycle. It measures the mutual coupling be-

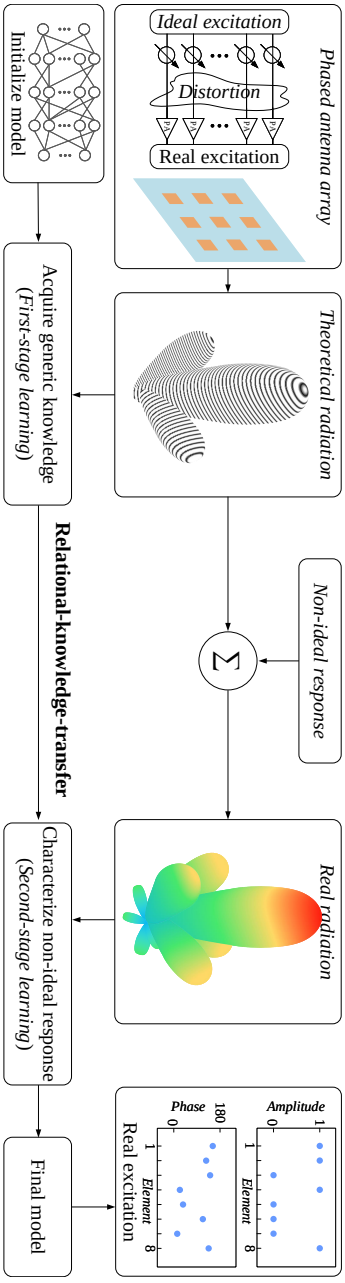


Fig. B.1: The proposed calibration method based on the surrogate model.

tween the target and reference elements to determine the excitation error. The phase toggling method [2] switches the phase shifter from 0 ° to 180 ° and measures complex array signals to determine the error of each element. If

the complex signals are not retrievable, the REV method is successful. The REV method [3–5] traverses all phase states of each element to determine the minimum and maximum power and the phase variation to determine the relative complex excitation field of the element. Typical REV methods require $2^m \times N$ measurement cycles to calibrate an N-element array with an m-bit phase shifter [4]. The phase resolution m must be at least 3 to achieve a clear sinusoidal representation.

The extended calibration methods [13–17] improve calibration efficiency by calibrating multiple elements simultaneously in each calibration cycle. Assuming M elements, a phased array antenna with N elements requires only $\lceil N/M \rceil$ measurement cycles. However, the number of elements in each calibration cycle is limited by the mathematical constraints and the number of bits of the phase shifter.

In [13], up to eight elements are calibrated simultaneously by applying a fast Fourier transform to sixteen measured signals, where the beamformer unit consists of sixteen phase states. The extended REV method [14, 15] successively shifts the phases of multiple elements with different phase intervals and measures the combined power change of the array. Then, the measured combined array power variation is converted into a form identical to the typical REV method using a Fourier transform.

While these extended methods improve the efficiency of calibration, still the improvement is limited due to the high mathematical computation requirements and phase resolution requirements. The number of elements calibrated simultaneously must be limited to ensure mathematical computation ability so that they provide independent contributions to the composite array radiation and can be successfully decomposed. The bit-number of the phase shifter usually determines the ability to be independent. For this reason, the extended methods often have non-negotiable hardware requirements for the bit-number of the phase shifter. In the case of the classic extended REV method, only two elements can be calibrated simultaneously when using a 4-bit phase shifter.

In this communication, a surrogate model is proposed to calibrate all elements of a phased antenna array at once. It requires much fewer measurement cycles than the existing REVs. This powerful technique does not require phase resolution, phase tuning, on-off elements, and equation manipulation. The surrogate model feeds from a few power-only features measured under the default phase state and directly determines the complex excitation field. A feature extraction scheme (FES) was thoroughly investigated to filter out the most informative features while keeping number of features small. As a result, measurement cycles could be significantly reduced. Training of the surrogate model alleviates the tedious data acquisition that existing machine-learning approaches often suffer from [18–20]. Unlike [18–20], which output only amplitude or phase, our model determines both the excitation amplitudes as well as phases for calibration. As shown in Fig. B.1, using relational-knowledge-

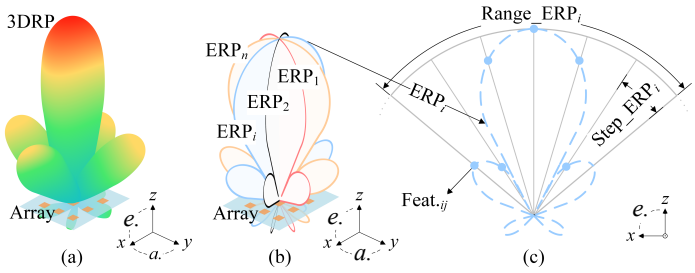


Fig. B.2: (a) 3DRP; (b) ERPs; (c) Sampling the front region of an ERP.

transfer learning, we divide the model training into two stages. The first-stage learning acquires knowledge about the general calibration from the array superposition theorem. The second-stage learning focuses on unexpected concrete properties such as coupling effects and other non-analytical behavior. We have verified the effectiveness and efficiency of the proposed method on several arrays.

The remainder is organized as follows. Section II examines the FES for feature selection. Section III explains the development and optimization of the surrogate model. Validation of the proposed method is presented in Section IV while conclusion is drawn in Section V.

2 Feature Extraction Scheme

The input of the surrogate model is the features extracted by the feature extraction scheme (FES), while the output of the surrogate model is the complex excitation field. This section explains the FES to extract the fewest and most informative radiation features.

2.1 Theoretical Basis

The phased antenna array regularly radiates according to two main rules: (a) the radiation intensity varies more dramatically with elevation (*e.*) angle θ than azimuth (*a.*) angle ϕ. Thus, the elevation radiation pattern (ERP) provides more informative features, and it is reasonable to represent radiation only by collecting multiple ERPs. And, (b) most of the energy is radiated forward, so the most important features of each ERP are mainly confined to the forward region.

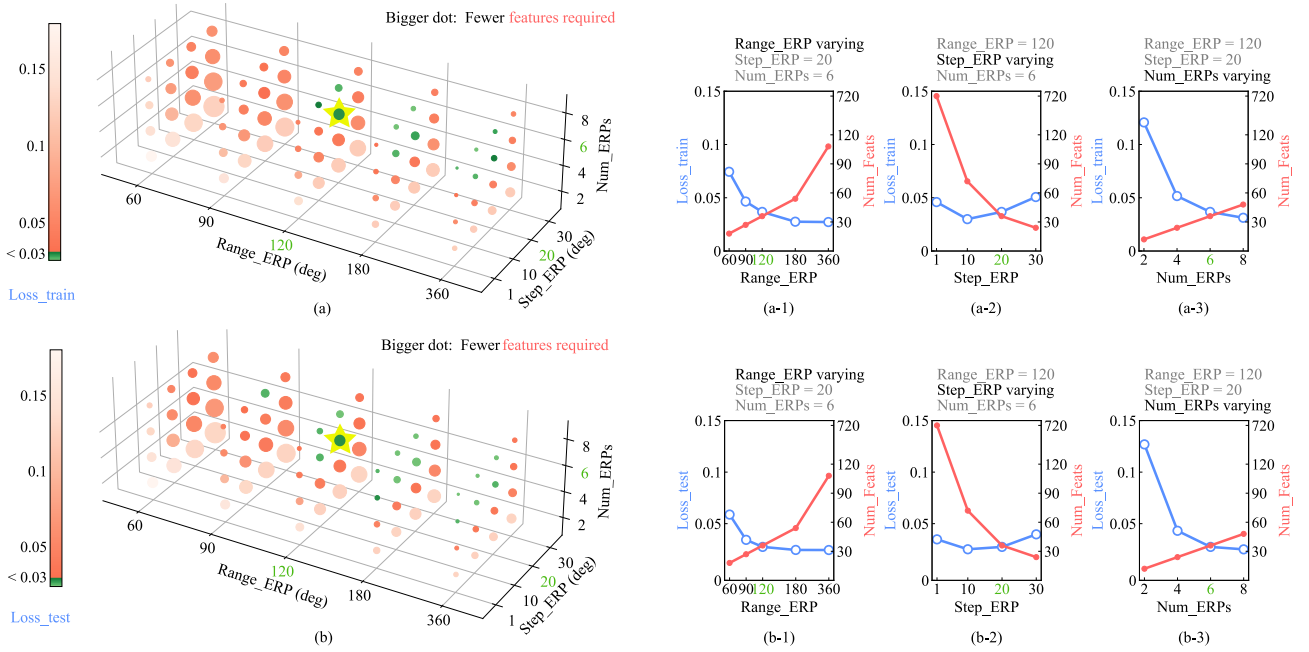


Fig. B.3: *Loss_train* with (a) *Num_ERPs*, *Range_ERP*, and *Step_ERP* varying, (a-1) *Range_ERP* varying, (a-2) *Step_ERP* varying, (a-3) *Num_ERPs* varying; *Loss_test* with (b) *Num_ERPs*, *Range_ERP*, and *Step_ERP* varying, (b-1) *Range_ERP* varying, (b-2) *Step_ERP* varying, (b-3) *Num_ERPs* varying.

2.2 Strategy of FES

The antenna array usually expresses its radiations in a three-dimensional far-field radiation pattern (3DRP) as shown in Fig. B.2 (a). A small number of features result in a small number of measurement cycles and high calibration efficiency. The above analysis leads to a general strategy of the FES: capturing several ERPs and sampling the front range, as exhibited in Fig. B.2 (b) and (c). We consider the radiation as a continuous function F versus the elevational angle θ and the azimuthal angle ϕ. Then each ERP can be expressed as $F_i(\theta, \phi_i)$. Each feature ($Feat_{ij}$) can be expressed as $F_{ij}(\theta_j, \phi_i)$, where θ_j is selected by both *Range_ERP* and *Step_ERP*. Here, *Num_ERPs* means the number of captured ERPs; *Range_ERP* defines the sampling range of each ERP; and *Step_ERP* decides the sampling intensity of each ERP. They customize the FES and decide the number of features (*Num_Feats*).

2.3 The Optimal FES

FES can be parameterized by three coefficients, i.e., *Range_ERP*, *Step_ERP*, and *Num_Feats*. Variation in these coefficients results in different FESs. There are two indicators to assess an FES, one is *Num_Feats* and the other is the consequent surrogate model's loss represented as L. Roughly, fewer features lead to worse performance. To quantify the trade-off between the two indicators, we evaluated 80 different FESs.

For simplicity and fairness, the fifty thousand samples are used by 80 different FESs to generate 80 corresponding training data sets. A deep neural network is trained separately with the 80 data sets. During training, the mean squared error between the prediction and label of the complex excitation field was set as the loss function. The final loss represents L.

The colored dots of different sizes represents the 80 FESs in Fig. B.3 (a) and (b). Here, the size indicates the number of features (*Num_Feats*) required as the bigger dot corresponds to the small number of required features. Whereas, the color represents the loss value as green means the loss value meets certain criteria (less than 0.03). The green color transforms to orange and gradually lighter orange as the loss value increases. Note that the threshold value for testing exceeds for training because models often behave slightly inferior during testing. It can be observed that the biggest green dots in Fig. B.3 (a) and (b) represent the optimal FES (*Num_ERPs* = 6, *Range_ERP* = 120, *Step_ERP* = 20).

The $2D$ plots on the right of Fig. B.3, shows the trade-off when two of the coefficients are fixed at their optimal values. Fig. B.3 (a-1), (a-2), and (a-3) represents trade-off for training while Fig. B.3 (b-1), (b-2), and (b-3) shows trade-off for testing. According to the optimal FES, six ERPs and six far-field radiation values on each ERP are selected as the input parameters for the surrogate model.

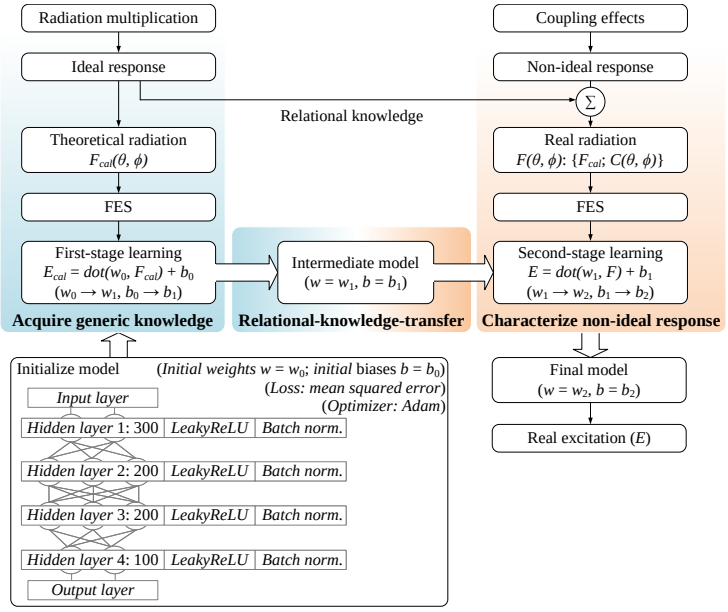


Fig. B.4: The workflow to develop the surrogate model.

For larger uniform planar arrays, the number of ERPs and the fineness of the FES proportionally rise to deliver the radiation of higher resolution. For non-uniform arrays with special arrangements, there is the need to adjust the ERPs distribution accordingly to fit the radiation distribution. The number of candidate FESs depends on the sweep ranges of *Num_ERPs*, *Range_ERP*, and *Step_ERP*.

3 The Surrogate Model

3.1 Workflow

The evolution of the surrogate model, as illustrated in Fig. B.4, arises from combining the conventional array theory and deep learning. The purpose of the surrogate model is to imitate the mapping from the power-only far-field radiation values to the excitation field. The mapping has its general relationship decided by the radiation multiplication theorem, and the coupling effects mainly generate undesirable variations.

The model can learn the generic knowledge of the radiation multiplication from theoretical calculation results; however, the non-ideal response originated from the coupling effects only exposes itself in simulation or measurement results. Thus, the relational-knowledge-transfer approach is used to de-

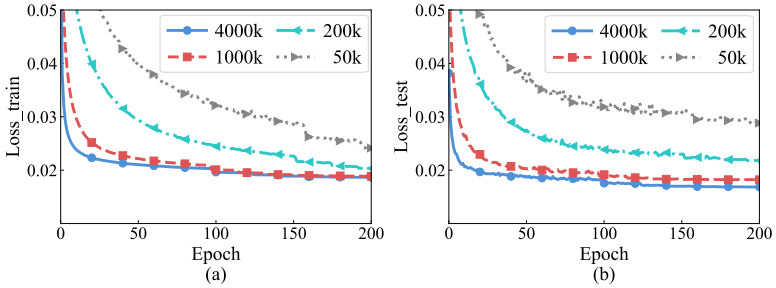


Fig. B.5: Loss versus epochs during the first stage using different amounts of calculation datasets. (a) Loss_train. (b) Loss_test.

compose the learning process into two stages. The first-stage learning grasps the generic knowledge from theoretical calculation results. The second-stage learning further characterizes the detailed non-ideal variations from a few simulation results.

This two-stage learning mode renders the model having generalizability, robustness, and interpretability. Also, our approach improves efficiency by significantly alleviating the need for stimulation by transferring knowledge from the calculation. The following section introduces the first/second-stage learning.

3.2 Acquire Generic Knowledge

After fine tuning and optimization, a fully-connected neural network is initialized, as shown in Fig. B.4, which consists of an input layer, an output layer, and four hidden layers of 300, 200, 200, and 100 neurons. Each hidden layer is attached with the leaky rectified linear unit (LeakyReLU) [21] as the activation function and a batch normalization layer. The Adam [22] is employed as the optimizer to upgrade the weights and biases of the neurons. The mean squared error between the predicted complex excitation field and its theoretical label is used as the loss function to evaluate the model during training.

In the first stage, the model is trained with theoretical data to grasp the generic knowledge where the deviations were excluded. The theoretical data were obtained via the formula established in [23]:

$$F_{cal}(\theta,\phi) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} F_{e,mn}(\theta,\phi)e^{-j\pi\sin\theta(m\cos\phi+n\sin\phi)}. \tag{B.1}$$

Here, $F_{e,mn}(\theta,\phi)$ represents the independent radiation pattern of the antenna element; m and n point at the index of the antenna element in two orthogonal directions while M and N are the numbers of the elements along with these two directions. The unbalanced magnitudes and phases for all the elements

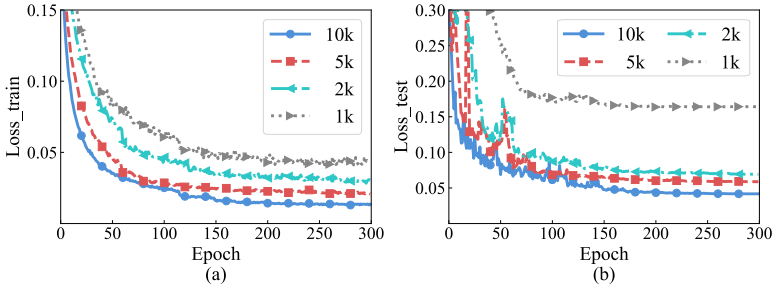


Fig. B.6: Loss versus epochs during the second stage using different amounts of simulation datasets. (a) Loss_train. (b) Loss_test.

were initialized to generate the first-stage training data and then fed into the Eq. B.1 to calculate the theoretical array of radiation. The amplitude can be deviated by ±3 dB, and the phase varies from −40° to 40°. The theoretical array radiation was the input, and the initialized magnitudes and phases were its output label.

The training and testing loss convergence curves versus the learning epoch are exhibited in Fig. B.5. Here, the loss indicates the mean squared error between the first-stage model estimation and the theoretical label of the complex excitation field. Generally, more data results in minor loss, consequently better performance. Nevertheless, this improvement is saturated when the data size reaches four million, as in Fig. B.5, and the curves indicate that four million could ensure good performance. Hence, four million calculation results are used for the first-stage training. After the first-stage learning, the initial model is referred as the intermediate model which has learned the generic knowledge of radiation multiplication.

3.3 Characterize Non-ideal Response

The second-stage learning focuses on characterizing the detailed non-ideal response, including the coupling effects and other non-analytical behavior. This stage relies on the learning by the intermediate model as it holds the generic calibration knowledge. Therefore, in the second stage, the aim is to capture the non-ideal characteristics due to mutual coupling and other non-analytical behavior. Similarly, the mean square error between the second-stage prediction and the simulation of the complex excitation field is set as the loss function for evaluating the second-stage model.

Alike the first stage, then unbalanced magnitudes and phases for all elements were initialized and randomly chosen for the second-stage data collection. They are fed into the simulation setup supported by Computer Simulation Technology®(CST) to yield the non-ideal array radiation. A few arbitrary

elements had the amplitudes deviated by ±3 dB, and the phases varied from -40° to 40°. The simulated array radiation was the input, and the initialized magnitudes and phases were its output.

Following this scheme, the model was trained with different numbers of simulation results as shown in Fig. B.6. Here, the loss represented the mean squared error between the simulated magnitudes and phases and the output during the second-stage iterations. As indicated by the curves, ten thousand simulation data are sufficient for the second-stage learning.

4 Validation and Discussion

4.1 Virtual Validation

The proposed method is applied on a three-by-three planar phased patch antenna array to verify its effectiveness and performance. At first, arbitrary excitation distortions were pre-imposed on the array as amplitudes were deviated by ±3 dB, and phases deviated from -40° to $40°$.

The optimal FES consisted of six ERPs. For each ERP, six power radiation points within the front region were collected. The collected features were fed into the well-established surrogate model to determine the actual excitation. Fig. B.7 exhibits 5 validation samples randomly chosen from 100 validation cases. The well-trained model is validated with 100 different unbalanced samples. The root mean square errors (RMSEs) of amplitude and phase of the 100 validation cases are 0.36 dB and 4.51°. As observed, the proposed surrogate model can precisely determine the unbalanced variations in amplitude and phase.

The method is firstly verified using a virtual scenario in a similar manner to other existing machine learning-based works [19, 20]. The measurement results agree with the simulation via CST as long as the setup is accurate. It is worth mentioning that a three-by-three phased antenna array was chosen to balance the time and computing recourse needed for validation and the complexity level of the array.

4.2 Experimental Validation

The robustness of the proposed method is validated by implementing it on a one-by-eight linear patch array at 3.5 GHz in an experimental scenario. The experimental setup is shown in Fig. B.8, with a one-by-eight linear patch array, 8 $Vaunix\ LPS-402$ phase shifters, a $Vaunix\ LDA-906V-8$ 8-channel attenuator, an 8-channel power divider, and a laptop to adjust the phase shifters and attenuator all placed in an anechoic chamber of 14 m×9.9 m×11.05 m. The elements are aligned in $\phi = 0°$ direction. Therefore, only took one ERP $\phi = 0°$

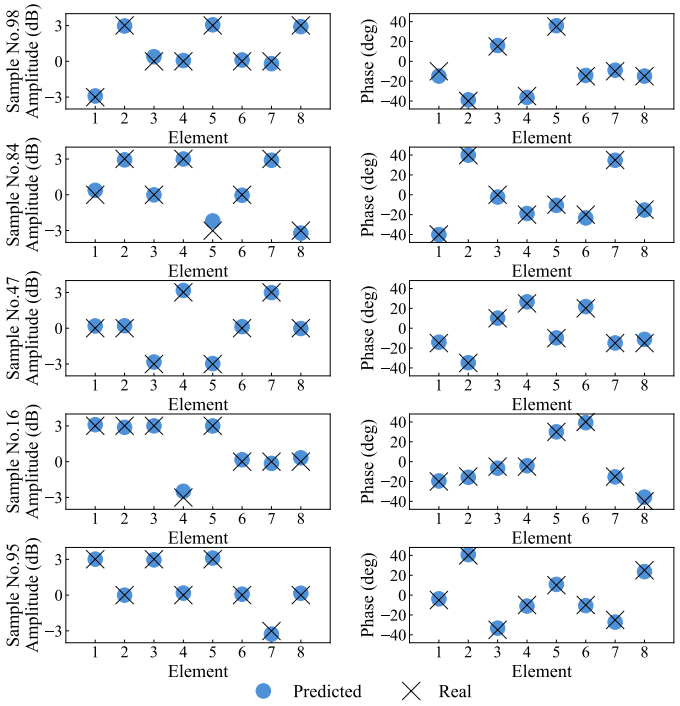


Fig. B.7: Implementation A: 5 samples randomly chosen from 100 cases.

was taken and had 32 features within this ERP because the radiation variation focused in the ERP of $\phi = 0°$.

We measured 100 unbalanced far-field array radiations in the chamber. The RMSE of 100 calibration results in the linear array case is 0.47 dB/5.37 ° in terms of amplitude/phase. Fig. B.9 exhibits 5 random cases. The increment of error depends on the accuracy of measurement.

4.3 Large Array Validation

The proposed method has the potential to calibrate any large arrays with sufficient training samples available. The main challenge is that, as the array size increases, the number of required training samples and the training time arise significantly because the number of possible array excitation combinations arises exponentially. In these validations, amplitudes were deviated by ±3 dB, and phases from −30 ° to 30 °.

It was applied on a four-by-four planar array, with the FES of *Range_ERP*=180°, *Step_ERP*=22.5°, and *Num_ERPs*=8. 4×10^7 samples were used for training, and the RMSE was 0.57 dB/6.14 ° in terms of amplitude/phase.

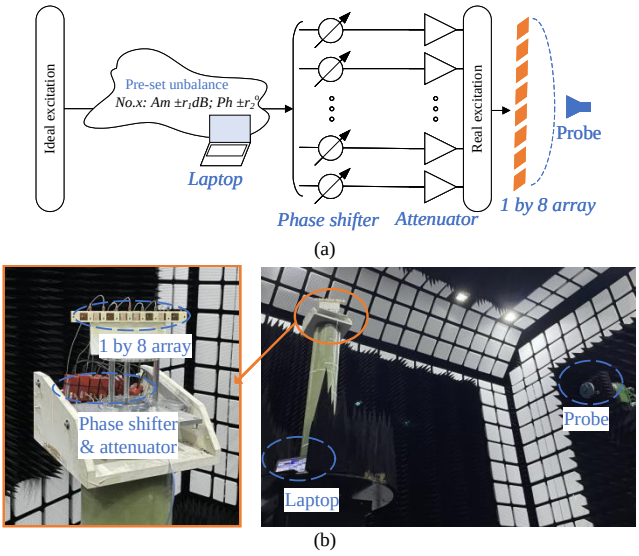


Fig. B.8: (a) Schematic diagram of the measurement setup. (b) Photograph of the measurement setup in an anechoic chamber.

For a linear array of 32 elements, the FES was set as *Range_ERP*=120°, *Step_ERP*=1°, and *Num_ERPs*=1. After being trained using 4×10^8 samples, the RMSE for 10 thousand test samples is 0.62 dB/6.65 ° in terms of amplitude/phase.

For a four-by-eight planar array, the FES was fixed at *Range_ERP*=180°, *Step_ERP*=15°, and *Num_ERPs*=10. It took 8×10^8 samples for training to arrive at the RMSE of 0.67 dB/6.84 ° for 10 thousand test samples. Planar arrays require more training samples than linear arrays of the same amount of elements, because planar arrays have to consider more complex $3D$ radiation variations than linear ones.

The propoosed method has the potential to calibrate larger arrays and the calibration loss can be further reduced if more data and more powerful computing resources are provided. However, powerful computing resources are not commonly available. In future work, we will focus on reduction of required training samples and the need for computing resources to facilitate large array calibration.

4.4 Comparison and Discussion

The proposed method is compared with existing ones in Table B.1. As authors in [4] claimed, a normal REV requires 2^m measurement times to calibrate each element using an m-bit phase shifter. Works in [5] and [15] reduced the aver-

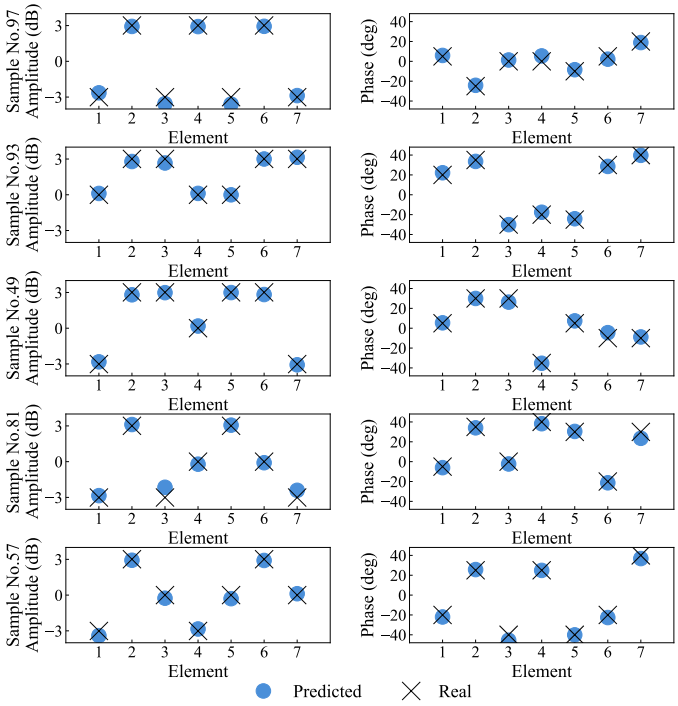


Fig. B.9: Implementation B: 5 samples randomly chosen from 100 cases.

Table B.1: The Comparison Between the Proposed Method and the Existing Power-only Methods

	Measurement times	RMSE (Amplitude/Phase)	phase shifter
[4]	$2^m \times N$	– / –	m-bit
[5]	$4.5N + 1$	0.1087 / 4.754 °	4-bit
[15]	$11.2N$	0.37 dB / 3.06 °	5-bit
This work	$\boldsymbol{4N}$	**0.07 (0.36 dB) / 4.51°**	**No requirement**
	$\boldsymbol{4N}$	**0.10 (0.47 dB) / 5.37°**	**No requirement**

age measurement cycles to around 4.5 and 11.2 by handling multiple elements simultaneously using 4 or 5-bit number phase shifters. Our method achieved comparable calibration accuracy with only 4 measurement cycles. Furthermore, it does not require phase shift resolution, repetitive phase shifting, or equation manipulation.

Theoretically, the proposed method can calibrate larger arrays by proportionally refining the FES. Consequently, the measurement cycles arise proportionally. The number of required training samples increases exponentially to provide sufficient informativeness and can exceeds the available computing

memory as the array size grows. In future work, we will focus on further reduction of measurement cycles and data requirements to facilitate calibration of larger-scaled arrays in $5G/6G$ scenarios.

5 Conclusion

In this communication, a calibration technique is presented using the surrogate model that can calibrate all the elements of phased antenna arrays at once. The surrogate model gets insights from the combination of conventional array theory and deep learning and gains its calibration ability through relational-knowledge-transfer learning. The proposed method requires a smaller number of measurement cycles than the existing REVs and avoids repetitive phase shifting or equation manipulation. The experimental results show that the approximation of the measurement times is $4N$. The optimal FES distinguishes the fewest points that deliver the most informative power-only features. Fed by these features, the surrogate model can directly determine amplitudes and phases for all the elements. It also breaks through the limitation that the conventional extended methods suffer from the bit-number of the phase shifter or the mathematical computability. Once trained, the surrogate model can serve convenient and efficient periodical calibration for phased antenna arrays in $5G/6G$ scenarios.

References

[1] H. Pawlak and A. F. Jacob, "An external calibration scheme for DBF antenna arrays," *IEEE Transactions on Antennas and Propagation*, vol. 58, no. 1, pp. 59–67, 2009.

[2] K.-M. Lee, R.-S. Chu, and S.-C. Liu, "A built-in performance-monitoring/fault isolation and correction (PM/FIC) system for active phased-array antennas," *IEEE Transactions on Antennas and Propagation*, vol. 41, no. 11, pp. 1530–1540, 1993.

[3] M. Tanaka, "On-orbit measurement of phased arrays in satellites by rotating element electric field vector method," *Electronics and Communications in Japan*, vol. 81, no. 1, pp. 1–13, 1998.

[4] T. Takahashi, H. Miyashita, Y. Konishi, and S. Makino, "Theoretical study on measurement accuracy of rotating element electric field vector (REV) method," *Electronics and Communications in Japan*, vol. 89, no. 1, pp. 22–33, 2006.

[5] H.-J. Yoon and B.-W. Min, "Improved rotating-element electric-field vector method for fast far-field phased array calibration," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 11, pp. 8021–8026, 2021.

[6] G. He, X. Gao, and H. Zhou, "Fast phased array calibration by power-only measurements twice for each antenna element," *International Journal of Antennas and Propagation*, vol. 2019, 2019.

[7] C.-N. Hu, "A novel method for calibrating deployed active antenna arrays," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 4, pp. 1650–1657, 2015.

[8] B. C. Ng and C. M. S. See, "Sensor-array calibration using a maximum-likelihood approach," *IEEE Transactions on Antennas and Propagation*, vol. 44, no. 6, pp. 827–835, 1996.

[9] W. T. Patton and L. H. Yorinks, "Near-field alignment of phased-array antennas," *IEEE Transactions on Antennas and Propagation*, vol. 47, no. 3, pp. 584–591, 1999.

[10] S. H. Son, S. Y. Eom, S. I. Jeon, and W. Hwang, "Automatic phase correction of phased array antennas by a genetic algorithm," *IEEE Transactions on Antennas and Propagation*, vol. 56, no. 8, pp. 2751–2754, 2008.

[11] M. D. Migliore, "A compressed sensing approach for array diagnosis from a small set of near-field measurements," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 6, pp. 2127–2133, 2011.

[12] C. He, X. Liang, J. Geng, and R. Jin, "Parallel calibration method for phased array with harmonic characteristic analysis," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 10, pp. 5029–5036, 2014.

[13] G. A. Hampson and A. B. Smolders, "A fast and accurate scheme for calibration of active phased-array antennas," in *IEEE Antennas and Propagation Society International Symposium. Digest*, vol. 2. IEEE, 1999, pp. 1040–1043.

[14] M. Liu and Z. Feng, "Combined rotating-element electric-field vector (CREV) method for nearfield calibration of phased array antenna," in *International Conference on Microwave and Millimeter Wave Technology* IEEE, 2007, pp. 1–4.

[15] T. Takahashi, Y. Konishi, S. Makino, H. Ohmine, and H. Nakaguro, "Fast measurement technique for phased array calibration," *IEEE Transactions on Antennas and Propagation*, vol. 56, no. 7, pp. 1888–1899, 2008.

[16] W. P. Keizer, "Fast and accurate array calibration using a synthetic array approach," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 11, pp. 4115–4122, 2011.

[17] Z. Wang, F. Zhang, H. Gao, O. Franek, G. F. Pedersen, and W. Fan, "Over-the-air array calibration of mmwave phased array in beam-steering mode based on measured complex signals," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 11, pp. 7876–7888, 2021.

[18] R. G. Ayestaran, F. Las-Heras, and L. F. Herran, "High-accuracy neural-network-based array synthesis including element coupling," *IEEE Antennas and Wireless Propagation Letters*, vol. 5, pp. 45–48, 2006.

[19] R. Lovato and X. Gong, "Phased antenna array beamforming using convolutional neural networks," in *IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting* IEEE, 2019, pp. 1247–1248.

[20] J. H. Kim and S. W. Choi, "A deep learning-based approach for radiation pattern synthesis of an array antenna," *IEEE Access*, vol. 8, pp. 226 059–226 063, 2020.

[21] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of the 30th International Conference on International Conference on Machine Learning*, vol. 30, no. 1, 2013.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, Y. Bengio and Y. LeCun, Eds., 2015, pp. 1–15.

[23] R. E. Collin, *Antenna theory*. McGraw-Hill, 1969, no. 1.

Paper C

Representation Learning-Driven Fully Automated Framework for the Inverse Design of Frequency-Selective Surfaces

Zhao Zhou, Zhaohui Wei, Jian Ren, Yingzeng Yin, Gert Frølund Pedersen, and Ming Shen

The paper has been published in the
IEEE Transactions on Microwave Theory and Techniques Vol. 71(6),
pp. 2409–2421, 2023.

The layout has been revised.

Abstract

Frequency selective surfaces (FSSs) refer to planar structures that behave with specific electromagnetic (EM) responses within a frequency range and are widely applied in wireless propagation systems. Given the fact that different EM responses correspond to distinguished topologies, conventional inverse design methods of FSSs are usually labor-intensive, as they rely on experienced human engineers to determine the topology and then rationally tune its structures. There have been great attempts using optimization algorithms (e.g., genetic algorithms) or machine learning to automate the second tuning stage after the initial EM topologies are determined by human engineers. However, the first topology selection stage still require engagements with experienced engineers. This paper proposes a fully-automated framework for the inverse design of FSSs. We achieved a fully-automated inverse design by establishing a machine-friendly mapping flow. The mapping flow derives its continuity and compactness from representation learning, which enables both auto-selection of the topology and auto-evolution of the unit cell based on the topology. The auto-selection stage automatically determines the appropriate topology by compressing the EM constraints through the principal component analysis (PCA) and classifying the topology using the support vector machine (SVM). Afterward, the auto-evolution system can efficiently evolve until it yields an optimal unit cell. We developed a self-monitor strategy to control the evolution and maximize the evolution efficiency by adaptively tuning the three modules within the auto-evolution system. We validated the presented framework with four FSS designs. The results proved its potential as a highly efficient fully-automated tool for the inverse design of FSSs.

1 Introduction

Metasurfaces refer to a periodical combination of overlapped planar metallic and dielectric layers in sub-wavelength size, which have unique abilities to manipulate microwave signals at desired frequencies. They possess capabilities of power-conserving transformations [1, 2], cloaking [3], absorber [4], spatial filtering [5], polarizer [6], radar cross-section reduction [7], to name just a few. Among them, the frequency selective surfaces (FSS) focus on the transmission or reflection capability. The typical inverse design of FSSs follows a performance-oriented process and takes two stages, topology selection and geometry optimization, as demonstrated in Fig. C.1. Conventional inverse design methods require experienced designers with domain knowledge to select the topology in advance. Afterward, the designers need to manually tune the parameters and check the performance through repetitive full-wave electromagnetic (EM) simulations, which is time-consuming and resource-demanding. The design efficiency depends on the designers' experience level and the computation capability. There have been many published methods

that automate the second optimization stage based on optimization algorithms or machine learning. However, most of them are not fully-automated and still require experienced engineers in the topology selection stage. A fully-automated inverse design method is in tremendous demand to reduce the need for experienced engineers for the inverse design of FSSs.

There have been many publications [8–22] utilizing machine learning to automatize the inverse design process to different degrees. Machine learning-based methods are popular due to their unique merits compared to optimization algorithm-based methods, such as genetic algorithm-based methods (GA) [23–25]. A number of full-wave simulations are inevitable for optimization methods [26]. By contrast, machine learning-based methods accumulate intelligence from historical data to form a surrogate model. The surrogate model can reduce or replace the need for full-wave simulation and facilitate the design for new constraints significantly. According to the automation level, we roughly divide them into semi-automated [9–19] and fully-automated methods [20–22]. Considering the strategy, there are two categories: some works [9–16, 21, 22] developed a forward mapping surrogate model and performed iterative optimization to find the optimal design; the others [17–20] directly established an inverse mapping surrogate model or utilized generative networks to produce the satisfying design.

Most of the optimization algorithm-based methods and ML-based methods [9–19] are semi-automated, because they mainly focused on automatic optimization in the second stage of the inverse design. Experienced designers are required to decide the topology in the first topology selection stage. For instance, in [9], an artificial neural network model was developed to determine the reflective phase for a given Minkowski reflectarray element, and then the model was integrated with an optimization algorithm to generate the optimal element for the desired phase. Prado *et al*. [10, 11] utilized support vector machines (SVMs) to substitute full-wave EM simulation tools to accelerate the inverse design of reflectarrays. The SVMs were trained with a sufficient number of simulation results in advance. Afterward, they were used to analyze reflectarray elements at a low time cost and were integrated with iterative optimization processes to determine the reflectarray element for a given phase shift. Abdullah *et al*. [12] applied a data-driven supervised-learning technique integrated with the sequential-search strategy on a crusader cross topology to maximize the radar cross-section (RCS) reduction of the coding FSSs. Hodge *et al*. [14] proposed deep convolutional generative adversarial networks (DCGANs) to realize the inverse design of metasurface elements. The DC-GANs were trained with elements selected from published literature. After training, a Generator, a Discriminator, and a Simulator were obtained. A three-fold process was taken to achieve given EM constraints: a) a random latent noise vector was initialized; b) generating an element by inputting the vector into the Generator and analyzing its EM response using the Simulator; c) adjust-

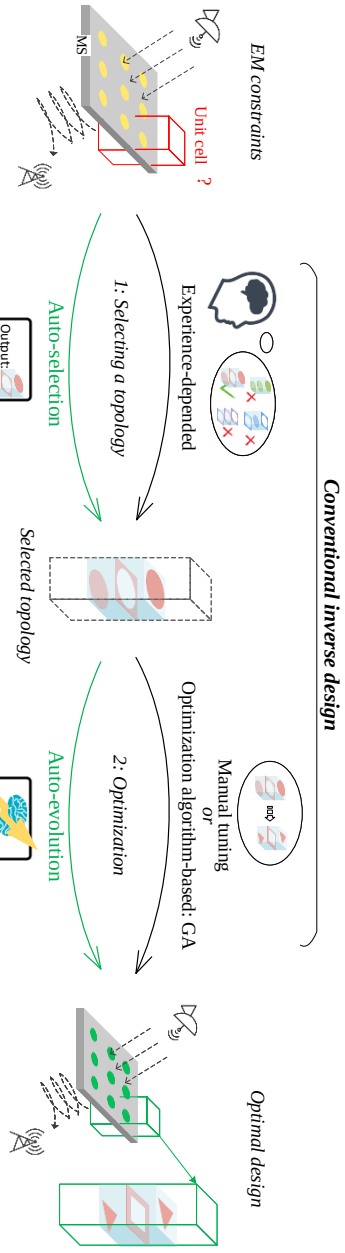


Fig. C.1: Inverse design of FSSs: conventional vs. fully-automated.

ing the vector and repeating steps (b-c) until the EM response meets the EM constraints. Xiao *et al*. [17] proposed an inverse learning system to predict geometry variables of selected metasurfaces for a desired EM behavior without

further optimization, which consists of a data classification technique and a cascade of two inverse learning machines for enhanced learning performance. Zhu *et al*. [18] artificially built up a modified Jerusalem Cross structure at first; afterward, they utilized a three-layer back-propagation neural network to project the reflection phase to the length of the metal arm. Likewise, Wei *et al*. [19] targeted the appropriate topology in advance through analysis of filter equivalent circuit before applying the genetic algorithm (GA) to reach an optimal design.

The works in [20–22] realized full automation. The work in [20] realized automated design by integrating a variational autoencoder, a predictor, and an optimizer. After training, they represented a metasurface into the latent variables using the encoder and optimized the variables by employing particle swarm optimizations integrated with the predictor. The decoder would decode the optimized variables into a final metasurface. In the implementations of dual-layer and three-layer metasurfaces, they simulated 17.5 thousand and 16.5 thousand structures to collect data for training the autoencoder and predictor. Naseri *et al*. [21] achieved a fully-automated inverse design of nonuniform bianisotropic metasurfaces in two steps. In the first step, they determined the surface parameters within several candidates using the alternating direction method of multipliers based on the two-dimensional method of moments. In the second step, they trained the machine learning surrogate model with 70 thousand samples and then utilized particle swarm optimization to optimize the surface parameters. Nadell *et al*. [22] utilized a twelve-layer deep neural network to model the forward mapping of an all-dielectric metasurface, which took 21 thousand simulation data for training. Afterward, they developed a novel fast-forward-dictionary-search method to solve the inverse modeling problem: they collected all combinations of geometric parameters on 13^8 spectra to generate a dictionary. Given the constraints, they searched through the dictionary to list all combinations that approximately satisfy the constraints and gave the best ones. One shared limitation among these works is that they require tens of thousands of simulation data to train the surrogate model, which is a commonly criticized disadvantage of most machine-learning approaches.

In this paper, we manage to realize the fully-automated inverse design without establishing a dictionary mapping through tremendous simulation data in advance. To achieve this, we need to develop a machine-readable mapping of continuity and compactness. The continuity of mapping determines the degree of automation. The works in [12, 13, 18, 19] achieved only semi-automation because their mappings did not start from the constraints, which is the initial state of inverse design. The compactness of the mapping impacts the difficulty to establish and the overall efficiency. The works in [20–22] managed to establish mappings from all the candidate geometries to their EM behaviors, which suffered from high complexity and required a huge amount

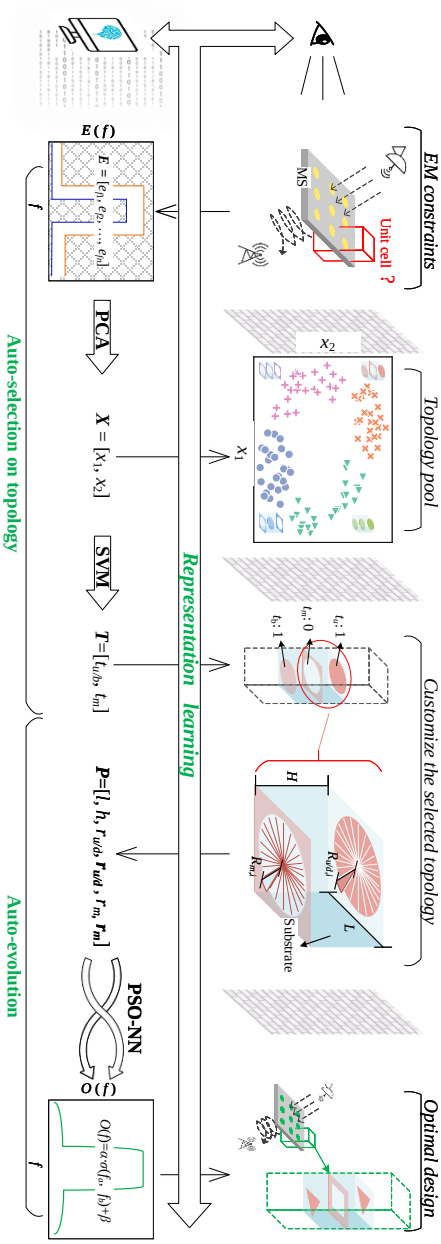


Fig. C.2: Proposed fully-automated inverse design framework based on representation learning.

of simulation data.

Representation learning [27, 28] may offer us inspiration to facilitate the fully-automated inverse design. Boulanger-Lewandowski *et al.* [28] have proven

that representation learning outperforms many state-of-art models aiming at the transcription of polyphonic music. Representation learning, also known as feature learning, allows the machine to automatically represent machine-unfriendly physical features into machine-friendly features and then smoothly project from the source space into the target space.

In this paper, a highly efficient fully-automated inverse design framework for FSSs is developed. Based on representation learning, we facilitate fully-automated inverse design by establishing a continuous and compact horizontal projection integrated with vertical representations of physical features. The horizontal projection consists of auto-selection and auto-evolution. The vertical representations include vectorization of the constraints and fan-based modeling. We have three main contributions:

1. compared with [12, 13, 18, 19], we establish a mapping from EM constraints to the topology candidates to enable auto-selection of the suitable topology, which can be accomplished without experienced designers;

2. one advantage over [20–22] is that we establish a cross-evolution system to customize the topology and evolve it into an optimal design instead of building a dictionary mapping in advance through tedious training, hence we take only hundreds of simulation cycles while they took tens of thousands of simulation cycles to collect data;

3. the other advantage is, unlike [20–22] that obtained the optimal design by optimizing several candidate geometries, the auto-evolution system can evolve the auto-selected topology into new geometries that satisfy various constraints, thanks to the proposed fan-based representation defined by only tens of parameters instead of the pixelated metal layer defined by hundreds of parameters [20].

We validate the proposed framework in four cases where different EM constraints were provided. It has been proven that, with only the EM constraints provided, the framework can automatically select the appropriate topology and yield the optimal designs in different scenarios.

The main content is arranged as follows. Section II demonstrates the workflow and components of the proposed fully-automated inverse design framework. After that, Section III describes the implementations we performed in different scenarios. The corresponding results are discussed and compared with the existing methods to clarify our advantages in Section IV. Section V lists our conclusions.

2 Fully-Automated Inverse Design Framework

The inverse design seeks a projection from EM constraints to an optimal design. Only a few states during the projection are physically visible to human engineers, and these states belong to various isolated spaces. Thus, conventional methods rely on experienced engineers to deal with these separate spaces. To enable fully-automated inverse design, we built up a continuous mapping flow to connect these states. As illustrated in Fig. C.2, this mapping flow derives from representation learning: the physical features in isolated states are represented in machine-readable features; the represented machine-readable features are projected from the current state to its posterior state. The vertical representation and horizontal projection follow adaptive rules in different states. The proposed framework can be roughly divided into two stages: auto-selection of the topology and auto-evolution of the selected topology.

2.1 Auto-Selection

The goal in the first stage is to select a suitable topology for given EM constraints. The EM constraints define how the FSS should behave. The EM constraints and possible topology are physically visible to human engineers. However, the EM constraints and possible topology are not machine-readable and are in isolated spaces. Thus, to enable auto-selection, we need to represent the EM constraints and possible topology into machine-readable features and establish a mapping that connects the EM constraints to the proper topology.

A single EM constraint can be represented as a curve (*E(f)*) of the coefficient (such as transmission, reflection, axial ratio, etc.) of interest versus the frequency within the desired band. We can discretely sample the *E(f)* and generate a vector $\boldsymbol{E}$ of a constant length that represents the constraint. Sometimes, there are multiple constraints on one FSS and consequently multiple *E(f)s*. In this case, we can discretely sample each constraint *E(f)s* and calculate their normalized weighted mean to form a final vector $\boldsymbol{E}$. The weights of multiple constraints can be identical if they are equally important. Otherwise, the weights can be unequally assigned according to their importance level. This final vector $\boldsymbol{E}$ represents all the constraints. As shown in Fig. C.2, $\boldsymbol{E}$ is defined as a one-dimensional vector,

$$\boldsymbol{E} = [e_{f1}, e_{f2}, ..., e_{fn}], \tag{C.1}$$

where each element (e_{fi}) of the vector $\boldsymbol{E}$ represents the constraints at a single frequency point. The size (n) of $\boldsymbol{E}$ equals the number of frequency points of interest. Generally, as n increases, $\boldsymbol{E}$ contains more detailed information. $\boldsymbol{E}$ would be taken as input for the proposed framework to impose the expected constraints on the inverse design process.

As for the topology of the unit cell, we took a combination of three overlapped metal layers and two substrate layers between every two metal layers, as seen in Fig. C.2. The up and bottom metal layers are identical as well as the two substrate layers, because the equivalent filtering circuit is symmetrical. We classified the topology according to whether each metal layer should be a patch layer or a slotted layer. The topology (T) can then be defined as,

$$\boldsymbol{T} = [t_{u/b}, t_m], \quad t_* = \begin{cases} 1, & if * is\ patch \\ 0, & if * is\ slotted, \end{cases} \tag{C.2}$$

where *u/m/b* means up, middle, and bottom metal layers, respectively.

By far, we had the EM constraints and possible topology represented as machine-readable vectors, $\boldsymbol{E}$ and $\boldsymbol{T}$. The next step is to connect them by projecting $\boldsymbol{E}$ to $\boldsymbol{T}$. In machine view, it is a classification task with $\boldsymbol{E}$s as the input features and $\boldsymbol{T}$s as the classes. After training, the machine would output the class of the highest possibility that satisfies the constraints.

To facilitate the classification, we performed the principal component analysis (PCA) [29] on $\boldsymbol{E}$ to compress the input features. The PCA can help reduce the column dimension n of $\boldsymbol{E}$ by projecting a high dimensional space into a low dimensional space. Therefore, it indeed projected $\boldsymbol{E}$ of high column dimension to $\boldsymbol{X}$ of low column dimension. Decreasing the low dimension can reduce the complexity of classification but may lower its accuracy. After tuning and comparison, we fixed the low dimension as 2 to balance the training complexity and classification accuracy. Here is how we performed it:

1. pre-defined N sets of constraints within a broad range and represented them as {*E(f)*s, $\boldsymbol{T}_s$}, where $N = 100$;
2. discretely sampled *E(f)*s (or the weighted mean in case of multiple constraints) into N sets of $\boldsymbol{E}$ and formed the data sets {$\boldsymbol{E}_s$, $\boldsymbol{T}_s$};
3. acquired the eigenvectors $\boldsymbol{W}_s$ of $\boldsymbol{E}_s$ by solving the equation $\boldsymbol{E}_s = \boldsymbol{W}_s \cdot \Lambda \cdot \boldsymbol{W}_s^{-1}$;
4. picked out two eigenvectors corresponding to the two largest eigenvalues and formed $\boldsymbol{W}_{s2}$;
5. projected $\boldsymbol{E}$ to $\boldsymbol{X}$ by transformation, $\boldsymbol{X} = \boldsymbol{W}_{s2}{}^T \cdot (\boldsymbol{E} - m)$, here m is the mean of $\boldsymbol{E}_s$;
6. the data set {$\boldsymbol{E}_s$, $\boldsymbol{T}_s$} was transformed into {$\boldsymbol{X}_s$, $\boldsymbol{T}_s$}.

After the PCA, $\boldsymbol{E}$ of size n was projected to $\boldsymbol{X} = [x_1, x_2]$ of size 2. To visualize the performance of the PCA, $\boldsymbol{X}$ is plotted as a two-dimensional scattering plot in Fig. C.3. It shows that four types of topology denoted by four different symbols are separated in the $\boldsymbol{X}$ space transformed through PCA.

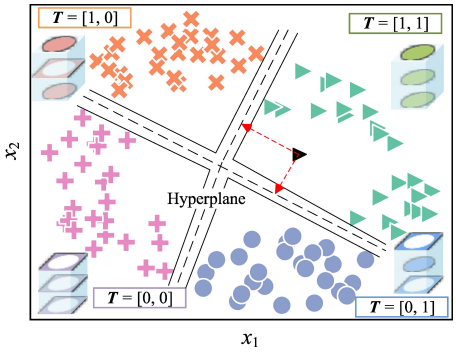


Fig. C.3: New classification space defined by $\boldsymbol{X}$ projected through PCA and the illustration of the support vector classification.

Now the classification problem was simplified to a projection problem from $\boldsymbol{X}$ to $\boldsymbol{T}$. We then automatized this projection by using SVMs [30] for classification. Here is the guideline to compute the SVM classifier:

1. performed the PCA on $\boldsymbol{E}_s$ from $\{\boldsymbol{E}_s, \boldsymbol{T}_s\}$ and formed training data $\{\boldsymbol{X}_s, \boldsymbol{T}_s\}$;

2. built an optimum hyperplane that separates the classes, its vector is as follows,

$$\boldsymbol{w} = \sum_{i=1}^{n} c_i \cdot \boldsymbol{T}_{si} \cdot \Phi(\boldsymbol{X}_{si}), \tag{C.3}$$

   where $\boldsymbol{X}_{si}$ and $\boldsymbol{T}_{si}$ are the input and label of each training data sample, Φ represents the projection from input to output;

3. the classification vector $\boldsymbol{w}$ was obtained by solving the linear kernel-based function,

$$\max_{w,b} \; f(c_i) = \sum_{i=1}^{n} c_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{T}_{si} \cdot c_i \cdot K(\boldsymbol{X}_{si}, \boldsymbol{X}_{sj}) \cdot \boldsymbol{T}_{si} \cdot c_i,$$

$$subject\ to\ \forall i, \; \sum_{j=1}^{n} c_i \cdot \boldsymbol{T}_{si} = 0, \; 0 \le c_i \le \frac{1}{2n\lambda}; \tag{C.4}$$

   where the linear kernel refers to $K(\boldsymbol{X}_{si}, \boldsymbol{X}_{sj}) = \sum_i^N \boldsymbol{X}_{si} \cdot \boldsymbol{X}_s$, and $\lambda > 0$ determines the trade-off between increasing margins between classes and ensuring samples being classified correctly;

4. the offset b is obtained by,

$$b = \boldsymbol{w}^T \cdot \Phi(\boldsymbol{X}_{si}) - \boldsymbol{T}_{si}; \tag{C.5}$$

5. after training, a soft margin was obtained, and the well-trained SVMs can be utilized to project a vector $\boldsymbol{X}$ to $\boldsymbol{T}$,

$$\boldsymbol{T} = \boldsymbol{w}^T \cdot \Phi(\boldsymbol{X}) - b. \tag{C.6}$$

In real design, an $\boldsymbol{E}$ vector of size n was projected to $\boldsymbol{X} = [x_1, x_2]$ of size 2 through PCA, which is represented as a black triangle in Fig. C.3. Afterward, $\boldsymbol{X}$ was projected to $\boldsymbol{T}$ using the well-trained SVM classifier. As shown in Fig. C.3, the black triangle was classified to $\boldsymbol{T}_s$= [1, 1] according to hyperplanes defined by the well-trained SVM classifier. Through PCA and SVM, the vector $\boldsymbol{E}$ that represents the constraints was projected to $\boldsymbol{T}$ that represents the selected topology. The auto-selection of topology was accomplished.

2.2 Auto-Evolution

Auto-evolution aims to search for an optimal design of the selected topology that satisfies the EM constraints. In machine view, it is to find a mapping from the design of the unit cell to its EM satisfaction level. The EM satisfaction level ($S(\boldsymbol{O})$) of a design indicates how much its EM behavior fits the EM constraints ($\boldsymbol{E}$). It is evaluated by measuring how much the design's EM behavior fits the EM constraints. At first, we need to represent the design and its EM satisfaction level in machine-readable features.

The design of the unit cell related to any topology is a combination of three overlapped metal layers with two substrate layers clamped in the middle, as exhibited in Fig. C.4. The metal layers define the topology type. According to the result $\boldsymbol{T}$ output from auto-selection in the first stage, the three metal layers can be either patch layer ($t_* = 1$) or slotted layer ($t_* = 0$). To generalize the instances of the topology, we divide the patch or slot into a fixed number of fans, as shown in Fig. C.4. Each fan shares an angle of $360°/N_f$ and there are N_f fans in total ($N_f = 72$). The length of each fan is adjustable to produce changeable capacitance and inductance. Multiple fans act as tunable resonant networks. The number of fans N_f decides the complexity and flexibility of the geometry of each metal layer. Simpler geometries can be obtained by reducing N_f to a smaller value, while more complex geometries can be obtained by increasing N_f to a larger value. The geometrical flexibility of each metal layer decreases or increases accordingly. By adjusting the length ($r_{u/m/b,i}$) of each fan, we can tune the capacitance and inductance to form various resonant networks. The metal layer can be customized into various planar geometries. Fan-based representation reaches a balance between increasing possibilities and reducing complexity. Unlike the traditional method [18] that swept the parameters of a specific geometry, fan-based representation creates new geometries and provides more possibilities and functionalities. Compared with [19] that pixelated the metal layer and involved hundreds of

parameters, fan-based representation creates sufficient structures by utilizing only tens of parameters, hence training complexity and the amount of training data are reduced significantly. To further reduce complexity, we divide the metal layer into several sectors according to filtering requirements in different directions. The fans within each sector are identical to the fans within any other sector. Suppose that the number of sectors is S, the number of fans to adjust will be reduced to N_f/S because all the sectors are identical. N_f/S is equivalent to the number of fans within each sector. The number of sectors (S) depends on the EM constraints: $S = 8$ fits symmetrical constraints; $S = 4$ is suitable when there are different constraints for orthogonally polarized signals; $S = 1$ or 2 is recommended for asymmetrical constraints in beam management scenarios. Fig. C.4 exhibits three examples of customized metal layers in case of $\boldsymbol{T} = [1, 0], N_f = 72, S = 8/4/2$. Now we represented each instance of the unit cell as a normalized vector $\boldsymbol{P}$,

$$\boldsymbol{P} = [l, h, r_{u/b}, \boldsymbol{r}_{u/b}, r_m, \boldsymbol{r}_m]. \tag{C.7}$$

The representation rules between the geometric parameters (mm) of the unit cell and its normalized $\boldsymbol{P}$ can be expressed as follows. l is the normalized length of the unit cell, L is the real length, f_c is the center frequency, and c is the speed of light in vacuum. The real length L can be represented as,

$$L = (4 \cdot f_c \cdot l/c - 0.8)/0.4. \tag{C.8}$$

The real height H of each substrate layer can be expressed as,

$$H = \begin{cases} 0.305, & 0 < h < 0.25 \\ 0.508, & 0.25 \leq h < 0.5 \\ 0.813, & 0.5 \leq h < 0.75 \\ 1.524, & 0.75 \leq h < 1, \end{cases} \tag{C.9}$$

according to the standard thickness of $Rogers4003C$ substrate from Rogers Corporation [31]. The real radius $R_{u/b/m}$ of each metal layer can be represented as,

$$R_{u/b/m} = H/18 \cdot (r_{u/b/m} + 8), \tag{C.10}$$

where $r_{u/b/m}$ is the normalized radius of the metal layer. The real length of each fan can be represented as,

$$R_{u/b/m,i} = R_{u/b/m} \cdot r_{u/b/m,i}. \tag{C.11}$$

Here, $r_{u/b/m,i}$ represents each element of the vector $\boldsymbol{r}_{u/b/m}$ and indicates the normalized length of each fan. $R_{u/b/m,i}$ represents each element of the vector $\boldsymbol{r}_{u/b/m}$ and indicates the real length of each fan. The size of vectors $\boldsymbol{r}_{u/b/m}$ and $\boldsymbol{r}_{u/b/m}$ equals to N_f/S.

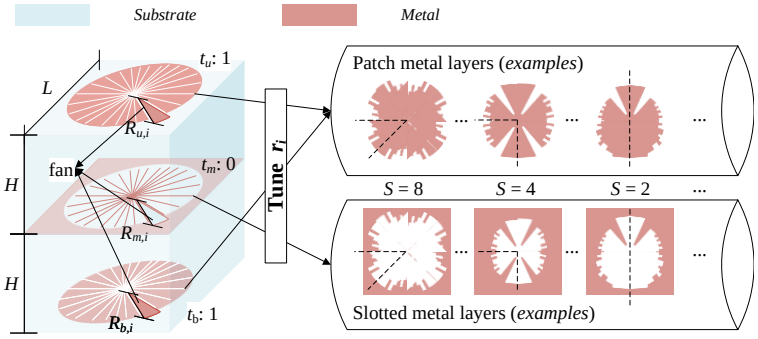


Fig. C.4: Customization of the unit cell (An example when $\boldsymbol{T} = [1, 0]$).

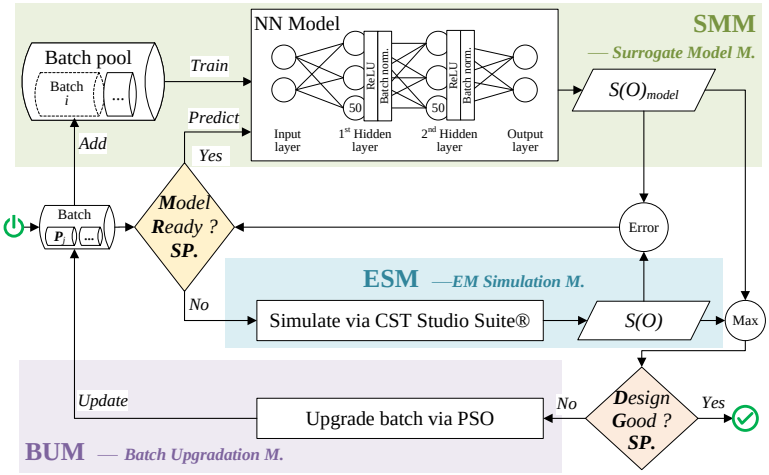


Fig. C.5: Auto-evolution workflow based on PSO-NN.

The EM satisfaction level ($S(\boldsymbol{O})$) of a design indicates how much its EM behavior fits the EM constraints ($\boldsymbol{E}$), and it is evaluated by measuring how much the design's EM behavior fits the EM constraints. To evaluate the satisfaction level $S(\boldsymbol{O})$ of a design, we represented its EM behavior as,

$$\boldsymbol{O} = [o_{f1}, o_{f2}, ..., o_{fn}]. \tag{C.12}$$

Then, the design's satisfaction level $S(\boldsymbol{O})$ is evaluated by measuring how much $\boldsymbol{O}[o_{fi}]$ fits $\boldsymbol{E}[e_{fi}]$. In practice, we decomposed $\boldsymbol{E}$ into an upper boundary, $\boldsymbol{E}_u[e_{fi,u}]$ and a lower boundary, $\boldsymbol{E}_d[e_{fi,d}]$. $S(\boldsymbol{O})$ can then be defined as,

$$S(\boldsymbol{O}) = \frac{1}{|\boldsymbol{L}(\boldsymbol{O})|}, \quad \boldsymbol{L}(\boldsymbol{O}) = \begin{cases} 0, & e_{fi,d} < o_{fi} < e_{fi,u} \\ \beta \cdot e_{fi,u} + (1-\beta) \cdot e_{fi,d}, & else \end{cases} \tag{C.13}$$

Here, β determines the trade-off between emphasizing the upper boundary and emphasizing the lower boundary.

After representation, now the auto-evolution is to find the optimal design $\boldsymbol{P}_{optimal}$ that maximizes the satisfaction level $S(\boldsymbol{O}(\boldsymbol{P}))$, which can be expressed as,

$$\boldsymbol{P}_{optimal} = \underset{\boldsymbol{P}}{argmax}\{S(\boldsymbol{O}(\boldsymbol{P}))\}. \tag{C.14}$$

To solve this, we set up a cross-evolution system, as illustrated in Fig. C.5. As its name implies, this system iteratively operates across three modules: a surrogate model module (SMM), an EM simulation module (ESM), and a batch upgradation module (BUM).

The SMM consists of a batch pool, a neural network (NN) model, and an output. The batch pool collects all the designs ($\boldsymbol{P}s$) produced by far and provides training data for the NN model. The NN model has two hidden layers, each equipped with 50 neurons and attached with a batch normalization [32]; it utilizes the Adaptive Moment Estimate (Adam) [33] as the optimizer; the Rectified Linear Unit (ReLU) [34] is employed as the activation function. The cost function of the surrogate model is the mean square error between the predicted $S(\boldsymbol{O})$ and the real $S(\boldsymbol{O})$ given by the ESM. The SMM operates in two states, training and predicting. In the training state, it utilizes the historical data from the batch pool to train a surrogate NN model; in the predicting state, it takes a new batch of the designs ($[\boldsymbol{P}_j]$) as input in each iteration and predicts their corresponding $S_i(\boldsymbol{O})$s, the $S(\boldsymbol{O})$s in iteration i. The SMM operates in the training state from the beginning, and it shifts to the predicting state when the NN model is well-trained. The NN model is considered well-trained when the predicted $S_i(\boldsymbol{O})$ is close enough to the real $S_i(\boldsymbol{O})$ given by the ESM. The SMM stops running when an optimal design has been obtained.

The ESM is supported by Computer Simulation Technology (CST) Studio Suite®, a full-wave EM modeling and simulation software. This module simulates each design within a batch ($[\boldsymbol{P}_j]$) in each iteration and yields its real $S_i(\boldsymbol{O})$. The real $S_i(\boldsymbol{O})$ is taken as the output label for training the NN model in the SMM. The real $S_i(\boldsymbol{O})$ is compared with the predicted $S_i(\boldsymbol{O})$ to evaluate the performance of the NN model in the SMM. The ESM stops running when the NN model has been well-trained.

At the end of each iteration, the BUM upgrades the batch and produces a new batch for the next iteration. The upgradation follows the rule of the particle swarm optimization (PSO) [35]. Suppose that the batch is $[\boldsymbol{P}_j]_i$ in iteration i, then a new batch $[\boldsymbol{P}_j]_{i+1}$ for next iteration $i+1$ can be expressed as,

$$[\boldsymbol{P}_j]_{i+1} = [\boldsymbol{P}_j]_i + \boldsymbol{U}_{i+1}. \tag{C.15}$$

Here, $\boldsymbol{U}_{i+1}$ depends on the batch-best design ($\boldsymbol{P}_{bb,i}$) and the historical-best design ($\boldsymbol{P}_{hb,i}$). The batch-best design ($\boldsymbol{P}_{bb,i}$) refers to the best design of current batch that maximizes $S_i(\boldsymbol{O}([\boldsymbol{P}_j]))$ in iteration i. The historical-best design refers

to the best design of all the batches ($\boldsymbol{P}_{hb,i}$) that maximizes $S(\boldsymbol{O}([\boldsymbol{P}]))$. $\boldsymbol{U}_{i+1}$ can be expressed as,

$$[\boldsymbol{U}_j]_{i+1} = w \cdot [\boldsymbol{U}_j]_i + c_1 \cdot r_1 \cdot \boldsymbol{P}_{bb,i} + c_2 \cdot r_2 \cdot \boldsymbol{P}_{hb,i}, \tag{C.16}$$

$$w = w_{max} + \frac{i}{I} \cdot (w_{max} - w_{min}), \tag{C.17}$$

where w_{max}, w_{min}, c_1, and c_2 are adjustable hyperparameters; i and I are the current and maximum iterations, respectively; r_1 and r_2 are random floats within the range [0, 1], as defined in [35]. The BUM keeps running until an optimal design has been reached.

The three modules cooperate. a) The batch pool in the SMM is expanded iteratively by gathering new batches generated by the BUM. b) The ESM provides the real $S_i(\boldsymbol{O})$ as the output label for the SMM to evaluate the performance of the surrogate NN model; the SMM trains the surrogate model to replace the ESM. c) in each iteration, the BUM generates a new batch according to $S_i(\boldsymbol{O})$ obtained from the SMM/ESM, and the new batch is sent to the SMM/ESM in the next iteration.

The running schedule of the three modules is controlled by the model-ready sign producer (MRSP) and the design-good sign producer (DGSP). The MRSP and DGSP monitor the auto-evolution system, and they activate or abort the modules by releasing the model-ready and design-good signs. The model-ready sign indicates that the model has been well-trained. The model-ready sign is released when the mean square error between the SMM's predicted $S_i(\boldsymbol{O})$ and the ESM's real $S_i(\boldsymbol{O})$ reaches a minimum threshold. The mean square error between the SMM's predicted $S_i(\boldsymbol{O})$ and ESM's real $S_i(\boldsymbol{O})$ is the cost function of the MRSP. The model-ready sign shifts the SMM from training to predicting state, and it stops the ESM. The design-good sign indicates that indicates the optimal design has been achieved. The cost function of DGSP is $S_i(\boldsymbol{O})$. The design-good sign is released to stop the whole system when the $S_i(\boldsymbol{O})$ reaches a maximum threshold. In reality, the actual running schedule of the system depends on the EM constraints.

The advantage of the cross-evolution system is to integrate three modules, upgrade the training data dynamically, and adjust working states adaptively. The cross-evolution system can maximize the efficiency to evolve the selected topology into an optimal design automatically.

3 Validation

We validated our framework in four common scenarios: band-pass, dual-band-pass, high-pass, and linear-to-circular polarizer. Note that the operating frequencies were set arbitrarily without any preference for fair validation.

Table C.1: Parameters of the Designed Band-Pass Unit Cell

Parameter	Value (mm)
L	12
H	1.524
$\boldsymbol{R}_{u/b}$	[8.40, 4.37, 8.51, 8.96, 9.30, 7.95, 4.82, 6.61, 10.64, 7.39]
$\boldsymbol{R}_m$	[3.06, 6.43, 4.94, 4.16, 2.98, 4.63, 1.57, 3.37, 4.39, 3.29]

3.1 Band-Pass

The band-pass behavior is widely required to raise the Signal to Noise (S/N) ratio and improve the sensitivity of wireless transmitters/receivers [36, 37]. A band-pass FSS acts as a spatial filter for incident microwave signals. It allows only the signals between a pre-defined frequency range to pass through and attenuates the remaining.

A band-pass FSS is evaluated by measuring its reflection coefficient ($S_{11/22}$) and transmission coefficient ($S_{21/12}$). As we proceeded our experiment in the ideal loss-free EM environment supported by CST, where the sum square of $S_{11/22}$ and $S_{21/12}$ remains constant and S_{21} equals S_{21}, we only utilized S_{21} to evaluate the unit cells.

For validation, we pre-defined the constraints as a band-pass behavior from 6 GHz to 7.5 GHz, an insert loss less than 2 dB, which can be expressed as,

$$S_{21}\begin{cases} < -15\,\mathrm{dB}, & f < 5\,\mathrm{GHz}; \\ > -2\,\mathrm{dB}, & 6\,\mathrm{GHz} < f < 7.5\,\mathrm{GHz}; \\ < -15\,\mathrm{dB}, & f > 9\,\mathrm{GHz}. \end{cases} \tag{C.18}$$

By discretely sampling 200 points from 2 GHz to 12 GHz, the constraints were represented into two vectors of length 200 that indicate upper and lower boundaries (Up_bound $\boldsymbol{E}_u$ and Low_bound $\boldsymbol{E}_d$), as given in (C.19) and (C.20). The $\boldsymbol{E}$ vector was defined as the normalized mean of $\boldsymbol{E}_u$ and $\boldsymbol{E}_d$, as given in (C.21).

$$\boldsymbol{E}_u = [\overbrace{-15, ..., -15}^{length:60}, \overbrace{0, ..., 0}^{length:80}, \overbrace{-15, ..., -15}^{length:60}]; \tag{C.19}$$

$$\boldsymbol{E}_d = [\underbrace{-30, ..., -30}_{length:80}, \underbrace{-2, ..., -2}_{length:30}, \underbrace{-30, ..., -30}_{length:90}]; \tag{C.20}$$

$$\boldsymbol{E} = \frac{(\boldsymbol{E}_u + \boldsymbol{E}_d)}{2 \times max(|E_{ui}|, |E_{di}|)}, \quad i = 1, 2, ..., 200. \tag{C.21}$$

According to $\boldsymbol{E}$, the auto-selection component pointed out the appropriate topology, $\boldsymbol{T}$= [1, 0]. The metal layers of the topology were divided into 8 sectors, and each sector had 10 fans, because the constraints were symmetrical. Once the topology was decided, the auto-evolution system activated and

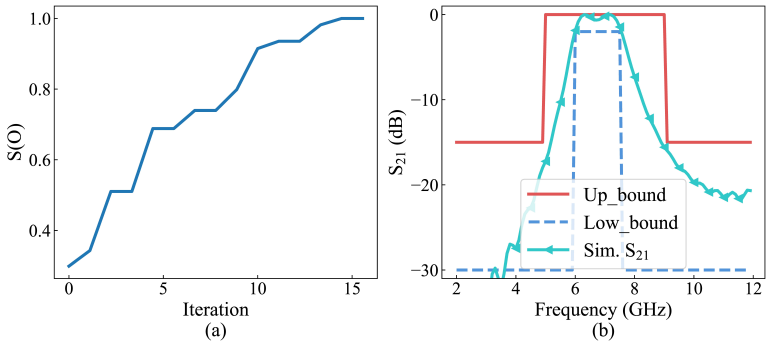


Fig. C.6: Evolution result of the band-pass unit cell: (a) The evolution record over the iterations; (b) The boundaries (Up_bound $\boldsymbol{E}_u$ and Low_bound $\boldsymbol{E}_d$) and simulated S_{21} of the designed unit cell.

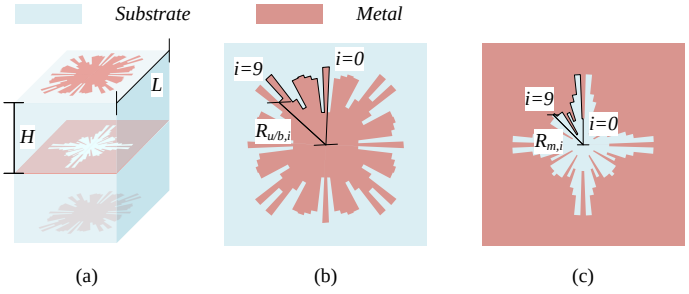


Fig. C.7: Designed unit cell for the band-pass FSS: (a) Its overall structure; (b) Its up/bottom metal layer; (c) Its middle metal layer.

yielded an optimal band-pass unit cell after only 14 iterations, as seen in Fig. C.6(a). There were 20 samples in each iteration, including 280 samples in total. The geometry of the designed unit cell is shown in Fig. C.7 and Table C.1 listed its geometric parameters. Fig. C.6(b) exhibits its simulated transmission coefficient (S_{21}, the green curve labeled as "Sim. S_{11}") of the output unit cell, where red and blue lines mark the upper boundary (Up_bound) and lower boundary (Low_bound). We can observe that the desired band-pass behavior was realized.

To validate the simulation results, a prototype of a band-pass surface that consists of 15 by 15 designed unit cells was fabricated and measured in an anechoic chamber. Three metal layers are etched on the surface of two *Rogers*4003C substrate layers with a thickness of 1.524 mm. The measurement setup is shown in Fig. C.8. A transmitter horn and a receiver horn were placed on two sides of the FSS prototype under test, they are opposite to each other and are connected to a network analyzer. The measured S_{21} is compared with the simulated S_{21} in Fig. C.9. A good agreement level between measure-

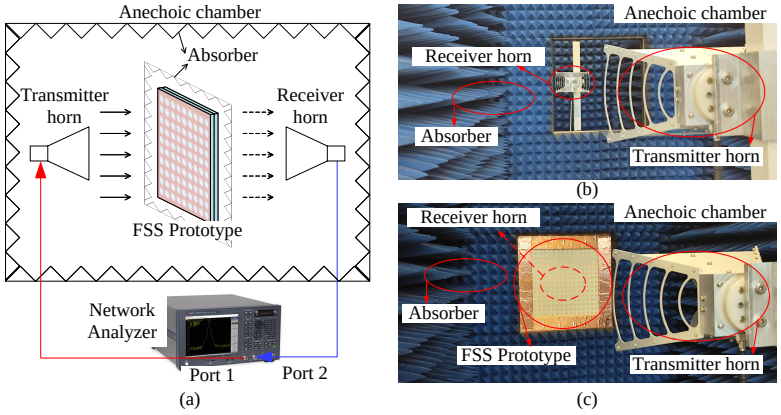


Fig. C.8: (a) Illustration of the measurement setup; (b) Photograph of the anechoic chamber; (c) Photograph of the FSS prototype in the anechoic chamber.

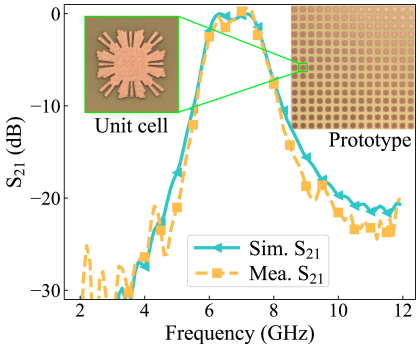


Fig. C.9: Measured and simulated S_{21}s of the band-pass prototype.

ment and simulation can be observed, which verifies the performance of this band-pass design.

3.2 Dual-Band-Pass

The significant impact of dual-band-pass FSSs [5] attracts both the industry and academia as the $5G/6G$ of wireless communication approaches. Dual-band-pass FSSs can decouple the shared aperture base station antennas and enable their co-existence.

A dual-band-pass FSS should serve as a filter that sieves the approaching microwave signals within two separate frequency ranges. Compared with the single-band-pass case, it increases the difficulty for human engineers, but it makes no difference in the perspective of the machine equipped with our framework.

Table C.2: Parameters of the Designed Dual-Band-Pass Unit Cell

Parameter	Value (mm)
L	18.98
H	0.813
$\boldsymbol{R}_{u/b}$	[3.17, 5.17, 4.90, 5.35, 2.81, 5.53, 5.26, 3.63, 3.81, 3.45]
$\boldsymbol{R}_m$	[1.48, 0.93, 4.69, 2.90, 5.00, 0.43, 2.41, 4.19, 3.15, 5.67]

Likewise, we pre-defined the constraints as a dual-band-pass behavior at 2.85 GHz and 8 GHz, with bandwidths of 300 MHz and 400 MHz, an insert loss of less than 5 dB, and a roll-off rate of 25 dB/GHz, which can be expressed as,

$$S_{21}\begin{cases} < -15\,\text{dB}, & f < 2.3\,\text{GHz}; \\ > -5\,\text{dB}, & 2.7\,\text{GHz} < f < 3\,\text{GHz}; \\ < -15\,\text{dB}, & 3.4\,\text{GHz} < f < 7.4\,\text{GHz}. \\ > -5\,\text{dB}, & 7.8\,\text{GHz} < f < 8.2\,\text{GHz}; \\ < -15\,\text{dB}, & f > 8.6\,\text{GHz}. \end{cases} \tag{C.22}$$

The constraints were discretely sampled and then represented into Up_bound $\boldsymbol{E}_u$ and Low_bound $\boldsymbol{E}_d$), as given in (C.23) and (C.24). The $\boldsymbol{E}$ vector was defined as the normalized mean of $\boldsymbol{E}_u$ and $\boldsymbol{E}_d$), as seen in (C.25).

$$\boldsymbol{E}_u = [\overbrace{\ldots,-15}^{length:46}, \overbrace{0,\ldots,\ 0}^{length:22}, \overbrace{-15,\ldots,-15}^{length:80}, \overbrace{0,\ldots,\ 0}^{length:24}, \overbrace{-15,\ldots}^{length:28}]; \tag{C.23}$$

$$\boldsymbol{E}_d = [\underbrace{\ldots,-30}_{length:54}, \underbrace{-5,\ldots,-5}_{length:6}, \underbrace{-30,\ldots,-30}_{length:96}, \underbrace{-5,\ldots,-5}_{length:8}, \underbrace{-30,\ldots}_{length:36}]; \tag{C.24}$$

$$\boldsymbol{E} = \frac{(\boldsymbol{E}_u + \boldsymbol{E}_d)}{2 \times max(|E_{ui}|, |E_{di}|)}, \qquad i = 1, 2, \ldots, 200. \tag{C.25}$$

According to $\boldsymbol{E}_u$ and $\boldsymbol{E}_d$, the topology of $\boldsymbol{T}$= [1, 0] was auto-selected based on the represented dual-band-pass constraints. The metal layers of the topology were divided into 8 sectors, and each sector had 10 fans, because the constraints were symmetrical. Afterward, it went through the auto-evolution system and produced an optimal unit cell to form the required dual-band-pass FSS. The evolutionary history was recorded in Fig. C.10(a), indicating that overall, 12 iterations were taken to arrive at the final design. No noticeable enhancement can be observed after 12 iterations. Fig. C.10(b) depicts the achieved transmission behavior at around 2.85 GHz and 8 GHz. The corresponding unit cell shows its geometry in Fig. C.11 and lists its detailed size in Table C.2.

Using 15 by 15 designed unit cells, a dual-band-pass surface was constructed and a prototype was fabricated and measured to verify the design

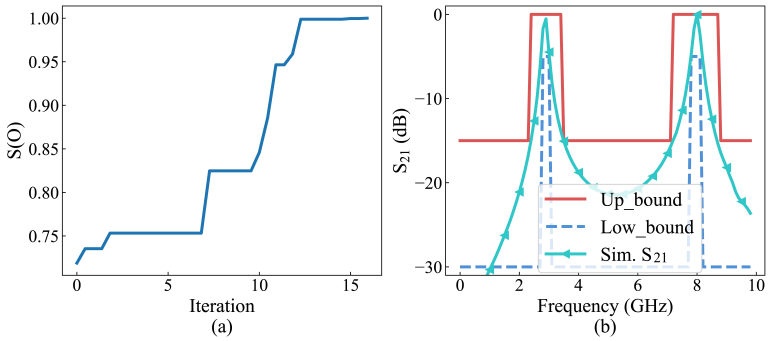


Fig. C.10: Evolution result of the dual-band-pass unit cell: (a) The evolution record over the iterations; (b) The boundaries (Up_bound $\boldsymbol{E}_u$ and Low_bound $\boldsymbol{E}_d$) and simulated S_{21} of the designed unit cell.

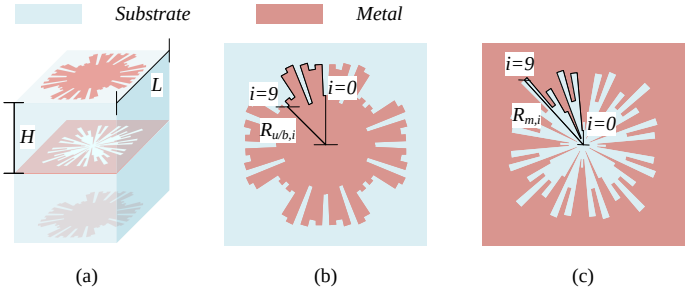


Fig. C.11: Designed unit cell for the dual-band-pass FSS: (a) Its overall structure; (b) Its up/bottom metal layer; (c) Its middle metal layer.

performance. Three metal layers are etched on the surface of two $Rogers4003C$ substrate layers with a thickness of 0.813 mm. The prototype was measured using the same measurement setup. The measured S_{21} is compared with the simulated S_{21} in Fig. C.12. S_{21} was measured from 2 GHz to 10 GHz due to the frequency limitation of the measurement system. Besides, frequencies below 2 GHz are out of the operating band. Within the main operating band, the measured S_{21} agrees well with the simulated S_{21}, which validates the performance of this dual-band-pass design.

3.3 High-Pass

The high-pass FSSs prevent signals below a cut-off frequency [38, 39], which is also a widely needed spatial filtering behavior.

Similarly, we pre-defined the constraints as a high-pass behavior, with a cut-off frequency at 7.5 GHz, an insert loss of less than 4 dB, and a roll-off rate

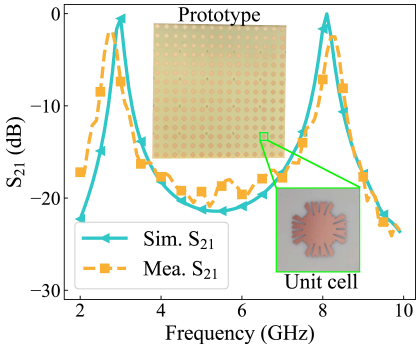


Fig. C.12: Measured S_{21} and simulated S_{21} of the dual-band-pass prototype.

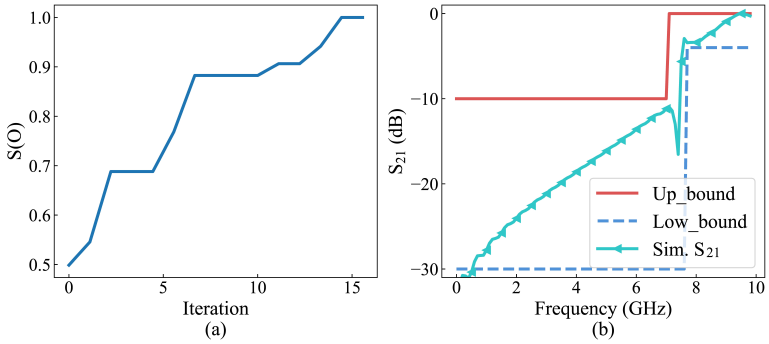


Fig. C.13: Evolution result of the high-pass case: (a) The evolution record over the iterations; (b) The boundaries (Up_bound $\boldsymbol{E}_u$ and Low_bound $\boldsymbol{E}_d$) and simulated S_{21} of the designed unit cell.

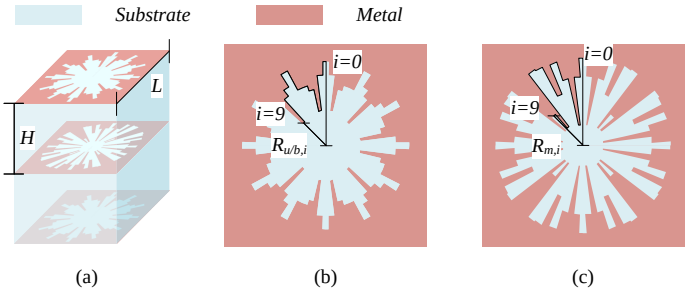


Fig. C.14: Designed unit cell for the high-pass FSS: (a) Its overall structure; (b) Its up/bottom metal layer; (c) Its middle metal layer.

of 15 dB/GHz, which can be expressed as,

$$S_{21}\begin{cases} < -10\ dB, & f < 7\ GHz; \\ > -4\ dB, & f > 7.5\ GHz. \end{cases} \tag{C.26}$$

Table C.3: Parameters of the Designed High-Pass Unit Cell

Parameter	Value (mm)
L	17.32
H	0.305
$\boldsymbol{R}_{u/b}$	[6.59, 5.54, 2.77, 4.48, 4.88, 4.88, 6.59, 5.45, 4.88, 2.52]
$\boldsymbol{R}_m$	[5.63, 4.44, 1.33, 4.37, 5.63, 3.56, 5.70, 5.78, 1.48, 2.67]

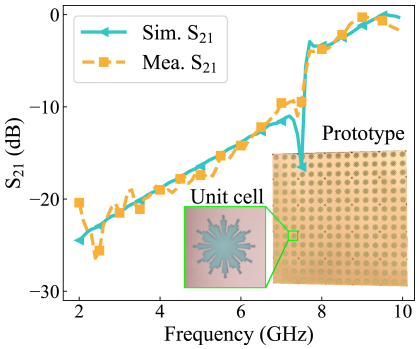


Fig. C.15: Measured and simulated S_{21}s of the high-pass prototype.

The constraints were discretely sampled and represented into two vectors that indicate upper and lower boundaries (Up_bound $\boldsymbol{E}_u$ and Low_bound $\boldsymbol{E}_d$) as given in (C.27) and (C.28). The $\boldsymbol{E}$ vector was fixed as the normalized mean of $\boldsymbol{E}_u$ and $\boldsymbol{E}_d$), as seen in (C.29).

$$\boldsymbol{E}_u = [\overbrace{-10, ..., -10}^{length:140}, \overbrace{0, ..., 0}^{length:60}]; \tag{C.27}$$

$$\boldsymbol{E}_d = [\underbrace{-30, ..., -30}_{length:150}, \underbrace{-4, ..., -4}_{length:50}]; \tag{C.28}$$

$$\boldsymbol{E} = \frac{(\boldsymbol{E}_u + \boldsymbol{E}_d)}{2 \times max(|E_{ui}|, |E_{di}|)}, \qquad i = 1, 2, ..., 200. \tag{C.29}$$

According to $\boldsymbol{E}_u$ and $\boldsymbol{E}_d$, a unit cell with three slotted metal layers ($\boldsymbol{T}$= [0, 0]) was auto-selected given the high-pass filtering constraint. The metal layers of the topology were divided into 8 sectors, and each sector had 10 fans, because the constraints were symmetrical. After a 15-iteration auto-evolution, the unit cell achieved a satisfying high-pass filtering behavior with a cut-off frequency at 6.85 GHz, as shown in Fig. C.13(b). The evolution history is recorded in Fig. C.13(a). The optimal high-pass design and corresponding parameters are given in Fig. C.14 and Table C.3, respectively.

A high-pass surface prototype was fabricated and measured to verify the

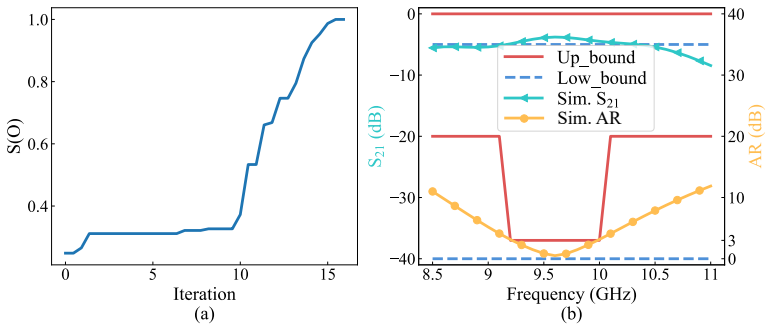


Fig. C.16: Evolution result of the polarizer case: (a) The evolution record over the iterations; (b) The boundaries (Up_bound $\boldsymbol{E}_u$ and Low_bound $\boldsymbol{E}_d$) and simulated S_{21} and AR of the designed unit cell.

simulation results, which consists of 15 by 15 designed high-pass unit cells. Three metal layers are etched on the surface of two $Rogers4003C$ substrate layers with a thickness of 0.305 mm. The prototype was measured using the same measurement setup. The measured and simulated S_{21}s are compared in Fig. C.15. Due to the frequency limitation of the measuring system, S_{21} was measured from 2 GHz to 10 GHz. Additionally, the operating frequency band does not include frequencies below 2 GHz. It can be observed that measurement and simulation agree well within the operating band, which validates the effectiveness of the high-pass design.

3.4 Polarizer

Circular polarization is widely preferred in satellite or point-to-point communication systems due to its immunity to the Faraday rotation effects, polarization mismatch, or multi-path fading issues. Mounting a linear-to-circular polarizer [40, 41] on top of a linearly polarized antenna is a valuable way to produce circular polarization.

We pre-defined the constraints as a linear-to-circular polarization transformation behavior at 9.6 GHz, with a bandwidth of 800 MHz, an insertion loss of less than 5 dB, and an axial ratio (AR) of less than 3 dB, which can be expressed as,

$$\begin{aligned} S_{21} > -5\,\text{dB}, \quad & 9.2\,\text{GHz} < f < 10\,\text{GHz}; \\ AR < 3\,\text{dB}, \quad & 9.2\,\text{GHz} < f < 10\,\text{GHz}. \end{aligned} \tag{C.30}$$

The main difference from the former cases is that now the EM constraints ($\boldsymbol{E}(f)$s) attribute to the transmission coefficient (S_{21}) and the axial ratio (AR) as well. Similarly, we discretely sampled 200 points from 8.5 GHz to 11 GHz and generated upper and lower boundaries (Up_bound $\boldsymbol{E}_{uS21}$&$\boldsymbol{E}_{uAR}$ and Low_bound

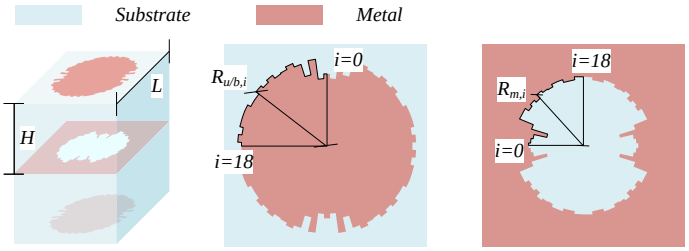


Fig. C.17: Designed unit cell for the polarizer FSS: (a) Its overall structure; (b) Its up/bottom metal layer; (c) Its middle metal layer.

$\boldsymbol{E}_{dS21}$&$\boldsymbol{E}_{dAR}$), as given in (C.31-C.34). The $\boldsymbol{E}$ vector was defined as the normalized weighted mean of the four boundaries, as given in (C.35).

$$\boldsymbol{E}_{uS21} = [\overbrace{0, \ldots, 0}^{length:200}]; \tag{C.31}$$

$$\boldsymbol{E}_{dS21} = [\underbrace{-5, \ldots, -5}_{length:200}]; \tag{C.32}$$

$$\boldsymbol{E}_{uAR} = [\overbrace{20, \ldots, 20}^{length:56}, \overbrace{3, \ldots, 3}^{length:64}, \overbrace{20, \ldots, 20}^{length:80}]]; \tag{C.33}$$

$$\boldsymbol{E}_{dAR} = [\underbrace{0, \ldots, 0}_{length:200}]; \tag{C.34}$$

$$\boldsymbol{E} = \frac{(\boldsymbol{E}_{uS21} + \boldsymbol{E}_{dS21} - \boldsymbol{E}_{uAR} - \boldsymbol{E}_{dAR})}{4 \times max(|E_{uS21i}|, |E_{dS21i}|, |E_{uARi}|, |E_{dARi}|)}, \quad i = 1, 2, \ldots, 200. \tag{C.35}$$

The polarizer should generate a 90 ° phase difference between the orthogonally polarized signals to form circular polarization. Therefore, the metal layers of the topology were divided into 4 sectors, and each sector has 19 fans. The proposed framework completed the design process after 16 iterations. As usual, we recorded the satisfaction level ($S(\boldsymbol{O})$) during evolution in Fig. C.16. S_{21} and AR of the optimal design are shown in Fig. C.16(a) and C.16(b), and its geometry and parameters are shown in Fig. C.17 and Table C.4. The designed polarizer has an insertion loss less than 5 dB and an AR less than 3 dB from 9.2 GHz to 10 GHz.

To validate the simulation results, a polarizer prototype that consists of 15 by 15 designed unit cells was fabricated and measured. Three metal layers are etched on the surface of two $Rogers4003C$ substrate layers with a thickness of 0.813 mm. The polarizer prototype was measured using the same mea-

Table C.4: Parameters of the Designed Polarizer Unit Cell

Parameter	Value (mm)
L	8.2
H	0.813
$\boldsymbol{R}_{u/b}$	[3.20, 3.02, 3.88, 3.24, 3.80, 3.96, 3.84, 3.82, 3.96, 3.86, 3.96, 3.82, 3.80, 3.96, 3.98, 3.96, 3.82, 3.98, 3.80]
$\boldsymbol{R}_m$	[1.26, 1.22, 1.16, 0.86, 1.24, 1.58, 1.60, 1.48, 1.54, 1.48, 1.58, 1.54, 1.52, 1.58, 1.60, 1.52, 1.54, 1.60, 1.58]

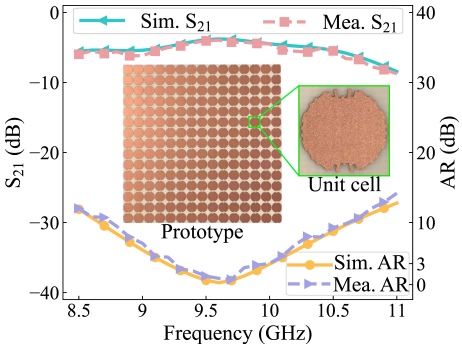


Fig. C.18: Measured and simulated S_{21}s and ARs of the polarizer prototype.

surement setup. Measured S_{21} and AR are compared with simulated ones in Fig. C.18(b). The performance of this polarizer design is confirmed by the measurement results, which are in good agreement with the simulation results.

4 Discussion

The validation results in Section 3 proved that the proposed framework could automatically inverse design a unit cell of the FSS to satisfy any given EM constraints in real scenarios. The implementations suggest that the whole inverse design process required no human experience and took only around 15 iterations and 100+300 samples on average. Noted that several geometries of the designed unit cell are strange, because N_f is set as 72 to show its potential of producing complex and flexible geometries. Such complex geometries may not be suitable for mm-wave or THz frequency bands due to fabrication limitations. For mm-wave or THz frequency bands, N_f can be reduced to generate simpler geometries. The smallest value of N_f is 8, in which case each metal layer would be a circular patch or slot. For example, a simpler band-pass unit cell is designed when N_f is set as 16. The designed simpler unit cell is shown in

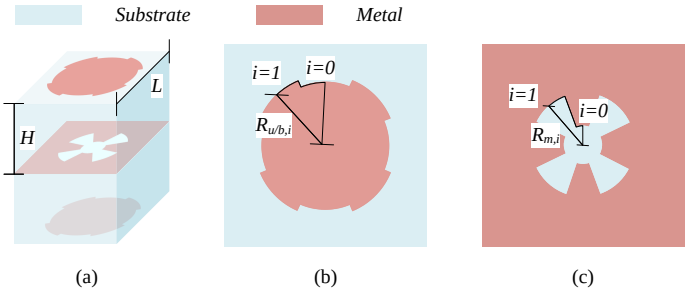


Fig. C.19: Simpler band-pass unit cell with $N_f = 16$: (a) Its overall structure; (b) Its up/bottom metal layer; (c) Its middle metal layer.

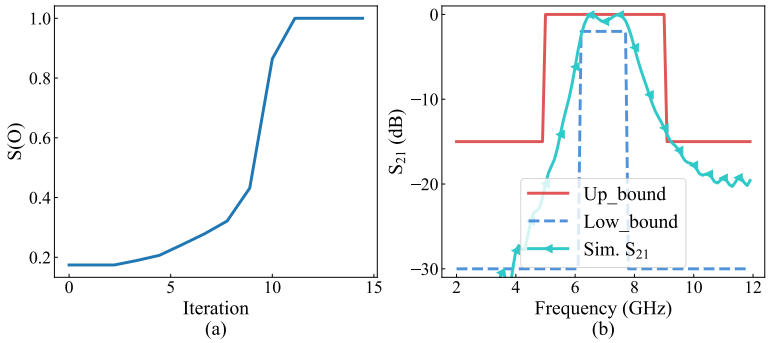


Fig. C.20: Evolution result of the simpler band-pass unit cell: (a) The evolution record over the iterations; (b) The boundaries (Up_bound $\boldsymbol{E}_u$ and Low_bound $\boldsymbol{E}_d$) and simulated S_{21} of the designed simpler unit cell.

Table C.5: Parameters of the Designed Simpler Band-Pass Unit Cell

Parameter	Value (mm)	Parameter	Value (mm)
L	15.23	$\boldsymbol{R}_{u/b}$	[6.53, 6.68]
H	1.524	$\boldsymbol{R}_m$	[1.31, 3.66]

Fig. C.19, and its geometric parameters are listed in Table C.5. A satisfying band-pass performance can be achieved, as shown in Fig. C.20. On the contrary, N_f can also be increased to generate more complex geometries for more complicated constraints in low frequency scenarios.

We compared our framework with existing methods in Table C.6. In our work, the amount of training data refers to the sum of the amount used for training the SVM (100) and the amount used for auto-evolution (300). References [12, 18] both required human engineers to determine the topology and involve intensive data pre-processing pipelines. Unlike them, our framework auto-selects the suitable topology.

Table C.6: Comparison between the Proposed Framework and Existing Methods

	Fully-automated	The amount of training data
[12]	No	588
[18]	No	1 thousand
[21]	Yes	70 thousand
[22]	Yes	21 thousand
The proposed work	Yes	100 + 300

Furthermore, the proposed method requires the smallest amount of training data to arrive at a satisfying design, whereas [21, 22] required a huge amount of training data. The reason is that existing methods often sweep over the whole parameter space to include all possibilities within the training data. We significantly reduce the number of required data thanks to the elaborately organized auto-evolution system. Instead of blindly sweeping over the solution space, our system optimizes the generation of the training data iteratively. The optimized training data have improved informativeness and quality. As a result, the number of required training data is reduced significantly.

The proposed fully-automated framework can be adopted by designers without domain knowledge to realize the inverse design of FSSs for given EM constraints. Thanks to the auto-evolution system, it takes only hundreds of simulation cycles to arrive at the final design. It offers inspiration for the fully-automated inverse design of other EM components. In future work, we will further integrate filter theory and equivalent circuit approach into machine learning-based methods to increase interpretability and to further understand the projection between geometries and responses.

5 Conclusion

We developed a fully-automated framework for the inverse design of FSSs. Unlike conventional inverse design methods that rely on experienced human engineers to determine the topology or involve intensive pre-processing pipelines, our framework can automatically read the physical EM constraints and then decide the appropriate topology. Afterward, the auto-evolution would evolve the selected topology into an optimal design. The auto-evolution system consists of three co-operated modules and two self-monitors to optimize the evolution process and maximize its efficiency. We have validated the effectiveness and efficiency of the presented framework in four inverse design scenarios. The validation results proved that our framework frees human engineers and requires a smaller amount of training data compared with existing inverse design methods. The proposed framework can fully automatize the inverse design of

FSSs and offer inspiration for the automation of other EM applications.

References

[1] S. Pearson and S. V. Hum, "Optimization of electromagnetic metasurface parameters satisfying far-field criteria," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 5, pp. 3477–3488, 2022.

[2] C. Wang, J. Zhang, S. Bai, X. Zhu, and Z. Zheng, "A harmonic suppression energy collection metasurface insensitive to load and input power for microwave power transmission," *IEEE Transactions Microwave Theory and Techniques*, vol. 70, no. 8, pp. 4036–4044, 2022.

[3] C. Y. Tay and Z. N. Chen, "Azimuthally inhomogeneous metasurface cloak for cylindrical objects," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 1, pp. 254–262, 2021.

[4] Y. Wei, J. Duan, H. Jing, Z. Lyu, J. Hao, Z. Qu, J. Wang, and B. Zhang, "A multiband, polarization-controlled metasurface absorber for electromagnetic energy harvesting and wireless power transfer," *IEEE Transactions Microwave Theory and Techniques*, vol. 70, no. 5, pp. 2861–2871, 2022.

[5] J. Zhang, S. Zhang, and G. F. Pedersen, "Dual-band structure reused antenna based on quasi-elliptic bandpass frequency selective surface for 5G application," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 11, pp. 7612–7617, 2020.

[6] W. Tang, S. Mercader-Pellicer, G. Goussetis, H. Legay, and N. J. Fonseca, "Low-profile compact dual-band unit cell for polarizing surfaces operating in orthogonal polarizations," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 3, pp. 1472–1477, 2017.

[7] H. Shan, K. Jiang, J. Xing, and T. Jiang, "BPSO and staggered triangle layout optimization for wideband rcs reduction of pixelate checkerboard metasurface," *IEEE Transactions Microwave Theory and Techniques*, vol. 70, no. 7, pp. 3406–3414, 2022.

[8] K. Yao, R. Unni, and Y. Zheng, "Intelligent nanophotonics: Merging photonics and artificial intelligence at the nanoscale," *Nanophotonics*, vol. 8, no. 3, pp. 339–366, 2019.

[9] F. Güneş, S. Nesil, and S. Demirel, "Design and analysis of minkowski reflectarray antenna using 3-D CST microwave studio-based neural network model with particle swarm optimization," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 23, no. 2, pp. 272–284, 2013.

[10] D. R. Prado, J. A. López-Fernández, M. Arrebola, and G. Goussetis, "Support vector regression to accelerate design and crosspolar optimization of shaped-beam reflectarray antennas for space applications," *IEEE Transactions on Antennas and Propagation*, vol. 67, no. 3, pp. 1659–1668, 2019.

[11] D. R. Prado, J. A. López-Fernández, M. Arrebola, M. R. Pino, and G. Goussetis, "Wideband shaped-beam reflectarray design using support vector regression analysis," *IEEE Antennas and Wireless Propagation Letters*, vol. 18, no. 11, pp. 2287–2291, 2019.

[12] M. Abdullah and S. Koziel, "Supervised-learning-based development of multibit RCS-reduced coding metasurfaces," *IEEE Transactions Microwave Theory and Techniques*, vol. 70, no. 1, pp. 264–274, 2022.

[13] S. Koziel and M. Abdullah, "Machine-learning-powered EM-based framework for efficient and reliable design of low scattering metasurfaces," *IEEE Transactions Microwave Theory and Techniques*, vol. 69, no. 4, pp. 2028–2041, 2021.

[14] J. A. Hodge, K. V. Mishra, and A. I. Zaghloul, "RF metasurface array design using deep convolutional generative adversarial networks," in *IEEE International Symposium on Phased Array Systems & Technology*, 2019, pp. 1–6.

[15] J. A. Fan, "Generating high performance, topologically-complex metasurfaces with neural networks," in *Conference on Lasers and Electro-Optics*, 2019, pp. 1–2.

[16] P. Naseri, G. Goussetis, N. J. Fonseca, and S. V. Hum, "Inverse design of a dual-band reflective polarizing surface using generative machine learning," in *European Conference on Antennas and Propagation*, 2022, pp. 1–5.

[17] L.-Y. Xiao, F.-L. Jin, B.-Z. Wang, Q. H. Liu, and W. Shao, "Efficient inverse extreme learning machine for parametric design of metasurfaces," *IEEE Antennas and Wireless Propagation Letters*, vol. 19, no. 6, pp. 992–996, 2020.

[18] R. Zhu, J. Wang, Y. Han, S. Sui, T. Qiu, Y. Jia, M. Feng, X. Wang, L. Zheng, and S. Qu, "Design of aperture-multiplexing metasurfaces via back-propagation neural network: Independent control of orthogonally-polarized waves," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 6, pp. 4569–4575, 2022.

[19] Z. Wei, Z. Zhou, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Equivalent circuit theory-assisted deep learning for accelerated generative design of metasurfaces," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 7, pp. 5120–5129, 2022.

[20] P. Naseri and S. V. Hum, "A generative machine learning-based approach for inverse design of multilayer metasurfaces," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 9, pp. 5725–5739, 2021.

[21] P. Naseri, S. Pearson, Z. Wang, and S. V. Hum, "A combined machine-learning/optimization-based approach for inverse design of nonuniform bianisotropic metasurfaces," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 7, pp. 5105–5119, 2022.

[22] C. C. Nadell, B. Huang, J. M. Malof, and W. J. Padilla, "Deep learning for accelerated all-dielectric metasurface design," *Optics Express*, vol. 27, no. 20, pp. 27 523–27 535, 2019.

[23] M. Ohira, H. Deguchi, M. Tsuji, and H. Shigesawa, "Multiband single-layer frequency selective surface optimized by genetic algorithm with geometry-refinement technique," in *IEEE Antennas and Propagation Society International Symposium*, 2003, pp. 833–836.

[24] M. Gingrich and D. H. Werner, "Synthesis of zero index of refraction metamaterials via frequency-selective surfaces using genetic algorithms," in *IEEE Antennas and Propagation Society International Symposium*, 2005, pp. 713–716.

[25] L. Lanuzza, A. Monorchio, and G. Manara, "Synthesis of zero index of refraction metamaterials via frequency-selective surfaces using genetic algorithms," in *IEEE Antennas and Propagation Society International Symposium*, 2002, pp. 364–367.

[26] M. O. Akinsolu, K. K. Mistry, B. Liu, P. I. Lazaridis, and P. Excell, "Machine learning-assisted antenna design optimization: A review and the state-of-the-art," in *European Conference on Antennas and Propagation*, 2020, pp. 1–5.

[27] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[28] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, "Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription," in *International Conference on Machine Learning*, 2012, pp. 1881–1888.

[29] R. Bro and A. K. Smilde, "Principal component analysis," *Analytical Methods*, vol. 6, no. 9, pp. 2812–2831, 2014.

[30] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Language*, vol. 20, no. 3, pp. 273–297, 1995.

[31] *High Frequency Electronics Product Selector Guide*, Rogers Corporation, 2021, pp. 14–15.

[32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015, pp. 448–456.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, Y. Bengio and Y. LeCun, Eds., 2015, pp. 1–15.

[34] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *International Conference on Machine Learning*, 2010, pp. 1–8.

[35] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *Proceedings of the IEEE Conference on Evolutionary Computation*, 1998, pp. 69–73.

[36] L. Zhou and Z. Shen, "Hybrid frequency-selective rasorber with low-frequency diffusion and high-frequency absorption," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 3, pp. 1469–1476, 2021.

[37] A. Lalbakhsh, M. U. Afzal, K. P. Esselle, and S. L. Smith, "All-metal wideband frequency-selective surface bandpass filter for te and tm polarizations," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 4, pp. 2790–2800, 2022.

[38] Z.-C. Hao and J.-S. Hong, "UWB bandpass filter using cascaded miniature high-pass and low-pass filters with multilayer liquid crystal polymer technology," *IEEE Transactions Microwave Theory and Techniques*, vol. 58, no. 4, pp. 941–948, 2010.

[39] Y.-J. Huang, C.-H. Hsieh, and Z.-M. Tsai, "A compact, high-selectivity, and wide passband semi-lumped 70 MHz high-pass filter," in *IEEE Asia Pacific Microwave Conference*, 2017, pp. 730–733.

[40] S. M. A. M. H. Abadi and N. Behdad, "Wideband linear-to-circular polarization converters based on miniaturized-element frequency selective surfaces," *IEEE Transactions on Antennas and Propagation*, vol. 64, no. 2, pp. 525–534, 2016.

[41] S. V. Samsonov, A. A. Bogdashov, and I. G. Gachev, "Waveguide linear-to-circular polarization converter with cross polarization below −40dB within 16% band," *IEEE Transactions Microwave Theory and Techniques*, vol. 70, no. 4, pp. 2108–2114, 2022.

Paper D

A High-Quality Data Acquisition Method for Machine-Learning-Based Design and Analysis of Electromagnetic Structures

Zhao Zhou, Zhaohui Wei, Abdullah Tahir, Jian Ren, Yingzeng Yin, Gert Frølund Pedersen, and Ming Shen

The paper has been published in the
IEEE Transactions on Microwave Theory and Techniques Vol. 71(10),
pp. 4295–4306, 2023.

The layout has been revised.

Abstract

Electromagnetic structures play a significant role in wireless communication, radar detection, medical imaging, etc. Machine learning has been increasingly applied to facilitate the design and analysis of electromagnetic structures. Data acquisition is a major bottleneck. Conventional methods blindly sweep geometric parameters on a uniform grid and acquire corresponding responses via simulation. Acquired data have unstable quality due to inconsistent informativeness of responses, leading to a low ratio of model performance to data amount. This paper proposes a high-quality data acquisition method to increase the ratio of model performance to data amount. It anticipates and generates high-quality data by analyzing distribution of existing data iteratively. Comparative analysis of four implementations proves, the proposed method reduces required data amount by around 40 % *for the same model performance, hence saves around* 40 % *simulation and computing resources. The proposed method benefits machine learning applications of metasurfaces, antennas, and many other microwave structures.*

1 Introduction

Wireless communication is substantially impacted by electromagnetic (EM) structures. EM structures are deliberately designed to satisfy EM constraints in a practical scenario. EM constraints refer to requirements for the structures, such as size, operating frequency, gain, axial ratio (AR), reflection, transmission, scattering coefficient, etc. Accordingly, EM response describes how a structure affects wireless signals within the concerning frequency range. The conventional design of EM structures relies on experienced human engineers. Firstly, engineers determine a topology and initialize its geometric parameters to form a draft structure for the topology according to the EM constraints and their experience. Afterward, they evaluate its EM response through full-wave simulation via EM simulation software (for example, Computer Simulation Technology®(CST)) and tune its geometric parameters based on their understanding of the correlation between parameters and EM responses iteratively. The number of iterations needed varies, depending on the designer's experience depth. Therefore, machine learning (ML) is increasingly being studied and applied in EM applications to improve the current EM solutions. However, data acquisition is a significant barrier for ML [1], especially for EM-related ML applications.

Most EM-related ML applications require [2–4] labeled data set consists of geometric parameters and corresponding EM responses. EM-related ML applications can be roughly divided into three categories, forward synthesis, inverse design, and generative method [5–14]. Forward synthesis establishes a model to imitate the projection from geometric parameters to EM

responses [15–22]. The forward model utilizes supervised learning because both geometric parameters and corresponding EM responses are needed as the training data set. After training, the model is used to replace EM simulation software and is often integrated with optimization algorithms to reach an optimal structure design. To start with, a draft structure is initialized with arbitrary geometric parameters, and the forward model synthesizes its responses. Afterward, the optimization algorithm updates the geometric parameters based on the difference between current responses and constraints. Inverse design develops a model to directly determine geometric parameters for given EM constraints, which also uses supervised learning [23–26]. The well-trained model acts as a dictionary that records the projection from EM responses to suitable geometric parameters. Generative method utilizes autoencoder (VAE) or generative adversarial network (GAN) to learn characteristics of real geometric parameters [27]. After training, the generative model referred to as generator can generate new structures with similar geometric parameters. However, the generated new structure still requires full-wave simulation to evaluate its EM responses. To reduce the need for full-wave simulation, the generator is sometimes integrated with a forward synthesis model in the real design process. To start with, a noise vector is randomly initialized as the first input for the generator, and the generator forms an initial structure accordingly. The EM responses of the structure are then evaluated using a forward synthesis model. By comparing its responses and constraints, the optimization algorithm updates the noise vector accordingly. The updated vector is taken as a new input for the generator to generate an updated structure, and it is then evaluated and updated again and again. The iteration stops when the updated structure fulfills the EM constraints.

There are two ways to acquire labeled training data for EM-related ML applications. The first way is to fabricate a prototype with respect to each setting of geometric parameters and then measure its EM responses to form each training data sample. Measuring EM responses require specific measurement devices, such as Vector Network Analyzers for reflection or transmission coefficients and anechoic chamber for gain or AR. To collect N training data samples, the designer need to adjust geometric parameters of the prototype and measure responses for N times if the prototype is adjustable, otherwise the designer has to fabricate N prototypes and measure their responses individually. Since acquiring EM data through measurement is costly and restricted, designers normally use full-wave simulation for collecting EM training data. Supported by simulation software such as CST, designers can build a virtual prototype and simulate its responses by using a computer. Full-wave simulation does not require fabrication or measurement devices, but relies on computation resources.

ML applications in EM are often criticized because it needs a large number of full-wave simulation cycles to generate sufficient training data, hence occu-

pies many computation resources. Most works define all settings of geometric parameters first and then simulate corresponding EM responses individually to collect the training data set. The defined geometric parameters usually distribute on a uniform grid within the parameter space, because designers cannot anticipate distribution of responses and the best policy is to uniformly cover the whole parameter space. However, EM responses are extremely sensitive to geometric parameters and do not distribute uniformly. With respect to different areas within the parameter space, the changes of EM responses may be slightly or significantly. Samples in parameter areas where the EM responses change significantly can greatly affect the model performance. These samples are referred to as high-quality data. By contrast, those in parameter areas where EM responses change slightly contribute little to the model performance, which are referred to as low-quality data. Low-quality samples occupy a number of unnecessary simulation cycles. However, designers cannot recognize low-quality data before simulation and avoid unnecessary simulation cycles.

An intelligent high-quality data acquisition method is demanded for EM-related ML. However, most data acquisition methods [28–32] are not suitable for EM-related ML applications. There have been many great works [33–41] that attempted to identify the most promising region of the parameter space and further tune the design by means of local routines. They improved the global optimization of expensive EM simulation models significantly. They focused on fast convergence of the promising region and the optimal design, instead of generation of a high-quality dataset for the ML model. A high-quality dataset should be informative and representative for the whole parameter space, so that the ML model can learn the intelligence of the whole parameter space. A high-quality data acquisition method is expected to generate a more informative and representative dataset with the smallest amount of samples. It helps ML models obtain the same performance by using a smaller amount of training data, resulting in the reduction of burden on simulation.

This paper proposes a high-quality data acquisition method. The objective is to improve the quality of data and reduce the need of simulation for ML-based design of EM structures. Quality of data for ML is measured based on the performance of the ML model. High-quality data can improve the performance of the ML model significantly. To start with, a small amount of training data samples are initialized, which are defined on a uniform grid within the parameter space. Existing data samples are analyzed with respect to the distribution of parameters and responses to recognize a parameter area where the EM responses change significantly. Afterward, a new data sample is generated by defining its geometric parameters through swarm operation in the selected area, and its EM responses are simulated through simulation. The new data sample is considered of high quality, because it is likely to improve the performance of the ML model significantly. The new high-quality

data sample is then added into the existing data samples, and a new round of analysis and generation begins. The existing data samples are iteratively analyzed and expanded, and an increasing number of high-quality data samples are generated. The iteration stops when sufficient data samples have been collected. Unlike conventional data acquisition methods that uniformly sweep geometric parameters on a constant grid of the parameter space, the proposed method adjusts parameter definition dynamically according to the quality of parameter area by analyzing the distribution of existing samples. The proposed method can maximize the quality of training data set with a reduced amount of simulation cycles. Based on the comparative results in four implementations, the proposed method significantly reduces the amount of training data samples required to reach the same model accuracy, hence a significant amount of simulation cycles are saved, and computation resources are greatly released.

The remaining content is arranged as follows: Section II introduces the algorithm of the proposed method; Section III validates the proposed method in four implementations; Section IV gives the conclusion.

2 Algorithm

Pseudo code for the proposed method is demonstrated in Algorithm 1 and 2. The proposed algorithm is established specially according to the requirements of data acquisition in the ML-based design of EM structures. Input features of this ML task in EM are represented as normalized vectors $\boldsymbol{X}$s, whereas, $\boldsymbol{Y}$s represent the output features. Here, each $\boldsymbol{Y}$ is obtained by full-wave EM simulation via CST for a given $\boldsymbol{X}$. The proposed method comprises of two major steps. The first step is to initialize the initial set of N_0 samples. The input vectors $\boldsymbol{X}$s of the N_0 samples are defined in a uniform manner. The values of $\boldsymbol{X}$ of the first sample are set as the minimum values within the parameter range. The values of $\boldsymbol{X}$s of the following samples gradually increase by a constant increment. The increment is decided by the parameter range and the number of initial samples N_0. The values of $\boldsymbol{X}$ of the last sample are the maximum values within the parameter range. The initial set of samples distribute uniformly within the parameter range. It ensures that the parameter space is represented and covered unbiasedly for avoiding uncertainty caused by initialization. Importantly, N_0 is significantly smaller than the number of samples required in common ML tasks. The second step is to analyze the existing samples and produce samples of high quality, iteratively. This step is integrated with online model training to abort the iteration as soon as adequate samples have been acquired and the model loss for the expected test set reaches the minimum threshold. It is worth noting that the proposed method can also be used independently without being integrated with the model training

Algorithm 1 The proposed data acquisition method (Part 1)

Require:

Variables to be fixed:
- $\boldsymbol{X}$, $\boldsymbol{Y}$: normalized input, output vector
- T: integer, maximum data acquisition iteration
- $\boldsymbol{X}_{step}$: vector, minimum step of input features
- *min_loss*: float, expected minimum loss
- N_0: integer, number of initial data samples

Built-in variables and functions:
- i: integer, index of elements in $\boldsymbol{X}$
- j: integer, index of sample within the data set
- t: integer, index of data acquisition iteration
- $\{(\hat{\boldsymbol{X}}_j, \hat{\boldsymbol{Y}}_j)\}$: reference samples
- N_t: integer, number of existing samples in iteration t
- k: integer, index of selected sample
- c_t: float, between 0 and 1, depend on t
- $(\boldsymbol{X}_*, \boldsymbol{Y}_*)$: new sample
- *model_loss*: float, model loss after training
- *Simulate*: full-wave EM simulation via CST
- *Dist*: calculate distance between two vectors

Initialize: ▷ Step 1

$\{(\boldsymbol{X}_j, \boldsymbol{Y}_j)\}_{N_0}, \quad j = 0, 1, ..., N_0$

$\boldsymbol{X}_j = [x_{j,0}, x_{j,1}, ...], \quad x_{j,i} \in \{0, X_{step,i}, 2X_{step,i}, ..., 1\}$

$\boldsymbol{Y}_j \leftarrow Simulate(\boldsymbol{X}_j)$

$c_t = 0$

process. In that case, the iteration stops when a sufficient number of samples are obtained.

For each iteration t in the second step, N_t existing samples $\{(\boldsymbol{X}_j, \boldsymbol{Y}_j)\}_{N_t}$ are analyzed to generate a new sample of high quality. The second step can be further divided into eight sub-steps (a-h) as follows.

(a) Pick a reference input vector $\hat{\boldsymbol{X}}_j$ for each existing input vector $\boldsymbol{X}_j$, by minimizing the distance between $\hat{\boldsymbol{X}}_j$ and $\boldsymbol{X}_j$, while making sure that all the elements of $\hat{\boldsymbol{X}}_j$ are equal to or bigger than those of $\boldsymbol{X}_j$ and at least one element of $\hat{\boldsymbol{X}}_j$ is bigger than that of $\boldsymbol{X}_j$.

(b) Form a list of reference samples $\{(\hat{\boldsymbol{X}}_j, \hat{\boldsymbol{Y}}_j)\}_{N_t}$ with respect to existing samples $\{(\boldsymbol{X}_j, \boldsymbol{Y}_j)\}_{N_t}$.

(c) Pick one input vector $\boldsymbol{X}_k$ by maximizing the distance ($Dist(\boldsymbol{Y}_k, \hat{\boldsymbol{Y}}_k)$) between its output $\boldsymbol{Y}_k$ and its reference output $\hat{\boldsymbol{Y}}_k$. The selected sample

Algorithm 2 The proposed data acquisition method (Part 2)

Acquisition: ▷ Step 2

for $t = N_0$ to $T-1$ **do**
  for $\hat{j} = 0$ to t **do**
    $\hat{\boldsymbol{X}}_{\hat{j}} = \boldsymbol{X}_j \leftarrow \underset{j, \boldsymbol{X}_j > \boldsymbol{X}_{\hat{j}}}{argmin}\{\{Dist(\boldsymbol{X}_j, \boldsymbol{X}_{\hat{j}})\}_{N_t}\}$ ▷ (a)
  end for
  $\{(\hat{\boldsymbol{X}}_j, \hat{\boldsymbol{Y}}_j)\}_{N_t} = \{(\hat{\boldsymbol{X}}_{\hat{j}}, \hat{\boldsymbol{Y}}_{\hat{j}})\}_{N_t}$ ▷ (b)
  while True **do**
    $\boldsymbol{X}_k, \hat{\boldsymbol{X}}_k = \underset{j}{argmax}\{Dist(\boldsymbol{Y}_j, \hat{\boldsymbol{Y}}_j)\}_{N_t}$ ▷ (c)
    if True in $Dist(\boldsymbol{X}_k, \hat{\boldsymbol{X}}_k) \geq 2\boldsymbol{X}_{step}$ **then** ▷ (d)
      $Break$
    else
      $Dist(\boldsymbol{X}_k, \hat{\boldsymbol{X}}_k) = 0$
    end if
  end while
  Generate new data set:
    $\boldsymbol{X}_* = c_t \cdot \boldsymbol{X}_k + (1 - c_t) \cdot \hat{\boldsymbol{X}}_k$ ▷ (e)
    $c_t = \frac{rand(0,1) + c_t \cdot (t - N_0)}{t - N_0 + 1}$ ▷ (f)
    $\boldsymbol{Y}_* = Simulate(\boldsymbol{X}_*)$ ▷ (g)
    $\{(\boldsymbol{X}_j, \boldsymbol{Y}_j)\}_{N_t}.append((\boldsymbol{X}_*, \boldsymbol{Y}_*))$ ▷ (h)
    $N_t = t + 1$
  $model\_loss \leftarrow$ **Train model** using $\{(\boldsymbol{X}_j, \boldsymbol{Y}_j)\}_{N_t}$
  if $model\_loss \leq min\_loss$ **then**
    Break
  end if ▷ (Optional) Integrated with model training
end for

$(\boldsymbol{X}_k, \boldsymbol{Y}_k)$ and its reference sample $(\hat{\boldsymbol{X}}_k, \hat{\boldsymbol{Y}}_k)$ point at an input space where the sample of high quality exists. The underlying reason is that selected samples have a large distance between their output vectors and a small distance between their input vectors, implying that they confine a parameter space where the output is sensitive to the input. A sensitive parameter space is difficult for prediction. Hence, it is likely that this space contribute a lot to the prediction error of the ML model. On the other hand, adding a new sample in this space can improve the model performance significantly. In this sense, a potential high-quality sample of high informativeness referred to as $(\boldsymbol{X}_*, \boldsymbol{Y}_*)$ can be generated from the selected sample $(\boldsymbol{X}_k, \boldsymbol{Y}_k)$ and its reference sample $(\hat{\boldsymbol{X}}_k, \hat{\boldsymbol{Y}}_k)$.

(d) Examine if the absolute difference between $\boldsymbol{X}_k$ and $\hat{\boldsymbol{X}}_k$ exceeds $2\times\boldsymbol{X}_{step}$.

If yes, it guarantees an input space large enough to generate a new input vector; otherwise, then repeat sub-step (c) after excluding these two samples.

(e) Generate a new input vector $\boldsymbol{X}_*$ from the selected input vector $\boldsymbol{X}_k$ and its reference vector $\hat{\boldsymbol{X}}_k$ through formula (D.1):

$$\boldsymbol{X}_* = c_t \cdot \boldsymbol{X}_k + (1 - c_t) \cdot \hat{\boldsymbol{X}}_k. \tag{D.1}$$

The input vector of the new sample is obtained through swarm operation, as given in formula (D.1), on the selected input vector $\boldsymbol{X}_k$ and its reference input vector $\hat{\boldsymbol{X}}_k$. Here, the input vector of the new sample is referred to as $\boldsymbol{X}_*$ and it is set as the weighted sum of the $\boldsymbol{X}_k$ and $\hat{\boldsymbol{X}}_k$. The weights for $\boldsymbol{X}_k$ and $\hat{\boldsymbol{X}}_k$ are c_t and $1 - c_t$, respectively.

(f) Determine the weight c_t through formula (D.2):

$$c_t = \frac{rand(0,1) + c_t \cdot (t - N_0)}{t - N_0 + 1}. \tag{D.2}$$

The definition of c_t reaches a balance between exploration and exploitation. Fig. D.1 shows that c_t has initial value between 0 and 1 and approaches 0.5 as the iteration t continues. This definition offers freedom to explore the input space in the earlier iterations as c_t can take any value between 0 and 1. With c_t being closer to 0, the new input vector $\boldsymbol{X}_*$ is closer to the selected input vector $\boldsymbol{X}_k$. As if c_t gets closer to 1, the new input vector $\boldsymbol{X}_*$ approaches $\hat{\boldsymbol{X}}_k$. As the iteration continues, c_t gets close to 0.5, and the input vector of the new sample $\boldsymbol{X}_*$ approaches the average of $\boldsymbol{X}_k$ and $\hat{\boldsymbol{X}}_k$. This ensures a minimum distance between the new input vector and the selected input vectors, because the space for exploration between $\boldsymbol{X}_k$ and $\hat{\boldsymbol{X}}_k$ decreases as the iteration continues. A new sample which is too close to an existing sample, lacks informativeness as it is likely to behave approximately the same as that existing sample.

(g) Obtain the output label vector $\boldsymbol{Y}_*$ for this new input vector $\boldsymbol{X}_*$ through full-wave simulation via CST.

(h) Form a new sample ($\boldsymbol{X}_*$, $\boldsymbol{Y}_*$), add this new sample into the existing data set, and repeat sub-steps (a-h) until sufficient data have been acquired.

In Algorithm 1 and 2, the proposed data acquisition method is integrated with model training. As the data set is being updated continuously by adding new high-quality samples, a model is trained in the meantime. After training, the model is tested on a fixed test set which is pre-defined according to the practical needs. The whole data acquisition procedure ceases as soon as the test loss reaches the desired minimum loss *min_loss*. Note that the proposed

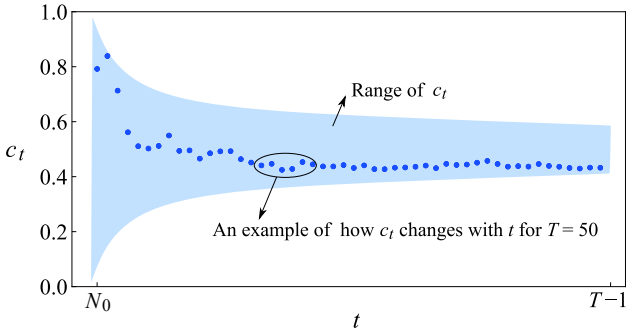


Fig. D.1: Values of c_t versus iteration t.

data acquisition method can also be employed without integration with model training. In that case, the iterations in the second step will stop when a sufficient number of samples have been generated.

3 Implementation

The proposed method is validated by comparing the results without and with the proposed method in four implementations. We utilize the proposed data acquisition method and re-implemente the design and analysis of Modified Jerusalem Cross (MJC) reflective surfaces [2] in Section 3.1, multi-bit coding metasurface for radar cross-section (RCS) reduction [3] in Section 3.2, array radiation synthesis [4] in Section 3.3, and larger array radiation synthesis in Section 3.4 to analyze the performance of our methodology. Comparison of training results in the implementations validates high quality data generation by the proposed method. It is worth noting that we directly use the prior information (e.g., griding space, parameter range, minimum loss, etc.) from the original implementations for fairly comparing with original results. When the proposed method is applied to a new unknown implementation, this prior information can be easily acquired according to domain knowledge and its concrete EM requirements.

3.1 Implementation A: MJC Reflective Surface

3.1.1 Implementation Description

The authors in [2] proposed a Modified Jerusalem Cross (MJC) reflective surface that offers independent control of orthogonally-polarized signals. The MJC reflective surface was designed to operate at 10 GHz. The structure of its

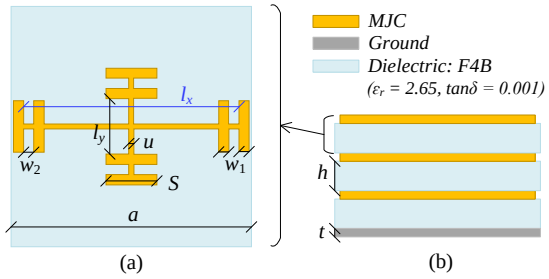


Fig. D.2: Implementation A: Structure of Modified Jerusalem Cross-based unit cell: (a) Top view; (b) Side view. [2]

Table D.1: Implementation A: Geometric Parameters of the Modified Jerusalem Cross-based Unit Cell [2]

Parameter	l_x	l_y	w_1	w_2	a	u	S	h	t
Value(mm)	[1.9, 4.4]	1.1	0.2	0.2	5	0.1	1	2	0.017

unit cell is given in Fig. D.2. The unit cell consists of three overlapped dielectric (F4B) layers, three identical metal MJCs printed on top of each dielectric layer, and a full metal layer as the ground at the bottom. A MJC is composed of two orthogonally-crossed metal bars. The length of each bar (l_x/l_y) can be adjusted independently to tune the reflective phase (φ_x/φ_y) for the corresponding polarization. Implementation A works on adjusting the length of the bar in x direction (l_x) independently for tuning the reflective phase in the x-polarization.

To conveniently design the length l_x for any desired phase φ_x, Zhu. et al. utilized a backpropagation neural network (BPNN) to learn the mapping from the reflective phase φ_x to the length l_x. l_x varies from 1.9 mm to 4.4 mm, while the rest of the geometric parameters were fixed as given in Table D.1. l_x is marked in blue color in Fig. D.2 and in Table D.1. The detailed architecture of BPNN is listed in Table D.2. It can be observed that the BPNN consists of an input layer which takes the phase as input, a hidden layer of 20 neurons with activation function as Tanh [42], and an output layer that outputs l_x. The BPNN uses the Mean Squared Error (MSE) of the predicted and real l_x as its loss function and Levenberg-Marquardt is set as the backpropagation algorithm. 1000 samples were acquired by sweeping the length at a constant step of 0.0025 mm, among which 700 and 150 samples were used for training and validation, and 150 samples for test, respectively. The minimum loss was 5.01×10^{-6}.

Table D.2: Implementation A: Architecture of the Backpropagation Neural Network (BPNN) [2]

No.	Layer	Neurons	Function
1	Input layer	1	Input: phase φ_x
2	Hidden layer	20	Fully-connected layer
3	Transfer function	-	Tanh
4	Output layer	1	Output: length l_x
5	Transfer function	-	Linear
-	**Loss function**	-	Mean Squared Error
-	**Algorithm**	-	Levenberg-Marquardt

3.1.2 Re-implementation

For re-implementation, the proposed method is used to generate high-quality data for training the same BPNN. Variables of the proposed method are fixed as listed in Table D.3. Here, the input vector $\boldsymbol{X}$ represents the phase φ_x of size 1 and the output vector $\boldsymbol{Y}$ represents the length l_x of size 1. Maximum number of iteration for data acquisition T is fixed at 700, which is also the number of training samples used in [2]. $\boldsymbol{X}_{step}$ is fixed at [0.0025 mm], which is the smallest step of l_x considered in [2] and a common fabrication tolerance. *min_loss* is set as 5.01×10^{-6}, which is also the test loss reported in [2]. The only adjustable variable left is N_0, which is marked in blue color as shown in Table D.3. Number of training samples used to reach minimum loss *min_loss* of 5.01×10^{-6} (which was actually achieved by [2]), with and without our method are considered for comparison.

The proposed method is performed four times for four different values of N_0s (50, 80, 100, 150). The model uses the architecture introduced in [2], as shown in Table D.2. The number of required training samples (Ns) and the final losses for 150 validation and 150 test samples (Ls) corresponding to 4 different N_0s are listed in Table D.4. The 150 validation and 150 test samples are generated by arbitrarily setting l_x between 1.9 mm and 4.4 mm with a step size of 0.0025 mm, in the same manner as introduced in [2]. The results suggest that $N_0 = 80$ requires the lowest number of training samples ($N = 80 + 65 = 145$, $L = 4.81 \times 10^{-6}$) to converge beneath the *min_loss* (marked in blue color in Table D.4). When $N_0 = 50$, it fails to converge towards *min_loss* because 50 initialized samples do not provide sufficient information. When N_0 rises up to 100 and 150, more training samples ($N = 100 + 61 = 161$ and $N = 150 + 52 = 202$) are required to reach the *min_loss*. The increment of training samples is mainly caused by increasing initialized samples because the amounts of new samples are approximately the same when N_0 is set as 80, 100, or 150. To sum up, N_0 should be large enough to offer adequate beginning data while still being as minimal as feasible to prevent taking up needless simulation cycles. N_0 is set as 80 in Implementation A. The $N_0 = 80$ samples

Table D.3: Re-implementation A: Variables of the Proposed Method

Variable	X	Y	T	X_{step}	min_loss	N_0
Value	$[l_x]$	$[\varphi_x]$	700	[0.0025 mm]	5.01×10^{-6}	80

Table D.4: Re-implementation A: Comparison of Results Without [2] and With the Proposed Data Acquisition Method with Various N_0

	Without	**With the proposed method with N_0 as:**			
	[2]	50	80	100	150
N	700	Fail	80 + 65	100 + 61	150 + 52
Time	7 h	-	1.45 h	1.61 h	2.02 h
***L*(10^{-6})**	5.01	Fail	4.81	4.89	4.90

Note: ***N*** is the number of training samples, which is 700 in [2];
Time is the time consumed for full-wave simulation.
L is the final loss, which is 5.01×10^{-6} in [2].

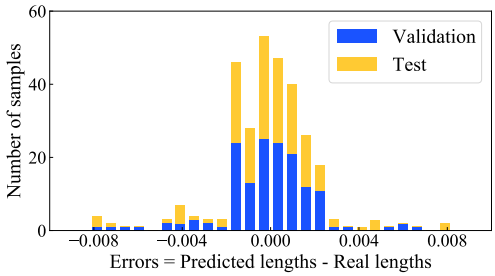


Fig. D.3: Re-implementation A: Number of samples versus the errors between predicted lengths and real lengths.

are initialized with l_x being set from 1.9 mm to 4.4 mm at a constant step.

3.1.3 Comparison of Training Results

Table D.4 compares the results obtained without [2] and with using the proposed method. Authors in [2] collected training samples by sweeping l_x at a constant step. 700 training samples were required in [2] to achieve an average loss of 5.01×10^{-6} and an error range of ±0.008 for 150 validation and 150 test samples. For comparison, the proposed method is integrated with model training to re-implement the work in [2]. The model is also tested for 150 validation and 150 test samples. At iteration $t = 144$, 145 training samples have been generated and are used to train the model. The average loss for 150 validation and 150 test samples is 4.81×10^{-6} and the error range is ±0.008 as shown in Fig. D.3. Note that only validation and test losses are compared for two reasons:

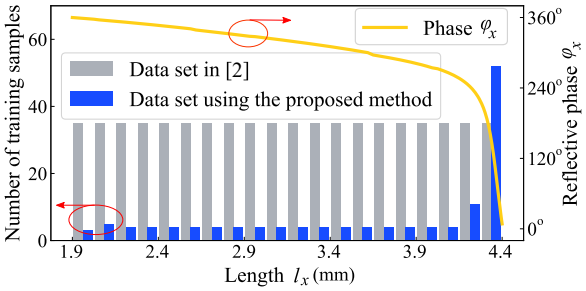


Fig. D.4: Re-implementation A: Comparison of distributions of the data set in [2] and the high-quality data set using the proposed method.

the same number of validation (150) and test (150) samples are collected in the same manner as in [2]; the number of training samples are different, which will lead to unfair comparison of the training losses. Numerical results suggest that only 20.71 % number of training samples are required to realize comparable model performance by using the proposed method. By comparison, the proposed data acquisition method saves 79.29 % training samples to achieve similar model performance, hence 79.29 % simulation cycles are saved in implementation A. Each simulation cycle takes around 36 seconds, with a computer equipped with 96 GB RAM and Intel® Xeon® Silver 4208 CPU @ 2.10 GHz (2 processors). Therefore, 5.55 hours full-wave simulation time is saved in implementation A by using the proposed method.

3.1.4 Analysis and Discussion

The distribution of the data set in [2] and the high-quality data acquired by using the proposed method are investigated to explore the underlying reason for saving 79.29 % training samples in Fig. D.4. The data set used in [2] was collected by sweeping at a uniform step of 0.0025 mm, hence following a uniform distribution. The high-quality data set is collected adaptively to the output distribution governed by the proposed data acquisition method. Phase φ_x decreases drastically When length l_x varies from 4.1 mm to 4.4 mm, while it shows incremental decrease as length l_x increases from 1.9 mm to 4.1 mm. Therefore, intensive samples are collected with l_x being fixed between 4.1 mm and 4.4 mm, and only a few samples are collected out of this range. This adaptive strategy provides more valuable information with a small number of samples. As a result, 79.29 % training samples are saved, hence 79.29 % simulation cycles are saved by using the proposed method.

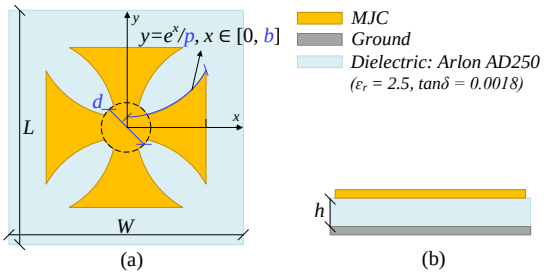


Fig. D.5: Implementation B: Structure of Crusader cross-based unit cell: (a) Top view; (b) Side view. [3]

3.2 Implementation B: RCS Reduction Metasurface

3.2.1 Implementation Description

The authors in [3] proposed a multi-bit coding metasurface for radar cross section (RCS) reduction. A x-bit metasurface consists of 2^x groups of unit cells. Different groups of unit cells have reflection phases incrementally increasing from 0 at a uniform step of $\frac{2\pi}{2^x}$, while unit cells within each group have identical reflection phases. The topology of unit cells resembles the Crusader cross, as shown in Fig. D.5. The structure of a unit cell is determined by three geometric parameters, p, b, and d. By adjusting p, b, and d, its structure can be modified, and its reflection phase can be changed accordingly. The unit cell's overall size and thickness were fixed as constant values, as shown in Table D.5.

To facilitate the multi-bit metasurface design process, the authors constructed a surrogate model to predict the reflection phase of any given unit cell. The input of the surrogate model was set as $[p, b, d]$ because the unit cell is determined by the three geometric parameters. $[p, b, d]$ was confined within a 3D parameter space defined by an upper bound $[3.5, 0.3, 0.2]$ and a lower bound $[10, 1.6, 2.4]$, as seen in Table D.5. The 3D space was divided uniformly into $7 \times 12 \times 7$, and 588 uniformly distributed inputs were defined. The 588 samples were simulated via full-wave simulation supported by CST. Each output was the reflection phase corresponding to each $[p, b, d]$ within a frequency range from 10 GHz to 35 GHz. Among the 588 samples, 85 % (500) were used for training, and 15 % (88) testing. The surrogate model was developed using kriging interpolation [43]. The loss function was the MSE between the predicted and simulated reflection phases. The average loss for the test samples was 0.86 ° with a standard deviation of 1.7 °.

3.2.2 Re-implementation

To re-implement the work in [3], the proposed method is used to generate training samples, and generated training samples are utilized to develop a

Table D.5: Implementation B: Geometric Parameters of the Crusader Cross-based Unit Cell [3]

Parameter		Value	Parameter		Value
p	(mm)	[3.5, 10]	W	(mm)	6
b	(mm)	[0.3, 1.6]	L	(mm)	6
d	(mm)	[0.2, 2.4]	h	(mm)	1.5

Table D.6: Re-implementation B: Variables of the Proposed Method

Variable	Value	Variable	Value
$\boldsymbol{X}$	$[p, b, d]$	T	500
$\boldsymbol{Y}$	[Reflection phase]	*min_loss*	0.86 °
$\boldsymbol{X}_{step}$	[0.5, 0.06, 0.18]	N_0	175

surrogate model using kriging interpolation. In this re-implementation, the variables for the proposed method are set as shown in Table D.6. Here, the input vector $\boldsymbol{X}$ consists of normalized values of p, b, and d, and the output vector $\boldsymbol{Y}$ refers to reflection phases from 10 GHz to 35 GHz. The maximum iteration of data acquisition T is fixed at 500, which is the same as the number of training samples used in [3]. $\boldsymbol{X}_{step}$ is set as [0.5, 0.06, 0.18], which is inversely proportional to the grid density ($7 \times 12 \times 7$) of the geometric parameter space defined in [3]. *min_loss* is set as 0.86 °, which is also the reported test loss in [3]. $N_0 = 175$ samples are initialized. The 175 input vectors distribute on a uniform grid of $5 \times 7 \times 5$. Their corresponding reflection phases from 10 GHz to 35 GHz are collected as output labels through full-wave simulation using CST.

Starting from the 175 initialized samples, new training samples are generated iteratively using the proposed method. At each iteration t, there are $N_t = t + 1$ existing training samples, including 175 initialized samples and $N_t - 175$ new training samples. All the N_t existing samples are used as training data to develop a surrogate model using kriging interpolation at iteration t. After training, the surrogate model is used to predict reflection phases of 88 test samples. The 88 test samples distribute randomly on a uniform grid of $7 \times 12 \times 7$ of the geometric parameter space, which is similar to the 88 test samples used in [3]. The mean square error L for the 88 test samples is compared with *min_loss*. The data acquisition process completes when L is less than or equal to *min_loss*. For re-implementation B, the data acquisition process finishes at iteration $t = 296$, and 297 data samples are collected for training.

3.2.3 Comparison of Training Results

The results of implementation B with or without using the proposed method are compared in Table D.7. Without the proposed method, 588 data samples were acquired on a uniform grid of $7 \times 12 \times 7$ of the geometric parameter space.

Table D.7: Re-implementation B: Comparison of Results Without [3] and With the Proposed Data Acquisition Method

	Without the proposed method	**With the proposed method**
N	500	175 + 122
Time	3.89 h	2.31 h
L	0.86°	0.82°

Note: N is the number of training samples, which is 500 in [3];
L is the test loss, which is 0.86° in [3].

The 588 data samples were arbitrarily separated into a training data set of 500 data samples and a test data set of 88 data samples. The surrogate model was trained using the training data set through kriging interpolation. Afterward, the well-trained model was tested on the test data set, and the mean square error on the test data set was 0.86 °. With the proposed method, 175 + 122 = 297 training samples are acquired and utilized to develop the surrogate model using kriging interpolation. A test data set of 88 data samples is formed by arbitrarily choosing 88 points on the uniform grid of $7 \times 12 \times 7$ of the geometric parameter space. The well-trained surrogate model is tested on the test data set, and the MSE for test data is 0.82 °. Without the proposed method [3], 500 training samples were required to reach a test loss of 0.86 °. With the proposed method, only 297 training samples are required to reach a test loss of 0.82 °. By comparison, using the proposed method reduces the number of required training samples by 40.6 % ($\frac{500-297}{500}$), hence 40.6 % full-wave simulation cycles are saved. It spends 28 seconds for each simulation cycle supported by a computer equipped with 96 GB RAM and Intel® Xeon® Silver 4208 CPU @ 2.10 GHz (2 processors). Therefore, 1.58 hours full-wave simulation time is saved in re-implementation B by using the proposed method.

3.2.4 Analysis and Discussion

Fig. D.6 illustrates how the generated training samples distribute in the geometric parameter space to illustrate the underlying reason why the number of training data samples is reduced by 40.6 %. The distributions over p, b, and d are shown in Fig. D.6(a), (b), and (c), respectively. In each sub-figure, blue rectangular bars represent the number of training samples generated by using the proposed method, gray rectangular bars represent the number of training samples used in [3], and each orange dot represents the average reflection phase from 10 GHz to 35 GHz of one data sample.

The data distribution over p and d is shown in Fig. D.6(a) and (c). As p increases from 3.5 mm to 10 mm and d increases from 0.2 mm to 2.4 mm, the variance of average reflection phases decreases incrementally, hence adding samples with p and d being fixed at smaller values can improve the model

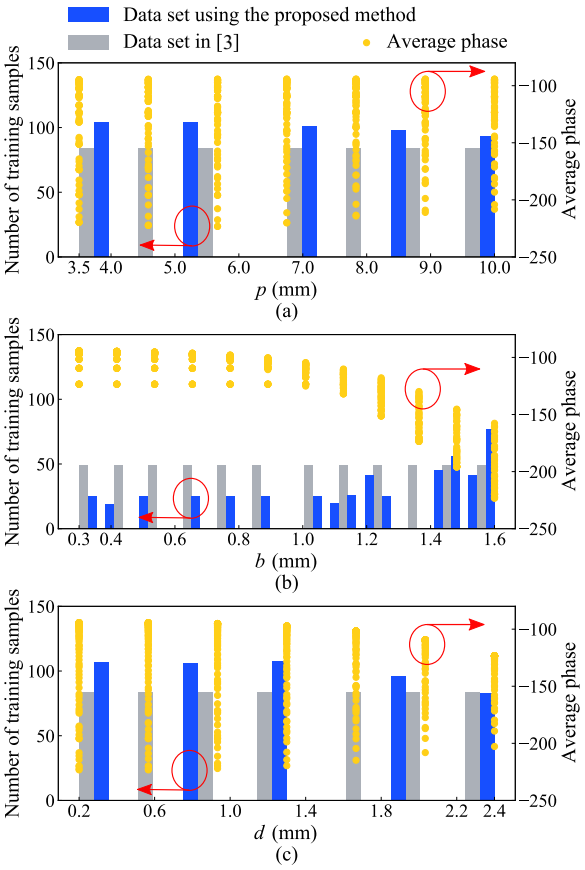


Fig. D.6: Implementation B: Comparison of distributions of the data set in [3] and the high-quality data set using the proposed method, while (a) p, (b) b, or (c) d varies. (Each orange dot indicates one data sample's average reflection phase from 10 GHz to 35 GHz.)

performance. Therefore, the proposed method tends to generate slightly more training samples as p and d decrease.

The data distribution over b is shown in Fig. D.6(b). The mean and variance of average reflection phases keep unchanged as b increases from 0.3 mm to 0.9 mm. As b increases from 0.9 mm to 1.6 mm, the mean decreases drastically from $-100\,^\circ$ to $-200\,^\circ$, and the variance increases from $\pm 30\,^\circ$ to $\pm 80\,^\circ$. Therefore, only a small number of training samples are generated uniformly by the proposed method as $b \in [0.3, 0.9]$, and a significantly increasing number of training samples are generated as b increases from 0.9 mm to 1.6 mm.

In a word, the proposed method tends to add more training samples where the average reflection phases have higher variance or unstable mean. The

reason is that the higher variance and unstable mean correspond to the more complex learning area. Adding training samples within this area can significantly improve the model performance. On the other hand, reducing training samples outside this area can save simulation cycles yet does not harm the model performance. It can be observed from Fig. D.6 that, the proposed method adjusts the number of training samples adaptively according to the variance and mean of average reflection phases, hence only 297 training samples are required to realize a test loss of 0.82 °. By contrast, the training samples were generated uniformly within the whole parameter space in [3], hence 500 training samples are required to realize a test loss of 0.86 °. The number of training samples is reduced by 40.6 %, and 40.6 % simulation cycles and 1.58 hours simulation time are saved by using the proposed method.

3.3 Implementation C: Array Radiation Synthesis

3.3.1 Implementation Description

Kim et al. in [4] utilized a deep neural network (DNN) to determine the phases of a 1×4 antenna array for various array radiation patterns. The operating frequency of the antenna array is at 2.4 GHz. The antenna array consists of four coaxial-fed patch antennas, as shown in Fig. D.7. The amplitudes for four elements were fixed at 1, and the phase for element 1 was set as 0 °. When the elements 2, 3, and 4 are fed with signals of different phases (α_2, α_3, and α_4), the combined array radiation pattern varies accordingly. In conventional scenarios, experienced engineers are required to decide the phases (α_2, α_3, and α_4) for a desired array radiation pattern. In [4], a DNN was trained to automatically determine the phases for desired radiation patterns without interference from human engineers.

The architecture of DNN used in [4] is given in Table D.8. There is an input layer, three hidden layers, and an output layer. The input layer and hidden layers have ReLU as activation functions, and the output layer uses the Linear activation function. Three hidden layers have neurons of 150, 100, and 80. Each input data for the input layer is of size 181 that represents a normalized radiation pattern with $\varphi = 0°$ and θ ranging from 0 ° to 180 ° in units of 1 °, which is referred to as $[R\{\varphi(0°),\ \theta[0°,\ 180°]\}]$. Each output data for the output layer is of size 6 that represents real and imaginary parts of complex excitation for elements 2, 3, and 4, which is referred to as $[cos\alpha_i, sin\alpha_i]$, $i = 2, 3, 4$. MSE is taken as the loss function for the DNN.

The authors in [4] generated 6859 samples for training. The amplitudes for four elements were fixed at 1, and the phase for element 1 was set as 0 °. The phases for element 2, 3, and 4 increased from 0 ° to 360 ° at a step of 20 °. The number of states for each element is 19 and the total amount of excitation combinations is 6859. The corresponding radiation patterns for all

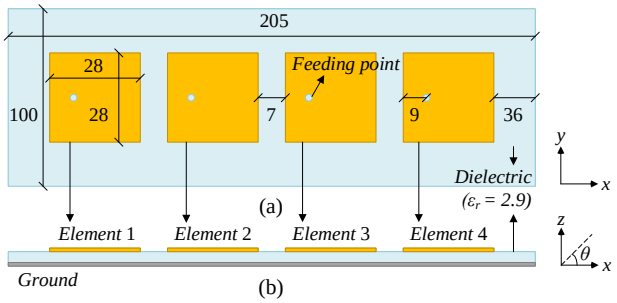


Fig. D.7: Implementation C: Array antenna structure: (a) Top view; (b) Side view. [4]

Table D.8: Implementation C: Architecture of the DNN [4]

No.	Layer	Neurons	Function
1	Input layer	181	Input: [radiation]
2	Activation function	-	ReLU
3	Hidden layer	150	Fully-connected layer
4	Activation function	-	ReLU
5	Hidden layer	100	Fully-connected layer
6	Activation function	-	ReLU
7	Hidden layer	80	Fully-connected layer
8	Activation function	-	ReLU
9	Output layer	6	Output: [excitation]
10	Activation function	-	Linear
-	**Loss function**	-	Mean Squared Error

6859 excitation combinations were acquired using CST. Similarly, 64 samples were generated as the validation data set. For validation, the phases for element 2, 3, and 4 increased from 10 ° to 130 ° at a step of 40 °.

The DNN was trained for 500 epochs using the 6859 training samples and the batch size was 100. After training, the model was validated on the 64 validation samples. The final training loss is 2.2×10^{-4} and the MSE for validation data is 2.6×10^{-4}.

3.3.2 Re-implementation

To re-implement the work in [4], the variables for the proposed method are determined as given in Table D.9. $\boldsymbol{X}$ is set as phases ([α_2, α_3, α_4]) of elements 2, 3, 4 during data acquisition and is converted to the real and imaginary format ([$cos\alpha_i$, $sin\alpha_i$], $i = 2, 3, 4$) for model training. $\boldsymbol{Y}$ is set as normalized radiation patterns with $\varphi = 0°$ and θ ranging from 0° to 180° in units of 1°. The minimum step of phases $\boldsymbol{X}_{step}$ is fixed at [10 °, 10 °, 10 °]. The minimum

Table D.9: Implementation C: Variables of the Proposed Method

Variable	Value	Variable	Value
$\boldsymbol{X}$	$[\alpha_2, \alpha_3, \alpha_4]$	T	6859
	$[cos\alpha_i, sin\alpha_i], i: 2, 3, 4$		
$\boldsymbol{Y}$	$[R\{\varphi(0°),\ \theta[0°,\ 180°]\}]$	min_loss	2.6×10^{-4}
$\boldsymbol{X}_{step}$	$[10\,°, 10\,°, 10\,°]$	N_0	2197

Table D.10: Re-implementation C: Comparison of Results Without [4] and With the Proposed Data Acquisition Method

	Without the proposed method	With the proposed method
$\boldsymbol{N}$	6859	2197 + 1923
Time	17.14 h	10.30 h
$\boldsymbol{L}$	2.6×10^{-4}	2.6×10^{-4}

Note: N is the number of training samples, which is 6859 in [4];
L is the test loss, which is 2.6×10^{-4} in [4].

loss *min_loss* is set as 2.6×10^{-4}, which equals the reported validation loss in [4]. The maximum data acquisition iteration is set as the number of training samples used in [4]. The amount of initialized samples N_0 is set as 2197 after comparing results corresponding to different N_0s.

3.3.3 Comparison of Training Results

The results of re-implementation using the proposed method are compared with results claimed in [4] in Table D.10. The authors in [4] collected 6859 training data samples by incrementally increasing elements' phases (α_2, α_3, and α_4) from 0 ° to 360 ° at a constant step of 20 °. Similarly, 64 test data samples were collected by incrementally increasing elements' phases (α_2, α_3, and α_4) from 0 ° to 360 ° at a constant step of 20 °. The DNN was trained using the 6859 training samples and was tested on the 64 test samples. The mean square error over the 64 test samples was 2.6×10^{-4}. For comparison, the proposed method is integrated with model training to re-implement the work in [4]. At iteration 4119, 4120 training samples are generated, and the DNN is trained using the 4120 training samples and is tested on the same 64 test samples. The mean square error equals the test loss reported in [4]. By comparison, the number of required training samples is reduced by 39.93 % by using the proposed method, hence 39.93 % simulation cycles are saved in implementation C. On average, it takes around 9 seconds for each simulation cycle supported by a computer equipped with 96 GB RAM and Intel® Xeon® Silver 4208 CPU @ 2.10 GHz (2 processors). Therefore, the proposed method saves 6.84 hours full-wave simulation time in implementation C.

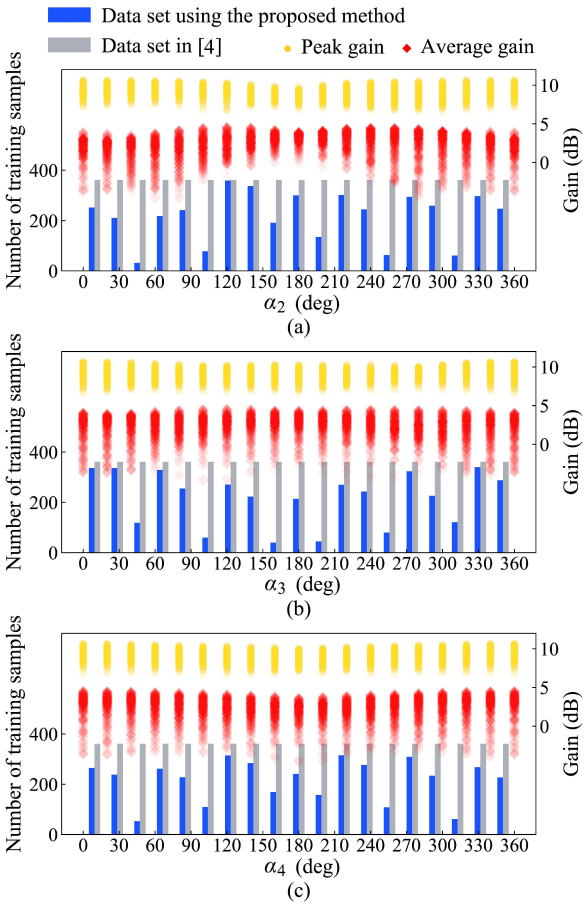


Fig. D.8: Implementation C: Comparison of distributions of the data set in [4] and the high-quality data set using the proposed method, while (a) α_2, (b) α_3, or (c) α_4 varies. (Each orange dot indicates one data sample's peak gain; each red diamond indicates one data sample's average gain.)

3.3.4 Analysis and Discussion

To get an insight into the reason why 39.93 % training data are saved, the distribution of the high-quality training data set acquired using the proposed method is compared with that of the training data set used in [4]. Fig. D.8(a), (b), and (c) exhibits the distribution over α_2, α_3, and α_4, respectively. Here, the high-quality data set is represented as blue rectangular bars, while the training data samples used in [4] are represented as gray rectangular bars. As it is difficult to plot the whole radiation pattern of each data sample, only the peak gain and average gain of the radiation pattern of each data sample

are plotted. Each orange dot represents the peak gain of one data sample, while each red diamond-shaped symbol represents the average gain of one data sample.

As can be observed in Fig. D.8, the peak gain and average gain fluctuate within a certain range as phases of elements (α_2, α_3, and α_4) change. The peak gain and average gain are just two compressed features of the radiation pattern. The changing tendency of radiation patterns is way more complex than that it can be plotted and observed from Fig. D.8. Heavier fluctuations can be expected for the changing tendency of the whole radiation patterns. Therefore, it is difficult to specifically clarify the distribution of the generated data set using the proposed method. An overall observation is that the generated data set using the proposed method distribute adaptively and the number of generated training data samples varies concerning different phases of elements. By contrast, the data set used in [4] distribute uniformly on a constant grid. The benefit and effectiveness of this adaptive sampling strategy are validated through numerical and comparative results. The number of training data samples required for the same model accuracy is significantly reduced by 39.93 % by using the proposed method.

ML-based radiation synthesis can be applied to a larger array with more than four elements, as long as powerful computation resources are available for simulating such large arrays, manipulating massive data, and training complicated models. The proposed method can also be used for this large array case, and the reduced amount will be proportional to the number of required training data acquired conventionally.

3.4 Implementation D: Enlarged Array Radiation Synthesis

The proposed method is validated in a three-dimensional parameter space in implementation C. To validate its performance in a higher-dimensional parameter space, we enlarge the four-element linear array in implementation C into an eight-element linear array, which is referred to as implementation D. Except for the number of elements, the array structure is the same. For the enlarged array, we use a new forward DNN for training, as shown in Table D.11. Here, phases of its excitation are set as input, and radiation patterns are set as output.

The DNN is first trained without using the proposed method. 78125 samples are collected for training by sweeping the phases of 7 elements. The amplitudes of eight elements are set as 1, and the phase of element 1 is set as $0\,^\circ$. The phases of the rest 7 elements 2-8 vary from $0\,^\circ$ to $180\,^\circ$ at a constant step of $45\,^\circ$. The corresponding 78125 radiation patterns are generated using CST. 128 samples are collected similarly as the validation data set, where the phases of 7 elements vary between $30\,^\circ$ and $150\,^\circ$. The MSE for validation data is around 0.73.

Table D.11: Implementation D: Architecture of the DNN

No.	Layer	Neurons	Function
1	Input layer	14	Input: [excitation]
2	Activation function	-	ReLU
3	Hidden layer	300	Fully-connected layer
4	Activation function	-	ReLU
5	Hidden layer	200	Fully-connected layer
6	Activation function	-	ReLU
7	Hidden layer	100	Fully-connected layer
8	Activation function	-	ReLU
9	Output layer	181	Output: [radiation]
10	Activation function	-	Linear
-	**Loss function**	-	Mean Squared Error

The proposed method is then used to re-implement implementation D for comparison. All the settings are similar to implementation C, and the variables are listed in Table D.12. 16384 samples are initialized by sweeping the phases of 7 elements from 0 ° to 180 ° at a constant step of 60 °. The generation of high-quality data stops in iteration $t = 46499$, and 46500 samples are generated in total. The test loss of the DNN for the 128 validation samples reaches $min\_loss = 0.73$ after being trained using the 46500 data acquired using the proposed method.

The results without and with using the proposed method are compared and listed in Table D.13. As can be observed, the number of required training data is reduced by 40.48 % with using the proposed method, hence 40.48 % simulation cycles are saved. The time needed for each simulation cycle is around 13 seconds. Thus, 114.2 hours simulation time is saved in implementation D with using the proposed method.

It proves that the proposed method is still effective as the dimension of parameter space increases from three to seven. Due to the limitations of computation resources, we cannot keep increasing the dimension and validating the proposed method in higher-dimensional spaces. If more powerful computation resources are available, it is meaningful to further investigate the effectiveness in higher-dimensional spaces, because it gets more complicated and may involve new challenges.

4 Discussion

The proposed solution addresses a common challenge in the ML-based design of electromagnetic structures. The issue is that ML often requires a vast amount of simulations to gather data, which is both time-consuming and com-

Table D.12: Implementation D: Variables of the Proposed Method

Variable	Value	Variable	Value
$\boldsymbol{X}$	$[\alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8]$	T	78125
	$[cos\alpha_i, sin\alpha_i], i: 2,3,4,5,6,7,8$		
$\boldsymbol{Y}$	$[R\{\varphi(0°),\ \theta[0°,\ 180°]\}]$	min_loss	0.73
$\boldsymbol{X}_{step}$	[10 °,10 °,10 °,10 °,10 °,10 °,10 °]	N_0	16384

Table D.13: Re-implementation D: Comparison of Results Without and With the Proposed Data Acquisition Method

	Without the proposed method	With the proposed method
$\boldsymbol{N}$	78125	16384 + 30116
Time	282.12 h	167.92 h
$\boldsymbol{L}$	0.73	0.73

Note: N is the number of training samples;
L is the test loss.

putationally expensive. Our approach reduces the number of simulations by evaluating the quality of data before simulation and prioritizing computation resources for high-quality data.

Designers can use the method to understand the sensitivity of the geometric parameters of electromagnetic structures. By analyzing the distribution of acquired high-quality data, as seen in Fig. 4, 6, and 8, designers can identify the most critical parameters and their respective ranges, helping them understand the motivation behind the geometry and how to adjust it to modify the electromagnetic response.

The effectiveness of the proposed method has been validated in low-dimensional implementations, and it remains effective as the dimension increases from three to seven. It is difficult to keep increasing the dimension and validate it in higher-dimensional spaces, due to the limitations of computation resources. Further investigation on higher-dimensional implementations is meaningful, as it gets more complicated and may involve new challenges. In future work, we will attempt to apply the proposed method in higher-dimensional implementations.

5 Conclusion

An intelligent high-quality data acquisition method for ML-related EM applications is proposed in this paper. Starting from a small uniformly initialized data set, the proposed method can intelligently generate high-quality data samples based on the analysis of existing data samples. Compared with con-

ventional EM-ML works that acquired training data by blindly sweeping the whole geometric parameter space on a constant and uniform grid, the proposed method adaptively adjusts the sampling density in different geometric parameter areas. The proposed method produces a data set that maximizes informativeness with the least number of simulation cycles. To validate its performance, the proposed method is utilized to re-implement four implementations. The comparative results without and with the proposed method show that the proposed method significantly reduces the number of training data required for the same model accuracy. Around 40 % training data are saved by using the proposed method, hence a huge number of full-wave simulation cycles and time are saved, and computing resources are significantly released.

References

[1] Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: A big data-AI integration perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1328–1347, 2019.

[2] R. Zhu, J. Wang, Y. Han, S. Sui, T. Qiu, Y. Jia, M. Feng, X. Wang, L. Zheng, and S. Qu, "Design of aperture-multiplexing metasurfaces via back-propagation neural network: Independent control of orthogonally-polarized waves," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 6, pp. 4569–4575, 2022.

[3] M. Abdullah and S. Koziel, "Supervised-learning-based development of multibit RCS-reduced coding metasurfaces," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 1, pp. 264–274, 2021.

[4] J. H. Kim and S. W. Choi, "A deep learning-based approach for radiation pattern synthesis of an array antenna," *IEEE Access*, vol. 8, pp. 226 059–226 063, 2020.

[5] D. Erricolo, P.-Y. Chen, A. Rozhkova, E. Torabi, H. Bagci, A. Shamim, and X. Zhang, "Machine learning in electromagnetics: A review and some perspectives for future research," in *International Conference on Electromagnetics in Advanced Applications*, 2019, pp. 1377–1380.

[6] Y. Kiarashinejad, S. Abdollahramezani, and A. Adibi, "Deep learning approach based on dimensionality reduction for designing electromagnetic nanostructures," *npj Computational Materials*, vol. 6, no. 1, pp. 1–12, 2020.

[7] N. A. Alderete, N. Pathak, and H. D. Espinosa, "Machine learning assisted design of shape-programmable 3D kirigami metamaterials," *npj Computational Materials*, vol. 8, no. 1, pp. 1–12, 2022.

[8] M. Sedaghat, R. Trinchero, Z. H. Firouzeh, and F. G. Canavero, "Compressed machine learning-based inverse model for design optimization of microwave components," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 7, pp. 3415–3427, 2022.

[9] Z. Zhou, Z. Wei, Y. Zhang, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Training of deep neural networks in electromagnetic problems: a case study of antenna array pattern synthesis," in *IEEE MTT-S International Wireless Symposium*, Nanjing, China, 2021, pp. 1-3.

[10] N. Kazemi, M. Abdolrazzaghi, and P. Musilek, "Comparative analysis of machine learning techniques for temperature compensation in microwave sensors," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 9, pp. 4223–4236, 2021.

[11] Z. Wei, Z. Zhou, Y. Zhang, P. Li, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Domain knowledge assisted training dataset generation for metasurface designs," in *IEEE MTT-S International Wireless Symposium*, Nanjing, China, 2021, pp. 1-3.

[12] C. Lewis, J. Bryan, N. Schwartz, J. Hale, K. Fanning, and J. S. Colton, "Machine learning to predict quasi TE_{001} mode resonances in double-stacked dielectric cavities," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 4, pp. 2135–2146, 2022.

[13] Z. Zhou, Z. Wei, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Representation learning-driven fully automated framework for the inverse design of frequency-selective surfaces," *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 6, pp.2409–2421, 2023.

[14] Z. Wei, Z. Zhou, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Fully automated design method based on reinforcement learning and surrogate modeling for antenna array decoupling," *IEEE Transactions on Antennas and Propagation*, vol. 71, no. 1, pp. 660-671, 2023.

[15] Z. Zhou, Z. Wei, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Two-order deep learning for generalized synthesis of radiation patterns for antenna arrays," *IEEE Transactions on Artificial Intelligence*, pp. 1–9, 2022.

[16] D. R. Prado, J. A. López-Fernández, M. Arrebola, and G. Goussetis, "Support vector regression to accelerate design and crosspolar optimization of shaped-beam reflectarray antennas for space applications," *IEEE Transactions on Antennas and Propagation*, vol. 67, no. 3, pp. 1659–1668, 2018.

[17] J. A. Fan, "Generating high performance, topologically-complex metasurfaces with neural networks," in *Conference on Lasers and Electro-Optics*, 2019, pp. 1–2.

[18] F. Güneş, S. Nesil, and S. Demirel, "Design and analysis of minkowski reflectarray antenna using 3-D CST microwave studio-based neural network model with particle swarm optimization," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 23, no. 2, pp. 272–284, 2013.

[19] J. A. Hodge, K. V. Mishra, and A. I. Zaghloul, "RF metasurface array design using deep convolutional generative adversarial networks," in *IEEE International Symposium on Phased Array Systems & Technology*, 2019, pp. 1–6.

[20] D. R. Prado, J. A. López-Fernández, M. Arrebola, M. R. Pino, and G. Goussetis, "Wideband shaped-beam reflectarray design using support vector regression analysis," *IEEE Antennas and Wireless Propagation Letters*, vol. 18, no. 11, pp. 2287–2291, 2019.

[21] S. Koziel and M. Abdullah, "Machine-learning-powered EM-based framework for efficient and reliable design of low scattering metasurfaces," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 4, pp. 2028–2041, 2021.

[22] P. Naseri, G. Goussetis, N. J. Fonseca, and S. V. Hum, "Inverse design of a dual-band reflective polarizing surface using generative machine learning," in *European Conference on Antennas and Propagation*, 2022, pp. 1–5.

[23] R. Zhu, T. Qiu, J. Wang, S. Sui, C. Hao, T. Liu, Y. Li, M. Feng, A. Zhang, C.-W. Qiu *et al.*, "Phase-to-pattern inverse design paradigm for fast realization of functional metasurfaces via transfer learning," *Nature Communications*, vol. 12, no. 1, pp. 1–10, 2021.

[24] P. Naseri and S. V. Hum, "A Generative Machine Learning-Based Approach for Inverse Design of Multilayer Metasurfaces," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 9, pp. 5725–5739, 2021.

[25] Z. Zhou, Z. Wei, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Transfer learning assisted multi-element calibration for active phased antenna arrays," *IEEE Transactions on Antennas and Propagation*, vol. 71, no. 2, pp. 1982–1987, 2022.

[26] L.-Y. Xiao, F.-L. Jin, B.-Z. Wang, Q. H. Liu, and W. Shao, "Efficient inverse extreme learning machine for parametric design of metasurfaces," *IEEE Antennas and Wireless Propagation Letters*, vol. 19, no. 6, pp. 992–996, 2020.

[27] Z. Wei, Z. Zhou, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Equivalent circuit theory-assisted deep learning for accelerated generative design of metasurfaces," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 7, pp. 5120–5129, 2022.

[28] J. Bouvette, H.-F. Liu, X. Du, Y. Zhou, A. P. Sikkema, J. da Fonseca Rezende e Mello, B. P. Klemm, R. Huang, R. M. Schaaper, M. J. Borgnia *et al.*, "Beam image-shift accelerated data acquisition for near-atomic resolution single-particle cryo-electron tomography," *Nature Communications*, vol. 12, no. 1, pp. 1–11, 2021.

[29] K. Jeong, M. Babović, V. Gorshkov, J. Kim, O. N. Jensen, and O. Kohlbacher, "FLASHIda enables intelligent data acquisition for top–down proteomics to boost proteoform identification counts," *Nature Communications*, vol. 13, no. 1, pp. 1–12, 2022.

[30] H. Garcia-Molina, M. Joglekar, A. Marcus, A. Parameswaran, and V. Verroios, "Challenges in data crowdsourcing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 4, pp. 901–911, 2016.

[31] X. Zhu and X. Wu, "Cost-constrained data acquisition for intelligent data preparation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 11, pp. 1542–1556, 2005.

[32] Y.-C. Lin, D.-N. Yang, H.-H. Shuai, and M.-S. Chen, "Data acquisition for probabilistic nearest-neighbor query," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 2, pp. 410–427, 2014.

[33] A. Pietrenko-Dabrowska, S. Koziel, and U. Ullah, "Reduced-cost two-level surrogate antenna modeling using domain confinement and response features," *Scientific Reports*, vol. 12, no. 1, Art. no. 1, 2022.

[34] S. Koziel and A. Pietrenko-Dabrowska, "Expedited variable-resolution surrogate modeling of miniaturized microwave passives in confined domains," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 11, pp. 4740–4750, 2022.

[35] A. Pietrenko-Dabrowska, S. Koziel, and L. Golunski, "Two-stage variable-fidelity modeling of antennas with domain confinement," *Scientific Reports*, vol. 12, no. 1, Art. no. 1, 2022.

[36] S. Koziel, N. Çalık, P. Mahouti, and M. A. Belen, "Reliable computationally efficient behavioral modeling of microwave passives using deep learning surrogates in confined domains," *IEEE Transactions on Microwave Theory and Techniques*, pp. 1–13, 2022.

[37] J. A. Tomasson, A. Pietrenko-Dabrowska, and S. Koziel, "Expedited globalized antenna optimization by principal components and variable-fidelity EM simulations: application to microstrip antenna design," *Electronics*, vol. 9, no. 4, Art. no. 4, 2020.

[38] S. Koziel and A. Pietrenko-Dabrowska, "Expedited feature-based quasi-global optimization of multi-band antenna input characteristics with Jacobian variability tracking," *IEEE Access*, vol. 8, pp. 83907–83915, 2020.

[39] A. Pietrenko-Dabrowska and S. Koziel, "Globalized parametric optimization of microwave components by means of response features and inverse metamodels," *Scientific Reports*, vol. 11, no. 1, Art. no. 1, 2021.

[40] S. Koziel, A. Pietrenko-Dabrowska, and M. Mahrokh, "Globalized simulation-driven miniaturization of microwave circuits by means of dimensionality-reduced constrained surrogates," *Scientific Reports*, vol. 12, no. 1, Art. no. 1, 2022.

[41] S. Koziel and A. Pietrenko-Dabrowska, "Low-cost quasi-global optimization of expensive electromagnetic simulation models by inverse surrogates and response features," *Scientific Reports*, vol. 12, no. 1, Art. no. 1, 2022.

[42] B. L. Kalman and S. C. Kwasny, "Why tanh: choosing a sigmoidal function," in *Proceedings of the International Joint Conference on Neural Networks*, vol. 4, 1992, pp. 578–581.

[43] T. W. Simpson, J. Poplinski, P. N. Koch, and J. K. Allen, "Metamodels for computer-based engineering design: survey and recommendations," *Eng. Comput.*, vol. 17, no. 2, pp. 129–150, 2001.

Paper E

Bayesian-Inspired Sampling for Efficient Machine-Learning-Assisted Microwave Component Design

Zhao Zhou, Zhaohui Wei, Jian Ren, Yingzeng Yin, Gert Frølund Pedersen, and Ming Shen

The paper has been published in the
IEEE Transactions on Microwave Theory and Techniques, Early Access, 2023.

The layout has been revised.

Abstract

Machine learning (ML) has demonstrated significant potential in accelerating the design of microwave components owing to its great ability to approximate the projection between geometric parameters and electromagnetic (EM) responses. A well-trained ML model can predict the EM responses of a microwave component with unseen geometric parameter settings accurately, or determine the parameter settings based on desired EM constraints in a matter of milliseconds. However, this ML-based design process often requires heavy simulation to collect a large amount of training data. To mitigate this issue, this paper proposes an efficient Bayesian inspired sampling assisted ML method for the design of microwave components. In contrast to typical ML-based design methods which use uniform and arbitrary sampling to extensively represent the entire parameter space, necessitating intensive simulation for generating training data, the proposed Bayesian inspired sampling strategy efficiently represents the entire parameter space by recognizing and emphasizing more promising parameter settings. This is achieved by defining a Bayesian-based expression for evaluating the probability of the outcome of adding a new data sample in a specific parameter area. During each iteration of the design process, new data is always added in the area with the highest probability of beneficial outcomes. Therefore, it optimizes the distribution of training data and reduces the amount of required training data and simulations. Results from three design case studies demonstrate that the proposed method can significantly reduce the number of required data and simulation by around 40 % *for the same model performance. This validates that the proposed Bayesian inspired sampling-aided ML method significantly improves overall efficiency.*

1 Introduction

Microwave components play an essential role in wireless communication systems. The conventional design of microwave components is knowledge-dependent, labor-intensive, and computational-expensive, hence of low efficiency. The upcoming beyond-fifth-generation (B5G) and sixth-generation (6G) wireless communication eras call for the speed-up and automation of microwave components design. Machine learning (ML) shows great potential for accelerating the design of microwave components, as shown in Fig. E.1.

There have been an increasing number of works [1–12] on ML-based electromagnetic solutions, such as on electromagnetic wave absorption of human head [8, 9], among which many works [1–7] use ML for automatizing and speeding up the design process of microwave components. Three major categories can be recognized: inverse, forward, and generative ML methods. The main difference is that inverse ML models aim to project electromagnetic (EM) responses to geometric parameters [13, 14], and forward ones doing the opposite [15–17]; generative ML models capture the distribution characteristics of

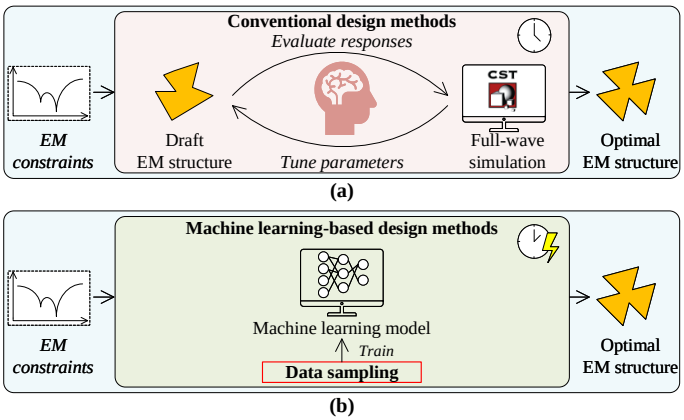


Fig. E.1: (a) Conventional design method; (b) ML-based design method.

geometric parameters and generate unseen parameter settings [18–21]. Generative ML models are often integrated with inverse or forward models for evaluating the generated new parameter settings.

The common nature of most ML-based design methods is to approximate the projection between geometric parameter settings and EM responses. The approximation is carried on by optimizing coefficients of numerical models based on the analysis of historical simulation data, so-called training data. Therefore, the performance of ML mainly depends on the simulation data for training. Typical ML methods often require a number of full-wave simulation runs for collecting sufficient training data, which is a significant obstacle for ML-based applications. Investigations on ML-based design methods that require less simulation time and data are quite meaningful.

Conventional ML methods usually adopt the uniform or arbitrary sampling strategy, as shown in Fig. E.2(a). Within the concerning parameter space, the parameter settings are defined on a uniform grid of the whole space to extensively cover all the possibilities. They usually take an arbitrary subset of all the uniformly distributed parameter settings if the size of the parameter space is too large to include all the uniform parameter settings. The defined parameter settings are then swept and simulated via full-wave EM simulation tools (e.g., Computer Simulation Technology (CST)) for collecting the corresponding EM responses. The combination of each parameter setting and its corresponding EM responses forms one training data sample. A training dataset including a number of data samples is taken for training an ML model. One tricky point is that the distribution of EM responses is non-linear and complicated. Therefore, different training data samples contribute unequally to improving the model accuracy. For example, as shown in Fig. E.2(a), the EM response could be more sensitive to larger parameter values, while it remains nearly consistent

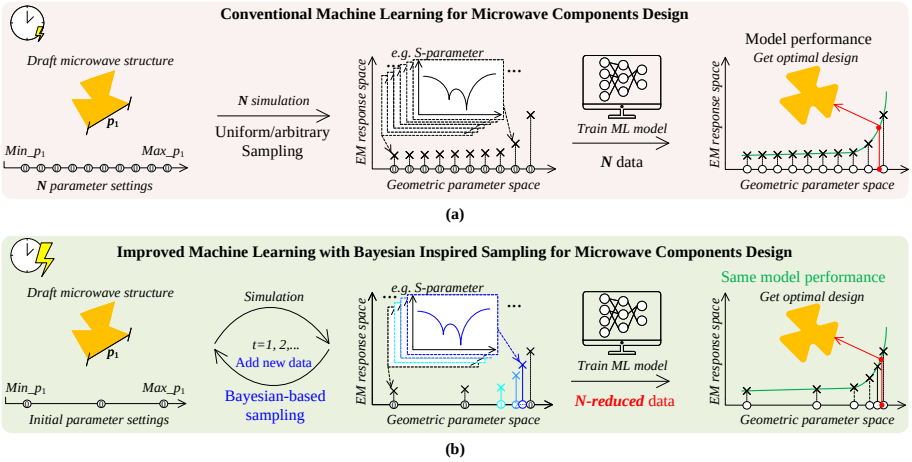


Fig. E.2: (a) Conventional ML method; (b) Proposed Bayesian inspired sampling assisted ML method.

for small parameter values. In this case, the training data samples related to small parameter values contribute little to improving the model performance, resulting in low overall efficiency. The inequality could be more complicated in real scenarios, especially as the dimension of the parameter space increases, which is difficult to evaluate and avoid before simulation.

There have been many works putting efforts into the optimization of data sampling. Koziel *et al.* [22–26] optimized the data sampling for global optimization by identifying the promising regions towards the optimal design. Unlike global optimization, the accuracy of ML models relies on not only these promising regions towards the optimal design but also other representative regions within the whole parameter space. [17, 27] utilized transfer learning to optimize data sampling for ML models by utilizing theoretical data to pre-train a model. Theoretical data can be easily acquired by executing well-established numerical equations assisted by EM domain knowledge, and it offers basic intelligence for the pre-trained model. Thus, it greatly reduced the simulation needed for collecting training data. However, it may be difficult to find well-established numerical equations for some design cases. There have also been great attempts that use transfer learning to reduce the required simulation data by pre-training the ML model with theoretical data. The drawback is that the required theoretical data may not be functional or available for some design scenarios. We propose a Bayesian inspired sampling method that is more robust and can be widely applied to many design scenarios, and it is of higher interpretability and more concise compared to the original method in [28].

This paper proposes an efficient ML method assisted by Bayesian inspired

sampling strategy for the design of microwave components, as shown in Fig. E.2(b). The proposed method addresses the challenge of data collection in ML-based design by reducing the required simulation cycles while maintaining model accuracy. It achieves this by estimating the quality of data before simulation and allocating computation resources to high-quality data samples. Unlike conventional uniform sampling or Latin hypercube sampling method that mainly depends on the parameter range, the proposed method jointly analyzes both parameter settings and EM responses of existing data for optimizing future data distribution. The proposed Bayesian sampling approach effectively measures the distributions of existing data, estimates the quality of new data, and leverages computation resources to simulate high-quality data. The proposed sampling strategy starts with a small number of parameter settings uniformly initialized within the whole parameter space. They are combined with their corresponding EM responses obtained through full-wave simulation to form an initial dataset. Afterward, we applied the Bayesian rule for evaluating the probability of the outcome of adding a new parameter setting in different parameter areas. A new parameter setting of the highest probability is then defined and forms new data together with its EM response extracted via simulation. After multiple iterations, an optimized dataset is generated and utilized for training a neural network (NN). The well-trained NN can then replace the full-wave simulation tool for predicting the EM responses of unseen parameter settings. The distribution of generated dataset shows that the sampling is dynamically and effectively guided by the proposed method, resulting in a high-quality dataset eventually. Experimental results validate the effectiveness of the proposed method in significantly reducing the required simulation data and improving the efficiency compared to the conventional uniform sampling and Latin hypercube sampling. Compared with existing ML methods, the proposed method has superior overall efficiency.

The following part is arranged as follows: Section II introduces the proposed Bayesian inspired ML method; Section III validates the proposed method in three implementations; Section IV gives the conclusion.

2 Proposed Method

The working principle of the proposed method is demonstrated and compared with conventional ML methods in Fig. E.2. The existing literature usually defines all the parameter settings first based on uniform sampling or Latin hypercube sampling and then simulates them individually. The conventional uniform sampling sweeps parameters on a fine and uniform grid within the whole parameter range. The Latin hypercube sampling adjusts the distribution of parameter settings only based on the parameter range. These traditional sampling methods cannot ensure the quality of data, because their sampling

strategies only depend on the parameter range and ignore the distribution of EM responses. The tricky thing is that the distribution of EM responses is non-uniform, unbalanced, and not aligned with the distribution of the parameter settings. Therefore, the collected data are of unbalanced quality. Some data contribute little to the model performance, which are named low-quality data; some data contribute more, which are called high-quality data.

Unlike conventional methods, we estimate the quality of data and adjust the data distribution by analyzing the distributions of both parameter settings and EM responses based on the proposed Bayesian sampling. The proposed Bayesian-inspired sampling iteratively measures the quality of existing data, predicts new high-quality data, and leverages computation resources to these high-quality data. By doing this, the distributions of generated data are optimized dynamically and effectively, resulting in a high-quality dataset. The high-quality dataset is of reduced size and takes reduced simulation for realizing the same model accuracy. Therefore, the proposed method has a significantly improved efficiency.

The proposed method takes ten major steps 1-10.

1. Determine the parameter space by estimating the minimum and maximum values of concerning geometric parameters.

2. Define a small size of parameter settings on a sparse, uniform grid of the whole parameter space, and obtain the corresponding EM responses via simulation. The parameter settings and EM responses are normalized to form an initial dataset. The initial sampling grid should be sparser than the typical sampling grid and include the minimum/maximum values.

3. For each data sample $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$ within the existing dataset, pick two reference data samples $(\boldsymbol{X}_i^{'}, \boldsymbol{Y}_i^{'})$ and $(\boldsymbol{X}_i^{''}, \boldsymbol{Y}_i^{''})$ from the rest of data samples excluding itself. Selection of $(\boldsymbol{X}_i^{'}, \boldsymbol{Y}_i^{'})$ obeys the following rules: all the elements of $\boldsymbol{X}_i^{'}$ should be incremental or equal compared to their counterparts of $\boldsymbol{X}_i$,

$$\Delta \boldsymbol{X} = \boldsymbol{X}_i^{'} - \boldsymbol{X}_i, \tag{E.1}$$

$$\forall \Delta x_i \in \Delta \boldsymbol{X}, \;\; \Delta x_i \geq 0; \tag{E.2}$$

the sum of increments of all the elements should be minimum,

$$\boldsymbol{X}_i^{'} = \underset{\boldsymbol{X}, \boldsymbol{X}_i^{'}}{argmin} \sum \Delta x_i, \;\; \Delta x_i \in \Delta \boldsymbol{X}; \tag{E.3}$$

at least one of the elements has an increment larger than a pre-defined threshold,

$$\exists \Delta x_i \in \Delta \boldsymbol{X}, \;\; \Delta x_i \geq \epsilon_i. \tag{E.4}$$

The thresholds for all the elements are denoted as a vector $\boldsymbol{X}_u$. $\boldsymbol{X}_u$ limits the minimum distance between any two data samples,

$$\boldsymbol{X}_u = [\epsilon_1,\ \epsilon_2,\ ...,\ \epsilon_i,\ ...]. \tag{E.5}$$

It is defined according to the tolerance of accuracy of specific components. Selection of $(\boldsymbol{X}_i^{''}, \boldsymbol{Y}_i^{''})$ makes sure that $(\boldsymbol{X}_i^{''}, \boldsymbol{Y}_i^{''})$ and $(\boldsymbol{X}_i^{'}, \boldsymbol{Y}_i^{'})$ follows the same rules. $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$ is skipped if there are no available reference data that fulfill the aforementioned rules. Each available combination of $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$, $(\boldsymbol{X}_i^{'}, \boldsymbol{Y}_i^{'})$, and $(\boldsymbol{X}_i^{''}, \boldsymbol{Y}_i^{''})$ defines a parameter area. A list of definitions for parameter areas should be generated after step 3.

4. The sampling process stops if the list inherited from the step 3 is empty. Otherwise, each parameter area is evaluated based on the Bayesian theorem, as given in formula (E.6),

$$P(\boldsymbol{A}|\boldsymbol{B}) = \frac{P(\boldsymbol{A}, \boldsymbol{B})}{P(\boldsymbol{B})}. \tag{E.6}$$

The evaluation measures the probability of the outcome of adding a new data sample within each parameter area, which is denoted as $P\{\boldsymbol{X}_i^{'}|[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\}$,

$$P\{\boldsymbol{X}_i^{'}|[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\} = P(\boldsymbol{A}|\boldsymbol{B}). \tag{E.7}$$

Here, $\boldsymbol{A}$ is referred to as $\boldsymbol{X}_i^{'}$, and $\boldsymbol{B}$ is referred to as $[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]$,

$$\boldsymbol{A} = \boldsymbol{X}_i^{'} \tag{E.8}$$

$$\boldsymbol{B} = [\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]. \tag{E.9}$$

Formula (E.6) is updated as formula (E.10),

$$P\{\boldsymbol{X}_i^{'}|[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\} = \frac{P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{'}, \boldsymbol{X}_i^{''}]\}}{P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\}}. \tag{E.10}$$

Here, $P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{'}, \boldsymbol{X}_i^{''}]\}$ is defined as,

$$P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{'}, \boldsymbol{X}_i^{''}]\} = P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{'}]\} + P\{[\boldsymbol{X}_i^{'}, \boldsymbol{X}_i^{''}]\}, \tag{E.11}$$

$$P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{'}]\} = |\boldsymbol{Y}_i - \boldsymbol{Y}_i^{'}|^2, \tag{E.12}$$

$$P\{[\boldsymbol{X}_i^{'}, \boldsymbol{X}_i^{''}]\} = |\boldsymbol{Y}_i^{'} - \boldsymbol{Y}_i^{''}|^2; \tag{E.13}$$

$P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\}$ is defined as,

$$P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\} = |\boldsymbol{Y}_i - \boldsymbol{Y}_i^{''}|^2. \tag{E.14}$$

$P\{\boldsymbol{X}_i^{'}|[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\}$ can be expressed as formula (E.15),

$$\begin{aligned} P\{\boldsymbol{X}_i^{'}|[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\} &= \frac{P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{'}, \boldsymbol{X}_i^{''}]\}}{P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\}} \\ &= \frac{P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{'}]\} + P\{[\boldsymbol{X}_i^{'}, \boldsymbol{X}_i^{''}]\}}{P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\}} \\ &= \frac{|\boldsymbol{Y}_i - \boldsymbol{Y}_i^{'}|^2 + |\boldsymbol{Y}_i^{'} - \boldsymbol{Y}_i^{''}|^2}{|\boldsymbol{Y}_i - \boldsymbol{Y}_i^{''}|^2}. \end{aligned} \tag{E.15}$$

Here, $\{\boldsymbol{X}_i, \boldsymbol{X}_i^{'}, \boldsymbol{X}_i^{''}\}$ represents geometrical parameters, and $\{\boldsymbol{Y}_i, \boldsymbol{Y}_i^{'}, \boldsymbol{Y}_i^{''}\}$ represents their corresponding EM responses. They form three data samples $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$, $(\boldsymbol{X}_i^{'}, \boldsymbol{Y}_i^{'})$, and $(\boldsymbol{X}_i^{''}, \boldsymbol{Y}_i^{''})$, which form the i-th available combination of data samples that fits the rules described in step 3. $P\{\boldsymbol{X}_i^{'}|[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\}$ represents the probability of the outcome of adding $\boldsymbol{X}_i^{'}$ within the parameter area defined by $\boldsymbol{X}_i$ and $\boldsymbol{X}_i^{''}$; $P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{'}, \boldsymbol{X}_i^{''}]\}$ represents the probability of the outcome of having $\boldsymbol{X}_i$, $\boldsymbol{X}_i^{'}$, and $\boldsymbol{X}_i^{''}$; $P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\}$ represents the probability of the outcome of having $\boldsymbol{X}_i$ and $\boldsymbol{X}_i^{''}$; $P\{[\boldsymbol{X}_i, \boldsymbol{X}_i^{'}]\}$ represents the probability of the outcome of having $\boldsymbol{X}_i$ and $\boldsymbol{X}_i^{'}$; $P\{[\boldsymbol{X}_i^{'}, \boldsymbol{X}_i^{''}]\}$ represents the probability of the outcome of having $\boldsymbol{X}_i^{'}$ and $\boldsymbol{X}_i^{''}$.

5. Calculate $P\{\boldsymbol{X}_i^{'}|[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\}$, as defined in formula (E.15), of all the available combinations that fit the rules in step 3. Select $[\boldsymbol{X}_i, \boldsymbol{X}_i^{'}, \boldsymbol{X}_i^{''}]$ that has the maximum value of $P\{\boldsymbol{X}_i^{'}|[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\}$, referred to as $[\boldsymbol{X}_*, \boldsymbol{X}_*^{'}, \boldsymbol{X}_*^{''}]$, as expressed in formula (E.16),

$$[\boldsymbol{X}_*, \boldsymbol{X}_*^{'}, \boldsymbol{X}_*^{''}] = argmax(P\{\boldsymbol{X}_i^{'}|[\boldsymbol{X}_i, \boldsymbol{X}_i^{''}]\}). \tag{E.16}$$

With their correlated EM responses, $\boldsymbol{Y}_*, \boldsymbol{Y}_*^{'}, \boldsymbol{Y}_*^{''}$, three data samples are selected, $(\boldsymbol{X}_*, \boldsymbol{Y}_*)$, $(\boldsymbol{X}_*^{'}, \boldsymbol{Y}_*^{'})$, and $(\boldsymbol{X}_*^{''}, \boldsymbol{Y}_*^{''})$. The objective is to select the parameter area that has the highest probability of the outcome of adding a new data sample within it.

6. Compare $|\boldsymbol{Y}_* - \boldsymbol{Y}_*^{'}|$ and $|\boldsymbol{Y}_*^{'} - \boldsymbol{Y}_*^{''}|$. $[\boldsymbol{X}_\alpha, \boldsymbol{X}_\beta]$ equals $[\boldsymbol{X}_*, \boldsymbol{X}_*^{'}]$ if $|\boldsymbol{Y}_* - \boldsymbol{Y}_*^{'}|$ is bigger than $|\boldsymbol{Y}_*^{'} - \boldsymbol{Y}_*^{''}|$, otherwise $[\boldsymbol{X}_*^{'}, \boldsymbol{X}_*^{''}]$, as formulated in formula (E.17),

$$[\boldsymbol{X}_\alpha, \boldsymbol{X}_\beta] = argmax(|\boldsymbol{Y}_* - \boldsymbol{Y}_*^{'}|, |\boldsymbol{Y}_*^{'} - \boldsymbol{Y}_*^{''}|). \tag{E.17}$$

It aims to pick out the sub-area within the selected parameter area that has a higher probability of the outcome of adding a new data sample within it.

7. A new parameter setting $\boldsymbol{X}_{new}$ is defined within the sub-area defined by $[\boldsymbol{X}_\alpha, \boldsymbol{X}_\beta]$. The definition is shown as formula (E.18),

$$\boldsymbol{X}_{new} = \omega \cdot \boldsymbol{X}_\alpha + (1 - \omega) \cdot \boldsymbol{X}_\beta. \tag{E.18}$$

Here, ω is a value between 0 and 1 (0 and 1 not included). In this paper, ω is fixed as 0.5. ω can be set as a random value if higher freedom is needed to more extensively exploit the parameter space. Obtain its corresponding EM responses $\boldsymbol{Y}_{new}$ through simulation. $\boldsymbol{X}_{new}$ and $\boldsymbol{Y}_{new}$ form a new data sample ($\boldsymbol{X}_{new}$, $\boldsymbol{Y}_{new}$).

8. The new data sample ($\boldsymbol{X}_{new}$, $\boldsymbol{Y}_{new}$) is added to the existing dataset, and repeat from step 3 until the dataset is sufficient for training. Whether it is sufficient or not can be tested by evaluating the accuracy of an ML model after training with the existing dataset.

9. After sufficient data have been obtained through the proposed sampling strategy, they are utilized to train the ML model. In this paper, we use forward and inverse ML models that project geometric parameter settings to corresponding EM responses and project EM responses to corresponding geometric parameter settings, respectively. Each ML model is composed of several dense layers. The last layer is followed with the Linear activation function, and all the rest layers are followed with the Rectified Linear Unit (ReLU) activation function [29]. For each data sample, EM responses and geometrical parameters are normalized and formulated into two vectors as input and output, respectively. Normalization means projecting the original values into a fixed range from 0 to 1 proportionally. The normalized values of each EM response are equal to, the differences between the original values and the minimum value of all the EM responses divided by the difference between the maximum and minimum values of all the EM responses. The normalized values of each geometrical parameter are equal to, the differences between the original values and the minimum value of this parameter divided by the difference between the maximum and minimum values of this parameter. The mean square error (MSE) of the output is taken as the loss function, and utilized for upgrading the weights and biases of the model by the Gaussian-based backpropagation process. The upgradation is ruled by the Adaptive Moment Estimate (Adam) optimizer [30]. All the aforementioned hyperparameters, including the number of layers, the number of neurons for each layer, the number of epochs, the batch size, and the learning rate, can be decided and adjusted according to specific implementations. The numbers of neurons for the input and output layers are decided by the vector sizes of geometric parameters and EM responses. The training process completes when it is the last epoch, the model loss stops converging, or the test loss meets a pre-defined threshold. It is worth mentioning that one can also use a generative model here.

10. After training, designers can use the forward model to replace the sim-

ulator or use the inverse model to determine the geometric parameter setting for given EM responses in milliseconds.

The core of the proposed Bayesian-based sampling is to estimate the quality of a new data sample in a parameter area by evaluating the quality of an existing data sample in the parameter area. The quality of a new data sample represents the contribution of adding this new data sample to the parameter area; the quality of an existing data sample represents the contribution of adding the existing data sample to the parameter area. In each parameter area defined by $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$, $(\boldsymbol{X}_i^{'}, \boldsymbol{Y}_i^{'})$, and $(\boldsymbol{X}_i^{''}, \boldsymbol{Y}_i^{''})$, $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$ and $(\boldsymbol{X}_i^{''}, \boldsymbol{Y}_i^{''})$ defines the past state, $(\boldsymbol{X}_i^{'}, \boldsymbol{Y}_i^{'})$ represents the existing data sample that has been added to the parameter area. $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$, $(\boldsymbol{X}_i^{'}, \boldsymbol{Y}_i^{'})$, and $(\boldsymbol{X}_i^{''}, \boldsymbol{Y}_i^{''})$ defines its current state. The contribution of adding $(\boldsymbol{X}_i^{'}, \boldsymbol{Y}_i^{'})$ is approximated as the contribution of the existing state $((\boldsymbol{X}_i, \boldsymbol{Y}_i)$, $(\boldsymbol{X}_i^{'}, \boldsymbol{Y}_i^{'})$, and $(\boldsymbol{X}_i^{''}, \boldsymbol{Y}_i^{''}))$ divided by the contribution of the past state $((\boldsymbol{X}_i, \boldsymbol{Y}_i)$ and $(\boldsymbol{X}_i^{''}, \boldsymbol{Y}_i^{''}))$, which approximates the contribution of adding a new data sample in this parameter area. All the combinations of $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$, $(\boldsymbol{X}_i^{'}, \boldsymbol{Y}_i^{'})$, and $(\boldsymbol{X}_i^{''}, \boldsymbol{Y}_i^{''})$ are considered and evaluated, and the quality of new data in all the parameter areas are estimated. In each iteration, a new data sample of the highest estimated quality is simulated and added to the existing dataset. The proposed Bayesian-based sampling leverages precious computation resources to simulate high-quality data iteratively, hence generating a higher-quality dataset and leading to a higher model accuracy using the same amount of simulation runs compared to conventional methods. Therefore, for realizing the same desired model accuracy, the proposed method can reduce the required simulation and improve the efficiency significantly compared to conventional methods.

3 Validation

We validate the efficiency of the proposed method by implementing it in two real designs and comparing training results with typical methods. The first design example is the design of a unit cell of metasurface in [31], and the other is the design of a microwave filter in [32]. We also extend example I to a five-dimensional example III for investigating the performance of the proposed method for higher dimensionality. Note that the simulation is supported by CST.

3.1 Example I: Unit Cell of Metasurface

We apply the proposed method for accelerating the design of a frequency-selective surface that operates from 40 GHz to 90 GHz. Its unit cell is developed based on aperture-coupled resonators, as shown in Fig. E.3. Two identical

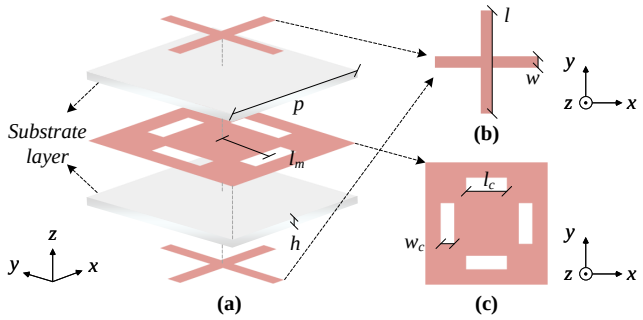


Fig. E.3: Example I: Structure of the unit cell of metasurface [31]. (a) 3D view; (b) Top and bottom metal layers; (c) Middle metal layer.

Table E.1: Example I: Geometric Parameters of the Unit Cell of Metasurface [31]

Parameter	Value (mm)	Parameter	Value (mm)
l_c	[0.4, 0.7]	l	1.87
w_c	[0.1, 0.4]	w	0.27
h	0.13	l_m	0.9
p	2.5		

Jerusalem-shaped crosses act as resonators and are etched on the top and bottom metal layers. Four rectangular slots are etched on the middle metal layer as its coupling aperture. Every two metal layers are separated by a $RogersRT5880$ substrate layer with relative permittivity of $\varepsilon_r = 2.2$ and loss tangent of $tan\delta = 0.004$. The geometric parameters are marked in Fig. E.3 and listed in Table E.1. Among them, l_c and w_c are the key parameters to adjust between specific ranges, while the rest parameters are fixed as constant values.

We define a forward model that projects the values of the two key geometric parameters (l_c and w_c) to the transmission coefficients ($|S_{21}|$) from 40 GHz to 90 GHz. The two adjustable geometric variables are represented as a 2-element vector ($[l_c, w_c]$). Fig. E.4 exhibits $|S_{21}|s$ of four samples with $[l_c, w_c]$ being fixed at four different settings, $[l_c, w_c]$ = [0.4 mm, 0.1 mm], [0.5 mm, 0.2 mm], [0.6 mm, 0.3 mm], and [0.7 mm, 0.4 mm]. $|S_{21}|$ of each sample is discretely sampled from 40 GHz to 90 GHz at a uniform frequency step of 1 GHz, and it is represented as a 51-element vector ($[|S_{21}|]$). Thus, the forward model has an input size of 1 by 2 and an output size of 1 by 51.

For comparison, the forward model is first developed from scratch by using the conventional method. Within the parameter space of l_c and w_c, a fine and uniform grid is defined as 31×31. On the defined grid, 31×31 = 961 parameter settings are acquired, and the corresponding 961 $|S_{21}|$ results are generated via full-wave simulation. Among these 961 data samples, 850 data samples are

Table E.2: Example I: Architecture of the Machine Learning Model

No.	Layer	Neurons	Function
1	Input layer	2	Input: $[l_c, w_c]$
2	Activation function	-	ReLU
3	Hidden layer	180	Fully-connected layer
4	Activation function	-	ReLU
5	Hidden layer	120	Fully-connected layer
6	Activation function	-	ReLU
7	Output layer	51	Output: $[\lvert S_{21}\rvert]$
8	Activation function	-	Linear
	Loss function		Mean Squared Error
	Optimizer		Adam
	Learning rate		0.001
	Batch size		20
	Epochs		200

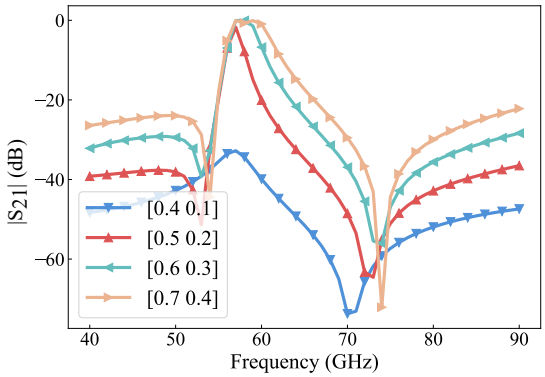


Fig. E.4: $|S_{21}|s$ of four samples with $[l_c, w_c]$ = [0.4 mm, 0.1 mm], [0.5 mm, 0.2 mm], [0.6 mm, 0.3 mm], and [0.7 mm, 0.4 mm].

arbitrarily chosen and grouped as a training dataset, and the rest 111 data samples form a test dataset. The allocation ratio of training and testing data is 88.4 % and 11.6 %. This allocation process will be repeated later on using different random seeds. We do not define a validation dataset, as ten-fold cross-validation is applied instead. The forward model is trained on the training dataset with ten-fold cross-validation and tested on the test dataset. Using ten-fold cross-validation and Bayesian optimization [33], we sweep and optimize the hyperparameters of the model, including the learning rate, the batch size, the number of neurons for each layer, and the number of epochs. The setting of hyperparameters that avoids overfitting and leads to the minimum loss is fixed as the final combination. Table E.2 introduces the optimized architecture of

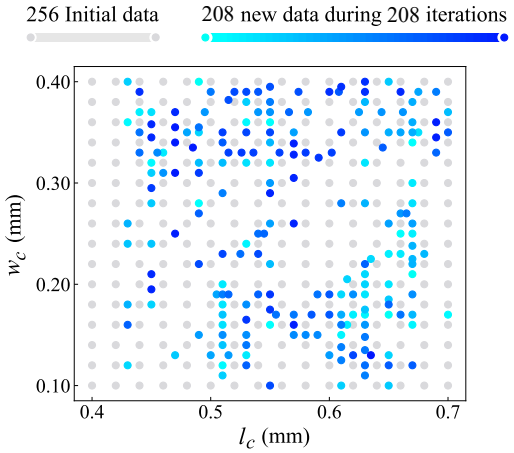


Fig. E.5: Example I: Training data distribution of the proposed method.

the forward model. The input is of size 2 for the two geometrical parameters l_c and w_c; the output is of size 51 for the 51 discretely sampled values of each $|S_{21}|$ curve from 40 GHz to 60 GHz. It consists of two fully-connected neural layers with 180 and 120 neurons as two hidden layers. ReLU follows each hidden layer as its activation function. Its loss function is MSE between the predicted and real values of normalized $[|S_{21}|]$. During each epoch, the MSE is utilized to upgrade the weights and biases of neurons guided by the optimizer Adam. The batch size is 20, the learning rate is 0.001, and the number of epochs is 200. For the sake of reliability, we repeat the data allocation and training process for five rounds using different random seeds, and the average results are taken as the final results. Its final test loss is 7.5×10^{-4} and is taken as the threshold value. It is worth noting that the same test dataset is used for evaluating both the model generated using the conventional method and the model generated using the proposed method later.

We use the proposed method to train the forward model again from scratch for comparing and validating the improvement of the proposed method. Note that this process is also repeated for five rounds using the aforementioned five random seeds, and the average results are taken as the final results. To start with, l_c and w_c are swept on a sparse, uniform grid 16×16 of the whole parameter space. 256 parameter settings are defined in total, and the corresponding $|S_{21}|$s from 40 GHz to 90 GHz are obtained through full-wave simulation. Taking the 256 data samples as an initial dataset, we iteratively generate new data samples guided by the proposed Bayesian inspired sampling. $\boldsymbol{X}_u$ is set as [0.02, 0.02] and normalized as [0.05, 0.05] during sampling. After 208 iterations, 208 new data samples are generated, and the initial dataset has been expanded into a larger dataset that consists of 464 data samples. The expanded dataset

	Uniform sampling				Proposed sampling			
	NN	SVRM	GRNN	PDRN	NN	SVRM	GRNN	PDRN
Training data		850			256 + 208 (↓ 386)	256 + 216 (↓ 378)	256 + 264 (↓ 330)	256 + 226 (↓ 368)
Simulation time (h)		19.15			10.45 (↓ 8.70)	10.63 (↓ 8.52)	11.72 (↓ 7.43)	10.86 h (↓ 8.29)
Iteration time (h)		– –			0.54 (↑ 0.54)	0.57 (↑ 0.57)	0.69 (↑ 0.69)	0.60 (↑ 0.60)
Total time (h)		19.15			10.99 (↓ 8.16)	11.20 (↓ 7.95)	12.41 (↓ 6.74)	11.46 (↓ 7.69)
Test loss (10^{-3})	0.75	1.61	1.01	1.43	0.75	1.61	1.01	1.43

Note: NN: Neural Network (default ML technique applied in Section II); SVRM: Support Vector Regression Machine; GRNN: Generalized Regression Neural Network; PDRN: Polynomial Design Response Network.

Table E.3: Example I: Comparison Between Results of the Proposed Sampling and Conventional Uniform Sampling Integrated with Different ML Techniques

is used to train the forward model. After training, the forward model is then

Table E.4: Example I: Comparison Between Results of Proposed Method and Conventional Method Using Wider Parameter Range

	Conventional method	**Proposed method**
Training data	900	289 + 218 (↓ 393)
Simulation time (h)	20.28	11.42 (↓ 8.86)
Iteration time (h)	−−	0.57 (↑ 0.57)
Total time (h)	20.28	11.99 (↓ 8.29)
Test loss (10^{-3})	3.96	3.96

Note: Proposed method: Proposed sampling with NN;
Conventional method: Uniform sampling with NN.

evaluated on the test dataset that consists of the 111 data samples. The final test loss is 7.5×10^{-4} and reaches the pre-obtained threshold value.

The distribution of generated training data is shown in Fig. E.5. Gray dots represent the 256 initial data. 208 new data are represented as blue dots: lighter blue dots correlate to the earlier iterations, and darker blue dots correspond to the later iterations. It can be observed that the distribution of training data is optimized. Sampling gets intensive where either l_c or w_c is fixed at large values.

The results of conventional methods (uniform sampling with NN) and the proposed method (proposed sampling with NN) are compared in Table E.3. Compared with the conventional method, the proposed method saves 386 training data and 8.70 h simulation time for the same model accuracy. The reduction of training data and simulation time is around 45.4 % ($\frac{386}{850} = \frac{8.70\,\text{h}}{19.15\,\text{h}} \approx 45.4\,\%$). Considering that the proposed method takes 0.54 h for executing the iterations, the total time reduction is 8.16 h (8.70 h − 0.54 h = 8.16 h). Therefore, the efficiency of the proposed method outperforms the conventional method by 42.6 % ($\frac{8.16\,\text{h}}{19.15\,\text{h}} \approx 42.6\,\%$).

We also investigate the results of our sampling method being integrated with other ML surrogate modeling techniques: Support Vector Regression Machine (SVRM), Generalized Regression Neural Network (GRNN), and 3-order Polynomial Design Response Network (PDRN). They are also compared with the results of uniform sampling being integrated with these three techniques, as can be seen in Table E.3. The reductions of total time in these three cases are 7.95 h, 6.74 h, and 7.69 h, hence improving the efficiency by 41.5 % ($\frac{7.95\,\text{h}}{19.15\,\text{h}} \approx 41.5\,\%$), 35.2 % ($\frac{6.74\,\text{h}}{19.15\,\text{h}} \approx 35.2\,\%$), and 40.2 % ($\frac{7.69\,\text{h}}{19.15\,\text{h}} \approx 40.2\,\%$), respectively.

To investigate the performance of the proposed method in case of a wider range of parameter space, we extend the variable ranges of l_c and w_c to [0.36 mm, 1 mm] and [0.06 mm, 0.7 mm], respectively, and then we reimplement the proposed method in example I. The comparative results (uniform sampling with NN and proposed sampling with NN) are listed in Table E.4.

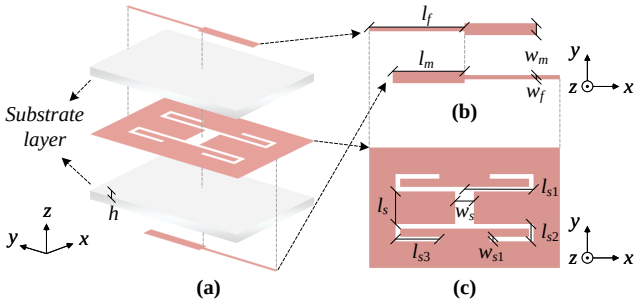


Fig. E.6: Example II: Structure of the microwave filter [32]. (a) 3D view; (b) Top and bottom metal layers; (c) Middle metal layer.

As can be observed, a 43.7 % ($\frac{393}{900} \approx 43.7\,\%$) reduction of simulation data and a 40.9 % ($\frac{8.29\,\mathrm{h}}{20.28\,\mathrm{h}} \approx 40.9\,\%$) improvement of efficiency can still be achieved when the variable range is extended.

By employing the well-trained forward model, $|S_{21}|$ results for new settings of l_c and w_c can be determined within seconds instead of repeating the time-consuming and tedious full-wave simulation, hence significantly improving the design process.

3.2 Example II: Microwave Filter

The proposed method is applied for the design of a microwave filter presented by Yang *et al.* [32] for validation. The structure of the microwave filter is exhibited in Fig. E.6. It is composed of three metal layers and two substrate layers. Two substrate layers are $RogersRO4003C$ with relative permittivity of $\varepsilon_r = 3.38$ and loss tangent of $tan\delta = 0.0027$. The top and bottom metal layers are open-ended two-order microstrip feedlines, and they are centrosymmetric to each other. The middle layer is rectangular and is etched with a symmetric four-branched slot. All the branches are folded and identical. Its geometric parameters are marked in Fig. E.6. Three geometric parameters (l_{s1}, l_{s2}, l_{s3}) are selected as adjustable variables, while the rest parameters are fixed as constant values. The minimum and maximum values for the three adjustable variables and the constant values for the rest parameters are listed in Table E.5.

An inverse model is defined to assist the design of this microwave filter. It aims to determine the three adjustable geometrical variables (l_{s1}, l_{s2}, l_{s3}) directly for yielding the desired $|S_{11}|$. Each setting of the three adjustable geometric variables is represented as a 3-element vector ($[l_{s1}, l_{s2}, l_{s3}]$). The frequency band of interest is from 0.1 GHz to 4.5 GHz, and the achieved operating band in [32] is around from 1 GHz to 3.3 GHz. $|S_{11}|s$ of four samples are shown in Fig. E.7, where $[l_{s1}, l_{s2}, l_{s3}]$ is fixed at four different settings, $[l_{s1}, l_{s2}, l_{s3}]$ = [11.5 mm, 2 mm, 9.5 mm], [11.7 mm, 2.7 mm, 9.6 mm], [12 mm,

Table E.5: Example II: Geometric Parameters of the Microwave filter [32]

Parameter	Value (mm)	Parameter	Value (mm)
l_{s1}	[11.5, 12.5]	w_f	1.5
l_{s2}	[2, 3]	w_s	1.5
l_{s3}	[9.5, 10.5]	w_{s1}	0.5
l_m	20.5	w_m	3.88
l_s	9.9	h	0.813
l_f	24		

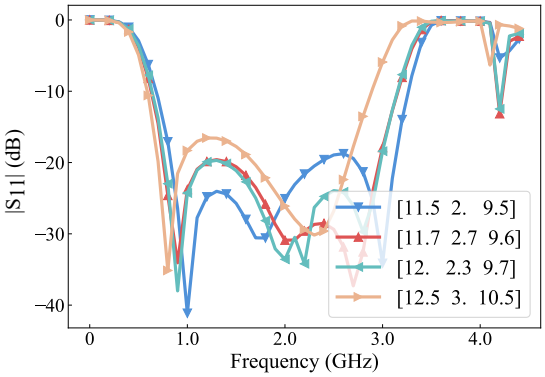


Fig. E.7: $|S_{11}|$s of four samples with [l_{s1}, l_{s2}, l_{s3}] = [11.5 mm, 2 mm, 9.5 mm], [11.7 mm, 2.7 mm, 9.6 mm], [12 mm, 2.3 mm, 9.7 mm], and [12.5 mm, 3 mm, 10.5 mm].

2.3 mm, 9.7 mm], and [12.5 mm, 3 mm, 10.5 mm]. The corresponding $|S_{11}|$ of each sample is discretely sampled at a uniform frequency step of 0.1 GHz and represented as a 45-element vector ([$|S_{11}|$]). Accordingly, the input size of the surrogate model is 1 by 45, and its output size is 1 by 3.

The inverse model is first developed from scratch by using the conventional method. Within the parameter space of l_{s1}, l_{s2} and l_{s3}, a fine and uniform grid of $11 \times 11 \times 11$ is defined. Based on the defined grid, $11 \times 11 \times 11 = 1331$ parameter settings are determined, and the corresponding $|S_{11}|$ results are then acquired via simulation and form 1331 data samples together with the correlated parameter settings. Among these 1331 data samples, 1200 data samples are arbitrarily selected and form a training dataset and the rest 131 data samples form a test dataset. The allocation ratio of data for training and testing is 90.2 % and 9.8 %. This allocation process will be repeated later on using different random seeds. We do not separate a validation dataset, as ten-fold cross-validation is utilized instead. Note that this test dataset is used to evaluate the surrogate model generated by both the conventional method and the proposed method. We sweep and optimize the hyperparameters of the surrogate model, including learning rate, batch size, the number of neurons

Table E.6: Example II: Architecture of the Machine Learning Model

No.	Layer	Neurons	Function
1	Input layer	45	Input: $[\lvert S_{11}\rvert]$
2	Activation function	-	ReLU
3	Hidden layer	120	Fully-connected layer
4	Activation function	-	ReLU
5	Hidden layer	160	Fully-connected layer
6	Activation function	-	ReLU
7	Output layer	3	Output: $[l_{s1}, l_{s2}, l_{s3}]$
8	Activation function	-	Linear
	Loss function		Mean Squared Error
	Optimizer		Adam
	Learning rate		0.001
	Batch size		10
	Epochs		160

for each layer, and the number of epochs, by using ten-fold cross-validation and Bayesian optimization [33]. The setting of hyperparameters that avoids overfitting and leads to the minimum loss is fixed as the final combination. The detailed final hyperparameters of the surrogate model are listed in Table E.6: it consists of two fully-connected neural layers with 120 and 160 neurons each as hidden layers; ReLU follows each layer as its activation function; MSE between the predicted and real values of normalized $[l_{s1}, l_{s2}, l_{s3}]$ is taken as the loss function; the obtained MSE is utilized to upgrade the weights and biases of neurons guided by the optimizer Adam; the learning rate is 0.001; the batch size is 10; the total number of epochs is 160. The inverse model is then trained using the training dataset and evaluated on the test dataset. To ensure reliability, we repeat the data allocation and training process for five rounds using different random seeds, and the average results are taken as the final results. Its final test loss 7.1×10^{-3} is set as the threshold value.

We then utilize the proposed method to develop the inverse model. Note that this process is also repeated for five rounds using the aforementioned five random seeds, and the average results are taken as the final results. 216 settings of geometric parameters are initialized by sweeping on a uniform grid of 6×6×6 within the three-dimensional parameter space confined by the minimum and maximum values of the three adjustable geometric parameters. $\boldsymbol{X}_u$ is defined as [0.2, 0.2, 0.2], and it is normalized into [0.2, 0.2, 0.2] during sampling. After 464 iterations, 464 new data samples are generated by executing the proposed Bayesian inspired sampling. The initial dataset has now been expanded into an optimized dataset with 680 data samples in total. The optimized dataset is used to train the inverse model. The well-trained inverse model is evaluated

	Uniform sampling				Proposed sampling			
	NN	SVRM	GRNN	PDRN	NN	SVRM	GRNN	PDRN
Training data		1200			216 + 464 (↓ 520)	216 + 488 (↓ 496)	216 + 472 (↓ 512)	– –
Simulation time (h)		74.80			42.39 (↓ 32.41)	43.88 (↓ 30.92)	42.89 (↓ 31.91)	– –
Iteration time (h)		– –			5.96 (↑ 5.96)	6.27 (↑ 6.27)	6.06 (↑ 6.06)	– –
Total time (h)		74.80			48.35 (↓ 26.45)	50.15 (↓ 24.65)	48.95 (↓ 25.85)	– –
Test loss (10^{-2})	0.71	2.49	2.15	NAN	0.71	2.49	2.15	NAN

Note: NN: Neural Network (default ML technique applied in Section II); SVRM: Support Vector Regression Machine; GRNN: Generalized Regression Neural Network; PDRN: Polynomial Design Response Network.

Table E.7: Example II: Comparison Between Results of the Proposed Sampling and Conventional Uniform Sampling Integrated with Different ML Techniques

on the same test dataset again. The final test loss equals 7.1×10^{-3} and realizes the threshold value. Therefore, the proposed method reduces the amount of required simulation data by 520 for the same model accuracy compared with the conventional method.

Table E.8: Example II: Comparison Between Results of Proposed Method and Conventional Method Using Wider Parameter Range

	Conventional method	Proposed method
Training data	600	125 + 218 (↓ 257)
Simulation time (h)	37.40	21.38 (↓ 16.02)
Iteration time (h)	−−	2.15 (↑ 2.15)
Total time (h)	37.40	23.53 (↓ 13.87)
Test loss (10^{-2})	2.12	2.12

Note: Proposed method: Proposed sampling with NN;
Conventional method: Uniform sampling with NN.

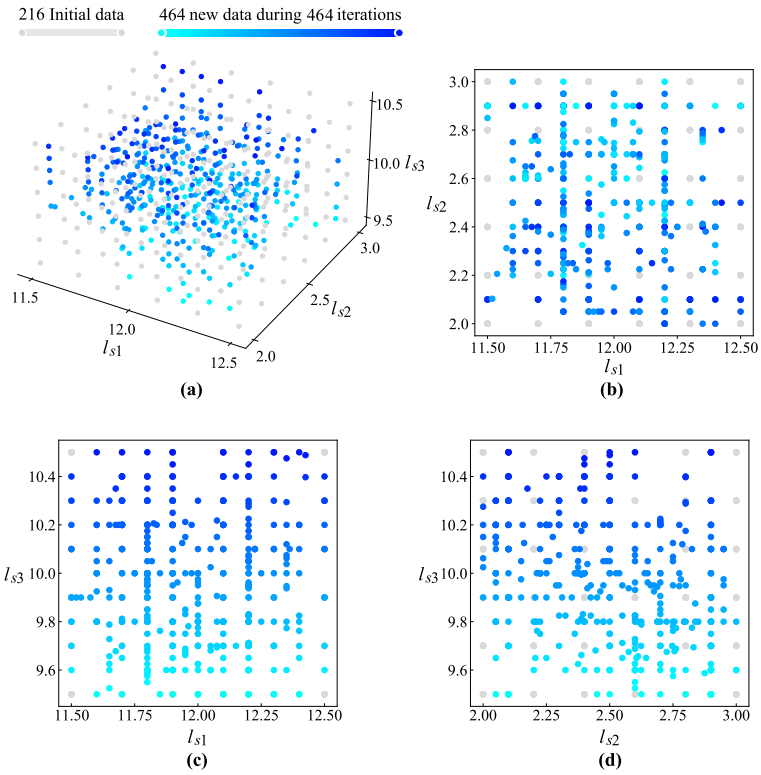


Fig. E.8: Example II: Training data distribution of the proposed method. (a) Distribution over l_{s1}, l_{s2}, and l_{s3}. (b) Distribution over l_{s1} and l_{s2}. (c) Distribution over l_{s1} and l_{s3}. (d) Distribution over l_{s2} and l_{s3}.

The distribution of the optimized training data is exhibited in Fig. E.8. Initial data are represented as gray dots, and new data are represented as blue dots. The lightness of blue dots corresponds to the order of new data being

Table E.9: Example III: Geometric Parameters of the Unit Cell of Metasurface [31]

Parameter	Value (mm)	Parameter	Value (mm)
l_c	[0.6, 1]	l	[1.7, 2.1]
w_c	[0.1, 0.5]	w	[0.1, 0.3]
h	[0.05, 0.15]	l_m	0.9
p	2.5		

generated. Data generated during earlier iterations are represented as lighter blue dots, and those during later iterations are represented as darker blue dots. It can be observed that the sampling densities of l_{s1}, l_{s2}, and l_{s3} are carefully manipulated by the proposed Bayesian inspired sampling strategy.

Training results of the proposed method (Proposed sampling method with NN) are shown and compared with conventional ML methods (uniform sampling with NN) in Table E.7. To converge beneath a prediction error around 7.1×10^{-3} , the proposed method requires 680 training data samples. In contrast, conventional ML methods using uniform sampling require 1200 training data samples. On average, each simulation cycle for simulating one parameter setting takes 3.74 min. Consequently, the required simulation time is reduced from 74.80 h to 42.39 h. Therefore, the proposed ML method realizes 43.3 % reduction of training data ($\frac{1200-680}{1200} \approx 43.3\,\%$) and simulation time ($\frac{74.80\,\mathrm{h}-42.39\,\mathrm{h}}{74.80\,\mathrm{h}} \approx 43.3\,\%$). Note that the execution of iterations takes 5.96 h, hence the total time reduction is 35.4 % ($\frac{74.80\,\mathrm{h}-42.39\,\mathrm{h}+5.96\,\mathrm{h}}{74.80\,\mathrm{h}} \approx 35.4\,\%$). Therefore, the efficiency improvement of the proposed method in example II is 35.4 %.

Again, we validate the proposed sampling method by integrating it with the three typical ML surrogate modeling techniques: SVRM, GRNN, and 3-order PDRN. Its corresponding results are listed in Table E.7 for comparison with the results of uniform sampling being integrated with these three techniques. The 3-order PDRN fails to converge in this case, as the input dimension (45) is relatively too high. The reductions of total time in the first two cases are 24.65 h and 25.85 h, and the efficiencies are improved by 33.0 % ($\frac{24.65\,\mathrm{h}}{74.80\,\mathrm{h}} \approx 33.0\,\%$) and 34.6 % ($\frac{25.85\,\mathrm{h}}{74.80\,\mathrm{h}} \approx 34.6\,\%$), respectively.

We validate the performance of the proposed method in case of a wider range of parameter space by extending the variable ranges of l_{s1}, l_{s2}, and l_{s3} to [10 mm, 14 mm], [2 mm, 6 mm], and [8 mm, 12 mm], respectively and reimplementing the proposed method in example II. The amount of training data is reduced to 600 for applying uniform sampling on a uniform grid of $9 \times 9 \times 9$ in the parameter space. The grid density for each dimension is fixed at 9 for setting the minimum parameter step as 0.5 mm (the range of each parameter is 4 mm). Among the $9 \times 9 \times 9 = 729$ data, 600 data are arbitrarily chosen as the training data, and the rest 129 data are taken as the test data. Note that the model loss increases from 0.71×10^{-2} to 2.12×10^{-2} because of

	Uniform sampling			Proposed sampling			Latin hypercube sampling		
	Size 1	Size 2	Size 3	Size 1	Size 2	Size 3	Size 1	Size 2	Size 3
Training data	300	500	800	108 + 64 = 172	162 + 132 = 294	243 + 226 = 469	172	294	469
Simulation time (h)	6.76	11.26	18.02	3.88	6.62	10.56	3.88	6.62	10.56
Iteration time (h)	– –	– –	– –	0.19	0.36	0.61	– –	– –	– –
Total time (h)	6.76	11.26	18.02	4.07	6.98	11.17	3.88	6.62	10.56
Test loss (10^{-2})	3.54	3.26	3.70	3.54	3.26	3.70	4.14	3.79	4.62
	Require more data			**Require reduced data for higher model accuracy**			**Lower model accuracy**		

Table E.10: Example III: The Proposed Sampling Compared to Conventional Uniform Sampling and Latin Hypercube Sampling for Datasets of Different Sizes

the reduction of training data. The increased model loss does not affect the validation, as we just need to validate that the proposed method can reduce

the required data while maintaining the model loss. The comparative results (uniform sampling with NN and proposed sampling with NN) are listed in Table E.8. As can be observed, a 42.8 % ($\frac{257}{600} \approx 42.8\,\%$) reduction of simulation data and a 37.1 % ($\frac{13.87\,\mathrm{h}}{37.40\,\mathrm{h}} \approx 37.1\,\%$) improvement of efficiency can still be achieved in example II with a wider parameter range.

The well-trained inverse model can then be utilized to conveniently determine an optimal geometric parameter setting $[l_{s1}, l_{s2}, l_{s3}]$ for generating a desired $|S_{11}|$. Designers just need to discretely sample the desired $|S_{11}|$ as a 1 by 45 vector $[|S_{11}|]$, then normalize and input it to the well-trained model. The well-trained model can directly determine an optimal setting of normalized $[l_{s1}, l_{s2}, l_{s3}]$. The actual parameter setting can be obtained after denormalization. It will save human designers significant time and energy by avoiding repetitive basic tasks, freeing them up to focus on more meaningful and innovative work.

3.3 Example III: Five-Dimensional Extension of Example I

To explore the performance of the proposed method as the dimensionality increases, we extend the 2-dimensional example I to the 5 dimensional example III. We increase the number of parameters to be adjusted from 2 to 5 and reimplement the proposed method. The original example I only has 2 adjustable parameters l_c and w_c. In the extended example III, there are 5 parameters to be adjusted: l_c, w_c, l, w, h. The parameter range is updated and listed in Table E.9.

We investigate the performance of the proposed method compared to conventional uniform sampling and Latin hypercube sampling for datasets of different sizes. For the sake of simplicity, we omit the descriptions of the detailed processes here, as they are basically repetitions of example I as described in Section III-A. The results are listed and compared in Table E.10.

As can be observed from Table E.10, a significant reduction of simulation data and improvement of efficiency can still be achieved for datasets of different sizes when the dimensionality increases from 2 to 5. Compared with uniform sampling using datasets of sizes 300, 500, and 800, the proposed sampling requires only 172, 294, and 469, respectively, hence improving the efficiency significantly. To compare with Latin hypercube sampling, we generate datasets of the same sizes of 172, 294, and 469 by using the Latin hypercube sampling technique. The three datasets are then used to train the surrogate model. We compare the results of the proposed method and Latin hypercube sampling using datasets of the same sizes of 172, 294, and 469 in Table E.10. Compared to Latin hypercube sampling, the proposed method also outperforms in terms of efficiency. Latin hypercube sampling results in on average 20 % higher model losses, which represents a 20 % deterioration of the model accuracy, when using the same-sized training datasets.

4 Conclusion

In conclusion, this paper proposes an efficient machine learning (ML) method for the design of microwave components, which utilizes Bayesian inspired sampling. Compared with conventional ML methods, the proposed method significantly reduces the required amount of data and simulation by optimizing the distribution of training data according to the Bayesian rule. We define a Bayesian inspired expression for evaluating the probability of the outcome of adding a new data sample in a specific parameter area. The parameter area with the highest probability of the outcome is selected for generating new data during each iteration. The proposed sampling strategy optimizes the distribution of training data dynamically within the parameter space, leading to a significant reduction of the amount of training data for the same model accuracy. The comparative results show that the proposed method achieves the same model accuracy with significantly reduced simulation data, indicating a higher efficiency compared to conventional methods. Other than the improvement of efficiency, the proposed sampling method may provide insights into the sensitivity of geometric parameters in electromagnetic structures. By analyzing the distribution of generated high-quality data, as shown in Fig. E.5 and E.8, it may assist human designers to identify the most sensitive parameters and ranges, enable a better understanding of the working principles, and guide the design decisions. The proposed method has great potential for practical applications in the design of microwave components and related fields. Further investigations on the extension of this method to other design problems and the exploration of its theoretical foundation are worth pursuing in future work.

References

[1] X. Fang, H. Li, and Q. Cao, "Design of reconfigurable periodic structures based on machine learning," *IEEE Trans. Microw. Theory Techn.*, early access, 2023, doi: 10.1109/TMTT.2022.3233740.

[2] Z. Zhou, Z. Wei, J. Ren, Y. Yin, G. F. Pedersen and M. Shen, "Representation learning-driven fully automated framework for the inverse design of frequency-selective surfaces," *IEEE Trans. Microw. Theory Techn.*, vol. 71, no. 6, pp. 2409–2421, Jun., 2023, doi: 10.1109/TMTT.2023.3235066.

[3] F. Feng, W. Na, J. Jin, J. Zhang, W. Zhang, and Q.-J. Zhang, "Artificial neural networks for microwave computer-aided design: the state of the art," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 11, pp. 4597–4619, Nov. 2022, doi: 10.1109/TMTT.2022.3197751.

[4] Z. Wei, Z. Zhou, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Fully automated design method based on reinforcement learning and surrogate modeling for antenna array decoupling," *IEEE Trans. Antennas Propag.*, vol. 71, no. 1, pp. 660-671, Jan. 2023, doi: 10.1109/TAP.2022.3221613.

[5] N. Kazemi, M. Abdolrazzaghi, P. Musilek, and M. Daneshmand, "A temperature-compensated high-resolution microwave sensor using artificial neural network," *IEEE Microw. Wireless Compon. Lett.*, vol. 30, no. 9, pp. 919–922, Sep. 2020, doi: 10.1109/LMWC.2020.3012388.

[6] H. M. Torun, A. C. Durgun, K. Aygün, and M. Swaminathan, "Causal and passive parameterization of S-parameters using neural networks," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 10, pp. 4290–4304, Oct. 2020, doi: 10.1109/TMTT.2020.3011449.

[7] L. Xiao, W. Shao, X. Ding, and B. Wang, "Dynamic adjustment kernel extreme learning machine for microwave component design," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 10, pp. 4452–4461, Oct. 2018, doi: 10.1109/TMTT.2018.2858787.

[8] P. Di Barba, L. Januszkiewicz, J. Kawecki, and M. E. Mognaschi, "Electromagnetic wave absorption in the human head: A virtual sensor based on a deep-learning model," *Sens.*, vol. 23, no. 6, pp. 1–19, Mar. 2023, doi: 10.3390/s23063131.

[9] J. Kawecki, L. Januszkiewicz, P. Di Barba, and K. Kropidlowski, "Eye shielding against electromagnetic radiation: Optimal design using a reduced model of the head," *Electron.*, vol. 12, no. 2, pp. 1–16, Jan. 2023, doi: 10.3390/electronics12020291.

[10] P. Di Barba, "Future trends in optimal design in electromagnetics," *IEEE Trans. Magn.*, vol. 58, no. 9, pp. 1–4, Sep. 2022, doi: 10.1109/TMAG.2022.3164204.

[11] P. Di Barba, M.E Mognaschi, and S. Wiak, "Neural metamodelling of fields: Towards a new deal in computational electromagnetics," *Int. Jnl. Appl. Electromagn. Mech.*, vol. 69, no. 2, pp. 127–137, Jun. 2022, doi: 10.3233/JAE-210222.

[12] Z. Wei, Z. Zhou, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Automated antenna design via domain knowledge-informed reinforcement learning and imitation learning," *IEEE Trans. Antennas Propag.*, vol. 71, no. 7, pp. 5549–5557, Jul. 2023, doi: 10.1109/TAP.2023.3266051.

[13] M. Sedaghat, R. Trinchero, Z. H. Firouzeh, and F. G. Canavero, "Compressed machine learning-based inverse model for design optimization of

microwave components," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 7, pp. 3415–3427, Apr. 2022, doi: 10.1109/TMTT.2022.3166151.

[14] Z. Zhou, Z. Wei, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Transfer learning assisted multi-element calibration for active phased antenna arrays," *IEEE Trans. Antennas Propag.*, vol. 71, no. 2, pp. 1982–1987, Oct. 2022, doi: 10.1109/TAP.2022.3216548.

[15] Z. Zhou, Z. Wei, Y. Zhang, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Training of deep neural networks in electromagnetic problems: a case study of antenna array pattern synthesis," in *2021 IEEE MTT-S International Wireless Symposium (IWS)*, Nanjing, China, May. 2021, pp. 1-3, doi: 10.1109/IWS52775.2021.9499638.

[16] S. Koziel and M. Abdullah, "Machine-learning-powered EM-based framework for efficient and reliable design of low scattering metasurfaces," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 4, pp. 2028–2041, Mar. 2021, doi: 10.1109/TMTT.2021.3061128.

[17] Z. Zhou, Z. Wei, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Two-order deep learning for generalized synthesis of radiation patterns for antenna arrays," *IEEE Trans. Artif. Intell.*, early access, Jul. 2022, doi: 10.1109/TAI.2022.3192505.

[18] J. A. Hodge, K. V. Mishra, and A. I. Zaghloul, "RF metasurface array design using deep convolutional generative adversarial networks," in *IEEE Int. Symp. Phased Array Syst. Technol.*, Oct. 2019, pp. 1–6.

[19] Z. Wei, Z. Zhou, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Equivalent circuit theory-assisted deep learning for accelerated generative design of metasurfaces," *IEEE Trans. Antennas Propag.*, vol. 70, no. 7, pp. 5120–5129, Feb. 2022, doi: 10.1109/TAP.2022.3152592.

[20] P. Naseri and S. V. Hum, "A Generative Machine Learning-Based Approach for Inverse Design of Multilayer Metasurfaces," *IEEE Trans. Antennas Propag.*, vol. 69, no. 9, pp. 5725–5739, Feb. 2021, doi: 10.1109/TAP.2021.3060142.

[21] P. Naseri, G. Goussetis, N. J. G. Fonseca, and S. V. Hum, "Synthesis of multi-band reflective polarizing metasurfaces using a generative adversarial network," *Sci. Rep.*, vol. 12, no. 1, pp. 1–14, Oct. 2022, doi: 10.1038/s41598-022-20851-y.

[22] S. Koziel, A. Pietrenko-Dabrowska, and M. Mahrokh, "Globalized simulation-driven miniaturization of microwave circuits by means of dimensionality-reduced constrained surrogates," *Sci. Rep.*, vol. 12, no. 1, Art. no. 1, Sep. 2022, doi: 10.1038/s41598-022-20728-0.

[23] S. Koziel and A. Pietrenko-Dabrowska, "Low-cost quasi-global optimization of expensive electromagnetic simulation models by inverse surrogates and response features," *Sci. Rep.*, vol. 12, no. 1, Art. no. 1, Nov. 2022, doi: 10.1038/s41598-022-24250-1.

[24] A. Pietrenko-Dabrowska and S. Koziel, "Globalized parametric optimization of microwave components by means of response features and inverse metamodels," *Sci. Rep.*, vol. 11, no. 1, Art. no. 1, Dec. 2021, doi: 10.1038/s41598-021-03095-0.

[25] J. A. Tomasson, A. Pietrenko-Dabrowska, and S. Koziel, "Expedited globalized antenna optimization by principal components and variable-fidelity EM simulations: Application to microstrip antenna design," *Electron.*, vol. 9, no. 4, Art. no. 4, Apr. 2020, doi: 10.3390/electronics9040673.

[26] S. Koziel and A. Pietrenko-Dabrowska, "Expedited feature-based quasi-global optimization of multi-band antenna input characteristics with Jacobian variability tracking," *IEEE Access*, vol. 8, pp. 83907–83915, May. 2020, doi: 10.1109/ACCESS.2020.2992134.

[27] Z. Wei, Z. Zhou, Y. Zhang, P. Li, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Domain knowledge assisted training dataset generation for metasurface designs," in *2021 IEEE MTT-S International Wireless Symposium (IWS)*, Nanjing, China, May. 2021, pp. 1-3, doi: 10.1109/IWS52775.2021.9499612.

[28] Z. Zhou, Z. Wei, A. Tahir, J. Ren, Y. Yin, G. F. Pedersen and M. Shen, "A high-quality data acquisition method for machine learning-based design and analysis of electromagnetic structures," *IEEE Trans. Microw. Theory Techn.*, early access, 2023, doi: 10.1109/TMTT.2023.3235066.

[29] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Int. Conf. Mach. Learn.*, 2010, pp. 1–8.

[30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Int. Conf. Learn. Repr.*, Y. Bengio and Y. LeCun, Eds., 2015, pp. 1–15.

[31] D. S. Wang, P. Zhao, and C. H. Chan, "Design and analysis of a high-selectivity frequency-selective surface at 60 GHz," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 6, pp. 1694–1703, Jun. 2016, doi: 10.1109/TMTT.2016.2557325.

[32] L. Yang, L. Zhu, R. Zhang, J. Wang, W. Choi, K. Tam, and R. Gómez-García, "Novel multilayered ultra-broadband bandpass filters on high-impedance slotline resonators," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 1, pp. 129–139, Jan. 2019, doi: 10.1109/TMTT.2018.2873330.

[33] N. Calik, F. Gunes, S. Koziel, A. P. Dabrowska, M. A Belen, and P. Mahouti, "Deep-learning-based precise characterization of microwave transistors using fully-automated regression surrogates," *Sci. Rep.*, vol. 13, no. 1, pp. 1–16, Jan. 2023, doi: 10.1038/s41598-023-28639-4.

ISSN (online): 2446-1628
ISBN (online): 978-87-7573-587-7

AALBORG UNIVERSITY PRESS