Computer Networks 257 (2025) 110963

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Survey paper

Physically-consistent EM models-aware RIS-aided communication — A survey

S. Bidabadi [a,*], M.A. Ouameur [a], M. Bagaa [a], D. Massicotte [a], F.D. Figueiredo [b], A. Chaaban [c]

[a] UQTR, Trois-Rivières, QC, Canada
[b] National Institute of Telecommunications – Inatel, Santa Rita do Sapucaí-MG, Brazil
[c] University of British Columbia Okanagan, Kelowna, Canada

ARTICLE INFO

Keywords:
Wireless communication
Reconfigurable intelligent surfaces
Electromagnetic models
Optimization techniques
Machine learning

ABSTRACT

The rapid development of reconfigurable intelligent surfaces (RISs) has sparked transformative advancements in wireless communication systems. These intelligent metasurfaces, adept at dynamically manipulating electromagnetic (EM) waves, hold vast potential for enhancing network capacity, coverage, and efficiency. However, to fully unleash the capabilities of RIS-aided communication systems, effective optimization is crucial. This article provides a recent development of RIS-assisted communication from the viewpoint of physically-consistent EM models. We delve into the realm of physically-consistent EM models, highlighting their pivotal role in achieving robust and efficient RIS designs. Furthermore, this paper offers a survey of the different optimization models utilized for RIS-assisted wireless communication systems, which consider various EM and physical aspects of RIS. We explore solution approaches aimed at optimizing different objectives like sum-rate/spectral efficiency and energy efficiency, spanning traditional optimization models to machine learning-based methods. Additionally, we discuss some open research issues in this field.

1. Introduction

Existing cellular generations will not be able to meet the extraordinary performance requirements needed by innovative applications anticipated in 2030, which will lead to a need for new technologies to enable 6G systems. 6G is expected to outperform 5G in various aspects, including ultra-high data rates, energy efficiency (EE), global coverage, connectivity, high reliability, and low latency, as illustrated in Fig. 1 [1–4]. While technologies like massive multiple-input multiple-output (MIMO), mmWave communication, and network densification have significantly advanced 5G performance, they face challenges such as increased hardware costs and energy consumption. These limitations highlight the need for new wireless technologies to efficiently achieve the key performance indicators (KPIs) of 6G, as listed in Fig. 2.

In this context, reconfigurable intelligent surfaces (RIS) have emerged as a promising solution, offering dynamic control over wireless environments and enhancing overall network performance for 6G. RIS are a natural evolution of earlier intelligent surface (IS) technologies [5]. The evolution of ISs can be traced back to reflectarrays, which used passive antennas for basic wave control. Metamaterials, with their ability to manipulate wave properties, further advanced the field. The concept of RIS emerged, enabling dynamic, programmable control of signal propagation. Subsequent innovations like large intelligent surfaces (LIS) and hybrid architectures expanded IS capabilities. More recently, the beyond-diagonal RIS (BD-RIS) architecture has emerged, offering even more sophisticated signal processing capabilities by encompassing previous IS architectures as special cases. This allows for enhanced control over signal propagation across IS elements. These advancements collectively pave the way for smarter, energy-efficient wireless networks [5–11].

Metasurfaces and patch-array designs represent two key approaches for implementing RIS technology. Metasurfaces consist of continuous, ultra-thin layers with sub-wavelength inclusions that manipulate electromagnetic (EM) waves in ways that natural surfaces cannot. They serve as two-dimensional equivalents of bulky metamaterials, providing similar capabilities to shape wave propagation while being more compact, easier to fabricate, and less lossy. This makes metasurfaces particularly effective for applications in wireless communications, where precise control of wave reflection, refraction, and scattering is crucial for optimizing signal transmission [12–18]. In contrast, many practical implementations of RISs use discrete components, such as patch antennas or PIN diodes, arranged in a grid-like structure. These elements can adjust the phase of impinging EM waves, similar to how reflectarrays operate. However, unlike conventional reflectarrays, which are typically static, RISs are equipped with tunable components

* Corresponding author.
E-mail address: samaneh.bidabadi@uqtr.ca (S. Bidabadi).

https://doi.org/10.1016/j.comnet.2024.110963
Received 25 June 2024; Received in revised form 9 November 2024; Accepted 28 November 2024
Available online 5 December 2024

Table 1
List of abbreviations.

Abbreviation	Description	Abbreviation	Description
ADC	Analog-to-Digital Converters	AI	Artificial Intelligence
ACO	Ant Colony Optimization	ANN	Artificial Neural Networks
AO	Alternating optimization	BC	Block Coordinate Descent
BPSO	Binary Particle Swarm Optimization	CCP	Convex-concave Procedure
CNNs	Convolutional Neural Networks	CR	Cutoff Rate
CSI	Channel State Information	cDC-GAN	Conditional Deep Convolutional Generative Adversarial Network
cDC-VAE	Conditional Deep Convolution Variational Autoencoder	DMT	Diversity-multiplexing tradeoff
DOA	Direction of Arrival	DRL	Deep Reinforcement Learning
DGMs	Deep Generative Models	DOFs	Degrees-of-Freedom
EHU	Energy Harvester User	EH	Energy Harvesting
EMI	Electromagnetic Interference	EMF	Electromagnetic Field Exposure
EM	Electromagnetic	EMOs	EM objects
FDTD	Finite-difference time-domain method	FEM	Finite Element Method
FL	Federated Learning	GA	Genetic Algorithm
IDBP	Information-Directed Branch-and-Prune	IDU	Information Decoder User
LHCP	Left-Handed Circular Polarization	LISs	Large Intelligent Surfaces
MAE	Mean Absolute Error	MEM	Mismatch Environment Model
MIMO	Massive Multiple-Input Multiple-Output	MI	Mutual Information
MISO	Multiple-input Single-output	ML	Machine Learning
MOLACO	Multi-Objective Lazy Ant Colony Optimization	MM	Majorization-Minimization Techniques
MTS	Metamaterial Surface	NOMA	Non-orthogonal Multiple Access
NNs	Neural Networks	OAM	Orbital Angular Momentum
OMA	Orthogonal Multiple Access	OP	Outage Probability
PIDL	Physics Informed Deep Learning	PINNs	Physics Informed Neural Networks
PGM	Projected Gradient Method	PSO	Particle Swarm Optimization
RL	Reinforcement Learning	SCA	Successive Convex Optimization
SAR	Specific Absorption Rate	SDR	Semidefinite Relaxation
SE	Spectral Efficiency	SIMO	Single-input Multiple-output
SISO	Single-input Single-output	SRR	Split-ring Resonator
SVM	Support Vector Machine	STAR	Simultaneous Transmission and Reflection
TEM	True Environment Model	ISAC	Integrated sensing and communication
SIMs	Stacked intelligent metasurfaces		

such as PIN diodes or varactors, enabling dynamic control of EM wave reflection. This software-defined control allows RISs to adapt in real-time to changing network conditions, providing flexibility in signal modulation and improving overall communication performance.

Despite the differences in hardware, both metasurface-based and patch-array-based RISs serve the same overarching purpose: they dynamically control the propagation of EM waves to optimize wireless communication environments. By adapting to varying network conditions and offering real-time configurability, RISs enable efficient signal management in next-generation communication systems, playing a critical role in the evolution toward 6G networks [12,13,19–25].

As the field of RIS-assisted communication systems rapidly evolves, it is essential to address the existing limitations and assumptions that have been prevalent in previous research efforts. For instance, the mutual coupling effects between adjacent RIS elements have been acknowledged as a critical consideration in the accurate modeling of RIS behavior. Likewise, the traditional phase shift resolution models, while providing valuable insights, often fail to capture the intricate nuances of real-world RIS scenarios [26]. Furthermore, the uncontrollable electromagnetic interference in dynamic environments introduces complexities that can significantly impact the performance of RIS-based communication systems. These challenges underscore the need for innovative solutions that can handle these practical constraints and push the boundaries of RIS deployment. Another instance of an often overlooked EM occurrence observed in intelligent surfaces involves what are known as Floquet modes. These are unintended reflections from surfaces that possess a periodic impedance, potentially causing interference at undesired angles and thus undermining system performance [27]. Furthermore, the quest to determine the ultimate theoretical boundaries cannot disregard the need for a precise understanding of the fundamental EM phenomena [28,29].

This comprehensive survey aims to provide an in-depth exploration of the various research dimensions concerning EM-based models of RIS in the context of wireless communication systems. Our focus centers on reviewing the formulation of optimization problems from EM perspectives. We also focus on investigating optimization techniques and EM-aware solution methodologies tailored to effectively harness the capabilities of RIS-assisted networks. By doing so, we strive to bridge the gap between theoretical insights and practical implementations and bring forward insights on EM model-aware RIS assisted communication.

In the pursuit of a comprehensive understanding, this survey draws from a wide range of literature sources and research contributions. It aims to provide researchers, practitioners, and decision-makers with valuable insights into the current state of EM-based RIS models and their implications for optimizing RIS-assisted communication systems.

Motivated by the importance of optimization techniques, in this paper we provide a comprehensive overview of optimization techniques for optimizing RIS-aided wireless communications from EM perspectives, including physics informed neural networks (PINN) approaches. There are several surveys devoted to the theory, design, analyses, and applications of RISs [26,31–38]. However, this work is different from existing surveys and tutorials since it systematically summarizes and analyzes EM-based models of RIS for optimizing RIS-aided communication systems. Specifically, as shown in Fig. 3, this paper focuses on the following aspects:

- First, the paper delves into the EM models of RIS employed in the literature to formulate and optimize EM-aware RIS-assisted communication problems.
- Second, the paper provides an extensive exploration of the various problem formulations related to optimizing RIS-enhanced wireless networks, all examined from the unique perspective of EM models. These formulations cover a wide range of objectives, including the maximization of sum-rate/capacity, energy efficiency, and the minimization of power consumption.
- Third, the paper presents a comprehensive review of model-centric, heuristic, and machine learning (ML) approaches used to optimize RIS-enhanced wireless networks from an EM perspective. The survey emphasizes the need for efficient optimization methods for solving the EM-based problems, notably

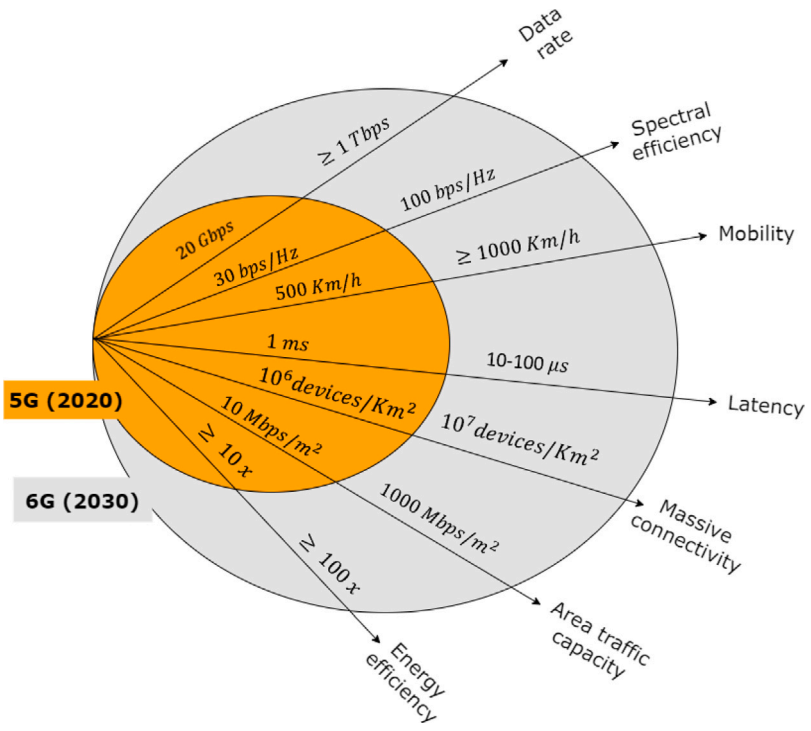

Fig. 1. Specifications for 6G wireless communication (© The Author(s) 2021, [1]).

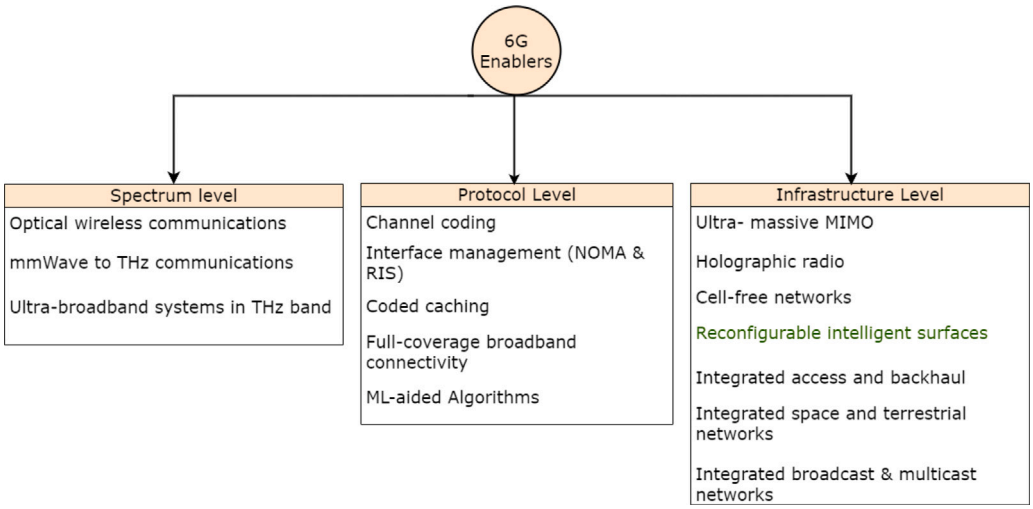

Fig. 2. 6G enablers (© 2021 IEEE, [30]).

highlighting PINN, which combines data-driven insights with governing differential equations to provide more realistic RIS behavior models.

Generally, the survey covers various aspects, including problem formulations, characteristics, and applications of different approaches, aiming to serve as a roadmap for researchers in this field. The paper is structured as follows: Section 2 reviews related work; Section 3 describes EM-based models of RIS; Section 4 presents EM model-aware problem formulations; and Section 5 reviews model-based, heuristic, and ML optimization algorithms used for optimizing RIS-assisted communications using physically consistent EM models of RIS. Section 6 discusses research gaps and identifies future challenges, and finally, Section 7 concludes the survey. The key themes of this paper are depicted in Fig. 3, and a comprehensive compilation of all abbreviations used in this survey can be found in Table 1.

2. Previous survey works and motivations

Numerous surveys have been dedicated to exploring the theory, design, analysis, optimization and practical applications of RISs and in particular meta-surfaces [26,31–34,34,34,36–41] as listed in Table 2. Some surveys delve into the exploration of physically-consistent EM models of RIS, while others examine RIS-aided communication without taking into account the physical models of RIS. This survey considers both perspectives, synthesizing them into a unified framework termed physically-consistent EM models-aware RIS-aided communication. The term "Physically-Consistent EM Models-Aware RIS-Aided Communication" is used to describe the integration of physically-consistent EM models with RIS-aided communication systems. This work aims to support researchers in wireless communication systems by offering more realistic communication models that incorporate the physical characteristics of RIS, which are essential for advancing future wireless

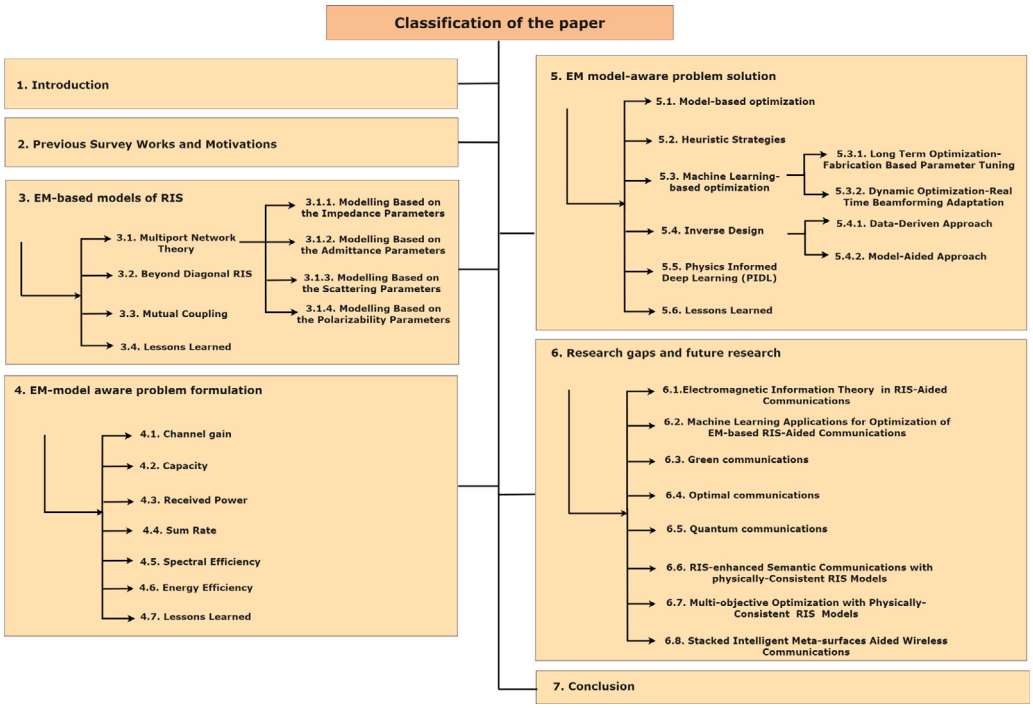

Fig. 3. The organization of this paper.

technologies, including 6G. We also find a gap in the existing literature that only focuses on the inverse design of metamaterials using ML and rarely studies EM-aware RIS-aided communication system optimization using ML techniques.

In [39], the authors investigate the potential applications of RIS technology for 6G communications. They begin with a comprehensive overview of RIS and then explore its various applications, including system and path loss models, asymptotic analysis, precoding techniques, channel state information (CSI) acquisition, prototypes, and its integration with other technologies, without focusing on EM-aware models of RIS. The paper provides practical examples to demonstrate the transformative impact of RIS on future wireless systems and highlights the potential benefits it offers. Furthermore, it identifies key challenges in the deployment and adoption of RIS technology and presents promising directions for future research. Authors in [26] provide a thorough review of RIS-assisted wireless communication, discussing fundamental principles like hardware architecture, control mechanisms, and channel models. They also analyze performance metrics to highlight the benefits of RIS deployment. They mention that, despite the advantages, the integration of RIS faces challenges due to its passive nature, including issues with channel estimation and deployment scenarios. However, the review does not cover physically-consistent EM models-aware RIS. The authors do identify mutual coupling as a significant research challenge and suggest it as a potential area for future exploration in RIS-aided wireless communication systems.

The survey in [31] offers a comprehensive exploration of machine learning algorithms applied to RIS-aided communications. It includes a concise introduction to RISs, a summarization of machine learning techniques used in RIS architecture, and a comparative analysis of these methods to elucidate their integration. The review, however, does not delve into physically-consistent EM Models of RIS in problem formulation and optimization. The authors do highlight that metasurfaces encounter challenges due to the absence of accurate models that consider their EM characteristics, which impacts their functionality beyond simple reflection. They identify this as a major research challenge and propose it as a promising avenue for future exploration in RIS-assisted wireless communication systems.

The work in [32] categorizes optimization techniques into model-based, heuristic, and ML algorithms, covering diverse problem formulations. It introduces model-based algorithms, heuristic optimization, and state-of-the-art ML applications, providing a comparative analysis across stability, robustness, and optimality. The survey also highlights RIS-aided applications in 6G networks and outlines emerging research challenges in this dynamic field, emphasizing RISs' pivotal role in shaping the future of wireless communication systems.

The survey in [33] offers a thorough overview of RISs and examines how reinforcement learning (RL) algorithms are used to optimize RIS parameters. This article specifically focuses on applying deep reinforcement learning (DRL) to improve aspects such as passive beamforming, phase shifting, and RIS placement. It also addresses various challenges in implementing RL algorithms for RIS technology in wireless communications and suggests possible solutions. Nevertheless, the study does not explore physically-consistent EM models of RISs in the problem formulation and restricts the optimization techniques to RL.

The work in [34] delves into the realm of optimizing RIS-aided wireless networks using a combination of heuristic algorithms and ML. It investigates various heuristic techniques, such as greedy algorithms, meta-heuristic algorithms, and matching theory, while also introducing innovative heuristic-ML hybrid approaches. These include heuristic DRL, heuristic-aided supervised learning, and heuristic hierarchical learning. Notably, the study demonstrates that heuristic DRL outperforms conventional DRL in terms of achieving higher data rates and faster convergence. The ultimate aim of the work is to present a novel perspective on optimizing RIS-aided wireless networks, capitalizing on the synergy between heuristic algorithms and ML methods.

The work in [35] discusses the combination of RISs with ML to enhance smart IoT systems, highlighting their use in managing radio propagation without traditional channel data. It examines specific ML techniques like deep reinforcement learning and federated learning, and their application in dynamic environments with challenges like user mobility and fluctuating wireless channels. The paper concludes with future challenges and the potential of RISs in next-generation IoT networks.

The work in [36] offers a comprehensive review of existing optimization techniques and artificial intelligence-driven methods employed to address the constraints associated with RISs. It conducts a

Table 2
Comparison of Related Surveys and Studies. [39][26][31][32][33][34][35][36][37][40][42][41][38].

Ref	Year	Main contributions	Physically-Consistent EM models	RIS-aided communication	Physically-consistent EM models-aware RIS-aided communication
[39]	2022	Provides an overview of RIS for 6G, including applications, system models, and future research directions.	○	●	○
[26]	2023	Comprehensive review of RIS-assisted wireless communication, covering hardware, control mechanisms, channel models, and highlighting challenges like channel estimation and deployment.	○	◐	○
[31]	2022	Comprehensive review of machine learning in RIS-aided communications, highlighting challenges with meta-surfaces due to the lack of accurate EM models and suggesting this as a key area for future research.	◐	●	○
[32]	2023	Explores model-based, heuristic, and ML optimization techniques for RIS-aided communication, provides a comparative analysis, and emphasizes emerging challenges and the critical role of RIS in future 6G networks.	○	●	○
[33]	2023	Examines deep reinforcement learning for optimizing RIS parameters and addresses implementation challenges.	○	◐	○
[34]	2023	Explores optimization of RIS-aided networks using heuristic and ML algorithms, highlighting the advantages of heuristic deep reinforcement learning over conventional DRL, and presents a novel approach combining heuristic and ML methods.	○	◐	○
[35]	2023	Explores RIS and ML integration for smart IoT systems, focusing on techniques like deep reinforcement learning and federated learning , and addresses future challenges in next-generation IoT networks.	◐	◐	○
[36]	2022	Reviews optimization and AI methods for RISs, compares solution quality and complexity, and addresses phase-shift optimization challenges.	○	●	○
[37]	2022	Examines state-of-the-art ML and DL techniques for RIS-enhanced communication, covering operational principles, channel estimation, and various ML applications in RIS-enhanced networks.	○	◐	○
[40]	2023	Provides a comprehensive review of STAR-RIS for 6G networks, covering fundamentals, protocols, applications, and performance metrics.	○	●	○
[42]	2024	Analyzes RIS control mechanisms and their optimization using AI, reviewing current methods, AI enhancements, and the challenges in meeting future network demands.	○	◐	○
[41]	2024	Comprehensive Review on advanced reconfigurable intelligent surfaces (ARIS), focusing on their advancements and applications in 6G networks. Examines various RIS technologies, their benefits, practical implementations, and identifies key research challenges and future directions.	○	◐	○
[38]	2022	Introduces a parametric model and optimization framework for RIS-assisted communications focusing on power maximization and surface impedance.	◐	◐	○
Current Survey	2024	Explores recent advancements in RIS-assisted communication through physically-consistent EM models, reviews optimization models for EM-aware RIS systems, including both traditional and ML-based approaches, and addresses open research issues in the field.	●	●	●

○ Not discussed in the paper.
◐ Mentioned in the paper, but not in detail.
● Discussed in detail in the paper.

comparative analysis, considering factors such as solution quality and computational complexity. Additionally, the paper outlines forthcoming challenges related to phase-shift optimization in the context of RISs and explores potential avenues for resolving these challenges.

Meanwhile, the survey in [37] offers a thorough examination of the current state-of-the-art in ML, with a particular emphasis on DL, within the context of RIS-enhanced communication. The review covers their operational principles, channel estimation (CE) techniques, and the diverse range of ML applications in RIS-enhanced wireless networks. Furthermore, it conducts a systematic survey of existing designs for RIS-enhanced wireless networks.

In [40], the authors deliver a comprehensive review of simultaneously transmitting and reflecting reconfigurable intelligent surface (STAR-RIS), with a particular emphasis on its applications in 6G networks. The paper begins by detailing the fundamentals of STAR-RIS, including its protocols, benefits, and various applications. The review categorizes STAR-RIS techniques across different use case scenarios, such as enhancing network coverage, improving physical layer security (PLS), maximizing sum rate, boosting EE, and reducing interference. It also explores various resource allocation strategies and performance evaluation metrics pertinent to STAR-RIS. However, the focus of this review is primarily on STAR-RIS. While it extensively surveys existing literature in this area, the paper does not delve into specific formulations or detailed solution approaches. Additionally, it does not address the EM-aware aspects of RIS, which is a key focus of our own work.

Another survey in [42] provides an in-depth analysis of RIS, focusing on the technology's control mechanisms and their optimization using artificial intelligence (AI). The paper reviews current RIS control methods, explores how AI can enhance performance, and highlights the limitations of AI-powered RIS control, as well as the significant challenges that must be addressed to meet future network demands.

A thorough examination of active RIS (ARIS) is given in [41], which highlights recent developments and their revolutionary effects on 6G networks. It provides a comparative examination of various schemes and solutions, covers numerous types and modes of RIS, and assesses ARIS applications for network parameter optimization. Additionally, it identifies key research challenges and future directions, highlighting ARIS's crucial role in advancing wireless communication, with a focus away from the physical aspects of RIS modeling and optimization.

Despite the extensive review of the works [26,31–34,34,34,36–41], there remains a gap in addressing the optimization of RIS-aided communications with an emphasis on EM-model-awareness. The work in [38] provides a thorough exploration of three pivotal communication models applied to RISs within wireless communication systems. These models are particularly relevant in scenarios involving in-homogeneous surface impedance boundaries, benefiting from their alignment with Maxwell's equations and consistent electromagnetic behavior, especially under practical approximations like physical optics. Delving deeper, the study navigates the intricacies of designing effective RISs, guided by local and global optimal criteria, and introduces an innovative approximate framework yielding reactive impedance boundaries. Supported by numerical examples, mathematical analysis, and simulations, the study offers a nuanced understanding of the models' strengths and limitations. A significant contribution lies in the introduction of a parametric model and optimization framework tailored for RIS-assisted communications, accommodating scenarios with gradually varying surface impedance while retaining the validity of the physical optics approximation. This model balances solution quality and implementation complexity through electromagnetic-consistent constraints. The study also emphasizes optimization scenarios where surface power efficiency is the objective, capturing optimal designs, corroborated by the power flux as a demonstrative measure of RIS steering capabilities. Placing findings within the context of wireless communications, the study paves the way for future research with streamlined numerical algorithms for electromagnetically consistent optimization problems and considerations for practical constraints in manufacturing-oriented applications. Ultimately, the study bridges theoretical models with tangible wireless communication systems, laying the groundwork for advancements that harmonize theory and application.

While the work in [38] presented a parametric model and an optimization framework for RIS-assisted communications, under the assumptions that the surface impedance changes slowly at the wavelength scale, and the physical optics approximation is valid, it did not extensively explore physically-consistent EM models-aware RIS-aided communication. The existing work focuss solely on the power maximization problem with respect to the surface impedance variable. In contrast, our work will address and comprehensively review EM-aware problems by taking into account both communication and physical variables.

Designing RIS-aided wireless networks requires paying attention to two important aspects. Firstly, it is important to consider EM-based designs where RISs perform fundamental radio wave transformations specified at the EM level. In this paradigm, engineers optimize RIS parameters to enhance network performance. Secondly, it is important to consider communication-based designs with focus on optimizing RISs to achieve specific communication goals, which must necessarily align with EM-based manipulations. In this paradigm, optimization is used to make RISs versatile for various performance metrics such as channel capacity and modulation schemes [12]. Compared to the surveys in Table 2, this comprehensive survey is dedicated to conducting a detailed exploration of the multifaceted research dimensions associated with RIS-aided networks while taking into account both aforementioned considerations. Our primary focus revolves around formulating optimization challenges concerning various objectives and constraints, including sum rate, received power, spectral efficiency (SE), and EE, all from an EM perspective. We also delve into the examination of the optimization techniques and methodologies applied to address these optimization challenges.

The exploration of RISs in wireless networks encompasses a diverse research landscape, spanning disciplines like communication theory, computer science, and electromagnetism. However, a significant drawback in the current state of RIS research lies in the absence of accurate models that effectively encapsulate the intricate EM characteristics of these adaptive meta-surfaces [43]. In the upcoming era, defined by the existence of reconfigurable EM environments, the primary goal will be to enhance system performance, potentially in real-time. This enhancement will be achieved by adjusting the parameters of EM objects (EMOs), such as the reflective properties of metasurface-based RISs. To achieve this goal, a holistic strategy is required, integrating physically-consistent models and design utilities that amalgamate signal processing with electromagnetic theory [44].

While existing studies often adopt a simplistic view of RISs as perfect reflectors, this oversimplification sidesteps crucial factors, such as incident angles, polarization, and material properties. To address these limitations and cultivate more realistic performance predictions, it is imperative to embrace physics and EM-compliant models [43].

In the pursuit of an accurate portrayal and optimization of RIS behavior, the incorporation of EM models proves indispensable. These models furnish a mathematical framework to decipher the intricate interplay between electromagnetic waves and the reconfigurable elements constituting the RIS. Physically-consistent EM models-aware RIS-aided communication approach, particularly demonstrated in the formulation of RIS-aided communication systems, offers significant advantages. These EM-aware models provide an accurate representation of the electromagnetic interactions between RIS elements and incoming waves, enabling more realistic predictions of RIS effects on signal propagation, reflection, and refraction, which is crucial for optimizing network performance. By incorporating these detailed aspects, the models enable precise tuning of RIS parameters, such as phase shifts and beamforming settings, resulting in improved performance metrics. This approach minimizes discrepancies between theoretical predictions and actual performance by accounting for real-world complexities. Additionally, EM models that integrate these concepts facilitate simulations of RIS-aided networks under realistic conditions, including environmental variations and complex wave interactions, which aids in assessing the feasibility of proposed solutions and making informed decisions about network deployment and configuration.

3. EM-based models of RIS

RIS are frequently reduced to perfect scatterers in optimization and signal processing designs for wireless communication systems. Nevertheless, because they ignore important details like phase–amplitude correlation in reflection coefficients and mutual coupling between scattering elements, these simplified models lack EM consistency. Incorporating realistic reradiation models that address the interaction between element-level scattering design and surface-level optimization is crucial, according to recent studies. Consequently, efforts are underway to develop electromagnetically consistent RIS models and efficient optimization algorithms. Traditionally, RIS behavior is analyzed using three main EM modeling approaches:

- Scattering (S-parameter) Model: This model describes how RIS elements scatter incoming signals, providing insights into the reflected signal's characteristics.
- Impedance (Z-parameter) Model: This model characterizes the impedance of the RIS elements, allowing analysis of how the elements interact with the incident signal in terms of impedance.

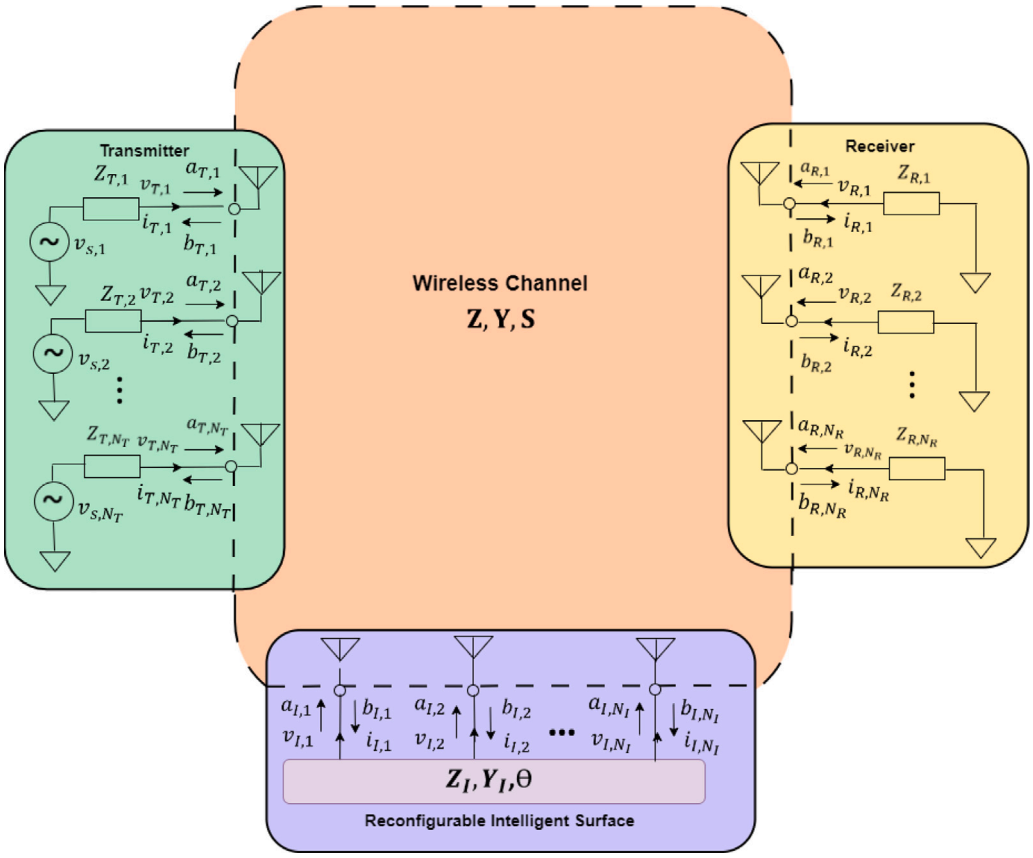

Fig. 4. Multiport representation of an RIS-assisted communication system (© 2024 IEEE, [46]).

- Admittance (Y-parameter) Model: This model describes the admittance characteristics of the RIS elements, offering an alternative perspective on how the elements interact with the incident signal.

All three models are ultimately derived from Maxwell's equations and provide equivalent descriptions of the RIS system.

Recently, another approach based on the polarizabilities of RIS elements has been proposed. The polarizability-based model offers simpler and more intuitive mathematical expressions, aligning better with the localized nature of RIS element reconfigurability. This approach complements the traditional models by providing a more flexible framework for RIS system analysis and design, particularly in capturing the interaction between RIS elements and electromagnetic waves.

3.1. Multiport network theory

Multiport network theory is emerging as a promising approach for modeling and optimizing RIS-aided channels due to its ability to capture electromagnetic coupling between closely spaced radiating elements, which is crucial for achieving advanced wave transformations with high power efficiency [45,46].

Consider a MIMO communication setup enhanced by an RIS, featuring N_T antennas at the transmitter, N_R antennas at the receiver, and N_I antennas at the RIS, as depicted in Fig. 4. The wireless channel is modeled as an N-port network, where $N = N_T + N_I + N_R$.

3.1.1. Modeling based on the impedance parameters

The wireless channel can be characterized by its impedance matrix $\mathbf{Z} \in \mathbb{C}^{N\times N}$, with the relationship:

$$\mathbf{v} = \mathbf{Z}\mathbf{i}, \tag{1}$$

where $\mathbf{v} \in \mathbb{C}^{N\times 1}$ and $\mathbf{i} \in \mathbb{C}^{N\times 1}$ are the voltages and currents at the N ports, respectively. We partition $\mathbf{v}$, $\mathbf{i}$, and $\mathbf{Z}$ as:

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}_T \\ \mathbf{v}_I \\ \mathbf{v}_R \end{bmatrix}, \quad \mathbf{i} = \begin{bmatrix} \mathbf{i}_T \\ \mathbf{i}_I \\ \mathbf{i}_R \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_{TT} & \mathbf{Z}_{TI} & \mathbf{Z}_{TR} \\ \mathbf{Z}_{IT} & \mathbf{Z}_{II} & \mathbf{Z}_{IR} \\ \mathbf{Z}_{RT} & \mathbf{Z}_{RI} & \mathbf{Z}_{RR} \end{bmatrix}, \tag{2}$$

where $\mathbf{v}_i \in \mathbb{C}^{N_i\times 1}$ and $\mathbf{i}_i \in \mathbb{C}^{N_i\times 1}$ for $i \in \{T, I, R\}$ denote the voltages and currents at the antennas of the transmitter, RIS, and receiver, respectively. Besides, $\mathbf{Z}_{TT} \in \mathbb{C}^{N_T\times N_T}$ is the impedance matrix for the transmitting antennas, with off-diagonal entries indicating mutual coupling among the antennas. $\mathbf{Z}_{II} \in \mathbb{C}^{N_I\times N_I}$ represents the impedance matrix for the RIS elements, where the off-diagonal terms reflect mutual coupling between different RIS elements. $\mathbf{Z}_{RR} \in \mathbb{C}^{N_R\times N_R}$ denotes the impedance matrix for the receiving antennas, capturing mutual coupling among these antennas. The transmission impedance matrices include $\mathbf{Z}_{TI} \in \mathbb{C}^{N_I\times N_T}$, which describes the interaction from transmitting antennas to the RIS; $\mathbf{Z}_{IR} \in \mathbb{C}^{N_R\times N_I}$, representing the interaction from the RIS to the receiving antennas; and $\mathbf{Z}_{RT} \in \mathbb{C}^{N_R\times N_T}$, which covers the interaction from receiving antennas to transmitting antennas. For a reciprocal wireless channel:

$$\mathbf{Z}_{TR} = \mathbf{Z}_{RT}^T, \quad \mathbf{Z}_{TI} = \mathbf{Z}_{IT}^T, \quad \mathbf{Z}_{IR} = \mathbf{Z}_{RI}^T. \tag{3}$$

At the transmitter, each antenna n_T is connected in series with a source voltage $\mathbf{v}_{s,n_T}$ and a source impedance $\mathbf{Z}_{T,n_T}$, where:

$$\mathbf{v}_T = \mathbf{v}_{s,T} - \mathbf{Z}_T\mathbf{i}_T, \tag{4}$$

with $\mathbf{v}_{s,T}$ as the source voltage vector and $\mathbf{Z}_T$ as a diagonal matrix.

At the RIS, the N_I scattering elements are connected to an N_I-port reconfigurable impedance network, related by:

$$\mathbf{v}_I = -\mathbf{Z}_I\mathbf{i}_I, \tag{5}$$

where $\mathbf{Z}_I$ is the impedance matrix of the reconfigurable network.

At the receiver, each antenna n_R is connected in series with a load impedance $\mathbf{Z}_{R,n_R}$, related by:

$$\mathbf{v}_R = -\mathbf{Z}_R \mathbf{i}_R, \tag{6}$$

where $\mathbf{Z}_R$ is a diagonal matrix.

3.1.2. Modeling based on the admittance parameters

The channel can also be characterized by its admittance matrix $\mathbf{Y} \in \mathbb{C}^{N\times N}$, where:

$$\mathbf{i} = \mathbf{Y}\mathbf{v}. \tag{7}$$

The admittance matrix $\mathbf{Y}$ is the inverse of $\mathbf{Z}$:

$$\mathbf{Y} = \mathbf{Z}^{-1}. \tag{8}$$

Similar to $\mathbf{Z}$, $\mathbf{Y}$ can be partitioned as:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_{TT} & \mathbf{Y}_{TI} & \mathbf{Y}_{TR} \\ \mathbf{Y}_{IT} & \mathbf{Y}_{II} & \mathbf{Y}_{IR} \\ \mathbf{Y}_{RT} & \mathbf{Y}_{RI} & \mathbf{Y}_{RR} \end{bmatrix}, \tag{9}$$

where $\mathbf{Y}_{ij} \in \mathbb{C}^{N_i \times N_j}$ for $i, j \in \{T, I, R\}$.

At the transmitter, receiver, and RIS, the currents $\mathbf{i}_T$, $\mathbf{i}_I$, and $\mathbf{i}_R$ and voltages $\mathbf{v}_T$, $\mathbf{v}_I$, and $\mathbf{v}_R$ are related by:

$$\mathbf{i}_T = \mathbf{i}_{s,T} - \mathbf{Y}_T \mathbf{v}_T, \quad \mathbf{i}_I = -\mathbf{Y}_I \mathbf{v}_I, \quad \mathbf{i}_R = -\mathbf{Y}_R \mathbf{v}_R, \tag{10}$$

where the source current vector $\mathbf{i}_{s,T}$ and diagonal matrix $\mathbf{Y}_T$ relate currents $\mathbf{i}_T$ and voltages $\mathbf{v}_T$ at the transmitter; the admittance matrix $\mathbf{Y}_I$ relates currents $\mathbf{i}_I$ and voltages $\mathbf{v}_I$ at the RIS; and the diagonal matrix $\mathbf{Y}_R$ relates currents $\mathbf{i}_R$ and voltages $\mathbf{v}_R$ at the receiver. Besides, the diagonal admittance matrices are related to the impedance matrices as $\mathbf{Y}_T = \mathbf{Z}_T^{-1}$, $\mathbf{Y}_I = \mathbf{Z}_I^{-1}$, and $\mathbf{Y}_R = \mathbf{Z}_R^{-1}$.

3.1.3. Modeling based on the scattering parameters

The channel can be characterized by its scattering matrix $\mathbf{S} \in \mathbb{C}^{N\times N}$, with the relationship:

$$\mathbf{b} = \mathbf{S}\mathbf{a}, \tag{11}$$

where $\mathbf{a} \in \mathbb{C}^{N\times 1}$ and $\mathbf{b} \in \mathbb{C}^{N\times 1}$ are the incident and reflected waves, respectively. The scattering matrix $\mathbf{S}$ can be partitioned as:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{TT} & \mathbf{S}_{TI} & \mathbf{S}_{TR} \\ \mathbf{S}_{IT} & \mathbf{S}_{II} & \mathbf{S}_{IR} \\ \mathbf{S}_{RT} & \mathbf{S}_{RI} & \mathbf{S}_{RR} \end{bmatrix}, \tag{12}$$

where $\mathbf{S}_{ij} \in \mathbb{C}^{N_i \times N_j}$ for $i, j \in \{T, I, R\}$.

The vectors $\mathbf{v}$ and $\mathbf{i}$ are related to $\mathbf{a}$ and $\mathbf{b}$ by:

$$\mathbf{v} = \mathbf{a} + \mathbf{b}, \quad \mathbf{i} = \frac{\mathbf{a} - \mathbf{b}}{Z_0} = Y_0(\mathbf{a} - \mathbf{b}), \tag{13}$$

where Z_0 is the characteristic impedance, and $Y_0 = \frac{1}{Z_0}$ is the characteristic admittance. The matrix $\mathbf{S}$ can be expressed in terms of $\mathbf{Z}$ as:

$$\mathbf{S} = (\mathbf{Z} + Z_0 I)^{-1}(\mathbf{Z} - Z_0 I). \tag{14}$$

At the transmitter, receiver, and RIS, the incident and reflected waves are related by:

$$\mathbf{a}_T = \mathbf{b}_{s,T} + \Gamma_T \mathbf{b}_T, \quad \mathbf{a}_I = \Theta \mathbf{b}_I, \mathbf{a}_R = \Gamma_R \mathbf{b}_R, \tag{15}$$

where $\mathbf{b}_{s,T}$ represents the source wave vector, Γ_T denotes the reflection coefficient matrix at the transmitter, Θ is the scattering matrix of the reconfigurable network at the RIS, and Γ_R is the reflection coefficient matrix at the receiver.

The summary of each model is presented in Table 3[45–49]. In the analysis of channel models utilizing Z, Y, and S parameters, several simplifications are applied under the assumptions of perfect matching and the absence of mutual coupling. Specifically, for the Z-parameters, the channel model simplifies to:

$$h = \frac{1}{2Z_0}\left[Z_{RT} - Z_{RI}(Z_I + Z_0 I)^{-1} Z_{IT}\right]. \tag{16}$$

For the Y-parameters, the model simplifies to:

$$h = \frac{1}{2Y_0}\left[-Y_{RT} + Y_{RI}(Y_I + Y_0 I)^{-1} Y_{IT}\right]. \tag{17}$$

For the S-parameters, the model is given by:

$$h = S_{RT} + S_{RI}(I - \Theta S_{II})^{-1}\Theta S_{IT}, \tag{18}$$

with further simplifications under the assumption of perfect matching at the RIS. The mappings between these parameter sets demonstrate that the simplified models are equivalent, confirming consistency across different parameter analyses [46,50].

3.1.4. Modeling based on the polarizability parameters

The majority of current signal-processing work makes the assumption that wireless channels and RIS configuration follow a linear relationship. But Maxwell's equations are frequently in conflict with this premise. A recently proposed physics-compliant model for RIS-parametrized channels focuses on expressing the reconfigurability of RIS elements through adjustable polarizability [51]. In this model, RIS elements and antennas are modeled as dipoles with tunable polarizabilities. An interaction matrix $\mathbf{M} \in \mathbb{C}^{N\times N}$ captures the local scattering characteristics of these dipoles along its diagonal and their interactions off-diagonal:

$$M_{i,j} = \begin{cases} \alpha_i^{-1} & \text{if } i = j \\ -G_{ij} & \text{if } i \neq j, \end{cases} \tag{19}$$

where $N = N_T + N_R + N_S$, with N_T transmitting antennas, N_R receiving antennas, and N_S RIS elements in the considered system model. Here, α_i denotes the polarizability of the ith dipole and G_{ij} is the Green's function, representing the electric field at the ith dipole due to a unit dipole moment at the jth dipole. The sub-indices T, R, A, and S denote the sets of indices associated with transmitting antennas, receiving antennas, all antennas, and RIS elements, respectively [52–56].

The vector $\mathbf{d}$ is defined as the vector of inverse polarizabilities for the RIS dipoles:

$$\mathbf{d} = \mathrm{diag}(M_{SS}), \tag{20}$$

where $\mathbf{d}$ includes the diagonal elements of the matrix M_{SS}.

Assuming identical transmitting and receiving antennas, the channel matrix $\mathbf{h}$ is proportional to the RT block of the inverse of the interaction matrix:

$$\mathbf{h} \propto [\mathbf{M}]_{RT}^{-1}. \tag{21}$$

To compute this efficiently, the reduced interaction matrix $\tilde{\mathbf{M}}$ is defined as:

$$\tilde{\mathbf{M}} = \mathbf{M}_{AA} - \mathbf{M}_{AS}\left(\mathbf{M}_{SS}\right)^{-1}\mathbf{M}_{SA}. \tag{22}$$

Then, the RT block of the inverse of $\tilde{\mathbf{M}}$ is given by:

$$\mathbf{H} \propto \left[\tilde{\mathbf{M}}^{-1}\right]_{RT} = -\left[\tilde{\mathbf{M}}_{RR} - \tilde{\mathbf{M}}_{RT}\tilde{\mathbf{M}}_{TT}^{-1}\tilde{\mathbf{M}}_{TR}\right]^{-1}\tilde{\mathbf{M}}_{RT}\tilde{\mathbf{M}}_{TT}^{-1}. \tag{23}$$

3.2. Beyond diagonal RIS

Traditionally, RISs have been implemented by adjusting each RIS element through a tunable load, resulting in a diagonal phase shift matrix with limited flexibility. To address this constraint, the concept of BD-RIS has recently emerged [57]. The key innovation of BD-RIS lies in the integration of tunable impedance components that interconnect the RIS elements, enhancing its flexibility.

The classification of existing BD-RIS systems, which focuses on the characteristics of the BD-RIS matrix, the supported modes, and the architectural designs, has been extensively explored in various studies conducted by Hongyu et al. [46,57–59].

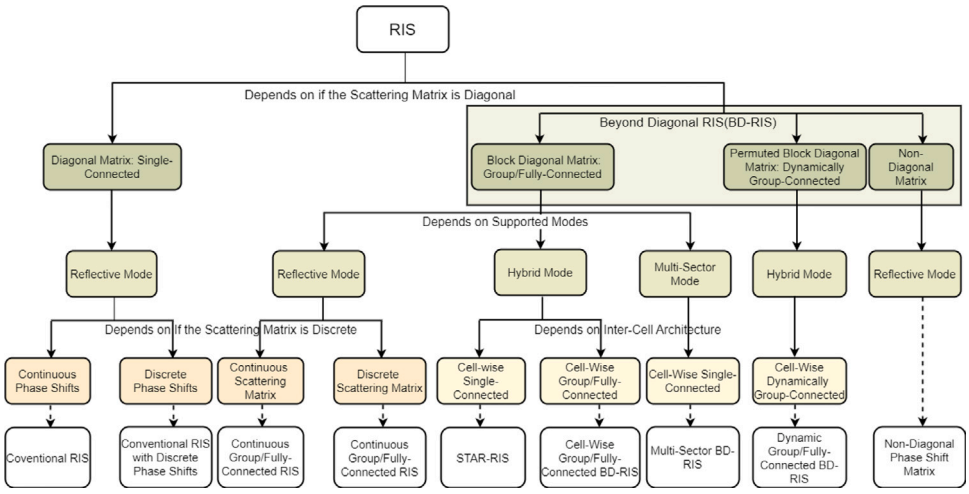

Fig. 5. RIS classification tree (© 2023 IEEE, [58]).

The three-layer classification scheme for BD-RIS systems, as described in [57], is depicted in Fig. 5. Each layer is explained in detail below.

First Layer: Characteristics of the Scattering Matrix Φ

1. **Diagonal Matrix-Single-Connected** The single connected RIS architecture, commonly found in the literature [13,60,61], follows a conventional approach. In this setup, each port of the reconfigurable impedance network connects to ground through a reconfigurable impedance and remains disconnected from the other ports.
2. **Block Diagonal Matrix:** This category involves dividing M antennas into G groups. Antennas within the same group are interconnected, but those across groups are not. This setup, known as group-connected RIS [57], employs a scattering matrix Φ structured as a block diagonal matrix. Each block is unitary, allowing manipulation of both phase and magnitude of incident waves. This configuration outperforms conventional RIS 1.0. When all antennas are connected (G = 1), it is termed fully connected RIS [57], resulting in a unitary scattering matrix. Single-connected RIS is a special case with M groups, having a diagonal scattering matrix [62].
3. **Permuted Block Diagonal Matrix:** Here, the grouping strategy adapts to channel state information (CSI), creating a permuted block diagonal matrix [59]. This flexibility enhances beam control compared to fixed group-connected RIS [63].
4. **Non-Diagonal Matrix:** Antennas are paired via phase shifters, causing signals on one antenna to purely reflect from another. This yields an asymmetric non-diagonal scattering matrix [46], offering higher power gain than conventional RIS 1.0 [64].

Second Layer: Supported Modes

1. **Reflective Mode:** Signals on one side of the RIS reflect toward the same side, covering half-space. Mathematically, it is characterized by a unitary constraint matrix Φ.
2. **Hybrid Mode:** In hybrid mode, the RIS allows signals to be partially reflected and transmitted to opposite sides, providing full-space coverage. This configuration, known as STAR-RIS or intelligent omni-surface (IOS), uses back-to-back antennas with unidirectional patterns, connected to a 2-port reconfigurable impedance network, to cover the entire space. Mathematically, the RIS operating in hybrid mode is described by two matrices, $\mathbf{\Phi}_{\text{ref}} \in \mathbb{C}^{M/2\times M/2}$ and $\mathbf{\Phi}_{\text{trns}} \in \mathbb{C}^{M/2\times M/2}$, satisfying the condition $\mathbf{\Phi}_H^{\text{ref}}\mathbf{\Phi}_{\text{ref}} + \mathbf{\Phi}_H^{\text{trns}}\mathbf{\Phi}_{\text{trns}} = I_{M/2}$ [59,65–69].
3. **Multi-Sector Mode:** This mode divides the full space into K sectors ($K \geq 2$), allowing signals impinging on one sector of the RIS to be partially reflected toward the same sector and partially scattered toward the other $K-1$ sectors . Each cell contains K antennas placed at each edge of an K-sided polygon, with each antenna having a unidirectional radiation pattern covering $1/K$ to avoid overlap among sectors. The K antennas are connected to an K-port fully-connected reconfigurable impedance network, resulting in coverage of the full space while providing higher performance gains than the hybrid mode. Mathematically, the RIS with multi-sector mode is characterized by K matrices, $\mathbf{\Phi}_k \in \mathbb{C}^{M/K\times M/K}$, for $k = 1,\ldots,K$, which satisfy $\sum_{k=1}^{K} \mathbf{\Phi}_k^H \mathbf{\Phi}_k = I_M$ [57,66,70].

Third Layer: Inter-Cell Architecture

This layer considers how cells are connected in BD-RIS with hybrid/multi-sector modes, resulting in single/group/fully connected architectures, affecting the resulting scattering matrices [57].

The authors in [58] also explored designs of single, group, and fully-connected RIS with discrete values. Their findings indicate that while four resolution bits are required in single-connected RIS, only one resolution bit suffices in fully-connected RIS. This simplification significantly streamlines the future development of these promising RIS architectures.

3.3. Mutual coupling

Mutual coupling-aware models are crucial for accurately simulating and understanding the interactions between elements in a RIS. These models enable precise predictions and optimizations of RIS system performance across different communication scenarios [71]. The primary models for considering mutual coupling in communications are generally split into two types: those that use impedance matrices and those that rely on scattering matrices.

The study in [72] examines a mutual coupling-aware communication model assisted by BD-RIS, utilizing analyses of scattering and impedance parameters. The proposed general RIS-aided channel, denoted as $\mathbf{h}_{\text{TR}} \in \mathbb{C}^{K\times N}$, which describes the relationship between the voltage at receiver ports and transmitter ports, is expressed as follows:

$$\mathbf{h}_{\text{TR}} = (\mathbf{\Gamma}_R + \mathbf{I}_K)^{-1}\mathbf{T}_{\text{RT}}(\mathbf{I}_N + \mathbf{\Gamma}_T\mathbf{T}_{\text{TT}} + \mathbf{T}_{\text{TT}})^{-1} \tag{24}$$

where $\mathbf{T}_{\text{TT}} \in \mathbb{C}^{N\times N}$ and $\mathbf{T}_{\text{RT}} \in \mathbb{C}^{K\times N}$ are submatrices of $\mathbf{T} = \mathbf{S}(\mathbf{I}_L - \mathbf{\Gamma}\mathbf{S})^{-1} \in \mathbb{C}^{L\times L}$ with $\mathbf{\Gamma} = \text{blkdiag}(\mathbf{\Gamma}_T, \mathbf{\Theta}, \mathbf{\Gamma}_R)$. Specifically, $\mathbf{T}_{\text{TT}} = [\mathbf{T}]_{1:N,1:N}$ and $\mathbf{T}_{\text{RT}} = [\mathbf{T}]_{N+M+1:L,1:N}$. Besides, $\mathbf{\Gamma}_T$ and $\mathbf{\Gamma}_R$ represents reflection coefficient matrix at the transmitter and receiver respectively.

Table 3
N-port network RIS-aided MIMO Communications using impedance, admittance, or scattering matrix [46].

Impedance parameters-based model	Scattering parameters-based model	Admittance parameters-based model
$\begin{bmatrix} \mathbf{v}_T \\ \mathbf{v}_I \\ \mathbf{v}_R \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{TT} & \mathbf{Z}_{TI} & \mathbf{Z}_{TR} \\ \mathbf{Z}_{IT} & \mathbf{Z}_{II} & \mathbf{Z}_{IR} \\ \mathbf{Z}_{RT} & \mathbf{Z}_{RI} & \mathbf{Z}_{RR} \end{bmatrix} \begin{bmatrix} \mathbf{i}_T \\ \mathbf{i}_I \\ \mathbf{i}_R \end{bmatrix}$	$\begin{bmatrix} \mathbf{b}_T \\ \mathbf{b}_I \\ \mathbf{b}_R \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{TT} & \mathbf{S}_{TI} & \mathbf{S}_{TR} \\ \mathbf{S}_{IT} & \mathbf{S}_{II} & \mathbf{S}_{IR} \\ \mathbf{S}_{RT} & \mathbf{S}_{RI} & \mathbf{S}_{RR} \end{bmatrix} \begin{bmatrix} \mathbf{a}_T \\ \mathbf{a}_I \\ \mathbf{a}_R \end{bmatrix}$	$\begin{bmatrix} \mathbf{Y}_{TT} & \mathbf{Y}_{TI} & \mathbf{Y}_{TR} \\ \mathbf{Y}_{IT} & \mathbf{Y}_{II} & \mathbf{Y}_{IR} \\ \mathbf{Y}_{RT} & \mathbf{Y}_{RI} & \mathbf{Y}_{RR} \end{bmatrix}$
$\mathbf{v}_T = \mathbf{v}_{s,I} - \mathbf{Z}_T\mathbf{i}_T$	$\mathbf{a}_T = \mathbf{b}_{s,I} + \mathbf{\Gamma}_T\mathbf{b}_T$	$\mathbf{i}_T = \mathbf{i}_{s,T} - \mathbf{Y}_T\mathbf{v}_T$
$\mathbf{v}_I = -\mathbf{Z}_I\mathbf{i}_I$	$\mathbf{a}_I = \mathbf{\Theta}\mathbf{b}_I$	$\mathbf{i}_I = -\mathbf{Y}_I\mathbf{v}_I$
$\mathbf{v}_R = -\mathbf{Z}_R\mathbf{i}_R$	$\mathbf{a}_R = \mathbf{\Gamma}_R\mathbf{b}_R$	$\mathbf{i}_R = -\mathbf{Y}_R\mathbf{v}_R$
$\mathbf{Z}_T = \mathrm{diag}(Z_{T,1}, \ldots, Z_{T,NT})$	$\mathbf{S} = (\mathbf{Z} + \mathbf{Z}_0\mathbf{I})^{-1}(\mathbf{Z} - \mathbf{Z}_0\mathbf{I})$	
$\mathbf{Z}_R = \mathrm{diag}(Z_{R,1}, \ldots, Z_{R,NR})$	$\mathbf{\Gamma}_T = (\mathbf{Z}_T + \mathbf{Z}_0\mathbf{I})^{-1}(\mathbf{Z}_T - \mathbf{Z}_0\mathbf{I})$	
	$\mathbf{\Theta} = (\mathbf{Z}_I + \mathbf{Z}_0\mathbf{I})^{-1}(\mathbf{Z}_I - \mathbf{Z}_0\mathbf{I})$	
	$\mathbf{\Gamma}_R = (\mathbf{Z}_R + \mathbf{Z}_0\mathbf{I})^{-1}(\mathbf{Z}_R - \mathbf{Z}_0\mathbf{I})$	

The proposed channel model, outlined in Eq. (24), effectively encompasses the impacts of antenna mismatching and mutual coupling at the transmitter, BD-RIS, and receiver. Nonetheless, its complexity makes it challenging to clearly understand the function of BD-RIS. Then, to streamline the overall communication model, they introduce a simplifying set of assumptions. Firstly, they assume equal source and load impedances ($\mathbf{Z}_T = Z_0\mathbf{I}_N, \mathbf{Z}_R = Z_0\mathbf{I}_K$), ensuring optimal power matching at the transmitter and receiver. Secondly, they make the unilateral approximation, assuming negligible links between devices due to sufficiently large distances ($\mathbf{Z}_{TI} \approx 0, \quad \mathbf{Z}_{TR} \approx 0, \quad \mathbf{Z}_{IR} \approx 0$). Thirdly, they posit perfectly matched antennas at both ends, minimizing mutual coupling effects ($\mathbf{Z}_{TT} = Z_0\mathbf{I}_N$ and $\mathbf{Z}_{RR} = Z_0\mathbf{I}_K$)[72].

Based on the above assumptions, and the relationship $\mathbf{S} = (\mathbf{Z} + Z_0\mathbf{I}_L)^{-1}(\mathbf{Z} - Z_0\mathbf{I}_L)$, the transmission scattering matrices from the receiver to the transmitter approach zero, i.e., $\mathbf{S}_{\mathrm{TI}} \approx 0, \quad \mathbf{S}_{\mathrm{TR}} \approx 0, \quad \mathbf{S}_{IR} \approx 0$. Additionally, the scattering matrices at the antenna arrays of the transmitter and receiver also approach zero, i.e., ($\mathbf{S}_{\mathrm{TT}} \approx 0, \quad \mathbf{S}_{\mathrm{RR}} \approx 0$) Consequently, the following simplified channel is resulted:

$$\mathbf{h}_{\mathrm{TR}} \approx \mathbf{S}_{\mathrm{RT}} + \mathbf{S}_{\mathrm{RI}}(\mathbf{I}_M - \mathbf{\Theta}\mathbf{S}_{\mathrm{II}})^{-1}\mathbf{\Theta}\mathbf{S}_{\mathrm{IT}} \tag{25}$$

As the matrix $\mathbf{\Theta}$ of reflection coefficients introduces complexity by appearing both inside and outside the inverse, further simplifications by leveraging the relationship between nonzero sub-matrices of $\mathbf{S}$ and those of $\mathbf{Z}$ are introduced. So, a more tractable and convenient expression equivalent to is suggested as:

$$\mathbf{h}_{\mathrm{TR}} \approx \frac{1}{2Z_0}(\mathbf{Z}_{\mathrm{RT}} - \mathbf{Z}_{\mathrm{RI}}(\mathbf{Z}_{\mathrm{II}} + \mathbf{Z}_{\mathrm{I}})^{-1}\mathbf{Z}_{\mathrm{IT}}) \tag{26}$$

The matrices $\mathbf{Z}_{\mathrm{RT}}$, $\mathbf{Z}_{\mathrm{RI}}$, and $\mathbf{Z}_{\mathrm{IT}}$ describe the channels from the transmitter to the receiver, from the BD-RIS to the receiver, and from the transmitter to the BD-RIS, respectively. The matrix $\mathbf{Z}_{\mathrm{II}}$ addresses the mismatching and mutual coupling effects at the RIS elements, where its diagonal entries represent the self-impedance of the elements and the off-diagonal entries account for mutual coupling. This coupling is influenced by the spacing between the elements, with wider separations typically leading to less mutual coupling.

Several other researchers have investigated mutual coupling-aware RIS-aided communications [48,73].

3.4. Lessons learned

- The EM effects are commonly characterized using S parameters and/or Z parameters in communication systems.
- The introduction of BD-RIS has transformed the conventional approach to RIS implementation by incorporating tunable impedance components for interconnecting RIS elements. This innovation enhances the flexibility of RIS configurations beyond the limitations imposed by diagonal phase shift matrices, marking a significant departure from traditional RIS setups.
- Understanding and modeling mutual coupling effects between RIS elements is crucial in optimizing RIS performance and overall system efficiency, especially in scenarios where the proximity of elements leads to mutual interference and complex electromagnetic interactions.

4. EM model-aware problem formulation

This section delves into diverse problem formulations pertaining to the optimization of RIS-enhanced wireless networks. The emphasis is distinctly on EM factors and considerations. These formulations cover a broad spectrum of objectives, including the improvement of channel gain and capacity, enhancement of SINR and maximization of sum rate, enhancement of energy efficiency, and minimization of power consumption.

4.1. Channel gain

Maximization of channel gain in BD-RIS aided communications with considering the mutual coupling effects between RIS elements is studied in [72]. A SISO system with focuses on the optimization of block diagonal impedance matrix $\mathbf{Z}_I$ of group-connected architecture is considered. The channel gain maximization problem considering the channel model in Eq. (26) is formulated as follows:

$$\max_{\mathbf{Z}_I} \quad \left|\mathbf{Z}_{\mathrm{RT}} - \mathbf{Z}_{\mathrm{RI}}(\mathbf{Z}_{\mathrm{II}} + \mathbf{Z}_I)^{-1}\mathbf{Z}_{\mathrm{IT}}\right|^2 \tag{27a}$$

subject to

$$\mathbf{Z}_I = \mathrm{blkdiag}(\mathbf{Z}_{I,1}, \ldots, \mathbf{Z}_{I,G}), \tag{27b}$$

$$\mathbf{Z}_{I,g} = \mathbf{Z}_{I,g}^T, \quad \Re\{\mathbf{Z}_{I,g}\} = 0, \quad \forall g \in \mathcal{G}. \tag{27c}$$

In this context, each block $\mathbf{Z}_{I,g} \in \mathbb{C}^{M' \times M'}$, for all $g \in \mathcal{G} = \{1, \ldots, G\}$, is characterized by being symmetric and purely imaginary in the context of reciprocal and lossless reconfigurable impedance networks . In particular, when $G = 1$ and $G = M$, it corresponds to the fully-connected and single-connected architectures of BD-RIS, respectively [72]. An iterative optimization technique, incorporating the Neumann series approximation is employed to tackle the computational hurdle arising from matrix inversion within the objective function mentioned in Eq. (27a).

4.2. Capacity

RIS's dynamic control over the propagation environment enables it to improve link quality, extend coverage, and ultimately enhance the overall capacity of wireless communication systems.

The research presented in [74] addresses the gap concerning wideband communications with BD-RIS, a topic previously explored only in [75]. Prior studies assumed linear variations of RIS phase-shifts across subcarriers and proposed a quasi-Newton method to optimize channel capacity, lacking a derived system model and a problem-specific algorithm. This paper fills these gaps by deriving the OFDM system model from fundamental principles, resulting in a unique expression. Additionally, it introduces a specialized algorithm to optimize wideband capacity under specific constraints. The resulting capacity expression provides detailed insights into the impact of propagation paths on cascaded channel expressions. Simulation results demonstrate that employing BD-RIS significantly enhances wideband capacity, particularly in non-line-of-sight (NLOS) channels.

The capacity of the BD-RIS-assisted SISO OFDM system, given a reflection matrix $\mathbf{\Psi}$, is expressed as:

$$C = \frac{B}{N_c + T} \sum_{\nu=0}^{N_c-1} \log_2 \left(1 + \frac{p_\nu^{\text{opt}} |\bar{h}_\nu + \text{tr}(\mathbf{\Psi}\mathbf{H}_\nu)|^2}{N_0} \right) \text{ bit/s} \tag{28}$$

where B denotes the bandwidth. Besides,

$$p_\nu^{\text{opt}} = \max \left(\mu - \frac{N_0}{|\bar{h}_\nu + \text{tr}(\mathbf{\Psi}\mathbf{H}_\nu)|^2}, 0 \right), \quad \nu = 0, \dots, N_c - 1 \tag{29}$$

Here, p_ν^{opt} denotes the power allocated to subcarrier ν using the water-filling method, and μ is chosen to meet the total power constraint $\sum_{\nu=0}^{N_c-1} p_\nu^{\text{opt}} = P_t N_c$. Moreover, P_t represents the total power available.

Considering the time and frequency domain representations of channels as follows:

$$c_s[\ell] = \sum_{i=1}^{L_s} a_{s,i} e^{-j2\pi f_c(\tau_{s,i}-\eta)} \text{sinc}\left(\ell + B(\eta - \tau_{s,i})\right), \tag{30}$$

$$c_{i,j}[\ell] = a_{r,1,j} a_{t,1,i} e^{-j2\pi f_c(\tau_{r,1,j}+\tau_{t,1,i}-\eta)} \times \text{sinc}\left(\ell + B(\eta - \tau_{r,1,j} - \tau_{t,1,i})\right), \tag{31}$$

$$\bar{c}_s[\nu] = \sum_{\ell=0}^{T} c_s[\ell] e^{-j2\pi\ell\nu/N_c}, \quad \nu = 0, \dots, N_c - 1, \tag{32}$$

$$\bar{c}_{i,j}[\nu] = \sum_{\ell=0}^{T} c_{i,j}[\ell] e^{-j2\pi\ell\nu/N_c}, \quad \nu = 0, \dots, N_c - 1, \tag{33}$$

where $a_{r,1,j}$, $a_{t,1,i} \in [0, 1]$ represent the attenuations, and $\tau_{t,1,i}$, $\tau_{r,1,j} \geq 0$ represent the propagation delays of the impulse responses to and from RIS element 1. The following channel formulation is obtained:

$$\bar{h}_{\mathbf{\Psi}}[\nu] = \bar{c}_s[\nu] + \text{tr}(\mathbf{\Psi} \sum_{j=1}^{L_r} \sum_{i=1}^{L_t} \bar{c}_{i,j}[\nu] \mathbf{a}(\phi_{i,i}, \theta_{i,i}) \mathbf{a}^T(\phi_{o,j}, \theta_{o,j})) \tag{34}$$

where

$$\bar{h}_\nu = \bar{c}_s[\nu], \tag{35}$$

$$\mathbf{H}_\nu = \sum_{j=1}^{L_r} \sum_{i=1}^{L_t} \bar{c}_{i,j}[\nu] \mathbf{a}(\phi_{i,i}, \theta_{i,i}) \mathbf{a}^T(\phi_{o,j}, \theta_{o,j}), \tag{36}$$

The channel model includes several parameters: L_t and L_r denote the number of propagation paths from the transmitter to the RIS and from the RIS to the receiver, respectively. T represents the length determined by the delay spread and PAM pulse, crucial for OFDM modulation. ν and ℓ are indices representing frequency and time, essential for analyzing temporal and spectral dynamics. The angles $(\phi_{i,i}, \theta_{i,i})$ and $(\phi_{o,j}, \theta_{o,j})$ describe incident and outgoing paths relative to the RIS's broadside direction. Finally, N represents the number of phase shifts associated with each path, influencing the interactions within the RIS-assisted communication system. Besides, the vectors $\mathbf{a}(\phi_{i,i}, \theta_{i,i})$ and $\mathbf{a}(\phi_{o,j}, \theta_{o,j})$ are the array response vectors for the incident and outgoing paths at the RIS, respectively, capturing the phase shifts across array elements due to path differences for the specified angles. Then, the capacity in Eq. (28) is maximized respect to a fully connected BD-RIS phase shift constraint as follows:

$$\mathbf{\Psi} = \mathbf{\Psi}^T, \tag{37}$$

$$\mathbf{\Psi}\mathbf{\Psi}^H = \mathbf{I}_N \tag{38}$$

Due to the complexity arising from the reflection matrix $\mathbf{\Psi}$ impacting all subcarriers, they adopt a novel divide-and-conquer approach to address the maximization problem [74].

4.3. Received power

The minimization of transmit power or the maximization of received power is another widely investigated topic in RIS-aided wireless communications. Transmission power reduction not only saves the power consumption of wireless networks but also reduces the interference on adjacent cells. Several papers on maximizaing received power are listed in Table 4. The authors in [76] consider a multiple-input single-output (MISO) system in which a RIS is deployed to enhance wireless communication from the transmitter to the receiver, as depicted in Fig. 6. It is assumed that the direct link between the transmitter and the UE is blocked. The transmitter is equipped with M antennas, forming uniform linear array (ULA) while the user is equipped with a single antenna.

The channel between the transmitter (T) and the receiver (R) to RIS (I) is presented as $\mathbf{h}_{TI}^H \in \mathbb{C}^{N\times M}$ and $\mathbf{h}_{IR}^H \in \mathbb{C}^{1\times N}$ respectively and formulated as follows:

$$\mathbf{h}_{TI}^H = \begin{bmatrix} a_{TI,1,1} e^{\theta_{TI,1,1}} & \cdots & a_{TI,1,N} e^{\theta_{TI,1,N}} \\ \vdots & \ddots & \vdots \\ a_{TI,L,1} e^{\theta_{TI,L,1}} & \cdots & a_{TI,L,N} e^{\theta_{TI,L,N}} \end{bmatrix} \tag{39}$$

$$\mathbf{h}_{RI}^H = \begin{bmatrix} a_{IR,1} e^{\theta_{IR,1}} & \cdots & a_{IR,1,L} e^{\theta_{IR,L}} \end{bmatrix} \tag{40}$$

Where $a_{T_{I,p,q}}$, and $\theta_{T_{I,p,q}}$ denotes the gain and phase of channel from pth antenna of transmitter and the qth reflective element of RIS. In the context of far-field operation, it is reasonable to assume that the amplitude gains from various antenna elements to distinct RIS elements are equal. Thus, the following assumption holds true:

$$a_{\text{TI,p,q}} \simeq a_{\text{TI}}, a_{\text{IR,q}} \simeq a_{\text{IR}}, p = 1, \dots, M, q = 1, \dots, N. \tag{41}$$

and the final amplitude gain from the transmitter to the receiver through individual reflective element of the RIS is $a_{\text{TI}} a_{\text{IR}} \triangleq a_{\text{TIR}}$. The values of a_{TIR} from physical and electromagnetic perspectives is formulated as follows:

$$\begin{aligned} a_{\text{TIR}} &= \sqrt{\frac{\|E^r\|^2}{2Z_{\text{air}}} A_r} \\ &= \sqrt{\frac{Z_{\text{air}} G_t G d_x d_y F(\theta_t, \phi_t) F(\theta_r, \phi_r) \Gamma^2 d_{\text{TI}}^{-2} d_{\text{IR}}^{-2}}{16\pi^2 Z_{\text{air}}} \frac{G_r \lambda^2}{4\pi}} \\ &\cong \delta_{\text{TIR}} d_{\text{TI}}^{-1} d_{\text{IR}}^{-1}. \end{aligned} \tag{42}$$

where $F(\theta_t, \phi_i) F(\theta_r, \phi_r)$ is the normalized power radiation pattern of the RIS, Z_{air} is the characteristic impedance of the air, G_t is the transmitter antenna gains, Γ is the RIS gain, (θ, ϕ) are the elevation and azimuth angles from the reference point of the RIS to the transmitter and receiver, F is the normalized power radiation pattern, and μ_{TI} is the complement angle of the direction of arrival (DOA) of the RIS at the transmitter. Then, the phase changes in the channel $\boldsymbol{h}_{\text{TI}}^H$ are equivalently written as:

$$\theta_{\text{TI},q,p} = 2\pi \frac{d_{\text{TI},q,p}}{\lambda} \cong 2\pi \frac{d_{\text{TI}} + \Delta d_{\text{TI},q,p}}{\lambda}, \tag{43}$$

$$\theta_{\text{IR},q} = 2\pi \frac{d_{\text{IR},p}}{\lambda} \cong 2\pi \frac{d_{\text{IR}} + \Delta d_{\text{IR},q}}{\lambda}, \tag{44}$$

where λ is the carrier wavelength, $\Delta d_{TI,q,p}, d_{TI}$, corresponds to the distance between the pth antenna and the reference point of the transmitter and the distance from the pth antenna element to the qth RIS reflective element, and $a_{TI,p,q}, a_{IR,q}$, stands for the amplitude gains from various antenna elements to different RIS elements.

The received power and its maximization problem respect to transmit and passive beam forming vectors $\mathbf{v}$ and $\mathbf{\Theta}$, the position of the RIS $\mathbf{r}_{I,pos}$, and the orientation of the RIS, $\mathbf{r}_{I,ori}$, is formulated as follows:

$$P_r = |\mathbf{h}_{IR}^H \mathbf{\Theta} \mathbf{h}_{TI}^H \mathbf{v}|^2. \tag{45}$$

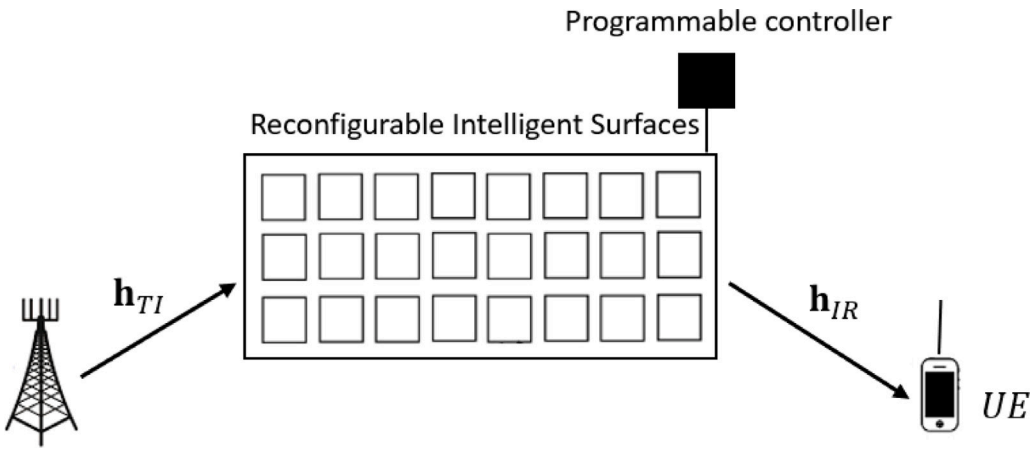

Fig. 6. Diagram of RIS-aided MISO communication system.

Table 4
Transmit power minimization or Received power maximization for EM-based models of RIS-assisted communications.

Ref.	Scenario	Objective	Design variables	Algorithms
[77]	BD-RIS aided MU-MISO network	Transmit power minimization	Transmit beamforming Phase shifts of the BD-RIS	BCD
[78]	Coordinated BD-RIS-aided SU MISO	Received power maximization	Transmit beamforming Phase shifts of the RIS	AO, Closed-form solutions
[79]	BD-RIS MISO	Received power maximization	Transmit beamforming and phase shifts of the RIS	Closed-form solutions
[80]	Multi-band RIS multi-BS MIMO	Received power maximization	phase shifts of the RIS	Closed-form solutions
[81]	Hybrid BD-RIS SU SISO, MISO,MIMO	Received power maximization	Phase shifts of the RIS	PDD
[76]	SISO	Received power maximization	Transmit beamforming ,phase shifts of the RIS, orientation and position of the RIS	AO, closed-form solutions
[82]	SISO	Received power maximization	Transmit beamforming and phase shifts of the RIS	AO, SOCP, geometry-based optimal phase control algorithm
[83]	SISO	Received power maximization	Tunable impedances at the RIS	Gradient-based algorithm
[84]	SISO	Received power maximization	Discrete phase shifts of RIS	Dynamic threshold phase quantization (DTPQ) and equal interval phase quantization (EIPQ)
[85]	SISO	Received power maximization	Discrete phase shifts of RIS	Group-based query algorithm, Closed-form solutions
[86]	STAR-RIS	Received power maximization	Transmit beamforming and phase shifts of the RIS	Closed-form solutions
[87]	BD-RIS-assisted SISO, MISO, SIMO	SNR maximization	RIS phase shift matrix	Closed-form solution
[88]	SISO	End-to-end SNR maximization	Tunable impedances of the RIS	Closed-form solution
[89]	MUs SIMO	End-to-end SINR maximization	Receive BF vectors at the BS, transmit power allocation at the users, and the RIS phase shift design	Projected gradient descent-based solution

$$\begin{aligned} \underset{\mathbf{v},\boldsymbol{\Theta},\mathbf{r}_{I,pos},\boldsymbol{r}_{I,ori}}{\text{maximize}} \quad & P_r \\ \text{subject to} \quad & \mathbf{v}^H\mathbf{v} \le P_t \\ & \boldsymbol{\Theta} \in \mathbb{B} \\ & \{\mathbf{r}_{I,pos}, \boldsymbol{r}_{I,ori}\} \in \mathbb{S}_0 \\ & \mathbf{r}_{I,pos} \in \mathbb{S}_1 \cap \mathbb{S}_2. \end{aligned} \tag{46}$$

where $\mathbb{S}_0$ is a set that guarantees the EM to propagate from transmitter to receiver through the RIS directly, which is specified by application environment, $\mathbb{S}_1$ can be expressed as $(d_{\mathrm{TI}} \ge r_0) \cap (d_{\mathrm{RI}} \ge r_1)$ where r_0 and r_1 are the minimum distance to guarantee the far-field condition, respectively, the region $\mathbb{S}_2$ defines the feasible area for fixing the RIS, while $\mathbb{B}$ characterizes the feasible set of Θ. It is assumed that the antennas in T and R are omni-directional.

The recent work in [78] proposes a coordinated RIS architecture that allows for optimized connection patterns between RIS elements and configurable impedances based on CSI. Unlike existing architectures, the coordinated RIS offers greater optimization flexibility. Simulation results show that it achieves higher power gain compared to

group-connected RIS, approaching the performance of fully-connected architectures while requiring fewer configurable impedances.

In [79], the authors introduces new BD-RIS models and architectures using graph theory, resulting in two efficient designs: tree-connected and forest-connected RIS. The tree-connected architecture, based on a tree graph, is the simplest yet reaches the performance peak for MISO systems. The forest-connected architecture offers a trade-off, reducing complexity more than the tree-connected model. For optimization, a closed-form solution is available for the tree-connected RIS, while the forest-connected RIS uses a simpler iterative method. Numerical tests show that the tree-connected RIS matches the performance of a fully-connected RIS and the forest-connected aligns with group-connected RIS, but with up to 16.4 times less complexity. Another work in [80] explores the frequency-dependent characteristics of BD-RISs, proposing a new reflection model for multi-band MIMO networks. It develops optimization strategies to enhance received power at various users connected to different base stations operating on separate frequencies. By leveraging matrix theory, the study derives simplified solutions for complex optimization challenges and implements these using codebook-based configurations. Simulation results demonstrate the effectiveness of BD-RISs over traditional models, particularly in their robustness to frequency variations, while also highlighting potential interference issues without proper synchronization between RISs and base stations. In [81], the authors present a closed-form solution for optimizing the scattering matrix in BD-RIS architectures, achieving theoretical performance limits across various channel setups. Initially developed for single-user SISO systems, the solution extends to MIMO and MISO configurations. The proposed algorithm simplifies and reduces the complexity compared to previous methods, scaling linearly with RIS elements in group connected architectures and cubically in fully connected ones.

The received power maximization is also studied in [83] [84,85]. The authors in [83] explore RISs using an EM-compliant model. Focusing on a SISO system with an RIS, the RIS is modeled as a collection of tunable load impedances controlling thin wire dipoles. A gradient-based algorithm is introduced to optimize the impedance of scattering elements in the presence of mutual coupling. The algorithm's convergence and computational complexity are analyzed, with numerical results demonstrating its superiority over a benchmark algorithm in terms of performance and convergence time.

The authors in [84] examine phase alignment and quantization for RIS in SISO systems. Their work uncovers phase distribution patterns that inform the discretization of phase shifts in the RIS. The study introduces dynamic threshold and equal interval phase quantization methods, demonstrating marked improvements in power gain, stability, and robustness over conventional methods. Additionally, the paper presents the first path loss scaling law under discrete phase shifts and validates these results through field trials. This research enhances RIS performance optimization and supports practical implementations in communication systems. In [85], the authors present a practical reflection coefficient model for RIS, departing from idealized assumptions. The model considers non-uniform amplitudes and limited phase shifts based on measurements. An algorithm is proposed to calculate the coefficients while accommodating these constraints, with analytical performance analysis. Simulation and experimental results validate the effectiveness of the algorithm, enhancing our understanding of RIS behavior and deployment possibilities. The work in [86] focus on metasurface-based RISs designed for simultaneous transmission and reflection (STAR). These structures, with tiny elements, differ from traditional patch-array RISs and are particularly suited for near-field scenarios. A continuous electric current distribution approach is employed for modeling their electromagnetic response. The research provides analytical insights for both single-user and multi-user settings. Notably, near-field channels offer higher degrees-of-freedom (DoFs) than far-field channels. STAR-RIS strategies are proposed and evaluated, showing the superiority of metasurface-based RISs and providing valuable insights into power scaling laws and optimal strategies for different scenarios. In [87], the authors discuss the use of BD-RIS architectures, which utilize unitary and symmetric RIS matrices to enhance wireless channel flexibility. The focus is on maximizing the SNR for single and multiple antenna systems using a BD-RIS. The problem of maximizing SNR is tackled by providing a solution in closed-form, achieved through the Takagi factorization of a complex symmetric matrix. This approach facilitates efficient resolution for SISO, SIMO, and MISO channel configurations. Besides, in [89], the authors presents a novel method for optimizing RIS in advanced communication systems. It uses channel statistics instead of instantaneous channel data, enhancing beamforming, user power allocation, and RIS phase adjustments to maximize the minimum SINR and ensure fairness among users. The problem is also formulated including electromagnetic field exposure (EMF)-aware constraints. The specific absorption rate (SAR), expressed in watts per kilogram of body weight (W/kg), gauges EMF in the immediate vicinity. It assesses how quickly the body absorbs energy when subjected to electromagnetic radiation.

The work in [88] introduces an optimization algorithm for RIS-assisted communications that maximizes end-to-end SNR by considering RIS tunable impedances and accounting for mutual coupling. The results show that this approach significantly improves SNR, primarily for single-antenna transceivers in the far-field.

4.4. Sum rate

Enhancing the sum-rate and channel capacity stands out as a prominent benefit of RIS. In contrast to direct transmission from the BS to the user, RISs offer an indirect line, potentially line-of-sight, resulting in reduced path loss and increased SINR [32].

In [90], the authors consider a MIMO interference channel that comprises N_T transmitter–receiver pairs. They adopt an electromagnetic-compliant communication model based on mutual impedances, particularly suited for RISs comprising thin wire dipoles made of perfectly conducting material. These dipoles are controllable through tunable load impedances, offering the flexibility to program and shape the scattered field within the RIS-assisted channel. The system model is shown in Fig. 7.

Under the following assumptions:

- The channel matrix between the jth transmitter and the kth receiver is determined by the internal impedances of the transmit antennas, the load impedances of the receive antennas, and the tunable impedances of the RIS.
- Denoting $\mathbf{Z}_I^{(i)} = \mathbf{Z}_{\text{tun}}^{(i)}$ as a $P \times P$ diagonal matrix representing the tunable impedances of the ith RIS, where $\boldsymbol{B}_i = \text{diag}(\boldsymbol{b}_i)$ is a diagonal matrix containing the adjustable impedances of all I RISs.

The following notation for channel matrices is derived:

$$\bar{\mathbf{h}}_{\text{TR},k,j} = (\boldsymbol{I}_L + \boldsymbol{Z}_{k,k}\boldsymbol{Z}_k^{-1})^{-1}\boldsymbol{Z}_{k,j}(\boldsymbol{Z}_{j,j} + \boldsymbol{Z}_j)^{-1} \in \mathrm{C}^{L\times M}, \tag{47a}$$

$$\bar{\boldsymbol{B}}_i = \boldsymbol{Z}_{i,i} \in \mathrm{C}^{P\times P}, \quad \boldsymbol{B}_i = \boldsymbol{Z}_I^{P\times P}, \tag{47b}$$

$$\tilde{\mathbf{h}}_{\text{TR},k,i,j} = -\boldsymbol{T}_{k,i}(\bar{\boldsymbol{B}}_i + \boldsymbol{B}_i)^{-1}\boldsymbol{S}_{i,j} \in \mathrm{C}^{L\times M}, \tag{47c}$$

$$\mathbf{h}_{\text{TR},k,j}(\mathcal{B}) \sim \bar{h}_{\text{TR},k,j} + \sum_{i=1}^{I} \tilde{\mathbf{h}}_{\text{TR},k,i,j}(\mathcal{B}) \in \mathrm{C}^{L\times M}. \tag{47d}$$

where $\bar{\mathbf{h}}_{\text{TR},k,j}$ accounts for the line-of-sight link and $\tilde{\mathbf{h}}_{\text{TR},k,i,j}$ accounts for the (virtual line-of-sight) link scattered by the ith RIS. To simplify notation, we use $\mathcal{B} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_I\}$ to collectively represent the I vectors b_i that are subject to optimization. Then, an EM-aware sum-rate maximization problem is formulated as follows:

$$\underset{\mathcal{V},\mathcal{B}}{\text{Max}} \quad \mathrm{R}_{\text{tot}}(\mathcal{V},\mathcal{B}) = \underset{\mathcal{V},\mathcal{B}}{\text{Max}} \sum_{k=1}^{N_T} \alpha_k R_k(\mathcal{V},\mathcal{B}) \tag{48a}$$

$$\text{subject to} \quad \text{tr}(\boldsymbol{V}_k\boldsymbol{V}_k^H) \le P_k, \quad \forall k = 1,\ldots,N_T \tag{48b}$$

$$\Re(b_{k,p}) = R_{\text{resistance},0}, \quad \forall i = 1,\ldots,I,\ p = 1,\ldots,P \tag{48c}$$

$$\Im(b_{k,p}) \in \mathbb{R}, \quad \forall i = 1,\ldots,I,\ p = 1,\ldots,P \tag{48d}$$

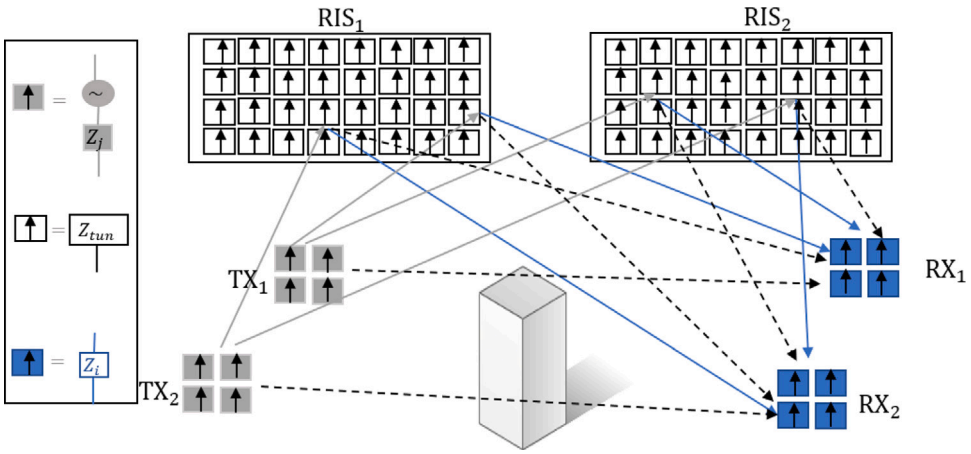

Fig. 7. Illustration of an RIS-assisted MIMO interference channel (© 2021 IEEE, [90]).

Table 5
Maximization of sum rate with physically-consistent models for the RIS in RIS-assisted communications.

Ref.	Scenario	Objective	Design variables	Algorithms
[63]	Hybrid BD-RIS-aided MUs-MISO	Sum rate maximization	Transmit beamforming , and BD-RIS matrix	Iterative algorithm, Closed form solution
[77]	BD-RIS aided SU-MISO network	Sum rate maximization	Transmit beamforming Phase shifts of the BD-RIS	block ascent (BCA) method
[90]	MIMO	Sum rate maximization	Active and passive beamforming	BCD, wMMSE
[91]	BD-RIS Aided SU/MUs-MISO	Sum rate maximization	Tunable impedance components of the BD-RIS	closed-form solution
[92]	BD-RIS RSMA Aided MUs-MISO	Sum rate maximization	Transmit beamforming and multi-sector BD-RIS configuration	Majorization-minimization (MM) and penalty-dual decomposition (PDD) method
[93]	FC BD-RIS MUs MISO ISAC system	Sum rate maximization	linear filter, transmit beamforming, and the BD-RIS configuration	RCG(Riemannian Conjugate Gradient), Manifold Based Solution
[94]	MU-MIMO	Sum rate maximization	SAR constraints	Iterative waterfilling algorithm
[95]	MISO	Sum rate maximization	Tunable impedances of the unit cells of the RIS, Passive beamforming	Iterative algorithm
[96]	RIS-aided MUs-MISO SDMA/NOMA	Weighted sum rate maximization	Multiple access precoding at the BS, and RIS configuration	AO, WMMSE precoding, Unsupervised ML

where

$$R_k(\mathcal{V}, \mathcal{B}) = \log\det(\mathbf{I}_L + \mathbf{V}_k^H \mathbf{h}_{\mathrm{TR},k,k}^H(\mathcal{B}) \bar{\mathbf{J}}_k^{-1} \mathbf{h}_{\mathrm{TR},k,k}(\mathcal{B}) \mathbf{V}_k) \tag{49}$$

and the interference-plus-noise covariance matrix is formulated as:

$$\bar{\mathbf{J}}k = \sum_{j=1, j\neq k}^{N_u} \mathbf{h}_{\mathrm{TR},k,j}^H(\mathcal{B}) \mathbf{V}j \mathbf{V}j^H \mathbf{h}_{\mathrm{TR},k,j}^H(\mathcal{B}) + \sigma_k^2 \mathbf{I}_L \tag{50}$$

where $\mathcal{V} = \{\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_{N_T}\}$ denotes the set of N_T precoding matrices.

Besides, $R_{\mathrm{resistance},0} \geq 0$ represents a resistance value that characterizes the losses within the tunable impedances of the RIS elements. Constraint (48c) is in place to guarantee that the RIS does not act as an amplifier for the incoming signal.

Table 5 summarizes sum-rate/channel capacity maximization works in RIS-aided wireless communications from EM perspectives, including scenario, objective, design variables, and algorithms.

The work in [91] explores optimizing the performance of group-connected BD-RIS. It introduces a static grouping strategy based on channel statistics to enhance performance while maintaining low circuit complexity. By solving optimization problems, the study demonstrates substantial performance improvements, particularly in scenarios with highly correlated channels, with up to a 60% increase in sum rate. On the other hand, the work in [63] introduces a new approach to enhance the performance of BD-RIS by dynamically adjusting antenna groupings based on CSI. Unlike existing static configurations, this dynamic grouping strategy optimizes the BD-RIS setup by adapting antenna subsets to variations in CSI, resulting in a more flexible and efficient system. The proposed method is evaluated using simulations, demonstrating its superiority over fixed group-connected architectures in multi-user multiple-input single-output systems.

In [93], the authors explore the use of fully-connected beyond-diagonal RIS (FC BD-RIS) to enhance the performance of integrated sensing and communication (ISAC) systems. The paper emphasizes the superior beamforming capabilities of BD-RIS compared to traditional diagonal RIS architectures. By applying the majorization-minimization (MM) and penalty-dual-decomposition (PDD) methods, the study develops an effective algorithm to address the challenges posed by orthogonality conditions and non-convex inequalities involving BD-RIS. The numerical results confirm the efficiency of the proposed solution and the advantages of incorporating BD-RIS in ISAC networks. The work in [92] investigates the combination of rate-splitting multiple access (RSMA) with a BD-RIS to enhance network performance and reduce antenna usage. Using a prism-shaped multi-sector BD-RIS in a multiuser system, the research develops joint transmit precoder and BD-RIS matrix designs under imperfect channel conditions. By transforming a

stochastic maximization problem into a deterministic one, the approach optimizes system performance iteratively. Simulations show that this integration outperforms traditional access methods and reduces the number of antennas needed.

In [94], the authors underscore the importance of incorporating user exposure constraints into the design of uplink transmission covariance matrices for advanced cellular systems such as 5G and beyond. The study emphasizes that ensuring near-field user exposure limitations impact the potential far-field data rates achievable by devices with multiple transmit antennas. The paper focuses on specific absorption rate (SAR) as a well-recognized metric for user exposure. It investigates a multi-user MIMO system, considering SAR restrictions at each user. The research analyzes the maximum achievable sum rates under different transmitter channel state information scenarios. The proposed SAR-aware MIMO transmission techniques, based on a modified waterfilling algorithm, are demonstrated to outperform conventional strategies for scenarios involving two users.

In [95], the authors focus on developing electromagnetic models for RISs, building upon a recent model involving mutually coupled loaded wire dipoles. The authors introduce an RIS-aided channel model suitable for more realistic scenarios. In this model, they represent scattering objects as loaded wire dipoles, allowing them to characterize the behavior of various natural and engineered materials by adjusting the dipole parameters. They also present an iterative algorithm that jointly optimizes RIS and transmitter configurations, aiming to maximize the system's overall data rate. Extensive numerical experiments confirm the considerable performance enhancements offered by their approach when compared to existing optimization methods.

In [96], Peng et al. employed ML techniques to WSR optimization of RIS-assisted communication systems, considering mutual coupling between RIS elements for the first time. The research begins by deriving an RIS-assisted channel model that incorporates mutual coupling effects. Subsequently, an unsupervised ML approach is proposed to optimize the RIS, utilizing a dedicated neural network architecture called RISnet with good scalability and desired symmetry. Furthermore, the study integrates ML-enabled RIS configuration with analytical precoding at the base station and introduces a variant of RISnet that requires partial channel state information from a subset of RIS elements [96].

4.5. Spectral efficiency

RISs are regarded as a promising technology for extending coverage and improving the spectral efficiency of wireless systems [97]. In a study by [98], the authors explore the design of SE maximization considering EM effects in multiuser MIMO uplink transmission assisted by RISs and digital metasurface antennas (DMAs). They propose a joint optimization of transmit precoding, RIS phases, and DMA weights to enhance SE while adhering to power and specific absorption rate (SAR) constraints at users. The system model is depicted in Fig. 8. The SE maximization problem concerning the transmission covariance matrix $\mathbf{Q}$ and EM exposure level is formulated as follows:

$$\underset{\mathbf{Q},\boldsymbol{\Phi},\boldsymbol{\Xi}}{\text{Maximize}} \quad \eta_{\text{SE}}(\boldsymbol{Q},\boldsymbol{\Phi},\boldsymbol{\Xi}) \tag{51a}$$

$$\text{subject to} \quad \text{tr}(\mathbf{Q}_k) \le P_{\max,k}, \quad \mathbf{Q}_k \succeq \mathbf{0} \tag{51b}$$

$$\text{tr}(\mathbf{R}_{k,i}\mathbf{Q}_k) \le D_{k,i}, \quad \forall k,i, \tag{51c}$$

$$\phi_n \in \mathcal{F}_1, \quad \forall n; \tag{51d}$$

$$\boldsymbol{\Xi} \in \mathcal{F}_3^{S\times M}. \tag{51e}$$

where $\varphi_n \in \mathcal{F}_1 = \{\varphi = e^{j\theta}, \theta \in [0, 2\pi)\}$. The DMA array is equipped at the BS consisting of M metamaterial units. These DMAs are composed of S microstrips, each of which contains L metamaterial units, i.e., $M = SL$. In practice, the DMA array can be regarded as a two-dimensional antenna array composed of a set of one-dimensional microstrips, and its configurable weight matrix $\boldsymbol{\Xi} \in \mathbb{C}^{S\times M}$ can be written as

$$\boldsymbol{\Xi}_{s1,(s2-1)L+l} = \begin{cases} \xi_{s1,l} \in \mathcal{F}_2 & \text{for } s_1 = s_2 \\ \mathbf{0} & \text{otherwise} \end{cases} \in \mathcal{F}_3^{S\times M} \tag{52}$$

where $s_1, s_2 \in \{1,\dots,S\}$, $l \in \{1,\dots,L\}$, and $\{\xi_{s_1,l}\}_{\forall s_1,l}$ are the weights of the DMA elements, and the feasible set of weight matrices is denoted as $\mathcal{F}_{S\times M}^3$. $\xi_{s_1,l}$ often satisfies certain constraints, e.g., its feasible set represented by $\mathcal{F}_2$. The input–output relationship of the system between the transmit signals and DMA's output can be written as:

$$\mathbf{z} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{\Xi}\mathbf{h}_{TI}\boldsymbol{\Phi}\mathbf{h}_{IR,k}\mathbf{x}_k + \tilde{\mathbf{n}} \in \mathbb{C}^{S\times 1}, \tag{53}$$

where the channel matrix from user k to the RIS is denoted as $\mathbf{h}_{IR,k} \in \mathbb{C}^{N_R\times N_k}$, and the channel matrix from the RIS to the BS as $\mathbf{h}_{TI} \in \mathbb{C}^{M\times N_R}$. $\tilde{\mathbf{n}} = \frac{1}{K}\boldsymbol{\Xi}\mathbf{n}$ is the equivalent channel noise. The constraints (51b) and (51c) keep the transmission power and EM exposure level below some specified values, and (51d) and (51e) prescribe the formats of RIS phase shift matrix $\boldsymbol{\Phi}$ and DMA weight matrix $\boldsymbol{\Xi}$, respectively.

4.6. Energy efficiency

Energy efficiency is a critical metric for green 5G and 6G networks. Different with power minimization problems, the objective of energy efficiency maximization includes both transmission rate and energy consumption metrics, which can better evaluate the power utilization efficiency. The main benefit of RISs lies in the capability of reshaping the signal propagation path with extremely low power consumption, making RISs a promising technique to improve energy efficiency. Energy efficiency maximization-related works are summarized in Table 6. In [99], a comprehensive model for MISO transmission with RIS assistance, incorporating both the end-to-end electromagnetic transfer model and the non-linear rectifier model is presented. The approach considers the hardware-related implications on EE, encompassing factors such as the radiation patterns of transmit antennas, the reflective properties of the RIS, and particularly the mutual coupling effects influenced by the arrangement of antenna arrays. The adopted model employs circuit impedances as optimization variables for configuring the RIS, making it more practical than ideal assumptions. The system model is shown in Fig. 9 and includes several separate information decoder user (IDU) and energy harvester user (EHU). Two kinds of users, IDUs and EHUs, are belong to the set $\mathcal{M}_{\mathcal{I}} = \{1,\dots,M_I\}$ and $\mathcal{M}_{\mathcal{E}} = \{1,\dots,M_E\}$, respectively. For the RIS-enhanced SWIPT network, the problem formulated to maximize the EE subject to the maximal transmitting power, the minimal rate of the IDU, the minimal DC output power of the EHU, and the hardware restrictions of the control circuit at the RIS. The formulated optimization problem is as follows:

$$\underset{\mathbf{w}_j,\mathbf{v}_l,\boldsymbol{\Omega}}{\text{Maximize}} \quad \eta_{\text{EE}}(\boldsymbol{\omega}_j,\mathbf{v}_l,\boldsymbol{\Omega}) \tag{54a}$$

$$\text{subject to} \quad R_j(\boldsymbol{\omega}_j,\boldsymbol{\Omega}) \ge R_j^{(D)}, \quad \forall j \tag{54b}$$

$$P_l(\boldsymbol{\omega}_j,\mathbf{v}_l) \le P_{\max}, \tag{54c}$$

$$P_{\text{RF},n}(\boldsymbol{\omega}_j,\mathbf{v}_l,\boldsymbol{\Omega}) \ge g_n(E_n^{(D)}), \quad \forall n; \tag{54d}$$

$$\Re(\boldsymbol{\Omega}_q) = R_0 \ge 0, \quad \forall q \tag{54e}$$

$$\Im(\boldsymbol{\Omega}_q) \in \mathbb{R}, \quad \forall q. \tag{54f}$$

where $\eta_{EE}(\boldsymbol{\omega}_j,\mathbf{v}_l,\boldsymbol{\Omega}) = \frac{\sum_{j=1}^{M_I} R_j(\boldsymbol{\omega}_j,\boldsymbol{\Omega})}{P(\boldsymbol{\omega}_j,\mathbf{v}_l)}$, $\boldsymbol{\omega}_j \in \mathbb{C}^{N_T\times 1}$ and $\mathbf{v}_l \in \mathbb{C}^{N_T\times 1}$ denote the transmitting beamforming vectors for the jth IDU and the lth EHU,(54), $R_j^{(D)}$ refers to the rate condition for the jth IDU. Regarding Eq. (54b), $P_{\max}$ represents the upper limit on the power budget for transmission. The value $E_n^{(D)}$ within Eq. (54c) signifies the minimum DC power output requirement for the nth EHU. Eq. (54c) and (54d) outline the constraints of the control circuit. In Eq. (54e), R_0 indicates the circuit's loss, with the requirement $R_0 \ge 0$ ensuring that the RIS functions in passive mode rather than active mode. Within the context of this paper, the reactances of the circuits mentioned in Eq. (54f) can be reconfigured, and we presume that their adjustable ranges encompass the real number set.

Furthermore, the total transfer model can be depicted by

$$\mathbf{h}(\boldsymbol{\Omega}) = \mathbf{h}_{\text{EM}}(\boldsymbol{\Omega}) + \mathbf{h}_{\text{Cluster}} \in \mathbb{C}^{1\times N_t} \tag{55}$$

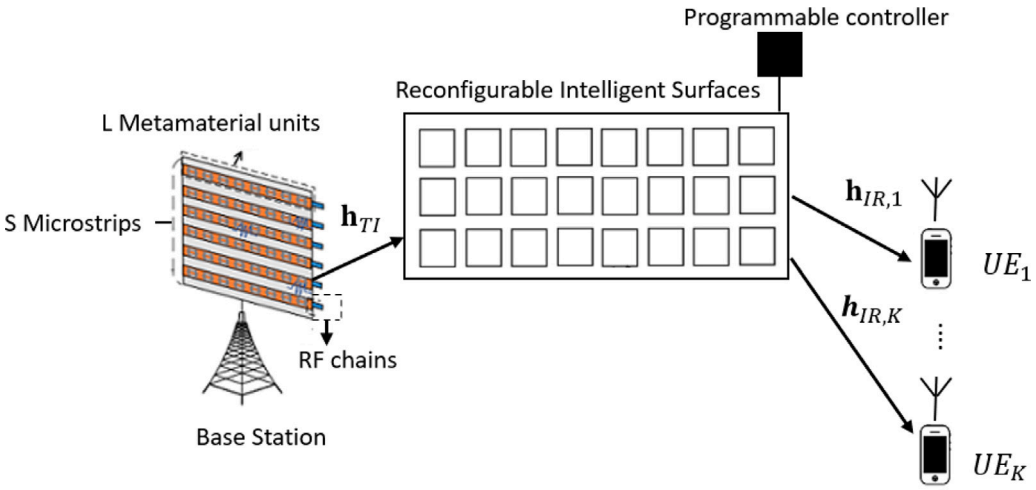

Fig. 8. Hybrid RIS and DMA assisted multiuser MIMO system (© 2022 IEEE, [98]).

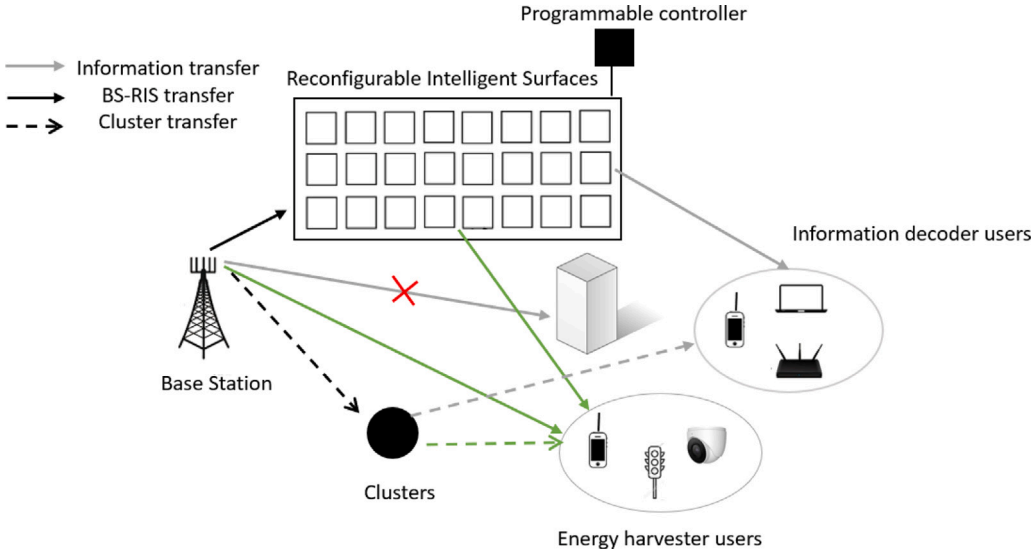

Fig. 9. The RIS-enhanced-SWIPT system (© 2023 IEEE, [99]).

where $\mathbf{h}_{\text{EM}}$ is the electromagnetic physical transfer model and $\mathbf{h}_{\text{Cluster}}$ is the clustered model. Regarding $\mathbf{h}_{\text{EM}}$, its behavior is influenced by the tunable component matrix $\mathbf{\Omega} \in \mathbb{C}^{N_i \times N_i}$. This matrix has a specific structure, denoted as $\mathbf{\Omega} = \text{diag}(\mathbf{\Omega}_1, \ldots, \mathbf{\Omega}_{N_i})$, with each entry representing the impedance of the backend circuits of the RIS elements. The choice of diodes or varactors allows for the configuration of $\mathbf{\Omega}$, with the RIS employing varactors to enable continuous adjustments of reactance. In more detail, the expression for $\mathbf{h}_{\text{EM}}$ is as follows:

$$\mathbf{h}_{\text{EM}}(\mathbf{\Omega}) = \frac{\mathbf{Z}_{\text{L}}\mathbf{S}_{\text{RT}}(\mathbf{\Omega})\left(\mathbf{Z}_{\text{G}} + \mathbf{S}_{\text{TT}}(\mathbf{\Omega})\right)^{-1}}{\mathbf{Z}_{\text{L}} + \mathbf{S}_{\text{RR}}(\mathbf{\Omega}) - \mathbf{S}_{\text{RT}}(\Omega)\left(Z_{\text{G}} + \mathbf{S}_{\text{TT}}(\mathbf{\Omega})\right)^{-1}\mathbf{S}_{\text{TR}}(\mathbf{\Omega})} \in \mathbb{C}^{1\times N_t}. \tag{56}$$

where $\mathbf{Z}_{\text{G}} \in \mathbb{C}^{N_t \times N_t}$ denotes the source impedance matrix, a diagonal matrix, $\mathbf{Z}_{\text{L}}$ is the load impedance, and $\mathbf{S}_{\text{RT}}, \mathbf{S}_{\text{TT}}, \mathbf{S}_{\text{RR}}, \mathbf{S}_{\text{TR}}$ denote the transfer matrices between the transmitter, RIS, and receiver.

The author in [100] addresses the challenge of optimizing EE in a RIS-assisted MIMO communications. The optimization problem involves maximizing EE while adhering to power constraints and limiting end-users' electromagnetic radiation exposure. The work jointly optimizes RIS phase shifts, transmit beamforming, receive filters, and transmit power using two convergent and low-complexity algorithms. One algorithm is applicable to general scenarios but may not achieve global optimality, while the other is proven optimal for cases with isotropic EMF constraints. Numerical results demonstrate that the RIS maintains comparable EE to scenarios without EMF constraints.

The work in [101] investigates the impact of LISs in millimeter-wave networks, demonstrating their potential to boost capacity and energy efficiency. However, it also highlights limitations when the base station density exceeds that of the LISs, particularly due to reflected interference and phase-shift energy consumption, suggesting that LIS deployment may not be necessary in such cases.

In the following section, we provide an extensive overview of model-based, heuristic, and ML methodologies employed in the optimization of RIS-enhanced wireless networks, with a specific focus on EM considerations.

4.7. Lessons learned

- Various problem formulations have been explored in the context of EM-aware RIS-aided communications, aiming to achieve different objectives such as increasing channel gain, enhancing SINR, maximizing sum rate, improving energy efficiency, and minimizing power consumption.

Table 6
EE Maximization in RIS-assisted communications from EM perspective.

Ref.	Scenario	Objective	Design variables	Algorithms
[77]	BD-RIS aided MU-MISO network	EE maximization	Transmit beamforming Phase shifts of the BD-RIS	block ascent (BCA) method
[99]	RIS-aided-SWIPT MISO	EE maximization	Impedance parameters of the RIS elements, Active beamforming vectors	AO, SDR, SCA, and Dinkelbach's algorithm
[100]	MIMO	EE maximization	RIS phase shifts, the transmit beamforming, the linear receive filter, and the transmit power	AO, Closed form solutions
[101]	LIS-assisted mmWave cellular network	EE maximization	Number of LIS elements	Stochastic geometry

- Literature has employed various optimization methods to tackle these problems, including traditional optimization techniques and heuristic algorithms.
- Machine learning and deep learning methods have been relatively underutilized for solving EM-aware RIS-aided communications problems, with few studies exploring their potential in this domain.

5. EM model-aware problem solution

5.1. Model-based optimization

Within the context of this study, model-based approaches refer to algorithms grounded in specific optimization models, underpinned by a comprehensive grasp of the defined problem. Such algorithms often require stringent prerequisites in terms of problem formulation characteristics like convexity, continuity, and differentiability. Among the highlighted model-based algorithms for optimizing physically-consistent EM models-aware RIS-aided communication are methodologies such as alternating optimization (AO), majorization-minimization (MM) techniques, successive convex optimization (SCA), semidefinite relaxation (SDR), Block Coordinate Descent (BCD) as listed in Tables 4, 5, 6.

5.2. Heuristic strategies

These strategies rely on heuristic rules to expedite problem-solving, offering efficient alternatives to conventional model-based methods. Although they may sacrifice optimality and precision in favor of reduced complexity and rapid solutions, heuristic algorithms prove instrumental, especially for tackling NP-hard problems or complementing other algorithms. Heuristic algorithms such as the convex-concave procedure (CCP) algorithm, various meta-heuristic algorithms, greedy algorithms, and the application of matching theory can all be employed to enhance the optimization of wireless networks that incorporate RIS. The work in [102] used heuristic method to solve physically-consistent EM models-aware RIS-aided communication.

- Genetic Algorithm (GA): GA is a versatile optimization method inspired by the principles of natural selection. It uses a population-based approach and evolves solutions over generations to find optimal or near-optimal solutions to complex problems in various domains. GA has been widely applied in the creation of pixelated coded MTS designs [103].
- Particle Swarm Optimization (PSO): PSO is a nature-inspired optimization algorithm that simulates the behavior of a flock of birds or a swarm of particles. In PSO, individual solutions, represented as particles, move through a search space to find the best solution by adjusting their positions based on their own experience and the collective knowledge of the swarm. PSO has been utilized to model EM waves with pixelized coded MTSs [103], while in [104], binary particle swarm optimization (BPSO) was applied to achieve a reflecting MTS featuring both left-handed circular polarization (LHCP) and right-handed circular polarization (RHCP) beams.
- Ant Colony Optimization (ACO): ACO is a metaheuristic algorithm inspired by the foraging behavior of ants. It models the way ants find the shortest path between their nest and food source by simulating the pheromone-based communication and exploration of multiple paths, leading to the discovery of optimal solutions in complex optimization problems. In [103], an inverse design approach for MTS is executed utilizing multi-objective lazy ant colony optimization (MOLACO) to create 3-D nano-antenna configurations. These designs aim to achieve low-loss transmission performance and extensive phase tunability.

5.3. Machine learning-based optimization

The complexity of optimization models in RIS-assisted communication systems stems from various factors such as the large number of reconfigurable elements, mutual coupling effects, and dynamic environmental conditions. Model-based approaches, while accurate, often face challenges due to the high dimensionality of the optimization space and the need for computationally intensive simulations. Heuristic algorithms offer alternative strategies for tackling these complex optimization problems. These methods provide efficient search mechanisms to explore the solution space and can often find near-optimal solutions in a reasonable amount of time. However, heuristic algorithms may struggle with scalability and may not always guarantee convergence to the global optimum.

In contrast, ML/DL techniques offer promising avenues for optimization in RIS-assisted communication systems. ML/DL methods can effectively handle the high-dimensional and nonlinear nature of the optimization problems. By learning patterns and relationships from data, these techniques can provide insights into the underlying structure of the optimization space, enabling more efficient exploration and exploitation of solutions. Additionally, ML/DL approaches can adapt to dynamic environments and incorporate real-time feedback, allowing for continual improvement and optimization.

In the context of RIS-empowered wireless networks, advanced ML techniques encompass supervised and unsupervised learning, RL, federated learning (FL), graph learning, transfer learning, and hierarchical learning that widely used by the researchers for optimizing RIS-aided communications. To the best of our knowledge, and as indicated in Tables 4, 5, 6, limited work like [96] have explored the application of machine learning for the optimization of physically-consistent EM model-aware RIS-assisted communication systems. When optimizing an RIS for wireless communications, certain parameters, like circuit parameters, require long-term optimization, typically during the fabrication process. This means that these optimizations can be conducted offline. However, other optimizations, such as passive beamforming optimization, need to be performed dynamically, on the fly. In the upcoming sections, we will delve into recent literature concerning on applying ML/DL methods for the RIS design and optimization and shed light on these two distinct time scales of optimization.

5.3.1. Long-term optimization: Fabrication-based parameter tuning

Some literatures investigate using inverse design and PINN for metamaterial design generally not RIS. Metamaterials are artificially engineered materials designed to exhibit properties that are not found in naturally occurring materials. These materials are composed of sub-wavelength structures (often referred to as meta-atoms) that interact with electromagnetic waves in ways that lead to unique properties, such as negative refractive indices or extreme wave manipulation. Metamaterials can be designed to control the propagation, absorption, reflection, and transmission of electromagnetic waves across different frequency ranges. They find applications in various fields, including optics, antennas, cloaking devices, and imaging systems.

Metamaterials and RISs share some similarities in their ability to control electromagnetic waves, but they serve different purposes. Metamaterials focus on altering the inherent properties of materials to achieve specific electromagnetic behaviors, while RISs aim to actively control the propagation of existing electromagnetic signals. In some cases, metamaterials can be integrated into the design of RIS elements to enhance their performance, such as achieving more precise control over phase and amplitude adjustments.

5.3.2. Dynamic optimization: Real-time beamforming adaptation

To the best of our knowledge, the optimization of EM-model aware RIS-aided communication systems using ML/DL techniques has only been explored in the work by [96].

Peng et al. introduce a specialized NN architecture called RISnet, notable for its outstanding scalability and desired symmetry properties. They integrate ML-based RIS configuration with analytical precoding at the BS, leveraging pre-existing schemes. This work represents an early effort in combining ML techniques with domain knowledge in communication for NN architecture design. Problem-specific ML methods are shown to achieve higher performance, lower complexity, and enhanced symmetry compared to generic ML approaches. After formulating two optimization problem for two multiple access schemes, space-division multiple access (SDMA) and non-orthogonal multiple access (NOMA), problems optimization using proposed unsupervised ML presented. The framework of unsupervised ML begins by defining the problem representation $\mathbf{\Gamma}$, which consists of CSI including the mutual coupling effect, user weights in SDMA, and rate requirements in NOMA. The solution sought is denoted by $\mathbf{\Phi}$, representing the RIS phase shifts aimed at maximizing the objective function f, expressed as $f(\mathbf{\Gamma}, \mathbf{\Phi})$. This function is fully determined by $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ and can be expanded to $f(\mathbf{\Gamma}, \mathbf{\Phi}) = f(\mathbf{\Gamma}, N_\theta(\mathbf{\Gamma}); \theta)$, indicating the dependency on the parameters θ of neural network N_θ. The neural network maps $\mathbf{\Gamma}$ to $\mathbf{\Phi}$ as $\mathbf{\Phi} = N_\theta(\mathbf{\Gamma})$.

To optimize this setup, a large dataset D of $\mathbf{\Gamma}$ is collected, and the optimization problem is formulated as:

$$\max_{\theta} K = \sum_{\mathbf{\Gamma} \in D} f(\mathbf{\Gamma}, N_\theta(\mathbf{\Gamma}); \theta). \tag{57}$$

The network N_θ is trained using gradient ascent, with the update rule:

$$\theta \leftarrow \theta + \eta \nabla_\theta K, \tag{58}$$

where η represents the learning rate. If N_θ is effectively trained, then for a new $\mathbf{\Gamma}' \notin D$, $\mathbf{\Phi}' = N_\theta(\mathbf{\Gamma}')$ should also be a satisfactory solution, mimicking how humans apply experience to solve new but similar problems.

5.4. Inverse design

Inverse design techniques are increasingly recognized as powerful alternatives to traditional methods for developing intricate functional metasurfaces. In this section, we review the literature on the inverse design of metasurfaces.

5.4.1. Data-driven approach

In the domain of modern nanophotonic devices, characterized by intricate nanostructures engineered to achieve advanced functionalities, traditional design methodologies grounded in optimization face formidable challenges. These conventional approaches typically commence with the generation of random designs, followed by resource-intensive electromagnetic simulations to evaluate their performance. The iterative refinement process involves adjusting structural parameters, often necessitating a substantial number of simulations. This becomes particularly impractical for intricate and complex devices. However, the field is currently experiencing a transformative shift toward data-driven methods, prominently leveraging artificial neural networks (NNs) as a promising alternative.

NNs offer two primary modes of operation: first, forward-modeling networks, which expedite the design process by predicting electromagnetic responses from structural parameters, thereby replacing computationally expensive EM simulations; and second, inverse-design networks, which directly generate structures based on desired responses, eliminating the need for iterative optimization. Both modes require significant training data, which involves EM simulations and computational resources as a one-time investment. This approach offers a distinct advantage over conventional optimization, which consumes extensive simulations for each design iteration. This shift toward data-driven approaches extends to the optimization of intricate metasurfaces, showcasing their potential to streamline design processes efficiently.

Within this context, inverse design techniques are emerging as potent solutions to the challenges faced by conventional methodologies when crafting complex functional metasurfaces. Furthermore, emerging research trends, such as scaling up inverse design methodologies for larger aperiodic metasurfaces and exploring end-to-end inverse design by co-optimizing photonic hardware and post-image processing, are poised to shape the future of metasurface design. By integrating these trends, metasurface design stands to achieve new levels of efficiency and effectiveness in crafting advanced nanophotonic devices [103,105–111].

When inverse design employed in the context of metamaterial surface (MTS) design, deep learning networks offer a significant advantage by acquiring a deep understanding of complex and unconventional connections between the structure of a metamaterial surface and its EM behavior. This comprehension is achieved through a collection of training examples [112]. Leveraging these trained datasets containing transmission or reflection response information, the learning algorithms have the capability to generate tailored, customized MTS configurations, including arbitrary wave-front MTS or metagratings [113,114].

Several more Refs. [104,115–120] have utilized various neural network architectures for MTS inverse design. These architectures encompass artificial neural networks (ANN), convolutional neural networks (CNNs), conditional deep convolutional generative adversarial network (cDC-GAN), and conditional deep convolution variational autoencoder (cDC-VAE), contributing to different aspects of MTS inverse design.

5.4.2. Model-aided approach

Another method, topology optimization, leverages advanced algorithms and physics principles to efficiently adjust and re-evaluate parameters to achieve the desired goals. Typically applied to simple meta-surface structures, machine learning's precision relies on numerical algorithms and data set size. However, it might lead to non-unique solutions in inverse design scenarios. To address these challenges, physics-informed neural networks incorporate physical laws during training to predict electromagnetic responses accurately. Topology optimization, which considers physical laws and fabrication constraints, can also be employed for improved precision in designing structure size and shape. This method yields the highest accuracy when the physical model is precise during optimization. The computational time required by machine learning techniques depends largely on factors such as

dataset size, metasurface complexity, and the number of optimization parameters. Strategies to reduce computational time include minimizing the training dataset size, simplifying the structure, and reducing the number of optimized parameters. However, these approaches may lead to a limited number of design parameters, potentially decreasing the diversity of structure designs. In contrast, PINNs integrate mathematical and physical principles, enabling more efficient training with fewer data samples while accommodating a greater number of structure parameters. This approach enhances both time efficiency and design flexibility [121].

Inverse RIS design revolves around identifying the optimal configuration, spanning geometry and material properties, of RIS elements to achieve specific electromagnetic behaviors or functionalities. This approach involves tackling design challenges by pinpointing RIS parameters that align with desired electromagnetic responses, such as reflection, refraction, or absorption, for given input signals. As this methodology finds application in diverse domains, such as antenna design and metamaterial engineering, it aims to establish parameter sets that align with known material and structural properties to achieve precise and targeted outcomes. In [122], the authors offer a summary of progress in utilizing deep learning methods for the inverse design of metasurfaces. These advancements encompass various approaches, such as a target-focused deep learning inverse design method, the fuzzy inverse design approach, the freeform inverse design technique, and the application of machine learning to aid in inverse design.

The work in [123] presents a novel approach to radar-signature control using a diffusion-absorption hybrid metasurface. Unlike traditional random metasurfaces, this design offers flexible and precise scattering-field reduction by combining absorption and diffusion mechanisms. Multi-objective optimization is used to fine-tune the design, achieving a significant reduction in scattering-field across a wide bandwidth. Experimental results confirm the effectiveness of the approach, making it a promising solution for versatile wavefront manipulation in radar applications.

In [124], the authors introduce a method using machine learning to design versatile elastic metasurfaces for manipulating mechanical wavefronts efficiently. This approach reduces the need for trial-and-error and computationally intensive methods by establishing a mapping between input parameters and desired properties. The trained machine learning network accelerates the design process and can be extended to other types of wave manipulation.

In [125], the authors presents a synopsis of deep learning techniques for both inverse RIS design and RIS-assisted wireless systems. The work in [126] offer an overview of deep learning methods applied to the inverse design and optimization of RIS, aiming to achieve the desired EM response required for upcoming wireless networks. They mentioned that DL-based inverse design is versatile in accommodating various RIS unit structures. Deep Generative Models (DGMs) are particularly valuable due to their capacity to generate novel designs that have not been documented in existing literature. In the schematic representation of an inverse RIS design, machine learning algorithms acquire and generalize intricate electromagnetic relationships between the physical structure of the RIS (left column) and its corresponding spectral response (right column). This learning process is facilitated through training data, as depicted in Fig. 10. In [127], the authors demonstrate deep learning-based metasurface synthesis using conditional generative adversarial networks for inverse design and surrogate model prediction.

A novel approach called AMID is introduced in [128], which leverages artificial intelligence (AI) techniques, specifically a combination of a CNN, autoencoder, and optimized support vector machine (SVM), to enable inverse design of metasurface structures. Unlike traditional methods that require expert knowledge and simulation-based trial and error, AMID allows designers to input desired design targets directly, simplifying the design process. This method aims to utilize machine learning to establish connections between metasurface structures and electromagnetic properties, facilitating efficient and automated design. Compared to traditional methods, AMID reduces computational requirements and eliminates the need for extensive expertise, offering a more streamlined and accessible approach to metasurface design. The solution in [129] introduces a fast and accurate method for designing metasurfaces using transfer learning with a neural network. This approach can predict meta-atom phases with 90% accuracy. It is validated for 2D focusing and abnormal reflection functions, demonstrating high design accuracy in both simulations and experiments. This method streamlines functional metasurface design and can help create a versatile library of meta-atoms.

The approach in [130] suggests a new method for designing metasurfaces, which offer unique optical properties and practical advantages over traditional optical devices. Traditional metasurface design is slow and computationally intensive. The proposed approach uses deep tandem neural networks to speed up the design process. By connecting a pretrained forward prediction model with an inverse design model, this method allows for rapid creation of metasurface structures that meet specific spectral targets. It presents a promising way to streamline metasurface design and help researchers tailor optical properties more efficiently.

The work in [131] employs DNN to design metasurfaces for TE and TM polarized waves up to 45 GHz. It uses a confined output configuration to streamline the process, reducing calculations and speeding up learning. The DNN approach achieves a 92% accuracy and directly generates the desired metasurface structures, eliminating the need for time-consuming optimization. This simplifies inverse metasurface design, allowing designers to focus on their goals.

The work in [121] provides an overview of recent developments in using ML/DL for designing the physical layer of RIS. The authors in [121] discussed the basic principle of using machine learning for metasurface design as shown in Fig. 11. The figure outlines a two-part process using neural networks. Initially, datasets representing EM responses for various parameter combinations of a basic cylindrical structure are generated through forward solver algorithms. These datasets train a deep neural network designed to predict the EM response from given inputs, known as a forward network. Simultaneously, an inverse network is developed where the input is the desired EM response, and the output is the corresponding structural parameters. Furthermore, they present a diagram depicting the process of using physics-informed neural networks for metasurface design (Fig. 12). The datasets are generated by simulating the structure using a forward solver. However, by integrating fundamental physical principles like Maxwell's equations and electromagnetic boundary conditions into the neural networks, they manage to decrease the necessary dataset size. This reduction in dataset size results in a notable reduction in computational time.

5.5. Physics informed deep learning (PIDL)

Recent research has utilized deep learning to model complex relationships between metamaterials and their behaviors, often governed by Maxwell's equations in electromagnetics. This includes tackling inverse problems to determine desired properties. Physics-informed neural networks have emerged as a revolutionary approach, enabling the incorporation of physics directly into neural networks. PINNs bridge the gap between physics and machine learning, facilitating accurate predictions of complex physical phenomena with limited training data, advancing scientific computing and data-driven modeling.

Physics-informed neural networks are capable of solving general differential equations of the form:

$$\mathcal{F}(u(z);\gamma) = f(z), \quad z \in \Omega, \tag{59}$$

subject to boundary conditions:

$$\mathcal{B}(u(z)) = g(z), \quad z \in \partial\Omega, \tag{60}$$

where $z := [x_1, \ldots, x_{d-1}, t]$ represents the space–time coordinate vector, u is the unknown solution, γ denotes parameters related to the physics,

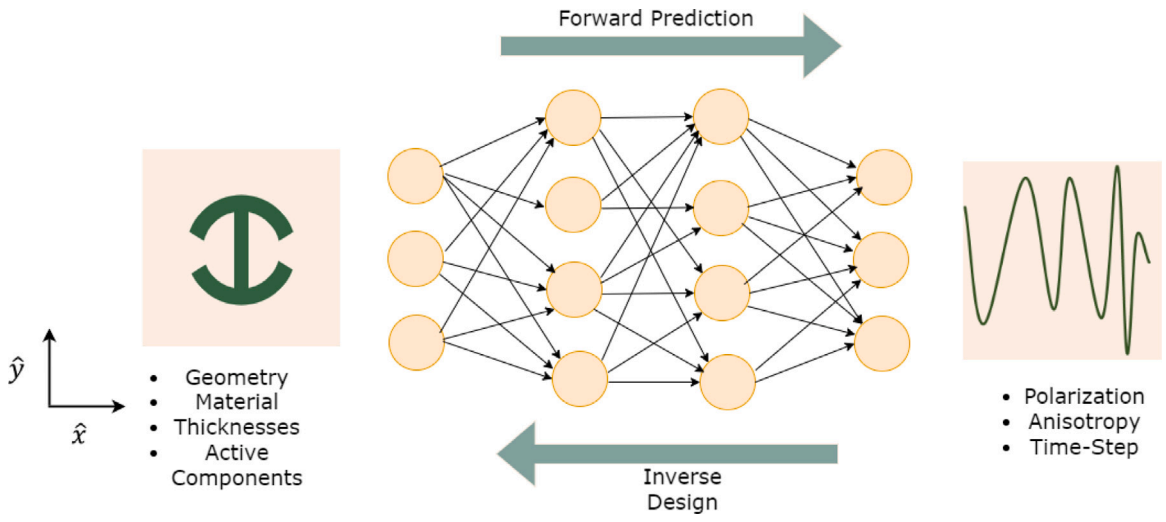

Fig. 10. Inverse RIS design: ML algorithms learn complex EM relationships between physical RIS structure (left) and spectral response (right) through training data (© The Author(s) 2022, [126]).

f is a function defining the problem's data, and $\mathcal{F}$ is a nonlinear differential operator. Besides, $\mathcal{B}$ represent the operator that signifies any unspecified initial or boundary conditions associated with the problem, and g denote the boundary function. In the PINN methodology, $u(z)$ is predicted computationally by a neural network parametrized by a set of parameters θ, leading to an approximation:

$$\hat{u}_\theta(z) \approx u(z;\theta), \tag{61}$$

where the notation $\hat{(\cdot)}_\theta$ signifies the neural network approximation performed using the parameters θ. In this context, the NN must learn to approximate the differential equations by determining the values of θ that define the NN. This is achieved by minimizing a loss function as follows:

$$\mathcal{L}(\theta) = \omega_{\text{physics}}\mathcal{L}_{\text{physics}} + \omega_{\text{Boundary}}\mathcal{L}_{\text{Boundary}} + \omega_{\text{data}}\mathcal{L}_{\text{data}} \tag{62}$$

$$\theta^* = \arg_\theta \min \mathcal{L}(\theta) \tag{63}$$

that includes three loss functions on:

1. **Labeled Data**

$$\mathcal{L}_{data}(\theta) = \text{MSE}_{\text{data}} = \frac{1}{N_{\text{data}}}\sum_{i=1}^{N_{\text{data}}} \left\|\hat{\mathbf{u}}_\theta(\mathbf{z}_i) - \mathbf{u}^*(i)\right\|^2 \tag{64}$$

2. **Boundary/ Initial Conditions**

$$\begin{aligned}\mathcal{L}_{\text{Boundary}}(\theta) &= \text{MSE}_{\text{Boundary}} \\ &= \frac{1}{N_{\text{Boundary}}}\sum_{i=1}^{N_{\text{Boundary}}} \left\|\mathcal{B}(\hat{\mathbf{u}}_\theta(\mathbf{z}))\right\|^2\end{aligned} \tag{65}$$

3. **Residual on PDE Equations**

$$\begin{aligned}\mathcal{L}_{\text{physics}}(\theta) = \text{MSE}_{\text{physics}} &= \frac{1}{N_{\text{physics}}}\sum_{i=1}^{N_{\text{physics}}} \left[\mathcal{F}(\hat{\mathbf{u}}_\theta(\mathbf{z}_i)) - \mathbf{r}_i\right]^2 \\ &= \frac{1}{N_{\text{physics}}}\left\|\sum_{i=1}^{N_{\text{physics}}} \left[\mathbf{r}_\theta(\mathbf{u}_i) - \mathbf{r}_i\right]\right\|\end{aligned} \tag{66}$$

The PINN's building block is shown in Fig. 13. In Eq. (62), the initial term $\mathcal{L}_{\text{physics}}$ signifies the loss resulting from a discrepancy with the governing differential equations $\mathcal{F}$. It enforces the differential equation $\mathcal{F}$ at the collocation points, which can be uniformly or unevenly distributed over the domain Ω of Eq. (59). The remaining two losses aim to accommodate the known data within the neural network. The loss arising from a mismatch with the data (i.e., the measurements of u) is represented as $\mathcal{L}_{\text{data}}(\theta)$. The second term typically compels $\hat{u}_\theta$ to match the measurements of u at provided points (z, u^*), which can be either synthetic data or actual measurements. The weight ω_{data} can adjust for the quality of such measurements. The other term is the loss due to a mismatch with the boundary or initial conditions, $\mathbf{B}(\hat{\mathbf{u}}_\theta) = g$ from Eq. (59) [132–137].

In the context of RIS, PIDL can involve training deep neural networks to approximate the behavior of electromagnetic fields based on known physics principles, such as Maxwell's equations. By incorporating these physical constraints, the model can better generalize to unseen scenarios and provide insights into the underlying physical phenomena.

In [133], the authors employ PINNs to achieve the "homogenization" of distinct meta atom designs, specifically a disc and a split-ring resonator (SRR). The objective is to predict their effective electric permittivity distributions inversely based on simulated scattered electric fields. Homogenization aims to ensure that the resultant electric fields match those produced by a hypothetical continuous medium with equivalent material properties, aligning with the well-established 'effective medium theory'.

SRRs, which combine inductive and capacitive elements, function as oscillators. When the split-ring resonates with incident light's frequency, significant local fields occur in the gap region. The enhancement of these local fields is governed by the effective electric permittivity coefficient of these non-magnetic metamaterials. Notably, this work marks the first instance of showcasing the impact of PINN-based inverse model accuracy on the intricate and resonant behavior of SRRs. The electric field is determined as a function of the metasurface's spatial coordinates (denoted as x and y) by employing the Helmholtz wave equation in the frequency domain for linear, non-magnetic, weakly inhomogeneous materials. This equation is presented as follows:

$$\nabla \times \nabla \times E(x,y) - k_0^2 \varepsilon_r E(x,y) = 0, \tag{67}$$

where, E represents the electric field, k_0 is the free space wave number, and ε_r signifies the relative electric permittivity distribution. Besides, $\nabla \times \nabla$ denotes the Laplacian operator. The mean squared error loss function, guided by physics principles, is derived from the residue of the Helmholtz equation and is computed as:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N} \left|F(x_i, y_i)\right|^2. \tag{68}$$

where, $F(x_i, y_i)$ is defined as:

$$F(x_i, y_i) = \frac{\partial^2 E(x_i, y_i)}{\partial x^2} + \frac{\partial^2 E(x_i, y_i)}{\partial y^2} + k_0^2 \varepsilon_r E(x_i, y_i). \tag{69}$$

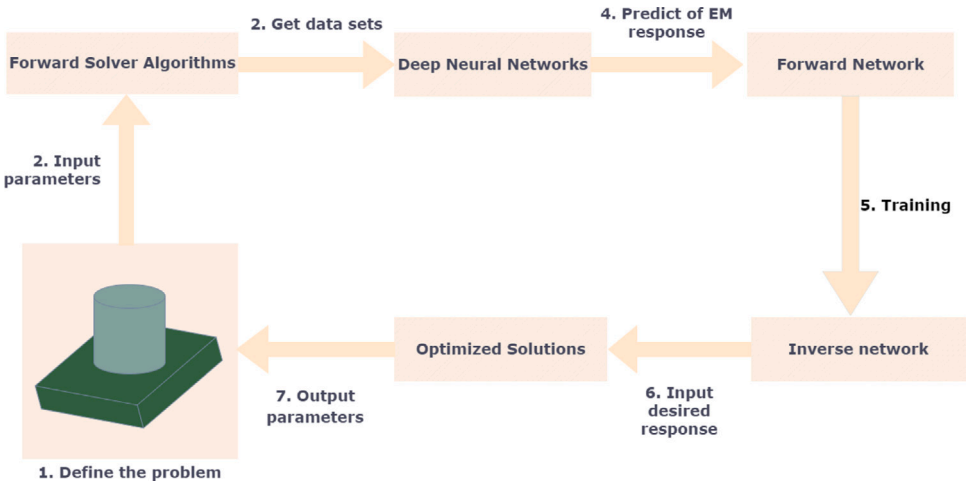

Fig. 11. Basic principle of using machine learning for metasurface design. Detailed steps are shown in the flowchart, including define the problem, collect data sets, pre-process data, train the model, validate the model, train inverse network, input desired response, iterate and optimization, and achieve desired performance (© The Author(s) 2023, [121]).

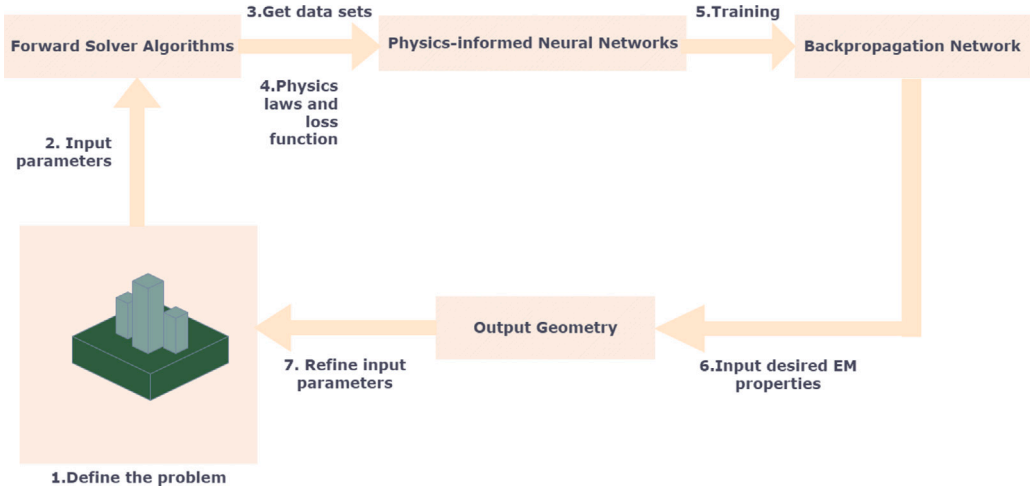

Fig. 12. Basic principle of using Physics-Informed Neural Networks for metasurface design. Detailed steps are demonstrated in the flowchart, including define the problem, create the training data, define physics principle and loss function, define the PINN architecture, validate the network, train backpropagation networks, input desired EM response, iterate the design process, and obtain desired metasurface designs (© The Author(s) 2023, [121]).

Eq. (68) is utilized to train the DNN and optimize its architecture to generate an effective spatial permittivity profile, resulting in an electric field that matches that of the original metasurface design. Fully connected feed-forward neural network, as shown in Fig. 14, with an input layer comprising of three neurons, namely, x, y, and normalized relative permittivity ϵ_r is employed. This is followed by 4 hidden layers of 250 neurons each and an output layer with a single neuron for the predicted E-field distribution $E(x, y)$. The input–output relationship of the network is denoted by $E(x, y, \theta)$ as a surrogate for the PDE solution to the Helmholtz equation, wherein θ is a vector containing the trainable weights and biases of the network.

One more paper [138] utilizes physical optics techniques to derive the pathloss expression for RIS configured to reflect waves toward a receiver in the far-field. The study highlights the limitations of using infinitely sized RISs and emphasizes the importance of coherent beamforming achieved through sub-wavelength-sized elements with unique phase shifts. The ideal phase shifts for beam creation follow the generalized Snell's law.

The work in [139] demonstrates the efficacy of a physics-guided deep learning approach with a tandem architecture (Fig. 15) for the efficient and interpretable design of metasurfaces in a supervised learning context. By harnessing neural networks' capabilities, they uncover the strong connection between device configuration and optical behavior. To address challenges posed by using deep neural networks for inverse problems, such as non-interpretability and non-uniqueness, they introduce a tandem architecture incorporating a physics-based constraint. Integrating both forward and inverse models in this architecture mitigates one-to-many prediction issues. By adding a physics-driven penalty term, their model generates interpretable and physically consistent predictions, while also enhancing generalizability and reducing data requirements. The physics penalty term is derived analytically by simplifying the metamaterial structure into a stratified medium, primarily applicable for sub-wavelength dimensions. To adapt the model to larger dimensions, they adjusted the physics-based loss function or employed a physics-guided deep learning architecture for prediction. The proposed model achieves high accuracy. The physics-guided DL framework that have developed to solve problem is shown in Fig. 16.

Tandem model structure consists of the forward EM equation approximated by a deep neural network and the inverse DL model. The networks are appended end-to-end so that the output of the inverse model is fed into the forward DL model. The inverse problem involves predicting the design based on a given response. In Fig. 16, the input of the pretrained forward DL model is connected to the output of the inverse model. This predicted design is then passed into the pretrained forward model to assess how well it can reconstruct

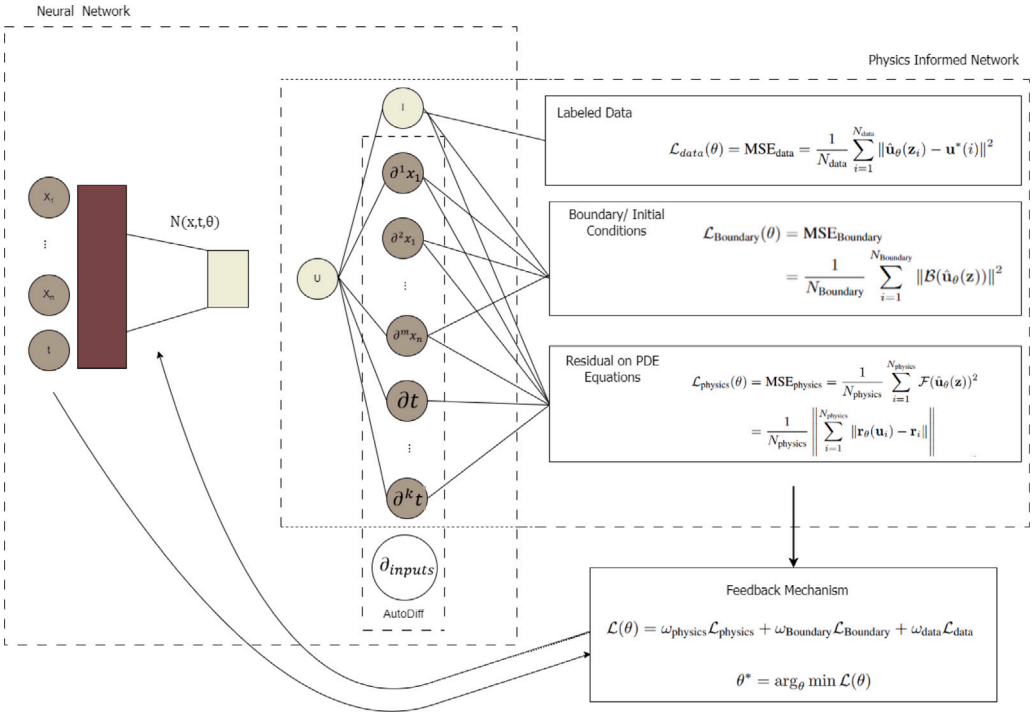

Fig. 13. The constituents of PINNs. PINNs are composed of residual terms derived from differential equations, as well as initial and boundary conditions. The network takes inputs, transforms them into field outputs (denoted as u), and is defined by parameters θ. The second component is the physics-informed network, which computes derivatives from the output field u based on provided equations. It also evaluates boundary and initial conditions (unless they are hardcoded in the neural network) and incorporates labeled data observations if available. The feedback mechanism is the final step, minimizing the loss with an optimizer and a specified learning rate to determine the NN parameters θ (© The Author(s) 2023, [137]).

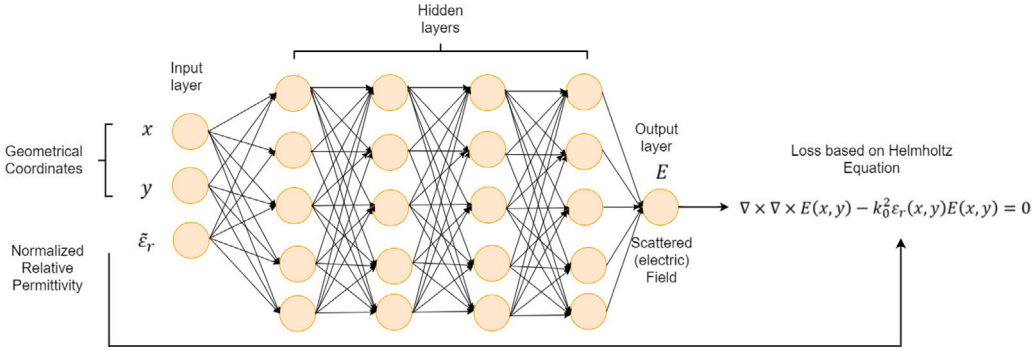

Fig. 14. Representative diagram of the physics-informed neural network model with 6 layers. Each hidden layer contains 250 neurons. A physics-driven loss function based on the Helmholtz equation is used to train the model (© The Author(s) 2020, [133]).

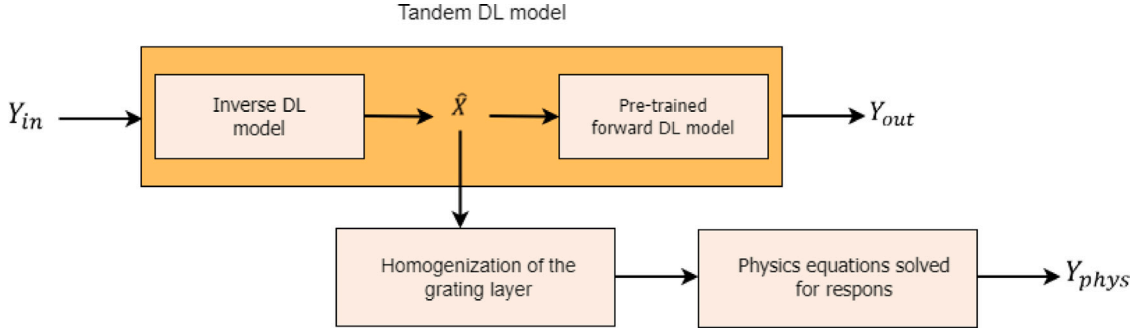

Fig. 15. Overview of the components of the physics-guided DL model. Y_{in} is the input optical response and the design, $\hat{X}$, predicted from the inverse model. The predicted design is then fed into a pre-trained forward model to ensure proper reconstruction of the optical response, Y_{out}. The structure of $\hat{X}$ is then simplified according to the homogenization principle and the physics equations are solved to obtain Y_{phys} that ensures physics guided learning (© The Author(s) 2023, [139]).

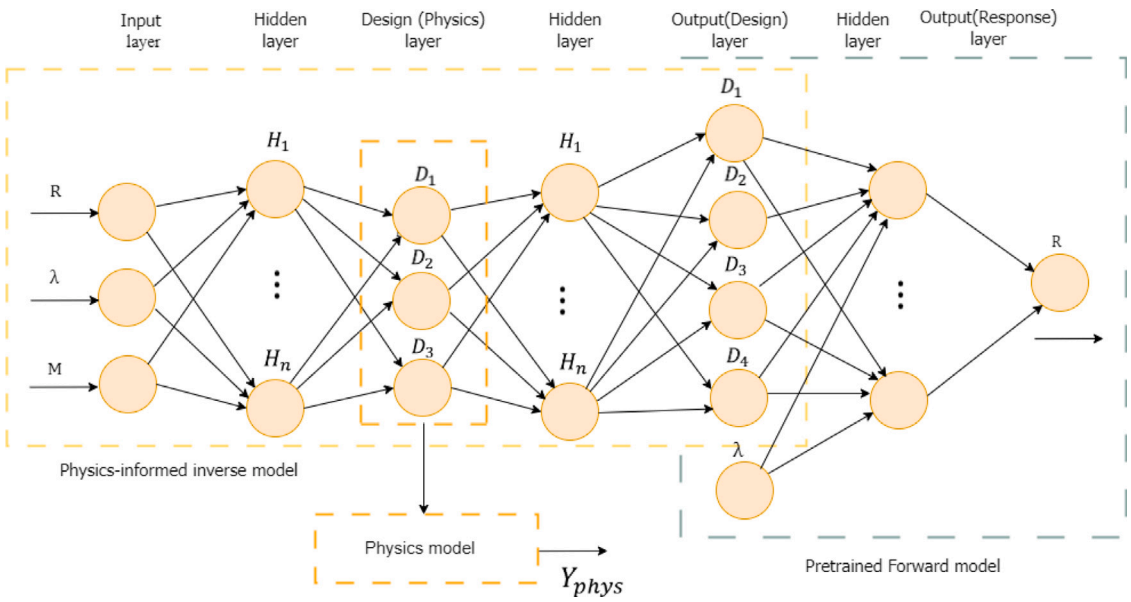

Fig. 16. Physics-informed DL architecture (PIA). The "Design (Physics) layer" includes intermediate variables that are consistent with the underlying physics. These variables are calculated based on the design parameters that result in a certain response, as predicted by the EM equations for the simplified homogenized structure. The "Output (Design) layer" then uses these intermediate variables to predict the original complex design parameters (© The Author(s) 2023, [139]).

the optical response. The entire structure, with the weights of the forward model fixed, undergoes training to minimize two key errors: the design prediction error (the deviation between the predicted design, denoted as $\hat{X}$, and the ground truth design, denoted as X) and the reconstruction error (the deviation between the response reconstructed by the predicted design, denoted as Y_{out}, and the ground truth response, denoted as Y_{in}).

Physics can be integrated into DL models through various approaches, including modifying the loss function, using residual modeling, initializing parameters, or adjusting the model architecture. In this study, the incorporation of physics was focusd on the loss function and DL architecture. By training a physics-guided DL model, the learning process adapts model parameters to align with physics principles, either for final or intermediate layer outputs. The inclusion of physics is manifested in Eq. (70) through a penalty term. This term is integrated either as a penalty for the final output layer or for the intermediate layer in the physics-guided loss function and the physics-guided architecture, respectively. This penalty corresponds to the disparity between the actual optical response, Y_{in}, and the response obtained through solving electromagnetic equations for a simplified design structure, denoted as Y_{phys}. The equation representing this incorporation is as follows:

$$\begin{aligned}\text{Objective} =& \omega_{\text{obs}_1}\text{Loss}(X;\hat{X})\\ &+ \omega_{\text{obs}_2}\text{Loss}(Y_{\text{in}};Y_{\text{out}})\\ &+ \omega_{\text{phys}}(X)\text{Loss}(Y_{\text{in}};Y_{\text{phys}}),\end{aligned} \tag{70}$$

In this context, the parameters ω_{obs_1}, ω_{obs_2}, and ω_{phys} represent the weighting of each loss term, acting as hyperparameters. The values for ω_{obs_1} and ω_{obs_2} are meticulously determined through grid search hyperparameter tuning to optimize the model's performance on the validation set. Simultaneously, ω_{phys} is directly proportional to $\frac{\lambda_0}{p}$, ensuring heightened emphasis on the physics-based loss function term when $\lambda_0 \gg p$, where the effective medium theory of homogenization holds true. Here, λ_0 denotes the wavelength, and p is a parameter associated with the metamaterial structure. The loss term in the equation corresponds to the mean absolute error (MAE) loss.

5.6. Lessons learned

- Model-based optimization techniques and heuristic algorithms often suffer from high computational complexity, scalability issues, and reliance on simplifying assumptions, limiting their effectiveness in solving complex EM-aware RIS optimization problems.
- In contrast, ML/DL methods offer advantages such as adaptability to dynamic environments, capability to handle large-scale optimization problems, and ability to learn complex patterns and relationships from data, making them promising for addressing the challenges of EM-aware RIS optimization in real-world communication scenarios.
- ML/DL methods has widely used for MTS design. However, dynamic optimization of EM-aware RIS-aided communications employing ML/DL algorithm rarely presented in the literatures.

6. Research gaps and future research

In this section, our goal is to delineate the areas of research that require further exploration and potential avenues for future research.

6.1. Electromagnetic information theory (EIT) in RIS-aided communications

In the realm of wireless communications, electromagnetic theory and information theory stand as fundamental principles, linked by antenna theory and wireless propagation channel modeling. While these concepts evolved separately until the era of 5G networks, the advent of 6G communication networks – spanning across space, air, ground, and sea – necessitates seamless coverage in three-dimensional space. This demands continuous acquisition of CSI and channel capacity calculation at any location and time. Electromagnetic theory enables precise modeling of how signals propagate through space, while information theory quantifies the capacity of these channels to carry information. By integrating these two theories, EIT provides a comprehensive framework that not only models the electromagnetic field distribution but also calculates the communication performance, allowing for more accurate and dynamic acquisition of CSI in 6G networks. Additionally, advanced 6G technologies like ultra-massive MIMO and holographic MIMO call for integrated modeling of antennas and wireless propagation environments. To address these challenges, the integration of electromagnetic theory, information theory, antenna theory, and wireless propagation channel modeling has become essential, leading to the emergence of EIT [140].

The classic electromagnetic theory, founded on Maxwell's equations, operates in continuous time and space, mainly addressing macroscopic fields. However, at microscopic scales, local field fluctuations become significant, especially in scenarios like high-mobility millimeter wave-aided wideband environments. Traditional Shannon's information theory, tailored for simple channels, lacks accuracy in complex, non-stationary settings demanded by emerging 6G technologies such as integrated sensing and communication (ISAC), RIS, massive MIMO, and holographic MIMO. Integrating electromagnetic theory with information theory expands the latter to encompass non-stationary channels, facilitating the evolution from discrete-space single-user information theory to continuous-space multi-user/network information theory. Continuous-space electromagnetic channel models, coupled with optimal antenna design, enhance channel capacity. By studying the mutual information between incident, reflected, transmitted, and scattered waves, Maxwell's equations find application in novel technologies like RIS, enabling precise performance and capacity descriptions. The concept of information metamaterials further integrates electromagnetic theory and information theory, aiming to manipulate electromagnetic waves within metamaterial structures for improved spectral efficiency and reduced system costs. However, conventional signal processing methods often fail to accurately characterize electromagnetic fields, underscoring the need for innovative approaches, such as leveraging RIS to understand electromagnetic wave interactions and realize end-to-end electromagnetic information transmission mechanisms in communication systems [140]. In [141], the concept of the digital metasurface, known as the information metasurface (IM), is introduced. The paper delves into its applications in wireless communications. IMs represent a significant paradigm shift, expanding the scope of metasurface research from the realm of physics into the realm of information.

Future research endeavors should aim to delve deeper into the possibilities and applications of EIT in RIS scenarios. This exploration could encompass theoretical developments, practical implementations, and experimentation to fully harness the potential benefits of EIT [142].

6.2. Machine learning applications for optimization of EM-based RIS-aided communications

Machine learning techniques for optimizing RIS-aided communications have been extensively examined in [32–35], all without the inclusion of physically consistent EM models. Machine learning techniques have yet to be thoroughly explored for optimizing RIS-aided communications from an EM perspective. While there are massive number of papers focusing on the inverse design of meta-surfaces using machine learning techniques (as discussed in Section 5), the optimization of EM-based RIS-aided communication and BD-RIS-assisted communication systems as a whole has been relatively untouched in the current literature. Furthermore, the integration of machine learning techniques into EIT-based RIS-aided communication systems is an uncharted territory in research. This represents a significant research gap that holds immense potential. Machine learning algorithms can play a pivotal role in enhancing the efficiency and accuracy of RIS-aided communications using EIT.

6.3. Green communications

In the evolving 5G and future 6G landscapes, the proliferation of network nodes is causing a surge in energy consumption and carbon emissions. Resource allocation, network planning and deployment, energy harvesting and transmission, and hardware solutions are the four areas under which the most popular methods for boosting the EE of wireless networks are categorized.

Energy harvesting (EH), which includes radio frequency (RF) energy harvesting, offers the dual advantage of supporting SWIPT while harnessing interference signals for energy collection. This concept has been explored in [99], as detailed in Section 4. When combined with SWIPT technology, RISs have the potential to enhance EH efficiency. Additionally, when integrated with orbital angular momentum (OAM) technology, RIS can serve the purpose of reflecting OAM waves obstructed by obstacles, establishing a direct path for OAM-based SWIPT. As a result, RIS-assisted OAM-SWIPT emerges as a promising technology for future wireless communication, offering energy and spectrum efficiency [143]. The works in [143–145] have delved into using OAM for RIS-assisted communications. Nonetheless, it is clear that additional research into the combined utilization of RIS and OAM technologies offers considerable potential for future advancements. To enable RIS to pass OAM waves effectively, RIS should be able to manipulate the electromagnetic field to accommodate the unique properties of OAM waves.

EM-aware resource allocation in RIS-assisted communications is another gap that should be addressed in future research. As elucidated in Section 4, the literature has offered several EM-aware problem formulations for RIS-assisted communications, addressing various aspects of this technology. These formulations have contributed significantly to our understanding and optimization of RIS deployments in diverse scenarios. However, one notable research gap pertains to resource allocation problems within the context of wideband communications. While existing studies have primarily focused on narrowband scenarios, the challenges and opportunities associated with wideband communication systems have yet to receive comprehensive attention. This represents a crucial area for further investigation, as wideband communications are increasingly prevalent in modern wireless networks, and tailored resource allocation strategies can significantly impact their performance and efficiency. In this regard, incorporating physical models of RIS systems into problem formulations can offer valuable insights and solutions.

By integrating physical models, researchers can develop resource allocation algorithms that are not only EM-aware but also consider the specific characteristics of RIS elements and their impact on wideband communication channels. This holistic approach, combining EM-awareness with physical realism, holds the potential to yield more robust and efficient resource allocation strategies for wideband RIS-aided communications.

6.4. Optical communications

The integration of optical communication with RIS-assisted communication presents a promising avenue for enhancing network performance and addressing modern communication challenges. Recent advancements, such as the introduction of optical simultaneous transmission and reflection reconfigurable intelligent surface (OSTAR-RIS), underscore the potential of this integration [146]. Authors in [147] seeks to bridge this gap by offering a comprehensive tutorial on indoor visible light communication (VLC) systems leveraging RIS technology. It delves into the fundamentals of VLC and RISs, explores the potential applications of optical RISs in mitigating line-of-sight blockages and device orientation issues, and proposes avenues for future research on integrating optical RISs with emerging technologies. However, there is currently limited exploration of RISs in optical wireless communication (OWC) systems. Novel BD-RIS architectures can be explored using the integration of optical communication. This integration can be as follows:

- Physical Consistency: Optical communication offers high-speed data transmission using light waves, which can be leveraged to accurately model the behavior of RIS elements. By considering the physical properties of light propagation, such as reflection, refraction, and diffraction, in conjunction with the characteristics of RIS surfaces, a more accurate and physically consistent model of RIS can be developed. This model can capture the intricate interactions between light waves and RIS elements, leading to improved performance predictions and optimization strategies.

- Innovative Architectures: The integration of optical communication with RIS-assisted communication opens up new possibilities for designing innovative RIS architectures, such as BD-RIS. By exploring the synergy between optical components and RIS elements, novel architectures that leverage both optical and electromagnetic properties can be proposed. For example, incorporating optical switches or modulators into RIS structures can enable dynamic reconfiguration of RIS elements, leading to enhanced flexibility and adaptability in network optimization.

6.5. Quantum communications

Quantum information science is rapidly advancing, promising enhanced security through quantum communication, transformative computational power with quantum computing, and revolutionary sensing capabilities. Digital metamaterials, also known as information metamaterials, enable the manipulation of digital data carried by electromagnetic waves, potentially revolutionizing information processing and high-performance wireless information systems. These metamaterials utilize binary coding for real-time control of electromagnetic waves, bridging the physical and digital realms. They offer significant potential for both fundamental science and engineering applications in the nonclassical electromagnetic field, closely tied to quantum information science. Electromagnetic metamaterials can further advance these quantum-inspired developments, giving rise to quantum metamaterials that extend classical concepts for manipulating and detecting nonclassical electromagnetic waves [148].

Viewing physical models of RIS from a quantum perspective introduces a fascinating and innovative dimension to the understanding of RIS. Quantum physics, with its principles of superposition, entanglement, and uncertainty, can provide new insights and possibilities in the design and operation of RIS.

To fully unlock the potential of RIS-enabled smart EM environments, it is essential to quickly identify the optimal configurations of the RIS to achieve specific objectives. While mathematical tools like generalized Snell's law aid in managing anomalous reflection, there is a lack of semi-analytical solutions for certain EM functionalities such as multi-beamforming, energy focusing, and diffusive scattering, as well as wireless functions like spatial diversity, data throughput, and physical-layer security. As a result, various optimization methods for RIS have been proposed, including genetic algorithms, impedance-based synthesis, electromagnetic inversion, machine learning, and dynamic optimization. However, despite these advancements, optimizing RIS configurations remains computationally challenging due to the vast number of possibilities, the complexity of EM scattering environments, and processing time constraints in wireless systems.

Recent progress in QC offers a promising solution to tackle these computational challenges. For EM applications, quantum algorithms like the quantum Fourier Transform have been applied to antenna array synthesis, while the Harrow/Hassidim/Lloyd (HHL) algorithm is used for characterizing interconnects. Variational quantum algorithms utilizing the finite difference method have also been developed for calculating waveguide modes. In the realm of wireless networks, quantum computing facilitates tasks such as multi-user MIMO detection, processing, and optimal resource allocation for 6G wireless networks [149].

Quantum communications and quantum computing for RIS-aided communications have been studied in some recent papers [149–152]. The possibilities of combining EM models-aware RIS-aided communication and quantum knowledge and computing is an interesting domain to be explored.

6.6. RIS-enhanced semantic communications with physically-consistent RIS models

Semantic communication, which prioritizes conveying the meaning of information rather than exact bit reconstruction, has garnered significant attention in recent years [153]. This approach revolutionizes the way we transmit and interpret data, focusing on the context and intent behind the information exchange. Recent studies have explored the integration of RIS in enhancing semantic communications, aiming to leverage their potential in shaping the electromagnetic environment for improved data transfer [153–156].

Despite these advancements, a promising avenue that remains largely unexplored is the incorporation of physically-consistent RIS models. These models, grounded in the principles of electromagnetic wave interactions, hold the key to achieving greater precision and efficiency in the context of RIS-enhanced semantic communications. By integrating EM models of RIS into the semantic communication framework, we can attain a deeper understanding of how RIS elements interact with electromagnetic waves and, in turn, impact the semantic content of the communication.

Such an integration offers numerous opportunities. It enables us to adapt RIS configurations dynamically, responding to real-time changes in the electromagnetic environment. This adaptability ensures that the conveyed meaning remains intact, even in the face of fluctuating signal quality and interference. Furthermore, physically-consistent RIS models facilitate the optimization of data transfer, allowing semantic content to be communicated with greater clarity and reliability.

6.7. Multi-objective optimization with physically-consistent RIS models

Developing a multi-objective optimization framework for RIS that integrates physically consistent models is crucial for enhancing RIS-based communication systems. This framework must balance optimizing communication capacity, minimizing interference, and conserving energy while adhering to the physical constraints and characteristics of RIS technology to ensure practical feasibility. Although significant strides have been made in multi-objective optimization for RIS, as noted in Refs. [157–162], there remains a notable opportunity to advance optimization strategies by incorporating EM phenomena.

In this context, recent research has expanded the scope of optimization by integrating RIS with ISAC systems. ISAC represents a paradigm shift that combines communication and sensing objectives into a unified framework through co-design approaches. This approach introduces a new dimension to the optimization problem, making it inherently multi-objective. Advanced signal processing techniques are being developed to design dual-functional waveforms that optimize both communication metrics—such as SINR and multi-user interference (MUI) and sensing metrics such as beam-pattern mean squared error (MSE), received echo power/signal-to-noise ratio (SNR), and Cramér–Rao bound (CRB) [163].

Further research can explore the advantages of EM-aware optimization in RIS-assisted ISAC systems. By incorporating a deeper understanding of EM phenomena into the optimization framework, this research avenue promises to unlock the full potential of RIS. It is likely to lead to more efficient and sophisticated applications of RIS in future wireless communication systems, paving the way for advanced solutions that better meet the demands of modern communication and sensing requirements.

6.8. Stacked intelligent metasurfaces aided wireless communication

Stacked intelligent metasurfaces (SIMs) represent a significant advancement in wireless communications by leveraging multiple layers of RIS to enhance signal processing. The study by Matteo and Clerckx [164] introduces a novel model utilizing multiport network theory to address SIM-assisted communications. Their approach provides a

physically consistent framework for understanding the EM coupling between transmitting antennas and metasurface layers, accurately predicting the transfer function of SIMs by considering the cascading effects of multiple metasurface layers on EM wave propagation. Despite this progress and some other limited recent literature [164–168], the exploration of SIMs remains limited, indicating a need for further investigation, particularly into their physical characteristics and modeling techniques.

7. Conclusion

In conclusion, the rapid evolution of RISs represents a remarkable leap forward in the realm of wireless communication systems. These adaptive metasurfaces, equipped with the ability to dynamically manipulate EM waves, hold great promise for enhancing network capacity, coverage, and efficiency. Throughout this survey, we have traversed the landscape of RIS-aided communication systems, exploring the multitude of EM-aware challenges and optimization avenues they present.

One key takeaway from our exploration is the pivotal role played by physically-consistent EM models. These models serve as the foundation for developing robust and efficient RIS designs, enabling precise control over EM wave propagation. By grounding our designs in these models, we can achieve superior performance and ensure the reliability of RIS-aided systems.

Furthermore, we have delved into the realm of machine learning approaches for RIS optimization. This exciting field has witnessed significant advancements, with recent developments showcasing the potential to simplify and expedite the design process. Machine learning techniques have the capacity to learn intricate relationships within RIS structures and EM responses, offering valuable tools for RIS engineers.

As we move forward, the integration of physically-consistent EM models with machine learning approaches promises to drive innovation in RIS-aided communications. These two paradigms, when harnessed synergistically, enable us to tackle complex EM-aware problems and optimize RIS performance effectively. However, it is important to remain cognizant of the challenges that lie ahead, including scalability, hardware constraints, and real-world deployment issues.

In essence, this survey paper has illuminated the path toward unlocking the full potential of RIS-aided communication systems. By embracing physically-consistent EM models and harnessing the power of machine learning, we can shape the future of wireless networks, ushering in an era of unprecedented capacity, coverage, and efficiency.

CRediT authorship contribution statement

S. Bidabadi: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **M.A. Ouameur:** Writing – review & editing, Supervision, Conceptualization. **M. Bagaa:** Writing – review & editing, Visualization, Supervision. **D. Massicotte:** Writing – review & editing, Validation. **F.D. Figueiredo:** Writing – review & editing. **A. Chaaban:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT to check the grammar of the original manuscript and improve readability. After using this tool/service, the authors reviewed and edited the content as needed, taking full responsibility for the publication's content.

References

[1] Ahmad A.A. Solyman, Khalid Yahya, Key performance requirement of future next wireless networks (6G), Bull. Electr. Eng. Inform. 10 (6) (2021) 3249–3255.

[2] Keysight Technologies, Next-Generation Wireless: A Guide to the Fundamentals of 6G, Keysight Technologies, USA, 2023, June 13, 2023, 1–39.

[3] Walid Saad, Mehdi Bennis, Mingzhe Chen, A vision of 6G wireless systems: Applications, trends, technologies, and open research problems, IEEE Netw. 34 (3) (2019) 134–142.

[4] Samaneh Bidabadi, Messaoud Ahmed Ouameur, Miloud Bagaa, Daniel Massicotte, Energy efficient resource allocation for re-configurable intelligent surface-assisted wireless networks, EURASIP J. Wireless Commun. Networking 2023 (1) (2023) 1–22.

[5] Qingqing Wu, Beixiong Zheng, Changsheng You, Lipeng Zhu, Kaiming Shen, Xiaodan Shao, Weidong Mei, Boya Di, Hongliang Zhang, Ertugrul Basar, Lingyang Song, Marco Di Renzo, Zhi-Quan Luo, Rui Zhang, Intelligent surfaces empowered wireless network: Recent advances and the road to 6G, 2024, URL https://arxiv.org/abs/2312.16918.

[6] Moazzam Shah Bukhari Syed, Hafiz Muhammad Attaullah, Sundus Ali, Muhammad Imran Aslam, Wireless communications beyond antennas: The role of reconfigurable intelligent surfaces, Eng. Proc. 32 (1) (2023) 10.

[7] Xiaodan Shao, Changsheng You, Rui Zhang, Intelligent reflecting surface aided wireless sensing: Applications and design issues, 2023, arXiv preprint arXiv:2302.05864.

[8] Yuanwei Liu, Xiao Liu, Xidong Mu, Tianwei Hou, Jiaqi Xu, Marco Di Renzo, Naofal Al-Dhahir, Reconfigurable intelligent surfaces: Principles and opportunities, IEEE Commun. Surv. Tutor. 23 (3) (2021) 1546–1577.

[9] Shimin Gong, Xiao Lu, Dinh Thai Hoang, Dusit Niyato, Lei Shu, Dong In Kim, Ying-Chang Liang, Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey, IEEE Commun. Surv. Tutor. 22 (4) (2020) 2283–2314, http://dx.doi.org/10.1109/COMST.2020.3004197.

[10] Ertugrul Basar, Marco Di Renzo, Julien de Rosny, Merouane Debbah, Mohamed-Slim Alouini, Rui Zhang, Wireless communications through reconfigurable intelligent surfaces, 2019.

[11] Xiaojun Yuan, Ying-Jun Angela Zhang, Yuanming Shi, Wenjing Yan, Hang Liu, Reconfigurable-intelligent-surface empowered wireless communications: Challenges and opportunities, IEEE Wirel. Commun. 28 (2) (2021) 136–143.

[12] Marco Di Renzo, Alessio Zappone, Merouane Debbah, Mohamed-Slim Alouini, Chau Yuen, Julien De Rosny, Sergei Tretyakov, Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead, IEEE J. Sel. Areas Commun. 38 (11) (2020) 2450–2525.

[13] Yuanwei Liu, Xiao Liu, Xidong Mu, Tianwei Hou, Jiaqi Xu, Marco Di Renzo, Naofal Al-Dhahir, Reconfigurable intelligent surfaces: Principles and opportunities, IEEE Commun. Surv. Tutor. 23 (3) (2021) 1546–1577, http://dx.doi.org/10.1109/COMST.2021.3077737.

[14] George C. Alexandropoulos, Geoffroy Lerosey, Merouane Debbah, Mathias Fink, Reconfigurable intelligent surfaces and metamaterials: The potential of wave propagation control for 6G wireless communications, 2020.

[15] Christopher L. Holloway, Edward F. Kuester, Joshua A. Gordon, John O'Hara, Jim Booth, David R. Smith, An overview of the theory and applications of metasurfaces: The two-dimensional equivalents of metamaterials, IEEE Antennas Propag. Mag. 54 (2) (2012) 10–35, http://dx.doi.org/10.1109/MAP.2012.6230714.

[16] Syed S. Bukhari, J. Vardaxoglou, William Whittow, A metasurfaces review: Definitions and applications, Appl. Sci. 9 (13) (2019) 2727.

[17] Mayurkumar Ladumor, Shreyas Charola, Shobhit K. Patel, Vigneswaran Dhasarathan, Graphene-based c-shaped metasurface broadband solar absorber, in: Physics and Simulation of Optoelectronic Devices XXVIII, vol. 11274, SPIE, 2020, pp. 7–12.

[18] Adnan Ali, Anirban Mitra, Brahim Aïssa, Metamaterials and metasurfaces: A review from the perspectives of materials, mechanisms and advanced metadevices, Nanomaterials 12 (6) (2022) 1027.

[19] Zhuqi Li, Yaxiong Xie, Longfei Shangguan, R. Ivan Zelaya, Jeremy Gummeson, Wenjun Hu, Kyle Jamieson, Programmable radio environments with large arrays of inexpensive antennas, GetMobile: Mob. Comput. Commun. 23 (3) (2020) 23–27.

[20] Chongwen Huang, Sha Hu, George C. Alexandropoulos, Alessio Zappone, Chau Yuen, Rui Zhang, Marco Di Renzo, Merouane Debbah, Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends, IEEE Wirel. Commun. 27 (5) (2020) 118–125.

[21] Marco Di Renzo, Konstantinos Ntontin, Jian Song, Fadil H. Danufane, Xuewen Qian, Fotis Lazarakis, Julien De Rosny, Dinh-Thuy Phan-Huy, Osvaldo Simeone, Rui Zhang, et al., Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison, IEEE Open J. Commun. Soc. 1 (2020) 798–807.

[22] Marco Di Renzo, Alessio Zappone, Merouane Debbah, Mohamed-Slim Alouini, Chau Yuen, Julien de Rosny, Sergei Tretyakov, Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and road ahead, 2020.

[23] Haris Gacanin, Marco Di Renzo, Wireless 2.0: Toward an intelligent radio environment empowered by reconfigurable meta-surfaces and artificial intelligence, IEEE Veh. Technol. Mag. 15 (4) (2020) 74–82, http://dx.doi.org/10.1109/MVT.2020.3017927.

[24] Qingqing Wu, Rui Zhang, Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network, IEEE Commun. Mag. 58 (1) (2019) 106–112.

[25] Cunhua Pan, Hong Ren, Kezhi Wang, Jonas Florentin Kolb, Maged Elkashlan, Ming Chen, Marco Di Renzo, Yang Hao, Jiangzhou Wang, A. Lee Swindlehurst, et al., Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions, IEEE Commun. Mag. 59 (6) (2021) 14–20.

[26] Saber Hassouna, Muhammad Ali Jamshed, James Rains, Jalil ur Rehman Kazim, Masood Ur Rehman, Mohammad Abualhayja, Lina Mohjazi, Tei Jun Cui, Muhammad Ali Imran, Qammer H Abbasi, A survey on reconfigurable intelligent surfaces: Wireless communication perspective, IET Commun. 17 (5) (2023) 497–537.

[27] Ana Díaz-Rubio, Sergei A. Tretyakov, Macroscopic modeling of anomalously reflecting metasurfaces: Angular response and far-field scattering, IEEE Trans. Antennas and Propagation 69 (10) (2021) 6560–6571, http://dx.doi.org/10.1109/TAP.2021.3076267.

[28] Özgecan Özdogan, Emil Björnson, Erik G. Larsson, Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling, IEEE Wirel. Commun. Lett. 9 (5) (2020) 581–585, http://dx.doi.org/10.1109/LWC.2019.2960779.

[29] Davide Dardari, Communicating with large intelligent surfaces: Fundamental limits and models, IEEE J. Sel. Areas Commun. 38 (11) (2020) 2526–2537, http://dx.doi.org/10.1109/JSAC.2020.3007036.

[30] Nandana Rajatheva, Italo Atzeni, Simon Bicais, Emil Bjornson, Andre Bourdoux, Stefano Buzzi, Carmen D'Andrea, Jean-Baptiste Dore, Serhat Erkucuk, Manuel Fuentes, Ke Guan, Yuzhou Hu, Xiaojing Huang, Jari Hulkkonen, Josep Miquel Jornet, Marcos Katz, Behrooz Makki, Rickard Nilsson, Erdal Panayirci, Khaled Rabie, Nuwanthika Rajapaksha, MohammadJavad Salehi, Hadi Sarieddeen, Shahriar Shahabuddin, Tommy Svensson, Oskari Tervo, Antti Tolli, Qingqing Wu, Wen Xu, Scoring the terabit/s goal:Broadband connectivity in 6G, 2021, URL https://arxiv.org/abs/2008.07220.

[31] K.M. Faisal, Wooyeol Choi, Machine learning approaches for reconfigurable intelligent surfaces: A survey, IEEE Access 10 (2022) 27343–27367.

[32] Hao Zhou, Melike Erol-Kantarci, Yuanwei Liu, H Vincent Poor, A survey on model-based, heuristic, and machine learning optimization approaches in RIS-aided wireless networks, 2023, arXiv preprint arXiv:2303.14320.

[33] Annisa Anggun Puspitasari, Byung Moo Lee, A survey on reinforcement learning for reconfigurable intelligent surfaces in wireless communications, Sensors 23 (5) (2023) 2554.

[34] Hao Zhou, Melike Erol-Kantarci, Yuanwei Liu, H. Vincent Poor, Heuristic algorithms for RIS-assisted wireless networks: Exploring heuristic-aided machine learning, 2023, arXiv preprint arXiv:2307.01205.

[35] Sree Krishna Das, Fatma Benkhelifa, Yao Sun, Hanaa Abumarshoud, Qammer H. Abbasi, Muhammad Ali Imran, Lina Mohjazi, Comprehensive review on ML-based RIS-enhanced IoT systems: basics, research progress and future challenges, Comput. Netw. 224 (2023) 109581.

[36] Zongze Li, Shuai Wang, Qingfeng Lin, Yang Li, Miaowen Wen, Yik-Chung Wu, H. Vincent Poor, Phase shift design in RIS empowered wireless networks: from optimization to AI-based methods, Network 2 (3) (2022) 398–418.

[37] Mohammad Abrar Shakil Sejan, Md Habibur Rahman, Beom-Sik Shin, Ji-Hye Oh, Young-Hwan You, Hyoung-Kyu Song, Machine learning for intelligent-reflecting-surface-based wireless communication towards 6G: A review, Sensors 22 (14) (2022) 5405.

[38] Marco Di Renzo, Fadil H. Danufane, Sergei Tretyakov, Communication models for reconfigurable intelligent surfaces: From surface electromagnetics to wireless networks optimization, Proc. IEEE 110 (9) (2022) 1164–1209.

[39] Muhammad Zain Siddiqi, Talha Mir, Reconfigurable intelligent surface-aided wireless communications: An overview, Intell. Converged Netw. 3 (1) (2022) 33–63, http://dx.doi.org/10.23919/ICN.2022.0007.

[40] Manzoor Ahmed, Abdul Wahid, Sayed Shariq Laique, Wali Ullah Khan, Asim Ihsan, Fang Xu, Symeon Chatzinotas, Zhu Han, A survey on STAR-RIS: Use cases, recent advances, and future research challenges, IEEE Internet Things J. 10 (16) (2023) 14689–14711, http://dx.doi.org/10.1109/JIOT.2023.3279357.

[41] Manzoor Ahmed, Salman Raza, Aized Amin Soofi, Feroz Khan, Wali Ullah Khan, Syed Zain Ul Abideen, Fang Xu, Zhu Han, Active reconfigurable intelligent surfaces: Expanding the frontiers of wireless communication-a survey, IEEE Commun. Surv. Tutor. (2024).

[42] Kinza Shafique, Mohammad Alhassoun, Going beyond a simple RIS: Trends and techniques paving the path of future RIS, IEEE Open J. Antennas Propag. 5 (2) (2024) 256–276, http://dx.doi.org/10.1109/OJAP.2024.3360900.

[43] Filippo Costa, Michele Borgese, Electromagnetic model of reflective intelligent surfaces, IEEE Open J. Commun. Soc. 2 (2021) 1577–1589, http://dx.doi.org/10.1109/OJCOMS.2021.3092217.

[44] Jieao Zhu, Zhongzhichao Wan, Linglong Dai, Mérouane Debbah, H. Vincent Poor, Electromagnetic information theory: Fundamentals, modeling, applications, and open problems, 2022, arXiv preprint arXiv:2212.02882.

[45] Andrea Abrardo, Alberto Toccafondi, Marco Di Renzo, Design of reconfigurable intelligent surfaces by using S-parameter multiport network theory – Optimization and full-wave validation, 2023.

[46] Matteo Nerini, Shanpu Shen, Hongyu Li, Marco Di Renzo, Bruno Clerckx, A universal framework for multiport network analysis of reconfigurable intelligent surfaces, 2023, arXiv preprint arXiv:2311.10561.

[47] Josef A. Nossek, Dominik Semmler, Michael Joham, Wolfgang Utschick, Physically consistent modelling of wireless links with reconfigurable intelligent surfaces using multiport network analysis, 2023, arXiv preprint arXiv:2308.12223.

[48] Gabriele Gradoni, Marco Di Renzo, End-to-end mutual coupling aware communication model for reconfigurable intelligent surfaces: An electromagnetic-compliant approach based on mutual impedances, IEEE Wirel. Commun. Lett. 10 (5) (2021) 938–942, http://dx.doi.org/10.1109/LWC.2021.3050826.

[49] D.M. Pozar, Microwave Engineering, John Wiley & Sons, 2011.

[50] Andrea Abrardo, Alberto Toccafondi, Marco Di Renzo, Design of reconfigurable intelligent surfaces by using s-parameter multiport network theory–optimization and full-wave validation, IEEE Trans. Wireless Commun. (2024).

[51] Philipp Del Hougne, Load impedances vs polarizabilities: On the compactness of physics-compliant models of RIS-parametrized wireless channels, in: 2024 18th European Conference on Antennas and Propagation, EuCAP, IEEE, 2024, pp. 1–4.

[52] Rashid Faqiri, Chloe Saigre-Tardif, George C. Alexandropoulos, Nir Shlezinger, Mohammadreza F. Imani, Philipp del Hougne, PhysFad: Physics-based end-to-end channel modeling of RIS-parametrized environments with adjustable fading, IEEE Trans. Wirel. Commun. 22 (1) (2023) 580–595, http://dx.doi.org/10.1109/twc.2022.3196834.

[53] Chloé Saigre-Tardif, Philipp del Hougne, Self-adaptive RISs beyond free space: Convergence of localization, sensing, and communication under rich-scattering conditions, IEEE Wirel. Commun. (ISSN: 1558-0687) 30 (1) (2023) 24–30, http://dx.doi.org/10.1109/mwc.001.2200192.

[54] Philipp del Hougne, Physics-compliant diagonal representation of beyond-diagonal RIS, 2024, arXiv preprint arXiv:2403.17222.

[55] Philipp del Hougne, RIS-parametrized rich-scattering environments: Physics-compliant models, channel estimation, and optimization, 2023, URL https://arxiv.org/abs/2311.11651.

[56] Jérôme Sol, Hugo Prod'homme, Luc Le Magoarou, Philipp Del Hougne, Experimentally realized physical-model-based frugal wave control in metasurface-programmable complex media, Nature Commun. 15 (1) (2024) 2841.

[57] Hongyu Li, Shanpu Shen, Matteo Nerini, Bruno Clerckx, Reconfigurable intelligent surfaces 2.0: Beyond diagonal phase shift matrices, 2023.

[58] Matteo Nerini, Shanpu Shen, Bruno Clerckx, Discrete-value group and fully connected architectures for beyond diagonal reconfigurable intelligent surfaces, IEEE Trans. Veh. Technol. 72 (12) (2023) 16354–16368, http://dx.doi.org/10.1109/TVT.2023.3299882.

[59] Hongyu Li, Shanpu Shen, Bruno Clerckx, Beyond diagonal reconfigurable intelligent surfaces: From transmitting and reflecting modes to single-, group-, and fully-connected architectures, IEEE Trans. Wireless Commun. 22 (4) (2023) 2311–2324, http://dx.doi.org/10.1109/TWC.2022.3210706.

[60] Ertugrul Basar, Marco Di Renzo, Julien De Rosny, Merouane Debbah, Mohamed-Slim Alouini, Rui Zhang, Wireless communications through reconfigurable intelligent surfaces, IEEE Access 7 (2019) 116753–116773, http://dx.doi.org/10.1109/ACCESS.2019.2935192.

[61] Qingqing Wu, Rui Zhang, Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network, IEEE Commun. Mag. 58 (1) (2020) 106–112, http://dx.doi.org/10.1109/MCOM.001.1900107.

[62] Shanpu Shen, Bruno Clerckx, Ross Murch, Modeling and architecture design of reconfigurable intelligent surfaces using scattering parameter network analysis, IEEE Trans. Wireless Commun. 21 (2) (2022) 1229–1243, http://dx.doi.org/10.1109/TWC.2021.3103256.

[63] Hongyu Li, Shanpu Shen, Bruno Clerckx, A dynamic grouping strategy for beyond diagonal reconfigurable intelligent surfaces with hybrid transmitting and reflecting mode, IEEE Trans. Veh. Technol. 72 (12) (2023) 16748–16753, http://dx.doi.org/10.1109/TVT.2023.3288690.

[64] Qingchao Li, Mohammed El-Hajjar, Ibrahim Hemadeh, Arman Shojaeifard, Alain A.M. Mourad, Bruno Clerckx, Lajos Hanzo, Reconfigurable intelligent surfaces relying on non-diagonal phase shift matrices, IEEE Trans. Veh. Technol. 71 (6) (2022) 6367–6383, http://dx.doi.org/10.1109/TVT.2022.3160364.

[65] Shuhang Zhang, Hongliang Zhang, Boya Di, Yunhua Tan, Zhu Han, Lingyang Song, Beyond intelligent reflecting surfaces: Reflective-transmissive metasurface aided communications for full-dimensional coverage extension, IEEE Trans. Veh. Technol. 69 (11) (2020) 13905–13909, http://dx.doi.org/10.1109/TVT.2020.3024756.

[66] Hongliang Zhang, Boya Di, Intelligent omni-surfaces: Simultaneous refraction and reflection for full-dimensional wireless communications, IEEE Commun. Surv. Tutor. 24 (4) (2022) 1997–2028, http://dx.doi.org/10.1109/COMST.2022.3202813.

[67] Jiaqi Xu, Yuanwei Liu, Xidong Mu, Joey Tianyi Zhou, Lingyang Song, H. Vincent Poor, Lajos Hanzo, Simultaneously transmitting and reflecting intelligent omni-surfaces: Modeling and implementation, IEEE Veh. Technol. Mag. 17 (2) (2022) 46–54, http://dx.doi.org/10.1109/MVT.2022.3157069.

[68] Hongliang Zhang, Shuhao Zeng, Boya Di, Yunhua Tan, Marco Di Renzo, Mérouane Debbah, Zhu Han, H. Vincent Poor, Lingyang Song, Intelligent omni-surfaces for full-dimensional wireless communications: Principles, technology, and implementation, IEEE Commun. Mag. 60 (2) (2022) 39–45, http://dx.doi.org/10.1109/MCOM.001.201097.

[69] Jiaqi Xu, Yuanwei Liu, Xidong Mu, Octavia A. Dobre, STAR-RISs: Simultaneous transmitting and reflecting reconfigurable intelligent surfaces, IEEE Commun. Lett. 25 (9) (2021) 3134–3138, http://dx.doi.org/10.1109/LCOMM.2021.3082214.

[70] Hongyu Li, Shanpu Shen, Bruno Clerckx, Beyond diagonal reconfigurable intelligent surfaces: A multi-sector mode enabling highly directional full-space wireless coverage, IEEE J. Sel. Areas Commun. 41 (8) (2023) 2446–2460, http://dx.doi.org/10.1109/JSAC.2023.3288251.

[71] Pinjun Zheng, Ruiqi Wang, Atif Shamim, Tareq Y. Al-Naffouri, Mutual coupling in RIS-aided communication: Model training and experimental validation, 2024.

[72] Hongyu Li, Shanpu Shen, Matteo Nerini, Marco Di Renzo, Bruno Clerckx, Beyond diagonal reconfigurable intelligent surfaces with mutual coupling: Modeling and optimization, IEEE Commun. Lett. 28 (4) (2024) 937–941, http://dx.doi.org/10.1109/LCOMM.2024.3361648.

[73] Ruoyan Ma, Jie Tang, Xiu Yin Zhang, Kai-Kit Wong, Jonathon A. Chambers, RIS-assisted SWIPT network for internet of everything under the electromagnetics-based communication model, IEEE Internet Things J. 11 (9) (2024) 15402–15415, http://dx.doi.org/10.1109/JIOT.2023.3347623.

[74] Özlem Tuğfe Demir, Emil Björnson, Wideband channel capacity maximization with beyond diagonal RIS reflection matrices, 2024.

[75] Hongyu Li, Matteo Nerini, Shanpu Shen, Bruno Clerckx, Wideband modeling and beamforming for beyond diagonal reconfigurable intelligent surfaces, 2024.

[76] Xin Cheng, Yan Lin, Weiping Shi, Jiayu Li, Cunhua Pan, Feng Shu, Yongpeng Wu, Jiangzhou Wang, Joint optimization for RIS-assisted wireless communications: From physical and electromagnetic perspectives, IEEE Trans. Commun. 70 (1) (2022) 606–620, http://dx.doi.org/10.1109/TCOMM.2021.3120721.

[77] Yuyan Zhou, Yang Liu, Hongyu Li, Qingqing Wu, Shanpu Shen, Bruno Clerckx, Optimizing power consumption, energy efficiency and sum-rate using beyond diagonal RIS — A unified approach, IEEE Trans. Wireless Commun. (2023) 1, http://dx.doi.org/10.1109/TWC.2023.3341262.

[78] Qingchao Li, Mohammed El-Hajjar, Ibrahim Hemadeh, Arman Shojaeifard, Lajos Hanzo, Coordinated reconfigurable intelligent surfaces: Non-diagonal group-connected design, IEEE Trans. Veh. Technol. (2024) 1–6, http://dx.doi.org/10.1109/TVT.2024.3376985.

[79] Matteo Nerini, Shanpu Shen, Hongyu Li, Bruno Clerckx, Beyond diagonal reconfigurable intelligent surfaces utilizing graph theory: Modeling, architecture design, and optimization, 2024.

[80] Arthur S. de Sena, Mehdi Rasti, Nurul H. Mahmood, Matti Latva-aho, Beyond diagonal RIS for multi-band multi-cell MIMO networks: A practical frequency-dependent model and performance analysis, 2024, arXiv preprint arXiv:2401.06475.

[81] Matteo Nerini, Shanpu Shen, Bruno Clerckx, Closed-form global optimization of beyond diagonal reconfigurable intelligent surfaces, IEEE Trans. Wireless Commun. 23 (2) (2024) 1037–1051, http://dx.doi.org/10.1109/TWC.2023.3285262.

[82] Zhuang Mao, Wei Wang, Qian Xia, Caijun Zhong, Xinhua Pan, Zhizhen Ye, Element-grouping intelligent reflecting surface: Electromagnetic-compliant model and geometry-based optimization, IEEE Trans. Wireless Commun. 21 (7) (2022) 5362–5376, http://dx.doi.org/10.1109/TWC.2021.3139611.

[83] Nemanja Stefan Perović, Le-Nam Tran, Marco Di Renzo, Mark F. Flanagan, Optimization of RIS-aided SISO systems based on a mutually coupled loaded wire dipole model, 2023.

[84] Jian Sang, Jifeng Lan, Mingyong Zhou, Boning Gao, Wankai Tang, Xiao Li, Xinping Yi, Shi Jin, Quantized phase alignment by discrete phase shifts for reconfigurable intelligent surface-assisted communication systems, 2023.

[85] Lin Cao, Haifan Yin, Li Tan, Xilong Pei, RIS with insufficient phase shifting capability: Modeling, beamforming, and experimental validations, 2023.

[86] Jiaqi Xu, Xidong Mu, Yuanwei Liu, Exploiting STAR-RISs in near-field communications, 2023.

[87] Ignacio Santamaria, Mohammad Soleymani, Eduard Jorswieck, Jesús Gutiérrez, SNR maximization in beyond diagonal RIS-assisted single and multiple antenna links, IEEE Signal Process. Lett. (2023).

[88] Xuewen Qian, Marco Di Renzo, Mutual coupling and unit cell aware optimization for reconfigurable intelligent surfaces, IEEE Wirel. Commun. Lett. 10 (6) (2021) 1183–1187.

[89] Athira Subhash, Abla Kammoun, Ahmed Elzanaty, Sheetal Kalyani, Yazan H. Al-Badarneh, Mohamed-Slim Alouini, Optimal phase shift design for fair allocation in RIS aided uplink network using statistical CSI, 2023.

[90] Andrea Abrardo, Davide Dardari, Marco Di Renzo, Xuewen Qian, MIMO interference channels assisted by reconfigurable intelligent surfaces: Mutual coupling aware sum-rate optimization based on a mutual impedance channel model, IEEE Wirel. Commun. Lett. 10 (12) (2021) 2624–2628, http://dx.doi.org/10.1109/LWC.2021.3109017.

[91] Matteo Nerini, Shanpu Shen, Bruno Clerckx, Static grouping strategy design for beyond diagonal reconfigurable intelligent surfaces, 2024.

[92] Hongyu Li, Shanpu Shen, Bruno Clerckx, Synergizing beyond diagonal reconfigurable intelligent surface and rate-splitting multiple access, IEEE Trans. Wireless Commun. (2024) 1, http://dx.doi.org/10.1109/TWC.2024.3353596.

[93] Zengrui Liu, Yang Liu, Shanpu Shen, Qingqing Wu, Qingjiang Shi, Enhancing ISAC network throughput using beyond diagonal RIS, IEEE Wirel. Commun. Lett. (2024) 1, http://dx.doi.org/10.1109/LWC.2024.3386155.

[94] Dawei Ying, David J. Love, Bertrand M. Hochwald, Sum-rate analysis for multi-user MIMO systems with user exposure constraints, IEEE Trans. Wireless Commun. 16 (11) (2017) 7376–7388, http://dx.doi.org/10.1109/TWC.2017.2748103.

[95] Placido Mursia, Sendy Phang, Vincenzo Sciancalepore, Gabriele Gradoni, Marco Di Renzo, SARIS: Scattering aware reconfigurable intelligent surface model and optimization for complex propagation channels, 2023.

[96] Bile Peng, Karl-Ludwig Besser, Shanpu Shen, Finn Siegismund-Poschmann, Ramprasad Raghunath, Daniel Mittleman, Vahid Jamali, Eduard A. Jorswieck, RISnet: A domain-knowledge driven neural network architecture for RIS optimization with mutual coupling and partial CSI, 2024.

[97] Bjorn Sihlbom, Marios I. Poulakis, Marco Di Renzo, Reconfigurable intelligent surfaces: Performance assessment through a system-level simulator, IEEE Wirel. Commun. (2022).

[98] Hanyu Jiang, Li You, Jue Wang, Wenjin Wang, Xiqi Gao, Hybrid RIS and DMA assisted multiuser MIMO uplink transmission with electromagnetic exposure constraints, IEEE J. Sel. Top. Sign. Proces. 16 (5) (2022) 1055–1069, http://dx.doi.org/10.1109/JSTSP.2022.3174701.

[99] Ruoyan Ma, Jie Tang, Xiuyin Zhang, Kai-Kit Wong, Jonathon A. Chambers, Energy efficiency optimization for mutual-coupling-aware wireless communication system based on RIS-enhanced SWIPT, IEEE Internet Things J. (2023) 1, http://dx.doi.org/10.1109/JIOT.2023.3241168.

[100] Alessio Zappone, Marco Di Renzo, Energy efficiency optimization of reconfigurable intelligent surfaces with electromagnetic field exposure constraints, IEEE Signal Process. Lett. 29 (2022) 1447–1451, http://dx.doi.org/10.1109/LSP.2022.3181532.

[101] Yongxu Zhu, Gan Zheng, Kai-Kit Wong, Stochastic geometry analysis of large intelligent surface-assisted millimeter wave networks, IEEE J. Sel. Areas Commun. 38 (8) (2020) 1749–1762, http://dx.doi.org/10.1109/JSAC.2020.3000806.

[102] Andrea De Jesus Torres, Luca Sanguinetti, Emil Björnson, Electromagnetic interference in RIS-aided communications, 2021.

[103] Sawyer D. Campbell, David Sell, Ronald P. Jenkins, Eric B. Whiting, Jonathan A. Fan, Douglas H. Werner, Review of numerical optimization techniques for meta-device design, Opt. Mater. Express 9 (4) (2019) 1842–1863.

[104] Qian Zhang, Che Liu, Xiang Wan, Lei Zhang, Shuo Liu, Yan Yang, Tie Jun Cui, Machine-learning designs of anisotropic digital coding metasurfaces, Adv. Theory Simul. 2 (2) (2019) 1800132.

[105] Zhaoyi Li, Raphaël Pestourie, Zin Lin, Steven G. Johnson, Federico Capasso, Empowering metasurfaces with inverse design: principles and applications, ACS Photonics 9 (7) (2022) 2178–2192.

[106] Dianjing Liu, Yixuan Tan, Erfan Khoram, Zongfu Yu, Training deep neural networks for the inverse design of nanophotonic structures, Acs Photonics 5 (4) (2018) 1365–1369.

[107] Sean Molesky, Zin Lin, Alexander Y. Piggott, Weiliang Jin, Jelena Vucković, Alejandro W. Rodriguez, Inverse design in nanophotonics, Nat. Photonics 12 (11) (2018) 659–670.

[108] Sunae So, Trevon Badloe, Jaebum Noh, Jorge Bravo-Abad, Junsuk Rho, Deep learning enabled inverse design in nanophotonics, Nanophotonics 9 (5) (2020) 1041–1057.

[109] Peter R. Wiecha, Arnaud Arbouet, Christian Girard, Otto L. Muskens, Deep learning in nano-photonics: inverse design and beyond, Photonics Res. 9 (5) (2021) B182–B200.

[110] Peter R. Wiecha, Alexander Yu Petrov, Patrice Genevet, Andrey Bogdanov, Inverse design of nanophotonics devices and materials, Photon. Nanostruct.: Fundam. Appl. (2022) 101084.

[111] Yurui Qu, Huanzheng Zhu, Yichen Shen, Jin Zhang, Chenning Tao, Pintu Ghosh, Min Qiu, Inverse design of an integrated-nanophotonics optical neural network, Sci. Bull. 65 (14) (2020) 1177–1183.

[112] Stefano Maci, Gabriele Minatti, Massimiliano Casaletti, Marko Bosiljevac, Metasurfing: Addressing waves on impenetrable metasurfaces, IEEE Antennas Wirel. Propag. Lett. 10 (2011) 1499–1502.

[113] Sandeep Inampudi, Hossein Mosallaei, Neural network based design of metagratings, Appl. Phys. Lett. 112 (24) (2018).

[114] Jiaqi Jiang, David Sell, Stephan Hoyer, Jason Hickey, Jianji Yang, Jonathan A Fan, Data-driven metasurface discovery, 2018, arXiv preprint arXiv:1811.12436.

[115] John Peurifoy, Yichen Shen, Li Jing, Yi Yang, Fidel Cano-Renteria, Brendan G. DeLacy, John D. Joannopoulos, Max Tegmark, Marin Soljačić, Nanophotonic particle simulation and inverse design using artificial neural networks, Sci. Adv. 4 (6) (2018) eaar4206.

[116] Tao Shan, Xiaotian Pan, Maokun Li, Shenheng Xu, Fan Yang, Coding programmable metasurfaces based on deep learning techniques, IEEE J. Emerg. Sel. Top. Circuits Syst. 10 (1) (2020) 114–125.

[117] Zhaocheng Liu, Dayu Zhu, Sean P. Rodrigues, Kyu-Tae Lee, Wenshan Cai, Generative model for the inverse design of metasurfaces, Nano Lett. 18 (10) (2018) 6570–6576.

[118] John A. Hodge, Kumar Vijay Mishra, Amir I. Zaghloul, RF metasurface array design using deep convolutional generative adversarial networks, in: 2019 IEEE International Symposium on Phased Array System & Technology, PAST, IEEE, 2019, pp. 1–6.

[119] Wei Ma, Feng Cheng, Yihao Xu, Qinlong Wen, Yongmin Liu, Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi-supervised learning strategy, Adv. Mater. 31 (35) (2019) 1901111.

[120] Jiaqi Jiang, Jonathan A. Fan, Global optimization of dielectric metasurfaces using a physics-driven neural network, Nano Lett. 19 (8) (2019) 5366–5372.

[121] Wenye Ji, Jin Chang, He-Xiu Xu, Jian Rong Gao, Simon Gröblacher, H. Paul Urbach, Aurèle J.L. Adam, Recent advances in metasurface design and quantum optics applications with machine learning, physics-informed neural networks, and topology optimization methods, Light: Sci. Appl. 12 (1) (2023) 169.

[122] Junjie Hou, Jing Jin, Hai Lin, Zixin Liu, Jiaping Fu, Feng Feng, An overview of deep learning techniques for inverse design of metasurface, in: 2023 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization, NEMO, IEEE, 2023, pp. 110–113.

[123] Zicheng Song, Ruicong Zhang, Pingping Min, Tianyu Wang, Wenxin Cao, Yurong He, Lin Wu, Jiaqi Zhu, Cheng-Wei Qiu, Inverse design of diffusion–absorption hybrid metasurfaces, Laser Photonics Rev. (2023) 2300280.

[124] Weijian Zhou, Shuoyuan Wang, Qian Wu, Xianchen Xu, Xinjing Huang, Guoliang Huang, Yang Liu, Zheng Fan, An inverse design paradigm of multi-functional elastic metasurface via data-driven machine learning, Mater. Des. 226 (2023) 111560.

[125] Kumar Vijay Mishra, Ahmet M. Elbir, Amir I. Zaghloul, Machine learning for metasurfaces design and their applications, 2022.

[126] John A. Hodge, Kumar Vijay Mishra, Amir I. Zaghloul, Deep inverse design of reconfigurable metasurfaces for future communications, 2021, arXiv preprint arXiv:2101.09131.

[127] John Adams Hodge II, Reconfigurable Intelligent Metasurfaces for Wireless Communication and Sensing Applications (Ph.D. thesis), Virginia Tech, 2022.

[128] Xin Shi, Tianshuo Qiu, Jiafu Wang, Xueqing Zhao, Shaobo Qu, Metasurface inverse design using machine learning approaches, J. Phys. D: Appl. Phys. 53 (27) (2020) 275105.

[129] Ruichao Zhu, Tianshuo Qiu, Jiafu Wang, Sai Sui, Chenglong Hao, Tonghao Liu, Yongfeng Li, Mingde Feng, Anxue Zhang, Cheng-Wei Qiu, et al., Phase-to-pattern inverse design paradigm for fast realization of functional metasurfaces via transfer learning, Nature Commun. 12 (1) (2021) 2974.

[130] Peng Xu, Jun Lou, Chenxia Li, Xufeng Jing, Inverse design of a metasurface based on a deep tandem neural network, J. Opt. Soc. Am. B 41 (2) (2024) A1–A5.

[131] Fardin Ghorbani, Javad Shabanpour, Sina Beyraghi, Hossein Soleimani, Homayoon Oraizi, Mohammad Soleimani, A deep learning approach for inverse design of the metasurface for dual-polarized waves, Appl. Phys. A 127 (11) (2021) http://dx.doi.org/10.1007/s00339-021-05030-6.

[132] Wei Peng, Jun Zhang, Weien Zhou, Xiaoyu Zhao, Wen Yao, Xiaoqian Chen, IDRLnet: A physics-informed neural network library, 2021.

[133] Prajith Pillai, Anirban Chaudhari, Parama Pal, Beena Rai, Physics-informed neural network for inversely predicting effective electric permittivities of metamaterials, in: Proceedings of the 35th Neural Information Processing Systems (NeurIPS) Machine Learning and the Physical Sciences Workshop, 2021.

[134] Tamara G. Grossmann, Urszula Julia Komorowska, Jonas Latz, Carola-Bibiane Schönlieb, Can physics-informed neural networks beat the finite element method? 2023, arXiv preprint arXiv:2302.04107.

[135] George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, Liu Yang, Physics-informed machine learning, Nat. Rev. Phys. 3 (6) (2021) 422–440.

[136] Chuizheng Meng, Sungyong Seo, Defu Cao, Sam Griesemer, Yan Liu, When physics meets machine learning: A survey of physics-informed machine learning, 2022, arXiv preprint arXiv:2203.16797.

[137] Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, Francesco Piccialli, Scientific machine learning through physics–informed neural networks: Where we are and what's next, J. Sci. Comput. 92 (3) (2022) 88.

[138] Steffen Limmer, Alberto Martinez Alba, Nicola Michailow, Physics-informed neural networks for pathloss prediction, 2022.

[139] Sulagna Sarkar, Anqi Ji, Zachary Jermain, Robert Lipton, Mark Brongersma, Kaushik Dayal, Hae Young Noh, Physics Guided Machine Learning for Inverse Design of Metamaterials, Department of Civil and Environmental Engineering, Carnegie Mellon University, 2022.

[140] Cheng-Xiang Wang, Yue Yang, Jie Huang, Xiqi Gao, Tie Jun Cui, Lajos Hanzo, Electromagnetic information theory: Fundamentals and applications for 6G wireless communication systems, 2024.

[141] Jun Yan Dai, Wankai Tang, Ming Zheng Chen, Chi Hou Chan, Qiang Cheng, Shi Jin, Tie Jun Cui, Wireless communication based on information metasurfaces, IEEE Trans. Microw. Theory Tech. 69 (3) (2021) 1493–1510.

[142] Zhongzhichao Wan, Jieao Zhu, Zijian Zhang, Linglong Dai, Chan-Byoung Chae, Mutual information for electromagnetic information theory based on random fields, IEEE Trans. Commun. 71 (4) (2023) 1982–1996.

[143] Runyu Lyu, Wenchi Cheng, Joint reflection and power splitting optimization for RIS-assisted OAM-SWIPT, in: GLOBECOM 2022 - 2022 IEEE Global Communications Conference, 2022, pp. 1073–1078, http://dx.doi.org/10.1109/GLOBECOM48099.2022.10001218.

[144] Decai Shen, Zijian Zhang, Linglong Dai, Joint beamforming design for RIS-assisted cell-free network with multi-hop transmissions, Tsinghua Sci. Technol. 28 (6) (2023) 1115–1127.

[145] Zhaohui Yang, Ye Hu, Zhaoyang Zhang, Wei Xu, Caijun Zhong, Kai-Kit Wong, Reconfigurable intelligent surface based orbital angular momentum: Architecture, opportunities, and challenges, IEEE Wirel. Commun. 28 (6) (2021) 132–137.

[146] Omar Maraqa, Sylvester Aboagye, Telex M.N. Ngatched, Optical STAR-RIS-aided VLC systems: RSMA versus NOMA, IEEE Open J. Commun. Soc. 5 (2024) 430–441, http://dx.doi.org/10.1109/OJCOMS.2023.3347534.

[147] Sylvester Aboagye, Alain R. Ndjiongue, Telex M.N. Ngatched, Octavia A. Dobre, H. Vincent Poor, RIS-assisted visible light communication systems: A tutorial, IEEE Commun. Surv. Tutor. 25 (1) (2023) 251–288, http://dx.doi.org/10.1109/COMST.2022.3225859.

[148] Jianwei You, Qian Ma, Lei Zhang, Che Liu, Jianan Zhang, Shuo Liu, Tiejun Cui, Electromagnetic metamaterials: From classical to quantum, Electromagn. Sci. 1 (1) (2023) 1–33.

[149] Qi Jian Lim, Charles Ross, Amitava Ghosh, Frederick Vook, Gabriele Gradoni, Zhen Peng, Quantum-assisted combinatorial optimization for reconfigurable intelligent surfaces in smart electromagnetic environments, IEEE Trans. Antennas and Propagation (2023).

[150] Takahiro Ohyama, Yuichi Kawamoto, Nei Kato, Resource allocation optimization by quantum computing for shared use of standalone IRS, IEEE Trans. Emerg. Top. Comput. (2023).

[151] Charles Ross, Gabriele Gradoni, Qi Jian Lim, Zhen Peng, Engineering reflective metasurfaces with ising Hamiltonian and quantum annealing, IEEE Trans. Antennas and Propagation 70 (4) (2021) 2841–2854.

[152] Charles Ross, Gabriele Gradoni, Zhen Peng, A hybrid classical-quantum computing framework for RIS-assisted wireless network, in: 2023 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization, NEMO, 2023, pp. 99–102, http://dx.doi.org/10.1109/NEMO56117.2023.10202166.

[153] Jiajia Shi, Tse-Tin Chan, Haoyuan Pan, Tat-Ming Lok, Reconfigurable intelligent surface assisted semantic communication systems, 2023, arXiv preprint arXiv:2306.09650.

[154] Peiwen Jiang, Chao-Kai Wen, Shi Jin, Geoffrey Ye Li, RIS-enhanced semantic communications adaptive to user requirements, 2023, arXiv preprint arXiv:2307.16100.

[155] Yiru Wang, Wanting Yang, Pengxin Guan, Yuping Zhao, Zehui Xiong, STAR-RIS-assisted privacy protection in semantic communication system, 2023, arXiv preprint arXiv:2306.12675.

[156] Hongyang Du, Jiacheng Wang, Dusit Niyato, Jiawen Kang, Zehui Xiong, Junshan Zhang, Xuemin Shen, Semantic communications for wireless sensing: RIS-aided encoding and self-supervised decoding, IEEE J. Sel. Areas Commun. 41 (8) (2023) 2547–2562, http://dx.doi.org/10.1109/JSAC.2023.3288231.

[157] Mengke Li, Bai Yan, Jin Zhang, Evolutionary multi-objective optimization for RIS-aided MU-MISO communication systems, Soft Comput. 27 (12) (2023) 8091–8106.

[158] Arman Azizi, Arman Farhang, RIS meets aerodynamic HAPS: A multi-objective optimization approach, 2023, arXiv preprint arXiv:2301.10682.

[159] Abdelhamed Mohamed, A. Zappone, Marco Di Renzo, Bi-objective optimization of information rate and harvested power in RIS-aided SWIPT systems, IEEE Wirel. Commun. Lett. 11 (10) (2022) 2195–2199, http://dx.doi.org/10.1109/LWC.2022.3196906.

[160] Kefeng Guo, Min Wu, Xingwang Li, Houbing Song, Neeraj Kumar, Deep reinforcement learning and NOMA-based multi-objective RIS-assisted IS-UAV-TNs: Trajectory optimization and beamforming design, IEEE Trans. Intell. Transp. Syst. 24 (9) (2023) 10197–10210, http://dx.doi.org/10.1109/TITS.2023.3267607.

[161] Jian Chen, Sujie Wang, Jie Jia, Qinghu Wang, Leyou Yang, Xingwei Wang, Multi-objective oriented resource allocation in reconfigurable intelligent surface assisted HCNs, Ad Hoc Netw. 140 (2023) 103066.

[162] Junwei Chai, Yunhui Yi, Xiandeng He, Zicheng Xing, Yuanxinyu Luo, Xingcai Zhang, Rate optimization and interference suppression in RIS-assisted MIMO systems, in: Proceedings of the 2023 10th International Conference on Wireless Communication and Sensor Networks, 2023, pp. 103–109.
[163] Rang Liu, Ming Li, Qian Liu, A. Lee Swindlehurst, Joint waveform and filter designs for STAP-SLP-based MIMO-DFRC systems, IEEE J. Sel. Areas Commun. 40 (6) (2022) 1918–1931.
[164] Matteo Nerini, Bruno Clerckx, Physically consistent modeling of stacked intelligent metasurfaces implemented with beyond diagonal RIS, IEEE Commun. Lett. (2024).
[165] Jiancheng An, Marco Di Renzo, Mérouane Debbah, Chau Yuen, Stacked intelligent metasurfaces for multiuser beamforming in the wave domain, in: ICC 2023-IEEE International Conference on Communications, IEEE, 2023, pp. 2834–2839.
[166] Jiancheng An, Chau Yuen, Chao Xu, Hongbin Li, Derrick Wing Kwan Ng, Marco Di Renzo, Mérouane Debbah, Lajos Hanzo, Stacked intelligent metasurface-aided MIMO transceiver design, IEEE Wirel. Commun. (2024).
[167] Hao Liu, Jiancheng An, Xing Jia, Shining Lin, Xianghao Yao, Lu Gan, Bruno Clerckx, Chau Yuen, Mehdi Bennis, Mérouane Debbah, Stacked intelligent metasurfaces for wireless sensing and communication: Applications and challenges, 2024, arXiv preprint arXiv:2407.03566.
[168] Guojun Huang, Jiancheng An, Zhaohui Yang, Lu Gan, Mehdi Bennis, Mérouane Debbah, Stacked intelligent metasurfaces for task-oriented semantic communications, 2024, arXiv preprint arXiv:2407.15053.

Samaneh Bidabadi received her B.Sc. degree in Electrical Engineering from Yazd University, Yazd, Iran, in 2011, and her M.Sc. degree from Isfahan University of Technology, Isfahan, Iran, in 2014. She is currently pursuing a Ph.D. in the Department of Electrical and Computer Engineering at the Université du Québec à Trois-Rivières (UQTR), Québec, Canada. Her doctoral research focus on AI-enabled resource allocation for RIS-aided communications. Her research interests include the optimization of next-generation wireless networks and AI-enabled resource allocation.

Messaoud Ahmed Ouameur received a bachelor's degree in electrical engineering from the Institute national d'electronique et d" electricit' e (INELEC),' Boumerdes, Algeria, in 1998, the M.B.A. degree from the Graduate School of International Studies, Ajou University, Suwon, South Korea, in 2000, and the master's and Ph.D. degrees (Hons.) in electrical engineering from the Universite du Qu' ebec' a Trois- ' Rivieres (UQTR), QC, Canada, in 2002 and 2006, ' respectively. He has been a Regular Professor at UQTR since 2018. His research interests include embedded real-time systems, parallel and distributed processing with applications to distributed Massive MIMO, deep learning and machine learning for communication system design, and the Internet of Things with an emphasis on end-to-end systems prototyping and edge computing.

Miloud Bagaa (Senior Member, IEEE) received the engineering, master's, and Ph.D. degrees from the University of Science and Technology Houari Boumediene, Algiers, Algeria, in 2005, 2008, and 2014, respectively. From 2009 to 2015, he was a Researcher with the Research Center on Scientific and Technical Information, Algiers. From 2015 to 2016, he was a Postdoctoral Researcher with the Norwegian University of Science and Technology, Trondheim, Norway. From 2016 to 2019, he was a Postdoctoral Researcher with Aalto University. From 2019 to 2020, he was a Senior Researcher with Aalto University. Last but not least, he was a Visiting Researcher with Aalto University and a Senior System Cloud Specialist with CSC, from October 2020 to November 2022. He is currently a Professor with the Department of Electrical and Computer Engineering, Université du Québec à Trois-Rivières (UQTR).

Daniel Massicotte (S'91–M'94–SM'08) received the B.Sc.A. and M.Sc.A. degrees in electrical engineering and industrial electronics in 1987 and 1990 respectively from the Universite du Qu ' ebec ' a Trois-Rivi ' eres (UQTR), Qc, Canada. He obtained ' the Ph.D. degree in electrical engineering in 1995 at the Ecole Polytechnique de Montr ' eal, Qc, Canada. ' In 1994, he joined the Department of Electrical and Computer Engineering, Universite du Qu ' ebec ' a Trois-Rivi ' eres, where he is currently a Full Pro- ' fessor. He is Founder of the Laboratory of Signal and Systems Integration. Since 2001, he has been Founding President and Chief Technology Officer of Axiocom Inc. He has been Head of the Industrial Electronic Research Group from 2011 to 2018 and Head of the Electrical and Computer Engineering Department, since 2014. Head of the Research Chair in Signals and Intelligence of High-Performance Systems. He received the Douglas R. Colton Medal for Research Excellence awarded by the Canadian Microelectronics Corporation, the PMC-Sierra High Speed Networking and Communication Award and the Second place at the Complex Multimedia/Telecom IP Design Contest from Europractice. His research interests include advanced VLSI implementation, digital signal processing for wireless communications, measurement, medical, real-time simulation and control problems for linear/nonlinear complex systems. He has proposed many methods based on modern signal processing, such as machine learning, transform domain, and metaheuristics. He is author/coauthor of more than 200 technical papers in international conferences and journals, as well as of 9 inventions. Dr. Massicotte is also member of the "Ordre des Ingenieurs du ' Quebec" and "Microsystems Strategic Alliance of Quebec" (ReSMiQ).

Felipe A. P. de Figueiredo, received the B.Sc. and M.Sc. degrees in telecommunication engineering from the National Institute of Telecommunications (Inatel) Brazil, in 2004 and 2011, respectively. He received his first Ph.D. degree from the State University of Campinas (UNICAMP), Brazil, in 2019 and the second one from the University of Ghent (UGhent), Belgium, in 2021. He has been working on the research and development of telecommunication systems for more than 15 years. His research interests include digital signal processing, digital communications, mobile communications, MIMO, multicarrier modulations, FPGA development, and machine learning.

Anas Chaaban (SM'17) received the Mâitrise ès Sciences degree in electronics from Lebanese University, Lebanon, in 2006, the M.Sc. degree in communications technology from the University of Ulm, Germany, in 2009, and the Dr.Ing. (Ph.D.) degree in electrical engineering and information technology from the Ruhr-University of Bochum, Germany, in 2013. From 2008 to 2009, he was with the Daimler AG Research Group on Machine Vision, Ulm, Germany. He was a Research Assistant with the Emmy-Noether Research Group on Wireless Networks, University of Ulm, from 2009 to 2011, which relocated to the Ruhr-University of Bochum in 2011. He was a Post-Doctoral Fellow with the Ruhr-University of Bochum from 2013 to 2014, and with the King Abdullah University of Science and Technology, Saudi Arabia, from 2015 to 2017. He joined the School of Engineering, The University of British Columbia as an Assistant Professor in 2018. His research interests include information theory and wireless communications with focus on relaying, multi-user communications, and interference management.